

**ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA
INFORMACION DE LA OPERACION Y ADMINISTRACION DEL MERCADO
ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRACTICO DE
ESTUDIO**

DIANA PAOLA BERNAL LOPEZ

**UNIVERSIDAD TECNOLOGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA INDUSTRIAL
CARTAGENA DE INDIAS D.T y C.**

2006

**ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA
INFORMACIÓN DE LA OPERACIÓN Y ADMINISTRACIÓN DEL MERCADO
ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRÁCTICO DE
ESTUDIO**

DIANA PAOLA BERNAL LOPEZ

**Trabajo de Grado para optar al título de
Ingeniero Industrial**

**Director
JAIME ACEVEDO CHEDID
Ingeniero Industrial**

**UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA INDUSTRIAL
CARTAGENA DE INDIAS D.T y C.**

2006

Cartagena, Abril 24 de 2006

AUTORIZACION

Yo Diana Paola Bernal López identificada con cedula de ciudadanía número 45.553.325 de Cartagena, autorizo a la Universidad Tecnológica de Bolívar a hacer uso de mi trabajo de grado titulado “ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA INFORMACION DE LA OPERACIÓN Y ADMINISTRACIÓN DEL MERCADO ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRACTICO DE ESTUDIO”, y publicarlo en el catalogo on line de la biblioteca.

DIANA PAOLA BERNAL LÓPEZ

C.C. 45.553.325 de Cartagena

Cartagena, Abril 24 de 2006

Señores

COMITÉ DE EVALUACION DE PROYECTOS
UNIVERSIDAD TECNOLÓGICA DE BOLIVAR
Facultad de Ingeniería Industrial
Ciudad.

Con la presente me permito presentar a ustedes para su estudio y consideración, el trabajo de grado titulado “ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA INFORMACION DE LA OPERACION Y ADMINISTRACIÓN DEL MERCADO ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRACTICO DE ESTUDIO”, para optar al título de Ingeniero Industrial.

Atentamente,

DIANA PAOLA BERNAL LÓPEZ
C.C 45553325 de Cartagena

Cartagena, Abril 24 de 2006

Señores

COMITÉ DE EVALUACION DE PROYECTOS
UNIVERSIDAD TECNOLOGICA DE BOLIVAR
Facultad de Ingeniería Industrial
Ciudad.

Con la presente me dirijo a ustedes para manifestarles que he dirigido el trabajo de grado titulado “ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA INFORMACION DE LA OPERACION Y ADMINISTRACION DEL MERCADO ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRACTICO DE ESTUDIO”, desarrollado por la estudiante Diana Paola Bernal López, como requisito para optar por al titulo de Ingeniero Industrial.

Atentamente,

JAIME ACEVEDO CHEDID
Ingeniero Industrial
DIRECTOR

NOTA DE ACEPTACION

PRESIDENTE DEL JURADO

JURADO

JURADO

Cartagena D.T y C., Abril de 2006

**A Dios, por que se que sin su ayuda este trabajo no seria una realidad;
a mi papá, que donde está se sentiría muy orgulloso de mí;
a mi mamá y mi hermano que son el motor de mi vida.**

CONTENIDO

INTRODUCCION

1 GENERALIDADES DE LA MINERA DE DATOS.....	26
1.1 DEFINICION	27
1.2 ORIGENES DE LA MINERIA DE DATOS.....	27
1.3 OBJETIVOS DE LA MINERIA DE DATOS.....	28
1.4 APLICACIONES DE LA MINERIA DE DATOS.....	28
1.4.1 APLICACIÓN EN EL SECTOR ELÉCTRICO	30
1.5 DESVENTAJAS DE UN PROYECTO DE MINERIA DE DATOS.....	32
1.6 TAREAS DE MINERIA DE DATOS.....	32
1.6.1 TAREAS PREDICTIVAS.....	33
1.6.1.1 Clasificación y Regresión	33
1.6.2 TAREAS DESCRIPTIVAS	33
1.6.2.1 Agrupamiento o Clustering	33
1.6.2.2 Reglas de Asociación	34
1.6.2.3 Análisis Correlacional	34
1.7 METODOS DE MINERIA DE DATOS	34
1.7.1 REDES NEURONALES.....	35
1.7.2 ÁRBOLES DE DECISIÓN.....	35
1.7.3 MÉTODO APRIORI	36
1.7.4 MÉTODOS BASADOS EN VECINDAD.....	37
1.7.4.1 K-Means	37
1.7.4.2 K-Vecinos	38
1.7.5 REDES BAYESIANAS.....	38
1.7.6 MÉTODOS DE REGRESIÓN	39
1.7.6.1 Regresión Lineal.....	39

1.7.6.2	Regresión Logística.....	39
1.7.7	ALGORITMOS GENÉTICOS.....	39
1.8	RELACION ENTRE TAREAS Y TECNICAS DE MINERIA DE DATOS	40
1.9	HERRAMIENTAS DE MINERIA DE DATOS.....	41
1.9.1	CLEMENTINE.....	42
1.9.2	WEKA.....	43
1.9.3	ENTERPRISE MINER.....	44
1.9.4	INTELLIGENT MINER.....	44
1.9.5	ODMS DARWIN.....	44
1.10	PROGRAMA DE MINERIA DE DATOS.....	45
1.10.1	METODOLOGÍA PARA LA IMPLEMENTACIÓN DE UN PROGRAMA DE MINERÍA DE DATOS: CRISP-DM.....	45
1.10.1.1	Comprensión Del Negocio.....	46
1.10.1.2	Comprensión De Los Datos.....	47
1.10.1.3	Preparación De Los Datos.....	48
1.10.1.4	Modelado.....	50
1.10.1.5	Evaluación.....	51
1.10.1.6	Despliegue.....	52
1.11	TRATAMIENTO DE LA INFORMACIÓN: CRÍTICA DE DATOS	52
1.11.1	FILTRADO, DETECCIÓN Y ELIMINACIÓN DE DATOS ANOMALOS.....	53
1.11.2	DETECCIÓN Y TRATAMIENTO DE VALORES FALTANTES.....	53
1.11.3	DETECCIÓN Y TRATAMIENTO DE VALORES ERRÓNEOS.....	55
1.11.4	DETECCION Y ELIMINACION DE VALORES DUPLICADOS.....	55
2	MINERIA DE DATOS EN XM.....	57
2.1	GENERALIDADES DE LA EMPRESA	57
2.1.1	MISIÓN.....	58
2.1.2	VISIÓN.....	58
2.1.3	AREA DONDE SE DESARROLLA EL PROYECTO.....	59
2.2	FASES DE LA MINERIA DE DATOS APLICADAS EN XM	60
2.2.1	COMPRESION DEL NEGOCIO.....	60
2.2.1.1	Necesidades del Negocio.....	60

2.2.1.2	Objetivos del Negocio.....	60
2.2.1.3	Objetivos de Minería de Datos	60
2.2.2	<i>COMPRESIÓN DE LOS DATOS</i>	61
2.2.2.1	Recopilación De Datos	61
2.2.2.2	Descripción de los Datos	61
2.2.3	<i>PREPARACIÓN DE LOS DATOS</i>	62
2.2.3.1	Selección de Datos.....	62
2.2.3.2	Limpieza de Datos	62
2.2.3.3	Construcción de Datos	62
2.2.4	<i>MODELADO</i>	64
2.2.4.1	Selección de la Técnica de Modelado	64
2.2.4.2	Selección de la Herramienta Tecnológica	64
2.2.4.3	Evaluación de Resultados	65
2.2.5	<i>EVALUACIÓN</i>	66
2.2.6	<i>DESPLIEGUE</i>	66
2.2.6.1	Planificación del Despliegue.....	66
2.3	<i>RESULTADOS Y ANALISIS DEL CASO DE ESTUDIO</i>	67
2.3.1	<i>REGLAS EXPORTADAS POR WEKA</i>	68
2.3.2	<i>RESULTADOS CASO GENERAL O CONJUNTO</i>	69
2.3.3	<i>ANÁLISIS CASO GENERAL O CONJUNTO</i>	73
2.3.4	<i>RESULTADOS CASOS INDIVIDUALES</i>	75
2.3.4.1	Resultados Departamento: ATLANTICO	75
2.3.5	<i>ANÁLISIS INDIVIDUAL</i>	85
2.3.5.1	Análisis Departamento: ATLANTICO	85
2.3.5.2	Resultados Departamento: BOLIVAR	88
2.3.5.3	Análisis Departamento: BOLIVAR	98
2.3.5.4	Resultados Departamento: CESAR.....	100
2.3.5.5	Análisis Departamento: CESAR	110
2.3.5.6	Resultados Departamento: CORDOBA.....	112
2.3.5.7	Análisis Departamento: CORDOBA	122
2.3.5.8	Resultados Departamento: LA GUAJIRA	124

2.3.5.9	Análisis Departamento: LA GUAJIRA.....	134
2.3.5.10	Resultados Departamento: MAGDALENA	136
2.3.5.11	Análisis Departamento: MAGDALENA	147
2.3.5.12	Resultados Departamento: SUCRE	149
2.3.5.13	Análisis Departamento: SUCRE	160
3	CRÍTICA DE DATOS	162
3.1	GENERACION IDEAL	164
3.1.1	INFORMACIÓN GENERAL	164
3.1.2	FORMATO DE REPORTE.....	164
3.1.3	GRÁFICAS SOPORTE.....	165
3.1.4	RESULTADOS	166
3.2	GENERACION REAL	167
3.2.1	INFORMACION GENERAL	167
3.2.2	FORMATO DE REPORTE.....	167
3.2.3	GRÁFICAS SOPORTE.....	168
3.2.4	RESULTADOS	169
3.3	PRECIO DE OFERTA	169
3.3.1	INFORMACION GENERAL	169
3.3.2	FORMATO DE REPORTE.....	170
3.3.3	RESULTADOS	171
3.4	DISPONIBILIDAD DE GENERACION	172
3.4.1	INFORMACION GENERAL	172
3.4.2	FORMATO DE REPORTE.....	172
3.4.3	GRAFICAS SOPORTE.....	173
3.4.4	RESULTADOS	174
3.5	CAPACIDAD REMUNERABLE REAL	175
3.5.1	INFORMACION GENERAL	175
3.5.2	FORMATOS DE REPORTE	175
3.5.3	GRÁFICAS SOPORTE.....	176
3.5.4	RESULTADOS	177
4	CONCLUSIONES Y RECOMENDACIONES.....	178
	BIBLIOGRAFIA	

LISTA DE CUADROS

Cuadro 1. Casos exitosos en el sector eléctrico.....	30
Cuadro 2. Tareas y técnicas de minería de datos	40
Cuadro 3. Estadísticas caso general.....	69
Cuadro 4. Distribución de Frecuencia caso general.....	69
Cuadro 5. Estadísticas Atlántico.....	75
Cuadro 6. Distribución de frecuencia Atlántico	75
Cuadro 7. Reglas Atlántico.....	78
Cuadro 8. Estadísticas Bolívar	88
Cuadro 9. Distribución de frecuencia Bolívar	88
Cuadro 10. Reglas Bolívar	91
Cuadro11. Estadísticas Cesar.....	100
Cuadro 12. Distribución de frecuencia Cesar	100
Cuadro 13. Reglas Cesar	103
Cuadro 14. Estadísticas Córdoba.....	112
Cuadro 15. Distribución de frecuencia Córdoba.....	112
Cuadro 16. Reglas Córdoba.....	115
Cuadro 17. Estadísticas La Guajira.....	124
Cuadro 19. Reglas La Guajira	127
Cuadro 20. Estadísticas Magdalena.....	136
Cuadro 21. Distribución de frecuencia Magdalena.....	136
Cuadro 22. Reglas Magdalena.....	139
Cuadro 23. Estadísticas Sucre	149
Cuadro 24. Distribución de frecuencia Sucre	149
Cuadro 26. Variables más consultadas en NEON.....	163
Cuadro 27. Reporte Generación Ideal.....	164

Cuadro 28. Reporte Generación Real	167
Cuadro 29. Reporte Precio de Oferta.....	170
Cuadro 30. Reporte Disponibilidad de Generación	172
Cuadro 31. Reporte Capacidad Remunerable Real	175

LISTA DE FIGURAS

Figura 1. Histograma de frecuencia caso general.....	71
Figura 2. Demanda departamentos diferenciada por periodos del día caso general	72
Figura 3. Demanda días de la semana diferenciada por departamentos caso general	72
Figura 4. Histograma de frecuencia (Atlántico)	77
Figura 5. Demanda días de la semana diferenciada por periodos del día (Atlántico)	84
Figura 6. Demanda meses del año diferenciada por días de la semana (Atlántico)	84
Figura 7. Demanda periodos del día (Atlántico)	85
Figura 8. Histograma de frecuencia (Bolívar)	90
Figura 9. Demanda días de la semana diferenciada por periodos del día (Bolívar)	97
Figura 10. Demanda meses del año diferenciada por periodos del día (Bolívar)	97
Figura 11. Demanda periodos del día (Bolívar)	98
Figura 12. Histograma de frecuencia (Cesar).....	102
Figura 13. Demanda días de la semana diferenciada por periodos del día (Cesar)	109
Figura 14. Demanda meses del año diferenciada por días de la semana (Cesar)	109
Figura 15. Demanda periodos del día (Cesar)	110
Figura 16. Histograma de frecuencia (Córdoba)	114
Figura 17. Demanda días de la semana diferenciada por periodos del día (Córdoba)	121

Figura 18. Demanda meses del año diferenciada por día de la semana (Córdoba)	121
Figura 19. Demanda periodos del día (Córdoba)	122
Figura 20. Histograma de frecuencia (La Guajira)	126
Figura 21. Demanda días de la semana diferenciada por periodos del día (La Guajira)	133
Figura 22. Demanda meses del año diferenciada por días de la semana (La Guajira)	133
Figura 23. Demanda periodos del día (La Guajira)	134
Figura 24. Histograma de frecuencia (Magdalena)	138
Figura 25. Demanda días de la semana diferenciada por periodos del día (Magdalena)	145
Figura 26. Demanda meses del año diferenciada por días de la semana (Magdalena)	145
Figura 27. Demanda periodos del día (Magdalena)	147
Figura 28. Histograma de frecuencia (Sucre)	151
Figura 29. Demanda días de la semana diferenciada por periodos del día (Sucre)	159
Figura 30. Demanda meses del año diferenciada por días de la semana (Sucre)	159
Figura 31. Demanda periodos del día (Sucre)	160
Figura 32. Generación Ideal mensual	165
Figura 33. Generación Ideal diaria	165
Figura 34. Generación Ideal horaria	165
Figura 35. Generación Real mensual	168
Figura 37. Generación Real horaria	168
Figura 38. Disponibilidad de Generación para la entidad Sistema	173
Figura 39. Disponibilidad de Generación para la entidad Central (LNN4-MNOR)	173
Figura 40. Capacidad Remunerable Real para la entidad Sistema	176
Figura 41. Capacidad Remunerable Real para la entidad Central (ALAG-BLL3)	176
Figura 42. Capacidad Remunerable Real para la entidad Central (CDF1-CHVR)	176

GLOSARIO

ALGORITMOS: generalmente proceden del campo del aprendizaje de máquinas, puesto que en estos realizamos una búsqueda (exhaustiva o heurística) con la cual tratamos de encontrar patrones que expliquen los datos introducidos.

ANÁLISIS EXPLORATORIO DE DATOS: uso de técnicas estadísticas tanto gráficas como descriptivas para aprender acerca de la estructura de un conjunto de datos.

ANÁLISIS MULTIDIMENSIONAL: es una técnica que permite ver la información corporativa de diferentes puntos de vista y las relaciones entre la misma. Permite tener diferentes perspectivas de los datos relacionados con los conceptos principales del plan de negocios.

ANÁLISIS PROSPECTIVO DE DATOS: análisis de datos que predice futuras tendencias, comportamientos o eventos basado en datos históricos.

ANÁLISIS RETROSPECTIVO DE DATOS: análisis de datos que provee una visión de las tendencias, comportamientos o eventos basado en datos históricos.

ARCHIVO ARFF: tipo de archivo plano, usado específicamente en la herramienta WEKA.

ARCHIVO CSV: tipo archivo plano separado por comas o tabuladores.

ARCHIVO PLANO: base de datos en formato ASCII que separa los registros mediante un carácter especial.

ARQUITECTURA DE INFORMACIÓN: el estudio de la organización de la información con el objetivo de permitir al usuario encontrar su vía de navegación hacia el conocimiento y la comprensión de la información.

ATRIBUTOS, CAMPOS O DIMENSIONES: variables correspondientes a las columnas de la tabla de la base de datos y observaciones a cada una de las filas de las mismas. Los atributos pueden ser continuos (numéricos) o discretos (cualitativos o nominales).

COEFICIENTE DE CORRELACIÓN: es una medida de dependencia estadística lineal entre dos características o variables. Es un número r que se encuentra en el intervalo $[-1, 1]$. Un valor cerca de uno indica que a medida que el valor de una variable aumenta, el valor de la otra también tiende a aumentar. Si el valor está cerca de -1 a medida que el valor de una variable aumenta, el valor de la otra variable tiende a disminuir. Una correlación de cero indica que no hay dependencia lineal estadística entre las dos variables, aunque no indica que las variables sean independientes (excepto en el caso en que las variables tengan distribución normal). Una correlación distinta de cero no es evidencia suficiente para concluir que hay una relación de causa y efecto entre las variables.

CONOCIMIENTO: capacidad de convertir datos e información en acciones efectivas.

CRÍTICA DE DATOS: proceso evaluativo que consiste en identificar la presencia de ruido en la información, originado por datos anómalos o outliers, valores duplicados, valores faltantes, datos erróneos y demás; que ocasionan problemas de calidad en los sistemas de información.

DATOS: medidas, cifras, nombres o hechos. Los datos de entrada a un proceso de minería de datos son típicamente tablas correspondientes a una base de datos o archivos planos.

DATOS ANÓMALOS O OUTLIERS: un ítem de datos cuyo valor cae fuera de los límites que encierran a la mayoría del resto de los valores correspondientes a la muestra.

DATOS ERRÓNEOS: datos que resultan de errores (por ejemplo: errores en el tipeado durante la carga).

DATOS FALTANTES: información no disponible.

DEMO: demostración. En la terminología de Internet, una presentación multimedia no interactiva.

DESVIACIÓN ESTÁNDAR: es una medida de distancia promedio de los valores observados a su media. La distancia de cada valor a la media se mide tomando el cuadrado de la diferencia entre ese valor y la media. Luego de obtener el promedio de esos cuadrados, tomamos la raíz cuadrada.

ESTADÍSTICA: esta disciplina ha proporcionado muchos de los conceptos, algoritmos o técnicas que se utilizan en minería de datos, como por ejemplo, la media, la varianza, las distribuciones, la regresión lineal y no lineal, las redes bayesianas y un largo etcétera. De hecho, algunos paquetes de análisis estadístico se comercializan como herramientas de minería de datos.

GESTIÓN DE CONOCIMIENTO: proceso de identificar, adquirir, usar y reutilizar información y conocimientos, tanto externos como internos, tácitos o explícitos, con el objetivo de mejorar la eficiencia y la eficacia de la empresa, para conservar y mejorar sus ventajas competitivas.

GESTIÓN DE LA INFORMACIÓN: conjunto de actividades realizadas con el fin de controlar, almacenar y, posteriormente, recuperar adecuadamente la información producida, recibida o retenida por cualquier organización en el desarrollo de sus actividades.

INFORMACIÓN: datos procesados u organizados.

JAVA: lenguaje de programación orientado a la elaboración de aplicaciones exportables a la Red. Puede ser usado para crear aplicaciones que funcionan tanto en un solo ordenador como en una distribución de ordenadores a partir de un servidor y sus usuarios de red.

MEDIA: computacionalmente es la suma de todos los valores observados dividida por el número de observaciones. Representa el centro de masa de los datos, un punto de equilibrio.

MODELO: descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas.

MODELO ANALÍTICO: una estructura y proceso para analizar un conjunto de datos.

MODELO LINEAL: un modelo analítico que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).

MODELO NO LINEAL: un modelo analítico que no asume una relación lineal en los coeficientes de las variables que son estudiadas.

MODELO PREDICTIVO: estructura y proceso para predecir valores de variables especificadas en un conjunto de datos

PARÁMETRO: es una característica medible que describe una población. Su valor es usualmente desconocido.

PATRONES: los patrones constituyen el resultado final del proceso de minería de datos y vienen a ser relaciones entre subconjuntos de los datos pasados como entrada.

PLATAFORMA: término general que normalmente designa una arquitectura de hardware, aunque a veces también se utiliza para sistemas operativos o para ambas cosas.

SISTEMA DE INFORMACIÓN: un sistema, típicamente computerizado, que permite la captura, almacenamiento, comprobación, manipulación, integración y/o análisis de datos.

SISTEMAS PARA LA TOMA DE DECISIÓN: son herramientas y sistemas informatizados que asisten a los directivos en la resolución de problemas y en la toma de decisiones. El objetivo es proporcionar la información necesaria para realizar decisiones efectivas en el ámbito empresarial o en tareas de diagnóstico.

SQL: lenguaje de consulta y programación de bases de datos utilizado para acceder a los datos y para consultar, actualizar y gestionar sistemas de bases de datos relacionales.

TECNOLOGÍA DE LA INFORMACIÓN: conjunto de sistemas físicos y lógicos especializados en el tratamiento automático de datos e información. Se diferencia y completa el concepto informática al extender su actividad al tratamiento de todo tipo de información, independientemente de su formato, basta con que este digitalizado.

TENDENCIA: es el comportamiento que presentan una serie de datos en base a un valor definido de tomar una posición muy estrecha con éste valor.

TOMA DE DECISIONES: se busca cumplir con la meta de solución de un problema. La toma de decisión se debe basar en información confiable y debe reflejar criterios o estándares que puedan ser comunicados públicamente y que puedan ser implementados.

TUPLAS, REGISTROS, INSTANCIAS, EJEMPLOS O FILAS: cada tupla representa un objeto, el cual se describe a través de los valores de sus atributos y se caracteriza por poseer una clave única o primaria que lo identifica.

VARIABLE: es una magnitud cuyos valores son objeto de estudio en la investigación.

VISTA MINABLE: tabla fruto de la combinación de varias tablas con la información que se requiere para cada tarea concreta de minería de datos.

VISUALIZACIÓN: interfaz propia de herramientas de minería de datos que permite a través de una construcción visual identificar situaciones en los datos.

RESUMEN

Título: ESTUDIO Y APLICACION DE LA MINERIA DE DATOS EN EL ANALISIS DE LA INFORMACIÓN DE LA OPERACIÓN Y ADMINISTRACIÓN DEL MERCADO ELECTRICO COLOMBIANO EN XM, A TRAVES DE UN CASO PRACTICO DE ESTUDIO.

Autor: Diana Paola Bernal López.

Objetivo General: Realizar un estudio e implementación de la metodología de Minería de Datos a la información de la operación y administración del mercado eléctrico colombiano, a través del uso de herramientas tecnológicas existentes en el mercado: WEKA y Clementine, que permitan descubrir patrones, modelos de comportamientos y relaciones ocultas entre los datos con un caso práctico y adicionalmente realizar una crítica de datos que sirva para detectar ruido en la información.

Metodología: Para la elaboración del proyecto se hizo necesario en primera instancia realizar una investigación profunda y exhaustiva alrededor de la Minería de Datos, a través de la consulta de diversas y variadas fuentes.

En la aplicación práctica para el caso de estudio de la demanda de energía eléctrica del sector minero en la Costa Atlántica, se utilizó la herramienta WEKA de libre distribución, a la que es posible tener acceso a través de la red; se siguió a demás un programa organizado de Minería de Datos con sus respectivas fases o etapas de desarrollo, según la metodología CRISP-DM.

Finalmente se solicitó una licencia de prueba (demo) a la compañía SPSS (multinacional de software de computadoras que proporciona tecnología), de la solución de minería de datos que ellos proveen: Clementine; con la cual se llevó a cabo la crítica de datos, aprovechando su capacidad en cuanto al manejo de gran cantidad de datos y las múltiples modelos con que trabaja para la limpieza y tratamiento de los mismos.

Resultados: A partir del desarrollo del presente trabajo de grado fueron obtenidos en términos generales los siguientes resultados:

Amplia sustentación teórica sobre el tema; que sirve de consulta para futuros trabajos referentes.

Estudio del comportamiento de la demanda de energía eléctrica del sector Minero en todos (se obtuvieron resultados para el sector de manera general) y cada uno de los departamentos de la Costa Atlántica (se logró analizar el sector para los departamentos de manera individual), con el descubrimiento de patrones y relaciones entre las variables implicadas, como también la hora, día de la semana y mes del año, durante el cual se presentan los valores mínimos o máximos de demandas del sector en cada uno de los respectivos departamentos.

Diagnóstico de la calidad de los datos almacenados en NEON, a través de la crítica de datos realizada a las cinco variables más consultadas, detectando en cada una de ellas numerosos y diversos causales de ruido, como el caso de información no disponible en un rango no inferior a un mes, valores erróneos como el caso de precios de ofertas y generación ideal igual a cero, precios de oferta relacionados a plantas no despachadas centralmente; entre muchos otras inconsistencias causales de ruido.

Director: Jaime Acevedo Chedid

INTRODUCCION

Con la denominada sociedad de la información se está produciendo un fenómeno curioso: día a día se multiplica la cantidad de datos almacenados. Sin embargo, contrariamente a lo que pudiera esperarse, esta explosión de datos no supone un aumento del conocimiento, puesto que resulta imposible procesarlos con los métodos clásicos.

Hace falta un tipo de análisis automatizado, que complemente al análisis manual realizado por personas expertas. Es decir, un sistema donde el computador proponga regiones de la base de datos dónde buscar valiosa información para el negocio; los criterios de interés se propongan automáticamente o estén predefinidos por el computador, y/o modificados por el usuario; las búsquedas se vayan efectuando poco a poco, y se vayan recopilando aquellos datos interesantes que resultaron usando criterios predefinidos en la máquina, o mejor aún, modificados por el usuario, a efecto de que el usuario pueda verlos y determinar por sí mismo qué tan valiosos o rutinarios en realidad son.

El enfoque de la investigación está orientado a la Minería de Datos, tecnología que intenta ayudar a comprender el contenido de una base de datos; se expondrán los resultados de un caso práctico de estudio sobre la información del mercado eléctrico colombiano que permitió la implementación de dicha metodología. Este caso consistió básicamente en analizar la información de la demanda de energía del sector Minero en los departamentos de la Costa Atlántica, con el fin de conocer los patrones de comportamiento y tendencias alrededor de todas las variables relacionadas y su interrelación.

El interés de la organización por este estudio en especial, se asocia fundamentalmente al alto crecimiento que ha tenido la demanda de energía durante los últimos años en la Costa Atlántica, y el peso relevante del sector Minero en la demanda nacional y en estos departamentos de manera especial.

Para este caso, se analizó la información a través del uso de WEKA, solución de Minería de Datos de libre distribución, suficientemente completa (es de hecho, la que tiene la mayor variedad de métodos).

A demás del caso práctico, se darán a conocer los resultados de la crítica de datos realizada a las cinco variables más consultadas en NEON: servicio de información sobre el Mercado de Energía Mayorista colombiano. Este servicio está soportado en la más grande bodega de datos del sector eléctrico; pero que al igual que otras fuentes de datos, sufre problemas como el ruido en la información; ocasionado por la presencia de datos anómalos, fuera de rango, erróneos, información no disponible, valores duplicados y demás.

Debido a que este aplicativo es de uso libre para todos los agentes del sector (generadores, transportadores, distribuidores y comercializadores) y partes interesadas (inversionistas, entes gubernamentales, entidades nacionales e internacionales) con licencia de acceso al mismo; la información aquí accedida debe ser oportuna y consistente con la realidad; por esta razón nació la necesidad de una crítica de datos que permitiera identificar estos causales de ruido, y así la organización poder proceder a los correctivos a que dieran lugar.

Para esta fase de la investigación se utilizó la herramienta Clementine, uno de los sistemas de minería de datos más populares del mercado; a través de una licencia limitada temporalmente (demo); accedida mediante el contacto con sus proveedores.

1 GENERALIDADES DE LA MINERA DE DATOS

Cada vez más, y gracias al desarrollo de las comunicaciones y la implementación y mejora de las redes informáticas; la cantidad de información que fluye en las empresas y la capacidad de acceso a la misma ha aumentado considerablemente, pero en cambio, cada vez se tiene menos tiempo y capacidad para asimilarla.

Muchas veces, las empresas no saben obtener información valiosa de toda la cantidad ingente de datos que tienen almacenados, aunque intuyen que el conocimiento que podrían extraer de ellos podría ser de gran ayuda en muchas de las áreas y facetas en que se desenvuelven. La competitividad de las empresas, y por lo tanto su supervivencia, depende de que este conocimiento pueda preservarse y utilizarse de forma eficiente.

Lógicamente, el tener un aceptable grado de automatización y Data Warehouse es requisito indispensable, ya que si no se dispone de la infraestructura necesaria para capturar y almacenar convenientemente la información, difícilmente se podrá obtener nada de ella.

Las herramientas de minería de datos son útiles en este momento, cuando ya se tiene un volumen de información importante y de buena calidad.

Los campos de aplicación de estas nuevas técnicas dentro de la industria son numerosas: control de calidad, identificación de sistemas, determinación de causas en fallos del proceso, detección de anomalías, etc.

1.1 DEFINICION

La Minería de Datos se crea por el reconocimiento de un nuevo potencial: el valor incalculable de la gran cantidad de datos almacenados en los sistemas de información de instituciones, empresas, gobiernos y particulares. Los datos pasan de ser un “producto” (el resultado histórico de los sistemas de información) a ser una “materia prima” que hay que explotar para obtener el verdadero “producto elaborado”: el conocimiento; un conocimiento que ha de ser especialmente valioso para la ayuda en la toma de decisiones sobre el ámbito en el que se han recopilado o extraído los datos.

En términos generales la minería de datos permite convertir los datos, recogidos durante el funcionamiento normal de la organización, en información valiosa, mediante la extracción de patrones, tendencias, modelos de comportamiento, relaciones entre diversas variables y regularidades que permiten describir y comprender dichos datos, ayudando a gestionar mejor el negocio o proceso.

1.2 ORIGENES DE LA MINERIA DE DATOS

A principios de la década de los ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros estadísticos e investigadores, empezaron a consolidar los términos de minería de datos, por las similitudes que existen entre buscar valiosa información de negocio en grandes bases de datos y minar una montaña para encontrar una veta de metales preciosos. Ambos procesos requieren examinar inteligentemente una inmensidad de material hasta encontrar algo que pueda resultar útil y valioso. A finales de los años ochenta sólo existían un par de empresas dedicadas a esta tecnología; al 2002 existían más de 100 empresas en el mundo que ofrecen alrededor de 300 soluciones.

Las listas de discusión sobre este tema las forman investigadores de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios.

1.3 OBJETIVOS DE LA MINERIA DE DATOS

Los objetivos alrededor de minería de datos consisten en extraer patrones, descubrir tendencias y regularidades, predecir comportamientos y en general, de sacar partido a la información computarizada que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades; permite a demás a los individuos y a las organizaciones comprender y modelar de una manera más eficiente y precisa el contexto en el que deben actuar y tomar sus decisiones. Estos objetivos expuestos pueden ser resumidos en dos únicos objetivos considerados de alto nivel:

Predicción: Consiste en utilizar algunas variables o campos de la base de datos para predecir valores desconocidos o futuros de otras variables de interés.

Descripción: Se centra en encontrar patrones interpretables por el ser humano, a partir de descripción de los datos.

1.4 APLICACIONES DE LA MINERIA DE DATOS

Actualmente existe un número creciente de organizaciones inmersas en proyectos de minería de datos. La tecnología se puede aplicar a cualquier organización que disponga de volúmenes de datos y que se proponga explotarlos para obtener reglas de negocio, mejorar sus procesos o el servicio que presta. Los siguientes

son sólo algunos ejemplos en los cuales las técnicas de minería de datos son apropiadas:

Marketing: Este es uno de los campos donde los éxitos de la minería de datos son más conocidos. El objetivo fundamental puede resumirse en determinar quien compra que, cuando y donde.

Predicción: A partir de los datos históricos almacenados y utilizando técnicas de minería de datos pueden elaborarse modelos que permitan estimar con precisión la evolución de una variable en el futuro. Esto puede resultar útil en: la detección de oportunidades, prevención de problemas, gestión óptima del personal, optimización de stocks, etc.

Reducción de Riesgos: La minería de datos permite construir sistemas de evaluación automática de riesgos, basados en la experiencia previa. Estos sistemas resultan de gran utilidad cuando la cantidad de casos a evaluar es excesiva para su procesamiento manual. Un ejemplo claro es el del riesgo crediticio, para el sector bancario.

Detección de Fraudes: Aplicando técnicas de minería de datos, pueden obtenerse modelos que permitan descubrir posibles fraudes, basándose en la detección de comportamientos anómalos, en comparación con los datos registrados anteriormente. Pueden ser detectados casos de fraudes como las transacciones fraudulentas realizadas con tarjeta de crédito, fraudes en distribuidores de una empresa multinacional, etc.

Procesos Industriales: Las técnicas de minería de datos permiten explotar la información disponible sobre un sistema o proceso y utilizar los modelos desarrollados para: automatizar y optimizar el control del proceso, optimizar su rendimiento, implementar programas de mantenimiento predictivo, etc.

Control de Calidad: Existen numerosos ejemplos en los que se han aplicado técnicas de minería de datos para desarrollar sistemas automáticos de control de calidad. Estos sistemas suponen un considerable ahorro en el proceso productivo, puesto que facilitan: la detección precisa de productos defectuosos, localización precoz de defectos, identificación de causa de fallos, análisis no destructivo, etc.

Astronomía: Clasificación de cuerpos celestes, aspectos climatológicos como la predicción de tormentas, etc.

Medicina: Caracterización y predicción de enfermedades, probabilidad de respuesta satisfactoria a tratamiento médico, entre otras aplicaciones.

1.4.1 APLICACIÓN EN EL SECTOR ELÉCTRICO

A continuación, en el Cuadro 1 se presentan algunos casos exitosos dentro del sector eléctrico mundial en el uso de herramientas de minería de datos.

Cuadro 1. Casos exitosos en el sector eléctrico

COMPañIA	DESAFIO	SOLUCION EN MINERIA DE DATOS
ISO NEW ENGLAND	Monitorear el mercado de energía mayorista de Nueva Inglaterra para analizar prácticas de ofertas anticompetitivas.	Provee las herramientas de análisis y reportes necesarias para encontrar actividades inusuales en tiempo real.
Copenhague Energy	Realizar un pronóstico ajustado de demanda para minimizar las pérdidas.	Duplica la precisión del pronóstico de demanda, ofrece ahorros hasta del 50% y mantiene un servicio de energía confiable y estable al usuario final.

Salt River Project	Incrementar el valor sabiendo cuando vender el exceso de electricidad por un mejor precio.	Ayuda a agrupar datos dispersos para convertirlos en inteligencia que alimenta modelos sofisticados de pronóstico que proveen resultados precios y a tiempo.
Dominion	Incrementar las ganancias y manejar el riesgo en el volátil mercado de energía.	Da a los ejecutivos y negociadores de Dominion la información que ellos necesitan para tomar las decisiones basadas en conocimiento dentro de sus límites de riesgo.
Reliant Energy	Empoderar a los decisores con la información necesaria para responder a la atención de las demandas de los consumidores de energía.	Poderosa consulta y análisis funcional sobre la web.
Entergy Solutions	Anticipar el riesgo para crecer la base de clientes y obtener ganancias en mercados deregulados de energía al por menor.	Ayuda a hacer ofertas en el mercado al por menor, tomar medidas y monitorear el riesgo sistemáticamente para todas las actividades.
Térmicas del Besós S.A IBERDROLA Premier Power Ltd.	Control y optimización de centrales térmicas: ❖ Térmica de TERBESA ❖ Central de Velilla del Rio Carrión ❖ Térmica de Ballylumford (Irlanda del Norte).	Desarrollo de un sistema de optimización/control de procesos complejos. El sistema ha sido aplicado con éxito en varias centrales térmicas, consiguiendo una mejora superior al 2% en el consumo de combustible.

1.5 DESVENTAJAS DE UN PROYECTO DE MINERÍA DE DATOS

- ❖ El número de posibles relaciones a descubrir durante el análisis es demasiado grande, y resulta prácticamente imposible validar cada una de ellas.
- ❖ La mayoría de herramientas funcionan mejor fijándoles objetivos de búsqueda concretos. Si bien la minería de datos da la impresión de que se puede simplemente aplicar como herramienta a los datos, se debe tener un objetivo, o al menos una idea general de lo que busca.
- ❖ El coste de esta prospección de datos debe ser coherente con el beneficio esperado.
- ❖ Es necesario trabajar en estrecha colaboración con expertos en el negocio para definir modelos. Su ausencia y/o disponibilidad marca el futuro y éxito del proyecto.
- ❖ Otro problema es que la información muchas veces está corrompida, tiene ruido, o simplemente le faltan datos. Los que ocasiona resultados equívocos, que podrían traducirse en consecuencias para la empresa.

1.6 TAREAS DE MINERÍA DE DATOS

Una tarea de minería de datos no es más que un problema de minería de datos a ser resuelto por un algoritmo.

Por ejemplo, “clasificar las piezas del proveedor X en óptimas, defectuosas reparables y defectuosas irreparables” es una tarea; concretamente una tarea predictiva de clasificación.

De acuerdo con los objetivos de alto nivel que definimos anteriormente, las tareas de minería de datos se dividen en dos grupos según el tipo de objetivo al que estén apuntando: predictivas y descriptivas.

