Project:

Customer Churn Analysis

An Exploratory Data Analysis using Python

Context

Portfolio Management Unit (PMU), a business unit of an MNC has comprised 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.

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

Hints:

1. You have 4 different datasets to which think of performing joins (Merge steps) to get the final dataset as per fig 1.
2. Perform correct joins
3. Check for duplicates
4. Perform the initial data transformation steps to clean your data like removing the unnecessary column
5. Use the pandas functions like duplicated(), describe(), unique(), isnull().sum(), skew(), corr() to get the details out of your dataframe.
6. List out missing values if any and first visualize its distribution across columns (you may use distplot)
7. Apply the missing values approach using any of the statistical methods which appear logical based on your analysis.
8. You may reuse this code to impute missing values

   missing_column = ['column_name']
   for i in missing_column:
   df.loc[df.loc[:,i].isnull(),i]=df.loc[:,i].**mean()**
   - You can replace mean with median or any other statistical approach
   - You may use Interpolation or neighbor imputations.

9. Plot different graphs using visualization libraries. Use the right graphs for categorical columns and non-categorical columns to extract stories out of it. You may use bar plot, count plot, histograms, pie charts, etc.

   Count plot: sns.countplot(x="column name",data=df, hue=" legends")
   Bar plot: sns.barplot(x=" column on x", y=" column on y",hue="legends ")
   Box plot: sns.boxplot(data=df, x="Column name")

   Please feel free to use any of the charts of your choice to present different numbers

10. Get the numbers out from your analysis like how many users are from different zones, how many of them have left the PUM services, and analyze the trend.

11. For the strategy buildup, try to get the correlation between the Exited column to the other columns.

    - use function corr() from pandas and you can use the below syntax to generate a correlation chart:

      ```
      plt.figure(figsize = (5, 5))
      corr = df.corr()
      sns.heatmap(corr, cmap="YlGnBu", annot = True,linewidths = 0.5,)
      plt.show()
      ```

12. Pandas, Numpy, Seaborn, and Matplotlib cheat sheets can be used to access the different functions easily.

Evaluation will be done on how you

1. Perform Transformations (Data Wrangling)
2. Handle Missing information
3. Present different key metrics using Visualization charts

Provide your analysis in a Jupyter Notebook along with a 1-2-page word document writeup.

We look forward to seeing your analysis!

Fig 1:

Customer Investment Snapshot

RowNumber	CustomerId	Tenure	CreditScore	Balance	EstimatedSalary
1	15634602	2	619		101348.88
2	15647311	1	608	83807.86	112542.58
3	15619304	8	502	159660.8	113931.57
4	15701354	1	699		93826.63

Customer Portfolio Snapshot

RowNumber	CustomerId	NumOfProducts	HasChckng	IsActiveMember
1288	15565701	1	0	0
4199	15565706	1	1	1

Customer Demographics

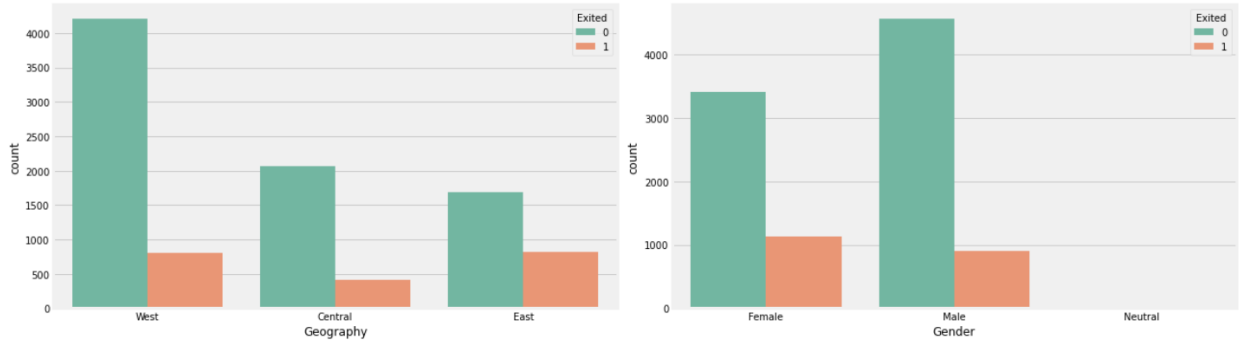
RowNumber	CustomerId	Surname	Geography	Gender	Age
1	15634602	Hargrave	West	Female	42
2	15647311	Hill	Central	Female	41
3	15619304	Onio	West	Female	42

Customer Attrition Status

RowNumber	CustomerId	Exited
1	15634602	1

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasChckng	IsActiveMember	EstimatedSalary	Exited
1	15634602	Hargrave	619	West	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Central	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	West	Female	42	8	159660.8	3	1	0	113931.57	1

Fig 2:


Exited
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
West
Central
East
Geography
Exited
0
1
count
4000
3000
2000
1000
0
Female
Male
Neutral
Gender