 ¿Qué orígenes de datos podemos utilizar para obtener dataset o enriquecer la información de nuestro dataset?  
- Se pueden importar cualquier tipo de datos, los más usados por las empresas son Excel debido a la costumbre, pero también se pueden importar datos desde cualquier tipo de base como sql o bien desde apis, que hoy en día como contienen la información más detallada y exacta son más fiables (ejemplo league of legends tiene su propia api con los datos de los jugadores o una api publica pero demasiado buena para hacer ejercicios es la de pokemon que contiene información de todos los pokemon, en que juegos aparecen y habilidades entre otras propiedades https://pokeapi.co/ )  
Los ejemplos anteriores son datos ya cargados o almacenados de alguna forma, pero, también se puede crear nuestro propio dataset haciendo web scraping o extrayendo la información de redes sociales, esto lo hacen las empresas para saber que tan popular o no están siendo, aunque también se usó en la presidencia de estados unidos para ve el comportamiento de los usuarios en las elecciones  
Dejo ejemplo aquí de cómo es extraer los tweets y clasificarlos con paralleldots (análisis de sentimientos) https://colab.research.google.com/drive/1JpKuoQGvFpz8YgRH6xHLyBKbgO2g92QM?usp=sharing  
  
• ¿Qué librería de phyton nos ayuda a realizar un buen análisis exploratorio de los datos?  
Las librerías que nos ayudan a hacer un buen análisis exploratorio son:  
- Numpy: es la librería que contiene funciones matemáticas, de hecho, su nombre es la abreviatura de numeric Python, la cual contiene desde la raíz cuadrada hasta la creación de arrays (matrices). Su funcion en el análisis es crear elementos vacíos para ser llenados más adelante.  
- Pandas: es la librería con la cual se hacen tablas de información, estas son llamadas datasets y son similares a las hojas de Excel. Su principal función en el análisis es poder visualizar de una mejor manera las bases de datos o archivos (Excel) ver si hay valores nulos, duplicados o erróneos, estos últimos pueden ser datos iguales pero que por una coma, letra o espacio el sistema los marque diferentes.

• ¿Qué librería de phyton nos ayuda a realizar un buen análisis exploratorio de los datos a nivel visual?  
- Matplotlib: es una de las librerías de visualización básicas para ver dentro de una columna los datos representados en barras, boxplot, histogramas, entre otros.  
- Plotly: Esta librería es más dinámica, es decir, el usuario puede seleccionar que datos ver y cuales no, cuando se pasa el mouse por el grafico de igual manera aparecen los datos que se seleccionan o en el que está el apuntador. Se usa principalmente para hacer dashboards en Python junto con la Liberia dash  
- Seaborn: Se podría decir que es el intermedio entre las dos anteriores, los gráficos tienen un mejor aspecto que matplotlib pero no son interactivos como con plotly  
  
• De las métricas existentes para evaluar modelos de regresión, ¿cuál consideras que es más sencilla de interpretar?  
- El error cuadrático medio considero que es el más fácil de interpretar debido a que nos dice la exactitud o aproximación que tiene el modelo a comparación de los datos reales  
  
• ¿Qué tipos de escalado de la librería de scikit-learn consideras más adecuados para aplicar en tus modelos de regresión?  
- Considero que el tipo de escalamiento dependerá de nuestros datos, así como del procesamiento que les demos a estos, siento que el mejor sería MaxAbsScaler porque va de 1 a -1 y podría darnos una mejor precisión, si es positivo, negativo o neutral pero esto dependerá de los datos y sus características