1.6.1 TAREAS PREDICTIVAS

Se trata de tareas en las que hay que predecir uno o más valores (clase, categoría o valor numérico).

1.6.1.1 Clasificación y Regresión

La regresión y la clasificación son las dos tareas más frecuentes en minería de datos. Hacen referencia al proceso por el cual tratamos de averiguar el valor de un determinado atributo a partir del valor de otros campos. Cuando el atributo que estamos considerando toma valores continuos, este proceso recibe el nombre de regresión, mientras que si son discretos se le conoce como clasificación.

1.6.2 TAREAS DESCRIPTIVAS

El objetivo de este tipo de tareas no es predecir nuevos datos como el caso de las tareas anteriores, sino describir el comportamiento de los datos ya existentes.

1.6.2.1 Agrupamiento o Clustering

Mediante esta tarea tratamos de identificar una serie de grupos o “clusters” en los datos, donde un “cluster” es una colección de datos similares; de manera que aquellos datos que pertenezcan a un “cluster” sean muy parecidos entre sí, pero

que difieran en gran medida con aquellos datos que se encuentran en otro “cluster”. Esta división se hace en base al concepto de distancia entre datos.

1.6.2.2 Reglas de Asociación

Una regla de asociación es una regla que implica ciertas relaciones de asociación entre distintos objetos o atributos categóricos de una base de datos, como puede ser: “ocurren juntos” o “uno implica el otro”.

Las reglas de asociación, se derivan de un tipo de análisis que extrae información por coincidencias. Este análisis permite descubrir correlaciones o concurrencias en los sucesos de los datos a analizar y se formaliza en la obtención de reglas de tipo: “Si el atributo X toma el valor de D entonces el atributo Y toma el valor de B”.

1.6.2.3 Análisis Correlacional

Se usa para examinar el grado de similitud entre dos variables numéricas. Una formula estándar para medir la correlación lineal es el coeficiente de correlación r , el cual es un valor real comprendido entre -1 y 1 .

1.7 METODOS DE MINERIA DE DATOS

Técnicas o algoritmos que permiten resolver tareas de minería de datos.

Cada una de las tareas descritas en la sección anterior requiere de métodos, técnicas o algoritmos para resolverlas. Una tarea puede tener muchos métodos diferentes para resolverla, y a la vez el mismo método puede resolver un gran abanico de tareas.

1.7.1 REDES NEURONALES

Las redes neuronales representan una forma especial de procesamiento de la información debido a que imitan la capacidad de la mente humana para encontrar patrones. Las redes neuronales, llamadas "redes de neuronas artificiales", son modelos bastante simplificados de las redes de neuronas que forman el cerebro, y al igual que éste, intentan "aprender" a partir de los datos que se le suministran.

Su ventaja principal es que, cuando están bien ajustadas, obtienen precisiones muy altas. Además son muy expresivas. Entre sus inconvenientes se suele mencionar su sensibilidad a valores anómalos, necesitan muchos ejemplos para el aprendizaje, son relativamente lentas y fundamentalmente su incomprendibilidad.

Pueden utilizarse para el reconocimiento de patrones, la comprensión de información y la reducción de la dimensionalidad, las tareas de agrupamiento, clasificación y regresión, la visualización, etc.

1.7.2 ÁRBOLES DE DECISIÓN

Los árboles de decisión son unos de los algoritmos clasificadores más conocidos y usados en las tareas de minería de datos ya que son una forma de representación sencilla para clasificar ejemplos de un número finito de clases. Se basan en la partición del conjunto de ejemplos según ciertas condiciones que se aplican a los valores de las características.

Los árboles están compuestos de nodos que nunca se unen en bucle cerrado. Se suele distinguir entre nodos terminales (pertenecientes al último nivel), nodo raíz (inicial) y nodos intermedios o no terminales. También se establecen relaciones entre ellos, de modo que se denomina padre de un nodo a su antecesor y descendiente a cualquiera de los que dependen de él.

Cada nodo tiene asociada una regla, de modo que se sigue hacia uno u otro descendiente en función de la respuesta a ésta, hasta que se llega a un nodo terminal, donde se efectúa realmente la clasificación.

Generalmente en cada nodo se evalúa cada atributo según una constante, aunque algunos árboles clasifican atributos con valores nominales, comparan dos o más atributos con otro, o utilizan alguna función que incluye uno o varios atributos.

De esta forma, si el atributo que se chequea es nominal, el número de hijos que salen de ese nodo será igual al de posibles valores de esa variable. En cambio, si el valor del mismo es continuo, lo más usual es que se divida en dos caminos: si es menor o igual que una constante, o si es mayor; aunque eso no signifique que se utilicen otros criterios de decisión que generen, en cada nodo, tres o más ramas.

1.7.3 MÉTODO APRIORI

El objetivo en todo algoritmo de búsqueda de reglas de asociación es encontrar todas las reglas que satisfagan las medidas de calidad dadas por la condición de confianza (porcentaje de veces que la regla se cumple cuando se puede aplicar) y soporte (número de instancias que la regla predice correctamente) mínimos, estipuladas previamente por el usuario, esto es necesario por cuanto sino es así la búsqueda se haría exhaustiva, encontrándose al final una explosión en el número de reglas generadas.

El algoritmo Apriori consta de dos pasos:

1. Encontrar todos los conjuntos frecuentes de ítems, es decir todos aquellos subconjuntos que cumplen con la condición: soporte \geq soporte mínimo.

2. Usar los conjuntos frecuentes de ítems obtenidos para generar reglas que cumplan con la condición confianza \geq confianza mínima.

La idea del algoritmo se basa en que si un conjunto de ítems cumple con la condición de soporte mínimo, entonces todo subconjunto de éste, también la cumplirá, en realidad se usa el contrarrecíproco; cuando se obtiene un conjunto de ítems se chequea si todos los subconjuntos de éste, medidos en la iteración anterior, cumplían con la condición de mínimo soporte, si se encuentra alguno que no la cumple se puede concluir, "a priori", que dicho conjunto no la cumplirá y con esto se quiere minimizar el tiempo total de acceso a la base de datos.

1.7.4 MÉTODOS BASADOS EN VECINDAD

Los métodos basados en vecindad tratan de resolver un problema a partir de la información extraída de un conjunto de ejemplos existentes previamente. Los ejemplos son los que aportarán la información necesaria para poder predecir el comportamiento de un nuevo dato no perteneciente al conjunto de ejemplos.

Reciben su nombre del hecho de que la predicción se basa fundamentalmente en la utilización del conjunto de ejemplos vecinos al dato que hay que procesar, o en el caso más general, por que la distancia entre cada ejemplo y el dato en cuestión es esencial en el proceso.

1.7.4.1 K-Means

Se trata de un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar, sin etiquetar. La idea de K-Medias es situar a los prototipos o centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares.

Todo ejemplo nuevo, una vez que los prototipos han sido correctamente situados, es comparado con estos y asociado a aquel que sea el más próximo. Este método únicamente se puede aplicar a atributos numéricos, y los outliers le pueden afectar muy negativamente.

1.7.4.2 K-Vecinos

Esta técnica de clasificación por vecindad simplemente asigna la clase del ejemplo más próximo, utilizando una función de distancia. Una variante de este método son los k vecinos más próximos, en el que se asigna la clase mayoritaria entre los k vecinos más próximos. La expresividad de este modelo es muy alta, y su principal problema es determinar un buen valor de k.

1.7.5 REDES BAYESIANAS

Método de representación del conocimiento con incertidumbre que permite un doble uso: descriptivo y predictivo.

En cuanto a su uso como modelo descriptivo, las redes bayesianas se centran en el descubrimiento de relaciones de independencia y/o relevancia entre sus variables, por lo que el modelo resultante refleja de forma explícita numerosas relaciones de interés.

En cuanto al uso predictivo, las redes bayesianas están enfocadas como clasificadores.

Las redes bayesianas no son más que un modelo probabilístico multivariado que relaciona un conjunto de variables aleatorias mediante un grafo dirigido, el cual indica explícitamente influencia causal.

Entre sus principales características se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos, evitan el sobre-ajuste y pueden manejar bases de datos incompletas.

1.7.6 MÉTODOS DE REGRESIÓN

Se habla de métodos de regresión cuando tanto la variable de respuesta o de salida como las variables explicativas o de entrada, son todas cuantitativas.

1.7.6.1 Regresión Lineal

Considerada la función de regresión más simple, donde cada variable explicativa participa de forma aditiva y constante para todo el dominio observado en la formación de la respuesta. Propia para tareas de regresión y análisis de correlaciones.

1.7.6.2 Regresión Logística

Modelo de regresión generalizada, es por cierto uno de los modelos generalizados más frecuentes, ya que modeliza una probabilidad. Útil para resolver problemas de clasificación, y de reglas de asociación.

1.7.7 ALGORITMOS GENÉTICOS

Un algoritmo genético es un algoritmo matemático o técnica de programación que imita a la evolución biológica de las especies y su base genético-molecular como estrategia para resolver problemas, mediante imitaciones de mutación,

reproducción y selección, proporcionando programas y optimizaciones que pueden ser utilizados en la construcción y entrenamiento de otras estructuras.

Aplicables a varias tareas de minería de datos como es el caso de la clasificación, regresión, agrupamiento, reglas de asociación, entre otras; aún cuando su aplicación más común ha sido la solución de problemas de optimización, en donde han mostrado ser muy eficientes y confiables.

1.8 RELACION ENTRE TAREAS Y TECNICAS DE MINERIA DE DATOS

Cuadro 2. Tareas y técnicas de minería de datos

NOMBRE	PREDICTIVO		DESCRPTIVO		
	Clasificación	Regresión	Agrupamiento	R. de Asociación	Correlaciones
Redes Neuronales	x	x	x		
Arboles de Decisión	x	x	x	x	
K-Means			x		
A priori				x	
Redes Bayesianas	x				
Regresión Lineal		x			x
Regresión Logística	x			x	

Vecinos más Próximos	x	x	x		
Algoritmos Genéticos	x	x	x	x	x

Como se puede observar en tabla anterior la correspondencia entre las tareas y técnicas o métodos de minería de datos es muy variada; algunas tareas pueden ser resueltas por muy diversas técnicas y algunas técnicas pueden aplicarse a la solución de más de una tarea. Es por esta variedad que resulta necesario conocer las capacidades de cada técnica, sus características funcionales, los ámbitos donde suele funcionar mejor, y en general su eficiencia.

1.9 HERRAMIENTAS DE MINERIA DE DATOS

Las herramientas, sistemas integrados, soluciones o paquetes de minería de datos como también se le conocen, se definen como entornos software diseñados con el fin de dar soporte al ejercicio de minería de datos; éstos pueden ser encontrados tanto en ámbitos comerciales como académicos. Cada uno de ellos posee una serie de características propias para realizar determinadas tareas o para analizar cierto tipo de datos.

Este tipo de herramientas integran en un mismo entorno capacidades para el preprocesado de datos, diferentes modelos de análisis, facilidades para el diseño de experimentos, y soporte gráfico para la visualización de resultados; a demás son capaces de trabajar con distintos formatos, de incorporar distintas técnicas, de obtener distintos tipos de conocimiento y de aplicarse a un gran abanico de tareas.

En términos generales se puede decir que su manejabilidad no se halla

condicionada al hecho que el usuario posea o no conocimientos de programación, ya que existe una interfaz (en la mayoría de los casos gráfica) que facilita la interacción entre el usuario y la herramienta.

A continuación se hará mención de algunos de los entornos de minería de datos más populares que actualmente se encuentran disponibles en el mercado. Se profundizará con mayor información en las herramientas Clementine y WEKA, ya que son éstas las soluciones utilizadas en la fase práctica del presente trabajo.

La selección de dichas herramientas fue basada en los resultados de una encuesta hecha en el conocido portal sobre minería de datos: Kdnuggets¹ en junio del 2002, donde se le preguntaba al encuestado sobre la herramienta de minería de datos que habitualmente usa. Los porcentajes más altos fueron obtenidos por estas dos herramientas: Clementine: (13%), y WEKA: (10%).

1.9.1 CLEMENTINE

Clementine es la mesa de trabajo líder en procesos de minería de datos según mineros y analistas de negocio a nivel mundial, ya que soporta todo el proceso de minería de datos entregando el máximo retorno de inversión de los datos en poco tiempo. Con Clementine los usuarios pueden acceder fácilmente y preparar sus datos para el modelaje, construir y comparar modelos rápidamente, y a la vez desplegarlos eficientemente en tiempo real a las personas que toman decisiones o realizan recomendaciones.

Fomenta la creatividad de los usuarios finales, a la vez que resulta muy fácil de usar sin necesidad de poseer conocimientos sobre redes neuronales, inducción de reglas u otros sistemas. El usuario no debe especificar parámetros técnicos en

¹ http://www.kdnuggets.com/polls/data_mining_tools_2002_june2.htm

Clementine, sólo debe preocuparse de aplicar, de una manera creativa su conocimiento sobre los datos, en un proceso interactivo con el ordenador.

Algunas de organizaciones que utilizan Clementine para maximizar los retornos en inversión son: la Caixa, Banco Popular, Caixa de Catalunya, Terra, Telefónica Móviles, Uni2, Sanitas, Winterthur, Hsbc Bank, U.S Government Intelligence Agency, U.S. Government Tax Agency, West Midlands Police, entre muchas otras.

Clementine funciona en todas las plataformas hardware y sistemas operativos UNIX Y VMS, incluso sobre Windows NT y en configuraciones cliente-servidor. Comercializado por la empresa SPSS, compañía multinacional de software de computadoras que proporciona tecnología que transforma los datos en conocimiento.

1.9.2 WEKA

Herramienta visual de libre distribución y difusión, compuesta por una extensa colección de algoritmos de máquinas de conocimiento desarrollados por la universidad de Waikato (Nueva Zelanda) implementados en Java; útiles para ser aplicados sobre datos mediante los interfaces que ofrece o para embeberlos dentro de cualquier aplicación.

Weka contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Está diseñada como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla.

Sin embargo, y pese a todas las cualidades que Weka posee, tiene algunos defectos entre los que vale mencionar el límite de trabajo sobre el volumen de los datos, y la escasa documentación orientada al usuario, lo que la convierte en una

herramienta difícil de comprender y manejar sin información adicional. Weka es independiente de la arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible.

1.9.3 ENTERPRISE MINER

Herramienta comercial que proporciona SAS Institute para tareas de minería de datos. Posee una arquitectura distribuida, en donde toda la funcionalidad del sistema es accesible mediante una potente interfaz gráfica de usuario. Los modelos proporcionados son: árboles de decisión, regresión lineal y logística y redes neuronales.

1.9.4 INTELLIGENT MINER

Herramienta comercial distribuida y desarrollada por IBM y pensada para explotar los masivos sistemas de información de las grandes corporaciones. Las tareas soportadas son: agrupamiento, asociaciones, patrones, clasificación, predicción, y análisis de serie temporales. Es especialmente útil para trabajar con grandes volúmenes de datos puesto que proporciona herramientas escalables. También posee un lenguaje de programación que permite al usuario especificar los experimentos a realizar y agilizar así el proceso de exploración de los datos.

1.9.5 ODMS DARWIN

Sistema de minería de datos desarrollado por Thinking Machines y adquirido y comercializado por Oracle. ODMS está diseñado sobre una arquitectura cliente/servidor y ofrece una elevada versatilidad para el acceso a grandes volúmenes de información.

Su interfaz gráfica potencia la interactividad con el usuario y el sistema. Los modelos de aprendizaje soportados son: redes neuronales para clasificación y regresión, regresión lineal, árboles de decisión, vecinos más próximos, redes bayesianas y técnicas de agrupamiento (K-medias).

1.10 PROGRAMA DE MINERÍA DE DATOS

Un programa de minería de datos es un uso sistemático, prolongado y racional de las técnicas de minería de datos en una determinada organización. La decisión de implantar un programa de este tipo debe preceder a cualquiera de las fases de la minería de datos; y depende de la identificación de necesidades que pueden ser cubiertas con su implementación.

Es necesario además, tener en cuenta que aunque la minería de datos funciona en general, no funciona igual de bien en todos los ámbitos, y evidentemente existirá organizaciones para las que será mayor el esfuerzo que el beneficio obtenido; es por esto que para tomar la decisión de si se debe o no implementar un programa de minería de datos se debe estudiar el caso en profundidad; y en algunas situaciones se sugerirá diseñar, aunque sea esquemáticamente, el programa de minería de datos, para saber si es factible o no.

1.10.1 METODOLOGÍA PARA LA IMPLEMENTACIÓN DE UN PROGRAMA DE MINERÍA DE DATOS: CRISP-DM

CRISP-DM (Proceso Estándar entre Industrias para Minería de Datos, por su sigla en inglés), es una metodología para el desarrollo de proyectos de minería de datos que se ha convertido en un estándar de facto. Esta metodología está descrita como un proceso jerárquico, compuesto por una serie de fases sucesivas,

considerado a demás como una guía de referencia a la hora de establecer una formulación o planificación de un programa de minería de datos.

A continuación se describirá brevemente las fases y las subfases de un programa de minería de datos según la metodología CRISP-DM.

1.10.1.1 Comprensión Del Negocio

La primera fase se centra en entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio, plasmando todo esto en una definición del problema de minería de datos y un plan preliminar diseñado a alcanzar los objetivos. Consta de tres subfases: establecimiento de las necesidades del negocio, objetivos de negocio y establecimiento de los objetivos de minería de datos.

Esta es la fase más relevante de todas ya que se considera como el punto de partida del programa pues establece las bases para la realización de las fases posteriores del proceso y por lo tanto el éxito o fracaso del mismo va a depender en gran medida de las decisiones que se adopten en esta etapa.

A continuación se detallarán cada una de sus subfases:

Identificación De Las Necesidades Del Negocio: Un aspecto a considerar si es o no necesario un programa de minería de datos es la identificación de las necesidades o problemas de la organización que pueden llegar a ser cubiertas con la minería de datos. Y es que, la necesidad debe dar lugar al programa, y no al revés, el programa dar lugar a una necesidad. Estas necesidades definidas y establecidas son las que van a marcar la siguiente subfase.

Establecimiento De Los Objetivos De Negocio: Es fundamental analizar cuáles son los objetivos del negocio, para abordar aquellas cuestiones que son realmente relevantes y beneficiosas para la organización. Si no existe suficiente conocimiento sobre el negocio, habrá que entrevistar e involucrar a los expertos en el tema.

Establecimiento De Los Objetivos De Minería De Datos: Esta subfase consiste en especificar los objetivos en términos de minería de datos, y no en términos de negocio. Es decir, es necesario que los objetivos de negocio establecidos en la fase anterior, sean plasmados y desglosados en tareas de minería de datos, donde se sepa a priori que tarea se va a resolver y que tipo de modelo se va a obtener.

1.10.1.2 Comprensión De Los Datos

Se trata de recopilar y familiarizarse con los datos, identificar los problemas de calidad que pudieran presentar y ver las primeras potencialidades o subconjuntos de datos que puede ser interesante analizar (para ello es muy importante haber establecido los objetivos de negocio en la fase anterior).

La fase de comprensión de los datos consta de cuatro subfases: recopilación inicial de datos, descripción de datos, exploración de datos y verificación de calidad de datos.

Para comenzar a analizar y extraer algo útil de los datos es preciso, en primer lugar disponer de ellos. En general, el problema de reunir un conjunto de datos que posibilite la extracción de conocimiento, es lo que da lugar a la subfase más relevante de la comprensión de datos: la recopilación, es por esto que se describirá a continuación.

Recopilación De Datos: Como se ha comentado, para poder llevar a cabo una búsqueda de conocimiento mediante las técnicas de minería de datos, se requiere de un sistema que permita la recopilación de la información. Esto implica, una necesidad clara de disponer de un sistema de adquisición, almacenamiento y manejo de la información lo suficientemente eficiente. Es aquí donde entra el Data Warehouse.

1.10.1.3 Preparación De Los Datos

El objetivo de esta fase es obtener la “vista minable”, aquí se incluye la integración, selección, limpieza y transformación de los datos. Consta de tres subfases: selección de datos, limpieza de datos y construcción de datos (atributos derivados, registros generados).

Selección De Datos: La selección de datos consiste en la identificación de los datos más relevantes del proceso así como su disponibilidad; esta tarea debe ser realizada por el analista, y consiste en la creación del conjunto de datos objetivo, enfocando la búsqueda en subconjuntos de variables y/o muestras de datos en donde realizar el proceso de análisis.

Limpieza De Datos: Como es sabido existen problemas de calidad de datos en los sistemas de información. La limpieza de datos, puede en muchos casos, detectar y solucionar estos problemas de calidad.

Construcción De Datos: El objetivo de la construcción de datos es la transformación del conjunto original de datos en un nuevo conjunto de datos más significativo y manejable, de modo que la información que se tiene pueda ser procesada por los algoritmos de minería de datos y además, reducir la cantidad de información redundante para simplificar las tareas posteriores.

Identificación y Conversión de Atributos

Las primeras tareas de la transformación de datos, son las más arduas ya que, generalmente, deben consistir en identificar, casi manualmente, los diferentes tipos de atributos existentes en la base de datos y convertirlos a otro tipo dependiendo de las necesidades posteriores.

Fundamentalmente, podemos clasificar los atributos en los siguientes dos grupos:

- ❖ Numéricos o Cuantitativos. También llamados “continuos”.
- ❖ Nominales o Cualitativos. También llamados “discretos” o “categóricos”.

Los atributos, según el tipo que sean, deben ser acomodados a los algoritmos que se vayan a utilizar. De esta forma, muchas veces resulta necesaria la conversión de los datos para que puedan ser tratados convenientemente. Es decir, es necesario comprender cómo trabajan los algoritmos de minería de datos que vamos a utilizar para saber como preparar los datos. Por ejemplo, un atributo nominal no puede ser tratado por una red neuronal o un clasificador basado en árboles puede necesitar que los datos sean nominales. Es por esto que los datos necesitan tratamientos como:

- ❖ La Discretización: Es la conversión de un valor numérico en un valor nominal.
- ❖ La Numerización: Es el proceso inverso de la discretización, aunque es menos común, también existen casos en los que puede ser extremadamente útil.

A demás de la conversión de atributos anteriormente expuesta, en la subfase de transformación o construcción de datos también se busca:

- ❖ La extracción de los atributos útiles de los datos (reducción de dimensionalidad): se tienen muchas dimensiones o atributos respecto a la cantidad de instancias o ejemplos.

- ❖ La transformación de los datos con el objetivo de proporcionar una representación de los datos más intuitiva y manejable.

Para conseguir esto, fundamentalmente podemos destacar tres tareas específicas:

- ❖ Reducción de los datos.
- ❖ Creación de datos derivados.
- ❖ Transformación de la distribución de los datos.

1.10.1.4 Modelado

Es la aplicación de técnicas de minería de datos propiamente dichas a las vistas minables preparadas en la fase anterior. Consta de cuatro subfases: selección de la técnica de modelado, diseño de la evaluación, construcción del modelo y evaluación del modelo.

Se hará únicamente descripción de las subfases de selección de la técnica de modelado y evaluación del modelo, ya que con esta información queda claro lo relacionado con las subfases restantes.

Selección De La Técnica De Modelado: Consiste en elegir el tipo de técnica o modelo apropiado, que permita cubrir la tarea de minería de datos correspondiente con lo objetivos definidos en la primera fase del programa.

Evaluación Del Modelo: Para entrenar y probar la calidad de un modelo es necesario dividir los datos en dos conjuntos: el de entrenamiento, que como su nombre lo indica es el conjunto de datos sobre los cuales se entrena o se genera el modelo, y el de prueba o de test, que es el conjunto de datos con los cuales se prueba la eficiencia y precisión del modelo entrando. Esta separación es necesaria para garantizar la validación de la precisión del modelo.

Si no se usan conjuntos diferentes de entrenamiento y prueba, la precisión del modelo será sobrestimada, es decir, se tendrán estimaciones muy optimistas.

En los modelos predictivos, el uso de esta separación entre entrenamiento y prueba es fácil de interpretar. Por ejemplo, para una tarea de clasificación, después de generar el modelo con el conjunto de entrenamiento, éste se puede usar para predecir la clase de los datos de prueba. Entonces, la precisión del modelo se obtiene dividiendo el número de clasificaciones correctas por el número total de instancias; siendo ésta una buena estimación de cómo se comportará el modelo para datos futuros similares a los de test.

La evaluación de los modelos descriptivos, como por ejemplo, los modelos de agrupamiento y las reglas de asociación, resulta un tanto complicado; esto se debe fundamentalmente a la ausencia de una clase donde medir el grado de acierto de un modelo. Es por ello que las medidas de evaluación de los modelos descriptivos se basan en conceptos tales como complejidad del modelo y de los datos a partir del modelo, o bien, en agrupamiento, el nivel de compactación de los diferentes grupos.

La mejor evaluación de este tipo de modelos es saber si el modelo resultado de la fase de aprendizaje tiene un comportamiento útil cuando se utilice en su área de aplicación.

1.10.1.5 Evaluación

Es necesario evaluar los modelos de la fase anterior, pero ya no sólo desde un punto de vista estadístico respecto a los datos, como se realiza en la última subfase de la fase anterior, sino ver si el modelo se ajusta a las necesidades establecidas en la primera fase, es decir, si el modelo sirve para responder a algunos de los requerimientos del negocio.

1.10.1.6 Despliegue

Se trata de explotar la potencialidad de los modelos, integrarlos en los procesos de toma de decisión de la organización, difundir informes sobre el conocimiento extraído, etc. Consta de dos subfases principales: la planificación del despliegue y la generación del informe final.

1.11 TRATAMIENTO DE LA INFORMACIÓN: CRÍTICA DE DATOS

Los beneficios del análisis y de la extracción de conocimiento a partir de los datos dependen en gran medida de la calidad de los datos recopilados.

El concepto de calidad de datos se asocia cada vez más a los sistemas de información; pero éstos han crecido de tal manera en las últimas décadas, que el problema de calidad, en vez de resolverse, en muchos casos se ha acentuado.

En definitiva, el éxito de un sistema de información con miras a la obtención de conocimiento con valor agregado para la organización depende, no solo de tener todos los datos necesarios (resultado de sus procesos a través de los años), sino de que éstos estén íntegros, completos y consistentes.

En la mayoría de las bases de datos existe mucha información que es incorrecta respecto al dominio de la realidad que se desea cubrir, y un número menor, pero a veces también relevante, de datos inconsistentes. Es a este tipo de situaciones a las que se le conoce como ruido en la información de una base de datos, y existen diferentes factores que lo pueden generar, como el caso de los datos anómalos o outliers como también se le conocen, los datos erróneos, faltantes, y los valores duplicados; los cuales a la vez, son originados por fuentes diversas.

1.11.1 FILTRADO, DETECCIÓN Y ELIMINACIÓN DE DATOS ANOMALOS

Se consideran datos anómalos o valores no usuales a un conjunto de datos cuya disposición espacial resulta sensiblemente extraña frente al comportamiento general del conjunto. Este tipo de datos se denominan outliers o espurios y deben ser tratados convenientemente ya que pueden afectar enormemente los resultados finales de los modelos generados, y peor aun pueden llegar a afectar los procesos que tiene relación directa con ellos. Los outliers pueden venir originados por errores en la medida, en la inserción de datos o en la transformación de los mismos, aunque también puede ser debidos a sucesos inusuales que pueden ser interesante estudiar, ya que un estudio detallado puede aportar información valiosa sobre los procesos implicados; es por esto la importancia de crear una alarma o reporte frente a su aparición en las fuentes de datos, para poder determinar su origen o procedencia, y de esta forma decidir como eliminarlos y a la vez tratar su fuente de generación. Entre las técnicas para detectar outliers en muestras normales multivariantes se pueden mencionar:

- ❖ Técnicas de regresión, para estimar el modelo que define los datos y determinar la desviación de los puntos frente al mismo según diversos criterios de calidad.
- ❖ Histogramas o gráficos boxplots para detectar los outliers gráficamente.
- ❖ Diagramas de dispersión que permitan identificar los datos anómalos de manera gráfica.
- ❖ Proyectores lineales.

1.11.2 DETECCIÓN Y TRATAMIENTO DE VALORES FALTANTES

Los valores faltantes, perdidos, ausentes, inexistentes o información no disponible como también se le suele conocer, son un mal endémico en las bases de datos de sistemas reales. Estos casos son muy comunes y pueden deberse a causas muy diversas, por ejemplo como al mal funcionamiento del dispositivo que hizo la

lectura del valor, a cambios efectuados en los procedimientos usados durante la colección de los datos o al hecho de que los datos se recopilen de desde fuentes diversas; pueden presentarse también en los procesos industriales donde se producen fallos de muestreo, se introducen mal los datos manuales, se realizan conversiones incorrectas o simplemente unas variables se muestren a diferente velocidad que otras.

A la hora de hablar de campos faltantes, debemos tratar de su detección y de su tratamiento. La detección de datos faltantes puede llegar a parecer sencilla. Si los datos proceden de una base de datos, basta mirar en la tabla de resumen de atributos y ver la cantidad de nulos que tiene cada atributo. El problema es que a veces los campos faltantes no están representados como nulos. Este tipo de situaciones complica sobremanera su detección. Tanto para su detección, como para su tratamiento posterior, es importante saber el por que de los valores faltantes.

- ❖ Algunos valores faltantes expresan características relevantes.
- ❖ Valores no existentes. Muchos valores faltantes existen en la realidad, pero otros no.
- ❖ Datos incompletos. Esto suele suceder cuando los datos vienen de fuentes diferentes.

Finalmente si se han conseguido establecer los datos faltantes e, idealmente sus causas, procederemos a su tratamiento. Las posibles acciones sobre datos faltantes son:

- ❖ Ignorar (dejar pasar).
- ❖ Eliminar (filtrar o reemplazar) toda la columna, es decir quitar el atributo para todos los ejemplos.
- ❖ Filtrar la fila.

- ❖ Reemplazar el valor por un valor que preserve la media para el caso de datos numéricos, o por el valor moda, en el caso de valores nominales.
- ❖ Modificar la política de calidad de los datos y esperar hasta que los datos faltantes estén disponibles.

1.11.3 DETECCIÓN Y TRATAMIENTO DE VALORES ERRÓNEOS

Este tipo de errores se consideran frecuentemente como inconsistencias del sistema o proceso del cual son resultado. Su detección se puede realizar básicamente de dos formas; teniendo en cuenta el formato y origen del campo, es decir la detección dependerá fundamentalmente de conocer los posibles valores del campo o aquellos datos que se ajusten al formato; y teniendo en cuenta su contenido. Para su detección se recomiendan técnicas como:

- ❖ El uso de histogramas.
- ❖ Las técnicas de Vecindad.
- ❖ Las técnicas de Clustering.

Para su tratamiento se recurre a acciones muy similares a las que se emprenden en el caso de datos faltantes:

- ❖ Ignorar.
- ❖ Filtrar (eliminar o reemplazar la columna).
- ❖ Filtrara la fila.
- ❖ Reemplazar el valor.

1.11.4 DETECCION Y ELIMINACION DE VALORES DUPLICADOS

Es el caso de ruido en la información más sencillo de describir y el menos común; como su nombre lo indica se presenta cuando un campo determinado toma más de una vez el mismo valor, o valor diferente para el mismo campo. Pueden ser

originados por una mala introducción de los datos manualmente, en el momento de la integración desde fuentes diferentes; entre otras muchas causas.

Su tratamiento es de los más sencillos pero tediosos de realizar, ya que se recomienda eliminarlos de manera manual.

Cuando el campo toma más de una vez el mismo valor, basta simplemente eliminar los valores repetidos; cuando los valores son diferentes se complica la cosa; ya que surge la necesidad de comprobar la validez de los mismos, eliminando aquellos que resultaron equívocos y esto se logra por conocimiento de causa o mediante la consulta de otras fuentes.

2 MINERIA DE DATOS EN XM

A continuación, en el presente capítulo se desarrollará la metodología CRISP-DM, expuesta en el capítulo anterior, aplicada al caso de estudio definido como objetivo de investigación del trabajo de grado desarrollado en XM. Se seguirán cada una de las fases de la metodología en mención, con sus respectivas subfases, con base en la realidad que encierra el caso de aplicación práctica.

2.1 GENERALIDADES DE LA EMPRESA

XM es una empresa del Grupo ISA que presta servicios integrales de operación, administración y desarrollo de mercados mayoristas eléctricos. Tiene el conocimiento y la experiencia práctica necesarios, para ayudar a industrias y gobiernos con cada uno de los aspectos inherentes a la evolución y desarrollo de mercados energéticos. Es una nueva organización basada en el conocimiento, orientada a la generación de valor a partir de la innovación. Tiene como compromiso entregar soluciones inteligentes y consolidar relaciones comerciales de largo plazo con sus clientes.

Con la creación de XM, el sector eléctrico colombiano se beneficiará con el fortalecimiento de relaciones comerciales de largo plazo con todos los agentes involucrados; al igual que se enfrentará a grandes retos como la administración de mercados de derivados financieros, la convergencia entre los sectores de gas y electricidad y la integración energética latinoamericana.

Al igual que todas las empresas del Grupo ISA los inspiran valores que le dan sentido a su tarea corporativa; la ética, la transparencia, la actitud de servicio, la efectividad, la innovación y la responsabilidad social, entre otros, alimentan su espíritu. Estos valores son pilares fundamentales que les permiten dedicarse a la disciplina del conocimiento; en esencia la energía que los impulsa es la innovación y en consecuencia encontrarse en la administración y gestión del conocimiento, la gran fuente de valor que puede generar una ventaja competitiva para la empresa y para el país en la región.

2.1.1 MISIÓN

Operar y Administrar Mercados Mayoristas de Energía y prestar Servicios Especializados Afines, contribuyendo al crecimiento del Grupo Empresarial ISA y generando valor para todos los grupos de interés.

Para la prestación de los servicios nos apoyamos en la tecnología y en el desarrollo integral de nuestra gente.

Actuamos con responsabilidad social y ética, y estamos comprometidos con el bienestar de la sociedad.

2.1.2 VISIÓN

En el año 2008, XM será reconocida por la prestación integrada de servicios de operación y administración de mercados energéticos en el ámbito latinoamericano.

Se destacará por el conocimiento y experiencia de su gente y por la confianza que inspira. Además, ofrecerá de manera competitiva servicios especializados afines para diferentes sectores productivos.

2.1.3 AREA DONDE SE DESARROLLA EL PROYECTO

El equipo de Seguimiento y Análisis de la Información de la Dirección de Servicios Asociados de la organización, es el equipo responsable de ofrecer servicios de información a los clientes del sector energético nacional.

Tiene como misión brindar servicios integrales de información sobre la operación y administración del mercado de energía a agentes y grupos de interés, mediante el análisis, monitoreo y divulgación de las variables del mercado.

Su visión, es llegar a ser reconocidos nacional e internacionalmente como creadores y proveedores de información inteligente que impulse el desarrollo del mercado de energía mayorista; apoyados en tecnologías informáticas y de telecomunicaciones, buscando la innovación, adaptabilidad, oportunidad, calidad y confiabilidad en los medios y en sus contenidos.

Para llegar a este cometido, cumpliendo con lo establecido dentro de la misión referente a brindar servicios integrales de información, a la visión de llegar a ser reconocidos como creadores y proveedores de información inteligente y al perfil corporativo de ser una organización basada en el conocimiento, y superar las expectativas, resulta necesario la adopción de tecnologías de información que estén a la vanguardia y sean acorde con el volumen y valor de la información manejada por el sector; el Mercado Eléctrico Colombiano, considerado hoy como uno de los más dinámicos y competitivos en Latinoamérica.

Por lo anterior, y con el objetivo de explotar toda esta fuente de información para obtener conocimiento con valor agregado y proporcionar información de alta calidad, siendo oportunos y consistentes con la realidad, se torna importante desarrollar nuevas habilidades en busca de una mayor competitividad, para lo cual el uso de técnicas de Minería de datos resulta de gran ayuda.

2.2 FASES DE LA MINERIA DE DATOS APLICADAS EN XM

2.2.1 COMPRESION DEL NEGOCIO

2.2.1.1 Necesidades del Negocio

Identificar patrones de variación en la demanda de energía eléctrica del sector Minero²:

- ❖ En ciertas horas del día
- ❖ Durante los días de la semana
- ❖ En los meses del año
- ❖ En los departamentos de la Costa Atlántica

2.2.1.2 Objetivos del Negocio

- ❖ Identificar los rangos de horario, los días de la semana, los meses del año y los departamentos de la Costa Atlántica, frente a los cuales se presentan patrones que inciden en variaciones del valor (mínimo, medio y máximo) de la demanda de energía del sector Minero.

- ❖ Identificar patrones de comportamiento de la demanda del sector Minero en la Costa Atlántica.

2.2.1.3 Objetivos de Minería de Datos

- ❖ Identificar relaciones no explícitas entre atributos como el horario, el día, el mes del año, y el espacio geográfico (departamentos de la Costa Atlántica), con el valor (mínimo, medio y máximo) de la demanda de energía del sector Minero.

- ❖ Determinar la dependencia del valor (mínimo, medio y máximo) de la demanda

de energía del sector Minero de variables como: hora del día, día de la semana, mes del año y departamento de la Costa Atlántica.

Con los objetivos de minería de datos planteados queda claro la tarea de minería de datos que permitirá su consecución y logro: Reglas de Asociación.

2.2.2 COMPRENSIÓN DE LOS DATOS

2.2.2.1 Recopilación De Datos

Los datos requeridos para el proceso están recopilados en la denominada base de datos de producción de XM Compañía de Expertos en Mercados, donde se encuentra almacenada toda la información confidencial de la organización; quien como administradora y operadora del mercado eléctrico colombiano maneja millones de datos que involucran directa e indirectamente a todos los agentes del sector en mención.

