



## **Data Collection and Preprocessing Phase**

Date	15 <sup>th</sup> July 2024
Team ID	739823
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										
	dataset								-		
		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
	<b>0</b> 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	
	<b>2</b> 12181264	18000	0.0	0.0	1.0	0.0	1.0	59527.35	2788.2	Alone	in .
<b>D</b>	<b>3</b> 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	jui
	<b>121852</b> 12173765	15750	0.0	1.0	1.0	0.0	0.0	104256	3388.05	Alone	340
	<b>121853</b> 12103937	8100	0.0	1.0	0.0	1.0	1.0	55107.9	2989.35	Alone	to.
	<b>121854</b> 12170623	38250	1.0	1.0	0.0	1.0	0.0	45000	2719.35	Alone	
	<b>121855</b> 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						





	Screenshots										
	#Reading the dataset										
	datasetzpd.read_c	sv(r°D:\Docum	nts\dataset	\train.csv"	)						
	C:\Users\Sharan\App@uta\Local\Temp\ipykernel_5684\3481427543.py:1: DtypeWarning: Columns (1,7,8,16,17,18,19,28,35) have mixed types. Specify dtype option import or set low_memory=False.  dataset-pd.read_csv(r*0:\locuments\dataset\train.csv*)										
	dataset										
Loading Data	ID	Client_Income	Car_Owned	Bike_Owned	f Active_Loan	House_Ow	n Child_Count	Credit_Amount	Loan_Annuity	Accompany_Client	Client_Perma
20	<b>0</b> 12142509	6750	0.0	0.0	1.0	0	0.0	61190.55	3416.85	Alone	10.2
	1 12138936	20250	1.0	0.0	1.0	Na	N 0.0	15282	1826.55	Alone	100
	2 12181264	18000	0,0	0.0			0 1.0		2788.2	Alone	
	<b>3</b> 12188929	15750	0.0	0.0	1.0	1	0.0	53870.4	2295.45	Alone	120
	4 12133385	33750	1.0	0.0	1.0	0	0 2.0	133988.4	3547.35	Alone	-
	-		-				-	-	.aus	-	m.
	<b>121851</b> 12207714	29250	0.0	0.0	) NaN	1	0.0	107820	3165,3	Relative	753
	# [Data Pre-Proce	ssing] -Handi	ing missing	values							
	dataset= dataset. dataset.head()	drop(['Credit	Bureau','So	cial_Circle	_Default','Ag	e_Days','E	mployed_Days',	'Score_Source_1	','Score_Sour	ce_2','Score_Source	_3','Registrat
11' M' ' D.	2	t Income. Car i	Owned Bike	Owned Act	ive Loan Hou	se Own Ch	ild Count Cred	it Amount Loan	Annuity Acco	mpany_Client Cl	lient Housing Ty
Handling Missing Data	<b>0</b> 12142509	6750	0.0	0.0	1.0	0.0	0.0	61190.55	3416.85	Alone	Ho
	1 12138936	20250	1.0	0.0	1.0	NaN	0.0	15282	1826.55	Alone	Но
	2 12181264	18000	0.0	0.0	1.0	0.0	1.0	59527.35	2788.2	Alone	Fan
	3 12188929	15750	0.0	0.0	1.0	1.0	0.0	53870,4	2295.45	Alone	Ho
	- WARRANGES										
	4 12133385	33750	1.0	0.0	1.0	0.0	2.0	133988.4	3547.35	Alone	Но
ata Transformation	dataset['Cli	ent_Income	e'] = pd.	to_numer	ic(datase	:['Clien	t_Income']	,errors=' <mark>co</mark>	erce')		
	dataset['Cre	dit_Amount	:'] = pd.	to_numer	ic(datase	['Credi	t_Amount']	,errors='co	erce')		
	dataset['Pop	ulation_Re	gion_Rel	ative']	= pd.to_nu	umeric(d	ataset['Po	pulation_Re	gion_Relat	ive'],errors=	'coerce')
	1.1	n_Annuity'	] = pd.t	o_numeri	c(dataset	'Loan_A	nnuity'],e	rrors='coer	ce')		
	dataset[ 'Loa										
		yalues and C	reating data	frame							
	# -Filling Missing	rocessing impo	rt Transacti	onEncoder						@ ↑	V ± ∓ 1
	# -Filling Missing	rocessing impo	rt Transacti	onEncoder	wned','Active	Loan', 'Hou	se_Own','Child	_Count','Credit	Amount','Loan	↑ _Annuity','Accompan	↓ ≛ 〒   sy_Client',"Cli
	# -Filling Missing from mixtend.prepicolumn_names=['10' #Create dataframe.	rocessing impo ','Client_Inco	rt Transacti me','Car_Own	onEncoder ed','Bike_O	wned/,'Active	Loan', 'Hou	se_Own','Child	Count','Credit	Amount', 'Loan	ē ↑ _Annuity','Accompan	↓ ± 〒   y_Client','Cli
	: # -Filling Missing : from mlxtend.preper column_names=['ID'	rocessing impo ','Client_Inco	rt Transacti me','Car_Own	onEncoder ed','Bike_O	wned', 'Active	Loan','How	se_Own','Child	_Count','Credit	Amount', Loan	個 ↑ _Annuity', Accompan	↓ ≛ ⊊ I
	# -Filling Missim from mlatend.preper column_namese['ID #Create datoframe: lnam_dataspd_Oatail loam_data ID	rocessing impo ','Client_Inco Frame(result,c	rt Transacti me','Car_Own	onEncoder ed','Bike_O n_names)						高 个 Annuity','Accompan	y_Client','Cli
