

Portafolio de Ciencia de Datos

Nombre del estudiante: Juan Pérez

Curso: Fundamentos de Ciencia de Datos

Fecha: 06/09/2025

Resumen Ejecutivo

Este portafolio tiene como objetivo demostrar las competencias adquiridas en el curso Fundamentos de Ciencia de Datos. A través de tres proyectos seleccionados, se muestra la capacidad para analizar información, aplicar modelos estadísticos y de machine learning, y comunicar hallazgos de manera clara y efectiva.

Las habilidades presentadas incluyen análisis exploratorio de datos, construcción de modelos predictivos, uso de herramientas de visualización, y dominio de lenguajes de programación como Python y SQL.

El portafolio refleja no solo los conocimientos técnicos adquiridos, sino también la capacidad de interpretar datos en contextos reales, proponer soluciones basadas en evidencia y comunicar resultados de forma visual y comprensible. Esto resulta esencial en un entorno donde las decisiones dependen cada vez más del análisis de datos.

Proyectos

Proyecto 1 – Análisis de ventas de una tienda online

Descripción del problema: Identificar patrones de compra en clientes de una tienda virtual para mejorar la segmentación de marketing.

Metodología: Python (pandas, matplotlib), análisis exploratorio, limpieza de datos, segmentación de clientes.

Resultados: Se encontró que el 60% de las compras provenían de clientes entre 25 y 35 años, y que los descuentos incrementaban un 20% las ventas en ese grupo.

Visualización: Gráfico de barras comparando ventas por rango de edad. Diagrama de dispersión entre monto gastado y frecuencia de compra.

Reflexión personal: Aprendí la importancia de limpiar datos antes del análisis y cómo las visualizaciones ayudan a comunicar resultados a públicos no técnicos.

Proyecto 2 – Modelo predictivo de churn en telecomunicaciones

Descripción del problema: Predecir la probabilidad de que un cliente abandone el servicio (churn).

Metodología: Python (scikit-learn), regresión logística, métricas de precisión y recall, división entrenamiento/prueba.

Resultados: El modelo alcanzó una precisión del 82% y permitió identificar que los clientes con contratos mensuales y facturas altas eran los más propensos a abandonar.

Visualización: Matriz de confusión del modelo. Gráfico de importancia de variables.

Reflexión personal: Entendí cómo elegir métricas adecuadas (precision vs recall) impacta en la interpretación del modelo y en decisiones empresariales.

Proyecto 3 – Visualización de datos de COVID-19

Descripción del problema: Analizar la evolución de contagios en América Latina y generar visualizaciones comprensibles para la ciudadanía.

Metodología: Tableau y Python (pandas), uso de datasets abiertos de la OMS, visualización de series temporales y mapas.

Resultados: Se observó que los picos de contagio coincidían con fechas de movilidad elevada (festivos). El tablero interactivo permitió explorar diferencias entre países.

Visualización: Mapa de calor por país. Línea de tiempo de casos diarios.

Reflexión personal: Aprendí que la visualización no solo es técnica, sino también comunicativa: la claridad y simplicidad son clave para llegar a audiencias diversas.

Habilidades Técnicas

- Lenguajes: Python, R, SQL
- Bibliotecas: pandas, NumPy, Matplotlib, scikit-learn
- Herramientas de visualización: Tableau, Power BI
- Otros: GitHub, Google Colab, Excel avanzado

Enlaces Adicionales

GitHub: <https://github.com/juanperez-data>

Tableau Public: <https://public.tableau.com/juanperez>

Blog personal: <https://juanperez-data.blogspot.com>