



Data Collection and Preprocessing Phase

Date	15 th July 2024
Team ID	739743
Project Title	Auto Foresight : A Predictive Model for Streamlining Car Loan Repayment Planning
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										
	dateset								-		
		ent Income Car	Owned Bike	Owned Act	ive Loan Ho	use Own Chi	ld Count Cr	edit Amount L	oan Annuity Acc	ompany Client	Client_Permanent
	0 12142509	6750	0.0	0.0	1.0	0.0	0.0	61190.55	3416.85	Alone	-
	1 12138936	20250	1.0	0.0	1.0	NaN	0.0	15282	1826.55	Alone	
D	2 12181264	18000	0.0	0.0	1.0	0.0	1.0	59527.35	2788.2	Alone	in .
	3 12188929	15750	0.0	0.0	1.0	1.0	0.0	53870.4	2295.45	Alone	30
Data Overview	4 12133385	33750	1.0	0.0	1.0	0.0	2.0	133988.4	3547.35	Alone	E
		-		- 44	1.000	7	***	199	140		20-
	121851 12207714	29250	0.0	0.0	NaN	1,0	0.0	107820	3165.3	Relative	pai
	121852 12173765	15750	0.0	1.0	1.0	0.0	0.0	104256	3388.05	Alone	340
	121853 12103937	8100	0.0	1.0	0.0	1.0	1.0	55107.9	2989.35	Alone	to.
	121854 12170623	38250	1.0	1.0	0.0	1.0	0.0	45000	2719.35	Alone	
	121855 12105610	9000	1.0	1.0	1.0	1.0	1.0	62428.95	4201.65	Alone	sai.
	121856 rows × 40 column	ns									
Univariate Analysis	freq Name: Clie print("Cre churn_cust Credit Amo count unique top freq Name: Cred sns.countplot(x	9566 322 13500 1058 ent_Inco ddit Amo comers[' bunt 9570.6 2251.6 45000.6 303.6 lit Amou	ome, dount") Credi	type: t_Amou	object nt'].d	: describ					
Bivariate Analysis	1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 -	Male		Female	nt Gend						