Es por esto, que según políticas y reglamentaciones internas de la compañía, guardando la confidencialidad de la información allí almacenada, los datos requeridos para el proceso de minería de datos serán accedidos mediante consultas “SQL” específicas, trabajadas sobre archivos planos y no a través de conexión directa de la herramienta tecnológica con dicha base de datos.

2.2.2.2 Descripción de los Datos

La información almacenada en la base de datos de producción está categorizada por un gran número de variables implicadas en el proceso operativo de la compañía como operadora del mercado eléctrico colombiano, registradas correspondientemente en los diversos periodos de tiempo: horaria, diaria, mensual

² Sector de la economía, que representa la explotación de minas y canteras.

y anual, con un alcance temporal que abarca desde el 20 de julio de 1995 (día de inicio de la bolsa de energía eléctrica) hasta la actualidad.

2.2.3 PREPARACIÓN DE LOS DATOS

2.2.3.1 Selección de Datos

La selección de las variables a trabajar depende directamente del objetivo de negocio que se pretenda alcanzar. Por lo cual, las variables a seleccionar según los objetivos planteados son:

Demanda del sector Minero (medida en Kwh.), discriminada por: hora del día (periodos de 0 a 23), fecha (de 01-01-03 hasta el 30-09-05) (días semana, mes del año), departamentos de la Costa Atlántica: Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena y Sucre.

2.2.3.2 Limpieza de Datos

Para garantizar la calidad de los datos resulta indispensable un trabajo de limpieza, relacionado con la verificación de los posibles valores de las variables implicadas en el análisis, ya que se puede llegar a presentarse por ejemplo valores de demanda negativa, datos faltantes, información duplicada, y en general ruido en los datos; el cual habrá que detectar y tratar siguiendo las técnicas para tratamiento de información vistas en el capítulo anterior.

2.2.3.3 Construcción de Datos

En esta fase del proceso se realizan los trabajos en los datos requeridos para garantizar, ya sea el buen funcionamiento del modelo a elegir, según los objetivos planteados y la tarea de minería de datos correspondiente, o para generar mayor información.

Tal es el caso de trabajos como la discretización, específicamente el del valor de la demanda, para el objetivo de minería de datos de identificar relaciones no explícitas entre los atributos, ya que para esta tarea (reglas de asociación) se emplean únicamente datos categóricos.

Este tipo de discretizado es posible realizarlo mediante la herramienta tecnológica seleccionada, ya que es un trabajo sencillo y automático, que bien, podría llevar a cabo cualquier tipo de herramienta; este discretizado es realizado a través de una agrupación de valores en rangos o intervalos previamente definidos, teniendo en cuenta por ejemplo el número de registros, la frecuencia de ocurrencia, el número de intervalos que se deseen obtener, y entre otras opciones particulares para cada herramienta.

El horario es una variable de tipo numérica que requiere igualmente un proceso de discretizado, pero éste es preferible realizarlo manualmente, para agregarle el toque propio al caso de estudio, donde se supone un conocimiento previo del negocio en el que se desenvuelve.

En este caso en particular, es necesaria la transformación de las horas del día en periodos de tiempo, definidos por rangos horarios según el comportamiento de la demanda nacional, durante las horas del día a través de los años.

La discretización quedaría de la siguiente forma:

Periodo Bajo 1: (1 – 4)

Periodo Bajo 2: (24)

Periodo Medio 1: (5 – 9)

Periodo Medio 2: (13 – 18)

Periodo Medio 3: (22 – 23)

Periodo Pico 1: (10 – 12)

Periodo Pico 2: (19 – 21)

La fecha, en términos generales es considerada por si sola una variable de muy poca información por lo que resulta necesario transformarla en nuevos atributos más significativos, como los días de la semana y meses del año.

2.2.4 MODELADO

2.2.4.1 Selección de la Técnica de Modelado

Para la selección de la técnica de modelado se hace necesario previamente señalar la tarea que corresponda con los objetivos de minería de datos definidos; y dependiendo de dicha tarea se decide el modelo apropiado para la misma.

En este caso en particular, las tareas asociadas a los objetivos de minería de datos definidos en la primera fase del programa y su correspondientemente técnica de modelado son:

TAREA	—————▶	METODO
Reglas de asociación		Método Apriori

2.2.4.2 Selección de la Herramienta Tecnológica

Muchos entornos y herramientas de minería de datos trabajan con archivos planos, es decir, archivos de texto con campos separados por comas, espacios u otros separadores. En realidad esta puede ser una opción para un primer programa piloto. No obstante un programa a largo plazo de minería de datos debe intentar permitir que la herramienta se conecte con la base de datos. La existencia de sistemas gratuitos es muy sugerente, sobre todo cuando se trata de sugerir el primer sistema. Un sistema gratuito puede permitir comenzar con un coste material muy bajo (limitándose al hardware).

Algunos sistemas gratuitos como el WEKA, son suficientemente completos (es, de hecho, el que tiene la mayor variedad de métodos) para poder comenzar y

avanzar en el programa de minería de datos, antes que llegue cierta necesidad de buscar otras herramientas más potentes; a demás esta herramienta trabaja tanto sobre archivos planos, como conectada a la base de datos, y por razones como las expuestas en la subfase de recopilación de datos, sobre la imposibilidad de conectar la herramienta seleccionada directamente a la base de datos de producción, por cuestiones de confidencialidad; esta primera característica la hace idónea para el caso de estudio, así como el hecho de ser un sistema gratuito.

2.2.4.3 Evaluación de Resultados

La evaluación de los modelos descriptivos, como por ejemplo, las reglas de asociación, resulta un tanto complicado; esto se debe fundamentalmente a la ausencia de una clase donde medir el grado de acierto del mismo. Es por ello que las medidas de evaluación de los modelos descriptivos se basan en conceptos tales como complejidad del modelo y de los datos a partir del modelo.

Para evaluar una regla de asociación, que es la tarea que nos concierne en este caso en particular, se suele trabajar con dos medidas que permiten conocer su calidad: la cobertura y la confianza. La cobertura se define como el número de instancias en las que la regla se puede aplicar. Y la confianza mide el porcentaje de veces que la regla se cumple cuando se puede aplicar. El valor de estas dos medidas dependen del grado de exactitud que el analista le quiera imprimir a los resultados arrojados, y de las opciones que permita la herramienta seleccionada; en este caso las medidas de calidad serán definidas de la siguiente manera: Confianza: >80% y Cobertura: >80%

Es decir, al definir directamente en WEKA estas medidas de calidad, la herramienta no generará ninguna regla que no cumpla con estas; lo que nos permite confiar en la eficiencia de las reglas obtenidas a partir del modelo; sin necesidad de hacer un procedimiento de evaluación posterior a su generación.

2.2.5 EVALUACIÓN

La mejor evaluación es saber si el modelo obtenido tiene un comportamiento útil cuando se utilice en su área de aplicación; o dicho en otras palabras, aquí lo importante es saber si el modelo obtenido si responde las necesidades del negocio que fueron planteadas en la primera fase de este programa.

2.2.6 DESPLIEGUE

2.2.6.1 Planificación del Despliegue

Finalmente en la fase de despliegue o difusión se hace uso del nuevo conocimiento y se hace participe de el a todos los posibles usuarios, ya sea por cauces habituales dentro de la organización, o a través de reuniones, grupos primarios, etc. programados especialmente con ese fin, con la intención de que se haga una integración de dicho conocimiento en el proceso de toma de decisiones.

La transferencia del conocimiento adquirido, durante la implementación de este primer programa de minería de datos, al equipo de trabajo de Seguimiento y Análisis de Información de XM, está planeado ser realizado a través de una capacitación dirigida por personal capacitado de la compañía SPSS distribuidora de CLEMENTINE, con experiencia en trabajos de minería de datos. La capacitación está programada a ser realizada básicamente en tres etapas que consisten en:

Una primera etapa, la cual abarcará todo el conocimiento teórico requerido para comprender la realidad que encierra un programa de minería de datos, su importancia, aplicación, fases de trabajo, etc. La segunda etapa en la cual se expondrá el manejo de CLEMENTINE, a través de ejemplos reales, que incluyan variables de análisis atractivas a la realidad de la organización. Finalmente en la última etapa del proceso de capacitación se expondrán los resultados obtenidos a través del análisis del caso de estudio de este programa.

2.3 RESULTADOS Y ANALISIS DEL CASO DE ESTUDIO

Para efectos de eficiencia en la obtención de los resultados arrojados por el modelo, este fue ejecutado de dos maneras diferentes, una, para un caso general o conjunto como lo llamaremos de aquí en adelante, donde se trabajó con toda la información del sector Minero en la Costa Atlántica, es decir, en este caso se incluyeron los datos de todos los departamentos; y la otra, para un caso individual, en donde se utilizaron los datos de cada departamento de manera individual. A continuación se mostrarán los resultados arrojados por la herramienta WEKA para el caso de estudio, que constan básicamente de: un formato original de la reglas que exporta la herramienta, (solo una pequeña muestra); la traducción de dichas reglas a un formato excel diseñado para facilitar su comprensión, ya que no son fácilmente identificables por aquel que no conozca su funcionamiento, a demás del hecho que su lectura se hace un tanto tediosa por su gran número; un resumen con las estadísticas suministradas durante la fase de preprocesamiento de datos, histogramas de frecuencias en excel y gráficas elaboradas en WEKA bajo el criterio propio del analista, con el fin de facilitar la comprensión de los resultados, y de corroborar la eficiencia de las reglas obtenidas.

Estos resultados se mostrarán para cada uno de los casos. Aunque para el caso general, existe una excepción para las reglas exportadas en el formato excel, ya que no se mostrarán de manera independiente, sino, que serán reportadas en cada formato de los casos individuales, es decir, para cada departamento; con el fin de comparar la eficiencia de los resultados obtenidos con ambos casos. Adicionalmente se analizará cada caso de manera independiente, dando a conocer como el primer caso: el general o conjunto, generó la necesidad de realizar el otro: el caso individual; con el fin de lograr una eficiencia en el estudio.

Antes de continuar resulta importante dejar claro dos puntos; el primero, tener en cuenta que este análisis no incluye razones de ocurrencia; ya que únicamente tendría este tipo de conclusiones un análisis realizado por un experto en el

negocio. El segundo, hace referencia al hecho que no todas las opciones de reglas (según el formato de traducción) fueron obtenidas; ya que la herramienta omite o excluye aquellas reglas que no cumplan con las medidas de calidad especificadas

2.3.1 REGLAS EXPORTADAS POR WEKA

```
=== Run information ===
Scheme: weka.associations.Apriori-N 800
Relation:  Minero-weka.filters.unsupervised.attribute.Remove-
R5-weka.filters.unsupervised.attribute.Discretize-B43-M-1-R1
Instances:    271900
Attributes:   5
              DEMANDA
              DIA SEMANA
              MES
              DEPTO
              HORA

=== Associator model (full training set) === A priori ===
Best rules found:
1.DEMANDA='(12838.825116-17118.433488]'    DIA    SEMANA=Martes
MES=Marzo 219 ==> DEPTO=LA_GUAJIRA 219    acc:(0.995)
2.DEMANDA='(171184.334884-175463.943256]'  HORA=Periodo_Bajo_1
218 ==> DEPTO=CORDOBA 218    acc:(0.995)
3.DIA    SEMANA=Sábado    MES=Enero    DEPTO=ATLANTICO
HORA=Periodo_Medio_2 168 ==> DEMANDA='(-inf-4279.608372]' 168
acc:(0.99499
```

Las reglas de asociación arrojadas por WEKA, como estas que acabamos de ver, pueden ser exportadas a cualquier archivo de texto.

2.3.2 RESULTADOS CASO GENERAL O CONJUNTO

Resumen De Estadísticas

Cuadro 3. Estadísticas caso general

Total de Datos	271.900	
Máximo	184.023,16	
Mínimo	0	
Rango	184.023,16	
No. Clases	42,54	43
Ancho Intervalo	4.279,60	
Media	15.217.39	
Desviación	42.264.56	

Cuadro 4. Distribución de Frecuencia caso general

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-4.279,608372]'	221.910	81,615
'(4.279,608372-8.559,216744]'	581	0,214
'(8.559,216744-12.838,825116]'	5.925	2,179
'(12.838,825116-17.118,433488]'	13.844	5,092
'(17.118,433488-21.398,04186]'	4.447	1,636
'(21.398,04186-25.677,650233]'	126	0,046
'(25.677,650233-29.957,258605]'	29	0,011
'(29.957,258605-34.236,866977]'	13	0,005
'(34.236,866977-38.516,475349]'	19	0,007
'(38.516,475349-42.796,083721]'	17	0,006
'(42.796,083721-47.075,692093]'	23	0,008
'(47.075,692093-51.355,300465]'	28	0,010
'(51.355,300465-55.634,908837]'	29	0,011
'(55.634,908837-59.914,517209]'	32	0,012
'(59.914,517209-64.194,125581]'	62	0,023
'(64.194,125581-68.473,733953]'	108	0,040

DEMANDA	FRECUENCIA	PORCENTAJE %
'(68.473,733953-72.753,342326]'	133	0,049
'(72.753,342326-77.032,950698]'	190	0,070
'(77.032,950698-81.312,55907]'	292	0,107
'(81.312,55907-85.592,167442]'	403	0,148
'(85.592,167442-89.871,775814]'	467	0,172
'(89.871,775814-94.151,384186]'	509	0,187
'(94.151,384186-98.430,992558]'	327	0,120
'(98.430,992558-102.710,60093]'	254	0,093
'(102.710,60093-106.990,209302]'	255	0,094
'(106.990,209302-111.269,817674]'	269	0,099
'(111.269,817674-115.549,426047]'	267	0,098
'(115.549,426047-119.829,034419]'	319	0,117
'(119.829,034419-124.108,642791]'	409	0,150
'(124.108,642791-128.388,251163]'	488	0,179
'(128.388,251163-132.667,859535]'	637	0,234
'(132.667,859535-136.947,467907]'	726	0,267
'(136.947,467907-141.227,076279]'	1.071	0,394
'(141.227,076279-145.506,684651]'	1.246	0,458
'(145.506,684651-149.786,293023]'	1.756	0,646
'(149.786,293023-154.065,901395]'	2.302	0,847
'(154.065,901395-158.345,509767]'	2.790	1,026
'(158.345,509767-162.625,11814]'	2.999	1,103
(162.625,11814-166.904,726512]'	2.904	1,068
'(166.904,726512-171.184,334884]'	2.230	0,820
'(171.184,334884-175.463,943256]'	1.171	0,431
'(175.463,943256-179.743,551628]'	264	0,097
'(179.743,551628-184.023,16]'	29	0,011

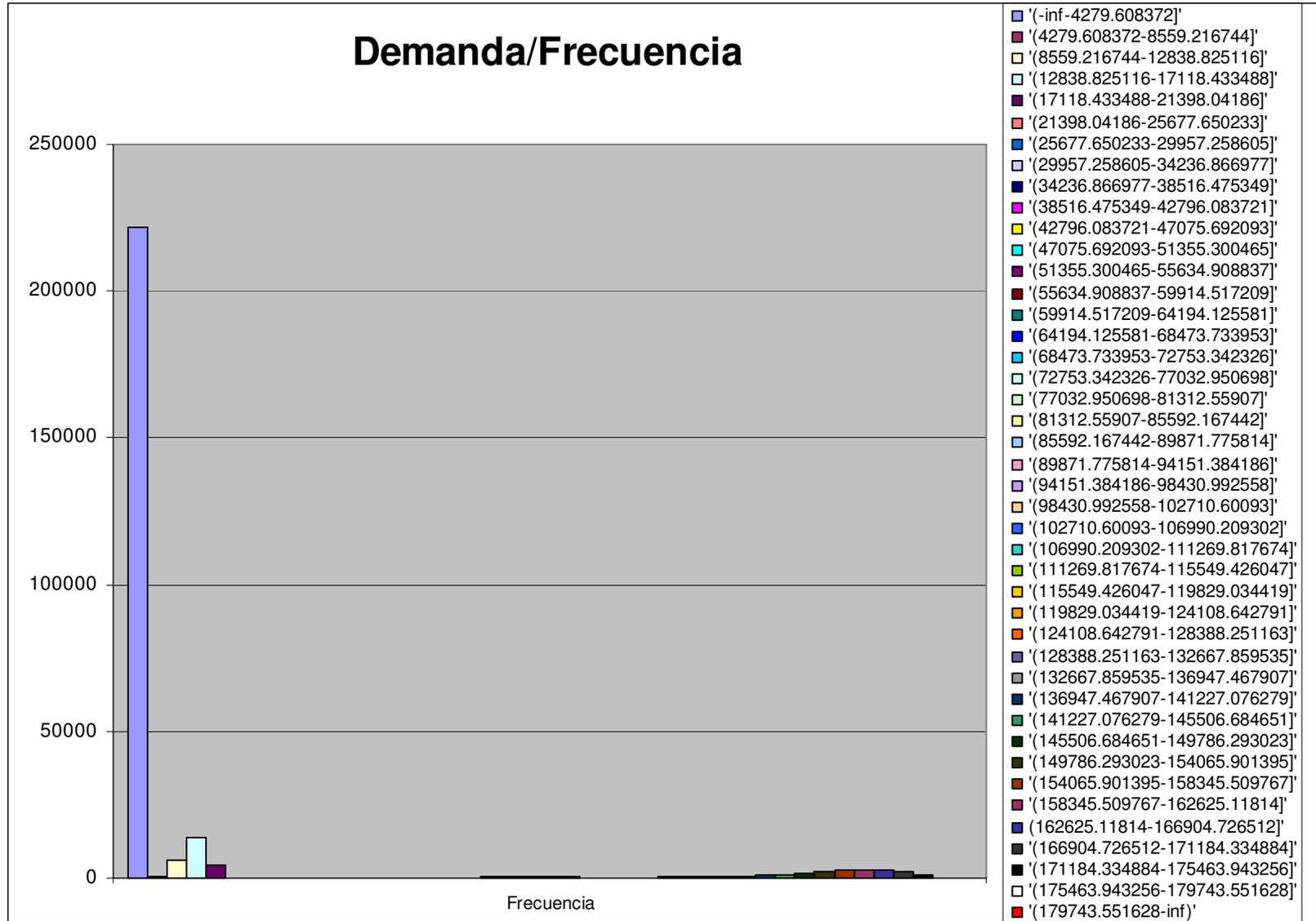


Figura 1. Histograma de frecuencia caso general

Gráficas

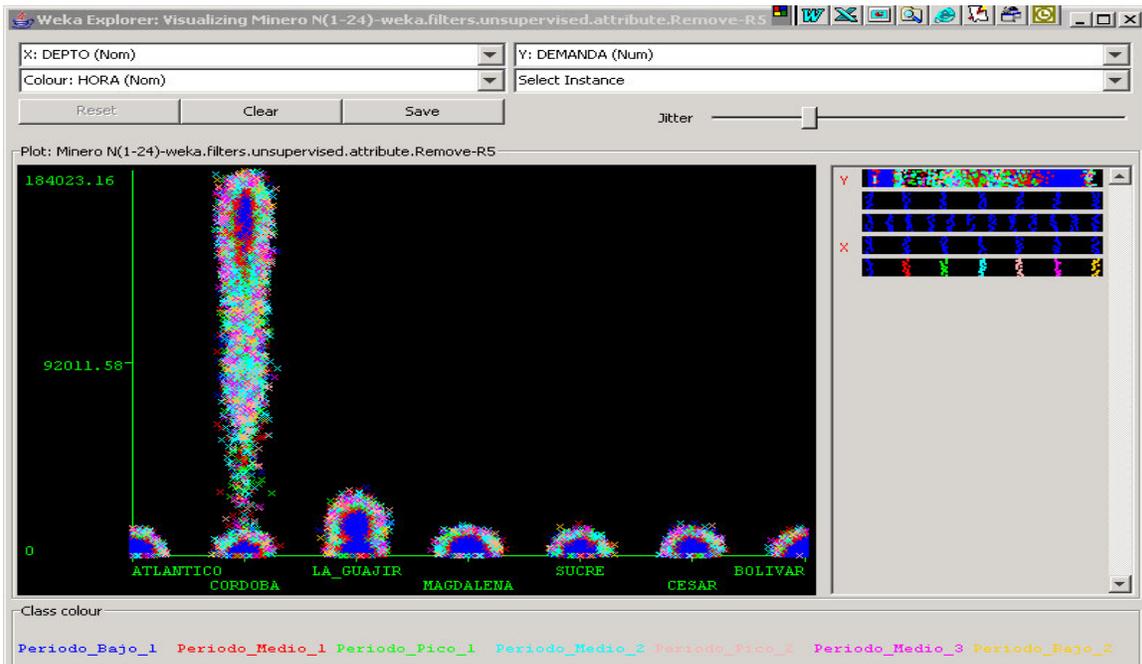


Figura 2. Demanda departamentos diferenciada por periodos del día caso general

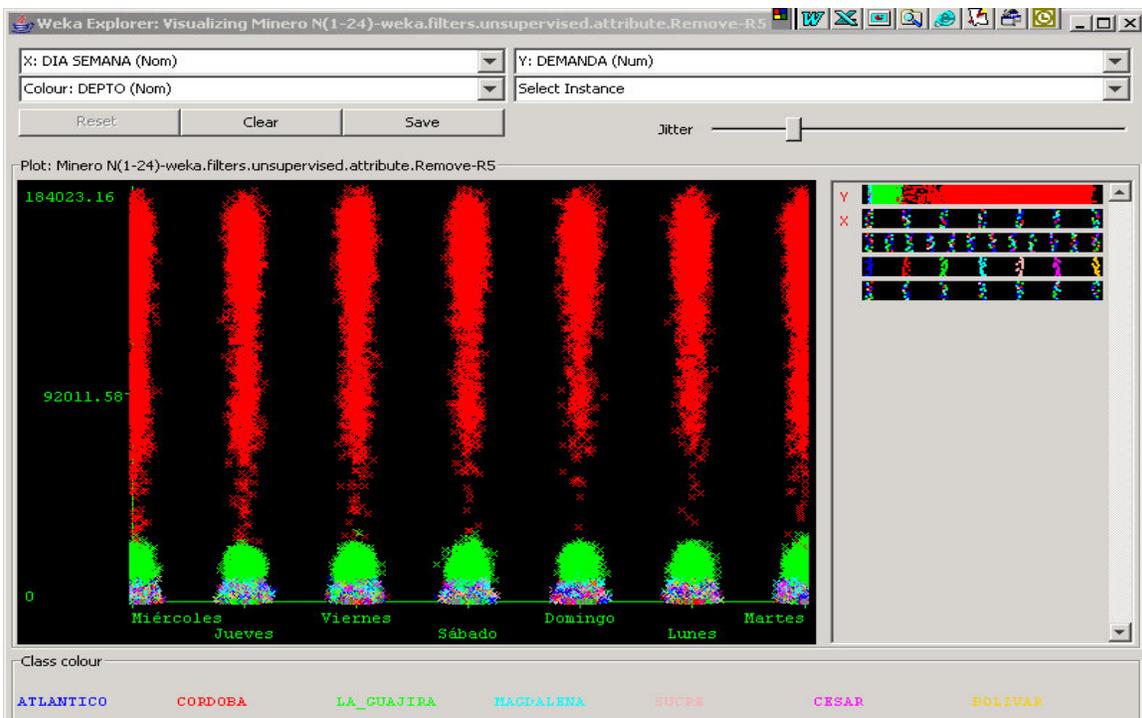


Figura 3. Demanda días de la semana diferenciada por departamentos caso general

2.3.3 ANÁLISIS CASO GENERAL O CONJUNTO

El análisis de los resultados del caso conjunto, consiste en la interpretación de gráficos y reglas obtenidas a partir de corridas de la aplicación, con todos los departamentos de la Costa Atlántica: Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena y Sucre; con los atributos previamente definidos: demanda horaria del sector Minero, periodo del día, día de la semana y mes del año; con el alcance temporal establecido: de 01-01-03 hasta el 30-09-05; a través de la tarea de minería de datos de reglas de asociación, por el método A priori.

Se realizaron las corridas con todas las combinaciones posibles de atributos; inicialmente con los cinco originales, hasta llegar a trabajar únicamente con pares de atributos.

Según resultados obtenidos a través de esta metodología de análisis, la demanda del sector Minero es jalonada en un alto porcentaje por un solo intervalo o rango de valores (entre los 43 definidos), de (0-4.279,60) kW/h, porcentaje que corresponde al 81,61% de la demanda de dicho sector, impidiendo por esta razón la uniformidad, el detalle, y con esto, la eficiencia de este tipo de análisis,

En palabras más claras y precisas, se puede afirmar que el 81,61% de la demanda del sector Minero oscila entre los (0-4.279,60) kW/h, demanda que corresponde a los departamentos de: Atlántico (departamento con la demanda más baja del sector), Bolívar, Cesar, Magdalena y Sucre, en un 100%, y el 50,23% de la demanda de La Guajira (siendo el departamento con la segunda demanda más alta; pero estando muy por debajo de la mayor demanda del sector).

El porcentaje restante, es decir el 18,39% de la demanda del sector, está representado por el 47,77% de la demanda del departamento de La Guajira, y el 100% de la demanda del departamento de Córdoba (departamento con la mayor demanda promedio del sector: 88.318,01kW/h).

Por lo anteriormente expuesto es sugerente señalar que existe una considerable variabilidad o dispersión en los datos, razón confirmada por el alto valor del estadístico de la desviación estándar: 42.264,56 kW/h.

En la Figura 2. Demanda departamentos diferenciada por periodos del día, es notorio el comportamiento especialmente marcado del periodo bajo 1; ya que para todos los departamentos, el valor de la demanda durante el periodo en mención, oscila con frecuencia considerable alrededor de los mismos rangos de valores dependiendo del departamento, y llega a ser de cierto modo excluyente con el resto de periodos, los cuales nunca llegan a tomar valores dentro de dichos intervalos.

La Figura 3. Demanda días de la semana diferenciada por departamento, muestra que para el departamento de Córdoba, los días miércoles y jueves son los días de menor demanda, seguido por los días viernes, sábado y domingo, donde en orden de frecuencia se presentan cada vez menos esos valores de demanda baja, terminando con lunes y martes donde nuevamente inicia la tendencia de los valores bajos.

Es claro no solo a nivel de gráficas, sino también a través de las estadísticas anteriormente expuestas y las reglas obtenidas para este caso, lo ineficiente que resulta incluir todos los departamentos para analizar de manera conjunta el sector Minero; ya que existe un departamento, para ser más concretos, el departamento de Córdoba, (cuya demanda se debe al Complejo minero y metalúrgico de Cerromatoso); el cual presenta una demanda promedio muy por encima de la del resto de departamentos; lo cual ocasiona alta variabilidad en los datos dando lugar a conclusiones erradas, ya que permite todo el análisis en base a este departamento únicamente (como quedo claro con la conclusión anterior), o en base a la realidad de los departamentos restantes como un todo, y no, a la del sector en general.

Por eso resulta importante tener en cuenta que aún cuando las medidas de calidad de las reglas de asociación: cobertura y confianza, se hayan definido con el fin de asegurar la calidad del análisis, esto no es suficiente; es necesario validar los resultados obtenidos con la realidad del negocio.

2.3.4 RESULTADOS CASOS INDIVIDUALES

2.3.4.1 Resultados Departamento: ATLANTICO

Resumen De Estadísticas

Cuadro 5. Estadísticas Atlántico

Total de datos	50.200	
Máximo	353,31	
Mínimo	0	
Rango	353,31	
No. Clases	36,93	37
Ancho Intervalo	9,54	
Media	105,95	
Desviación	83,41	

Cuadro 6. Distribución de frecuencia Atlántico

DEMANDA	FRECUENCIA	PORCENTAJE %
'(-0-9.548919]'	1.560	3,11
'(9.548919-19.097838]'	5.848	11,65
'(19.097838-28.646757]'	7.440	14,82
'(28.646757-38.195676]'	3.906	7,78
'(38.195676-47.744595]'	1.761	3,51
'(47.744595-57.293514]'	9.56	1,90
'(57.293514-66.842432]'	6.44	1,28
'(66.842432-76.391351]'	6.17	1,23

DEMANDA	FRECUENCIA	PORCENTAJE %
'(76.391351-85.94027]'	5.44	1,08
'(85.94027-95.489189]'	6.42	1,28
'(95.489189-105.038108]'	1.157	2,30
'(105.038108-114.587027]'	1.838	3,66
'(114.587027-124.135946]'	2.632	5,24
'(124.135946-133.684865]'	2.608	5,20
'(133.684865-143.233784]'	1.606	3,20
'(143.233784-152.782703]'	1.152	2,29
'(152.782703-162.331622]'	1.310	2,61
'(162.331622-171.880541]'	1.363	2,72
'(171.880541-181.429459]'	1.466	2,92
'(181.429459-190.978378]'	1.530	3,05
'(190.978378-200.527297]'	1.317	2,62
'(200.527297-210.076216]'	1.084	2,16
'(210.076216-219.625135]'	1.133	2,26
'(219.625135-229.174054]'	1.310	2,61
'(229.174054-238.722973]'	1.162	2,31
'(238.722973-248.271892]'	696	1,39
'(248.271892-257.820811]'	572	1,14
'(257.820811-267.36973]'	564	1,12
'(267.36973-276.918649]'	482	0,96
'(276.918649-286.467568]'	423	0,84
'(286.467568-296.016486]'	389	0,77
'(296.016486-305.565405]'	275	0,55
'(305.565405-315.114324]'	107	0,21
'(315.114324-324.663243]'	56	0,11
'(324.663243-334.212162]'	28	0,06
'(334.212162-343.761081]'	16	0,03
'(343.761081-353,31)'	6	0,01

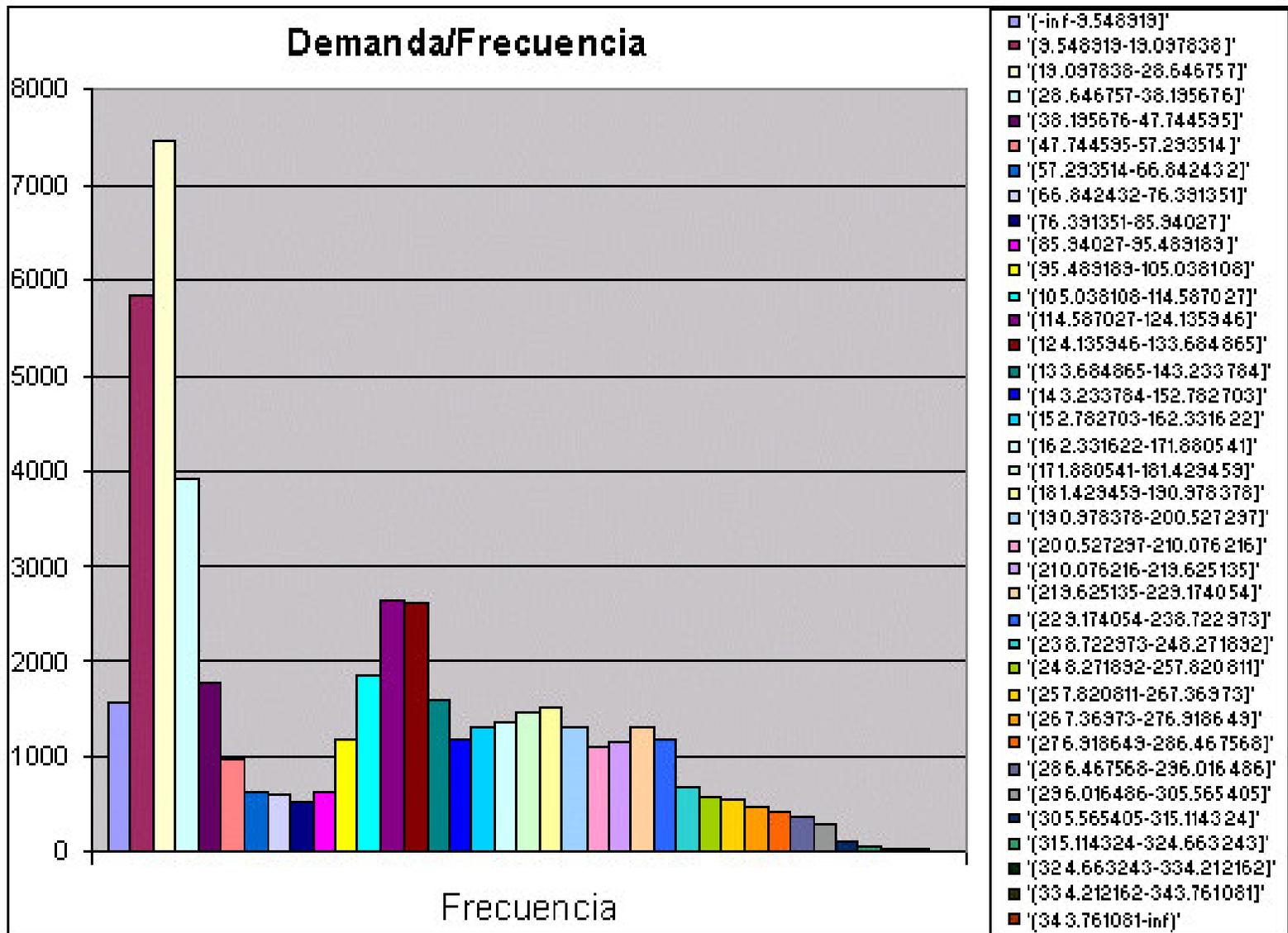


Figura 4. Histograma de frecuencia (Atlántico)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Cuadro 7. Reglas Atlántico

Análisis Individual

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
		(76,39-85,94)					(210,07-219,62)	(190,97-200,52)					
Bajo 1 (1-4)													
Bajo 2 (24)							(0-4279,60)						
Medio 1 (5-9)													
Medio 2 (13-18)	(324,66-334,21)	(248,27-257,82)											
Medio 3 (22-23)							(334,21-343,76)	(296,01-305,56)					
Pico 1 (10-12)													
Pico 2 (19-21)							(324,66-334,21)						
Lunes													
Bajo 1 (1-4)		(0-9,54)	(57,29-66,84)						(105,03-114,58)				
Bajo 2 (24)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(85,94- 95,48)			(57,29- 66,84)				(286,46- 296,01)					
Medio 2 (13-18)	(315,11- 324,66)	(257,82- 267,86)	(276,91- 286,46)	(0- 4279,60)		(0- 4279,60)			(0- 4279,60)			(57,29- 66,84)	
Medio 3 (22-23)							(267,36- 276,91)	(229,17- 238,72)					
Pico 1 (10-12)													
Pico 2 (19-21)													
Martes	(95,48- 105,03)												
Bajo 1 (1-4)			(57,29- 66,84)		(76,39- 85,94)					(114,58- 124,13)			(0-9,54)
Bajo 2 (24)													
Medio 1 (5-9)	(76,39- 85,94)												
Medio 2 (13-18)				(0- 4279,60)					(0- 4279,60)				
Medio 3 (22-23)													
Pico 1 (10-12)	(286,46- 296,01)	(0-9,54)											

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)			(286,46- 296,01)					(219,62- 229,17)					
Miércoles													
Bajo 1 (1-4)								(219,62- 229,17)			(114,58- 124,13)	(76,39- 85,94)	
Bajo 2 (24)													
Medio 1 (5-9)	(76,39- 85,94)												
Medio 2 (13-18)		(0- 4279,60)		(267,36- 276,91)	(0- 4279,60)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)			
Medio 3 (22-23)						(105,03- 114,58)							
Pico 1 (10-12)		(0-9,54)											
Pico 2 (19-21)													
Jueves													
Bajo 1 (1-4)							(66,84- 76,39)		(76,39- 85,94)	(57,29- 66,84)		(124,13- 133,68)	
Bajo 2 (24)	(0- 4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(76,39,85 ,94)												
Medio 2 (13-18)	(315,11- 324,66)				(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)		(0- 4279,60)			
Medio 3 (22-23)								(286,46- 296,01)					
Pico 1 (10-12)				(0- 4279,60)					(0- 4279,60)		(0-9,54)		
Pico 2 (19-21)								(57,29- 66,84)					
Viernes	(181,42- 190,97)						(343,76- 353,31)						
Bajo 1 (1-4)			(76,39- 85,94)			(57,29- 66,84)				(124,13- 133,68)			
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)	(85,94- 95,48)												
Medio 2 (13-18)	(343,76- 353,31)	(0- 4279,60)		(0- 4279,60)	(0- 4279,60)		(343,76- 353,31)	(343,76- 353,31)					(238,72- 248,27)
Medio 3 (22-23)								(238,72- 248,27)		(38,19- 47,74)			
Pico 1 (10-12)		(0-9,54)											

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)									(219,62- 229,17)		(57,29- 66,84)		
Sábado	(162,33- 171,88)						(334,21- 343,76)	(276,91- 286,46)					
Bajo 1 (1-4)					(66,84- 76,39)			(248,27- 257,82)					(105,03- 114,58)
Bajo 2 (24)													
Medio 1 (5-9)	(85,94- 95,48)		(162,33- 171,88)										
Medio 2 (13-18)	(334,21- 343,76)	(0-9,54)	(66,84- 76,39)			(0- 4279,60)		(0- 4279,60)				(85,94- 95,48)	
Medio 3 (22-23)	(334,21- 343,76)							(248,27- 257,82)					
Pico 1 (10-12)								(47,74- 57,29)					
Pico 2 (19-21)													
Domingo								(248,27- 257,82)					
Bajo 1 (1-4)		(0-9,54)				(114,58- 124,13)							(57,29- 66,84)
Bajo 2 (24)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)													(47,74- 57,29)
Medio 2 (13-18)				(114,58- 124,13)		(0- 4279,60)		(85,94- 95,48)	(0- 4279,60)	(19,09- 28,64)			
Medio 3 (22-23)													
Pico 1 (10-12)	(85,94- 95,48)									(38,19- 47,74)			
Pico 2 (19-21)										(9,54- 19,09)	(47,74- 57,29)		

