



Universidad  
Tecnológica  
de Bolívar



Escuela de  
Transformación  
Digital

Universidad Tecnológica de Bolívar  
Escuela de Transformación Digital  
Programa de Ciencia de Datos

# **Data Impact Day** **Edición (2026-01)**

Feria de Proyectos de Ciencia de Datos

Memorias de pósters y presentaciones

26 de mayo de 2026  
Campus Tecnológico, Ternera

### **Organizadores**

David Sierra Porta – Universidad Tecnológica de Bolívar  
César Augusto Viloría Núñez – Universidad Tecnológica de Bolívar  
Humberto Manuel Marbello Peña – Universidad Tecnológica de Bolívar  
Fernando Enrique Salcedo Mejía – Universidad Tecnológica de Bolívar  
Skarling Amira Aguilar Ramos – Universidad Tecnológica de Bolívar  
Gloria Mercedes Osorio Viana – Universidad Tecnológica de Bolívar

### **Rector**

Alberto Roa Varelo

### **Vicerrector Académico**

Andrés Guillermo Marrugo Hernández

### **Vicerrectora Administrativa**

María del Rosario Gutiérrez de Piñeres Perdomo

### **Secretaría General**

Ana María Horrillo Caraballo

### **Decano de la Escuela de Transformación Digital**

César Augusto Viloría Núñez

### **Director Programa de Ciencia de Datos – Escuela de Transformación Digital**

David Sierra Porta

### **Compilador de memorias**

David Sierra Porta

ISSN: XXXX-XXXX

Cartagena de Indias, D. T. y C. – Colombia  
www.utb.edu.co  
2026

# DATA IMPACT DAY

## TU PROYECTO, TUS DATOS, TU IMPACTO

#ConDatosUTB #DatoFestUTB  
#EnDatosUTB #DataImpactUTB

Un espacio de socialización y aprendizaje en el que los estudiantes presentan sus proyectos finales de cursos de Ciencia de Datos, conectando resultados técnicos con impacto en la industria y la sociedad.



### El desarrollo de competencias analíticas

- Inteligencia Artificial.
- Desarrollo de modelos con datos reales.



### Conectar la formación con necesidades reales.

- Logística, ambiente, salud, finanzas, energía, ciudad inteligente, etc.



### Buenas prácticas

- Ética de datos.
- Comunicación clara.
- Enfoque en valor.



### Trabajo en equipo

- Fomenta la interdisciplinariedad.
- Ecosistema UTB-empresa.

Inspira trayectorias profesionales con ejemplos concretos de casos de uso de datos



Universidad  
Tecnológica  
de Bolívar



Escuela de  
Transformación  
Digital

✦ PROGRAMA  
DE CIENCIA DE  
DATOS ✦

## Presentación

En el marco del *Data Impact Day*, feria de proyectos de Ciencia de Datos de la Universidad Tecnológica de Bolívar, se presentan en este documento las memorias de los pósters y trabajos desarrollados por los estudiantes de los diferentes cursos del programa.

La socialización de los proyectos en formato de póster se llevó a cabo en los pasillos del Módulo A2 del primer piso del edificio principal.



Figura 1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day.

# Índice general

<b>Presentación</b>	<b>4</b>
<b>1 Ciencia de Datos para transformar realidades</b>	<b>7</b>
1.1 La ciencia de datos en el mundo real	7
1.2 Ética, responsabilidad y cultura de datos	8
1.3 La ciencia de datos en la UTB y la Escuela de Transformación Digital	9
1.4 Líneas y proyectos que conectan datos y territorio	9
1.5 El <i>Data Impact Day</i> como laboratorio de aprendizaje	10
1.6 Mirando hacia adelante	10
<b>2 Áreas temáticas y compromisos con el desarrollo sostenible</b>	<b>11</b>
2.1 Distribución por área temática	11
2.2 Proyectos y Objetivos de Desarrollo Sostenible	12
<b>3 Proyectos del curso de Introducción a la Ciencia de Datos</b>	<b>14</b>
3.1 Impacto de la inteligencia artificial en el aprendizaje de estudiantes de la UTB	14
3.2 Si la UTB tuviera 100 estudiantes ¿Cuántos comerían bien?	14
3.3 Vida académica y bienestar estudiantil	15
3.4 Una foto de la vida universitaria en la UTB	15
3.5 Si la UTB tuviera 100 estudiantes: así usarían la IA y manejarían el estrés	15
3.6 Si la UTB tuviera 100 estudiantes.	16
3.7 Si la UTB tuviese 100 personas	16
3.8 ¿Qué hacen los estudiantes de la UTB cuando no estudian?	17
3.9 Si la UTB Tuviera 100 Estudiantes: Análisis Exploratorio de Datos Académicos y Percepción Estudiantil	17
3.10 Si la UTB tuviera 100 estudiantes como sería la participación en la universidad	18
3.11 ¿Qué hacen los 100 estudiantes de la UTB cuando no estudian?	18
3.12 Si la UTB tuviera 100 estudiantes	18
<b>4 Proyectos del curso de Programación para Ciencia de Datos</b>	<b>20</b>
4.1 Análisis de tasa de accidentalidad en Barranquilla	20
4.2 Análisis de Datos Educativos: Saber Pro 2023 y Factores Socioeconómicos	20
4.3 SummaryData: Implementación de una función de análisis descriptivo avanzado aplicada a datos de COVID-19	21
4.4 Tasas de Interés Activas en el Sistema Financiero Colombiano	21
4.5 DATEnidos	21
4.6 Evaluaciones agropecuarias municipales	22
4.7 Conflicto Armado, Paz y Víctimas en Colombia Análisis Exploratorio de Datos – 2002–2023	22
4.8 Popularidad en el anime	23
4.9 ¿Qué variables macroeconómicas explican la actividad económica de Colombia?	23
4.10 Impacto de las comorbilidades en la gravedad de casos de COVID-19 en México	23
4.11 RIESGO LABORAL, MUNICIPIO A MUNICIPIO	24
4.12 Análisis y Limpieza de Datos del Dataset Lending Club	24
4.13 Estrategias de Optimización Comercial para Starbucks: Un Enfoque Basado en el Análisis de Datos	25
4.14 Crocs Analytcs	25
4.15 Uso y Dependencia de la Inteligencia artificial.	25

4.16	Dashboard ICFES - Análisis Educativo . . . . .	26
4.17	E-commerce . . . . .	26
4.18	Análisis de Vehículos Usados y Factores que Influyen en su Precio de Venta. . . . .	26
<b>5</b>	<b>Proyectos del curso de Modelos de Regresión y Series de Tiempo</b>	<b>28</b>
5.1	Predictur . . . . .	28
5.2	HYPATIA: Sistema Híbrido de Predicción de Trayectorias de Objetos Cercanos a la Tierra . . . . .	28
<b>6</b>	<b>Proyectos del curso de Sensado y modelado de Sistemas Físicos</b>	<b>30</b>
6.1	Clasificación de imágenes satelitales . . . . .	30
6.2	Modelado de la fragmentación del bosque andino en el Valle del Cocora mediante imágenes Sentinel-2 . . . . .	30
<b>7</b>	<b>Proyectos del curso de Machine Learning</b>	<b>32</b>
7.1	Veles . . . . .	32
7.2	EmotiSpeech: Detección Automática de Emociones en Voz para la Optimización de la Atención al Cliente en Call Centers . . . . .	32
7.3	Clasificador inteligente de billetes colombianos mediante visión por computador . . . . .	33
<b>8</b>	<b>Proyectos del curso de Toma de Decisiones Basadas en Datos</b>	<b>35</b>
8.1	Veles . . . . .	35
8.2	Data4Mind . . . . .	35
<b>9</b>	<b>Proyectos del curso de StoryTelling y Narrativa de Datos</b>	<b>36</b>
9.1	Predictur . . . . .	36
9.2	El paraíso que se borra . . . . .	36
<b>10</b>	<b>Reconocimientos y premios</b>	<b>38</b>
	Premios a los primeros lugares . . . . .	38
	Menciones especiales . . . . .	39
	<b>Cierre y agradecimientos</b>	<b>41</b>

# 1 Ciencia de Datos para transformar realidades

Vivimos en un momento en que la capacidad de leer datos se ha vuelto tan fundamental como la de leer texto. Cada sensor, cada transacción, cada interacción digital deja una huella que, bien interpretada, puede revelar patrones invisibles, anticipar crisis, orientar recursos o simplemente ayudarnos a entender mejor el mundo en que vivimos. La ciencia de datos no inventó esa posibilidad, pero le dio un conjunto de herramientas, un lenguaje compartido y una disciplina metodológica que permiten convertir esa huella en conocimiento accionable.

Lo que distingue a la ciencia de datos de otras prácticas analíticas no es solo la escala o la potencia computacional: es la forma de pensar. Formular la pregunta correcta antes de tocar un dataset. Entender el contexto del que provienen los datos antes de modelarlos. Comunicar los hallazgos con honestidad, incluyendo sus limitaciones. Esa combinación de rigor estadístico, destreza computacional y juicio crítico es lo que define a un científico o científica de datos capaz de operar en el mundo real.

Los proyectos reunidos en estas memorias lo demuestran con claridad. Desde la detección de emociones en voz hasta el modelado orbital de asteroides, pasando por el análisis de la fragmentación del bosque andino, el seguimiento de colombianos detenidos en el exterior o la radiografía alimentaria de los propios estudiantes de la UTB, esta edición muestra que la ciencia de datos no tiene un único dominio ni un único perfil de practicante. Tiene, en cambio, un método y una actitud.

## 1.1. La ciencia de datos en el mundo real

La presencia de la ciencia de datos en la vida cotidiana y en los grandes sistemas que organizan la sociedad ya no es tendencia: es realidad consolidada. Su alcance se extiende por prácticamente todos los sectores de actividad humana.

- **Salud y calidad de vida:** los modelos predictivos apoyan el diagnóstico temprano, la gestión de recursos hospitalarios y el diseño de intervenciones en salud pública. El reto en 2026 no es solo construir modelos precisos, sino garantizar que sean equitativos, explicables y éticamente responsables en sus criterios de decisión.
- **Ambiente, territorio y cambio climático:** la integración de imágenes satelitales, sensores en campo y registros históricos permite monitorear ecosistemas, anticipar riesgos y cuantificar el impacto de la actividad humana sobre el territorio con una resolución y una frecuencia sin precedentes.
- **Industria, logística y energía:** la analítica predictiva y el mantenimiento basado en datos han dejado de ser exclusivos de grandes corporaciones. Hoy son accesibles para empresas medianas y organismos públicos que buscan operar con mayor eficiencia y menor incertidumbre.
- **Finanzas y economía:** desde la detección de fraude hasta la modelación macroeconómica, los datos permiten leer señales débiles en entornos volátiles y fundamentar decisiones que antes dependían exclusivamente del criterio experto.
- **Sector público y políticas basadas en evidencia:** los gobiernos y las entidades territoriales disponen hoy de más datos que nunca sobre movilidad, educación, seguridad y bienestar ciudadano. El desafío es traducir esa abundancia en intervenciones concretas, focalizadas y evaluables.



Figura 1.1: Figura tomada de Pexels.com: <https://www.pexels.com/photo/computer-monitor-with-stock-exchange-figures-in-an-office-8204311/>. Modern office with financial trading screens and a diverse team discussing strategies.

- **Ciencia, educación y sociedad:** la analítica de trayectorias estudiantiles, la medición de impacto de programas educativos y el análisis de dinámicas sociales son ya campos maduros que informan políticas institucionales y agendas de investigación.

