



Data Collection and Preprocessing Phase

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

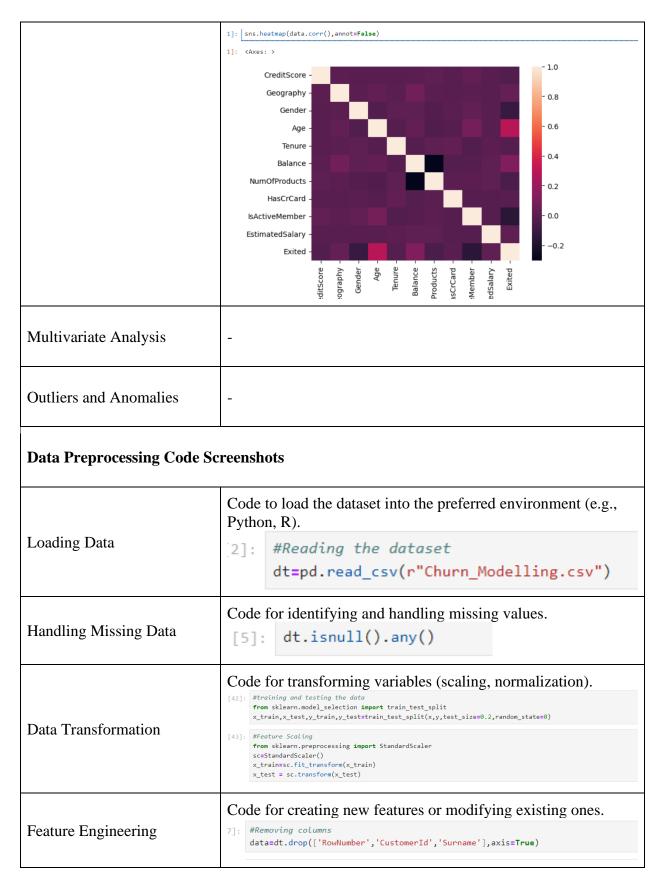
Data Exploration and Preprocessing Template