1 8	Screenshots											
	Meading the dataset											
	dataset=pd.read_c	sv(r*D:\Docume	nts\dataset\tra	n.csv")								
	C:\Users\Sharan\AppNata\Loca\Yemp\ipykernel_5604\3481427543.py:1: DtypeWarning: Columns (1,7,8,16,17,18,19,20,35) have mixed types. Specify dtype opti n import or set low memory=false.											
	<pre>dataset.pd.read_csv(n^D:\Documents\dataset\train.csv*) dataset</pre>											
L 10 D - 4 -		Client Income	C C	O 4	abortono t		child cours	C Et 1	Lana Amerika	Accompany Client	Client Brown	
Loading Data	0 12142509	6750	0.0	0.0	1.0	0.0	0.0	61190.55	3416.85	Alone		
	1 12138936	20250	1.0	0.0	1.0	NaN	0.0	15282		Alone		
	2 12181264	18000	0,0	0.0	1,0	0.0	1.0	59527.35		Alone		
	3 12188929	15750	0.0	0.0	1.0	1.0	0.0	53870.4	2295.45	Alone	22	
	4 12133385	33750	1.0	0.0	1.0	0.0	2.0	133988.4	3547.35	Alone		
	30.76.30000	337.00	1.0			914.	A	1223001	, 20,37120	T. T		
	121851 12207714	29250	0.0	0.0	NaN	1.0	0.0	107820	3165.3	Relative	-	
		×2.02.010.10.000000000000000000000000000										
	# [Data Pre-Proce dataset= dataset.				ault','Age_	Days','Emplo	oyed_Days',	'Score_Source_1	L','Score_Sourc	e_2','Score_Sourc	e_3','Registra	
	dataset.head()				26 200	_						
Handling Missing Data	CO POLICIONARY									mpany_Client C	lient_Housing_Ty	
5 5	0 12142509	6750	0.0	.0	1.0	0.0	0.0	61190.55	3416.85	Alone	Но	
	1 12138936	20250	1.0	.0	1.0	NaN	0.0	15282	1826.55	Alone	Ho	
	2 12181264	18000	0.0	.0	1.0	0.0	1.0	59527.35	2788.2	Alone	Far	
	3 12188929	15750	0.0	.0	1.0	1.0	0.0	53870,4	2295.45	Alone	Ho	
	4 12133385	33750	1.0	.0	1.0	0.0	2.0	133988.4	3547.35	Alone	Но	
	5 rows × 25 columns											
Data Transformation	dataset['Cli	dit_Amount	'] = pd.to_	numeric(dataset['Credit_ <i> </i>	Amount']	,errors='cc	perce')	ive'],errors:	·'coence')	
	dataset['Loa									1,51,012	,	
	# -Filling Missin	Values and Cr	reating data fra							© ↑	↓ ≐ ♀ ii	
	: from mixtend, prep column_names=['ID	','Client_Incom			,'Active_to	an','House_(Dwn','Child	count, Credit	Amount , Loan	remotely of necompo	ny_Client','Cli	
	column_names=['ID	, Client_Incom	e','Car_Owned',	Bike_Owned	,'Active_Lo	aan','House_(hm','Child	count, create	Amount , Loan	, , , , , ,	ny_Client','Cli	
	column_names=['ID #Create dataframe loan_dataspd.Data loan_data	','Client_Incom	e','Car_Owned',	Bike_Owned							ny_Client',"Cli	
	column_names=['ID #Create dataframe : loan_datampd.Data loan_data	rane(result,co	e','Car_Owned', clumns=column_na	Bike_Owned es) Owned A	ctive_Loan H	iouse_Own C	hild_Count	Credit_Amount	Loan Annuity A	Accompany_Client .	ny_Client','Cli	
	column_names=['ID #Create dataframe loan_dataspd.Data loan_data	Franc(result,co	e','Car_Owned',	Bike_Owned							. Client Housing	
eature Engineering	column_names=['ID #Create dataframe : loan_dataspd.Data loan_data : ID 0 12132045.0	Frame(result,come 27000.0 13500.0	clumns=column_na Car_Owned Bik	es) Owned Ac	ctive_Loan H	louse_Own C	hild_Count	Credit_Amount 60750.00	Loan Annuity A	Accompany_Client .	. Client Housing	
Ceature Engineering	column_names['ID' #Create dataframe loan_datampd.Data loan_data 10 12320450 1 121920450 2 12201738.0 3 1213195.0	Client_Income 27000.0 13500.0 15750.0	car_Owned Bik	es) Owned A 1.0 1.0 0.0 1.0	1.0 1.0 1.0 1.0	1.0 1.0 1.0 1.0	0.0 0.0 0.0 0.0	Credit_Amount 60750.00 28440.00 18000.00 59301.00	Loan_Annuity A 7222.50 1851.30 900.00 1746.90	1.0 1.0	. Client Housing	
Ceature Engineering	column_namese[*ID #Create dataframe loan_datampd.Data loan_data 10 1239655.0 1 1239655.0 2 12201738.0 3 1213195.0 4 12214557.0	Client_Income 27000.0 13500.0 15750.0	Car_Owned Bik	es) Owned Ad 1.0 1.0 0.0	1.0 1.0 1.0	1.0 1.0 1.0	hild_Count 0.0 0.0	Credit Amount 60750.00 28440.00 18000.00	Loan_Annuity A 7222.50 1851.30 900.00	1.0	. Client Housing	
Feature Engineering	column_namese[*ID #Create dataframe loan_datampd.Data loan_data 10 0 12132045.05 1 12196654.0 2 122017380.0 3 12131395.0	Client_Income 2700.0 13500.0 13500.0	Car_Owned Bik	Owned At 1.0 1.0 0.0 1.0	1.0 1.0 1.0 1.0 1.0	1.0 1.0 1.0 1.0 1.0	hild_Count 0.0 0.0 0.0 0.0	Credit_Amount 60750.00 28440.00 18000.00 59301.00 30234.15	Loan_Annuity A 7222.50 1851.30 900.00 1746.90 1840.05	1.0	. Client Housing	
Feature Engineering	column_namese[*ID #Create dataframe loan_datampd.Data loan_data 10 1239655.0 1 1239655.0 2 12201738.0 3 1213195.0 4 12214557.0	Client_Income 27000.0 13500.0 15750.0 13500.0	car_Owned Bik	es) Owned A 1.0 1.0 0.0 1.0	1.0 1.0 1.0 1.0	1.0 1.0 1.0 1.0	0.0 0.0 0.0 0.0	Credit_Amount 60750.00 28440.00 18000.00 59301.00	Loan_Annuity A 7222.50 1851.30 900.00 1746.90	1.0 1.0	. Client Housing	
Feature Engineering	column_namese[*10 #Create dataframe loan_datarpd_flata loan_data 10 0 12132045.01 1 12196654.0 2 12201738.0 3 12131950.0 4 12214557.0 224017 12136406.0	Frame(result,cc) Client_Income 27000.0 13500.0 15750.0 12150.0 15750.0	Car_Owned ', Car_Owned Bik 0.0 0.0 1.0 0.0 0.0	Owned Ad 1.0 1.0 1.0 1.0 - 0.0 1.0 - 0.0	1.0 1.0 1.0 1.0 1.0 1.0	1.0 1.0 1.0 1.0 0.0	hild_Count 0.0 0.0 0.0 0.0 0.0 0.0 0.0	Credit Amount 60750.00 28440.00 18000.00 59301.00 30234.15	Loan_Annuity A 7222.50 1851.30 900.00 1746.90 1840.05 — 2383.65	1.0	Client Housing	
Feature Engineering	column_names["ID" #Create dataframe loan_datampd.Bata loan_data ID 0 1232040.0 1 12190654.0 2 12201738.0 3 1213195.0 4 12214557.0 224017 12136406.0 224018 12173765.0	Client_Income 27000.0 13500.0 13500.0 15750.0 15750.0 15750.0 15750.0 15750.0 15750.0 15750.0 15750.0	Car_Owned Sile 0.0 0.0 1.0 0.0 0.0 0.0 0.0 0.	Owned Ad 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.	1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0	1.0 1.0 1.0 1.0 0.0	hild_Count	Credit_Amount 60750.00 28440.00 18000.00 59301.00 30234.15 78192.00 104256.00	7222.50 1851.30 900.00 1746.90 1840.05 — 2383.65 3388.05	1.0 - 1.0 -	Client ', 'Cli	