Gráficos

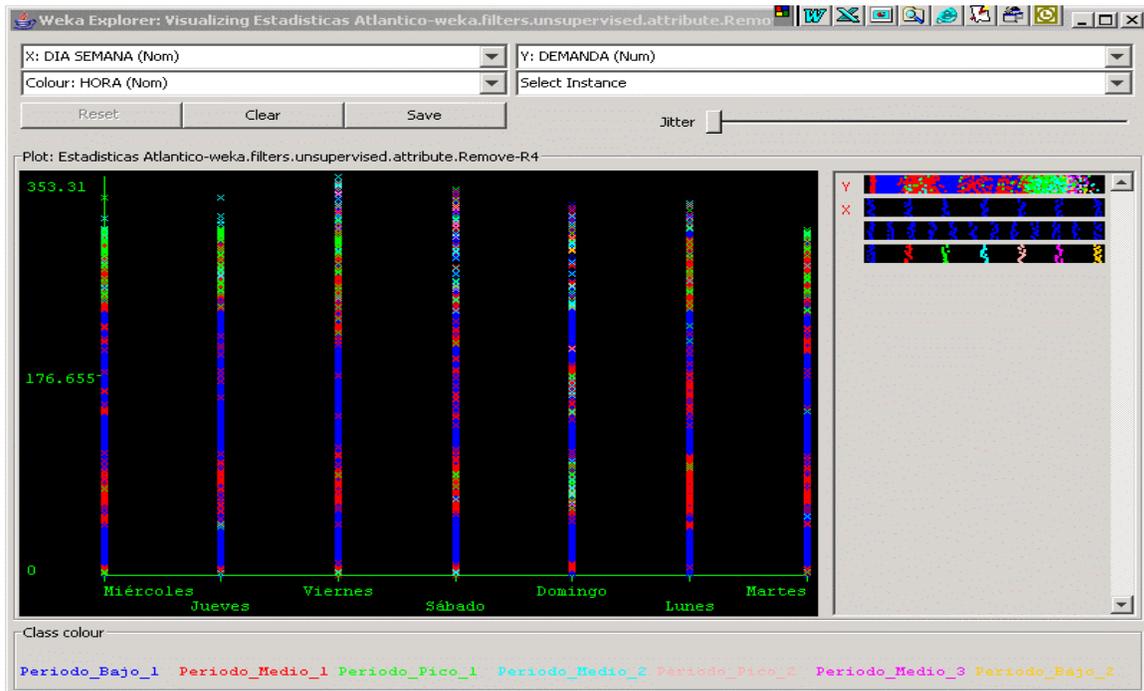


Figura 5. Demanda días de la semana diferenciada por periodos del día (Atlántico)

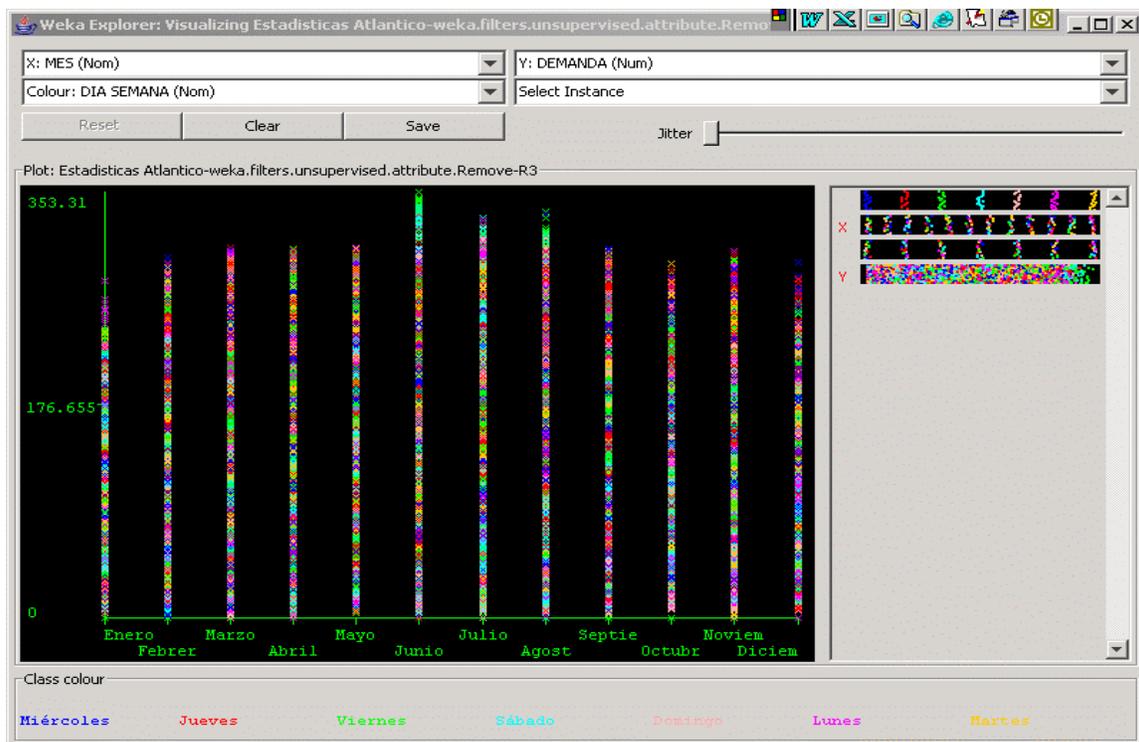


Figura 6. Demanda meses del año diferenciada por días de la semana (Atlántico)

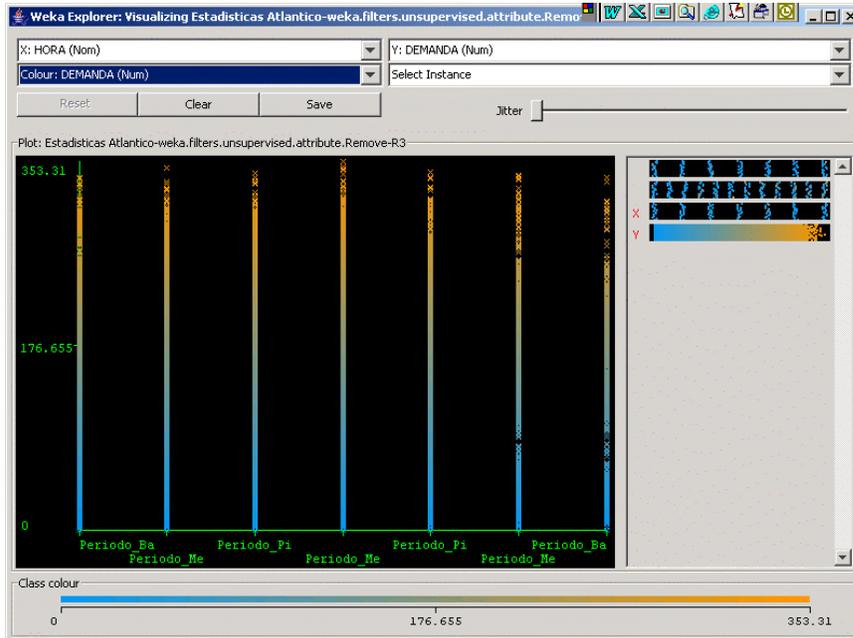


Figura 7. Demanda periodos del día (Atlántico)

2.3.5 ANÁLISIS INDIVIDUAL

Para este tipo de análisis la metodología seguida consistió básicamente en corridas de la aplicación con las especificaciones descritas anteriormente para el análisis conjunto, difiriendo de éste, en que se trabajan con los departamentos de manera independiente, profundizando así el detalle y precisión del análisis, y de esta forma mejorando la calidad del mismo.

2.3.5.1 Análisis Departamento: ATLANTICO

Atlántico, como se mencionó anteriormente, es el departamento de la Costa Atlántica con la menor demanda de energía eléctrica del sector Minero; con media: 105,95 kW/h y desviación estándar: 83,41 kW/h. Con la mayor proporción de demanda concentrada en los rangos:

(9,54-19,09) kW/h, que representa el 11,65% de su demanda,

(19,09-28,64) kW/h, que representa el 14,82% de su demanda,
(28,64-38,19) kW/h, que representa el 7,78% de su demanda,

Correspondiendo estos intervalos en su conjunto, al 34,25% de la demanda del departamento; en cuanto al resto de valores, éstos se encuentran distribuidos de manera casi plana y uniforme, en porcentajes relativamente pequeños que van desde el 0,01% al 5,24% de la demanda total del departamento en este sector.

Los datos arrojados por las reglas de asociación y los gráficos construidos por la aplicación, nos dejan saber que los días de la semana donde se presentan los valores de demanda más altos en el Atlántico, son el viernes (D. Media: 181,42-190,97) kW/h, cuya máxima se presenta durante el periodo medio 2 (343,76-353,31) kW/h, seguido por el sábado (D. Media: 162,33-171,88) kW/h y cuya demanda máxima se presenta durante los periodos medio 2 y medio 3 (334,21-343,76) kW/h. Mientras tanto, el día de la semana donde se presenta la menor demanda del departamento es el martes (D. Media: 95,48-105,03) kW/h, cuya demanda máxima se presenta exactamente durante el periodo pico 1 (286,46-296,01) kW/h.

En cuanto al comportamiento de la demanda por periodos del día, resulta clara la tendencia del mercado periodo bajo 1 de tomar valores de demanda tanto en rangos bajos:

(0-9,54) kW/h

(57,29-66,84) kW/h

(66,84-76,39) kW/h

(76,39-85,94) kW/h

como en medios altos:

(105,03-114,58) kW/h

(114,58-124,13) kW/h

(124,13-133,68) kW/h

Con mayor frecuencia de ocurrencia; desmintiendo de esta forma la concepción de que en este periodo solo se dan valores de demanda bajos; y a la vez corroborando la conclusión expuesta en el caso general, de ser este un periodo de comportamiento especial.

Otro caso que resulta interesante mencionar, del cual las reglas y lo gráficos hacen clara alusión en el comportamiento de la demanda de este departamento, es el del periodo medio 1, el cual presenta mayor frecuencia de ocurrencia dentro de unos mismos rangos de valores (76,39-85,94) kW/h y (85,94-95,48) kW/h existiendo valores de intersección con el periodo bajo 1, en todos los días de la semana, exceptuando el domingo; día en el cual el rango de valores propio de este periodo es tomado durante el periodo pico 1.

En términos generales, durante el periodo medio 2 se presentan los valores promedio de demanda más altos del departamento, exactamente durante los días: lunes y jueves (D. Máxima: 315,11-324,66) kW/h, viernes (D. Máxima: 343,76-353,31) kW/h y sábado (D. Máxima: 334,21-343,76) kW/h.

El mes del año con la más alta demanda del Atlántico en el sector minero es el mes de Junio (D. Media: 210,07-219,62) kW/h y cuya máxima se presenta los viernes (343,76-353,31) kW/h; seguido por el mes de Julio con demanda promedio de (190,97-200,52) kW/h,.

Por otro lado, el mes de menor demanda es Enero (D. Media: 76,39-85,94), cuya demanda máxima se presentan los días lunes durante el periodo medio 2 (257,82-267,86) kW/h.

Para concluir, unos valores interesantes de la demanda del departamento:

Demanda Máxima: 353,31 kW/h, propia de los viernes de Junio (el día de la semana y el mes del año con mayor demanda).

Demanda Mínima: (0-0,94), propia de:

Lunes, Periodo Bajo 1, Enero
 Martes, Periodo Pico 1 y Bajo 1, Enero y Diciembre respectivamente
 Miércoles: Periodo Pico 1, Enero
 Jueves: Periodo Pico 1, Octubre
 Viernes: Periodo Pico 1, Enero
 Sábado: Periodo Medio 2, Enero
 Domingo: Periodo Bajo 1, Enero.

2.3.5.2 Resultados Departamento: BOLIVAR

Resumen De Estadísticas

Cuadro 8. Estadísticas Bolívar

Total de Datos	29.075	
Máximo	1.467,34	
Mínimo	0	
Rango	1.467,34	
No. Clases	35,12	36
Ancho Intervalo	40,75	
Media	503,04	
Desviación	287,33	

Cuadro 9. Distribución de frecuencia Bolívar

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-40,759444]'	4.604	15,83
'(40,759444-81,518889]'	278	0,96
'(81,518889-122,278333]'	55	0,19
'(122,278333-163,037778]'	50	0,17
'(163,037778-203,797222]'	47	0,16
'(203,797222-244,556667]'	48	0,17
'(244,556667-285,316111]'	67	0,23

DEMANDA	FRECUENCIA	PORCENTAJE %
'(285,316111-326,075556]'	103	0,35
'(326,075556-366,835]'	459	1,58
'(366,835-407,594444]'	7.761	26,69
'(407,594444-448,353889]'	2.269	7,80
'(448,353889-489,113333]'	17	0,06
'(489,113333-529,872778]'	31	0,11
'(529,872778-570,632222]'	31	0,11
'(570,632222-611,391667]'	169	0,58
'(611,391667-652,151111]'	835	2,87
'(652,151111-692,910556]'	857	2,95
'(692,910556-733,67]'	1.483	5,10
'(733,67-774,429444]'	4.048	13,92
'(774,429444-815,188889]'	3.116	10,72
'(815,188889-855,948333]'	1.985	6,83
'(855,948333-896,707778]'	327	1,12
'(896,707778-937,467222]'	41	0,14
'(937,467222-978,226667]'	32	0,11
'(978,226667-1.018,986111]'	35	0,12
'(1.018,986111-1.059,745556]'	31	0,11
'(1.059,745556-1.100,505]'	18	0,06
'(1.100,505-1141,2.64444]'	6	0,02
'(1.141,264444-1.182,023889]'	3	0,01
'(1.182,023889-1.222,783333]'	15	0,05
'(1.222,783333-1.263,542778]'	54	0,19
'(1.263,542778-1.304,302222]'	45	0,15
'(1.304,302222-1.345,061667]'	50	0,17
'(1.345,061667-1.385,821111]'	51	0,18
'(1.385,821111-1.426,580556]'	42	0,14
'(1.426,580556-1.467,34]'	12	0,04

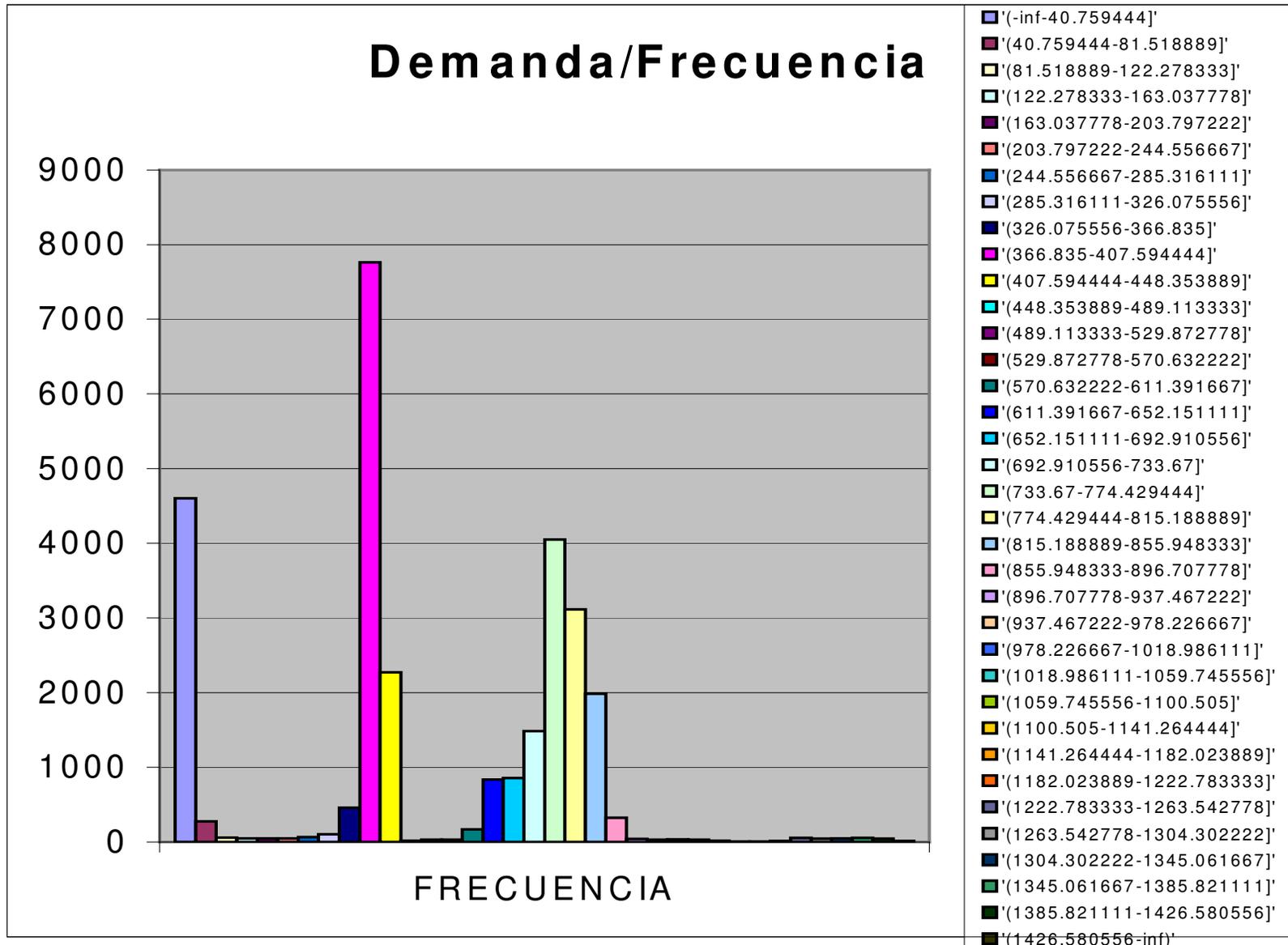


Figura 8. Histograma de frecuencia (Bolívar)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Cuadro 10. Reglas Bolívar

Análisis Individual

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
						(896.70-937.46)				(448.35-489.11)	(489.11-529.87)	(448.35-489.11)	
Bajo 1 (1-4)				(0-40.75)	(570.63-611.39)			(815.18-855.94)			(0-4279.60)	(0-4279.60)	(0-4279.60)
Bajo 2 (24)								(815.18-855.94)					
Medio 1 (5-9)					(529.87-570.63)	(1182.02-1222.78)							
Medio 2 (13-18)	(733.67-774.42)							(896.70-937.46)		(937.46-978.22)		(163.03-203.79)	
Medio 3 (22-23)	(611.39-652.15)	(0-4279.60)	(122.27-163.03)	(489.11-529.87)	(733.67-774.42)			(489.11-529.87)	(489.11-529.87)	(203.79-244.55)	(0-4279.60)	(448.35-489.11)	(448.35-489.11)
Pico 1 (10-12)		(0-4279.60)	(0-4279.60)	(0-4279.60)	(652.15-692.91)			(815.18-855.94)			(0-4279.60)	(0-4279.60)	(0-4279.60)
Pico 2 (19-21)		(0-4279.60)	(529.87-570.63)	(0-4279.60)		(611.39-652.15)		(448.35-489.11)		(652.15-692.91)	(529.87-570.63)	(529.87-570.63)	(529.87-570.63)
Lunes	(815.18-855.94)	(611.39-652.15)									(611.39-652.15)	(611.39-652.15)	(0-4279.60)
Bajo 1 (1-4)					(0-40.75)			(774.42-815.18)				(570.63-611.39)	
Bajo 2 (24)	(0-4279.60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)												(652.15- 692.91)	
Medio 2 (13-18)		(652.15- 692.91)							(40.75- 81.51)		(978.22- 1018.98)	(652.15- 692.91)	
Medio 3 (22-23)	(1385.82- 1426.58)	(326.07- 366.83)				(489.11- 529.87)		(855.94- 896.70)				(448.35- 489.11)	
Pico 1 (10-12)					(652.15- 692.91)				(692.91- 733.67)				
Pico 2 (19-21)													(652.15- 692.91)
Martes	(855.94- 896.70)	(0- 4279.60)									(0- 4279.60)	(244.55- 285.31)	(0- 4279.60)
Bajo 1 (1-4)				(0-40.75)		(570.63- 611.39)		(774.42- 815.18)				(244.55- 285.31)	
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)						(285.31- 326.07)						(203.79- 244.55)	
Medio 2 (13-18)	(1059.74- 1100.50)												
Medio 3 (22-23)	(1426.58- 1467.34)					(285.31- 326.07)		(855.94- 896.70)			(1100.50- 1141.26)		
Pico 1 (10-12)						(285.31- 326.07)			(652.15- 692.91)			(203.79- 244.55)	

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)						(285.31- 326.07)					(652.15- 692.91)		(652.15- 692.91)
Miércoles		(0- 4279.60)									(0- 4279.60)	(896.70- 937.46)	(0- 4279.60)
Bajo 1 (1-4)	(570.63- 611.39)		(0-40.75)					(774.42- 815.18)					
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)					(570.63- 611.39)								
Medio 2 (13-18)	(1426.58- 1467.34)				(40.75- 81.51)	(1385.82- 1426.58)		(855.94- 896.70)				(1018.98- 1059.74)	
Medio 3 (22-23)	(1385.82- 1426.58)				(733.67- 774.42)			(855.94- 896.70)					
Pico 1 (10-12)									(692.91- 733.67)				
Pico 2 (19-21)											(652.15- 692.91)	(652.15- 692.91)	
Jueves		(0- 4279.60)				(1182.02- 1222.78)		(855.94- 896.70)			(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)				(570.63- 611.39)				(774.42- 815.18)		(0-40.75)			
Bajo 2 (24)	(0- 4279.60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)													
Medio 2 (13-18)						(652.15- 692.91)						(40.75- 81.51)	
Medio 3 (22-23)								(855.94- 896.70)					
Pico 1 (10-12)					(652.15- 692.91)				(692.91- 733.67)				
Pico 2 (19-21)								(652.15- 692.91)			(652.15- 692.91)	(652.15- 692.91)	(652.15- 692.91)
Viernes	(529.87- 570.63)	(0- 4279.60)				(1182.02- 1222.78)		(855.94- 896.70)			(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)						(1182.02- 1222.78)		(774.42- 815.18)		(570.63- 611.39)			(0-40.75)
Bajo 2 (24)	(0- 4279.60)						(652.15- 692.91)						
Medio 1 (5-9)													
Medio 2 (13-18)	(1385.82- 1426.58)										(40.75- 81.51)		
Medio 3 (22-23)					(652.15- 692.91)					(40.75- 81.51)			
Pico 1 (10-12)					(652.15- 692.91)				(692.91- 733.67)				

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)								(652.15- 692.91)			(652.15- 692.91)	(652.15- 692.91)	(652.15- 692.91)
Sábado			(529.87- 570.63)			(529.87- 570.63)		(896.70- 937.46)		(1018.98- 1059.74)	(529.87- 570.63)	(529.87- 570.63)	(529.87- 570.63)
Bajo 1 (1-4)	(774.42- 815.18)		(774.42- 815.18)	(774.42- 815.18)	(570.63- 611.39)	(774.42- 815.18)	(774.42- 815.18)	(774.42- 815.18)	(0-40.75)	(774.42- 815.18)	(774.42- 815.18)	(774.42- 815.18)	
Bajo 2 (24)	(815.18- 855.94)			(774.42- 815.18)				(815.18- 855.94)			(774.42- 815.18)	(774.42- 815.18)	
Medio 1 (5-9)	(815.18- 855.94)	(774.42- 815.18)	(815.18- 855.94)		(815.18- 855.94)			(815.18- 855.94)			(774.42- 815.18)	(774.42- 815.18)	
Medio 2 (13-18)	(570.63- 611.39)		(855.94- 896.70)		(855.94- 896.70)			(896.70- 937.46)					
Medio 3 (22-23)	(611.39- 652.15)				(855.94- 896.70)	(285.31- 326.07)		(896.70- 937.46)					
Pico 1 (10-12)	(815.18- 855.94)	(774.42- 815.18)		(774.42- 815.18)		(244.31- 326.07)	(774.42- 815.18)	(815.18- 855.94)	(774.42- 815.18)	(774.42- 815.18)	(774.42- 815.18)	(774.42- 815.18)	
Pico 2 (19-21)	(570.63- 611.39)	(570.63- 611.39)	(529.87- 570.63)	(692.91- 733.67)	(692.91- 733.67)	(529.87- 570.63)	(692.91- 733.67)	(652.15- 692.91)		(570.63- 611.39)	(652.15- 692.91)	(529.87- 570.63)	(529.87- 570.63)
Domingo						(529.87- 570.63)				(937.46- 978.22)	(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)	(692.91- 733.67)	(570.63- 611.39)		(692.91- 733.67)			(0-40.75)	(692.91- 733.67)	(774.42- 815.18)	(692.91- 733.67)	(692.91- 733.67)	(692.91- 733.67)	(692.91- 733.67)
Bajo 2 (24)	(0- 4279.60)		(733.67- 774.42)		(733.67- 774.42)	(733.67- 774.42)	(733.67- 774.42)	(733.67- 774.42)					

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)			(733.67- 774.42)			(570.63- 611.39)							
Medio 2 (13-18)	(1263.54- 1304.30)		(733.67- 774.42)	(774.42- 815.18)		(733.67- 774.42)	(774.42- 815.18)	(855.94- 896.70)					
Medio 3 (22-23)	(1304.30- 1345.06)					(774.42- 815.18)		(855.94- 896.70)		(774.42- 815.18)		(774.42- 815.18)	
Pico 1 (10-12)		(692.91- 733.67)		(692.91- 733.67)		(692.91- 733.67)	(692.91- 733.67)	(692.91- 733.67)		(692.91- 733.67)	(692.91- 733.67)	(692.91- 733.67)	
Pico 2 (19-21)	(815.18- 855.94)		(855.94- 896.70)	(855.94- 896.70)	(855.94- 896.70)	(855.94- 896.70)	(855.94- 896.70)	(855.94- 896.70)	(774.42- 815.18)			(855.94- 896.70)	(855.94- 896.70)

Gráficos

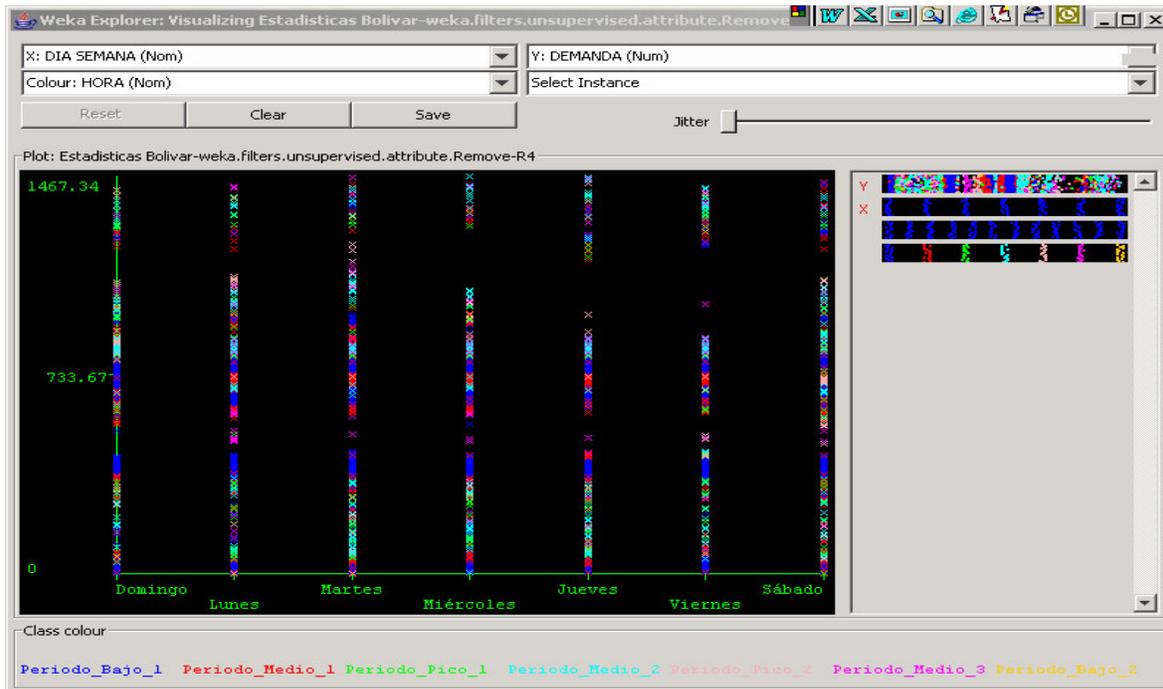


Figura 9. Demanda dias de la semana diferenciada por periodos del dia (Bolívar)

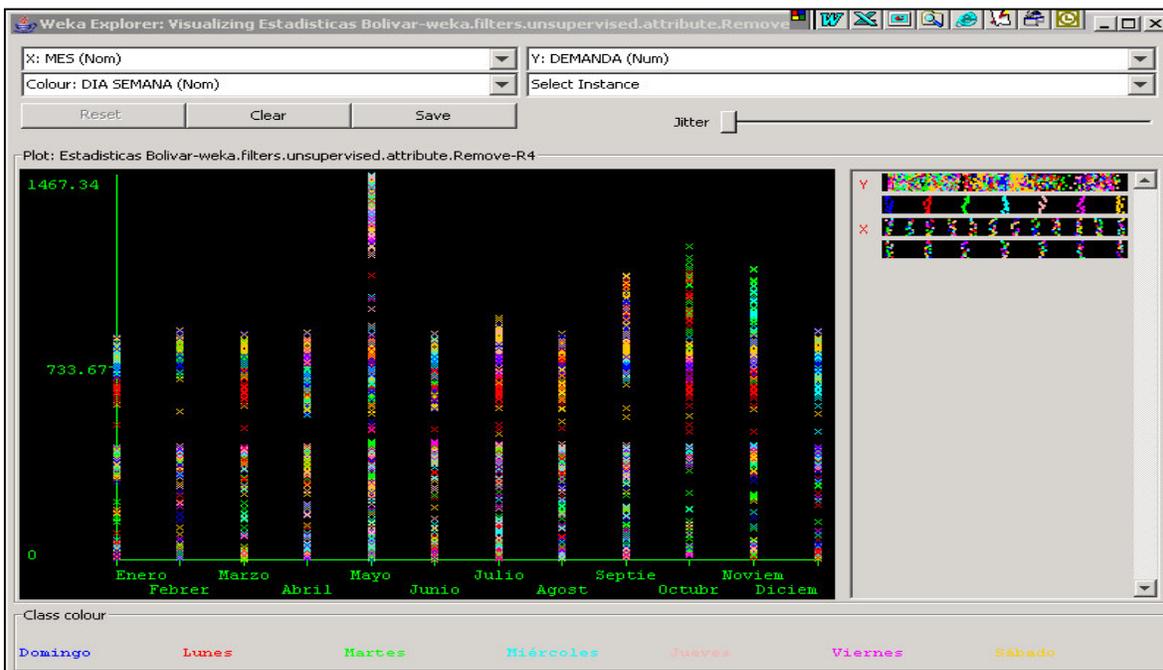
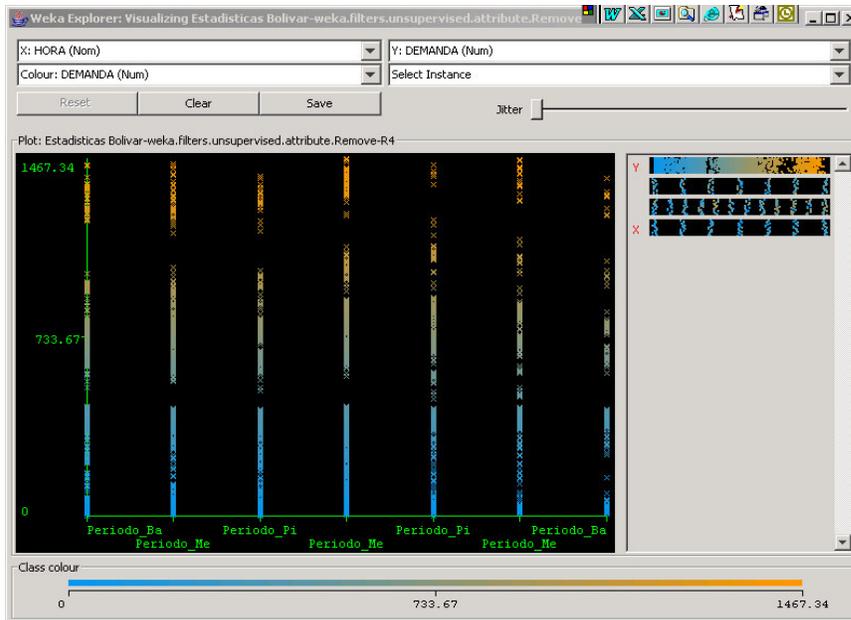


Figura 10. Demanda meses del año diferenciada por periodos del día (Bolívar)



**ORDEN
SECUENCIAL**

P. Bajo 1: (1–4)
P. Medio 1: (5–9)
P. Pico 1: (10–12)
P. Medio 2: (13–18)
P. Pico 2: (19–21)
P. Medio 3: (22–23)
P. Bajo 2: (24)

Figura 11. Demanda periodos del día (Bolívar)

2.3.5.3 Análisis Departamento: BOLIVAR

La demanda del departamento de Bolívar es considerada como una de las más bajas del sector; para ser exactos es la tercera más baja, con valor medio de 503,04 kW/h y desviación estándar igual a 287,33 kW/h. Concentrada en las más altas proporciones en los siguientes intervalos:

- (0-40,75) kW/h en un 15,83%,
- (366,83-407,59) kW/h en un 26,69%,
- (733,67-774,42) kW/h en un 13,92%,
- (774,42-815,18) kW/h en un 10,72%,

Representando en su conjunto el 67,17% de la demanda del sector minero en Bolívar.

El día de la semana con la mayor demanda promedio es el martes (D. Media: 855,94-896,70) kW/h y (D. Máxima: 1.426,58-1.467,34) kW/h presente durante el periodo medio 3, según datos arrojados por las reglas de asociación. Le sigue muy

de cerca el día lunes con (D. Media: 815,18-855,94) kW/h, cuya máxima se presenta igualmente durante el periodo medio 3 (1.385,82-1.426,58) kW/h.

Mientras tanto, durante los viernes se presenta la menor demanda promedio del sector en este departamento (529,87-570,63) kW/h, y la demanda máxima del día se presenta durante el periodo medio 2 (1.385,82-1.426,58) kW/h.

En cuanto a los periodos del día, el de mayor demanda promedio es el periodo medio 2 (D. Media: 733,67-774,42) kW/h, seguido por el periodo medio 3 (D. Media: 611,39-652,15) kW/h; ambos con demanda máxima de (1.426,58-1.467,34) kW/h, durante los días miércoles y martes respectivamente.

Durante el periodo bajo 1 se presentan los valores de demanda más bajos del departamento, es decir, los rangos correspondientes a (0-40,75) kW/h y (40,75-81,51) kW/h independiente del día de la semana; situación que resulta clara visualizar a través de la Figura 9. Demanda días de la semana diferenciada por periodo del día; aún cuando existen valores de intersección con otros periodos, relacionados directamente al correspondiente día de la semana.

A demás de los valores inferiores; durante el periodo bajo 1 la demanda tiene tendencia a tomar valores tanto medios (570,63-611,39) kW/h como medios altos (774,42-815,18) kW/h, en mayores proporciones en cuanto a frecuencia de ocurrencia.

El mes del año que en grandes proporciones tiene la mayor demanda promedio de energía en este departamento, es Mayo, con demanda media de (896,70-937,46) kW/h, le sigue Octubre (D. Media: 489,11-529,87) kW/h, Septiembre y Noviembre con igual demanda promedio (448,35-489,11) kW/h; mientras que los meses Enero, Febrero, Marzo, Abril, Junio, Agosto, y Diciembre, presentan una demanda media que resulta casi plana, y cuya máxima general llega a oscilar alrededor de los valores contenidos en el intervalo (855,94-896,70) kW/h.