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

ic stat ata.describe(Credits ount 10000.00 nean 650.52 std 96.65 stmin 350.00 25% 584.00 652.00 75% 718.00 max 850.00) icore 0000 10000 10000 38800 38. 3299 10. 0000 18. 0000 32. 0000 37. 0000 44.	Age	Tenure 000.000000 5.012800 2.892174 0.000000 3.000000 5.000000	76485.8 62397.4	alance !	NumOfProducts 10000.000000 1.530200 0.581654 1.000000	HasCrCard 10000.00000 0.70550 0.45584	of the d IsActiveMember 10000.000000 0.515100 0.499797	ata. EstimatedSalary 10000.000000 100090.239881 57510.492818	Exited 10000.000000 0.203700
nean 650.52 std 96.65 min 350.00 25% 584.00 50% 652.00 718.00	0000 10000. 8800 38. 3299 10. 0000 18. 0000 32. 0000 37.	.000000 100 .921800 .487806 .000000 .000000	5.012800 2.892174 0.000000 3.000000	76485.8 62397.4	000000 889288 405202	10000.000000 1.530200 0.581654	10000.00000 0.70550 0.45584	10000.000000 0.515100	10000.000000	0.203700
std 96.65 min 350.00 25% 584.00 50% 652.00 75% 718.00	8800 38. 3299 10. 0000 18. 0000 32. 0000 37.	.921800 .487806 .000000 .000000	5.012800 2.892174 0.000000 3.000000	76485.8 62397.4 0.0	889288 405202	1.530200 0.581654	0.70550 0.45584	0.515100	100090.239881	0.203700
std 96.65 min 350.00 25% 584.00 50% 652.00 75% 718.00	3299 10. 0000 18. 0000 32. 0000 37.	.000000 .000000 .000000	2.892174 0.000000 3.000000	62397.4	105202	0.581654	0.45584			
min 350.00 25% 584.00 50% 652.00 75% 718.00	0000 18. 0000 32. 0000 37. 0000 44.	.000000	0.000000 3.000000	0.0				0.499797	57510.492818	0.402769
25% 584.00 50% 652.00 75% 718.00	0000 32. 0000 37. 0000 44.	.000000	3.000000		000000	1.000000	0.00000			
50% 652.00 75% 718.00	0000 37. 0000 44.	.000000		0.0			0.00000	0.000000	11.580000	0.000000
75% 718.00	0000 44		5.000000		000000	1.000000	0.00000	0.000000	51002.110000	0.000000
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7		000000		97198.5	40000	1.000000	1.00000	1.000000	100193.915000	0.000000
max 850.00		.000000	7.000000	127644.2	240000	2.000000	1.00000	1.000000	149388.247500	0.000000
	0000 92.	.000000	10.000000	250898.0	90000	4.000000	1.00000	1.000000	199992.480000	1.000000
ta.corr()	reditScore 6	Seography	Gender	Age	Tenure	Balance Numí	OfProducts H	asCrCard Is&ctiveM	emher EstimatedSa	arv Exited
CreditScore	1.000000	,		-		0.006268				•
Geography	0.007888	1.000000	0.004719	0.022812	0.003739	0.069408	0.003972 -	0.008523 0.0	-0.001	369 0.03594
Gender	-0.002857	0.004719	1.000000 -0	0.027544	0.014733	0.012087	-0.021859	0.005766 0.0	22544 -0.008	112 -0.106512
Age	-0.003965	0.022812 -	0.027544	1.000000 -	0.009997	0.028308	-0.030680 -	0.011721 0.0	185472 -0.007	201 0.28532
Tenure	0.000842					-0.012254				
	0.0.12200									
ActiveMember	0.025651					-0.010084				
timatedSalary	-0.001384	-0.001369 -	-0.008112 -0	0.007201	0.007784	0.012797		0.009933 -0.0		
Exited	-0.027094	0.035943 -	-0.106512	0.285323 -	0.014001	0.118533	-0.047820 -	0.007138 -0.1	56128 0.012	097 1.000000
ti	CreditScore Geography Gender Age Tenure Balance nOfProducts HasCrCard tiveMember matedSalary Exited	CreditScore 1.000000 Geography 0.007888 Gender -0.002857 Age -0.003965 Tenure 0.00628 Balance 0.005288 MOFProducts 0.002288 HasCrCard -0.005459 tweMember 0.025651 matedSalary -0.001384 Exited -0.027094	Geography 0.007888 1.000000 Gender -0.002857 0.004719 Age -0.003965 0.022812 Tenure 0.00042 0.003739 Balance 0.006268 0.06940 nOFProducts 0.012238 0.009972 HasCrCard -0.005458 -0.008523 thveMember 0.025651 0.006724 matedSalary -0.001384 -0.001309 Exited -0.027704 0.035943	CreditScore 1.00000 0.007888 -0.02857 Geography 0.007888 1.00000 0.04719 Gender -0.002857 0.004719 1.00000 Age -0.003965 0.022812 -0.027544 Tenure 0.00642 0.003739 0.014733 Balance 0.006268 0.06940 0.012807 HasCrCard 0.002548 -0.003972 -0.02576 tiveMember 0.02561 0.006724 0.02544 matedSalary 0.001384 -0.001369 -0.008112	CreditScore 1.000000 0.007888 -0.002857 -0.003965 Geography 0.007888 1.000000 0.04719 0.022812 Gender -0.002857 0.004719 1.000000 -0.027544 Age -0.003965 0.022812 -0.027544 1.000000 Tenure 0.000842 0.003293 -0.014733 -0.00990 Balance 0.006288 0.069408 0.012877 -0.028308 nOFProducts 0.012238 0.003972 -0.025695 -0.011721 tweMember 0.025651 0.006724 0.02544 0.08472 - matedSalary -0.001384 -0.001369 -0.00112 -0.00701 Exited -0.027094 0.035943 -0.106512 0.285323	