

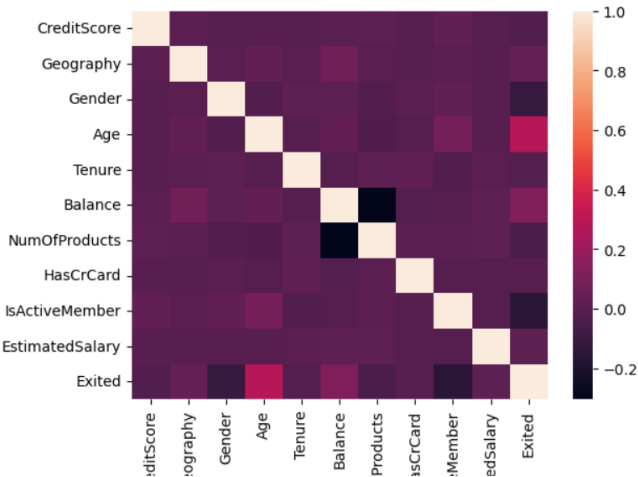
Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	739699
Project Title	Telecom Customer Churn Prediction
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																																																																																																																																																																																																																										
Data Overview	<div>Basic statistics, dimensions, and structure of the data.</div> <div><pre>[22]: data.describe()</pre><table><thead><tr><th></th><th>CreditScore</th><th>Age</th><th>Tenure</th><th>Balance</th><th>NumOfProducts</th><th>HasCrCard</th><th>IsActiveMember</th><th>EstimatedSalary</th><th>Exited</th></tr></thead><tbody><tr><td>count</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td><td>10000.000000</td></tr><tr><td>mean</td><td>650.528800</td><td>38.921800</td><td>5.012800</td><td>76485.889288</td><td>1.530200</td><td>0.70550</td><td>0.515100</td><td>100090.239881</td><td>0.203700</td></tr><tr><td>std</td><td>96.653299</td><td>10.487806</td><td>2.892174</td><td>62397.405202</td><td>0.581654</td><td>0.45584</td><td>0.499797</td><td>57510.492818</td><td>0.402769</td></tr><tr><td>min</td><td>350.000000</td><td>18.000000</td><td>0.000000</td><td>0.000000</td><td>1.000000</td><td>0.000000</td><td>0.000000</td><td>11.580000</td><td>0.000000</td></tr><tr><td>25%</td><td>584.000000</td><td>32.000000</td><td>3.000000</td><td>0.000000</td><td>1.000000</td><td>0.000000</td><td>0.000000</td><td>51002.110000</td><td>0.000000</td></tr><tr><td>50%</td><td>652.000000</td><td>37.000000</td><td>5.000000</td><td>97198.540000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td><td>100193.915000</td><td>0.000000</td></tr><tr><td>75%</td><td>718.000000</td><td>44.000000</td><td>7.000000</td><td>127644.240000</td><td>2.000000</td><td>1.000000</td><td>1.000000</td><td>149388.247500</td><td>0.000000</td></tr><tr><td>max</td><td>850.000000</td><td>92.000000</td><td>10.000000</td><td>250898.090000</td><td>4.000000</td><td>1.000000</td><td>1.000000</td><td>199992.480000</td><td>1.000000</td></tr></tbody></table></div> <div><pre>[30]: data.corr()</pre><table><thead><tr><th></th><th>CreditScore</th><th>Geography</th><th>Gender</th><th>Age</th><th>Tenure</th><th>Balance</th><th>NumOfProducts</th><th>HasCrCard</th><th>IsActiveMember</th><th>EstimatedSalary</th><th>Exited</th></tr></thead><tbody><tr><td>CreditScore</td><td>1.000000</td><td>0.007888</td><td>-0.002857</td><td>-0.003965</td><td>0.000842</td><td>0.006268</td><td>0.012238</td><td>-0.005458</td><td>0.025651</td><td>-0.001384</td><td>-0.027094</td></tr><tr><td>Geography</td><td>0.007888</td><td>1.000000</td><td>0.004719</td><td>0.022812</td><td>0.003739</td><td>0.069408</td><td>0.003972</td><td>-0.008523</td><td>0.006724</td><td>-0.001369</td><td>0.035943</td></tr><tr><td>Gender</td><td>-0.002857</td><td>0.004719</td><td>1.000000</td><td>-0.027544</td><td>0.014733</td><td>0.012087</td><td>-0.021859</td><td>0.005766</td><td>0.022544</td><td>-0.008112</td><td>-0.106512</td></tr><tr><td>Age</td><td>-0.003965</td><td>0.022812</td><td>-0.027544</td><td>1.000000</td><td>-0.009997</td><td>0.028308</td><td>-0.030680</td><td>-0.011721</td><td>0.085472</td><td>-0.007201</td><td>0.285323</td></tr><tr><td>Tenure</td><td>0.000842</td><td>0.003739</td><td>0.014733</td><td>-0.009997</td><td>1.000000</td><td>-0.012254</td><td>0.013444</td><td>0.022583</td><td>-0.028362</td><td>0.007784</td><td>-0.014001</td></tr><tr><td>Balance</td><td>0.006268</td><td>0.069408</td><td>0.012087</td><td>0.028308</td><td>-0.012254</td><td>1.000000</td><td>-0.304180</td><td>-0.014858</td><td>-0.010084</td><td>0.012797</td><td>0.118533</td></tr><tr><td>NumOfProducts</td><td>0.012238</td><td>0.003972</td><td>-0.021859</td><td>-0.030680</td><td>0.013444</td><td>-0.304180</td><td>1.000000</td><td>0.003183</td><td>0.009612</td><td>0.014204</td><td>-0.047820</td></tr><tr><td>HasCrCard</td><td>-0.005458</td><td>-0.008523</td><td>0.005766</td><td>-0.011721</td><td>0.022583</td><td>-0.014858</td><td>0.003183</td><td>1.000000</td><td>-0.011866</td><td>-0.009933</td><td>-0.007138</td></tr><tr><td>IsActiveMember</td><td>0.025651</td><td>0.006724</td><td>0.022544</td><td>0.085472</td><td>-0.028362</td><td>-0.010084</td><td>0.009612</td><td>-0.011866</td><td>1.000000</td><td>-0.011421</td><td>-0.156128</td></tr><tr><td>EstimatedSalary</td><td>-0.001384</td><td>-0.001369</td><td>-0.008112</td><td>-0.007201</td><td>0.007784</td><td>0.012797</td><td>0.014204</td><td>-0.009933</td><td>-0.011421</td><td>1.000000</td><td>0.012097</td></tr><tr><td>Exited</td><td>-0.027094</td><td>0.035943</td><td>-0.106512</td><td>0.285323</td><td>-0.014001</td><td>0.118533</td><td>-0.047820</td><td>-0.007138</td><td>-0.156128</td><td>0.012097</td><td>1.000000</td></tr></tbody></table></div>		CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	count	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	mean	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700	std	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769	min	350.000000	18.000000	0.000000	0.000000	1.000000	0.000000	0.000000	11.580000	0.000000	25%	584.000000	32.000000	3.000000	0.000000	1.000000	0.000000	0.000000	51002.110000	0.000000	50%	652.000000	37.000000	5.000000	97198.540000	1.000000	1.000000	1.000000	100193.915000	