**UNIVERSIDAD CATÓLICA BOLIVIANA”SAN PABLO”  
MAESTRÍA EN CIENCIA DE DATOS V2  
PRIMERA VERSIÓN  
MINERÍA DE DATOS 1   
DATASET DE PRACTICA**

**ROSSMAN**

1. Quitar filas duplicadas
2. Quitar columnas duplicadas
3. Identificar columnas que tienen valores constantes
4. Identifique el porcentaje de valores faltantes en las columnas
5. Analice el caso de los atributos que no serían apropiaos para en dataset, por medio de gráficas Cuales decidiría usted de transformar o en su caso quitar, con el propósito de mejorar el dataset. Genere las siguientes gráficas:
   1. Histograms
   2. Quartiles
   3. Scatter plots
   4. Distributions
   5. Heartmaps
6. Obtenga la matriz de correlación de las variables numéricas e identifique si algunas se pudieran quitar.
7. Mapear Day of Week con los días en literal.
8. Obtenga la tabla de relación de frecuencias entre los atributos StoreType y Assortment
9. Mapear el atributo CompetitionOpenSinceMonth con los meses del año en literal y rellenar los valores faltantes “?” con el mes que tenga la mayor frecuencia.
10. Identifique atributos que no serían de utilidad por su nivel de variación, si hubiera. Usted identifique su nivel de utilidad, dando razones para ello.
11. Identifique los campos que tienen valores faltantes y decida como reemplazarlos, haciendo un análisis en cada caso.
12. Obtenga por día de semana (DayOfWeek) el total en ventas (Sales) y ordénelo por ventas en forma descendente.
13. Obtenga por día de semana (DayOfWeek) y tienda el total y el promedio en ventas (Sales) y ordénelo por ventas en forma ascendente.

**seaborn Scatter plot**

**x = sb.PairGrid(iris)**

**x = x.map(plt.scatter)**

**x = sb.PairGrid(iris, hue = 'species', palette='coolwarm')**

**x = x.map\_diag(plt.hist)**

**x = x.map\_offdiag(plt.scatter)**

**x = x.add\_legend()**