2.3.5.4 Resultados Departamento: CESAR

Resumen De Estadísticas

Cuadro11. Estadísticas Cesar

Total de Datos	26.046	
Máximo	2.710,83	
Mínimo	0	
Rango	2.710,83	
No. De Clases	34,75	35
Media	1.435,84	
Desviación	558,89	

Cuadro 12. Distribución de frecuencia Cesar

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-77,452286]'	517	1,98
'(77,452286-154,904571]'	1.325	5,09
'(154,904571-232,356857]'	23	0,09
'(232,356857-309,809143]'	26	0,10
'(309,809143-387,261429]'	52	0,20
'(387,261429-464,713714]'	63	0,24
'(464,713714-542,166]'	80	0,31
'(542,166-619,618286]'	130	0,50
'(619,618286-697,070571]'	235	0,90
'(697,070571-774,522857]'	414	1,59
'(774,522857-851,975143]'	556	2,13
'(851,975143-929,427429]'	711	2,73
'(929,427429-1.006,879714]'	830	3,19
'(1.006,879714-1.084,332]'	934	3,59
'(1.084,332-1.161,784286]'	978	3,75
'(1.161,784286-1.239,236571]'	998	3,83
'(1.239,236571-1.316,688857]'	1.183	4,54

DEMANDA	FRECUENCIA	PORCENTAJE %
'(1.316,688857-1.394,141143]'	1.287	4,94
'(1.394,141143-1.471,593429]'	1.389	5,33
'(1.471,593429-1.549,045714]'	1.631	6,26
'(1.549,045714-1.626,498]'	1.733	6,65
'(1.626,498-1.703,950286]'	1.818	6,98
'(1.703,950286-1.781,402571]'	1.672	6,42
'(1.781,402571-1.858,854857]'	1.600	6,14
'(1.858,854857-1.936,307143]'	1.401	5,38
'(1.936,307143-2.013,759429]'	1.215	4,66
'(2.013,759429-2.091,211714]'	945	3,63
'(2.091,211714-2.168,664]'	757	2,91
'(2.168,664-2.246,116286]'	634	2,43
'(2.246,116286-2.323,568571]'	385	1,48
'(2.323,568571-2.401,020857]'	264	1,01
'(2.401,020857-2.478,473143]'	161	0,62
'(2.478,473143-2.555,925429]'	63	0,24
'(2.555,925429-2.633,377714]'	27	0,10
'(2.633,377714-2.710,83)'	9	0,03

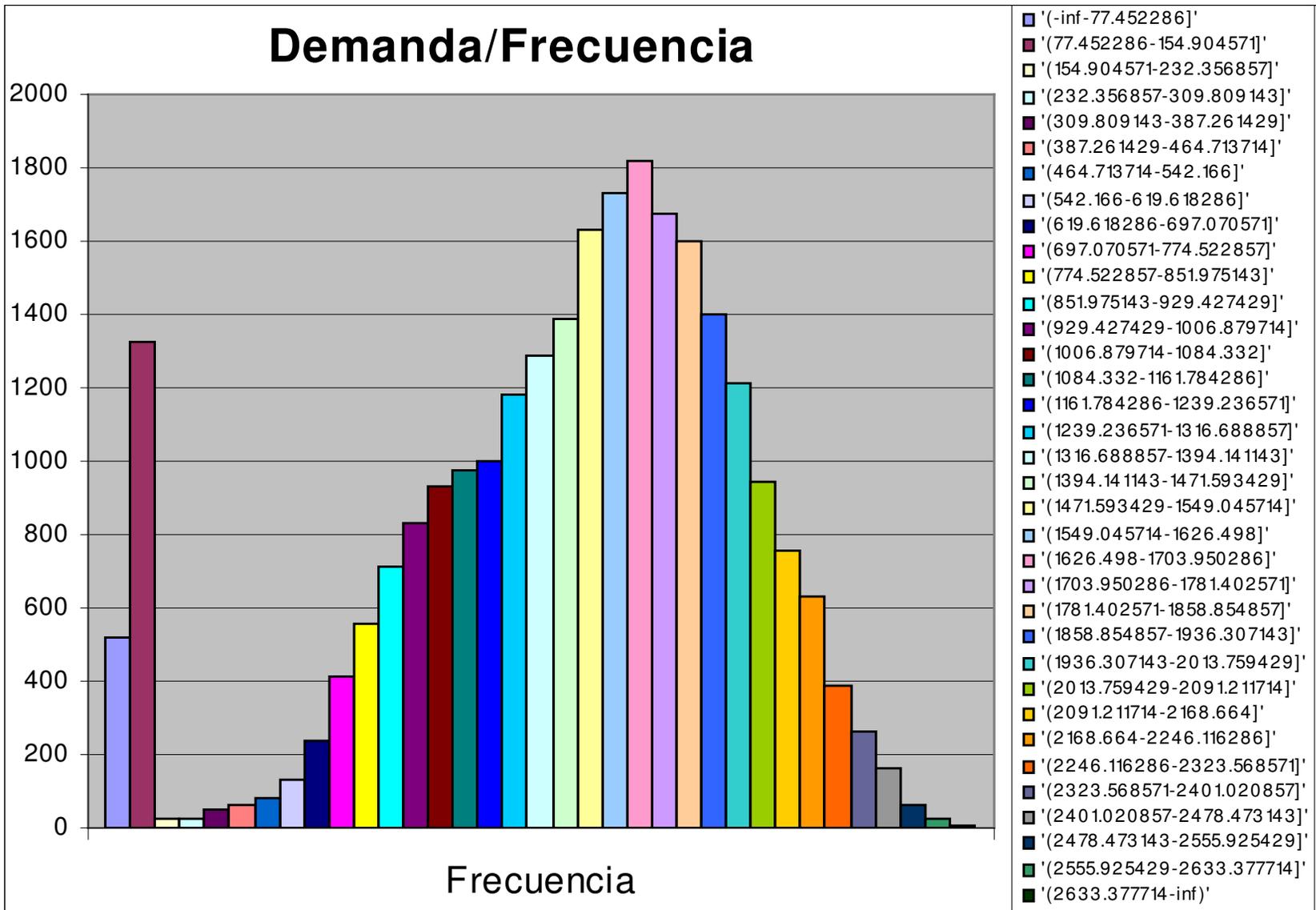


Figura 12. Histograma de frecuencia (Cesar)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Cuadro 13. Reglas Cesar

Análisis Individual

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
					(1936.30-2013.75)	(1703.95-1781.40)		(1084.33-1161.78)		(929.42-1006.87)			
Bajo 1 (1-4)											(0-4279.60)	(0-4279.60)	(0-4279.60)
Bajo 2 (24)													
Medio 1 (5-9)	(1006.87-1084.33)												
Medio 2 (13-18)	(851.97-929.42)												
Medio 3 (22-23)		(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)			(0-4279.60)	(0-4279.60)	(0-4279.60)
Pico 1 (10-12)	(1626.49-1703.95)	(0-4279.60)	(0-4279.60)	(232.35-309.80)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)			(0-4279.60)	(0-4279.60)	(0-4279.60)
Pico 2 (19-21)	(1549.04-1626.49)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)	(0-4279.60)			(0-4279.60)	(0-4279.60)	(0-4279.60)
Lunes	(1239.23-1316.68)	(0-4279.60)				(2633.37-2710.83)					(0-4279.60)	(0-4279.60)	(0-4279.60)
Bajo 1 (1-4)			(387.26-464.71)	(542.16-619.61)			(774.52-851.97)			(0-77.45)			
Bajo 2 (24)	(0-4279.60)					(2246.11-2323.56)							

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)										(619.61- 697.07)			
Medio 2 (13-18)				(2013.75- 2091.21)									
Medio 3 (22-23)							(0-77.45)						
Pico 1 (10-12)													
Pico 2 (19-21)	(2633.37- 2710.83)												
Martes	(1394.14- 1471.59)	(0- 4279.60)									(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)							(0-77.45)						
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)							(0-77.45)						
Medio 2 (13-18)									(0-77.45)				
Medio 3 (22-23)													
Pico 1 (10-12)	(2633.37- 2710.83)	(0-77.45)											

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)										(2168.66- 2246.11)			
Miércoles	(1471.59- 1549.04)	(0- 4279.60)			(2633.37- 2710.83)			(1858.85- 1936.30)			(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)													
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)		(774.52- 851.97)											
Medio 2 (13-18)	(2633.37- 2710.83)												
Medio 3 (22-23)	(387.26- 464.71)				(2401.02- 2478.47)								(929.42- 1006.87)
Pico 1 (10-12)													
Pico 2 (19-21)						(0-77.45)							
Jueves	(1239.23- 1316.68)	(0- 4279.60)						(0-77.45)			(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)		(387.26- 464.71)								(0-77.45)			
Bajo 2 (24)	(0- 4279.60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)								(542.16- 619.61)					
Medio 2 (13-18)													
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)	(2633.37- 2710.83)							(0-77.45)					
Viernes	(1239.23- 1316.68)									(1858.85- 1936.30)		(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)													
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)													
Medio 2 (13-18)								(697.07- 774.52)		(1936.30- 2013.75)	(2091.21- 2168.66)		
Medio 3 (22-23)		(542.16- 619.61)											(542.16- 619.61)
Pico 1 (10-12)	(2633.37- 2710.83)		(774.52- 851.97)									(2401.02- 2478.47)	

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)													(542.16- 619.61)
Sábado											(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)		(464.71- 542.16)											
Bajo 2 (24)	(0- 4279.60)												
Medio 1 (5-9)					(619.61- 697.07)	(2401.02- 2478.47)							(1006.87- 1084.33)
Medio 2 (13-18)		(464.71- 542.16)											
Medio 3 (22-23)								(619.61- 697.07)	(0-77.45)				
Pico 1 (10-12)							(2323.56- 2401.02)						
Pico 2 (19-21)									(0-77.45)				
Domingo	(1084.33- 1161.78)	(0- 4279.60)									(0- 4279.60)	(0- 4279.60)	(0- 4279.60)
Bajo 1 (1-4)									(0-77.45)			(0-77.45)	
Bajo 2 (24)	(0- 4279.60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)											(0-77.45)		
Medio 2 (13-18)													
Medio 3 (22-23)				(464.71- 542.16)			(697.07- 774.52)			(0-77.45)			
Pico 1 (10-12)	(2478.47- 2555.92)										(2091.21- 2168.66)		
Pico 2 (19-21)													

Gráficos

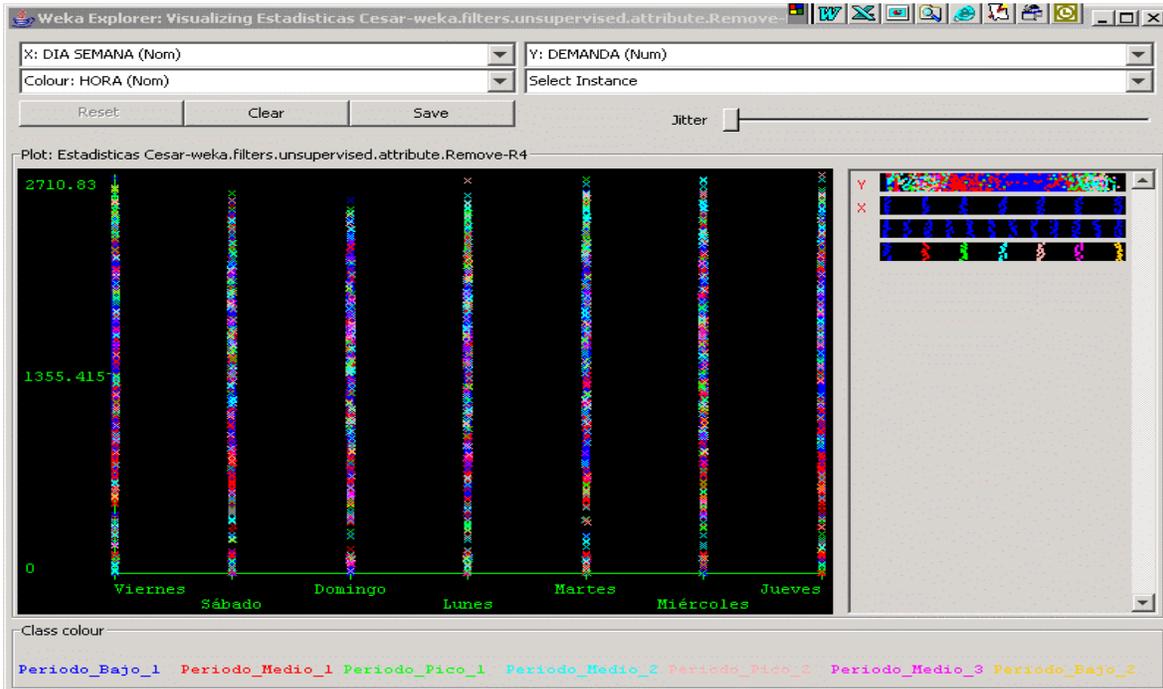


Figura 13. Demanda días de la semana diferenciada por periodos del día (Cesar)

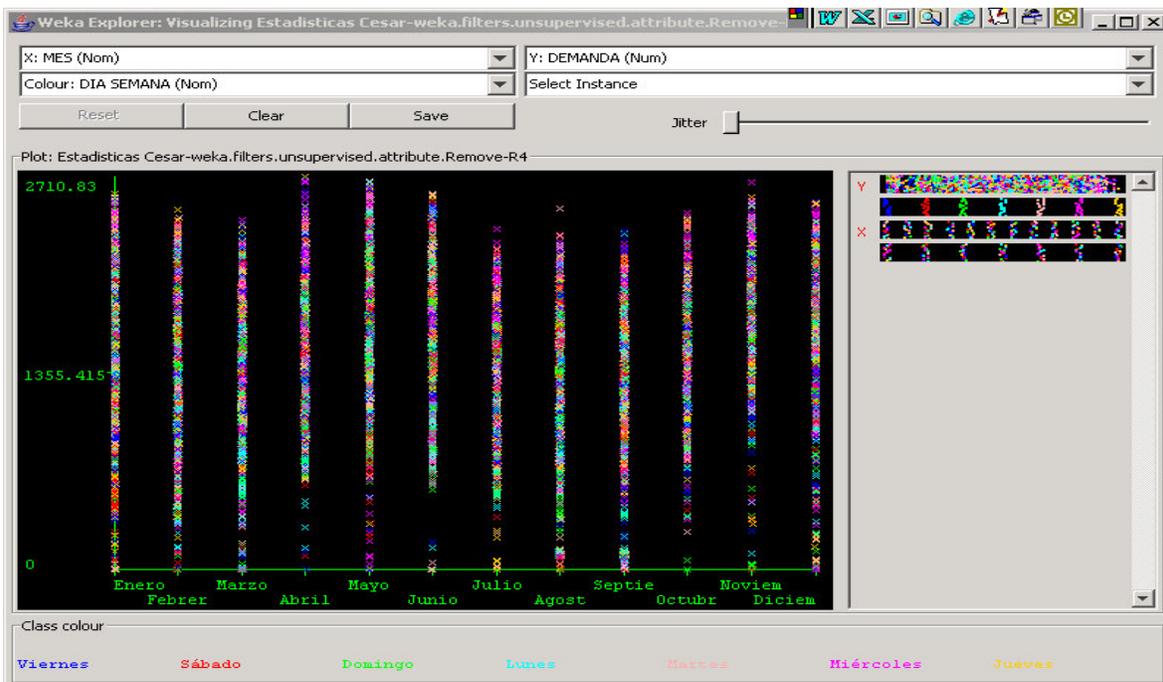
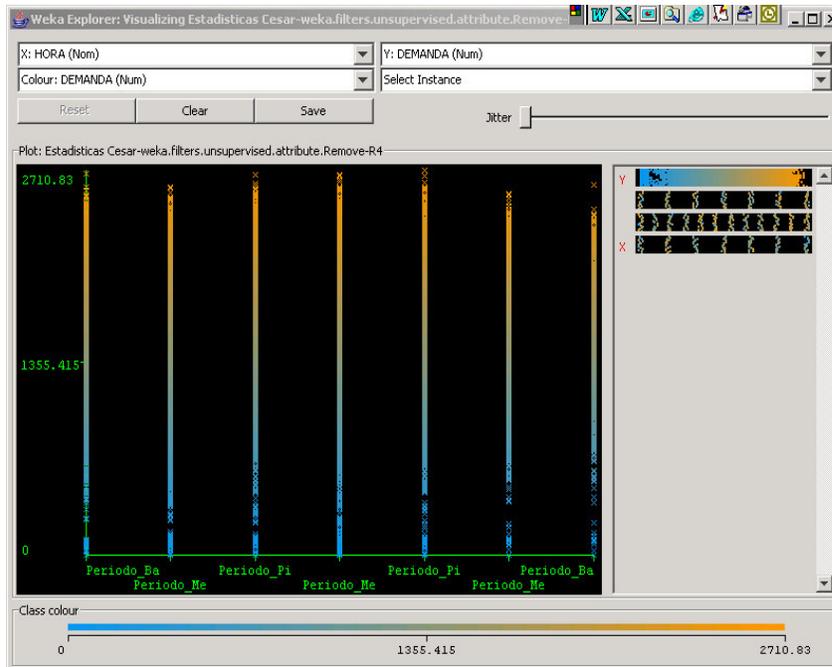


Figura 14. Demanda meses del año diferenciada por días de la semana (Cesar)



**ORDEN
SECUENCIAL**

P. Bajo 1: (1–4)
P. Medio 1: (5–9)
P. Pico 1: (10–12)
P. Medio 2: (13–18)
P. Pico 2: (19–21)
P. Medio 3: (22–23)
P. Bajo 2: (24)

Figura 15. Demanda periodos del día (Cesar)

2.3.5.5 Análisis Departamento: CESAR

La demanda de este departamento es la tercera más alta del sector, con demanda promedio de 1.435,84 kW/h y desviación estándar de 558,92 kW/h. Representada por los yacimientos de carbón de la Loma (Chiriguaná), cuya explotación equivale en volumen a 1/3 del total de la reserva del Cerrejón, La Jaqua de Ibirico, El Descanso y Alenturitas en el municipio de el Paso.

Los mayores porcentajes de frecuencia de ocurrencia se presentan en los rangos de valores cercanos a la demanda media, a excepción de un porcentaje representativo correspondiente a los intervalos (0-77,45) kW/h y (77,45-154,90) kW/h, que en su conjunto representan el 7,07% de la demanda; situación que resulta fácil de visualizar a través de la Figura12. Histograma de frecuencia.

Lo anterior quiere decir que la demanda del Cesar en el sector Minero oscila alrededor de los siguientes valores, en términos de una mayor frecuencia:

(1.471,59-1.549,04) kW/h

(1.549,04-1.626,49) kW/h

(1.626,49-1.703,95) kW/h

(1.703,95-1.781,40) kW/h

(1.781,40-1.858,85) kW/h

Con porcentajes muy parejos, que corresponden desde el 6,14% al 6,98% de la demanda.

En relación al comportamiento de la demanda por día de la semana, el día con la menor demanda promedio es el domingo (D. Media: 1.084,33-1.161,78) kW/h, cuya máxima (2.478,47-2.555,92) kW/h se presenta exactamente durante el periodo pico 1. Mientras que el día con la mayor demanda media es el miércoles (1.471,59-1.549,04) kW/h, seguido por el martes (D. Media: 1.394,14-1.471,59) kW/h, ambos con demanda máxima de (2.633,37-2.710,83) kW/h, durante los periodos medio 2 y pico 1 respectivamente. Esta misma demanda máxima se da en los días lunes, jueves y viernes; lunes y jueves durante el periodo pico 2 y el viernes durante el periodo pico 1; días en los cuales se presenta a demás igual demanda media (1.239,23-1.316,68) kW/h.

El mes de Abril, presenta una demanda promedio superior a la del resto de los meses del año de (1.936,30-2.013,75) kW/h, se encuentra a demás entre los de menor proporción de valores de demanda inferiores, y cuya demanda máxima se presenta durante los días miércoles (2.633,37-2.710,83) kW/h; a este mes le sigue muy de cerca la demanda del mes de Mayo (D. Media: 1.703,95-1.781,40) kW/h, con igual demanda máxima, presentada durante los días lunes.

Los meses durante los cuales se presentan los valores de demanda promedio más bajos del sector son los meses de Julio y Septiembre en ese mismo orden, con demandas medias de (1.084,33-1.161,78) kW/h y (929,42-1.006,87) kW/h respectivamente. Su demanda máxima se encuentra dentro del mismo intervalo (1.858,85-1.936,30) kW/h, y se presenta en los días miércoles de Julio y viernes de Septiembre.

El periodo del día durante el cual se presentan los menores valores de demanda promedio de energía del departamento en este sector es el periodo medio 1 (D. Media: 1.006,87-1.084,33) kW/h, seguido por el periodo medio 2 (D. Media: 851,97-929,42) kW/h. Mientras que los valores máximos promedios se presentan en los periodos pico 1 (D. Media: 1.626,49-1.703,95) kW/h y pico 2, (1.549,04-1.626,49) kW/h en ese mismo orden.

2.3.5.6 Resultados Departamento: CORDOBA

Resumen De Estadísticas

Cuadro 14. Estadísticas Córdoba

Total Datos	41.075	
Máximo	184.023,16	
Mínimo	0	
Rango	184.023,16	
No. De Clases	36.56	37
Media	88.318,01	
Desviación	73.607,83	

Cuadro 15. Distribución de frecuencia Córdoba

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-4.973,598919]'	15.978	38,90
'(4.973,598919-9.947,197838]'	2	0,00
'(9.947,197838-14.920,796757]'	18	0,04
'(14.920,796757-19.894,395676]'	9	0,02
'(19.894,395676-24.867,994595]'	13	0,03
'(24.867,994595-29.841,593514]'	17	0,04
'(29.841,593514-34.815,192432]'	16	0,04
'(34.815,192432-39.788,791351]'	20	0,05
'(39.788,791351-44.762,39027]'	26	0,06

DEMANDA	FRECUENCIA	PORCENTAJE %
'(44.762,39027-49.735,989189]'	25	0,06
'(49.735,989189-54.709,588108]'	37	0,09
'(54.709,588108-59.683,187027]'	36	0,09
'(59.683,187027-64.656,785946]'	70	0,17
'(64.656,785946-69.630,384865]'	127	0,31
'(69.630,384865-74.603,983784]'	185	0,45
'(74.603,983784-79.577,582703]'	285	0,69
'(79.577,582703-84.551,181622]'	411	1,00
'(84.551,181622-89.524,780541]'	535	1,30
'(89.524,780541-94.498,379459]'	581	1,41
'(94.498,379459-99.471,978378]'	364	0,89
'(99.471,978378-104.445,577297]'	304	0,74
'(104.445,577297-109.419,176216]'	278	0,68
'(109.419,176216-114.392,775135]'	316	0,77
'(114.392,775135-119.366,374054]'	373	0,91
'(119.366,374054-124.339,972973]'	465	1,13
'(124.339,972973-129.313,571892]'	583	1,42
'(129.313,571892-134.287,170811]'	764	1,86
'(134.287,170811-139.260,76973]'	999	2,43
'(139.260,76973-144.234,368649]'	1.375	3,35
'(144.234,368649-149.207,967568]'	1.914	4,66
'(149.207,967568-154.181,566486]'	2.630	6,40
'(154.181,566486-159.155,165405]'	3.309	8,06
'(159.155,165405-164.128,764324]'	3.450	8,40
'(164.128,764324-169.102,363243]'	3.123	7,60
'(169102,363243-174075,962162]'	1.844	4,49
'(174.075,962162-179.049,561081]'	546	1,33
'(179.049,561081-184.023,16)'	47	0,11

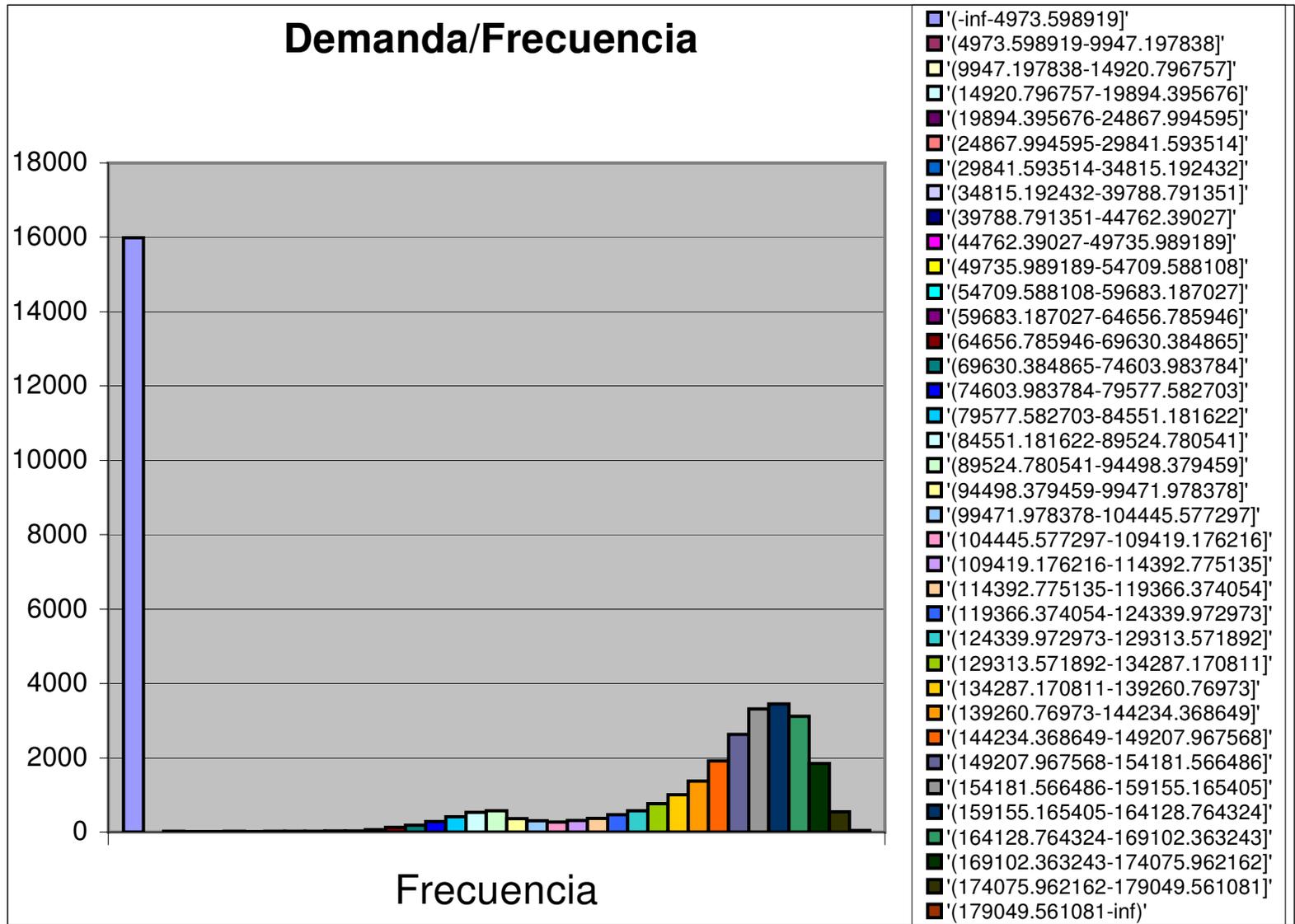


Figura 16. Histograma de frecuencia (Córdoba)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Análisis Individual

Cuadro 16. Reglas Córdoba

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
		(124339,97-129313,57)	(154065,90-158345,50)	(154065,90-158345,50)	(158345,50-162625,11)	(166904,72-171184,33)	(54709,58-59683,18)	(162625,11-166904,72)	(158345,50-162625,11)	(162625,11-166904,72)	(158345,50-162625,11)	(158345,50-162625,11)	(134287,17-139260,76)
Bajo 1 (1-4)	(124339,97-129313,57)												
Bajo 2 (24)	(119366,37-124339,97)									(24867,99-29841,59)			
Medio 1 (5-9)	(141227,07-145506,68)												
Medio 2 (13-18)	(69630,38-74603,98)		(29841,59-34815,19)				(9947,19-14920,79)						(179049,56-184023,16)
Medio 3 (22-23)	(149786,29-154065,90)						(29841,59-34815,19)		(59683,18-64656,78)	(19894,39-24867,99)	(24867,99-29841,59)		
Pico 1 (10-12)	(59683,18-64656,78)											(34815,19-39788,79)	
Pico 2 (19-21)	(149786,29-154065,90)												
Lunes	(109419,17-114392,77)												
Bajo 1 (1-4)		(159155,16-164128,76)					(0-4973,59)						
Bajo 2 (24)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(174075,96- 179049,56)												
Medio 2 (13-18)	(154065,90- 158345,50)		(119366,37- 124339,97)										
Medio 3 (22-23)	(174075,96- 179049,56)								(99471,97- 104445,57)				
Pico 1 (10-12)													
Pico 2 (19-21)	(158345,50- 162625,11)												
Martes	(145506,68- 149786,29)												
Bajo 1 (1-4)											(0-4973,59)		(144234,36- 149207,96)
jo 2 (24)	(12838,82- 17118,43)												
Medio 1 (5-9)	(158345,50- 162625,11)												
Medio 2 (13-18)	(158345,50- 162625,11)	(129313,57- 134287,17)											
Medio 3 (22-23)	(162625,11- 166904,72)												
Pico 1 (10-12)	(34815,19- 39788,79)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)													
Miércoles	<i>(79577,58-84551,18)</i>			<i>(12838,82-17118,43)</i>		<i>(54709,58-59683,18)</i>	<i>(9947,19-14920,79)</i>			<i>(19894,39-24867,99)</i>			
Bajo 1 (1-4)						<i>(119366,37-124339,97)</i>	<i>(0-4973,59)</i>			<i>(124339,97-129313,57)</i>			
Bajo 2 (24)													
Medio 1 (5-9)	<i>(174075,96-179049,56)</i>												
Medio 2 (13-18)	<i>(154065,90-158345,50)</i>		<i>(109419,17-114392,77)</i>			<i>(69630,38-74603,98)</i>							
Medio 3 (22-23)	<i>(19894,39-24867,99)</i>				<i>(74603,98-79577,58)</i>					<i>(19894,39-24867,99)</i>			
Pico 1 (10-12)													
Pico 2 (19-21)	<i>(174075,96-179049,56)</i>												
Jueves	<i>(79577,58-84551,18)</i>												
Bajo 1 (1-4)						<i>(119366,37-124339,97)</i>	<i>(0-4973,59)</i>						
Bajo 2 (24)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	<i>(174075,96-179049,56)</i>												
Medio 2 (13-18)	(158345,50-162625,11)						<i>(74603,98-79577,58)</i>						
Medio 3 (22-23)	(158345,50-162625,11)												
Pico 1 (10-12)		<i>(84551,18-89524,78)</i>											
Pico 2 (19-21)	<i>(174075,96-179049,56)</i>												
Viernes	(145506,68-149786,29)										<i>(9947,19-14920,79)</i>		
Bajo 1 (1-4)		(139260,76-144234,36)								<i>(0-4973,59)</i>			
Bajo 2 (24)													
Medio 1 (5-9)	(162625,11-166904,72)												
Medio 2 (13-18)	(154065,90-158345,50)	(84551,18-89524,78)											
Medio 3 (22-23)	(162625,11-166904,72)								<i>(109419,17-114392,77)</i>				
Pico 1 (10-12)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)	(158345,50- 162625,11)												
ábado	(109419,17- 114392,77)					(12838,82- 17118,43)							
Bajo 1 (1-4)						(134287,17- 139260,76)				(0-4973,59)			(159155,16- 164128,76)
Bajo 2 (24)													
Medio 1 (5-9)	(174075,96- 179049,56)												
Medio 2 (13-18)	(162625,11- 166904,72)												
Medio 3 (22-23)	(174075,96- 179049,56)												
Pico 1 (10-12)													
Pico 2 (19-21)	(154065,90- 158345,50)												
Domingo	(119366,37- 124339,97)												
Bajo 1 (1-4)		(154181,56- 159155,16)						(0-4973,59)					(164128,76- 169102,36)
Bajo 2 (24)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(158345,50- 162625,11)												
Medio 2 (13-18)	(158345,50- 162625,11)										(174075,96- 179049,56)		
Medio 3 (22-23)	(154065,90- 158345,50)												
Pico 1 (10-12)	(179049,56- 184023,16)												
Pico 2 (19-21)	(162625,11- 166904,72)												

Gráficos

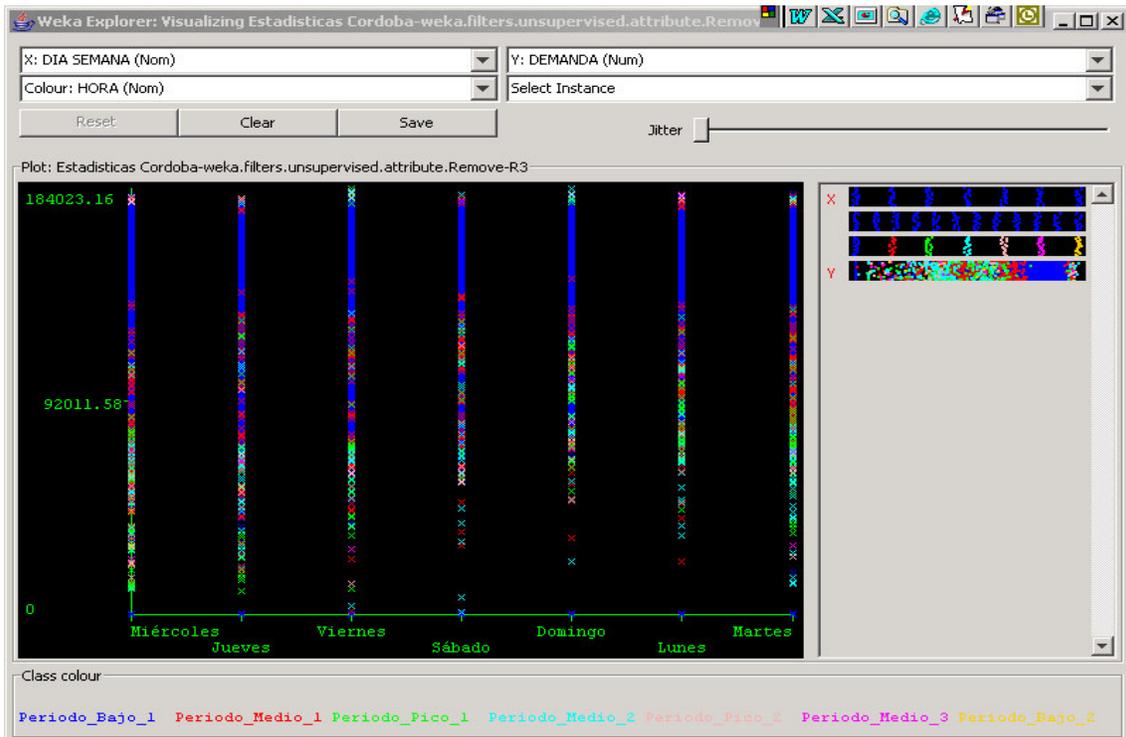


Figura 17. Demanda días de la semana diferenciada por periodos del día (Córdoba)

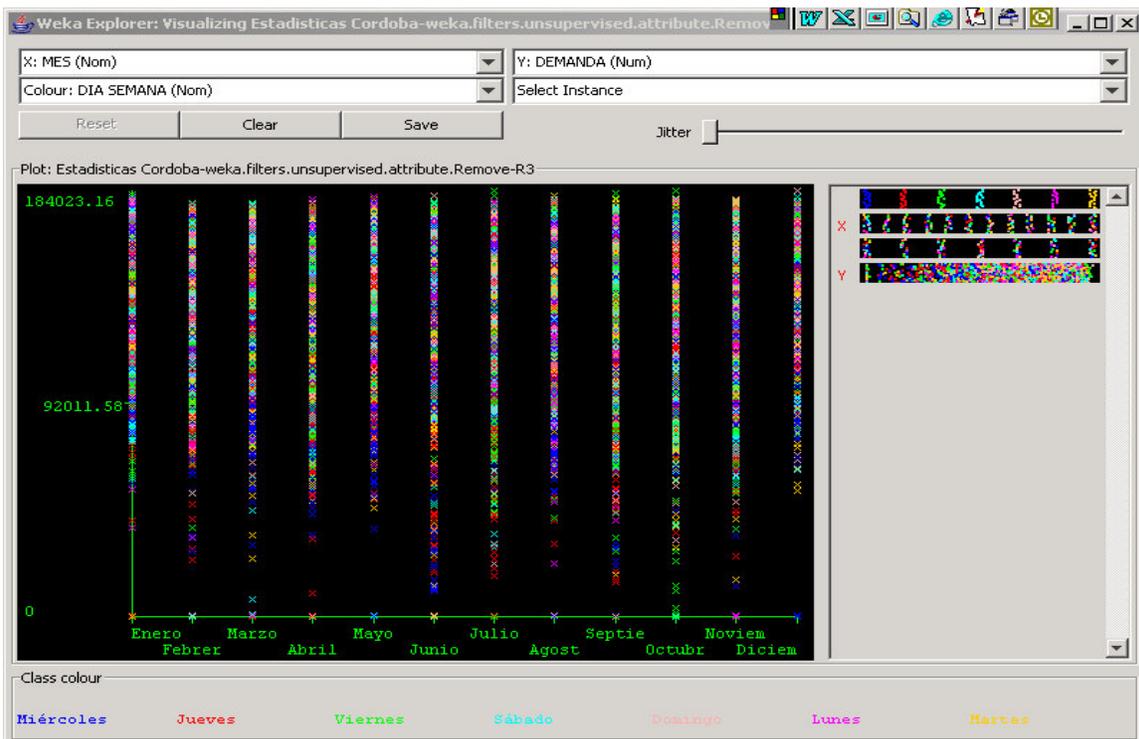


Figura 18. Demanda meses del año diferenciada por día de la semana (Córdoba)

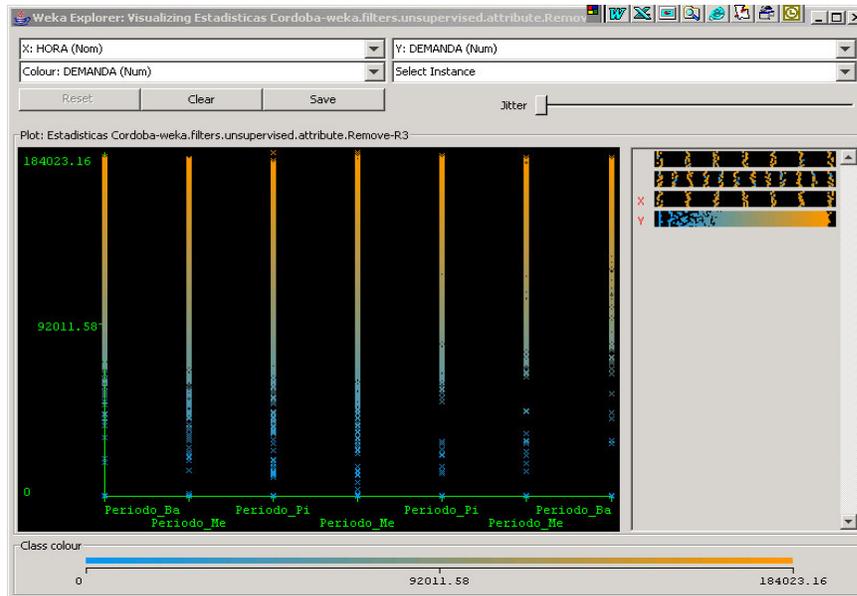


Figura 19. Demanda periodos del día (Córdoba)

- ORDEN
SECUENCIAL**
- P. Bajo 1: (1–4)
 - P. Medio 1: (5–9)
 - P. Pico 1: (10–12)
 - P. Medio 2: (13–18)
 - P. Pico 2: (19–21)
 - P. Medio 3: (22–23)
 - P. Bajo 2: (24)

2.3.5.7 Análisis Departamento: CORDOBA

Este departamento tiene la mayor demanda del sector Minero con media de 88.318,01 kW/h y desviación estándar de 73.607,83 kW/h; demanda que supera en un 97,88% la demanda promedio del resto de departamentos de la Costa Atlántica. Representado por los yacimientos de níquel de Planeta Rica, Ure y de manera significativa Cerromatoso, Complejo minero y metalúrgico, ubicado al sur del departamento (Montelíbano); actualmente es una de las diez empresas más grandes de Colombia, la tercera empresa minera después de ECOPETROL y Cerrejón, y la única productora de níquel en el país.

