

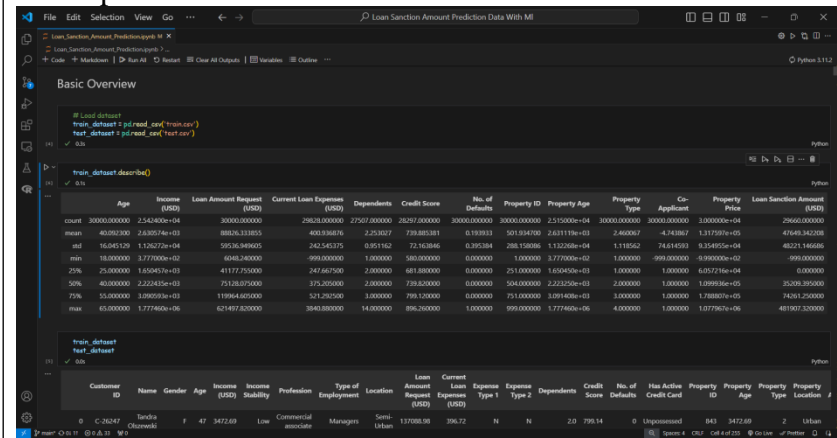
Data Collection and Preprocessing Phase

Date	15 June 2024
Team ID	739926
Project Title	Loan Sanction Amount Prediction Data With ML
Maximum Marks	6 Marks

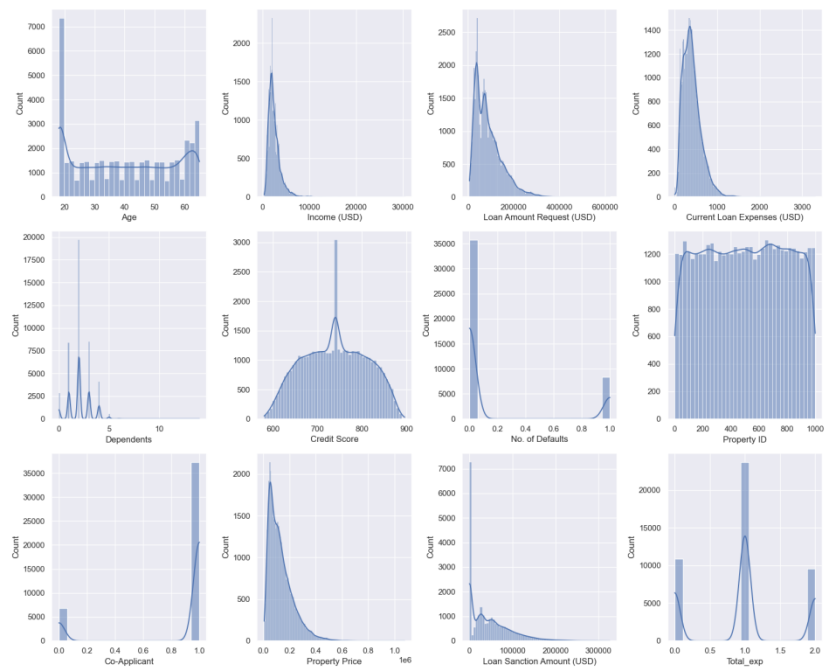
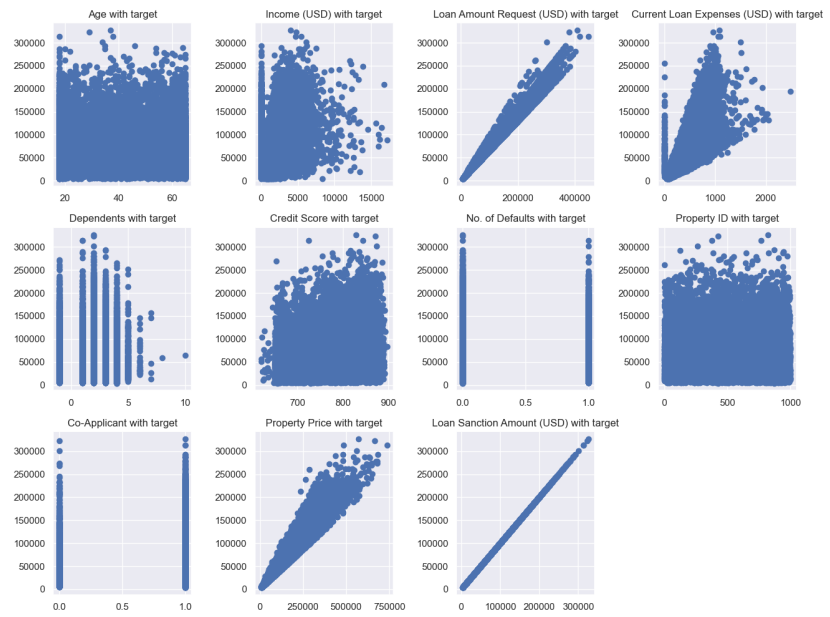
Data Exploration and Preprocessing Template