En todos estos casos, la ciencia de datos funciona como mediador entre la realidad y la decisión: no la reemplaza, pero la hace más informada, más transparente y más reproducible.

## 1.2. Ética, responsabilidad y cultura de datos

A medida que los modelos algorítmicos adquieren mayor influencia sobre decisiones que afectan a personas —acceso a crédito, contratación, diagnóstico médico, vigilancia urbana—, la pregunta por la responsabilidad ética se vuelve insoslayable. Un modelo no es nunca neutral: incorpora supuestos sobre qué variables importan, qué patrones históricos son deseables de replicar y qué tipo de error es más costoso. Ignorar eso no es ingenuidad técnica; es una decisión con consecuencias.

Formar en ciencia de datos en 2026 implica, por tanto, formar también en la capacidad de auditar decisiones algorítmicas, comunicar incertidumbre con honestidad, documentar procedimientos de forma reproducible y reconocer cuándo un modelo no debería usarse aunque funcione bien en métricas estándar. La sofisticación técnica sin reflexión crítica produce herramientas poderosas al servicio de objetivos equivocados.

El *Data Impact Day* refuerza esa cultura desde la práctica: al exponer un proyecto ante pares, docentes e invitados externos, los equipos se ven obligados a defender no solo sus resultados sino sus elecciones. Esa exposición al escrutinio es uno de los aprendizajes más valiosos del evento.

### 1.3. La ciencia de datos en la UTB y la Escuela de Transformación Digital

La **Escuela de Transformación Digital (ETD)** de la Universidad Tecnológica de Bolívar es hoy uno de los espacios académicos más activos de la región Caribe en formación analítica y digital. Sus programas de Ciencia de Datos, Ingeniería en Sistemas y Computación, Comunicación Social y Marketing, y Transformación Digital comparten una convicción: que la transformación digital no es una cuestión de herramientas sino de capacidades humanas para entender, cuestionar y mejorar los sistemas con los que interactuamos.

El **Programa de Ciencia de Datos** articula cuatro dimensiones de formación que se complementan:

- Fundamentos matemáticos y estadísticos que permiten entender por qué funcionan los modelos, no solo cómo usarlos.
- Competencias computacionales en programación científica, ingeniería de datos y arquitecturas de procesamiento.
- Técnicas de aprendizaje automático y aprendizaje profundo aplicadas a problemas con datos reales, ruidosos e incompletos.
- Formación en comunicación de resultados, ética algorítmica y trabajo interdisciplinario con actores no técnicos.

A lo largo de su formación, los estudiantes abordan problemas anclados en el territorio: calidad del aire en ciudades costeras, dinámica de uso del suelo en la región Caribe, series de tiempo económicas y financieras, datos clínicos y epidemiológicos, y fenómenos sociales que requieren datos para ser comprendidos en toda su complejidad.

### 1.4. Líneas y proyectos que conectan datos y territorio

La actividad investigativa del Programa de Ciencia de Datos de la UTB se despliega en líneas que articulan la capacidad analítica con los retos específicos de la región y del país:

- **Calidad del aire y ambiente urbano:** redes de sensado fijo y móvil, integración de fuentes heterogéneas y construcción de indicadores para la gestión ambiental en ciudades del Caribe.
- **Uso del suelo, teledetección y cambio climático:** procesamiento de imágenes satelitales multiespectrales para estudiar cobertura vegetal, expansión urbana, cuerpos de agua y ecosistemas costeros, con implicaciones directas en planificación territorial y gestión del riesgo.
- **Salud, ambiente y bienestar:** análisis de la relación entre variables ambientales y condiciones de salud, modelado de factores de riesgo clínico y evaluación de intervenciones con enfoque preventivo y basado en evidencia.
- **Inteligencia artificial aplicada:** diseño e implementación de modelos de clasificación, predicción y recomendación en sectores como industria, finanzas, salud y servicios, con énfasis en interpretabilidad y robustez.
- **Educación, datos y equidad:** analítica de trayectorias estudiantiles, identificación de factores de riesgo de deserción y apoyo a la toma de decisiones académicas e institucionales con criterios de equidad y pertinencia.

Estas líneas se nutren mutuamente y dialogan con las apuestas de la UTB por una formación pertinente, socialmente responsable y orientada al desarrollo sostenible de la región.

## 1.5. El *Data Impact Day* como laboratorio de aprendizaje

El *Data Impact Day* no es una vitrina de proyectos terminados: es un espacio donde el aprendizaje sigue ocurriendo. En el momento en que un equipo presenta su trabajo ante una audiencia que no participó en su construcción, comienza una fase nueva del proceso: la de someter las decisiones tomadas al escrutinio de otros, defender los supuestos adoptados y recibir perspectivas que el trabajo interno no pudo generar.

Esa experiencia de exposición pública tiene efectos formativos que ningún examen o entrega puede replicar. Los equipos aprenden a:

- Sintetizar meses de trabajo en una narrativa coherente y accesible para audiencias con distintos niveles de familiaridad técnica.
- Responder preguntas sobre decisiones metodológicas que, en el calor del proceso, parecían obvias pero que, vistas desde afuera, merecen justificación.
- Identificar conexiones entre su proyecto y el de otros equipos, descubriendo temas transversales —calidad de datos, equidad algorítmica, comunicación de incertidumbre— que atraviesan dominios muy distintos.
- Reconocerse como parte de una comunidad de práctica con capacidad real de contribuir a la comprensión y transformación de su entorno.

En esta décima edición, esa diversidad es especialmente visible: los proyectos reunidos abarcan desde la predicción de demanda turística en el Caribe colombiano hasta la clasificación de billetes mediante visión por computador, pasando por el análisis del conflicto armado colombiano en dos décadas de datos, la pérdida de manglar en zonas costeras de Cartagena y la detección de emociones en voz para entornos de servicio al cliente. Son 41 proyectos, siete cursos y un denominador común: datos al servicio de preguntas que importan.

## 1.6. Mirando hacia adelante

La ciencia de datos de 2026 no es la de 2020. Los modelos de lenguaje de gran escala, la computación cuántica incipiente, la proliferación de sensores en infraestructuras críticas y la creciente demanda de explicabilidad en los sistemas de decisión están redefiniendo el campo con una velocidad que desafía cualquier currículo estático. El reto para un programa de formación no es perseguir cada novedad tecnológica, sino cultivar en sus estudiantes la capacidad de adaptarse a ellas con criterio: entender los fundamentos, cuestionar los límites y reconocer cuándo una herramienta nueva resuelve un problema real y cuándo solo añade complejidad innecesaria.

Las memorias de esta edición son un registro de ese proceso en curso. Cada proyecto es, al mismo tiempo, un resultado concreto y un punto de partida: para el equipo que lo construyó, para los cursos que vendrán, para las organizaciones que puedan beneficiarse de sus hallazgos y para el programa que sigue aprendiendo de lo que sus estudiantes son capaces de hacer cuando se les da un problema real, las herramientas adecuadas y el espacio para explorar.

El *Data Impact Day* seguirá siendo ese espacio. Y estas memorias, su evidencia.

## 2 Áreas temáticas y compromisos con el desarrollo sostenible

Los proyectos del *Data Impact Day 202610* no son ejercicios académicos aislados: se inscriben en problemas concretos del entorno social, económico y ambiental de Colombia y de la región Caribe. Esta sección ofrece una lectura transversal de los 41 proyectos presentados en esta edición a partir de dos ejes analíticos: las áreas temáticas que abordan y los Objetivos de Desarrollo Sostenible (ODS) con los que se conectan.

### 2.1. Distribución por área temática

Cada proyecto fue clasificado en una de siete áreas temáticas de acuerdo con el dominio de aplicación central de su problema de investigación. La figura 2.1 muestra la distribución resultante.

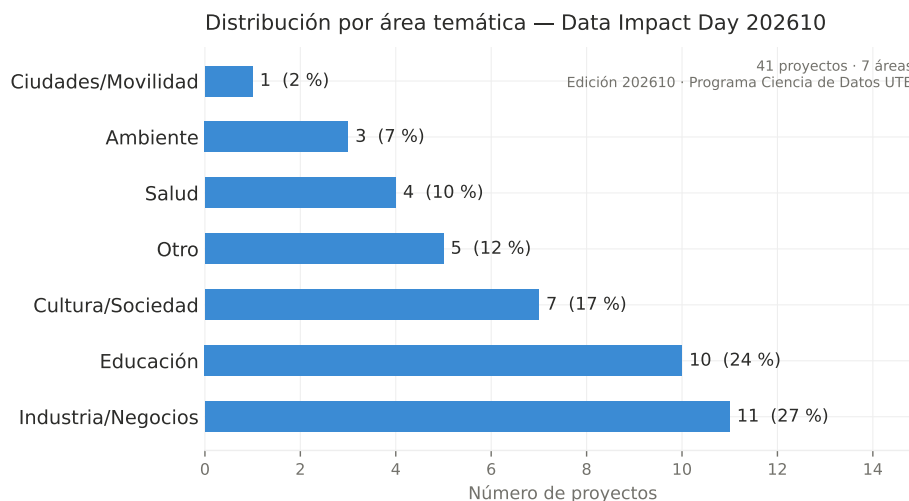


Figura 2.1: Distribución de los 41 proyectos del *Data Impact Day 202610* según área temática.

La mayor concentración de proyectos se ubica en **Industria y Negocios** (11 proyectos, 27 %) y **Educación** (10 proyectos, 24 %), lo que refleja tanto la diversidad de cursos de la Escuela de Transformación Digital como el interés natural de los estudiantes por problemas cercanos a su experiencia universitaria. La categoría **Cultura y Sociedad** reúne siete proyectos (17 %) que analizan fenómenos como el conflicto armado colombiano, el consumo cultural digital y la vida cotidiana en el campus. Los proyectos de **Salud** (cuatro) y **Ambiente** (tres) concentran algunos de los trabajos técnicamente más ambiciosos de la edición, incluyendo análisis de comorbilidades con COVID-19, modelado de fragmentación forestal con imágenes Sentinel-2 y narrativa de pérdida costera en Cartagena. Cierra el mapa un proyecto de **Ciudades y Movilidad** centrado en la accidentalidad vial en Barranquilla.

Esta distribución evidencia que la ciencia de datos en el Programa de la UTB no tiene un perfil temático único: abarca desde el análisis macroeconómico hasta la teledetección ambiental, pasando por el comportamiento estudiantil y la industria musical.

## 2.2. Proyectos y Objetivos de Desarrollo Sostenible

Más allá de su área temática, cada proyecto puede leerse en clave de los Objetivos de Desarrollo Sostenible de la Agenda 2030 de Naciones Unidas. El cuadro 2.1 sistematiza la asociación entre cada proyecto y sus ODS primarios. Un mismo proyecto puede estar vinculado a más de un ODS cuando su problema abarca varias dimensiones del desarrollo sostenible.