La demanda de Córdoba en el sector Minero está concentrada en grandes proporciones un intervalo de valores: (0-4.973,59) kW/h, el cual representa el 38,90% de la demanda del departamento. A pesar del porcentaje concentrado alrededor de dichos valores, las gráficas en general dejan saber que realmente existe baja proporción de datos alrededor de los valores de demanda más cercanos a cero; lo cual quiere decir, que este porcentaje de frecuencia representa los valores de demanda aproximados al límite superior del intervalo en mención.

En términos generales resulta claro que la demanda del departamento es muy estable en los valores de demanda superiores, tanto cuando se analiza por día de la semana, como para los meses del año, y periodos del día; esta misma estabilidad no suele presentarse para los límites inferiores; ya que su frecuencia de ocurrencia depende directamente del mes, el día y el periodo.

El día de mayor demanda promedio es el domingo (D. Media: 119.366,37-124.339,97) kW/h, día con la menor proporción de valores de demanda inferiores; su demanda máxima (179.049,56-184.023,16) kW/h se presenta durante el periodo pico 1, a este día le siguen con igual demanda media los días lunes y sábado (D. Media: 109.419,17-114.392,77) kW/h, y cuya máxima se presenta para ambos con igual valor durante los periodos medio 1 y medio 3 (D. Máxima: 179.049,56-184.023,16) kW/h. Los días de menor demanda son los miércoles y jueves con igual demanda media (79.577,58-84.551,18) kW/h y demanda máxima (174.075,96-179.049,56) kW/h, presentada esta última en ambos casos durante los periodos medio 1 y pico 2; siendo estos días los de mayor proporción de valores inferiores de demanda.

En cuanto a los periodos horarios, los valores de demanda superior se presentan con mayor frecuencia durante el periodo bajo 1 (D. Media: 124.339,97-129.313,57) kWh, aún cuando durante este mismo periodo se presentan de manera notoria los valores de demanda inferiores pertenecientes al intervalo (0-4.973,59) kW/h exactamente de demanda cero, para todos los días de la semana sin excepción alguna; vale agregar a demás que este es el periodo con la menor proporción de valores inferiores de demanda. En términos de demanda promedio, le sigue muy de cerca la demanda del periodo bajo 2 con demanda media (119.366,37-124.339,97) kW/h. Mientras tanto, el periodo del día con la menor demanda promedio es el periodo medio 2 (D. Media: 69.630,38-74.603,98) kW/h, seguido por el periodo pico 1 (D. Media: 59.683,18-64.656,78) kW/h; siendo estos los meses con la mayor proporción de valores inferiores de demanda.

Los meses en términos generales tienen una demanda muy pareja en cuanto a los valores superiores de demanda, como afirmamos anteriormente; Diciembre y Enero, en ese mismo orden, son los meses con las mayores demandas promedio del departamento, (134.287,17-139.260,76) kW/h y (124.339,97-129.313,57) kW/h, respectivamente; mientras que Junio es el mes con la mayor proporción de valores de demandas bajas, lo cual lo convierte en el mes con la menor demanda promedio del departamento en el sector (D. Media: 54.709,58-59.683,18) kW/h.

2.3.5.8 Resultados Departamento: LA GUAJIRA

Resumen De Estadísticas

Cuadro 17. Estadísticas La Guajira

Total Datos	50.200	
Máximo	27.962,42	
Mínimo	0	
Rango	27.962,42	
No, De Clases	36,93	37
Media	7.446,98	
Desviación	7.362,01	

Cuadro 18. Distribución de frecuencia La Guajira

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-755,741081]'	24.655	49,11
'(755,741081-1.511,482162]'	401	0,80
'(1.511,482162-2.267,223243]'	43	0,09
'(2.267,223243-3.022,964324]'	73	0,15
'(3.022,964324-3.778,705405]'	44	0,09
'(3.778,705405-4.534,446486]'	137	0,27
'(4.534,446486-5.290,187568]'	63	0,13
'(5.290,187568-6.045,928649]'	60	0,12

DEMANDA	FRECUENCIA	PORCENTAJE %
'(6.045,928649-6.801,66973]'	95	0,19
'(6.801,66973-7.557,410811]'	120	0,24
'(7.557,410811-8.313,151892]'	146	0,29
'(8313,151892-9068,892973]'	198	0,39
'(9.068,892973-9.824,634054]'	374	0,75
'(9.824,634054-10.580,375135]'	660	1,31
'(10.580,375135-11.336,116216]'	1.190	2,37
'(11.336,116216-12.091,857297]'	1.564	3,12
'(12.091,857297-12.847,598378]'	2.004	3,99
'(12.847,598378-13.603,339459]'	2.323	4,63
'(13.603,339459-14.359,080541]'	2.598	5,18
'(14.359,080541-15.114,821622]'	2.676	5,33
'(15.114,821622-15.870,562703]'	2.669	5,32
'(15.870,562703-16.626,303784]'	2.342	4,67
'(16.626,303784-17.382,044865]'	1.743	3,47
'(17.382,044865-18.137,785946]'	1.401	2,79
'(18.137,785946-18.893,527027]'	1.118	2,23
'(18.893,527027-19.649,268108]'	779	1,55
'(19.649,268108-20.405,009189]'	410	0,82
'(20.405,009189-21.160,75027]'	154	0,31
'(21.160,75027-21.916,491351]'	68	0,14
'(21.916,491351-22.672,232432]'	28	0,06
'(22.672,232432-23.427,973514]'	19	0,04
'(23.427,973514-24.183,714595]'	10	0,02
'(24.183,714595-24.939,455676]'	12	0,02
'(24.939,455676-25.695,196757]'	9	0,02
'(25.695,196757-26.450,937838]'	9	0,02
'(26.450,937838-27.206,678919]'	2	0,004
'(27.206,678919-27.962,42)'	3	0,01

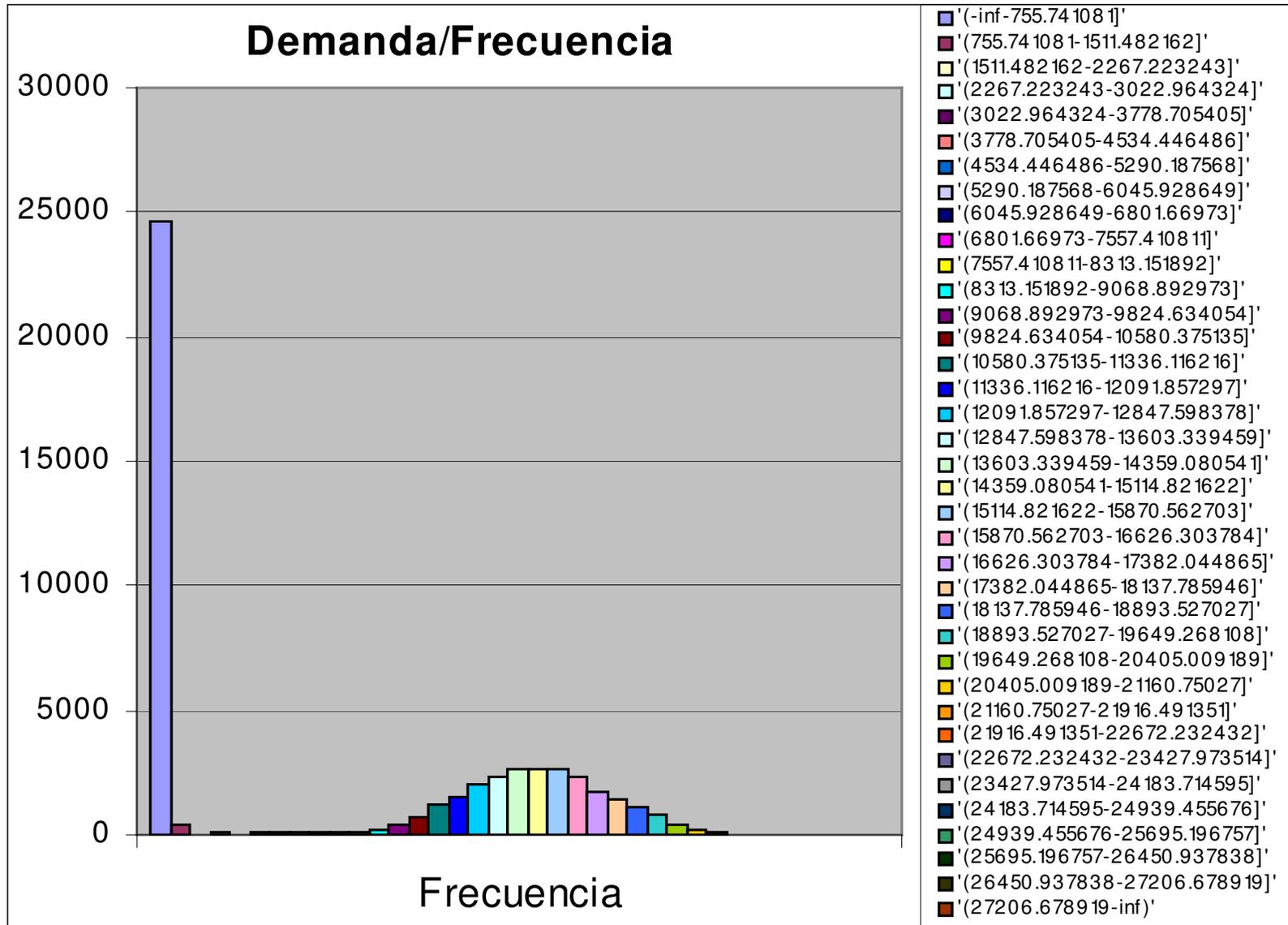


Figura 20. Histograma de frecuencia (La Guajira)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Cuadro 19. Reglas La Guajira

Análisis Individual

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
		(17118,43-21398,04)	(17118,43-21398,04)	(9824,63-10580,37)			(9824,63-10580,37)				(14359,08-15114,82)	(17118,43-21398,04)	(6045,92-6801,66)
Bajo 1 (1-4)		(12838,82-17118,43)		(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)		(12838,82-17118,43)	(12838,82-17118,43)		(12838,82-17118,43)	(3022,96-3778,70)
Bajo 2 (24)	(17118,43-21398,04)												
Medio 1 (5-9)		(12838,82-17118,43)	(12838,82-17118,43)		(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(8559,21-12838,82)	(12838,82-17118,43)
Medio 2 (13-18)	(15114,82-15870,56)	(12838,82-17118,43)	(12838,82-17118,43)	(2267,22-3022,96)	(8559,21-12838,82)	(17118,43-21398,04)	(17118,43-21398,04)	(8559,21-12838,82)	(8559,21-12838,82)	(12838,82-17118,43)	(12838,82-17118,43)	(8559,21-12838,82)	(12838,82-17118,43)
Medio 3 (22-23)	(4279,60-8559,21)	(4534,44-5290,18)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(21916,49-22672,23)	(12838,82-17118,43)	(5290,18-6045,92)
Pico 1 (10-12)	(14359,08-15114,82)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)		(12838,82-17118,43)		(12838,82-17118,43)
Pico 2 (19-21)	(4279,60-8559,21)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)		(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(23427,97-24183,71)	(12838,82-17118,43)	(3778,70-4534,44)
Lunes	(9824,63-10580,37)	(12838,82-17118,43)	(20405,009-21160,75)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(12838,82-17118,43)	(158345,50-162625,18)	(5290,18-6045,92)
Bajo 1 (1-4)	(8559,21-12838,82)		(0-755,74)			(755,74-1511,48)					(14359,08-15114,82)		
Bajo 2 (24)	(12838,82-17118,43)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(17118,43- 21398,04)					(755,74- 1511,48)					(18893,52- 19649,26)		
Medio 2 (13-18)	(8559,21- 12838,82)												
Medio 3 (22-23)	(12838,82- 17118,43)								(21160,75- 21916,49)				
Pico 1 (10-12)	(22672,23- 23427,97)												
Pico 2 (19-21)	(12838,82- 17118,43)												
Martes	(12847,59- 13603,33)	(12838,82- 17118,43)	(12838,82- 17118,43)	(2267,22- 3022,96)	(12838,82- 17118,43)	(12838,82- 17118,43)	(17118,43- 21398,04)	(12838,82- 17118,43)	(6801,66- 7557,41)	(12838,82- 17118,43)	(27206,67- 27962,42)		(3778,70- 4534,44)
Bajo 1 (1-4)			(12847,59- 13603,33)			(15870,56- 16626,30)				(0-755,74)			(755,74- 1511,48)
Bajo 2 (24)													
Medio 1 (5-9)	(8559,21- 12838,82)												
Medio 2 (13-18)	(27206,67- 27962,42)			(2267,22- 3022,96)							(24183,71- 24939,45)		(3022,96- 3778,70)
Medio 3 (22-23)	(12838,82- 17118,43)												
Pico 1 (10-12)	(12838,82- 17118,43)	(8313,15- 9068,89)											(6045,92- 6801,66)

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)	(12838,82- 17118,43)									(21160,75- 21916,49)			
Miércoles	(4279,60- 8559,21)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)		(12838,82- 17118,43)	(3778,70- 4534,44)	(12838,82- 17118,43)	(22672,23- 23427,97)		(2267,22- 3022,96)
Bajo 1 (1-4)		(3778,70- 4534,44)		(12847,59- 13603,33)		(755,74- 1511,48)	(0-755,74)			(13603,33- 14359,08)			(6045,92- 6801,66)
Bajo 2 (24)	(12838,82- 17118,43)												
Medio 1 (5-9)	(8559,21- 12838,82)										(9068,89- 9824,63)		(6045,92- 6801,66)
Medio 2 (13-18)	(17118,43- 21398,04)												
Medio 3 (22-23)	(12838,82- 17118,43)										(19649,26- 20405,00)		(3778,70- 4534,44)
Pico 1 (10-12)	(12838,82- 17118,43)												
Pico 2 (19-21)	(12838,82- 17118,43)						(8313,15- 9068,89)						
Jueves	(4279,60- 8559,21)	(12838,82- 17118,43)	(3778,70- 4534,44)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)		(8559,21- 12838,82)	(1511,48- 2267,22)	(12838,82- 17118,43)
Bajo 1 (1-4)					(0-755,74)				(12847,59- 13603,33)			(15870,56- 16626,30)	
Bajo 2 (24)	(12838,82- 17118,43)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(8559,21- 12838,82)										(19649,26- 20405,009)		
Medio 2 (13-18)	(17118,43- 21398,04)		(4534,44- 5290,18)	(9824,63- 10580,37)					(10580,37- 11336,11)			(8313,15- 9068,89)	
Medio 3 (22-23)	(12838,82- 17118,43)												
Pico 1 (10-12)	(12838,82- 17118,43)												
Pico 2 (19-21)	(12838,82- 17118,43)												
Viernes	(10580,37- 11336,11)	(12838,82- 17118,43)	(12838,82- 17118,43)	(24183,71- 24939,45)	(12838,82- 17118,43)		(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)		(12838,82- 17118,43)	(12838,82- 17118,43)
Bajo 1 (1-4)		(3778,70- 4534,44)		(0-755,74)	(12091,85- 12847,59)			(16626,30- 17382,04)					(755,74- 1511,48)
Bajo 2 (24)	(12838,82- 17118,43)												
Medio 1 (5-9)	(8559,21- 12838,82)												
Medio 2 (13-18)	(25695,19- 26450,93)												(4534,44- 5290,18)
Medio 3 (22-23)	(12838,82- 17118,43)						(9068,89- 9824,63)						
Pico 1 (10-12)	(12838,82- 17118,43)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)	(12838,82- 17118,43)												
Sábado		(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(1511,48- 2267,22)	(12838,82- 17118,43)	(24939,45- 25695,19)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(4534,44- 5290,18)	(12838,82- 17118,43)	(2267,22- 3022,96)
Bajo 1 (1-4)				(14359,08- 15114,82)	(755,74- 1511,48)	(12091,85- 12847,59)	(8313,15- 9068,89)					(0-755,74)	
Bajo 2 (24)	(12838,82- 17118,43)												
Medio 1 (5-9)	(8559,21- 12838,82)	(3778,70- 4534,44)											
Medio 2 (13-18)	(8559,21- 12838,82)				(1511,48- 2267,22)								
Medio 3 (22-23)					(1511,48- 2267,22)								
Pico 1 (10-12)	(12838,82- 17118,43)												
Pico 2 (19-21)					(1511,48- 2267,22)								
Domingo		(12838,82- 17118,43)	(8559,21- 12838,82)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(2267,22- 3022,96)	(12838,82- 17118,43)	(12838,82- 17118,43)	(12838,82- 17118,43)	(19649,26- 20405,00)
Bajo 1 (1-4)					(14359,08- 15114,82)			(0-755,74)			(755,74- 1511,48)		
Bajo 2 (24)	(12838,82- 17118,43)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(8559,21- 12838,82)								(2267,22- 3022,96)				
Medio 2 (13-18)	(8559,21- 12838,82)			(8313,15- 9068,89)									(5290,18- 6045,92)
Medio 3 (22-23)	(12838,82- 17118,43)												
Pico 1 (10-12)	(12838,82- 17118,43)												
Pico 2 (19-21)	(2267,22- 3022,96)								(2267,22- 3022,96)				

Gráficos

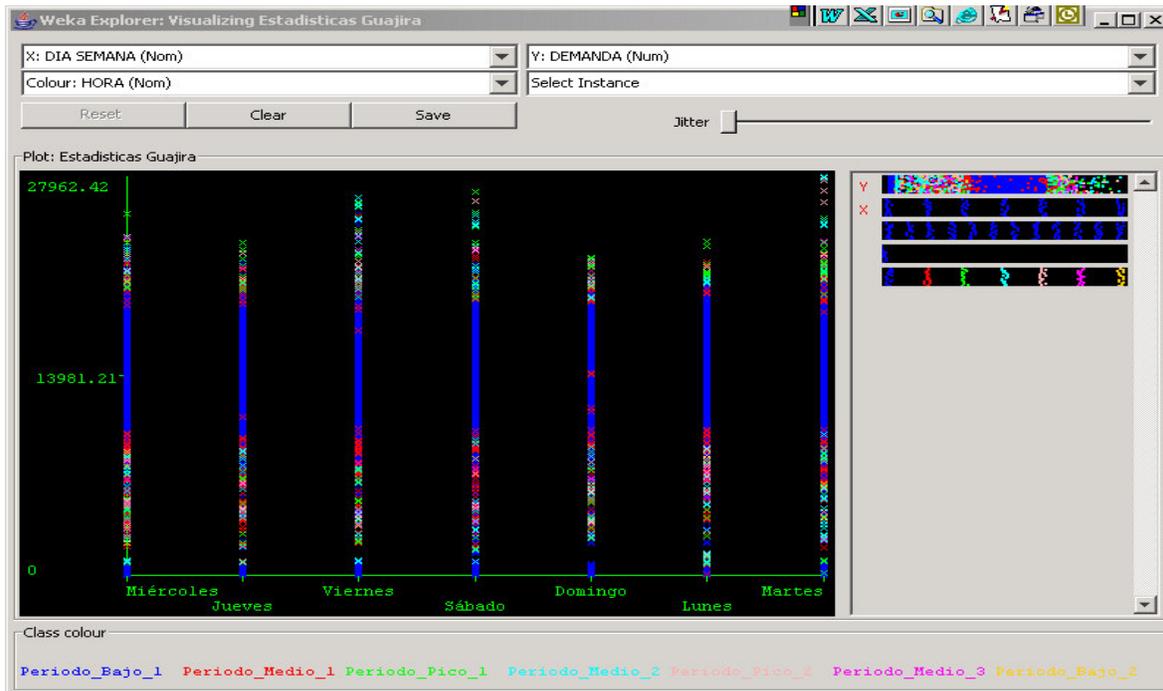


Figura 21. Demanda días de la semana diferenciada por periodos del día (La Guajira)

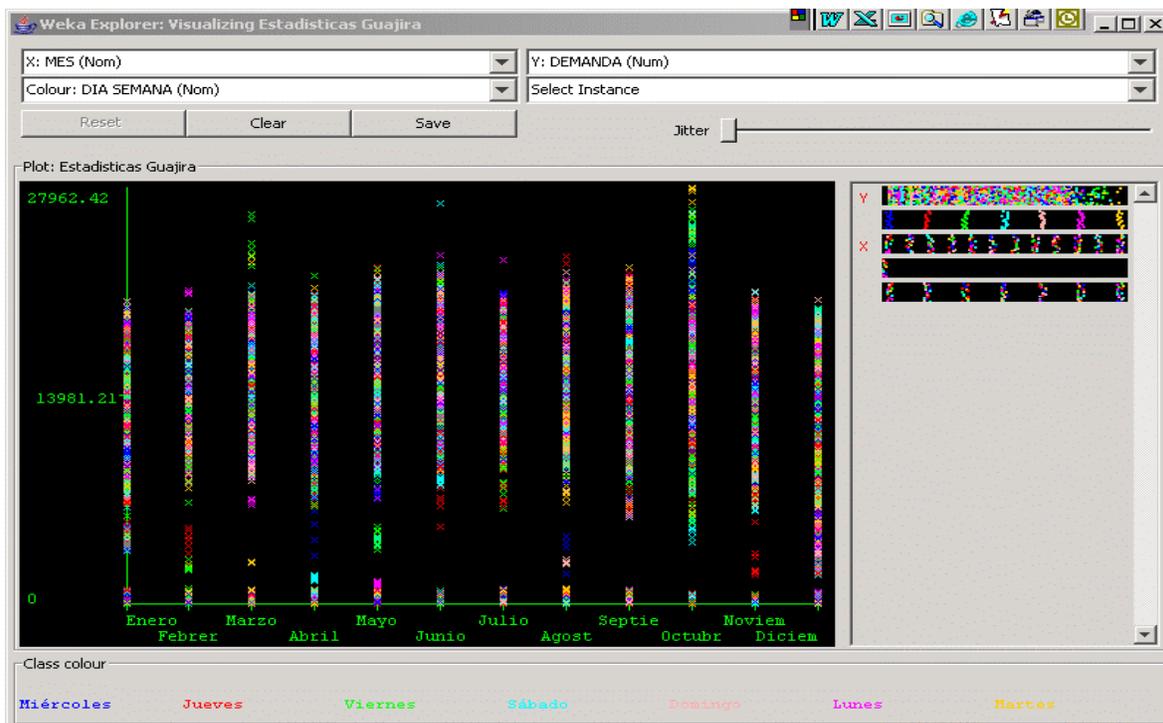
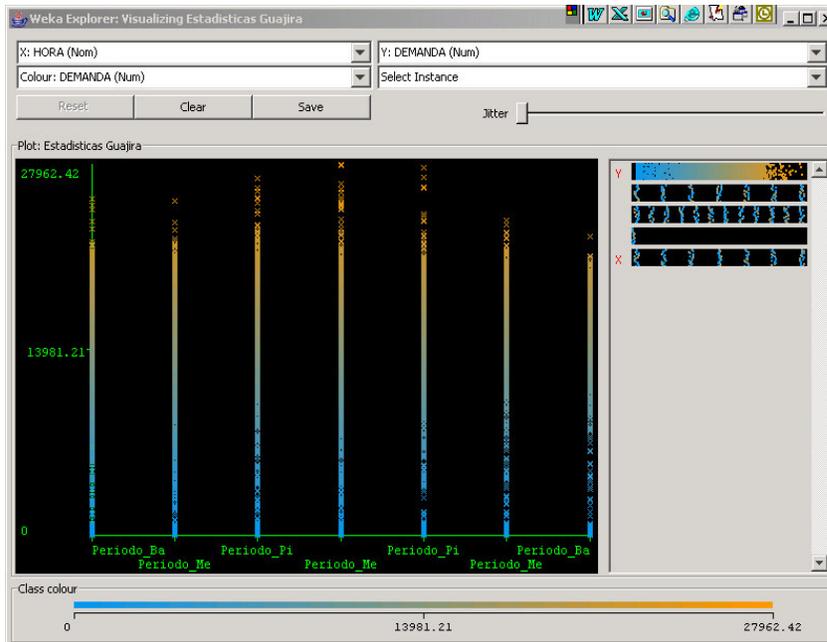


Figura 22. Demanda meses del año diferenciada por días de la semana (La Guajira)



**ORDEN
SECUENCIAL**

P. Bajo 1: (1–4)
P. Medio 1: (5–9)
P. Pico 1: (10–12)
P. Medio 2: (13–18)
P. Pico 2: (19–21)
P. Medio 3: (22–23)
P. Bajo 2: (24)

Figura 23. Demanda periodos del día (La Guajira)

2.3.5.9 Análisis Departamento: LA GUAJIRA

La demanda del departamento de La Guajira es la segunda demanda más alta del sector, con media: 7.446,98 kW/h, y desviación estándar: 7.362,01 kW/h, superando en un 47,81% la demanda promedio del resto de departamentos; sin incluir, claro está, el departamento de Córdoba.

El sector minero en el departamento de La Guajira está representado por la empresa el Cerrejón; reserva de carbón mineral calculada en 3.670 millones de toneladas, con un 89,4% de la producción nacional; ubicada en el sur del departamento.

La mayor proporción de la demanda está concentrada en un porcentaje representativo correspondiente al 49,11% de la demanda del departamento en el intervalo de valores de (0-755,74) kW/h.

La demanda promedio más baja del departamento es la del día lunes (D. Media: 9.824,63-10.580,37) kW/h, con demanda máxima de (22.672,23-23.427,97) kW/h, presente durante el periodo pico 1.

Los días de mayor demanda promedio son los martes, con demanda media de (12.847,59-13.603,33) kW/h y demanda máxima de (27.206,67-27.962,42) kW/h en el periodo medio 2; y los viernes con demanda media de (10.580,37-11.336,11) kW/h y demanda máxima (25.695,19-26.450,93) kW/h igualmente durante el periodo medio 2.

En este departamento el periodo bajo 1 deja ver nuevamente su comportamiento especial en el sector Minero, tomando valores en rangos inferiores, específicamente (0-755,74) kW/h y (755,74-1,511,48) kW/h y en rangos medios altos como es el caso de los rangos (12,847,59-13,603,33), (13,603,33-14,359,08) kW/h y (14,359,08-15,114,82) kW/h, entre otros intervalos cercanos.

Durante los días de la semana, los valores de demanda máxima se presentan durante los periodos medio 2 (martes y viernes) y pico 1 (lunes, miércoles, jueves, sábado y domingo); resultando ser a demás, los periodos durante los cuales se presenta la mayor demanda promedio del departamento.

Aún cuando en el periodo pico 1 durante mayor número de días se presentan valores de demanda máxima que en el periodo medio 2; este último presenta una demanda promedio superior al primero, Periodo pico 1 (D. Media: 14.359,08-15.114,82) kW/h y periodo medio 2 (D. Media: 15.114,82-15.870,56).

En cuanto a los meses del año, el de mayor demanda promedio es Octubre (D. Media: 14.359,08-15.114,82) kW/h, cuya demanda máxima (27.206,67-27.962,42) kW/h se da durante los martes; en segundo lugar están los meses de Marzo y Junio con igual demanda media (9.824,63-10.580,37) kW/h y demandas máximas

de (24.183,71-24.939,45) kW/h durante los viernes y (24.939,45-25.695,19) kW/h durante el sábado, respectivamente.

Mientras tanto, el mes con menor demanda promedio es Diciembre (D. Media: 6.045,92-6.801,66) kW/h, con demanda máxima de (19.649,26-20.405,00) kW/h presente durante los domingos. Diciembre es a demás el mes del año en el que se presentan con mayor frecuencia, los valores de demanda de rangos inferiores.

2.3.5.10 Resultados Departamento: MAGDALENA

Resumen De Estadísticas

Cuadro 20. Estadísticas Magdalena

Total de Datos	50.200	
Máximo	3.678,58	
Mínimo	0	
Rango	3.678,58	
No. De Clases	36,93	37
Media	1.340,59	
Desviación	913,44	

Cuadro 21. Distribución de frecuencia Magdalena

INTERVALOS	FRECUENCIA	PORCENTAJE %
'(0-99,421081]'	1.095	2,18
'(99,421081-198,842162]'	2.832	5,64
'(198,842162-298,263243]'	1.912	3,81
'(298,263243-397,684324]'	2.212	4,41
'(397,684324-497,105405]'	2.874	5,73
'(497,105405-596,526486]'	3.025	6,03
'(596,526486-695,947568]'	3.109	6,19
'(695,947568-795,368649]'	2.985	5,95

INTERVALOS	FRECUENCIA	PORCENTAJE %
'(795,368649-894,78973]'	2.492	4,96
'(894,78973-994,210811]'	1.952	3,89
'(994,210811-1.093,631892]'	1.487	2,96
'(1.093,631892-1.193,052973]'	1.101	2,19
'(1.193,052973-1.292,474054]'	992	1,98
'(1.292,474054-1.391,895135]'	846	1,69
'(1.391,895135-1.491,316216]'	587	1,17
'(1.491,316216-1.590,737297]'	540	1,08
'(1.590,737297-1.690,158378]'	560	1,12
'(1.690,158378-1.789,579459]'	689	1,37
'(1.789,579459-1.889,000541]'	899	1,79
'(1.889,000541-1.988,421622]'	1.163	2,32
'(1.988,421622-2.087,842703]'	1.733	3,45
'(2.087,842703-2.187,263784]'	1.590	3,17
'(2.187,263784-2.286,684865]'	1.729	3,44
'(2.286,684865-2.386,105946]'	1.949	3,88
'(2.386,105946-2.485,527027]'	2.277	4,54
'(2.485,527027-2.584,948108]'	2.076	4,14
'(2.584,948108-2.684,369189]'	1.786	3,56
'(2.684,369189-2.783,79027]'	1.386	2,76
'(2.783,79027-2.883,211351]'	828	1,65
'(2.883,211351-2.982,632432]'	557	1,11
'(2.982,632432-3.082,053514]'	392	0,78
'(3.082,053514-3.181,474595]'	231	0,46
'(3.181,474595-3.280,895676]'	152	0,30
'(3.280,895676-3.380,316757]'	97	0,19
'(3.380,316757-3.479,737838]'	41	0,08
'(3.479,737838-3.579,158919]'	15	0,03
'(3.579,158919-3.678,58]'	9	0,02

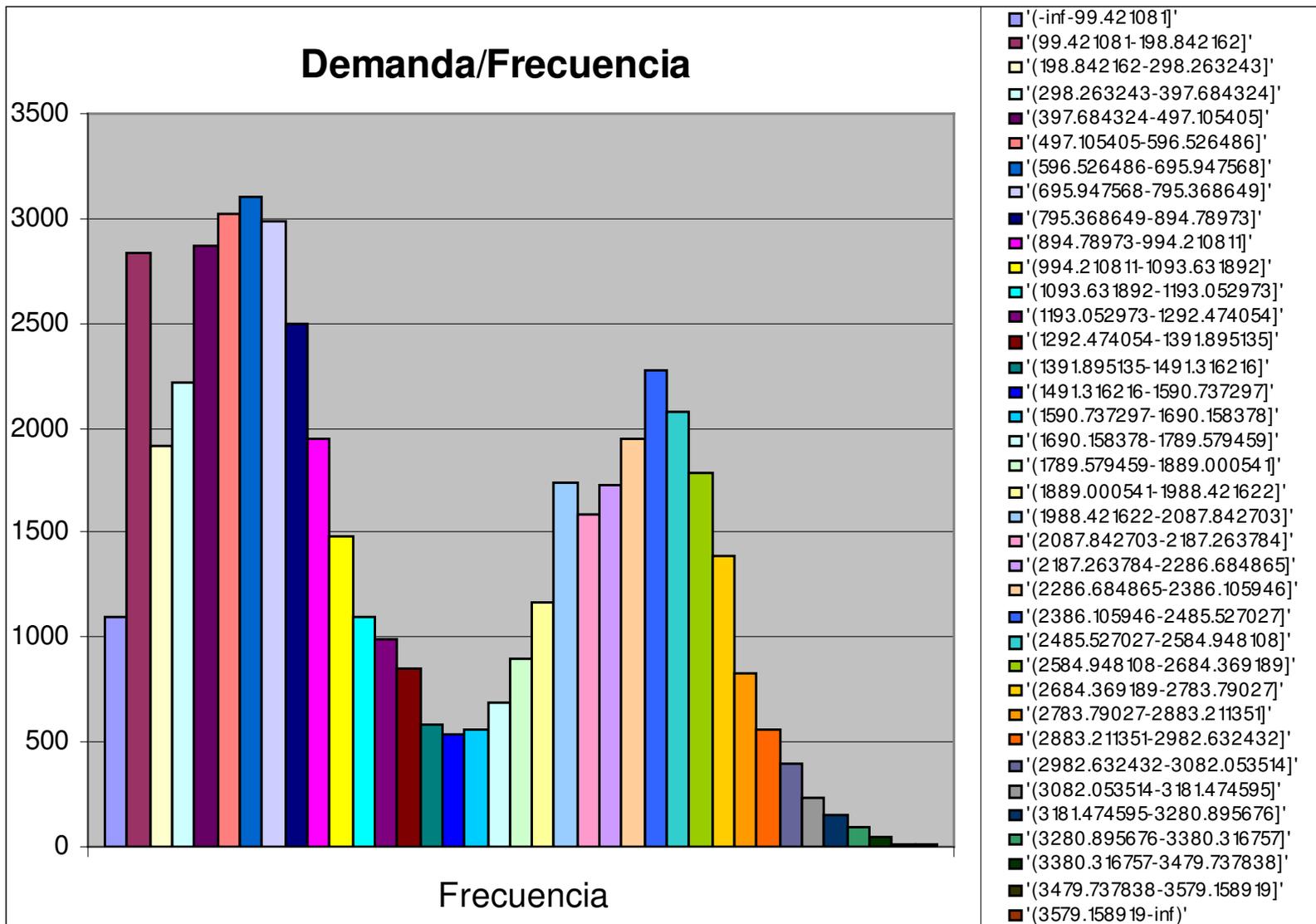


Figura 24. Histograma de frecuencia (Magdalena)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Cuadro 22. Reglas Magdalena

Análisis Individual

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
				(1590,73- 1690,15)	(1491,31- 1590,73)	(1491,31- 1590,73)	(1690,15- 1789,57)					(1193,05- 1292,42)	(1193,05- 1292,42)
Bajo 1 (1-4)													
Bajo 2 (24)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)			
Medio 1 (5-9)	(1789,57- 1889,00)												
Medio 2 (13-18)													
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Lunes	(1590,73- 1690,15)						(3479,73- 3579,15)						
Bajo 1 (1-4)		(298,26- 397,68)					(2883,21- 2982,63)						(0-99,42)
Bajo 2 (24)	(0- 4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(3479,73- 3579,15)												
Medio 2 (13-18)				(0- 4279,60)		(0- 4279,60)			(0-99,42)				
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Martes	(1292,47- 1391,89)				(3579,15- 3678,58)	(3579,15- 3678,58)							
Bajo 1 (1-4)	(3579,15- 3678,58)				(2187,26- 2286,68)							(0-99,42)	
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)													
Medio 2 (13-18)				(0- 4279,60)					(0- 4279,60)				
Medio 3 (22-23)	(3579,15- 3678,58)												
Pico 1 (10-12)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)													
Miércoles													
Bajo 1 (1-4)		<i>(0-99,42)</i>	<i>(397,68-497,10)</i>					<i>(1889,00-1988,42)</i>		<i>(2783,79-2883,21)</i>			
Bajo 2 (24)	(0-4279,60)												
Medio 1 (5-9)													
Medio 2 (13-18)		(0-4279,60)		(0-4279,60)	(0-4279,60)		(0-4279,60)	(0-4279,60)	(0-4279,60)	(0-4279,60)			
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Jueves													
Bajo 1 (1-4)					<i>(298,26-397,68)</i>				<i>(2982,63-3082-05)</i>				<i>(0-99,42)</i>
Bajo 2 (24)	(0-4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)													
Medio 2 (13-18)		(0- 4279,60)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)		(0- 4279,60)			
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Viernes													
Bajo 1 (1-4)				(298,26- 397,68)			(2982,63- 3082,05)				(0-99,42)		
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)	(3479,73- 3579,15)												
Medio 2 (13-18)		(0- 4279,60)			(0- 4279,60)			(0- 4279,60)					
Medio 3 (22-23)													
Pico 1 (10-12)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)													
Sábado	(1590,73- 1690,15)						(3479,73- 3579,15)					(3280,89- 3380,31)	(3280,89- 3380,31)
Bajo 1 (1-4)				(397,68- 497,10)	(2883,21- 2982,63)	(2982,63- 308205)						(0-99,42)	
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)													
Medio 2 (13-18)		(0- 4279,60)				(0- 4279,60)		(0- 4279,60)					
Medio 3 (22-23)	(3579,15- 3678,58)												
Pico 1 (10-12)													
Pico 2 (19-21)													
Domingo	(1193,05- 1292,47)												
Bajo 1 (1-4)							(2783,79- 2883,21)		(298,26- 397,68)		(0-99,42)		
Bajo 2 (24)	(0- 4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)													
Medio 2 (13-18)						(0- 4279,60)		(99,42- 198,84)	(0- 4279,60)				
Medio 3 (22-23)	(3380,31- 3479,73)												
Pico 1 (10-12)								(99,42- 198,84)					
Pico 2 (19-21)													

