

Guía 4

Ejercicios Jupyter Notebook

Observaciones: Todos los ejercicios debe realizarlos con jupyter notebook, además debe utilizar la función “markdown” para crear notebooks que se puedan interpretar e interpretar los datos de manera mas fácil, siempre pensando las entregas como “informes formales”

Ejercicio 1: Análisis exploratorio del dataset de venta de autos

- Cargar y limpiar datos con pandas.
- Realizar un análisis descriptivo (estadísticas, valores nulos).
- Graficar la relación entre precio y kilometraje con scatter plots.
- Interpretar las visualizaciones y sacar conclusiones preliminares.

Ejercicio 2: Regresión lineal simple para predecir precio de autos

- Seleccionar una variable predictora (e.g. kilometraje).
- Entrenar un modelo de regresión lineal simple con scikit-learn.
- Visualizar la línea de regresión sobre el scatter plot.
- Evaluar el modelo con métricas como error cuadrático medio.

Ejercicio 3: Regresión lineal múltiple para predecir precio usando varias variables

- Usar variables como año, kilometraje, y tipo de transmisión.
- Entrenar un modelo de regresión múltiple con scikit-learn.
- Comparar resultados con la regresión simple, interpretando coeficientes.
- Visualizar los resultados y analizar residuos.

Ejercicio 4: Análisis del dataset veterinaria con pandas y visualización

- Explorar y limpiar el dataset veterinaria.
- Realizar agrupaciones y cálculos estadísticos para características por especie.
- Crear gráficos comparativos (boxplots, barras) para variables como edad, peso, y enfermedades más comunes.
- Realizar predicción de una variable numérica con regresión lineal, si aplica.

Ejercicio 5: Dataset de actividad volcánica

Usa el dataset de SERNAGEOMIN.

- Agrupa por año y cuenta el número de eventos por nivel de peligro.
- Grafica las tendencias temporales (por colores).
- Analiza si existen años más activos o niveles predominantes.
- Usar regresión lineal para analizar relación entre variables geológicas y magnitud de erupciones.
- Crear gráficos avanzados que resuman el análisis, y redactar un informe en Markdown documentando el proyecto completo.

Ejercicio 6: Recolectar y analizar datos desde una API pública.

Instrucciones:

- Extrae los primeros 150 Pokémon desde <https://pokeapi.co/api/v2/pokemon>.
- Guarda nombre, altura, peso, tipo.
- Calcula:
 - Promedio de peso y altura por tipo.
 - Pokémon más pesado y más liviano.
- Exporta como pokemones .csv.

Ejercicio 7: Proyecto de Web Scraping y Análisis con "Books to Scrape"

En este ejercicio, utilizarás la página "Books to Scrape" para hacer web scraping, procesar datos y aplicar análisis usando las herramientas vistas en clase.

Objetivos

- Practicar web scraping con requests y BeautifulSoup.
- Manipular y limpiar datos con pandas.
- Hacer análisis estadístico y visualizaciones con matplotlib.
- Aplicar regresión lineal simple o múltiple usando scikit-learn.
- Documentar el trabajo en un Jupyter Notebook con Markdown explicativo.

Parte 1: Web Scraping

1. Accede a **Books to Scrape**.
2. Extrae los datos de todos los libros listados en la página principal y, si te animas, recorre las páginas siguientes para obtener más datos (paginación).
3. Para cada libro, captura al menos estas variables:
 - Título
 - Precio
 - Calificación en estrellas (puedes extraerla de la clase CSS)
 - Disponibilidad (por ejemplo, "In stock")
 - Categoría (puedes obtenerla al entrar en la página del libro o navegando el menú lateral)

Parte 2: Limpieza y Explorar Datos

1. Guarda los datos obtenidos en un DataFrame de pandas.
2. Limpia los datos:
 - Convierte precios a números (float).
 - Extrae la calificación en estrellas como número (por ejemplo, "Three" a 3).
 - Trata valores nulos o erróneos.
3. Realiza un análisis descriptivo:

- Estadísticas básicas de precios y calificaciones.
- Distribución del número de libros por categoría.

Parte 3: Visualización

1. Genera gráficos:

- Histograma o boxplot de precios.
- Gráfico de barras con cantidad de libros por categoría.
- Scatter plot precio vs calificación.

Parte 4: Modelado con Regresión (opcional, si hay sentido)

1. Define una variable objetivo para predecir (por ejemplo, precio).
2. Usa variables predictoras numéricas, como la calificación en estrellas, o codifica categorías si quieres.
3. Entrena un modelo de regresión lineal simple o múltiple para predecir el precio.
4. Evalúa el modelo usando error cuadrático medio y muestra resultados.

Parte 5: Documentación

1. Explica cada fase con Markdown:

- Descripción del problema
- Estrategia de scraping
- Procesamiento de datos
- Resultados y análisis
- Conclusiones