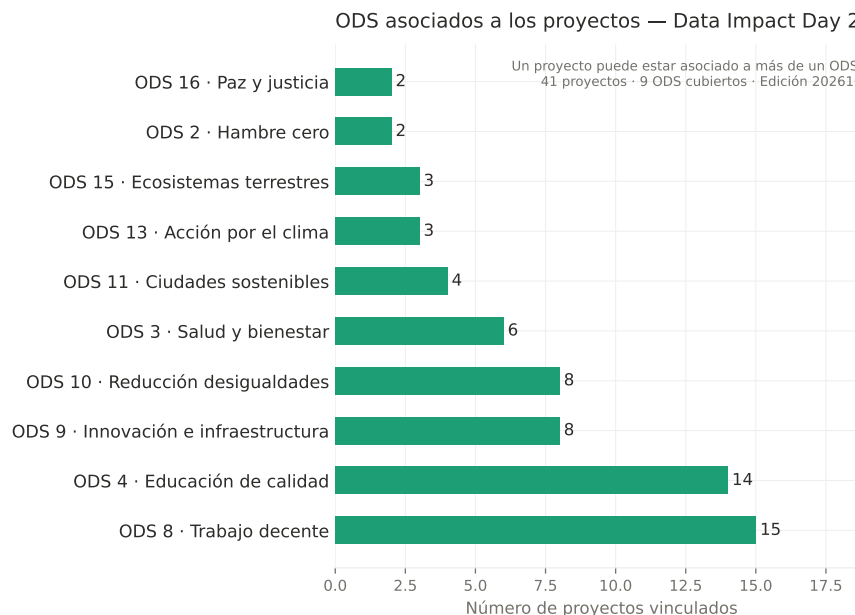


Figura 2.2: Frecuencia de los ODS referenciados en los proyectos del *Data Impact Day 202610*. Un proyecto puede estar asociado a más de un ODS.

Los ODS con mayor presencia son el **ODS 8** —*Trabajo decente y crecimiento económico*— con 15 proyectos, y el **ODS 4** —*Educación de calidad*— con 14. Esta concentración es coherente con los perfiles de los cursos participantes y con el contexto socioeconómico de Colombia: la analítica financiera, el estudio del mercado laboral, la gestión comercial basada en datos y el análisis de factores socioeconómicos asociados al desempeño académico son preguntas recurrentes tanto en el sector productivo como en la política pública.

El **ODS 9** —*Industria, innovación e infraestructura*— y el **ODS 10** —*Reducción de las desigualdades*— empatan en ocho proyectos cada uno. En el primer caso destacan iniciativas como HYPATIA, EmotiSpeech y Veles, que aplican aprendizaje automático avanzado a problemas de frontera. En el segundo, proyectos como el análisis de las pruebas Saber Pro, el dashboard del ICFES y el estudio de tasa de interés revelan cómo los datos pueden iluminar brechas estructurales en educación, finanzas y trabajo.

Con seis proyectos, el **ODS 3** —*Salud y bienestar*— concentra los trabajos más directamente orientados a la salud pública: COVID-19, comorbilidades, bienestar estudiantil y salud mental adolescente. Los ODS de acción ambiental (ODS 13 y ODS 15) están representados por los tres proyectos de sensado y narrativa ambiental —clasificación satelital, fragmentación del bosque andino y pérdida de manglar en Cartagena— que constituyen la línea de trabajo más cercana a las ciencias de la Tierra dentro del programa. Finalmente, el **ODS 16** —*Paz, justicia e instituciones sólidas*— aparece en los proyectos sobre colombianos detenidos en el exterior y sobre el conflicto armado 2002–2023, que utilizan datos abiertos para analizar fenómenos de alta relevancia política y social.

En conjunto, esta edición del *Data Impact Day* toca directamente nueve de los diecisiete ODS, con énfasis claro en educación, trabajo, innovación y equidad. La ausencia de proyectos en ODS como energía asequible (ODS 7), agua limpia (ODS 6) o ciudades inclusivas en dimensión más amplia (ODS 11 más allá de la movilidad) señala, al mismo tiempo, líneas de trabajo que el programa podría

Cuadro 2.1: Asociación de proyectos con ODS primarios.

Proyecto	Curso	ODS
Análisis de accidentalidad en Barranquilla	PCD	ODS 11
Análisis Saber Pro 2023	PCD	ODS 4, ODS 10
SummaryData: COVID-19	PCD	ODS 3
Tasas de Interés Activas	PCD	ODS 8, ODS 10
IA en el aprendizaje UTB	ICD	ODS 4
DATENidos	PCD	ODS 16
Veles	ML / TDBD	ODS 8, ODS 9
Evaluaciones agropecuarias	PCD	ODS 2, ODS 8
UTB: ¿Cuántos comerían bien?	ICD	ODS 2, ODS 3
Vida académica y bienestar	ICD	ODS 3, ODS 4
Conflicto armado 2002–2023	PCD	ODS 16
Predictur	MRST / SND	ODS 8, ODS 11
Foto de la vida universitaria	ICD	ODS 4
Data4Mind	TDBD	ODS 3, ODS 10
EmotiSpeech	ML	ODS 8, ODS 9
UTB: IA y estrés	ICD	ODS 3, ODS 4
Clasificación imágenes satelitales	SMSF	ODS 13, ODS 15
Si la UTB tuviera 100 estudiantes	ICD	ODS 4
HYPATIA	MRST	ODS 9
El paraíso que se borra	SND	ODS 13, ODS 15
Si la UTB tuviese 100 personas	ICD	ODS 4
¿Qué hacen los estudiantes?	ICD	ODS 4
Popularidad en el anime	PCD	ODS 11
Variables macroeconómicas Colombia	PCD	ODS 8, ODS 10
Bosque andino Valle del Cocora	SMSF	ODS 13, ODS 15
Clasificador billetes colombianos	ML	ODS 8, ODS 9
UTB: análisis exploratorio	ICD	ODS 4
UTB: participación universitaria	ICD	ODS 4, ODS 10
Comorbilidades COVID-19 México	PCD	ODS 3
Riesgo laboral municipio a municipio	PCD	ODS 8, ODS 10
Lending Club	PCD	ODS 8, ODS 10
UTB: ¿qué hacen cuando no estudian?	ICD	ODS 4
Starbucks analytics	PCD	ODS 8, ODS 9
UTB: si tuviera 100 estudiantes	ICD	ODS 4
Crocs Analytics	PCD	ODS 8
Uso y dependencia de la IA	PCD	ODS 9, ODS 4
E-commerce Online Retail	PCD	ODS 8, ODS 9
Vehículos usados y precio de venta	PCD	ODS 8
Dashboard ICFES	PCD	ODS 4, ODS 10

ICD: Introducción a la Ciencia de Datos PCD: Programación para Ciencia de Datos

MRST: Modelos de Regresión y Series de Tiempo SMSF: Sensado y Modelado de Sistemas Físicos

ML: Machine Learning TDBD: Toma de Decisiones Basadas en Datos SND: StoryTelling y Narrativa de Datos

explorar en ediciones futuras.

## 3 Proyectos del curso de Introducción a la Ciencia de Datos

### 3.1. Impacto de la inteligencia artificial en el aprendizaje de estudiantes de la UTB

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** Alejandra Guerrero - T00088573- aleguerrero@utb.edu.co, Mathias Franceschi - T00098842- mfranceschi@utb.edu.co Elías Castro -T00098342 - eliacastro@utb.edu.co, Roiver Tapia - T00098425- rotapia@utb.edu.co, Leidy Velásquez - T00077780 - levelasquez@utb.edu.co

**Líder del equipo:** Leidy Laura Velasquez - T00077780 - levelasquez@utb.edu.co - T00077780

**Área temática:** Educación

**Palabras clave:** inteligencia artificial, aprendizaje, estudiantes universitarios, pensamiento crítico, educación digital

#### Resumen

Este proyecto analiza el impacto de la inteligencia artificial en los hábitos de aprendizaje y el desempeño académico de estudiantes de la Universidad Tecnológica de Bolívar. Mediante una encuesta aplicada a estudiantes de diferentes programas y semestres, se identificó cómo utilizan herramientas de IA como ChatGPT en actividades académicas, qué beneficios perciben y cuáles son los principales riesgos asociados a su uso. El proyecto también explora dimensiones como el pensamiento crítico, la dependencia tecnológica, la postura docente y la percepción sobre el uso responsable de estas herramientas en el entorno universitario. Los resultados se comunican mediante visualizaciones y proporciones normalizadas a 100 estudiantes, presentadas en formato de reel tipo Instagram.

### 3.2. Si la UTB tuviera 100 estudiantes ¿Cuántos comerían bien?

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00082823 -Maria Restrepo Licona - liconam@utb.edu.co; T00082522 - Mauricio Fernandez - maufernandez@utb.edu.co; T00083123 - Jerry Buelvas Mejia- jebuelvas@utb.edu.co; T00082556 - Joshua Reyes Leones - joshreyes@utb.edu.co; T00082818 - Luis Perez Diaz - lpdiaz@utb.edu.co

**Líder del equipo:** T00082823 -Maria Restrepo Licona - liconam@utb.edu.co

**Área temática:** Salud

**Palabras clave:** hábitos alimenticios, estudiantes universitarios, omisión de comidas, alimentación saludable, encuesta

#### Resumen

Este proyecto analiza los hábitos alimenticios de estudiantes de la Universidad Tecnológica de Bolívar con el objetivo de responder: si la UTB tuviera 100 estudiantes, ¿cuántos comerían bien? Mediante una encuesta de 18 variables se recopilieron datos sobre frecuencia de comidas, omisión de tiempos alimenticios, consumo de frutas, verduras, comida rápida y bebidas azucaradas, así como el impacto del estrés académico en la alimentación. Los resultados revelan que el 72 % realiza tres comidas principales al día y el 68 % almuerza siempre; sin embargo, el desayuno es

la comida más omitida, con un 52 %. La falta de tiempo y la ausencia de hambre son las causas principales. Solo el 24 % consume frutas o verduras diariamente, mientras que el 51 % ingiere comida rápida entre una y tres veces por semana.

### 3.3. Vida académica y bienestar estudiantil

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** Yosmith samahara díaz-T00083642, Kevin Mendoza Martínez T00100515, Isaac Pérez Ramírez T00097295, Valentina Julio Pitalua T00081186, Luis Antonio Tamara Velásquez T00098776

**Líder del equipo:** Yosmith Díaz- T00083642

**Área temática:** Educación

**Palabras clave:** estrés académico, redes sociales, sueño, ocio, bienestar estudiantil

#### Resumen

Este proyecto analiza la relación entre la vida académica y el bienestar de los estudiantes de la UTB. A partir de una encuesta, se exploraron variables como el tiempo libre disponible, el uso de redes sociales, la calidad del descanso, los hábitos de sueño y los niveles de estrés académico. Los resultados evidencian que, aunque la mayoría de los estudiantes cuenta con espacios de ocio, muchos no logran desconectarse realmente de las obligaciones académicas durante ese tiempo. El proyecto busca visibilizar que el descanso efectivo es un factor determinante en el rendimiento académico y propone reflexionar sobre la importancia del equilibrio entre estudio y bienestar personal.

### 3.4. Una foto de la vida universitaria en la UTB

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00082178 - Valeria Berrio Payares - vpayares@utb.edu.co; T00077164 - Alejandro Villarreal Imitola - imitolaa@utb.edu.co; T00070606 - Fabián Camilo Quintero Pareja - parejaf@utb.edu.co; T00067791 - Isabella Sofia Arrieta Guardo - arrietai@utb.edu.co; T00082473 - Santiago José Cuesta Maza - cuestas@utb.edu.co; T00098251 - Katlin Gómez Hernández - Katgomez@utb.edu.co; T00070603 - Santiago Quintero Pareja - squintero@utb.edu.co

**Líder del equipo:** T00067791 - Isabella Sofia Arrieta Guardo - arrietai@utb.edu.co

**Área temática:** Educación

**Palabras clave:** utb, vida universitaria, estadísticas, campus, estudiantes

#### Resumen

¿Cómo llega un estudiante de la UTB a clases? ¿Qué come al mediodía? ¿Cómo ocupa su tiempo libre? Este proyecto responde esas preguntas con datos reales. A través de una encuesta aplicada a 100 estudiantes, se analizaron variables como género, transporte, alimentación, participación en grupos culturales y deportivos, satisfacción con las instalaciones y pertenencia a programas de bienestar. Los resultados ofrecen un retrato estadístico del día a día universitario, útil tanto para comprender la experiencia estudiantil como para identificar áreas de mejora en el campus.