0.000000	75%	718.000000	44.000000	7.000000	127644.240000	2.000000	1.000000	1.000000	149388.247500	0.000000	max	850.000000	92.000000	10.000000	250898.090000	4.000000	1.000000	1.000000	199992.480000	1.000000		CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	CreditScore	1.000000	0.007888	-0.002857	-0.003965	0.000842	0.006268	0.012238	-0.005458	0.025651	-0.001384	-0.027094	Geography	0.007888	1.000000	0.004719	0.022812	0.003739	0.069408	0.003972	-0.008523	0.006724	-0.001369	0.035943	Gender	-0.002857	0.004719	1.000000	-0.027544	0.014733	0.012087	-0.021859	0.005766	0.022544	-0.008112	-0.106512	Age	-0.003965	0.022812	-0.027544	1.000000	-0.009997	0.028308	-0.030680	-0.011721	0.085472	-0.007201	0.285323	Tenure	0.000842	0.003739	0.014733	-0.009997	1.000000	-0.012254	0.013444	0.022583	-0.028362	0.007784	-0.014001	Balance	0.006268	0.069408	0.012087	0.028308	-0.012254	1.000000	-0.304180	-0.014858	-0.010084	0.012797	0.118533	NumOfProducts	0.012238	0.003972	-0.021859	-0.030680	0.013444	-0.304180	1.000000	0.003183	0.009612	0.014204	-0.047820	HasCrCard	-0.005458	-0.008523	0.005766	-0.011721	0.022583	-0.014858	0.003183	1.000000	-0.011866	-0.009933	-0.007138	IsActiveMember	0.025651	0.006724	0.022544	0.085472	-0.028362	-0.010084	0.009612	-0.011866	1.000000	-0.011421	-0.156128	EstimatedSalary	-0.001384	-0.001369	-0.008112	-0.007201	0.007784	0.012797	0.014204	-0.009933	-0.011421	1.000000	0.012097	Exited	-0.027094	0.035943	-0.106512	0.285323	-0.014001	0.118533	-0.047820	-0.007138	-0.156128	0.012097	1.000000
		CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited																																																																																																																																																																																																																																	
count	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000																																																																																																																																																																																																																																		
mean	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700																																																																																																																																																																																																																																		
std	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769																																																																																																																																																																																																																																		
min	350.000000	18.000000	0.000000	0.000000	1.000000	0.000000	0.000000	11.580000	0.000000																																																																																																																																																																																																																																		
25%	584.000000	32.000000	3.000000	0.000000	1.000000	0.000000	0.000000	51002.110000	0.000000																																																																																																																																																																																																																																		
50%	652.000000	37.000000	5.000000	97198.540000	1.000000	1.000000	1.000000	100193.915000	0.000000																																																																																																																																																																																																																																		
75%	718.000000	44.000000	7.000000	127644.240000	2.000000	1.000000	1.000000	149388.247500	0.000000																																																																																																																																																																																																																																		
max	850.000000	92.000000	10.000000	250898.090000	4.000000	1.000000	1.000000	199992.480000	1.000000																																																																																																																																																																																																																																		
	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited																																																																																																																																																																																																																																
CreditScore	1.000000	0.007888	-0.002857	-0.003965	0.000842	0.006268	0.012238	-0.005458	0.025651	-0.001384	-0.027094																																																																																																																																																																																																																																
Geography	0.007888	1.000000	0.004719	0.022812	0.003739	0.069408	0.003972	-0.008523	0.006724	-0.001369	0.035943																																																																																																																																																																																																																																
