

Evaluación del módulo

Consigna del proyecto 

Proyecto: Análisis exploratorio de datos para decisiones comerciales

Evaluación del módulo ANÁLISIS EXPLORATORIO DE DATOS

Situación inicial

Unidad solicitante: Dirección de análisis estratégico de la empresa **ComercioYA**. La empresa ComercioYA, dedicada a ventas en línea, ha recolectado un conjunto de datos históricos de comportamiento de sus clientes (compras, visitas, montos, devoluciones, reseñas). Actualmente se enfrenta a la necesidad de comprender estos datos para tomar decisiones basadas en evidencia. Sin embargo, el equipo directivo no tiene claridad sobre cómo interpretar los datos ni qué variables son realmente relevantes para mejorar la estrategia de ventas y marketing.

Tu equipo ha sido convocado para explorar y visualizar los datos, detectar patrones, relaciones y valores atípicos, y entregar una propuesta clara y fundamentada que ayude a mejorar la toma de decisiones.

Nuestro objetivo

El objetivo del proyecto es aplicar técnicas de análisis exploratorio de datos (EDA) sobre un conjunto de datos reales del negocio, con el fin de obtener hallazgos relevantes que permitan comprender el comportamiento de los clientes y sustentar decisiones estratégicas.

Al finalizar el proyecto, se espera entregar un informe técnico con visualizaciones, análisis estadístico y recomendaciones clave, demostrando dominio del EDA y herramientas como Seaborn y Matplotlib en Python.

Requerimientos

- Analizar tipos de variables (categóricas, numéricas) y su distribución
- Identificar medidas de tendencia central y dispersión
- Detectar puntos atípicos y relaciones entre variables

- Calcular e interpretar correlaciones
- Ajustar un modelo de regresión lineal
- Visualizar los datos utilizando Seaborn y Matplotlib
- Documentar hallazgos relevantes en un informe técnico visualmente efectivo

Paso a paso

Este proyecto refiere exclusivamente al **módulo 4: Análisis Exploratorio de Datos**, y se compone de **6 etapas (lecciones)**, las cuales podrás avanzar de forma progresiva y escalonada con la ayuda de los manuales teóricos y los contenidos desarrollados en las clases en vivo.

Ten en cuenta de invertir **tiempo asincrónicos** para el desarrollo de cada etapa a modo de poder finalizar el módulo y realizar la entrega formal de tu propuesta. Cualquier consulta que surja compártela en los espacios sincrónicos para resolver las dudas en equipo.

A continuación encontrarás las consignas y tareas a desarrollar:

- **Lección 1: ANÁLISIS EXPLORATORIO DE DATOS**
 -  **Objetivo:** Comprender el propósito del EDA y diferenciarlo del análisis inicial.
 -  **Tareas a desarrollar:**
 - Generar un dataset para poder analizarlo
 - Distinguir variables cuantitativas y categóricas
 - Detectar valores faltantes o inconsistencias (IDA)
 - Documentar primeros hallazgos en Visual Studio Code
- **Lección 2: CONCEPTOS BÁSICOS DE ESTADÍSTICA DESCRIPTIVA**

-  **Objetivo:** Aplicar conceptos de tendencia central, dispersión y medidas de posición.
-  **Tareas a desarrollar:**
 - Calcular media, mediana, moda, varianza, desviación estándar
 - Determinar cuartiles y percentiles
 - Generar histogramas y boxplots
 - Identificar valores atípicos y documentar su impacto
- **Lección 3: CORRELACIÓN**
 -  **Objetivo:** Detectar y cuantificar relaciones entre variables numéricas
 -  **Tareas a desarrollar:**
 - Construir scatterplots y matrices de correlación
 - Calcular el coeficiente de Pearson (R)
 - Detectar correlaciones espurias y justificar
 - Documentar visual y estadísticamente las asociaciones encontradas
- **Lección 4: REGRESIONES LINEALES**
 -  **Objetivo:** Implementar un modelo de regresión simple y múltiple para explicar variables dependientes.
 -  **Tareas a desarrollar:**
 - Aplicar modelos con statsmodels
 - Calcular R², MSE, MAE
 - Evaluar significancia de predictores

- Interpretar coeficientes e inferencias básicas
- Visualizar la regresión con Seaborn
- Lección 5: ANÁLISIS VISUAL DE DATOS
 -  **Objetivo:** Representar relaciones complejas y distribuciones mediante Seaborn.
 -  **Tareas a desarrollar:**
 - Crear gráficos de tipo pairplot, violinplot, jointplot y heatmap
 - Construir visualizaciones que combinen múltiples variables
 - Ajustar parámetros de estilo, colores, tamaños
 - Utilizar FacetGrid para segmentar por categorías
 - Documentar insights visuales
- Lección 6: LIBRERÍA MATPLOTLIB
 -  **Objetivo:** Crear visualizaciones personalizadas y exportables.
 -  **Tareas a desarrollar:**
 - Crear figuras y subplots
 - Personalizar títulos, etiquetas, leyendas, ticks
 - Usar anotaciones, límites y formatos de archivo
 - Generar una presentación con gráficos finales
 - Entregar informe EDA con visualizaciones embebidas

¿Qué vamos a validar? 

- Identificación y tratamiento correcto de tipos de variables

- Aplicación de medidas estadísticas adecuadas
- Correcta interpretación de correlaciones y regresiones
- Calidad visual y técnica de los gráficos entregados
- Uso correcto de Seaborn y Matplotlib
- Documentación clara, estructurada y con foco analítico
- Coherencia del informe técnico: hallazgos fundamentados
- Profesionalismo y estética de la presentación final

Referencias

- <https://seaborn.pydata.org/>
- <https://matplotlib.org/stable/users/index.html>
- <https://www.statsmodels.org/stable/index.html>
- <https://pandas.pydata.org/docs/>

Recursos

- <https://www.kaggle.com/learn/data-visualization>
- Dataset públicos similares (Kaggle o UCI ML Repository)

Entregables

- Cuaderno Visual Studio Code con todo el desarrollo paso a paso
- Gráficos exportados en PNG/PDF
- Informe técnico de análisis (PDF o presentación)
- Código fuente comentado y limpio

- Documento final con insights y recomendaciones
- Subida a GitHub + link entregado en Moodle

Portafolio

Este proyecto es ideal para incluir en tu portafolio profesional de ciencia o análisis de datos.

Recomendamos:

- Subir el notebook con visualizaciones interactivas a plataformas como GitHub
- Redactar un README claro explicando el contexto, los pasos y las conclusiones
- Incluir capturas de pantalla de los gráficos más representativos
- Enlazar el informe como PDF junto con el código
- Destacar tu capacidad para analizar datos reales y comunicar visualmente hallazgos relevantes

¡Éxitos!

Nos vemos más adelante