### 3.5. Si la UTB tuviera 100 estudiantes: así usarían la IA y manejarían el estrés

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00095621 - Victor Manuel Julio Jimenez - vijulio@utb.edu.co ;T00100276 - Wilson José Muñoz Gutiérrez - Munozw@utb.edu.co ; T00097950 - Hillary Feria Liñan -

feriah@utb.edu.co ;T00100507 - Leonel Almanza Meza - Lealmanza@utb.edu.co;T00097192 - Santiago Ariza Cadena - cadenas@utb.edu.co

**Líder del equipo:** T00097192 - Santiago Ariza Cadena - cadenas@utb.edu.co

**Área temática:** Educación

**Palabras clave:** estudiantes universitarios, estrés académico, tiempo libre, inteligencia artificial, tareas

#### **Resumen**

Este proyecto presenta los resultados de una encuesta sobre la cotidianidad de los estudiantes universitarios, abordando dimensiones como el estrés, la carga de tareas, el uso de herramientas de inteligencia artificial, la situación económica y el tiempo libre disponible. Los hallazgos se comunican en formato de video tipo reel, representando los datos de forma proporcional sobre una población de 100 estudiantes para facilitar su comprensión e identificación con la realidad universitaria.

### **3.6. Si la UTB tuviera 100 estudiantes.**

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00099079 - Esteban Tolosa Salcedo - etolosa@utb.edu.co; T00097539 - Danzel Archibold Pianeta - darchibold@utb.edu.co; T00083195 - Abraham Márquez Fuentes - abmarquez@utb.edu.co; T00082548 - Danna Zuluaga Hernández- dazuluaga@utb.edu.co

**Líder del equipo:** T00097539 - Danzel Archibold Pianeta - darchibold@utb.edu.co

**Área temática:** Educación

**Palabras clave:** experiencia universitaria, perspectivas estudiantiles, bienestar estudiantil, vulnerabilidad, encuesta

#### **Resumen**

Este trabajo presenta, mediante la metáfora de 100 estudiantes, una radiografía de la UTB en 2026. La encuesta de 24 preguntas evidencia paridad casi perfecta de género, una población muy joven —94 % tiene 23 años o menos— y un perfil académico fuertemente técnico, con 87 % en áreas de ingeniería, arquitectura, diseño y transformación digital. Sin embargo, detrás de esa fortaleza aparecen alertas importantes: 34 % reporta agotamiento casi diario, 68 % presenta malestar emocional clínico, 82 % duerme menos de lo recomendado y 57 % no cuenta con mentor. Aun así, 64 % declara estar satisfecho, lo que revela una paradoja entre valor percibido y bienestar real. El estudio concluye con recomendaciones sobre aspectos institucionales a fortalecer.

### **3.7. Si la UTB tuviese 100 personas**

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00082523- Elias Daniel Ramos Herrera- elramos@utb.edu.co ;T00083411- Victoria Montes Burgos- Montesv@utb.edu.co

**Líder del equipo:** T00069449 - David Sierra Porta - dporta@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** películas, streaming, cine, entretenimiento, preferencias audiovisuales

#### **Resumen**

Este proyecto presenta los resultados de una encuesta aplicada a 100 estudiantes universitarios para analizar sus hábitos y preferencias en consumo audiovisual. La investigación aborda la frecuencia con que los estudiantes ven películas, los géneros cinematográficos preferidos, las plataformas de streaming más utilizadas y la vigencia de la experiencia en sala de cine. Los datos recopilados permiten identificar tendencias de consumo y comportamientos comunes, y se

presentan mediante gráficos y análisis estadísticos descriptivos.

### 3.8. ¿Qué hacen los estudiantes de la UTB cuando no estudian?

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00097122 - Carolina Esalas Moreno - caesalas@utb.edu.co; T00097708 - Juan José Herrera - ortizjua@utb.edu.co; T00098446 - Gabriel David Torres Montiel - gmontiel@utb.edu.co; T00097320 - Juan Sebastián Merlano Merlano - juamerlano@utb.edu.co; T00098105 - José Angel Gomez Agamez - gomezjo@utb.edu.co; T00099111 - Diego Alejandro Salamanca Martinez - disalamanca@utb.edu.co; T00097406 - Orlando Contreras Leguia - leguiao@utb.edu.co; T00097215 - Diego Alejandro Tous Aponte - daponte@utb.edu.co

**Líder del equipo:** T00097122 - Carolina Esalas Moreno - caesalas@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** hobbies, tiempo libre, actividades extracurriculares, entretenimiento, vida universitaria

#### Resumen

Este proyecto encuestó a 100 estudiantes de la UTB con el fin de conocer sus actividades durante el tiempo libre. A partir de los datos recolectados, se identificaron los pasatiempos más comunes, incluyendo el entretenimiento digital, la música, el deporte y otras actividades recreativas. El objetivo es comprender mejor cómo los estudiantes invierten su tiempo fuera de las aulas y qué papel juegan estas actividades en su bienestar y desarrollo personal.

### 3.9. Si la UTB Tuviera 100 Estudiantes: Análisis Exploratorio de Datos Académicos y Percepción Estudiantil

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00082325 - Dawin Andres Jimenez Pombo -dpombo@utb.edu.co; T00979492 - Miranda Mardini Sarruf - mmardini@utb.edu.co, T00077913 - Santiago Cardona Rendon - cardonas@utb.edu.co

**Líder del equipo:** T00082654 - Emmanuel Ascendra Perez -eascendra@utb.edu.co

**Área temática:** Educación

**Palabras clave:** Utb, Percepcion Estudiantil, Patrones, Normalizacion

#### Resumen

En este proyecto se realizó un análisis exploratorio de datos a partir de encuestas y estadísticas relacionadas con estudiantes de la UTB. Utilizando herramientas básicas de ciencia de datos y estadística descriptiva, se organizaron, interpretaron y visualizaron diferentes variables académicas y sociales para identificar patrones y comportamientos dentro de la población estudiantil. Además, los resultados fueron transformados en ejemplos cotidianos y fáciles de comprender, utilizando comparaciones como “si la UTB tuviera 100 estudiantes”, con el fin de comunicar la información de manera más clara, dinámica y cercana al público.

### 3.10. Si la UTB tuviera 100 estudiantes como sería la participación en la universidad

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00081696 - Sebastián Peniche Gonzalez - peniches@utb.edu.co

**Líder del equipo:** T00081696 - Sebastián Peniche Gonzalez - peniches@utb.edu.co

**Área temática:** Educación

**Palabras clave:** participación estudiantil, ecosistema universitario, actividades extracurriculares, semilleros, grupos culturales

#### Resumen

Este proyecto analiza, bajo la metáfora de 100 estudiantes, el nivel de participación de la comunidad universitaria de la UTB en actividades extracurriculares. Se busca determinar cuántos estudiantes se vinculan a grupos estudiantiles, semilleros de investigación, selecciones deportivas o colectivos culturales, en contraste con quienes mantienen una experiencia estrictamente académica. Los resultados permiten reflexionar sobre la diversidad del ecosistema universitario y el papel que tiene la participación activa en la formación integral del estudiantado.

### 3.11. ¿Qué hacen los 100 estudiantes de la UTB cuando no estudian?

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00097122 - Carolina Esalas Moreno - caesalas@utb.edu.co; T00097708 -

Juan José Herrera - ortizjua@utb.edu.co; T00098446 - Gabriel David Torres Montiel - gmontiel@utb.edu.co; T00097320 - Juan Sebastián Merlano Merlano - juamerlano@utb.edu.co;

T00098105 - José Angel Gomez Agamez - gomezjo@utb.edu.co; T00099111 - Diego Alejandro

Salamanca Martinez - disalamanca@utb.edu.co; T00097406 - Orlando Contreras Leguia - le-

guiiao@utb.edu.co; T00097215 - Diego Alejandro Tous Aponte - daponte@utb.edu.co

**Líder del equipo:** T00097122-Carolina Esalas Moreno-caesalas@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** hobbies, tiempo libre, vida universitaria, encuesta, actividades recreativas

#### Resumen

Este proyecto indaga sobre las actividades que realizan los estudiantes de la UTB en su tiempo libre, representando los resultados sobre una muestra normalizada de 100 personas. Mediante una encuesta, se identificaron los pasatiempos más frecuentes, como el entretenimiento, la música, el deporte y otras actividades recreativas, con el fin de comprender mejor cómo los estudiantes se desconectan de las obligaciones académicas y qué dimensiones del bienestar se ven reflejadas en esas elecciones.

### 3.12. Si la UTB tuviera 100 estudiantes

**Curso:** Introducción a la Ciencia de Datos

**Autores (equipo):** T00098497 - jubenitez@utb.edu.co - Julian Andres Benitez Peñaloz;

T00082617- antvasquez@utb.edu.co - Antonio David Vasquez Lora; T00082458 - German Dario

Bonilla Rocha - gbonilla@utb.edu.co; T00089521 - Jose Angel Machacon Blanquicett - jomacha-

con@utb.edu.co

**Líder del equipo:** T00098497 - jubenitez@utb.edu.co - Julian Andres Benitez Peñaloz

**Área temática:** Educación

**Palabras clave:** tecnología educativa, hábitos de estudio, población estudiantil, bienestar universitario, análisis de datos

## Resumen

Este proyecto analiza la relación entre la población estudiantil de la Universidad Tecnológica de Bolívar y su vínculo con la tecnología en hábitos de aprendizaje, representando los resultados sobre una muestra proporcional de 100 estudiantes. Se examinaron variables como hábitos de estudio, uso de herramientas tecnológicas, salud mental y manejo del tiempo libre. Los hallazgos se presentan en proporciones claras que facilitan la comprensión de los patrones predominantes en la comunidad universitaria.



Figura 3.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de Introducción a la Ciencia de Datos.

## 4 Proyectos del curso de Programación para Ciencia de Datos

### 4.1. Análisis de tasa de accidentalidad en Barranquilla

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00095703 - Egduer Ovalles Rojas - eovalles@utb.edu.co - T00095653 - Johan Ardila Gómez - joardila@utb.edu.co - T00094826 - Diego Rátiva Pérez - drativa@utb.edu.co - T00097295 - Isaac Pérez Ramírez - isperez@utb.edu.co - T00096422 - Aura Guerrero Erazo - auraguerrero@utb.edu.co

**Líder del equipo:** T00095653 - Johan Ardila Gómez - joardila@utb.edu.co

**Área temática:** Ciudades/Movilidad

**Palabras clave:** accidentalidad vial, barranquilla, análisis exploratorio, movilidad, seguridad vial

#### Resumen

Este proyecto analiza el dataset de tasa de accidentalidad en la ciudad de Barranquilla con el objetivo de identificar patrones geospaciales y demográficos asociados a los accidentes de tránsito. El análisis permite determinar las ubicaciones con mayor concentración de incidentes, los grupos etarios más afectados y otros factores relevantes. A partir de estos resultados, se proponen medidas orientadas a mejorar la seguridad vial y la movilidad urbana en la ciudad.

