

Tarea 1 – Búsqueda de Dataset

Elegí 3 posibles opciones de datos para poder desarrollar mi proyecto final, todos se encuentran, abajo se enlistan de forma prioritaria:

- Health Insurance Marketplace:
 - <https://www.kaggle.com/datasets/hhs/health-insurance-marketplace?select=BenefitsCostSharing.csv>
 - El CSV principal tiene 32 columnas y casi 6m de registros
 - Toda la base de datos pesa cerca de 4 gb
 - Se basa en los seguros médicos en USA, donde se mencionan los beneficios, la edad de los beneficiarios, sus estados, etc.
 - Fue una elección porque me llama mucho la atención la Ciencia de Datos aplicada a la salud. Fue de los pocos que no era para Computer Vision.
- Zillow House Price Data:
 - <https://www.kaggle.com/datasets/paultimothymooney/zillow-house-price-data>
 - Es información de precios de casas en todo USA.
 - Zillow publica información del mercado de real estate (el cual reemplazaría a los datos de Kaggle)
 - <https://www.zillow.com/research/data/>
- Credit Card Fraud Detection:
 - <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
 - Es un dataset con 31 columnas y casi 300k registros.
 - Lo malo de este dataset es que sus columnas no tienen nombre definido y no hay un diccionario de datos.