	# -Filling Missing from mlatend, propt column_namese['ID #Create dataframe loan_datampd_Ostal loan_data   ID   12132045.0	rocessing impo	rt Transacti me','Car_Own blumns=colum Car_Owned	onEncoder ed','Bike_0 n_names) Bike_Owned	Active_Loan	House_Ow	Child_Count	Credit_Amount 60750.00	Loan Annuity 7222.50	Annuity', 'Accompan  Accompany_Client	y_Client','Cli
eature Engineering	# -Filling Missim from mlatend.preper column_namese['ID #Create datoframe: lnam_dataspd_Oatail loam_data ID	rocessing impo	rt Transacti me','Car_Own  clumns=colum  Car_Owned  0.0  0.0	onEncoder ed', 'Bike_O  n_names)  Bike_Owned  1.0	d Active_Loan 1.0 1.0	House_Ow	n Child_Count 0 0.0 0 0.0	Credit_Amount	Loan_Annuity	Annuity', 'Accompan	vy_Client', 'Cli
eature Engineering	# -Filling Missin; frem mlatend, prop; column_names=['10  #Create dataframe: loan_datampd_Ostal loan_data    ID  0 12132045.0  1 12196554.0	rocessing impo	rt Transacti me','Car_Own plumns=colum Car_Owned 0.0 0.0	onEncoder ed','Bike_O  n_names)  Bike_Ownec  1.0	d Active_Loan 1.0 1.0 1.0	House_Ow	n Child_Count 0 0.0 0 0.0 0 0.0	Credit_Amount 60750.00 28440.00	7222.50 1851.30	Annuity', 'Accompan  Accompany_Client  1.0 =	y_Client', 'Cli
eature Engineering	# -Filling Missing from mixtend.prept column_namess['ID #Create data/rame: loan_dataspd.Ostal loan_data  1	rocessing impo	rt Transacti se', 'Car_Own clumnszcolum Car_Owned 0.0 0.0 1.0	onEncoder ed','Bike_0 n_names) Bike_Ownec 1.4 1.4 1.4	d Active_Loan 1.0 1.0 1.0 1.0 1.0	House_Ow 1. 1.	n Child_Count 0 0.0 0 0.0 0 0.0 0 0.0	Credit Amount 60750.00 28440.00 18000.00	7222.50 1851.30 900.00	Accompany_Client  1.0 = 1.0 = 1.0 =	y_Client', "Cli
eature Engineering	# -Filling Missim from mixtend.preps column_namese[*ID #Create dataframe: lnam_datampd.Datail loan_data 1	rocessing impo	rt Transacti ne', 'Car_Oun Olumnszcolum Car_Owned 0.0 0.0 0.0	onEncoder ed','8ike_0 n_names)  Bike_Ownec  1.4  0.0  1.4	d Active_Loan 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0	1. 1. 1. 0.	n Child_Count 0 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 / 7222.50 1851.30 900.00 1746.90 1840.05	Annuity', 'Accompan  Accompany_Client  1.0 _  1.0 _  1.0 _  1.0 _  1.0 _  -  1.0 _	y_Client', 'Cli
eature Engineering	# -Filling MissInju from mlatend.preps column_names['ID #Create dataframe loan_datampd.Datal loan_data  1	rocessing important processing	Car_Owned  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0	onEncoder onEncoder on_names)  Bike_Ownec  1.4  1.4   0.4	1 Active_Loan 1 1.0 1 1.	House_Ow 1. 1. 1.	n Child_Count 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0	Credit_Amount 60750,00 28440,00 18000,00 59301,00	7222.50 1851.30 900.00 1746.90	Annuity', 'Accompan'  Accompany.Client  1.0 -  1.0 -  1.0 -  1.0 -  1.0 -	Client Housing
eature Engineering	# -Filling Missing from mixtend.prept column_namess[*10 #Create data/rame: loan_data.pd 10 12132045 1 1219654.0 2 12201738.0 3 12131195.0 4 12214557.0	rocessing important processing	Car_Owned  0.0  1.0  0.0  0.0  0.0  0.0  0.0  0.	ontncoder ed', '81ke_0  n_names)  Bike_Ownec  1.4  1.4  -  0.0  1.4	1 Active_Loan 1 1.0	House_Ow 1. 1. 1. 1. 1. 0. 0. 0.	n Child_Count  0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 0.0 0 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 78192,00	7222.50 1851.30 900.00 1746.90 1840.05	Accompany_Client  1.0 _ 1.0 _ 1.0 _ 1.0 _ 1.0 _ 1.0 _ 1.0 _	Client Housing
eature Engineering	# -Filling Missing from mixtend.preps column_namess[*ID #Create data/rame: loan_dataspd.Oatal loan_data  1D 0 12320450.0 1 12196054.0 2 12201738.0 3 1213195.0 4 12214557.0 224017 12136406.0 224018 12173765.0	rocessing Laponovariant Client Income  Client Income  27000.0  13500.0  13500.0  15750.0  18750.0  18750.0  18750.0  18750.0  18750.0  18750.0  18750.0  18750.0  18750.0	Car_Owned  Car_Owned  0.0  1.0  0.0  0.0  0.0  1.0  1.0  1.	ontncoder ed', '8ike_0 n_names) Bike_Ownec 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4	i Active_Loan  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.	1. 1. 1. 0. 0. 0. 0.	n Child_Count 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.	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	Accompany.Client 1.0 1	Client Housing