### 4.2. Análisis de Datos Educativos: Saber Pro 2023 y Factores Socioeconómicos

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00082648-Deiner Gonzalez Paredes-paredesd@utb.edu.co;T00082297-Luis Eduardo Mendoza-luiangulo@utb.edu.co;T00082473-Santiago José Cuesta Maza-cuestas@utb.edu.co;T00082178-Valeria Berrío Payares-vpayares@utb.edu.co; T00082548-Danna Valentina Zuluaga Hernández-dazuluaga@utb.edu.co

**Líder del equipo:** T00082648-Deiner Gonzalez Paredes-paredesd@utb.edu.co

**Área temática:** Educación

**Palabras clave:** rendimiento académico, contexto socioeconómico, icfes, saber pro, datos abiertos

#### Resumen

Este proyecto analiza el desempeño de los estudiantes que presentaron las pruebas Saber Pro en Colombia en 2023, con énfasis en la posible relación entre el contexto socioeconómico y los resultados obtenidos. Se empleó Python como lenguaje principal, apoyado en la librería Pandas para el tratamiento y limpieza de los datos, y Matplotlib para la elaboración de visualizaciones estadísticas. Los datos provienen del portal oficial de datos abiertos del ICFES, lo que garantiza la trazabilidad y verificabilidad del análisis.

### 4.3. SummaryData: Implementación de una función de análisis descriptivo avanzado aplicada a datos de COVID-19

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00082229 - Sebastian Garcia Devoz - devozs@utb.edu.co; T00097539 - Danzel Washington Archibold Pianeta ; T00082654 - Emmanuel Ascendra Perez - eascendra@utb.edu.co; T00082325 - Dawin Andres Jimenez Pombo - dombo@utb.edu.co; T00083799 - Isabel Sofia Paez Matallana - paezi@utb.edu.co

**Líder del equipo:** T00082654 - Emmanuel Ascendra Perez - eascendra@utb.edu.co

**Área temática:** Salud

**Palabras clave:** Pandas, análisis estadístico descriptivo, Python, COVID-19, limpieza de datos

#### Resumen

Este proyecto desarrolla `summary_data`, una función en Python con Pandas que extiende el método `describe()` con métricas adicionales como media, mediana, moda, desviación estándar, varianza, cuantiles y valores faltantes, todas calculadas manualmente sin depender de funciones estadísticas nativas. Se aplica sobre el dataset oficial de casos positivos de COVID-19 en Colombia del INS, con millones de registros entre 2020 y 2025. Incluye además curtosis y coeficiente de asimetría, verificadas contra los métodos nativos de Pandas. Los resultados se presentan en un dashboard interactivo con Quarto que incluye gráficos y tablas de frecuencia sobre variables clave del dataset como departamento, sexo, tipo de contagio y estado de recuperación.

### 4.4. Tasas de Interés Activas en el Sistema Financiero Colombiano

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00082182 - Maria Camila Chaustre - mchaustre@utb.edu.co; T00082628 - Yugeimi Andrea Granados - ygranados@utb.edu.co

**Líder del equipo:** T00082628 - Yugeimi Andrea Granados - ygranados@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** crédito, tasa de interés, sistema financiero colombiano, superintendencia financiera, análisis descriptivo

#### Resumen

Las tasas de interés activas son uno de los indicadores más relevantes del sistema financiero colombiano, pues determinan el costo del crédito para personas naturales y empresas de distintos tamaños. Comprender su comportamiento según el tipo de crédito, el perfil del solicitante y otras variables asociadas es clave tanto para la toma de decisiones financieras como para el diseño de políticas públicas más equitativas. Este proyecto analiza dicho comportamiento a partir de datos oficiales de la Superintendencia Financiera de Colombia, disponibles en el Portal de Datos Abiertos, aplicando técnicas de análisis descriptivo y correlación.

### 4.5. DATEnidos

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00082342 - Carlos David Manrique Fals - cfals@utb.edu.co; T00082674 - Nerlis Nikol Otero Perez - oteron@utb.edu.co

**Líder del equipo:** T00082674 - Nerlis Nikol Otero Perez - oteron@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** datos abiertos, colombianos en el exterior, detención, análisis exploratorio, dashboard

#### **Resumen**

DATENidos es un proyecto de análisis de datos sobre colombianos privados de libertad en el exterior, basado en registros oficiales de la Cancillería de Colombia entre 2018 y 2025. A través de un dashboard interactivo, se exploran patrones geográficos, demográficos y temporales para responder tres preguntas clave: ¿en qué países y por qué delitos se concentran los casos?, ¿existen diferencias según género y edad?, y ¿cómo ha evolucionado la cifra a lo largo del tiempo?

## **4.6. Evaluaciones agropecuarias municipales**

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00078528 - Alan Oswaldo Tapias Melgarejo - amelgarejo@utb.edu.co; T00077913 - Santiago Andres Cardona - cardonas@utb.edu.co; T00061359 - Juan David Colón - colonj@utb.edu.co; T00098776 - Luis Tamara - Lutamara@utb.edu.co

**Líder del equipo:** T00077913 - Santiago Andres Cardona - cardonas@utb.edu.co

**Área temática:** Otro

**Palabras clave:** producción agrícola, rendimientos, cultivos, municipios colombianos, evaluaciones agropecuarias

#### **Resumen**

Este proyecto analiza el dataset de Evaluaciones Agropecuarias Municipales (EVA), que registra datos de producción agrícola desagregados por municipio para el período 2006–2018. Cada registro incluye variables como área sembrada y cosechada (ha), producción (t), rendimiento (t/ha), tipo de cultivo, estado físico y ciclo productivo. El conjunto de datos cubre 32 departamentos y más de 1.000 municipios de Colombia, lo que permite identificar patrones territoriales de productividad agrícola y analizar la diversidad de cultivos a nivel regional.

## **4.7. Conflicto Armado, Paz y Víctimas en Colombia Análisis Exploratorio de Datos – 2002–2023**

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00100515 - Kevin Mendoza Martínez - Kemendoza@utb.edu.co; T00076352 - Juan Diego Serpa Medina - juserpa@utb.edu.co; T00081186 - Valentina Julio Pitalua - Vpitalua@utb.edu.co; T00071291 - Jorge Morales Blanco - jorgmorales@utb.edu.co; T00071286 - Javier Mercado - jbolivar@utb.edu.co

**Líder del equipo:** T00076352 - Juan Diego Serpa Medina - juserpa@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** Conflicto armado, análisis de datos, construcción de paz, víctimas, Colombia

#### **Resumen**

Este proyecto analiza indicadores del conflicto armado, paz y víctimas en Colombia entre 2002 y 2023 utilizando datos abiertos de entidades oficiales como el DNP, UARIV, SIJU y UNODC. A través de técnicas de análisis exploratorio de datos, se estudiaron variables relacionadas con minas antipersonal, cultivos de coca, víctimas por ocurrencia, reincorporación e Índice ICA, con el fin de identificar patrones territoriales y temporales asociados a la construcción de paz. Se realizó limpieza, normalización y visualización de más de 86 mil registros provenientes de 1153 municipios del país. Los resultados muestran desigualdades regionales persistentes y permiten reconocer territorios con mayores rezagos o avances en indicadores de paz. El proyecto busca aportar herramientas visuales y analíticas para apoyar la toma de decisiones públicas y académicas sobre

el posconflicto en Colombia.

## 4.8. Popularidad en el anime

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00097122 - Carolina Esalas Moreno - caesalas@utb.edu.co; T00099111 - Diego Alejandro Salamanca Martinez - disalamanca@utb.edu.co

**Líder del equipo:** T00097122 - Carolina Esalas Moreno - caesalas@utb.edu.co

**Área temática:** Cultura/Sociedad

**Palabras clave:** anime, tendencias históricas, géneros, popularidad, análisis de datos

### Resumen

Este proyecto analiza un dataset sobre el ecosistema del anime con el objetivo de identificar tendencias históricas, géneros dominantes y estudios de producción más destacados. El análisis abarca la evolución del anime desde 1980, los títulos más populares y favoritos según distintas métricas, así como los patrones de consumo a lo largo del tiempo. Los resultados se presentan mediante visualizaciones que permiten explorar la diversidad y el crecimiento de este formato audiovisual a nivel global.

## 4.9. ¿Qué variables macroeconómicas explican la actividad económica de Colombia?

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00078350-Luz Angel Martinez Ahumada-Ahumadal@utb.edu.co; T00081936-Ornella Gómez Meusburger- meusburger@utb.edu.co

**Líder del equipo:** T00078350-Luz Angel Martinez Ahumada-Ahumadal@utb.edu.co

**Área temática:** Otro

**Palabras clave:** series de tiempo macroeconómicas, regresión ols, ise colombia, análisis económico, choques estructurales

### Resumen

Este proyecto analiza la relación entre cuatro variables macroeconómicas fundamentales —tasa de cambio (TRM), inflación, tasa de política monetaria y desempleo— y el Índice de Seguimiento de la Economía (ISE) de Colombia. A partir de una base de datos de frecuencia mensual que abarca más de dos décadas, se aplica un proceso riguroso de limpieza mediante interpolación y propagación temporal, preservando los valores atípicos que corresponden a eventos históricos verificables como la crisis financiera de 2008 y la pandemia de COVID-19. El análisis emplea regresión OLS para estimar la capacidad explicativa de cada variable sobre la dinámica económica del país.

## 4.10. Impacto de las comorbilidades en la gravedad de casos de COVID-19 en México

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00098251 - Katlin celen Gomez Hernandez - katgomez@utb.edu.co  
T00099067 - Suying Mo Wu - mos@utb.edu.co

**Líder del equipo:** T00099067 - Suying Mo Wu - mos@utb.edu.co

**Área temática:** Salud

**Palabras clave:** covid-19, comorbilidades, análisis exploratorio de datos, salud pública, factores

de riesgo

### Resumen

Este proyecto analiza cómo factores como obesidad, diabetes, hipertensión y edad influyen en la gravedad y mortalidad de pacientes con COVID-19 en México. Se empleó un dataset de más de un millón de registros con 21 variables clínicas y demográficas, aplicando técnicas de análisis exploratorio (EDA), cálculo de correlaciones y visualización con Python, pandas, seaborn y matplotlib. Los resultados indican que la edad —especialmente en mayores de 50 años— presenta la correlación más alta con el ingreso a UCI, seguida por diabetes e hipertensión con correlación moderada, mientras que la obesidad muestra una correlación baja. El estudio aporta evidencia cuantitativa sobre factores de riesgo clínico con aplicación en salud pública.

## 4.11. RIESGO LABORAL, MUNICIPIO A MUNICIPIO

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00094997 - Simón Bedoya Acevedo - sibedoya@utb.edu.co; T00089521 - Jose Angel Machacon - jomachacon@utb.edu.co; T00068237 - Isaac Navarro Jiménez - isnavarr@utb.edu.co; T00058442 - Daniel Villalba Hoyos - dhoyos@utb.edu.co

**Líder del equipo:** T00058442 - Daniel Villalba Hoyos - dhoyos@utb.edu.co

**Área temática:** Otro

**Palabras clave:** riesgo laboral, análisis exploratorio, siniestralidad, colombia, ciencia de datos

### Resumen

Este proyecto analiza el comportamiento de la afiliación laboral y la siniestralidad en Colombia a partir de registros oficiales de Positiva Compañía de Seguros con corte a noviembre de 2025. Evaluando una muestra de 60.684 registros en más de mil municipios, el estudio caracteriza la distribución territorial y sectorial del riesgo laboral en el país. Los resultados principales demuestran que el tamaño de la fuerza laboral no predice directamente el nivel de riesgo: Bogotá concentra el 31 % de los trabajadores afiliados pero registra una tasa de accidentalidad por debajo de la media nacional, mientras que departamentos como Antioquia y Meta presentan niveles de siniestralidad que superan ampliamente ese promedio.