Gráficos

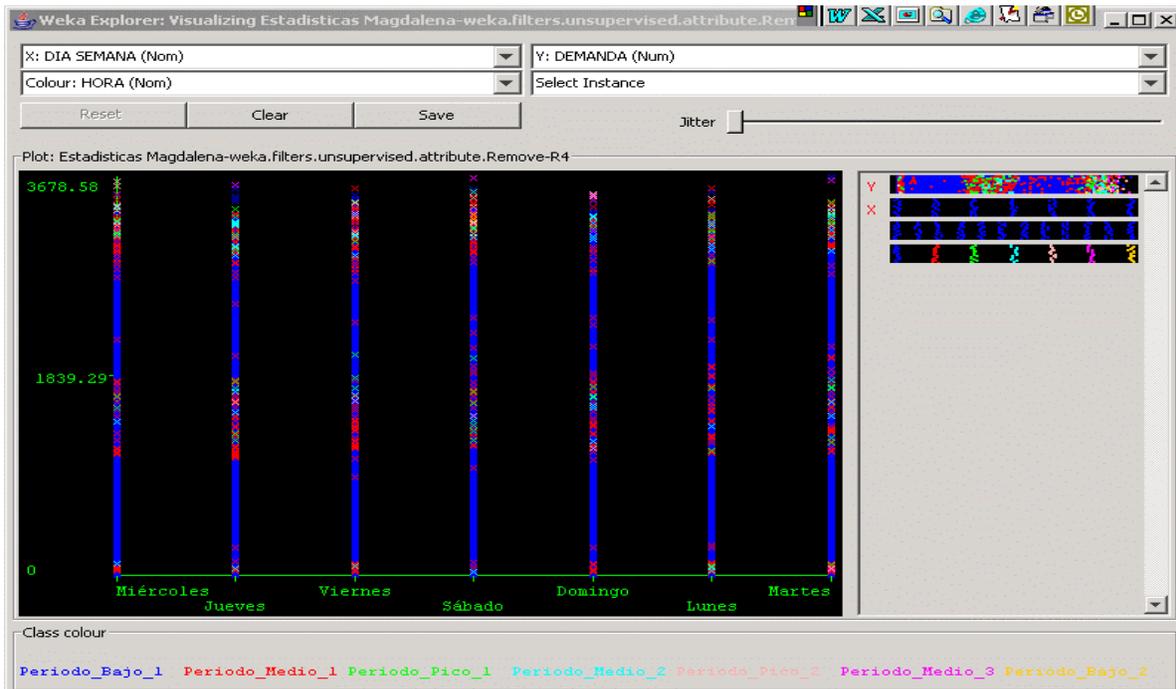


Figura 25. Demanda días de la semana diferenciada por periodos del día (Magdalena)

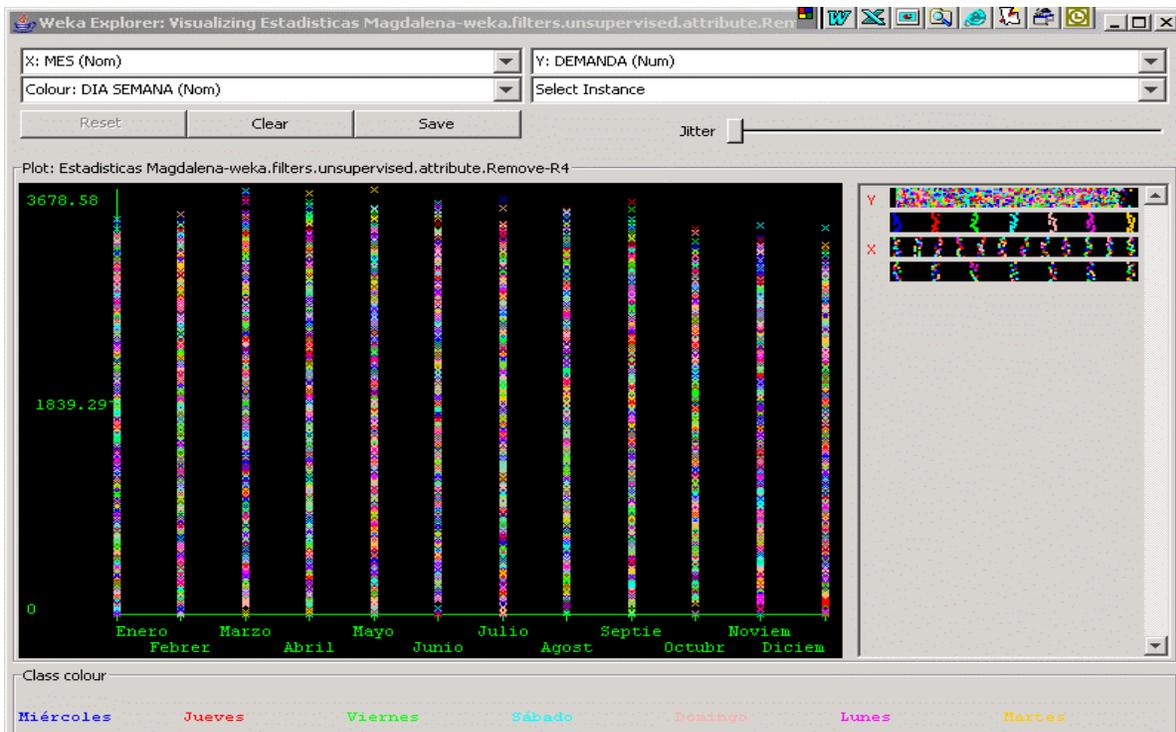
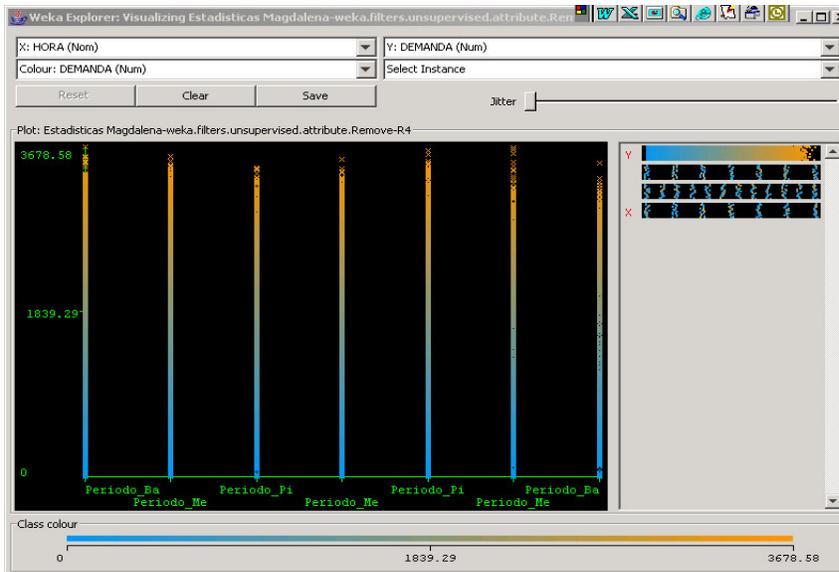


Figura 26. Demanda meses del año diferenciada por días de la semana (Magdalena)



ORDEN SECUENCIAL	
P. Bajo 1:	(1–4)
P. Medio 1:	(5–9)
P. Pico 1:	(10–12)
P. Medio 2:	(13–18)
P. Pico 2:	(19–21)
P. Medio 3:	(22–23)
P. Bajo 2:	(24)

Figura 27. Demanda periodos del día (Magdalena)

2.3.5.11 Análisis Departamento: MAGDALENA

La demanda del departamento del Magdalena ocupa el cuarto lugar en ponderación de menor a mayor demanda de los siete departamentos de la Costa Atlántica, lo cual nos impide inferir mayores conclusiones, en cuanto a si se encuentra o no dentro de las más bajas o altas demandas del sector. Tiene una demanda media de 1.340,59 kW/h y desviación estándar de 913,44 kW/h.

Se encuentra distribuida en 37 intervalos de valores que van desde (0-99,42) kW/h hasta (3.579,15-3.678,58) kW/h, con porcentajes de frecuencia de 0,02% (el menor) correspondiendo al último intervalo en mención y 6,19% (el mayor) que corresponde al rango de (596,52-695,94) kW/h.

Los días de mayor demanda promedio del departamento son los lunes y sábados con igual demanda media (1.590,73-1.690,15) kW/h y demandas máximas de (3.479,73-3.579,15) kW/h durante el periodo medio 1 y (3.579,15-3.678,58) kW/h durante el periodo medio 3, respectivamente.

Mientras tanto, el día de menor demanda es el martes (D. Media: 1.292,47-1.391,89) kW/h y cuya demanda máxima (3.579,15-3.678,58) kW/h se presenta exactamente durante los periodos bajo 1 y medio 3. A este día le sigue muy de cerca la demanda promedio del domingo (1.193,05-1.292,47) kW/h, con demanda máxima (3.380,31-3.479,73) kW/h durante el periodo medio 3.

Resulta claro a través de la grafica de demanda contra día de la semana diferenciada por periodo, que en este departamento aparece nuevamente el comportamiento del periodo bajo 1 especialmente marcado para los valores de demanda bajos y medios altos, corroborando lo expuesto desde el análisis conjunto, siguiendo con el recorrido individual de los departamentos.

En valores de demanda inferiores, el periodo bajo 1 se concentra con mayor frecuencia en los intervalos (0-99,42) kW/h, (298,26-397,68) kW/h y (397,68-497,10) kW/h; entre los medios altos se pueden mencionar los rangos (2.187,26-2.286,68) kW/h, (2.783,79-2.883,21) kW/h y (2.883,21-2.982,63) kW/h.

El periodo del día con mayor demanda promedio es el periodo medio 1, con demanda media de (1.789,57-1.889,00) kW/h, con demanda máxima para los días lunes y viernes (3.479,73-3.579,15) kW/h.

En cuanto a los meses del año, el de mayor demanda es el mes de Junio (D. Media: 1.690,15-1.789,57) kW/h, cuya demanda máxima se presenta con igual valor durante los días lunes y sábados (3.479,73-3.579,15) kW/h. Le sigue el mes de Marzo con demanda media de (1.590,73-1.690,15) kW/h y los meses de Abril y Mayo con igual demanda promedio (1.491,31-1.590,73) kW/h y demanda máxima durante los martes (3.579,15-3.678,58) kW/h. Los meses de menor demanda del sector en el departamento de Magdalena son con igual valor, los meses de Noviembre y Diciembre (D. Media: 1.193,05-1.292,42) kW/h, ambos con demanda máxima durante los sábados (3.280,89-3.380,31) kW/h.

2.3.5.12 Resultados Departamento: SUCRE

Resumen De Estadísticas

Cuadro 23. Estadísticas Sucre

Total de Datos	25.100	
Máximo	2.767,61	
Mínimo	0	
Rango	2.767,61	
No, De Clases	34,63	35
Media	456,76	
Desviación	596,91	

Cuadro 24. Distribución de frecuencia Sucre

DEMANDA	FRECUENCIA	PORCENTAJE %
'(0-79,074571]'	167	0,67
'(79,074571-158,149143]'	887	3,53
'(158,149143-237,223714]'	10.639	42,39
'(237,223714-316,298286]'	5.420	21,59
'(316,298286-395,372857]'	1.875	7,47
'(395,372857-474,447429]'	2.444	9,74
'(474,447429-553,522]'	886	3,53
'(553,522-632,596571]'	174	0,69
'(632,596571-711,671143]'	53	0,21
'(711,671143-790,745714]'	29	0,12
'(790,745714-869,820286]'	24	0,10
'(869,820286-948,894857]'	28	0,11
'(948,894857-1.027,969429]'	19	0,08
'(1.027,969429-1.107,044]'	19	0,08
'(1.107,044-1.186,118571]'	14	0,06
'(1.186,118571-1.265,193143]'	61	0,24
'(1.265,193143-1.344,267714]'	129	0,51

DEMANDA	FRECUENCIA	PORCENTAJE %
'(1.344,267714-1.423,342286]'	55	0,22
'(1.423,342286-1.502,416857]'	34	0,14
'(1.502,416857-1.581,491429]'	34	0,14
'(1.581,491429-1.660,566]'	18	0,07
'(1.660,566-1.739,640571]'	13	0,05
'(1.739,640571-1.818,715143]'	22	0,09
'(1.818,715143-1.897,789714]'	20	0,08
'(1.897,789714-1.976,864286]'	18	0,07
'(1.976,864286-2.055,938857]'	31	0,12
'(2.055,938857-2.135,013429]'	22	0,09
'(2.135,013429-2.214,088]'	37	0,15
'(2.214,088-2.293,162571]'	255	1,02
'(2.293,162571-2.372,237143]'	777	3,10
'(2.372,237143-2.451,311714]'	588	2,34
'(2.451,311714-2.530,386286]'	128	0,51
'(2.530,386286-2.609,460857]'	114	0,45
'(2.609,460857-2.688,535429]'	55	0,22
'(2.688,535429-2.767,61]'	11	0,04

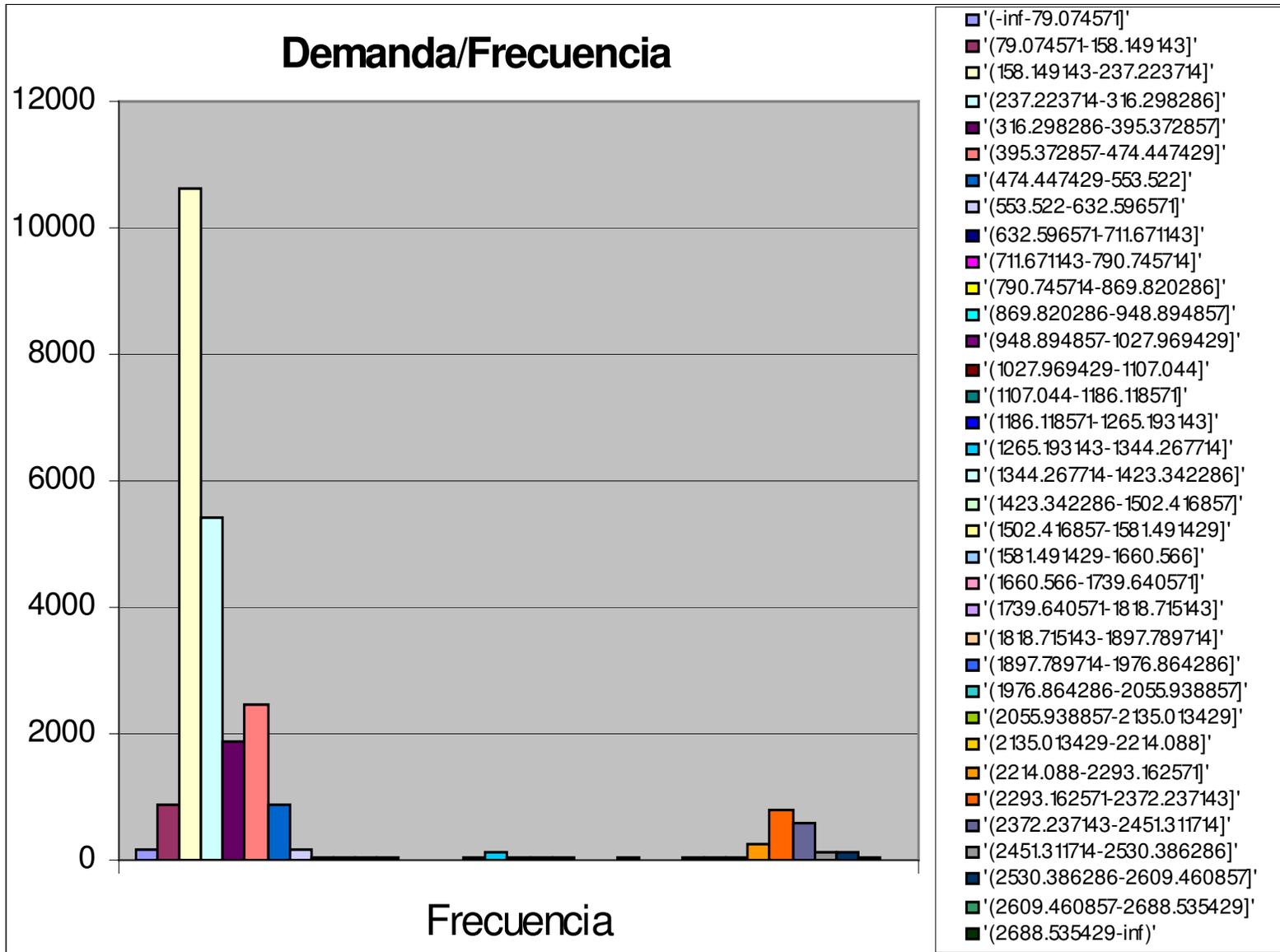


Figura 28. Histograma de frecuencia (Sucre)

Formato Excel: Traducción De Reglas

Análisis Conjunto

Análisis Individual

Cuadro 25. Reglas Sucre

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
				(1186,11- 1265,19)		(1265,19- 1344,26)			(711,67- 790,74)	(632,59- 711,67)	(632,59- 711,67)		
Bajo 1 (1-4)							(553,52- 632,59)		(790,74- 869,82)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 2 (24)					(711,67- 790,74)								
Medio 1 (5-9)	(869,82- 948,89)				(1107,04- 1186,11)						(869,82- 948,89)		(948,89- 1027,96)
Medio 2 (13-18)				(948,89- 1027,96)			(1976,86- 2055,93)	(869,82- 948,89)		(2451,31- 2530,38)			
Medio 3 (22-23)	(1265,19- 1344,26)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(948,89- 1027,96)	(2688,53- 2767,61)	(2609,46- 2688,53)	(790,74- 869,82)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Pico 1 (10-12)	(869,82- 948,89)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0-79,07)
Pico 2 (19-21)	(1423,34- 1502,41)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(632,59- 711,67)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Lunes	(632,59- 711,67)	(2530,38- 2609,46)							(2609,46- 2688,53)		(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 1 (1-4)			(553,52- 632,59)				(79,07- 158,14)			(632,59- 711,67)			
Bajo 2 (24)	(0- 4279,60)										(158,14- 237,22)		

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)				(79,07- 158,14)									(79,07- 158,14)
Medio 2 (13-18)	(1976,86- 2055,93)								(2214,08- 2293)				
Medio 3 (22-23)	(2530,38- 2609,46)						(237,22- 316,29)		(2293,16- 2372,23)				
Pico 1 (10-12)			(474,44- 553,52)										
Pico 2 (19-21)	(2530,38- 2609,46)												
Martes	(948,89- 1027,96)			(2688,53- 2767,61)	(948,89- 1027,96)						(0- 4279,60)		(0- 4279,60)
Bajo 1 (1-4)						(2214,08- 2293,16)			(632,59- 711,67)	(0-79,07)		(553,52- 632,59)	
Bajo 2 (24)	(0- 4279,60)								(1265,19- 1344,26)				
Medio 1 (5-9)											(158,14- 237,22)		
Medio 2 (13-18)	(2688,53- 2767,61)										(2530,38- 2609,46)	(158,14- 237,22)	
Medio 3 (22-23)													
Pico 1 (10-12)													(0-79,07)

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)	(2688,53- 2767,61)												
Miércoles											(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 1 (1-4)				(553,52- 632,59)	(2214,08- 2293,16)				(2372,23- 2451,31)		(0-79,07)	(632,59- 711,67)	
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)					(2530,38- 2609,46)								(79,07- 158,14)
Medio 2 (13-18)					(2451,31- 2530,38)								
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Jueves	(1107,04- 1186,11)					(2688,53- 2767,61)	(1581,49- 1660,56)				(0- 4279,60)	(0- 4279,60)	
Bajo 1 (1-4)				(2214,08- 2293,16)					(632,59- 711,67)				
Bajo 2 (24)	(0- 4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)	(1344,26- 1423,34)												
Medio 2 (13-18)					(158,14- 237,22)				(1265,19- 1344,26)				
Medio 3 (22-23)	(2688,53- 2767,61)									(316,29- 395,37)			
Pico 1 (10-12)					(158,14- 237,22)								
Pico 2 (19-21)	(2688,53- 2767,61)												
Viernes	(632,59- 711,67)										(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 1 (1-4)							(632,59- 711,67)					(553,52- 632,59)	(2214,08- 2293,16)
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)													
Medio 2 (13-18)													
Medio 3 (22-23)	(2530,38- 2609,46)												
Pico 1 (10-12)													

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pico 2 (19-21)	(2530,38- 2609,46)												(2372,23- 2451,31)
Sábado				(553,52- 632,59)							(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 1 (1-4)		(2214,08- 2293,16)						(632,59- 711,67)					(532,52- 632,59)
Bajo 2 (24)	(0- 4279,60)												
Medio 1 (5-9)				(632,59- 711,67)									
Medio 2 (13-18)		(1265,19- 1344,26)											
Medio 3 (22-23)													
Pico 1 (10-12)													
Pico 2 (19-21)													
Domingo						(1581,49- 1660,56)					(0- 4279,60)	(0- 4279,60)	(0- 4279,60)
Bajo 1 (1-4)		(1344,26- 1423,34)	(632,59- 711,67)		(0-79,07)	(532,52- 632,59)	(2214,08- 2293,16)						
Bajo 2 (24)	(0- 4279,60)												

Mes		Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Medio 1 (5-9)		<i>(1265,19- 1344,26)</i>											
Medio 2 (13-18)													
Medio 3 (22-23)	<i>(2530,38- 2609,46)</i>												
Pico 1 (10-12)			<i>(79,07- 158,14)</i>										
Pico 2 (19-21)	<i>(2530,38- 2609,46)</i>								<i>(553,52- 632,59)</i>				

Gráficos

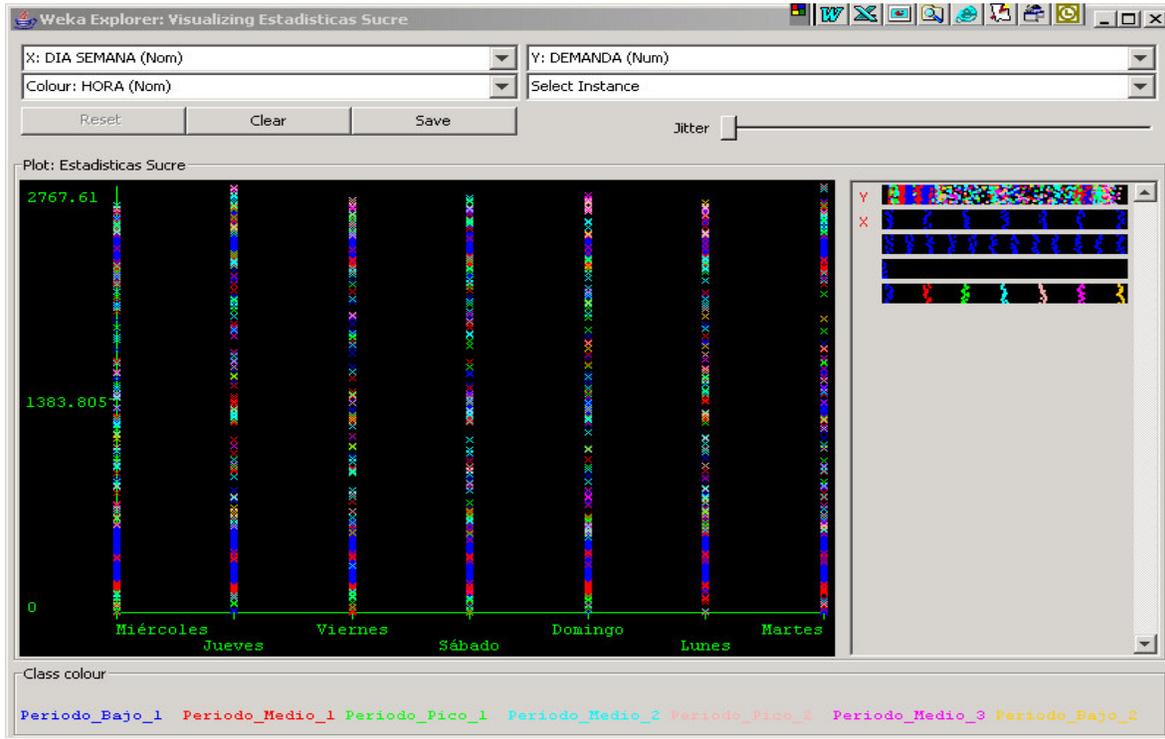


Figura 29. Demanda días de la semana diferenciada por periodos del día (Sucre)

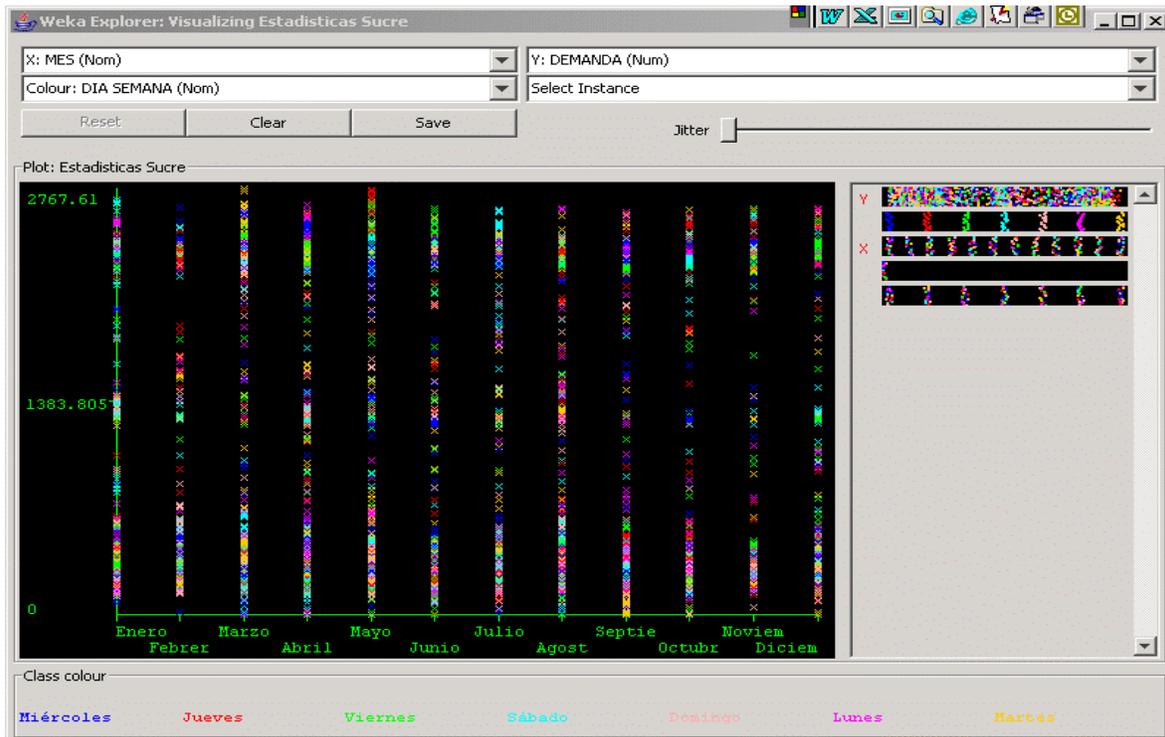
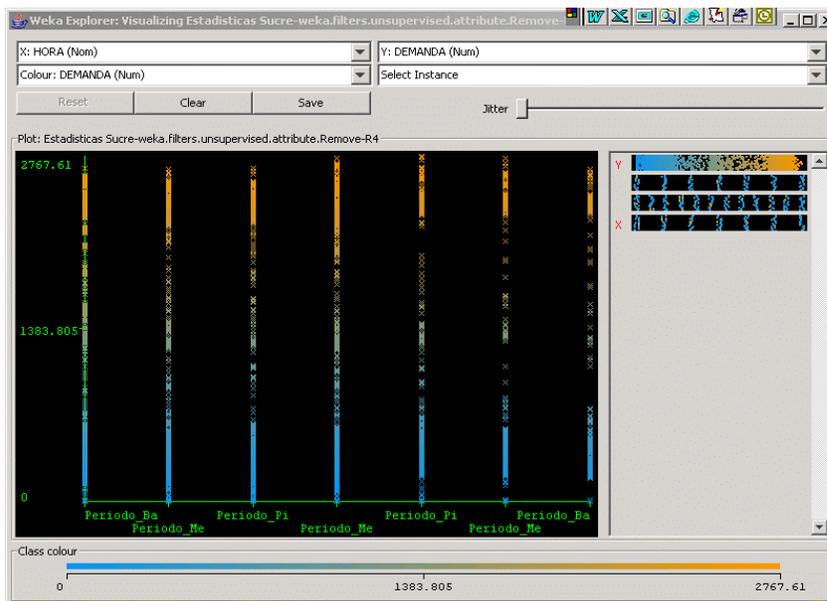


Figura 30. Demanda meses del año diferenciada por días de la semana (Sucre)



**ORDEN
SECUENCIAL**

P. Bajo 1: (1–4)
P. Medio 1: (5–9)
P. Pico 1: (10–12)
P. Medio 2: (13–18)
P. Pico 2: (19–21)
P. Medio 3: (22–23)
P. Bajo 2: (24)

Figura 31. Demanda periodos del día (Sucre)

2.3.5.13 Análisis Departamento: SUCRE

La demanda del departamento de Sucre es la segunda demanda más baja del sector Minero, con media de 456,76 kW/h y desviación estándar igual a 596,91 kW/h.

La mayor proporción de datos está concentrada en los intervalos (158,14-237,22) kW/h y (237,22-316,29) kW/h, que representan el 42,39% y el 21,59% de la demanda, respectivamente.

El día de la semana con la mayor demanda promedio del departamento es el jueves con demanda media de (1.107,04-1.186,11) kW/h y demanda máxima de (2.688,53-2.767,61) kW/h durante los periodos pico 2 y medio 3.

A este día le sigue la demanda del martes (D. Media: 948,89-1.027,96) kW/h, cuya máxima se presenta con igual valor durante los periodos medio 2 y pico 2 (2.688,53-2.767,61) kW/h.

Con las menores demandas promedio están los días lunes y viernes con igual demanda media (632,59-711,67) kW/h, y ambos con demanda máxima de (2.530,38-2.609,46) kW/h, durante los periodos pico 2 y medio 3. Siendo estos últimos los periodos con la mayor demanda promedio del departamento, con medias de (1.423,34-1.502,41) kW/h y (1.265,19-1.344,26) kW/h, respectivamente.

Durante estos periodos, se presentan a demás, las demandas máximas de los días jueves (D. Máxima: 2.688,53-2.767,61) kW/h y domingos (D. Máxima: 2.530,38-2.609,46) kW/h.

Mientras tanto, los periodos del día con menor demanda promedio resultan ser los periodos medio 1 y pico 1 con igual demanda media (869,82-948,89) kW/h.

Por otra parte, pero continuando con el comportamiento de la demanda durante los periodos del día, cabe señalar que el periodo bajo 1 en el departamento de Sucre presenta mayor frecuencia de ocurrencia dentro de los siguientes intervalos de valores de manera particular: (553,52-632,59) kW/h, (632,59-711,67) kW/h y (2.214,08-2.293,16) kW/h.

En lo referente a los meses del año, el de mayor demanda es el mes de Mayo, con demanda media de (1.265,19-1.344,26) kW/h, cuya demanda máxima se presenta durante los jueves (2.688,53-2.767,61) kW/h. Le sigue muy de cerca la demanda media del mes de Marzo (1.186,11-1.265,19) kW/h, con demanda máxima de (2.688,53-2.767,61) kW/h, presente durante los días martes.

La menor demanda promedio mensual, se presenta durante Agosto (D. Media: 711,67-790,74) kW/h, Septiembre y Octubre, con igual valor medio (632,59-711,67) kW/h.

3 CRÍTICA DE DATOS

En este capítulo serán expuestos los resultados de la crítica de datos realizada a las cinco variables más consultadas en NEON: servicio de información sobre el Mercado de Energía Mayorista Colombiano, que permite obtener en forma interactiva, información que facilita el seguimiento y comprensión de la dinámica de dicho mercado.

Este servicio está soportado en la más grande bodega de datos del sector eléctrico colombiano: AQUARIUS; pero al igual que otras fuentes de datos, sufre problemas como el ruido en la información; ocasionado por la presencia de datos anómalos, fuera de rango, erróneos, información no disponible, valores duplicados y demás.

Debido a que este aplicativo es de uso libre para todos los agentes del sector eléctrico (generadores, transportadores, distribuidores y comercializadores) y partes interesadas (inversionistas, entes gubernamentales, entidades nacionales e internacionales) con licencia de acceso al mismo a través de un contrato; la información aquí accedida debe ser oportuna y consistente con la realidad.

Esta fase de la investigación se llevó a cabo mediante la utilización de la herramienta Clementine, uno de los sistemas de minería de datos más populares del mercado, con alta utilidad en el preprocesamiento de datos. El uso de esta herramienta se realizó en línea, es decir, en conexión directa con la base de datos.

Los resultados obtenidos a través de la crítica de datos consisten básicamente en formatos excel que permiten el repote de los causales de ruido, a demás de gráficas arrojadas por la herramienta que facilitan su visualización.

Variables más consultadas en NEON

El aplicativo NEON emite anualmente un reporte con el número de consultas promedio hechas por los usuarios a cada una de las variables almacenadas en el Data Warehouse.

Teniendo en cuenta el número de consultas hechas en el 2004 y 2005, las siguientes, resultaron ser las variables más consultadas por los usuarios del aplicativo.

Variable	Número de Consultas	
	2004	2005
Generación Ideal	74.719	32.521
Generación Real	28.773	66.968
Precio de Oferta	72.144	25.033
Disponibilidad de Generación	8.818	8.195
Capacidad Remunerable Real	3.389	10.670

Cuadro 26. Variables más consultadas en NEON

La crítica de datos fue realizada a cada una de las variables teniendo en cuenta su Entidad (subvariable, que permite mayor detalle y profundidad al análisis), y Resolución (periodicidad de tiempo en el que se tiene información acerca de la variable), con un alcance temporal, del inicio del Mercado: 20 de Julio de 1995 hasta la fecha del análisis.

Únicamente serán expuestos los resultados para cada variable según su respectiva Entidad y/o Resolución para la cual fue detectado algún tipo de traumatismo.

3.1 GENERACION IDEAL

3.1.1 INFORMACIÓN GENERAL

Descripción: Generación obtenida luego de realizar un despacho por estricto orden de precio, hasta logra el cubrimiento total de la demanda.

Reglas del Negocio: Solo existe información para Generadores. La Generación Ideal para el Sistema nunca, por ningún motivo debe ser cero, ni menor que cero.

Unidad de medida: kW/h.

Entidad: Sistema.

Resolución: Mensual, diaria y horaria.

Alcance Temporal: 01/08/1995 - 30/11/05.

