**CONSIDERACIONES EN EL TRATAMIENTO DE LA DATA**

**Preprocesamiento de data:**

* Desde Excel se eliminaron algunos datos que incumplían con la estandarización y lógica de la base de datos, registros que su tipo de formato se había registrado mal. Por ejemplo, algunos registros contaban con descripciones en la columna de valor total siendo esta una columna para expresar cifras numéricas. Así mismo, se eliminaron atributos poco relevantes para centralizar la atención en los que si lo eran.
* Se reviso que se hubiese cargado correctamente todos los atributos de los registros junto con sus formatos correspondientes
* Se eliminaron registros a los que les hacía falta algún atributo ya que estos eran muy mínimos dentro de todo el conjunto de datos.
* Se cambio el nombre de las columnas para simplificar la llamada en el código posterior. Así mismo, se crearon columnas nuevas para año y fecha con el mismo propósito.

**Graficas:**

Se obtuvieron dos clases de graficas:

* Graficas de Acuerdo a los Valores Totales más Representativos: Estas graficas se obtienen filtrando inicialmente el atributo vs los mayores costos asociados a cada uno de los valores del atributo, es decir, para los clientes se seleccionaron los clientes con mayores valores totales del DataSet.

Estos datos se interpretan como las compras con mayor representación económica.

* Graficas de acuerdo a las mayores frecuencias de Compras: Estas graficas se obtienen obteniendo las frecuencias de cada atributo según los datos registrados.

Estos datos se pueden considerar como compras que a pesar de no ser representativas económicamente por si mismas, son muy frecuentes y en conjunto pueden representar grandes compras a lo largo del tiempo.