## 4.12. Análisis y Limpieza de Datos del Dataset Lending Club

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00098446- Gabriel David Torres Montiel- gmontiel@utb.edu.co; Jose Angel Gomez Agamez - gomezjo@utb.edu.co - T00098105

**Líder del equipo:** T00098446- Gabriel David Torres Montiel- gmontiel@utb.edu.co.

**Área temática:** Industria/Negocios

**Palabras clave:** limpieza de datos, valores nulos, lending club, préstamos p2p, imputación

### Resumen

Este proyecto realiza un proceso completo de análisis y limpieza de datos sobre el dataset público de Lending Club, obtenido de Kaggle, que contiene 2,26 millones de registros de préstamos entre personas (P2P) emitidos entre 2007 y 2018 en Estados Unidos. Utilizando Python y la librería Pandas, se identificaron y eliminaron 44 columnas con más del 50 % de valores nulos, reduciendo el dataset de 151 a 107 columnas. Las columnas restantes con valores faltantes fueron imputadas empleando medianas para variables numéricas y moda para variables de fecha, preservando la distribución original. Se verificó asimismo la ausencia total de registros duplicados en el conjunto completo de datos.

### 4.13. Estrategias de Optimización Comercial para Starbucks: Un Enfoque Basado en el Análisis de Datos

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00090449 - Néstor Gómez López - negomez@utb.edu.co; T00088573 - Alejandra Guerrero Gómez - aleguerrero@utb.edu.co

**Líder del equipo:** T00088573 - Alejandra Guerrero Gómez - aleguerrero@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** Análisis de Venta, Limpieza de Datos, Dashboards, Programación en Python

#### Resumen

Este proyecto se enfoca en la creación de una herramienta analítica interactiva diseñada para diagnosticar y mejorar el rendimiento comercial y operativo de las tiendas Starbucks. El proyecto centraliza la información de facturación de la compañía en un tablero de control dinámico que le permite a la gerencia evaluar el comportamiento del negocio en tiempo real. La solución está diseñada para responder a tres necesidades críticas: identificar oportunidades para elevar el gasto promedio por cliente, medir el impacto real de las estrategias de fidelización por sucursal y detectar cuellos de botella en los tiempos de servicio en barra. Así, el proyecto transforma datos crudos en una guía estratégica para aumentar los ingresos y optimizar la atención al cliente.

### 4.14. Crocs Analytcs

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00098425-Roiver Alejandro tapia García - rotapia@utb.edu.co ,T00098842-Mathias Franceschi - mfranceschi@utb.edu.co

**Líder del equipo:** Humberto Manuel Marbello

**Área temática:** Industria/Negocios

**Palabras clave:** análisis de datos, limpieza de dataset, precios, ratings, satisfacción del cliente

#### Resumen

Este proyecto presenta un dashboard de análisis comercial construido sobre un dataset depurado de 426 registros de productos Crocs, obtenido a partir de un conjunto original de 1.458 registros. El análisis abarca el comportamiento de precios, los ratings de satisfacción del cliente, la participación en ventas por categoría y los ingresos por año. A partir de estos resultados, se formulan recomendaciones de optimización comercial basadas en los patrones identificados en los datos.

### 4.15. Uso y Dependencia de la Inteligencia artificial.

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00082458 - Germán Bonilla Rocha.

**Líder del equipo:** T00082458 - Germán Bonilla Rocha.

**Área temática:** Cultura/Sociedad

**Palabras clave:** dependencia de ia, toma de decisiones, autonomía cognitiva, análisis exploratorio, python

#### Resumen

Este proyecto analiza cómo la inteligencia artificial influye en los procesos de toma de decisiones humanas, utilizando el dataset *AI Decision Dependence & Cognitive Caution 2025*. Mediante herramientas como Pandas y visualizaciones en Python, se estudian patrones de dependencia, confianza y autonomía cognitiva asociados al uso frecuente de sistemas de IA. Los resultados permiten reflexionar sobre los límites entre el apoyo tecnológico y la delegación excesiva de juicio

en algoritmos automatizados.

## 4.16. Dashboard ICFES - Análisis Educativo

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00098342 - Elías Castro Marrugo - eliacastro@utb.edu.co , T00097918 - Harold Padilla - harpadilla@utb.edu.co

**Líder del equipo:** T00098342 - Elías Castro Marrugo - eliacastro@utb.edu.co

**Área temática:** Educación

**Palabras clave:** icfes, visualización de datos, análisis socioeconómico, pruebas saber 11, python

### Resumen

Este proyecto analiza los resultados de las Pruebas Saber 11 de 946 estudiantes colombianos entre 2021 y 2024, aplicando técnicas de limpieza, transformación y visualización de datos con Python. A partir de un dataset de 3.150 registros, se depuraron duplicados, valores nulos y valores atípicos. El análisis examina el rendimiento por área académica, género, jornada, estrato socioeconómico, nivel educativo de los padres, evolución temporal y desempeño departamental. Los resultados muestran que Lectura Crítica es el área más sólida y el Inglés el mayor reto, y que el estrato socioeconómico no determina de forma lineal el desempeño académico. Los hallazgos se presentan en un dashboard interactivo desarrollado en HTML.

## 4.17. E-commerce

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00097320-juan Sebastián Merlano Mendoza - jmerlano@utb.edu.co ; T00097215-Diego Alejandro Tous Aponte - daponte@utb.edu.co

**Líder del equipo:** T00097320-juan Sebastián Merlano Mendoza

**Área temática:** Industria/Negocios

**Palabras clave:** e-commerce, limpieza de datos, análisis exploratorio, transacciones en línea, online retail

### Resumen

Este proyecto desarrolla un proceso exhaustivo de limpieza, tratamiento y análisis exploratorio sobre el dataset Online Retail II, que contiene más de 1.067.371 registros reales de transacciones de una empresa minorista en línea del Reino Unido durante el período 2009-2011. A través de una metodología estructurada en cuatro fases, se identificaron y corrigieron valores nulos, registros duplicados y valores atípicos, obteniendo un dataset limpio y confiable. El análisis exploratorio permitió identificar patrones de comportamiento de ventas, concentración geográfica, estacionalidad en los ingresos y segmentación de clientes, generando conclusiones estratégicas orientadas a la toma de decisiones empresariales.

## 4.18. Análisis de Vehículos Usados y Factores que Influyen en su Precio de Venta.

**Curso:** Programación para Ciencia de Datos

**Autores (equipo):** T00098497 - jubenitez@utb.edu.co - Julian Andres Benitez Peñaloza; T00098224 - pcortecero@utb.edu.co - Pablo Cortecero Heroza

**Líder del equipo:** T00098497 - jubenitez@utb.edu.co - Julian Andres Benitez Peñaloza

**Área temática:** Industria/Negocios

**Palabras clave:** vehículos usados, precio de venta, análisis exploratorio de datos, mercado automotriz, regresión lineal

### Resumen

Este proyecto analiza un dataset de vehículos usados para identificar los factores que más influyen en su precio de venta. Se utilizó el Car Details Dataset de Kaggle, que incluye variables como año del vehículo, kilometraje, tipo de combustible, transmisión, tipo de vendedor y número de propietarios anteriores. El estudio comprende procesos de limpieza y organización de datos, análisis exploratorio mediante gráficos y estadísticas, y la aplicación de técnicas de regresión lineal y agrupamiento. Los resultados permiten comprender el comportamiento del mercado automotriz de segunda mano y determinar las variables con mayor incidencia en el valor de los vehículos.



Figura 4.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de Programación para Ciencia de Datos.

# 5 Proyectos del curso de Modelos de Regresión y Series de Tiempo

## 5.1. Predictur

**Curso:** Modelos de Regresión y Series de Tiempo

**Autores (equipo):** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co; T00068420 - Eduardo Rafael Mardini Serrano - emardini@utb.edu.co; T00064965 - Michael Andres Taboada Naranjo - mtaboada@utb.edu.co; T00086039 - Carlos Miguel Toro Torres - toroc@utb.edu.co

**Líder del equipo:** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** series de tiempo, turismo, storytelling, machine learning, pipeline

### Resumen

Predictur es un sistema analítico integrado (*end-to-end*) diseñado para el pronóstico de la demanda turística en el Caribe colombiano, operando bajo una arquitectura local de archivos planos. Mediante la ingesta automatizada y limpieza por interpolación de datos mensuales de la encuesta EMA del DANE (2020–2026), el sistema unifica métricas de ocupación, oferta e ingresos reales. Tras evaluar las series con técnicas de descomposición estacional, la plataforma entrena y compara cuatro enfoques predictivos: ETS, Prophet, SARIMAX y un modelo *ensemble* ponderado. Los resultados se presentan en una aplicación interactiva desarrollada en Streamlit con gráficos dinámicos en Plotly, transformando métricas analíticas complejas en alertas y recomendaciones de negocio accionables.

## 5.2. HYPATIA: Sistema Híbrido de Predicción de Trayectorias de Objetos Cercanos a la Tierra

**Curso:** Modelos de Regresión y Series de Tiempo

**Autores (equipo):** T00083005 - Daniel Andres Jimenez Povea - andres.ajp31@gmail.com; T00082136 - Josue David Morelo Castro - jomorelo@utb.edu.co

**Líder del equipo:** T00083005 - Daniel Andres Jimenez Povea - andres.ajp31@gmail.com

**Área temática:** Otro

**Palabras clave:** Efecto Yarkovsky, predicción orbital, ecuaciones diferenciales, machine learning, defensa planetaria

### Resumen

HYPATIA es un sistema computacional híbrido que predice trayectorias de asteroides cercanos a la Tierra incorporando el efecto Yarkovsky, una perturbación térmica ignorada en modelos gravitacionales clásicos que acumula desviaciones orbitales significativas en décadas. El sistema integra tres componentes: un integrador numérico de ecuaciones diferenciales ordinarias del problema de  $N$  cuerpos extendido con la fuerza de Yarkovsky (RK45), un módulo de series de tiempo y regresión bayesiana que detecta y estima el parámetro  $da/dt$  desde residuos orbitales históricos, y un modelo de aprendizaje automático (XGBoost cuantílico) que infiere dicho parámetro desde propiedades físicas observables del asteroide. El caso de estudio principal es Apophis, cuyo acercamiento confirmado en abril de 2029 a menos de 32.000 km de la Tierra lo convierte en referencia de validación ideal.



Figura 5.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de Modelos de Regresión y Series de Tiempo.

## 6 Proyectos del curso de Sensado y modelado de Sistemas Físicos

### 6.1. Clasificación de imágenes satelitales

**Curso:** Sensado y modelado de Sistemas Físicos

**Autores (equipo):** T00068765-German De Armas Castaño-gdearmas@utb.edu.co;T00068186-Angel Vega Rodriguez-anvega@utb.edu.co;T00064965 - Michael Andres Taboada Naranjo-mtaboada@utb.edu.co

**Líder del equipo:** T00064965 - Michael Andres Taboada Naranjo - mtaboada@utb.edu.co

**Área temática:** Ambiente

**Palabras clave:** machine learning, sentinel-2, teledetección, clasificación de cobertura terrestre, random forest

#### Resumen

Este proyecto presenta una aplicación web que automatiza y visualiza de forma interactiva el procesamiento de imágenes satelitales para la clasificación de cobertura terrestre. La plataforma se conecta al catálogo STAC de Microsoft Planetary Computer para consultar y descargar imágenes Sentinel-2 L2A, sobre las cuales se aplican algoritmos de aprendizaje automático —en particular Random Forest— para identificar y categorizar diferentes tipos de cobertura del suelo de forma eficiente y reproducible.

