EDA[Exploratory Data Analysis] for E-retail factors for customer activation and retention.

Step1:

1. Collecting the dataset from the given e-comerce.xlsx file.

2.read the xlsx file using the command **pd.read\_excel**

3.Display the first 5 rows of the dataset using **head ()** command and understand the features name.

4.check how many rows and columns using **shape** command.

There are 269 rows and 71 columns/features.

5. checking for null values in a dataset using **isnull()** function

There is no null values in the dataset.

6. For more Analysis,getting some more information about dataset using **info()** function.

It displays complete details about dataset

There are 70 object type and 1 int type of data.

Memory usage is 149kb.

It gives non null values.

7. One of the important step is to find some statistical measure in the dataset using **describe()** function.

It displayed some statistical measures like

Count,mean,25%,50%,75%,min,max,etc.

From this measures we found what is the mean,median,missing values,max values etc.

8.Since dataset contains bit large features,we need to know how many columns are there.so we have accessed all the features by applying for loops and print the columns.

9.Since all the dataset contains both numerical and categorical data, here verified how many numerical features and categorical fetaures are there in dataset using **select.dtype** function .

There are only one numerical feature and,

Remaining 70 columns are categorical .

10.Checking for duplicate datas using **duplicate()** function

There are totally 166 duplicate records.

11. getting the index values for the same duplicate records using **index.values()**

12.Try to remove duplicate records using **drop.duplicate()** function