# Customer Churn Analysis

We have 4 excel sheets with data regarding customers. All the 4 datasets have redundant column ‘row number’ which can be safely dropped. After that we can investigate any null values still present in each column of all datasets. Dropped those rows which have all values as null. Now join all datasets using merge method from pandas.

After joining all 4 datasets into 1, we can explore for how many unique values are there in each column, drop any duplicate rows and check for any null values. We can see that there are null values in 2 columns. I have replaced them with the median value of the respective columns. We can use mean or mode as well instead of median as per requirement. Median is the safest option to consider.

Now our dataset looks clean in proper structure, without null values and duplicates. We can perform EDA on this data using different statistical methods like describe() to gain insight into numeric columns. Correlation helps in measuring the relation between target variable (Exited) and other independent variables.

We can investigate the distribution of data in different columns. We can also explore the relationship between different columns using visualizations like barplot, histogram, heatmaps etc.,

We can use groupby to gain insights in data based on categorical variables. We can use crosstab, pivotable to check for the distribution of data against 2 or more categorical variables. Melt is useful in reverting the changes of pivotable as needed.