### 6.2. Modelado de la fragmentación del bosque andino en el Valle del Cocora mediante imágenes Sentinel-2

**Curso:** Sensado y modelado de Sistemas Físicos

**Autores (equipo):** T00067613 - Isaac David Sánchez Sánchez - sanchezi@utb.edu.co; T00067539 - Santiago Andrés Orejuela Cueter - sorejuela@utb.edu.co; T00067756 - María Valentina Serna González - maserna@utb.edu.co; T00069220 - Luis Mario Díaz Martínez - diazluis@utb.edu.co

**Líder del equipo:** T00067613 - Isaac David Sánchez Sánchez - sanchezi@utb.edu.co

**Área temática:** Ambiente

**Palabras clave:** Sentinel-2, bosque andino, palma de cera, fragmentación, Cocora

#### Resumen

Este proyecto aplica sensado remoto y modelado físico para evaluar la fragmentación del bosque andino en el Valle del Cocora, Salento, Quindío, como indicador del deterioro del hábitat de regeneración de la palma de cera. A partir de imágenes Sentinel-2 L2A se calculan índices espectrales como NDVI, EVI, NDMI, BSI, NBR y NDRE para analizar vigor vegetal, humedad del dosel, suelo expuesto y perturbación de la cobertura. El estudio compara un periodo base con uno reciente, clasifica el paisaje en bosque denso, vegetación intermedia, pasturas y suelo expuesto, y estima métricas de fragmentación. El objetivo es demostrar cómo un paisaje aparentemente verde puede perder estructura, humedad y conectividad ecológica.



Figura 6.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de Sensado y Modelado de Sistemas Físicos.

# 7 Proyectos del curso de Machine Learning

## 7.1. Veles

**Curso:** Machine Learning

**Autores (equipo):** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co; T00083873 - Jasen Mihovil Yukopila Escobar - jyukopila@utb.edu.co; T00058985 - Dariem Antonio García Cardona- dargarcia@utb.edu.co; T00082259 - Katlyn Gutierrez Cardona - katgutierrez@utb.edu.co; T00070606 - Fabian Camilo Quintero Pareja - parejaf@utb.edu.co; T00067791 - Isabella Sofia Arrieta Guardo - arrietai@utb.edu.co; T00070603 - Santiago Quintero Pareja - squintero@utb.edu.co; T00086039 - Carlos Miguel Toro Torres - toroc@utb.edu.co; T00082962 - Shalom Jhoanna Arrieta Marrugo - sharrieta@utb.edu.co

**Líder del equipo:** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co

**Área temática:** Otro

**Palabras clave:** inteligencia artificial, machine learning, clasificación musical, recomendación, industria musical

### Resumen

Veles es una plataforma inteligente que transforma la industria musical mediante inteligencia artificial y aprendizaje automático. El sistema analiza pistas en tiempo real para clasificar géneros musicales con alta precisión, extrayendo patrones técnicos no perceptibles al oído humano. Esta capacidad permite generar recomendaciones personalizadas basadas en el perfil acústico real de cada usuario. Paralelamente, Veles apoya la toma de decisiones estratégicas en empresas del sector musical mediante un tablero de control que visualiza tendencias de consumo y proyecciones de mercado, convirtiendo métricas analíticas complejas en estrategias comerciales rentables.

## 7.2. EmotiSpeech: Detección Automática de Emociones en Voz para la Optimización de la Atención al Cliente en Call Centers

**Curso:** Machine Learning

**Autores (equipo):** T00068175-Valentina Zuñiga Vasquez- vzuniga@utb.edu.co; T00065092-Samuel David Diaz Llamas-llamass@utb.edu.co; T00064027-Samuel Eduardo Sanmartin Safarssanmartin@utb.edu.co; T00064315-Valentina Alfaro Padilla-valfaro@utb.edu.co; T00068420-Eduardo Rafael Mardini Serrano-emardini@utb.edu.co; T00064965- Michael Taboada Naranjo-mtaboada@utb.edu.co; T00089314-Geisler Santiago Garcia de la Rosa-geigarcia@utb.edu.co

**Líder del equipo:** T00064965-Michael Taboada Naranjo-mtaboada@utb.edu.co;T00064965

**Área temática:** Industria/Negocios

**Palabras clave:** Emociones, clasificador, Embeddings de audio, Balanced Accuracy

### Resumen

EmotiSpeech es un clasificador de emociones en voz entrenado con un dataset propio de 146 audios en español colombiano. Extrajimos embeddings con wav2vec2 y comparamos seis algoritmos de Machine Learning bajo validación cruzada leave-one-out, alcanzando una balanced accuracy del 73.8 % con Regresión Logística, superando ampliamente el azar del 33 %. El modelo recibe un audio, procesa sus patrones acústicos profundos y predice si la persona está enojada, Feliz o

triste, con aplicación directa en la detección temprana de clientes en riesgo dentro de un call center.

### 7.3. Clasificador inteligente de billetes colombianos mediante visión por computador

**Curso:** Machine Learning

**Autores (equipo):** T00067613 - Isaac David Sánchez Sánchez - sanchezi@utb.edu.co; T00067539 - Santiago Andrés Orejuela Cueter - sorejuela@utb.edu.co; T00069220 - Luis Mario Díaz Martínez - diazluis@utb.edu.co; T00078330 - Zulianys Liseth Orozco Chavez - orozcoz@utb.edu.co

**Líder del equipo:** T00067613 - Isaac David Sánchez Sánchez - sanchezi@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** visión por computador, aprendizaje automático, billetes colombianos, clasificación de imágenes, denominación

#### Resumen

Este proyecto desarrolla un sistema de reconocimiento de billetes colombianos basado en visión por computador y aprendizaje automático. La aplicación permite capturar o cargar una imagen de un billete y estimar su denominación, incluso en condiciones adversas como baja iluminación, recortes parciales, fondos complejos o uso de cámara web. Para mejorar la precisión, el sistema combina extracción de características visuales, clasificación supervisada y verificación por coincidencia visual, reduciendo así los falsos positivos en ausencia de un billete claro. La solución busca apoyar procesos de identificación rápida de efectivo en contextos educativos, comerciales y de prueba tecnológica.



Figura 7.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de Machine Learning.

## 8 Proyectos del curso de Toma de Decisiones Basadas en Datos

### 8.1. Veles

**Curso:** Toma de Decisiones Basadas en Datos

**Autores (equipo):** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co; T00083873 - Jasen Mihovil Yukopila Escobar - jyukopila@utb.edu.co; T00058985 - Dariem Antonio García Cardona- dargarcia@utb.edu.co; T00082259 - Katlyn Gutierrez Cardona - katgutierrez@utb.edu.co; T00070606 - Shalom Johanna Arrieta Marrugo - sharrieta@utb.edu.co - T00082880 - Karla Andrea Barraza Torres - barrazak@utb.edu.co

**Líder del equipo:** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** inteligencia artificial, machine learning, clasificación musical, recomendación, toma de decisiones

#### Resumen

Veles es una plataforma inteligente que transforma la industria musical mediante inteligencia artificial y aprendizaje automático. El sistema analiza pistas en tiempo real para clasificar géneros musicales con alta precisión, extrayendo patrones técnicos no perceptibles al oído humano. Esta capacidad permite generar recomendaciones personalizadas basadas en el perfil acústico real de cada usuario. Paralelamente, Veles apoya la toma de decisiones estratégicas en empresas del sector musical mediante un tablero de control que visualiza tendencias de consumo y proyecciones de mercado, convirtiendo métricas analíticas complejas en estrategias comerciales rentables.

### 8.2. Data4Mind

**Curso:** Toma de Decisiones Basadas en Datos

**Autores (equipo):** T00082575 - María José Álvarez Peña - alvarezma@utb.edu.co; T00083870 - Alana Valentina Guerrero Quiroz - alguerrero@utb.edu.co; T00077885 - Isabella Márquez Vides - imarquez@utb.edu.co

**Líder del equipo:** T00077885 - Isabella Márquez Vides - imarquez@utb.edu.co

**Área temática:** Salud

**Palabras clave:** ciencia de datos, salud mental, bienestar digital, toma de decisiones, redes sociales

#### Resumen

Data4Mind es un proyecto de ciencia de datos orientado a analizar cómo los hábitos digitales y los estilos de vida influyen en la salud mental de los adolescentes. A partir de variables como tiempo en redes sociales, calidad del sueño, niveles de ansiedad, estrés, actividad física y bienestar digital, el sistema identifica patrones y factores de riesgo mediante técnicas de análisis predictivo y aprendizaje automático. El objetivo es transformar datos en información accionable para apoyar la toma de decisiones preventivas en entornos educativos y familiares, promoviendo el uso responsable de la tecnología y el equilibrio entre la vida digital y el bienestar emocional en los jóvenes.

## 9 Proyectos del curso de StoryTelling y Narrativa de Datos

### 9.1. Predictur

**Curso:** StoryTelling y Narrativa de Datos

**Autores (equipo):** T00078181 - Alejandro Patron Montero - patrona@utb.edu.co; T00068420- Eduardo Rafael Mardini Serrano - emardini@utb.edu.co; T00064965 - Michael Andres Taboada Naranjo - mtaboada@utb.edu.co; T00086039 - Carlos Miguel Toro Torres - toroc@utb.edu.co T00082962 - Shalom Jhoanna Arrieta Marrugo - sharrieta@utb.edu.co; T00082259 - Katlyn Gutierrez Cardona - katgutierrez@utb.edu.co; T00065092 - Samuel David Díaz Llamas - lla-mass@utb.edu.co; T00064027 - Samuel Eduardo Sanmartin Safar - ssanmartin@utb.edu.co; T00082880 - Karla Andrea Barraza Torres - barrazak@utb.edu.co

**Líder del equipo:** T00082259 - Katlyn Gutierrez Cardona - katgutierrez@utb.edu.co

**Área temática:** Industria/Negocios

**Palabras clave:** series de tiempo, turismo, storytelling, machine learning, pipeline

#### Resumen

Predictur es un sistema analítico integrado (*end-to-end*) diseñado para el pronóstico de la demanda turística en el Caribe colombiano, operando bajo una arquitectura local de archivos planos. Mediante la ingesta automatizada y limpieza por interpolación de datos mensuales de la encuesta EMA del DANE (2020–2026), el sistema unifica métricas de ocupación, oferta e ingresos reales. Tras evaluar las series con técnicas de descomposición estacional, la plataforma entrena y compara cuatro enfoques predictivos: ETS, Prophet, SARIMAX y un modelo *ensemble* ponderado. Los resultados se presentan en una aplicación interactiva en Streamlit con gráficos dinámicos en Plotly, transformando métricas analíticas complejas en alertas y recomendaciones de negocio accionables.

### 9.2. El paraíso que se borra

**Curso:** StoryTelling y Narrativa de Datos

**Autores (equipo):** T00083810 - Juan Mateo Anaya Amador - juanaya@utb.edu.co; T00063215 - Daniel Enrique Pupo gallego - dgallego@utb.edu.co; T00083960 - Jesús David Pérez García - perezjesu@utb.edu.co

**Líder del equipo:** T00083005 - Daniel Andres Jimenez Povea - dpovea@utb.edu.co

**Área temática:** Ambiente

**Palabras clave:** deforestación costera, narrativa de datos, manglar, turismo, cartagena

#### Resumen

Este proyecto analiza la pérdida de cobertura forestal y de manglar en cuatro zonas costeras de Cartagena de Indias entre 2010 y 2024: Cartagena urbana, Tierra Bomba, La Boquilla/Manzanillo del Mar y Barú/Playas Blancas. Mediante técnicas de narrativa de datos y visualización interactiva, se documenta cómo el crecimiento acelerado del turismo —que alcanzó aproximadamente 3,5 millones de visitantes en 2024, con un aumento del 22 % en turismo internacional— ha impulsado la destrucción de ecosistemas costeros que protegen la ciudad frente a la erosión. Los resultados se presentan en una plataforma web con mapas geográficos, gráficas temporales interactivas e índices de presión ambiental por zona, construida con datos verificados de fuentes como IDEAM,

CARDIQUE, Aerocivil, INVEMAR y AFP.



