# Customer Churn Analysis

**Context**

Portfolio Management Unit (PMU), a business unit of an MNC, comprises a user-based portfolio of which a dedicated banker offers comprehensive financial planning and investment management advice to individuals.

**Prompt**

A key priority for the Portfolio Management Unit is building and maintaining long-term relationships with clients. Management has asked you to develop analytically based measures to generate a report on the possible factor/s on which clients have discontinued the services of PUM. The goal of the exercise is to do exploratory analysis and provide your insights along with strategies to build to stem attrition/customer churn.

**Procedure**

* We have 4 different datasets in different csv files so, initially import the dataset in different data frames and perform to joins (Merge steps) to get the final dataset in a single data frame.
* All the 4 datasets have redundant column ‘row number’ which can be safely dropped. Below table shows the final joined datasets in a data frame.

**Dataset**

The final Dataset will consist of 14 variables:

|  |  |  |
| --- | --- | --- |
| **Column** | **Data Type** | **Description** |
| RowNumber | Integer | Approx. 10K customers |
| CustomerID | Integer | Unique Identifier for client |
| Surname | String | Client Surname |
| CreditScore | Integer | Ranging from 350 to 850 |
| Geography | String | PMU Sales region |
| Gender | String |  |
| Age | Integer |  |
| Tenure | Integer | Length of client relationship in years |
| Balance | Decimal | Investment balance snapshot |
| Number of Products | Integer | Number of products with PMU |
| HasChckng | Integer | 1 = Has a checking account 0 = No checking account |
| IsActiveMember | Integer | 1 = Digitally Active 0 = Digitally Inactive |
| EstimatedSalary | Decimal | Salary Exited Integer 1 = Exited 0 = Not Exited |

* After that we can investigate any missing values present in each column of data frame. Need to remove the missing values presented in the CustomerId (because represents the Key element) rows which have all values as null.
* After joining all 4 datasets into 1, Explore number of unique values presented in each column, then drop the duplicate rows. After removing the unique data samples in a data frame check for any missing values presented in a data frame.
* If any of the missing values presented, then as per the requirement do cleaning either remove those missing values or apply the missing values approach using any of the statistical methods which appear logical based on your analysis. Such as impute them using mean, median, mode or use interpolation/neighbour imputations.
* Here we found that the missing cells presented in 2 columns. For understanding purpose here, used different imputation techniques such as mean, median, and interpolation because the missing cells presented in numeric/continuous data samples. If missing samples presented in labeled/categorical column, then median/mode is suitable for imputation.
* We can use mean or mode as well instead of median as per requirement. Median is the safest option to consider.
* If needed rearrange the columns based on the requirement if samples had independent and dependent samples the rearrange them otherwise keep it like.
* After the data transformation we can see the clean data in proper structure without any missing/unique data samples. Later perform Exploratory Data Analysis (EDA) for the cleaned 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.
* Investigate the distribution of data in different columns. Also explore the relationship between different columns using visualizations like bar plot, histogram, heatmaps etc.,
* Use group by to gain insights in data based on categorical variables. 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.