Customer Churn EDA - Project:
Venkataramanan T

The assignment or project was to do an Exploratory Data analysis on the Customer Churn datasets provided.

The following are the files provided :

- Customer Demographics.csv
- Customer Investment Snapshot.csv
- Customer Portfolio Snapshot.csv
- Customer Attrition Status.csv

The first task was to merge all these files together to create a single csv file that can be used further for the analysis.

I merged all the csv files into one using the pandas merge function and using the inner join.

The next part was to remove columns that are not required for the analysis : the column RowNumber was not required as the index was already available and self generated by Pandas, so the column RowNumber was removed.

Then the following operations were performed:

1. Removal of Duplicates
2. Check for Missing values : Balance had about 3617 and Credit Score had 3 missing values, these missing values were then imputed using the median values for these columns.
3. The correlation matrix was then printed.
4. The total data points across Geography was show as a graph
5. There is a strong correlation between “Exited’ and “Age’, with the customers exiting having a median age of 45 whereas customers not exiting having a median age of 35.
6. The exited and not exited values were the printed across regions, it can be determined from the graphs that for the region ‘East’ out of a total of around 2500 values, there are about 750 exits and the number who have not exited is about 1750, this region has the highest number of exits with respect to the total number in that region , some interventions will be required to understand the exits and act upon them further.
7. It is also further seen that the number of female exits is more than the male when doing a comparison between Gender and Exits as can be seen in the graph.