CreditScore 1.000000 0.007888 -0.002857 -0.003955 0.000842 Geography 0.007888 1.000000 0.004719 0.022812 0.003739 Gender -0.002857 0.004719 1.00000 -0.027544 0.014733 Age -0.003965 0.022812 -0.027544 1.00000 -0.009997 Tenure 0.000842 0.003473 -0.00397 1.00000 -0.012254 nOFProducts 0.002288 0.069408 0.012087 0.028306 -0.013244 HasCrCard -0.005459 -0.008523 0.005766 -0.011721 0.022583 tweMember 0.025651 0.006724 0.022544 0.085472 -0.023862 matedSalary -0.01384 -0.003699 -0.005112 0.007701 0.007784 Exited -0.027094 0.035943 -0.106512 0.285323 -0.014001	CreditScore 1.00000 0.007888 -0.02857 -0.00395 0.009842 0.006268 Georgraphy 0.007888 1.00000 0.04719 0.022812 0.003739 0.069408 Gender -0.002857 0.004719 1.00000 -0.027544 0.014733 0.012087 Age -0.003865 0.022812 -0.027544 1.000000 -0.009997 0.028308 Tenure 0.000842 0.003793 0.014733 -0.009997 1.000000 -0.012254 Balance 0.006268 0.066408 0.012793 -0.028308 -0.012254 1.000000 nOFFroducts 0.012238 0.003972 -0.021859 -0.036680 0.013444 -0.304180 HasCrCard -0.005458 -0.008523 0.005766 -0.011721 0.022383 -0.014859 tiveMember 0.025651 0.007244 0.022544 0.085472 -0.028362 -0.01084	Creditscore 1.00000 0.07888 -0.02857 -0.03955 0.00842 0.006268 0.012238 -0.02187 Geography 0.007888 1.00000 0.04719 0.02212 0.003739 0.069408 0.003972 - Gender -0.002857 0.004719 1.00000 -0.02744 0.014733 0.012087 -0.021859 Age -0.003965 0.022812 -0.02744 1.00000 -0.09997 0.003838 -0.039680 - Tenure 0.006268 0.069408 0.012274 0.009997 1.00000 -0.12254 0.014444 onOFProducts 0.012238 0.003972 -0.012897 -0.028308 -0.012444 -0.304180 -1.000000 HasCrCard 0.005458 -0.008523 0.00576 -0.011721 0.022583 -0.014838 0.003183 tiveMember 0.025651 0.006724 0.02544 0.08472 -0.028362 -0.01084 0.00912 -	Creditiscore 1.00000 0.07888 -0.02857 -0.03955 0.00842 0.006268 0.012238 -0.05459 0.0 Geography 0.007888 1.00000 0.004719 0.022812 0.003739 0.069408 0.003972 -0.008523 0.0 Gender -0.002857 0.004719 1.00000 -0.027544 0.014733 0.012087 -0.021859 0.005766 0.0 Age -0.003955 0.022812 0.027544 1.000000 -0.009997 0.028808 -0.013680 0.011721 0.0 Tenure 0.0064268 0.003973 0.024733 -0.02910 -0.002254 0.0012254 0.003444 0.022583 -0.0 noPfProducts 0.012238 0.093772 -0.028308 -0.012254 1.000000 -0.03183 0.0 0.00 HasCrCard -0.02561 0.00823 0.005766 -0.011721 0.022583 -0.014858 0.003183 1.00000 -0.03183 0.00000 -0.00000 -0.001889 0.001818 0.001818 0.001818 </td <td>CreditScore 1.00000 0.00788 -0.02857 -0.03955 0.00842 0.006288 0.012238 -0.005458 0.025651 -0.001 Geography 0.007888 1.00000 0.04719 0.02212 0.003739 0.069408 0.003972 -0.008523 0.006724 -0.001 Gender -0.002857 0.00479 1.00000 -0.02744 0.014733 0.012087 -0.021859 0.005766 0.022544 -0.008 Age -0.003955 0.022312 0.02774 1.00000 -0.003808 -0.038680 -0.017211 0.085472 -0.007 Tenure 0.006680 0.003973 0.014733 -0.00997 1.00000 -0.012544 0.012444 -0.02383 -0.028682 -0.018482 -0.010842 0.007 Balance 0.006268 0.006972 -0.02890 -0.012254 1.00000 -0.304180 -0.014858 -0.010844 0.01248 -0.010844 -0.01284 -0.010844 -0.014858 -0.010844 -0.01284 -0.010844 -0.014858 -0.010844</td>	CreditScore 1.00000 0.00788 -0.02857 -0.03955 0.00842 0.006288 0.012238 -0.005458 0.025651 -0.001 Geography 0.007888 1.00000 0.04719 0.02212 0.003739 0.069408 0.003972 -0.008523 0.006724 -0.001 Gender -0.002857 0.00479 1.00000 -0.02744 0.014733 0.012087 -0.021859 0.005766 0.022544 -0.008 Age -0.003955 0.022312 0.02774 1.00000 -0.003808 -0.038680 -0.017211 0.085472 -0.007 Tenure 0.006680 0.003973 0.014733 -0.00997 1.00000 -0.012544 0.012444 -0.02383 -0.028682 -0.018482 -0.010842 0.007 Balance 0.006268 0.006972 -0.02890 -0.012254 1.00000 -0.304180 -0.014858 -0.010844 0.01248 -0.010844 -0.01284 -0.010844 -0.014858 -0.010844 -0.01284 -0.010844 -0.014858 -0.010844











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["Geography"]] le.fit_transform(data["Geography"]) #data["Gender"]] #data["Age"]=le.fit_transform(data["Age"]) #data["Age"]] #data["Balance"]=le.fit_transform(data["NumOfProducts"]) #data["Balance"]=le.fit_transform(data["NumOfProducts"]) #data["IndexCrand"]] #data["IndexCrand"]=le.fit_transform(data["IndexCrand"]) #data["IndexCrand"]=le.fit_transform(data["IndexCrand"]) #data["Exited"]=le.fit_transform(data["Exited"]) #data["Exited"]=le.fit_transform(data["Exited"])