Figura 9.1: Todos nuestros chicos de la Feria de Ciencias de Datos en el Data Impact Day de los cursos de StoryTelling y Narrativa de Datos.

## 10 Reconocimientos y premios

Con el propósito de destacar el rigor, la creatividad y el compromiso demostrados por las y los estudiantes durante el *Data Impact Day*, se llevó a cabo un proceso de evaluación de los pósters y presentaciones por parte de docentes de la Escuela de Transformación Digital y de distintas unidades académicas de la Universidad Tecnológica de Bolívar.

La evaluación consideró criterios como la claridad en la formulación del problema, la calidad del análisis de datos, la pertinencia de la metodología empleada, la solidez de las conclusiones y la calidad visual y narrativa del póster. A partir de esta valoración, se otorgaron reconocimientos generales y menciones especiales en categorías específicas.

### Premios a los primeros lugares

Los siguientes equipos fueron reconocidos por obtener las puntuaciones globales más altas en la evaluación integral de sus proyectos. En esta edición, el primer lugar fue declarado **en empate** entre dos proyectos que alcanzaron la máxima puntuación:

**1. Primer lugar (empate) – Veles**

Equipo: T00078181 - Alejandro Patron Montero - patrona@utb.edu.co; T00083873 - Jasen Mihovil Yukopila Escobar - jyukopila@utb.edu.co; T00058985 - Dariem Antonio García Cardona - dargarcia@utb.edu.co; T00082259 - Katlyn Gutierrez Cardona - katgutierrez@utb.edu.co; T00070606 - Fabian Camilo Quintero Pareja - parejaf@utb.edu.co; T00067791 - Isabella Sofia Arrieta Guardo - arrietai@utb.edu.co; T00070603 - Santiago Quintero Pareja - squintero@utb.edu.co; T00086039 - Carlos Miguel Toro Torres - toroc@utb.edu.co; T00082962 - Shalom Jhoanna Arrieta Marrugo - sharrieta@utb.edu.co  
Curso: Machine Learning

**2. Primer lugar (empate) – EmotiSpeech: Detección Automática de Emociones en Voz para la Optimización de la Atención al Cliente en Call Centers**

Equipo: T00068175 - Valentina Zuñiga Vasquez - vzuniga@utb.edu.co; T00065092 - Samuel David Díaz Llamas - llamass@utb.edu.co; T00064027 - Samuel Eduardo Sanmartin Safar - ssanmartin@utb.edu.co; T00064315 - Valentina Alfaro Padilla - valfaro@utb.edu.co; T00068420 - Eduardo Rafael Mardini Serrano - emardini@utb.edu.co; T00064965 - Michael Andres Taboada Naranjo - mtaboada@utb.edu.co; T00089314 - Geisler Santiago García de la Rosa - geigarcia@utb.edu.co  
Curso: Machine Learning

**3. Segundo lugar – SummaryData: Implementación de una función de análisis descriptivo avanzado aplicada a datos de COVID-19**

Equipo: T00082229 - Sebastian Garcia Devoz - devozs@utb.edu.co; T00097539 - Danzel Washington Archibold Pianeta - darchibold@utb.edu.co; T00082654 - Emmanuel Ascendra Perez - eascendra@utb.edu.co; T00082325 - Dawin Andres Jimenez Pombo - dpombo@utb.edu.co; T00083799 - Isabel Sofia Páez Matallana - paezi@utb.edu.co  
Curso: Programación para Ciencia de Datos

**4. Tercer lugar – ¿Qué variables macroeconómicas explican la actividad económica de Colombia?**

Equipo: T00078350 - Luz Ángel Martínez Ahumada - ahumadal@utb.edu.co; T00081936 - Ornella Gómez Meusburger - meusburgero@utb.edu.co  
Curso: Programación para Ciencia de Datos

Estos proyectos se destacaron por integrar de manera ejemplar el ciclo completo de ciencia de datos: desde el planteamiento del problema y la preparación de los datos, hasta el modelado, la interpretación de resultados y la comunicación clara y rigurosa de sus hallazgos.

## Menciones especiales

Además de los premios generales, se otorgaron menciones especiales en categorías que reflejan dimensiones clave del trabajo en ciencia de datos:

### ■ **Mención Impacto**

*HYPATIA: Sistema Híbrido de Predicción de Trayectorias de Objetos Cercanos a la Tierra*

Equipo: T00083005 - Daniel Andrés Jiménez Povea - andres.ajp31@gmail.com; T00082136

- Josué David Morelo Castro - jomorelo@utb.edu.co

Curso: Modelos de Regresión y Series de Tiempo

Reconocimiento al proyecto cuyo problema, enfoque y resultados muestran el mayor potencial de impacto en su dominio de aplicación, destacándose por la ambición científica de integrar modelado orbital clásico, regresión bayesiana y aprendizaje automático en un problema de relevancia para la defensa planetaria.

### ■ **Mención Mejor Análisis de Datos**

*Análisis de tasa de accidentalidad en Barranquilla*

Equipo: T00095703 - Egduer Ovalles Rojas - eovalles@utb.edu.co; T00095653 - Johan Ar-

dila Gómez - joardila@utb.edu.co; T00094826 - Diego Rátiva Pérez - drativa@utb.edu.co;

T00097295 - Isaac Pérez Ramírez - isperrez@utb.edu.co; T00096422 - Aura Guerrero Era-

zo - auraguerrero@utb.edu.co

Curso: Programación para Ciencia de Datos

Reconocimiento al equipo que demostró mayor profundidad analítica, identificando patrones geoespaciales y demográficos con rigor metodológico y traduciendo los resultados en recomendaciones concretas para la seguridad vial urbana.

### ■ **Mención Mejor Narrativa**

*Si la UTB tuviera 100 estudiantes: ¿Cuántos comerían bien?*

Equipo: T00082823 - María Del Mar Restrepo Licon - liconam@utb.edu.co; T00082522 -

Mauricio Fernández - maufernandez@utb.edu.co; T00083123 - Jerry Buelvas Mejia - jebuel-

vas@utb.edu.co; T00082556 - Joshua Reyes Leones - joshreyes@utb.edu.co; T00082818 -

Luis Pérez Díaz - lpdiaz@utb.edu.co

Curso: Introducción a la Ciencia de Datos

Reconocimiento al equipo que presentó la historia de sus datos de forma más clara, cercana y convincente, articulando hallazgos concretos —con cifras precisas y un formato accesible— que conectan directamente con la experiencia cotidiana de la comunidad universitaria.

### ■ **Mención Mejor Visualización Geoespacial**

*Modelado de la fragmentación del bosque andino en el Valle del Cocora mediante imágenes Sentinel-2*

Equipo: T00067613 - Isaac David Sánchez Sánchez - sanchezi@utb.edu.co; T00067539 -

Santiago Andrés Orejuela Cueter - sorejuela@utb.edu.co; T00067756 - María Valentina Serna

González - maserna@utb.edu.co; T00069220 - Luis Mario Díaz Martínez - diazluis@utb.edu.co

Curso: Sensado y modelado de Sistemas Físicos

Reconocimiento al proyecto que integró con mayor rigor técnico el procesamiento de imágenes satelitales multiespectrales, combinando índices como NDVI, EVI, NDMI, BSI, NBR y NDRE en un análisis temporal comparativo que va más allá de la visualización superficial para cuantificar la pérdida de estructura, humedad y conectividad ecológica en un ecosistema andino de alta relevancia ambiental.

- **Mención Conciencia Ambiental**

*El paraíso que se borra*

Equipo: T00083810 - Juan Mateo Anaya Amador - juanaya@utb.edu.co; T00063215 - Daniel Enrique Pupo Gallego - dgallego@utb.edu.co; T00083960 - Jesús David Pérez García - Perezjesu@utb.edu.co

Curso: StoryTelling y Narrativa de Datos

Reconocimiento al proyecto que con mayor eficacia articuló datos verificados y narrativa visual para documentar un problema urgente y local: la destrucción de ecosistemas costeros en Cartagena de Indias impulsada por el crecimiento turístico. Su plataforma web con mapas geográficos, gráficas temporales e índices de presión ambiental por zona demuestra que contar bien una historia con datos puede ser, en sí mismo, un acto de incidencia.

- **Mención Relevancia Social**

*Data4Mind*

Equipo: T00082575 - María José Álvarez Peña - alvarezma@utb.edu.co; T00083870 - Alana Valentina Guerrero Quiroz - alguerrero@utb.edu.co; T00077885 - Isabella Márquez Vides - imarquez@utb.edu.co

Curso: Toma de Decisiones Basadas en Datos

Reconocimiento al proyecto que eligió atacar uno de los problemas más urgentes y menos visibles del entorno juvenil contemporáneo: la relación entre hábitos digitales y salud mental en adolescentes. La pertinencia del problema, la solidez del enfoque analítico y el potencial directo de sus resultados para apoyar decisiones preventivas en entornos educativos y familiares distinguen a este equipo por encima del desafío técnico.

Estos reconocimientos no solo celebran el desempeño sobresaliente de los equipos premiados, sino que buscan resaltar dimensiones fundamentales de la práctica en ciencia de datos: la capacidad de transformar datos en conocimiento con impacto real, la rigurosidad en el análisis y la habilidad de contar buenas historias con evidencia.

---

## Cierre y agradecimientos

La décima edición del *Data Impact Day* cierra con una certeza que se renueva cada semestre: la ciencia de datos tiene mucho más que decir cuando sale del aula y se enfrenta a preguntas reales. Los 41 proyectos reunidos en estas memorias son evidencia de ello. Detrás de cada póster hay decisiones metodológicas, datos difíciles de limpiar, hipótesis que no siempre se confirmaron y narrativas que hubo que construir desde cero. Ese proceso —invisible en el producto final— es precisamente el núcleo de la formación en ciencia de datos.

Desde el Programa de Ciencia de Datos de la Universidad Tecnológica de Bolívar, agradecemos a todas las personas que hicieron posible esta edición: al cuerpo docente que acompañó a los equipos en cada etapa del proyecto, al estudiantado que respondió al reto con dedicación y apertura crítica, y a las directivas y dependencias de apoyo de la UTB cuyo respaldo institucional sostiene este tipo de iniciativas.

Un reconocimiento muy especial merece el trabajo de **Skarling Amira Aguilar Ramos** y **Gloria Mercedes Osorio Viana**, quienes se encargaron de la coordinación operativa y logística del evento con la misma meticulosidad que exigimos en el análisis de datos. Sin su gestión, el *Data Impact Day* no habría sido posible.

Agradecemos también a los invitados externos, aliados institucionales y asistentes que participaron activamente: sus preguntas, observaciones y perspectivas enriquecieron el diálogo y recordaron a los equipos que sus resultados interesan más allá del campus.

Estas memorias quedan como registro público de lo que esta cohorte fue capaz de construir. Que sirvan de referencia, de inspiración y de punto de comparación para quienes vengan después. El *Data Impact Day* seguirá siendo el espacio donde la ciencia de datos de la UTB demuestra que aprender a trabajar con datos es, sobre todo, aprender a hacerse mejores preguntas sobre el mundo que nos rodea.