3.1.2 FORMATO DE REPORTE

Cuadro 27. Reporte Generación Ideal

ENTIDAD	Resolución			Sistema		
	FECHA	H	D	M	Valor	
Anómalo					Erróneo	ND
dic-95/jun-96			x			x
oct-97/feb-98			x			x
01/12/95 - 30/06/96		x				x
11/10/97 - 28/02/98		x				x
jul-95			x		0	
ago-95			x		113.069.370,00	
20/07/95 - 31/07/95		x			0	
02/08/95 - 31/08/95		x			0	
16/03/2002 23:00	x			1.335.884,57		
17/03/2002 0:00	x			1.687.799,80		
17/03/2002 1:00	x			2.534.978,20		

3.1.3 GRÁFICAS SOPORTE

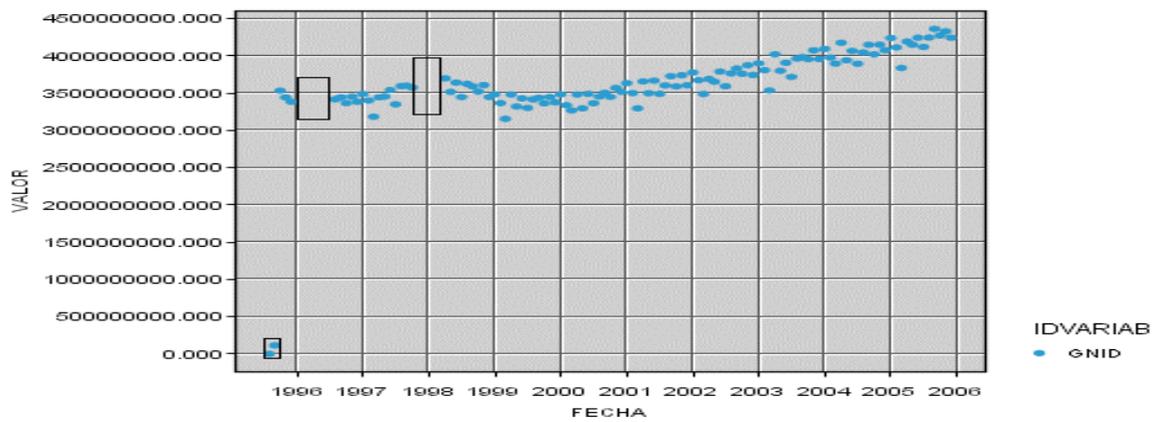


Figura 32. Generación Ideal mensual

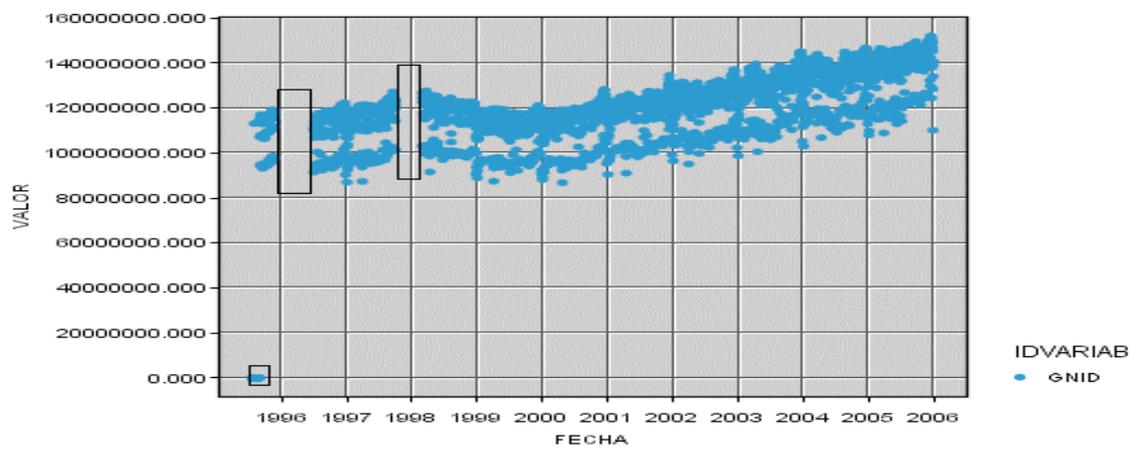


Figura 33. Generación Ideal diaria

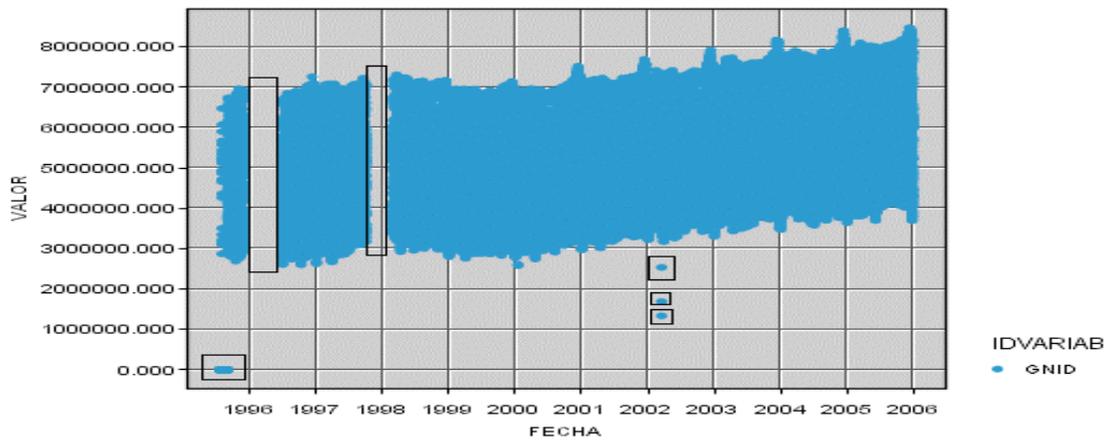


Figura 34. Generación Ideal horaria

3.1.4 RESULTADOS

En términos generales fueron detectados 3 causales de ruido en la formación referente a Generación Ideal:

Información No Disponible, en los siguientes rangos de tiempo:

01-12-95 / 30-06-96 y 11-10-97 / 28-02-98

Información Errónea, que corresponde a la Generación Ideal con valor de cero en los intervalos de tiempo entre:

02-08-95 / 31-08-95, y el correspondiente al mes de Julio de 1995 (20-07-95 / 31-07-95), el cual debería ser inferior al resto de valores mensuales, ya que su registro inicia desde el día 20 (comienzos del Mercado); pero nunca debería ser igual a cero para el Sistema, según reglas del negocio.

Y el valor erróneo del mes de Agosto, que registra 113.069.370 kW/h mensuales, valor muy por debajo de la Generación Ideal mensual promedio, la cual es superior a los 3.000.000.000 kW/h. Esta situación se debe a que durante este mes el único valor de Generación Ideal reportado fue el del 01-08-95 = 113.069.370 kW/h, y el resto de días del mes presentan valores de cero.

Valores Anómalos, con sus respectivas fechas de ocurrencia:

16/03/2002 23:00 = 1.335.884,57 kW/h

17/03/2002 0:00 = 1.687.799,80 kW/h

17/03/2002 1:00 = 2.534.978,20 kW/h

Los cuales podrían resultar erróneos, ya que se encuentran por debajo del valor horario promedio de la Generación Ideal para el año 2002, el cual es superior a los 3.000.000 kW/h; o podrían deberse a un suceso inusual específico por corresponder a un intervalo de tiempo continuo. Ver Figura 34. Generación Ideal horaria.

3.2 GENERACION REAL

3.2.1 INFORMACION GENERAL

Descripción: Generación neta que entrega cada una de las plantas al Sistema en la operación real. Incluye las importaciones internacionales.

Reglas del Negocio: Solo existe información para Generadores.

La Generación Real para el Sistema nunca, por ningún motivo debe ser cero, ni menor que cero.

Unidad de medida: kW/h.

Entidad: Sistema.

Resolución: Diaria y horaria.

Alcance Temporal: 20/07/95 - 30/11/05

3.2.2 FORMATO DE REPORTE

Cuadro 28. Reporte Generación Real

ENTIDAD	Resolución			Sistema		
	H	D	M	Valor		
FECHA				Anómalo	Erróneo	ND
oct-97/ feb-98			x			x
11/10/97 - 28/02/98		x				x
16/03/2002 23:00	x			1.335.884,57		
17/03/2002 0:00	x			1.687.799,80		
17/03/2002 1:00	x			2.534.978,20		

3.2.3 GRÁFICAS SOPORTE

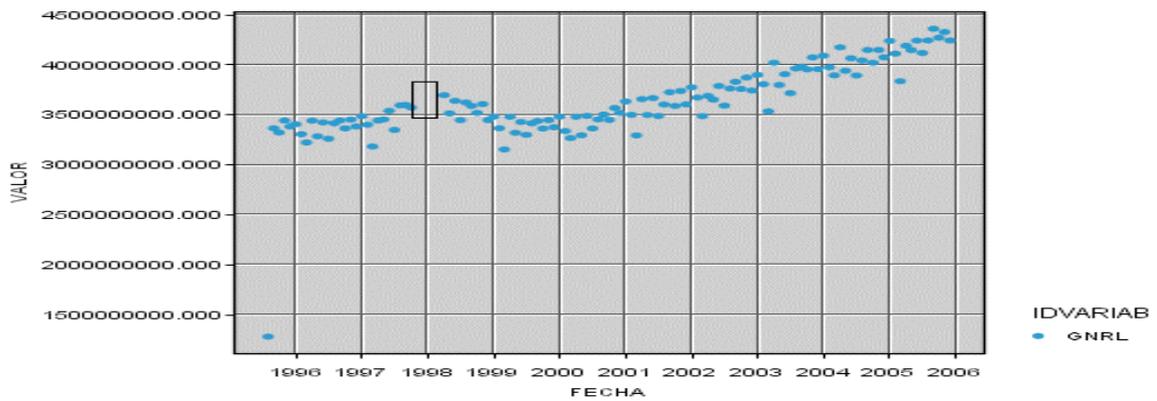


Figura 35. Generación Real mensual

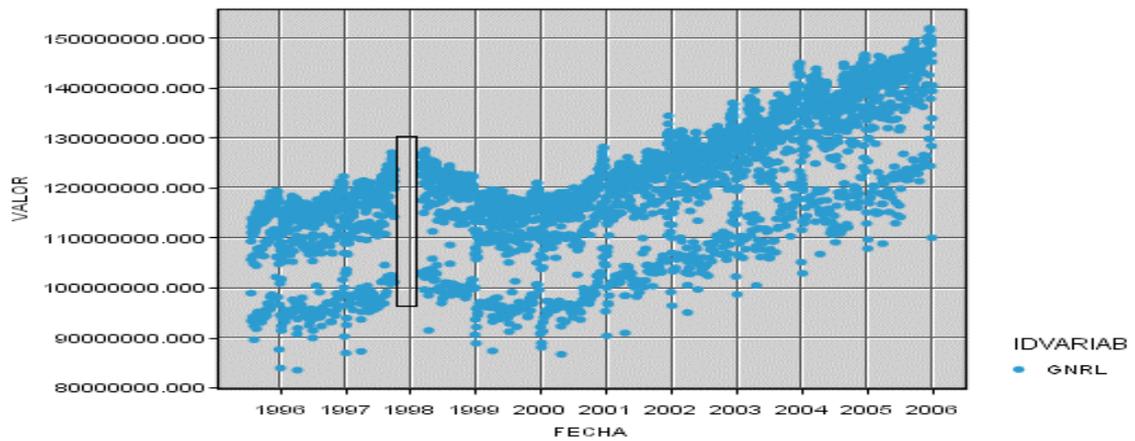


Figura 36. Generación Real diaria

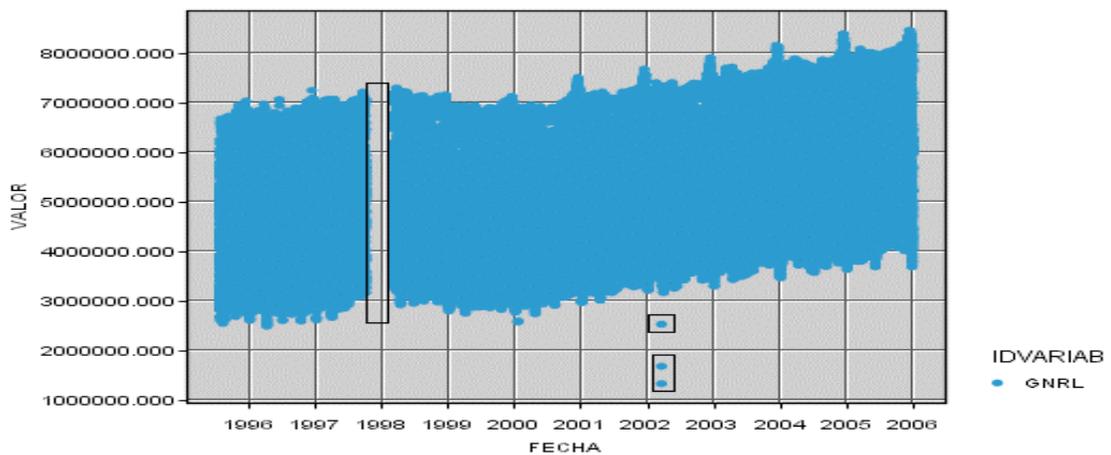


Figura 37. Generación Real horaria

3.2.4 RESULTADOS

Los causales de ruido detectados en el análisis de la variable Generación Real fueron los siguientes:

Información No Disponible para las fechas comprendidas entre:

11/10/97 - 28/02/98.

Datos Anómalos, que podrían resultar siendo erróneos, a menos que un suceso inusual se halla presentado ocasionando durante los correspondientes periodos de tiempo, valores por fuera del límite del comportamiento normal de la variable.

Los datos relacionados a sus respectivas fechas de ocurrencia, son los siguientes:

16/03/2002 23:00 = 1335884,57 kW/h

17/03/2002 0:00 = 1687799,8 kW/h

17/03/2002 1:00 = 2534978,2 kW/h

Ver Figura 37. Generación Real horaria.

3.3 PRECIO DE OFERTA

3.3.1 INFORMACION GENERAL

Descripción: Precio diario al cual está dispuesto a vender la energía al Sistema un recurso de generación. Es el mismo para las 24 horas del día y su oferta se realiza diariamente.

Reglas del Negocio: Sobre este tema, existe información disponible únicamente para Centrales Generadoras despachadas centralmente, es decir, aquellas plantas con Capacidad Efectiva Neta (máxima cantidad de potencia neta que puede suministrar una unidad de generación) superior a los 20 MW. Por ningún motivo el Precio de Oferta puede ser cero; mínimo debe ser igual al Costo Equivalente de Energía, según resolución CREG (Comité Regulatorio de Energía y Gas).

Unidad de medida: \$/kWh.

Entidad: Central.

Resolución: Horaria.

Alcance Temporal: 01/08/1995 - 05/01/06.

Observaciones: Debido al hecho de que esta variable solo presenta información para Centrales (250, reportadas con valor) de manera horaria, no se elaboraron gráficas soporte (no resultan claras por el número de centrales), ni se mostrará el formato de reporte completo por ser absolutamente extenso (se dará a conocer solo una pequeña muestra), pero en su caso, en el apartado de resultados serán expuestos todos los causales de ruido detectado para esta variable.

3.3.2 FORMATO DE REPORTE

Plantas no Despachadas Centralmente.

Cuadro 29. Reporte Precio de Oferta

ENTIDAD		Resolución	Central			
FECHA		H	ID	Valor		
				Anómalo	Erróneo	ND
05/03/1997 11:00	03/08/1998 23:00	x	ALAG		0	
01/09/2000 0:00	05/01/2006 23:00	x	AMR1		0	
20/09/2000 0:00	05/01/2006 23:00	x	ASN1		0	
21/10/2000 0:00	05/01/2006 23:00	x	AYR1		0	
01/08/1998 0:00	05/01/2006 23:00	x	BIOE		0	
05/03/1997 11:00	03/08/1998 23:00	x	BJAG		0	
01/11/1997 0:00	04/02/1998 23:00	x	BLL1		0	
26/05/1997 8:00	17/04/1998 23:00	x	BLL2		0	
01/09/2000 0:00	05/01/2006 23:00	x	BLL3		0	
21/01/2000 0:00	05/01/2006 23:00	x	BLM1		0	
09/07/1997 3:00	13/04/1998 23:00	x	BRN1		0	
14/09/1997 18:00	09/04/1998 23:00	x	BRN2		0	
25/08/1997 0:00	30/01/2005 23:00	x	BRN3		0	
01/01/2003 0:00	05/01/2006 23:00	x	BSQ1		0	
01/09/2000 0:00	05/01/2006 23:00	x	BYN1		0	
15/08/2004 0:00	05/01/2006 23:00	x	CAS1		0	

27/09/2000 0:00	05/01/2006 23:00	x	CCN1		0	
02/05/1997 19:00	30/09/1997 23:00	x	CDF1		0	
07/07/1997 2:00	10/07/1997 8:00					
12/07/1997 21:00	12/07/1997 23:00					
17/07/1997 17:00	17/07/1997 20:00					
25/07/1997 18:00	28/07/1997 17:00					
10/08/1997 13:00	10/08/1997 20:00					
08/09/1997 20:00	08/09/1997 22:00					
02/06/1998 10:00	05/06/1998 23:00					
13/06/1998 17:00	13/06/1998 23:00					
28/06/1998 0:00	01/07/1998 23:00					
13/07/1998 9:00	01/08/1998 23:00					
03/08/1998 0:00	11/08/1998 23:00	x	CHBG		0	
01/01/2003 0:00	05/01/2006 23:00	x	CHL1		0	
01/11/1997 00:00	22/03/1999 23:00	x	CHN4		0	
01/11/1997 00:00	22/03/1999 23:00	x	CHN5		0	

3.3.3 RESULTADOS

A la variable Precio de Oferta, le fue detectado básicamente el siguiente causal de ruido, con su respectiva descripción.

Datos Erróneos: Violando las reglas de negocio de tipo regulatorio, se le relaciona Precio de Oferta a plantas no despachadas centralmente entre las que podemos mencionar: AMR1, ASN1, AYR1, BIOE, BLL3, BLM1, BSQ1, BYN1, CAS1, CCN1, CHL1, CMN1, CMP2, CPT1, CQT1, CRC1, HRD1, INC1, INP1, INS1, INT1, IQU1, IQU2, IRG1, JBV1, JNC1, JPR1, LBR1, LIM1, LMR1, entre muchas otras; y a plantas en periodos inactivos.

Se registran Precios de Oferta con valor de cero, lo cual es incorrecto desde la resolución 55 de la CREG de 1996, que dice “Para efectos de Precio de Oferta el CEE (Costo Equivalente de Energía) debe incluirse como costo variable del generador, el cual debe estar reflejado en los precios que las empresas generadoras ofrezcan diariamente”, razón por la cual, el Precio de Oferta debería ser mínimo igual al CEE. Situación que se presenta para 190 centrales de las 250 reportadas.

3.4 DISPONIBILIDAD DE GENERACION

3.4.1 INFORMACION GENERAL

Descripción: Máxima cantidad de potencia neta que un generador puede suministrar al sistema durante un intervalo de tiempo determinado.

Reglas del Negocio: El valor del Sistema se obtiene de la suma de la disponibilidad de todos los recursos de generación para la fecha, por lo tanto, por ningún motivo puede llegar a ser cero, ni menor a este.

No incluye plantas menores ni cogeneradores (plantas generadoras no despachadas centralmente).

Unidad de medida: MW.

Entidad: Sistema y Central.

Resolución: Diaria.

Alcance Temporal: 11/08/1997 - 06/22/2006

Observaciones: Únicamente serán expuestos los resultados de la crítica para la entidad Sistema, más sin embargo, se hará mención de los traumatismos detectados para la Entidad Central.

3.4.2 FORMATO DE REPORTE

Cuadro 30. Reporte Disponibilidad de Generación

ENTIDAD		Resolución	Sistema		
FECHA		D	Valor		
			Anómalo	Erróneo	ND
16/08/1997 0:00	13/09/1997 0:00	x			
01/10/1997 0:00	04/10/1997 0:00				x
17/05/1998 0:00				158	
30/08/1999 0:00				15.130,66	

3.4.3 GRÁFICAS SOPORTE

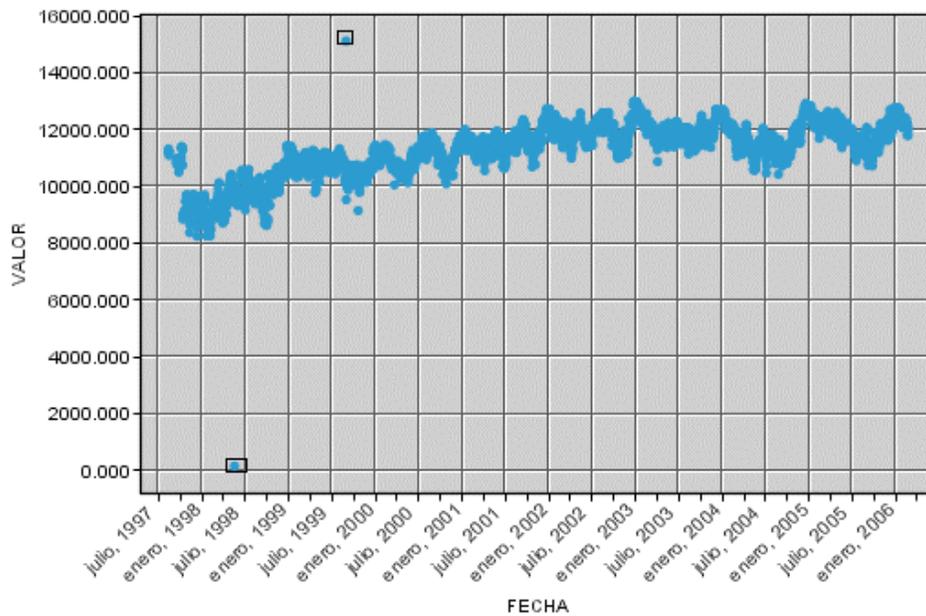


Figura 38. Disponibilidad de Generación para la entidad Sistema

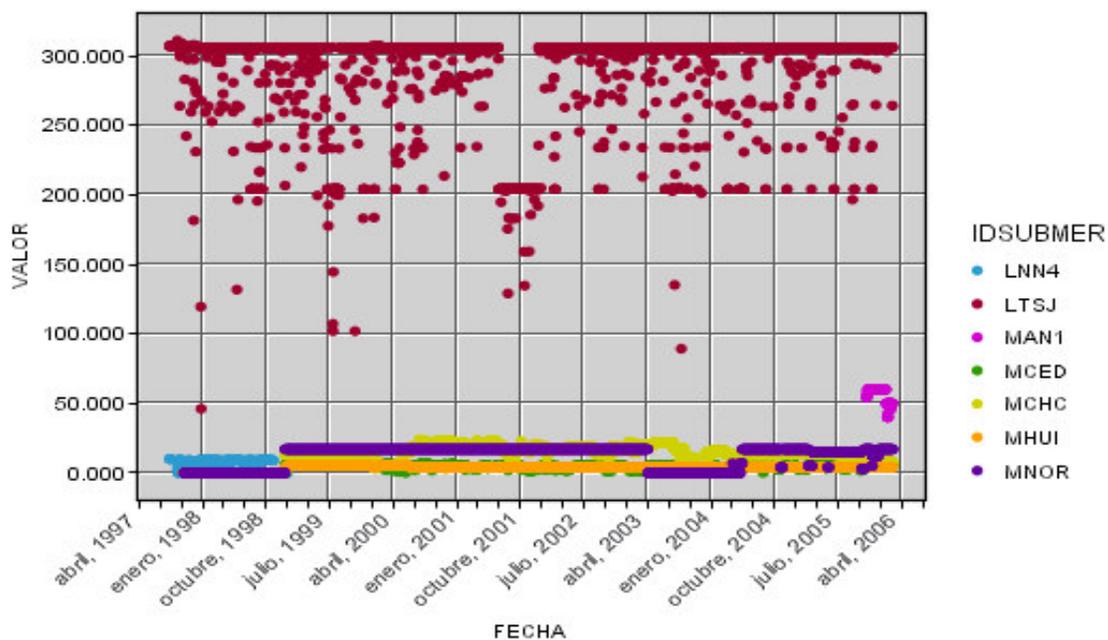


Figura 39. Disponibilidad de Generación para la entidad Central (LNN4-MNOR)

3.4.4 RESULTADOS

El ruido detectado en la información almacenada de esta variable fue básicamente causado por:

Información No Disponible, para los intervalos de tiempo de:

16/08/1997 0:00 al 13/09/1997 0:00

01/10/1997 0:00 al 04/10/1997 0:00

Datos Erróneos, con su respectiva fecha de ocurrencia:

17/05/1998 0:00 = 158 MW

30/08/1999 0:00 = 15.130,66 MW

Ver Figura 38. Disponibilidad de Generación para la entidad Sistema.

Entre los errores detectados para la Resolución de Central, está el hecho de que de manera equívoca (definido por reglas de negocio) en esta variable se incluye información para plantas menores y cogeneradores, como es el caso de las plantas INS1, MAN1, MCED, MCHC, MHUI, MNOR, MTOL y TPD1. Ver Figura 39. Disponibilidad de Generación para la entidad Central (LNN4-MNOR).

Incluye a demás información para plantas inactivas como: CHN1, CHN2 y RMAR, cuya fecha de retiro ha sido anterior al inicio del mercado, y a plantas que durante el desarrollo del mismo se han ido retirando, sin ser excluidas del proceso de carga de información del Data Warehouse, tal es el caso de plantas como: CLDR, CSP1, CSP2, CSP3, CSP4, CSP5, GLD1, RNGR, TBU1, TBU2, TBU3, TOP1 y TOP2, plantas que aún después de estar inactivas se le relacionando valor de Disponibilidad de Generación.

Otro de los errores se relaciona a la existencia de plantas cuyo único valor de Disponibilidad a través de sus años de actividad en el Mercado es cero, como las plantas CHN4, CHN7, CSP1, CSP2, CSP3, ECU2, LNN3 y TBU2.

3.5 CAPACIDAD REMUNERABLE REAL

3.5.1 INFORMACION GENERAL

Descripción: Es la parte de la Capacidad Remunerable Teórica (Capacidad de generación que cada planta despachada centralmente aporta en un despacho al abastecimiento de la demanda en condiciones hidrológicas críticas) que estuvo disponible para el abastecimiento de la demanda.

Reglas del Negocio: Solo aplica a Generadores.

No incluye plantas menores ni cogeneradores (plantas no despachadas centralmente).

Unidad de medida: Kw.

Entidad: Sistema y Central.

Resolución: Mensual.

Observaciones: Únicamente serán expuestos los resultados de la crítica para la entidad Sistema, más sin embargo, se hará mención de los traumatismos detectados para la Entidad Central.

3.5.2 FORMATOS DE REPORTE

Cuadro 31. Reporte Capacidad Remunerable Real

ENTIDAD		Resolución	Sistema		
FECHA			M	Valor	
			Anómalo	Erróneo	ND
mar-97		x		254.000	
oct-97	dic-97	x			x

3.5.3 GRÁFICAS SOPORTE

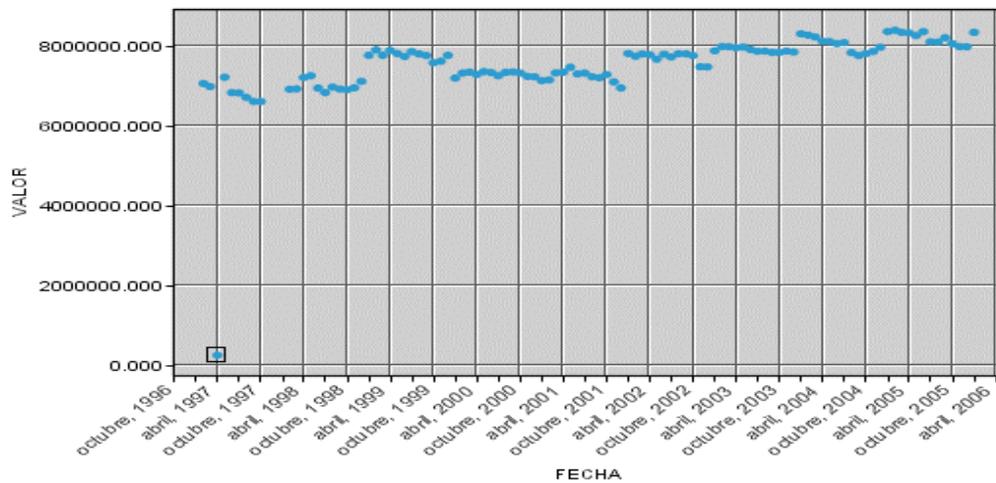


Figura 40. Capacidad Remunerable Real para la entidad Sistema

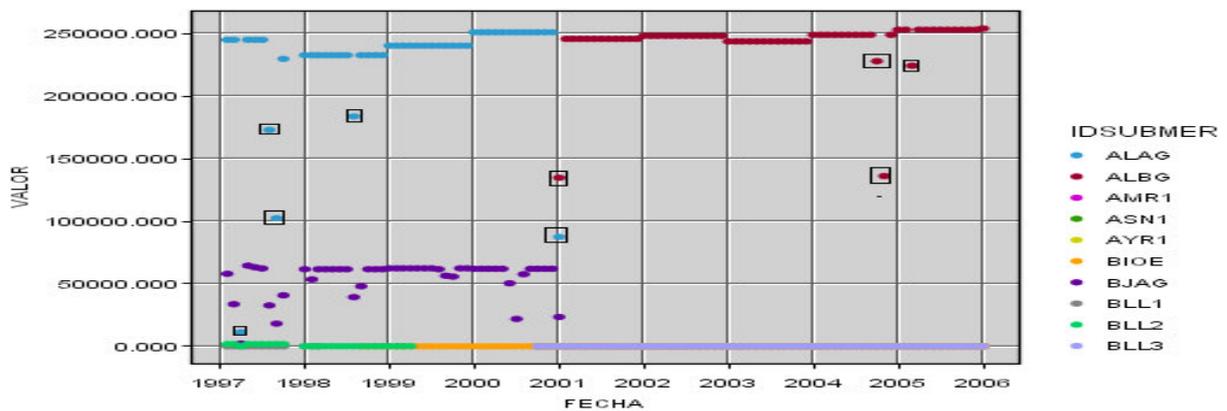


Figura 41. Capacidad Remunerable Real para la entidad Central (ALAG-BLL3)

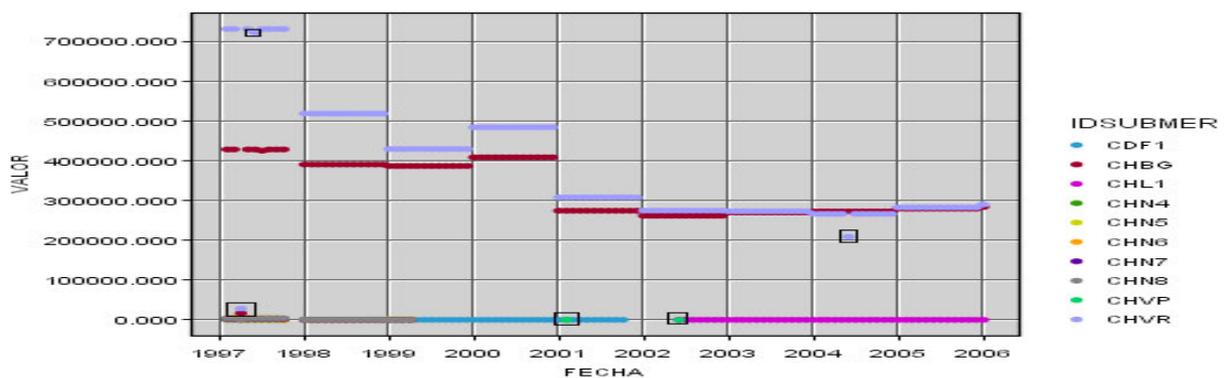


Figura 42. Capacidad Remunerable Real para la entidad Central (CDF1-CHVR)

3.5.4 RESULTADOS

El ruido detectado en la información de la variable Capacidad Remunerable Real consiste básicamente en:

La presencia de un dato erróneo de 254.000 kW durante el mes de Marzo de 1997, donde el orden de valores normales oscila entre los 7.000.000 kW y los 8.500.000 kW.

Información No Disponible para el intervalo de tiempo comprendido entre el mes de Octubre de 1997 a Diciembre de 1997.

Ver Figura 40. Capacidad Remunerable Real para la entidad Sistema.

En el análisis de la variable para la Entidad Central fueron detectadas a demás de la presencia de datos anómalos independientes a cada central, las siguientes inconsistencias:

Son incluidas las plantas no despachadas centralmente relacionándoles valor de cero, como es el caso de las plantas: AMR1, ASN1, AYR1, BIOE, BLL3, BLM1, BSQ1, BYN1, CAS1, CCN1, CHL1, CMN1, CMP2, CPT1, CQT1, CRC1, CRZO, CRZ1, CRZ2, CRZ3, CSC1, HRD1, INC1, INP1, INS1, entre muchas otras. Ver Figura 41. Capacidad Remunerable Real para la entidad Central (ALAG-BLL3).

Existen a demás plantas despachadas centralmente a las que se les es asignado como valor único cero, a todo lo largo de sus años de actividad en el Mercado; estas plantas son: BLL1, CDF1, CHVP, ECD1, PLQ4, ST24, TOP1, TOP2 y URAP. Ver Figura 42 Capacidad Remunerable Real para la entidad Central (CDF1-CHVR).

4 CONCLUSIONES Y RECOMENDACIONES

En este trabajo de grado se propone un enfoque nuevo, mediante el uso de herramientas de Minería de Datos para el análisis de la información de la operación y administración del mercado eléctrico colombiano a través de un caso de estudio, y la crítica de datos a las variables que resultaron de mayor interés para los usuarios del aplicativo NEON.

Los primeros pasos realizados, consistieron en una prospección bastante amplia de las técnicas y conceptos de Minería de Datos más actuales, desarrollándose unos esquemas de clasificación que pueden resultar de gran ayuda para comprender el “estado del arte” del Data Mining.

En este trabajo, se ha intentado demostrar que el uso de estas técnicas de Minería de Datos como tecnología en el manejo de la información, resultan útil no solo en la adquisición de conocimiento a partir de los datos almacenados a lo largo de la vida de una organización, sino, que permiten a demás garantizar la calidad de la información almacenada en sus bodegas de datos, que posiblemente es manejada como aporte a los procesos internos o publicada para los entes externos; en ambos casos apoyando la toma de decisiones.

Los análisis efectuados a partir del caso de estudio han pretendido conocer más a fondo el comportamiento de cada uno de los departamentos de la Costa Atlántica (Región del país con la mayor demanda de energía eléctrica a Febrero de 2006: 716 GWh/mes) en el sector de la economía correspondiente a actividades de explotación de minas y canteras (actividad con el mayor crecimiento de demanda de energía eléctrica a Febrero de 2006: 25,3%); encontrando resultados puntuales

que llegan a ser conocidos por un experto en el negocio después de años de experiencia, pero que gracias a técnicas y herramientas de Minería de Datos pueden descubrirse después de su aplicación.

Los resultados referentes a la crítica de datos, reflejan además de la utilidad de las herramientas de Minería de Datos para la evaluación de la calidad de los mismos, el grado de inconsistencia y traumatismo de la información almacenada en el Data Warehouse (las 5 variables evaluadas: las más consultadas, presentan ruido), la cual es publicada externamente a través del aplicativo NEON, y es consultada por millares de usuarios entre los más frecuentes y a la vez importantes del sector se puede mencionar La Super (Superintendencia de Servicios Públicos), La CREG (Comité Regulatorio de Energía y Gas) y La UPME (Unidad de Planeación Minero Energética).

Es por esto que resulta de suma importancia la aplicación de este tipo de herramientas que evalúen el estado actual de la información almacenada descubriendo la presencia de ruido, para luego trabajar sobre sus causales e impedir su reincidencia; y de esta forma asegurar la calidad de los datos utilizados tanto en los procesos internos de la compañía como aquellos que son emitidos al público en general.

Los resultados obtenidos en el presente trabajo de investigación han sido muy buenos, aunque pueden ser mejorados si para el análisis de la información, se trabaja la herramienta en línea con la base de datos permitiendo con esto un mayor número de registros, y si se intenta con otros tipos de técnicas o formas de entrenamiento de los modelos, dependiendo del caso de estudio a seguir, las necesidades y objetivos de negocio establecidos.

En cuanto a la crítica de datos, la Dirección de Servicios Asociados de XM (dirección donde se desarrolló el estudio) tiene a su cargo el proyecto: “Herramientas Inteligentes para la divulgación de variables operativas y

comerciales del Mercado”, una de sus etapas es el rediseño de la bodega de datos y el levantamiento de reglas de negocio sobre los datos a cargar, es por esto que resulta oportuno tener en cuenta la importancia de continuar la evaluación de la variables restantes del Data Warehouse, ya que los resultados obtenidos permitirán la construcción de nuevas reglas de negocio, asegurando la calidad de la información de la nueva bodega de datos.

BIBLIOGRAFIA

ABAJO, Nicolás. Optimización mediante Data Mining de modelos para el diagnóstico de calidad en hojalata. Tesis Doctoral. Universidad de Oviedo. Gijón, España (2001).

CANTÚ, Francisco. Minería de Datos e Inteligencia de Negocios
Centro de Sistemas Inteligentes.

CLEMENTINE, Spss. Dirección Web:
<http://www.spss.com/spssbi/clementine>.

CRISP-DM. Step-by-step data mining guide. Dirección Web:
<http://www.crisp-dm.org>.

CUADRADO, Abel. Supervisión de procesos complejos mediante técnicas de Data Mining con Incorporación de Conocimiento Previo. Tesis Doctoral. Universidad de Oviedo. Gijón, España (2002).

DAEDALUS. Minería de Datos: Conceptos y Objetivos. Dirección Web:
<http://www.daedalus.es>.

Data Minig: Torturando a los datos hasta que confiesen. Dirección Web:
<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>.

HERNANDEZ, José; RAMIREZ, Maria; FERRI, Cesar. Introducción a la Minería de Datos. PEARSON EDUCATION, S.A., Madrid 2004.

KDNUGGETS. Portal de Minería de Datos. Dirección Web:
<http://www.kdnuggets.com>. KDNuggets.

MAOJO, Víctor. Adquisición de Conocimientos. Departamento de Inteligencia Artificial. Universidad Politécnica de Madrid. (2002).

MARTÍNEZ, F.J.; PERNÍA, A.; CASTEJÓN, M.; GONZÁLEZ, A. Minería de Datos: Herramientas, Técnicas y Metodologías. Memorias: VI International Congress on Project Engineering. Barcelona, España (2002).

Minería y Almacenes de Datos. Dirección Web:
<http://usuarios.lycos.es/sachavir>.

MORALES, E. Descubrimiento de Conocimiento en Bases de Datos. Curso KDD On-Line. Dirección Web: <http://dns1.mor.itesm.mx/~emorales/Cursos/KDD01/>. (2000).

MORENO, Maria; MIGUEL, Luis; GARCIA, Francisco; POLO, Maria. Aplicación de técnicas de Minería de Datos en la construcción y validación de modelos predictivos y asociativos a partir de especificaciones de requisitos de software. Universidad de Salamanca.

SAS. Casos Exitosos. Dirección Web:
<http://www.sas.com/ctx/success/index.jsp>

TORRES, José. Bodegas de Datos como apoyo a la toma de decisiones.

WEKA (Waikato Environment for Knowledge Analysis. Dirección Web:
<http://www.cs.waikato.ac.nz/~ml/WEKA>.