

Laboratorio #3: Regresiones

Para el dataset adjunto al laboratorio deberá realizar los siguientes pasos para completar el análisis de regresión donde el Target será la variable **Price (in USD)**:

- 1) Cargar el dataset.
- 2) Determinar y listar las variables numéricas continuas, numéricas categóricas y categóricas.
- 3) Realizar un análisis descriptivo para cada variable del dataset, que incluya lo siguiente:
 - a. Para las variables numéricas:
 - i. Gráfica de densidad.
 - ii. Análisis de correlación entre todas las variables numéricas.
 - iii. Scatter plot entre cada variable numérica vrs el target.
 - iv. Dar una conclusión escrita sobre cada una de las variables numéricas y su relación con la variable target.
 - b. Para las variables categóricas y discretas:
 - i. Mostrar una gráfica de mosaico con la cantidad por cada valor o categoría.
 - ii. Gráfica de Box-plot y violín entre cada variable y la variable target.
 - iii. Dar una conclusión escrita sobre cada una de las variables categórica o discreta y su relación con la variable target.
- 4) Separar el dataset en dos parte, con una 85% para entrenamiento y un 15% para prueba.
- 5) Entrenar al menos 5 modelos distintos para regresión haciendo uso de validación cruzada sobre el dataset de entrenamiento.
- 6) Mostrar una tabla ordenada del promedio de los valores de RMSE calculados en cada corrida de la validación cruzada, la tabla debe incluir el nombre del modelo y el valor del RMSE y debe estar ordenada de menor a mayor RMSE.
- 7) Re-entrenar los dos primeros modelos (top 2 ganadores) con todo el set de entrenamiento es decir con el 85% apartado al inicio.
- 8) Calcular el RMSE del modelo entrenado con el 85% utilizando los datos apartados para prueba, es decir el 15% apartado inicialmente.

- 9) Realizar una gráfica de histograma de las predicciones y los valores reales.
 - a. Mostrar una conclusión escrita sobre el resultado mostrado.
- 10) Mostrar una gráfica de scatter plot entre las predicciones y los valores reales.
 - a. Mostrar una conclusión escrita sobre el resultado mostrado.
- 11) Mostrar los resultados finales sobre el RMSE y dar una conclusión sobre los dos modelos ganadores.
- 12) Dar una conclusión final sobre su análisis y la factibilidad del modelo para usarlo para realizar predicciones.