Gender	-0.002857	0.004719	1.000000	-0.027544	0.014733	0.012087	-0.021859	0.005766	0.022544	-0.008112	-0.106512																																																																																																																																																																																																																																
Age	-0.003965	0.022812	-0.027544	1.000000	-0.009997	0.028308	-0.030680	-0.011721	0.085472	-0.007201	0.285323																																																																																																																																																																																																																																
Tenure	0.000842	0.003739	0.014733	-0.009997	1.000000	-0.012254	0.013444	0.022583	-0.028362	0.007784	-0.014001																																																																																																																																																																																																																																
Balance	0.006268	0.069408	0.012087	0.028308	-0.012254	1.000000	-0.304180	-0.014858	-0.010084	0.012797	0.118533																																																																																																																																																																																																																																
NumOfProducts	0.012238	0.003972	-0.021859	-0.030680	0.013444	-0.304180	1.000000	0.003183	0.009612	0.014204	-0.047820																																																																																																																																																																																																																																
HasCrCard	-0.005458	-0.008523	0.005766	-0.011721	0.022583	-0.014858	0.003183	1.000000	-0.011866	-0.009933	-0.007138																																																																																																																																																																																																																																
IsActiveMember	0.025651	0.006724	0.022544	0.085472	-0.028362	-0.010084	0.009612	-0.011866	1.000000	-0.011421	-0.156128																																																																																																																																																																																																																																
EstimatedSalary	-0.001384	-0.001369	-0.008112	-0.007201	0.007784	0.012797	0.014204	-0.009933	-0.011421	1.000000	0.012097																																																																																																																																																																																																																																
Exited	-0.027094	0.035943	-0.106512	0.285323	-0.014001	0.118533	-0.047820	-0.007138	-0.156128	0.012097	1.000000																																																																																																																																																																																																																																
Univariate Analysis	Exploration of individual variables (mean, median, mode, etc..)																																																																																																																																																																																																																																										
Bivariate Analysis	Relationships between two variables (correlation, scatter plots).																																																																																																																																																																																																																																										

	<pre>1]: sns.heatmap(data.corr(),annot=False)</pre> <p>1]: <Axes: ></p> 
Multivariate Analysis	-
Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	<p>Code to load the dataset into the preferred environment (e.g., Python, R).</p> <pre>[2]: #Reading the dataset dt=pd.read_csv(r"Churn_Modelling.csv")</pre>
Handling Missing Data	<p>Code for identifying and handling missing values.</p> <pre>[5]: dt.isnull().any()</pre>
Data Transformation	<p>Code for transforming variables (scaling, normalization).</p> <pre>[42]: #training and testing the data from sklearn.model_selection import train_test_split x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0) [43]: #Feature Scaling from sklearn.preprocessing import StandardScaler sc=StandardScaler() x_train=sc.fit_transform(x_train) x_test = sc.transform(x_test)</pre>
Feature Engineering	<p>Code for creating new features or modifying existing ones.</p> <pre>7]: #Removing columns data=dt.drop(['RowNumber','CustomerId','Surname'],axis=True)</pre>

Save Processed Data

Code to save the cleaned and processed data for future use.

```
[23]: #LabelEncoding

[24]: from sklearn.preprocessing import LabelEncoder
      le=LabelEncoder()

[25]: #data["CreditScore"]=le.fit_transform(data["CreditScore"])
      data["Geography"]=le.fit_transform(data["Geography"])
      data["Gender"]=le.fit_transform(data["Gender"])
      #data["Age"]=le.fit_transform(data["Age"])
      #data["Tenure"]=le.fit_transform(data["Tenure"])
      #data["Balance"]=le.fit_transform(data["Balance"])
      #data["NumOfProducts"]=le.fit_transform(data["NumOfProducts"])
      #data["HasCrCard"]=le.fit_transform(data["HasCrCard"])
      #data["IsActiveMember"]=le.fit_transform(data["IsActiveMember"])
      #data["EstimatedSalary"]=le.fit_transform(data["EstimatedSalary"])
      #data["Exited"]=le.fit_transform(data["Exited"])
```