Data Exploration and Preprocessing Template

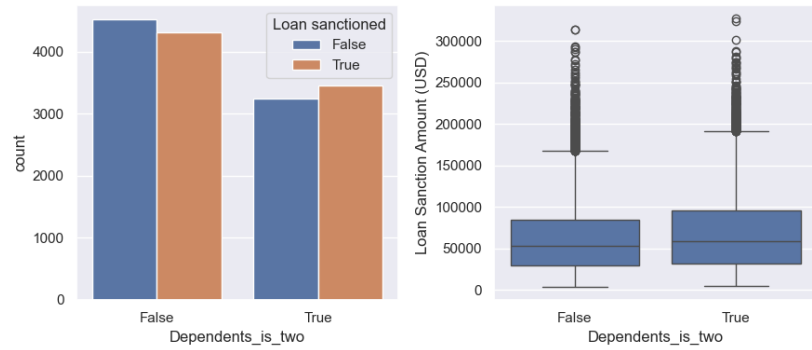
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	<p>Dimension: 8 rows x 131 columns</p> <p>Descriptive statistics:</p>  <p>The screenshot displays a Jupyter Notebook interface with the following content:</p> <ul style="list-style-type: none"> Basic Overview: A code cell showing the loading of the dataset into 'train_default' and 'test_default' variables. train_default.describe(): A table showing summary statistics for the training dataset across 131 columns. The columns include Age, Income, Loan Amount Request, Current Loan Expense, Dependents, Credit Score, No. of Defaults, Property ID, Property Age, Property Type, Co-Applicant, Property Price, and Loan Sanction Amount (USD). The statistics provided are count, mean, std, min, 25%, 50%, 75%, and max. test_default.describe(): A table showing summary statistics for the testing dataset across 131 columns. The columns include Customer ID, Name, Gender, Age, Income, Income Stability, Profession, Type of Employment, Location, Loan Amount Request, Current Loan Expense, Expense Type 1, Expense Type 2, Dependents, Credit Score, No. of Defaults, Has Active Credit Card, Property ID, Property Age, Property Type, and Property Location. The statistics provided are count, min, max, and a few specific values.

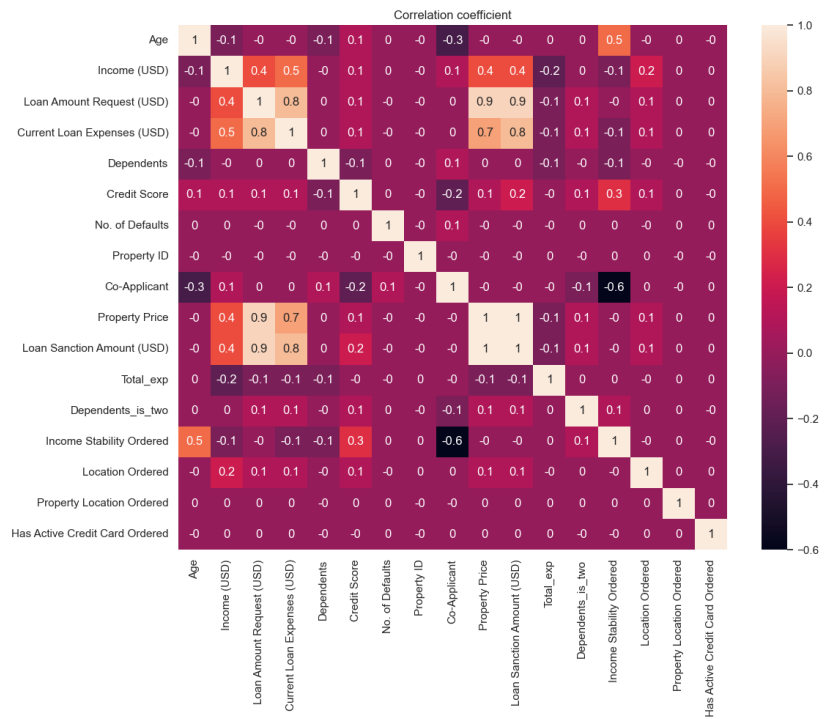
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading train Data

FileEditSelectionViewGo

Loan Sanction Amount Prediction Data With ML

2 Loan Sanction Amount Prediction Data With ML

Loan Sanction Amount Prediction Data With ML

File Edit Selection View Go

Python 3.8.5

Train, default

Test, default

20000 rows × 23 columns

Customer ID	Name	Gender	Age	Income (USD)	Marital Status	Profession	Type of Employment	Location	Loan Request (USD)	Current Loan (USD)	Exposure Type 1	Exposure Type 2	Dependents	Credit Score	No. of Defaults	Has Active Credit Card	Property ID	Property Age	Property Type	Property Location	Co-Applicant	Property Price			
0	C-20247	Tanisha Choudhary	F	40	1477.00	Low	Commercial associate	Manager	Urban	117080.00	296.72	N	N	2.0	786.14	0	Unimproved	841	1477.00	2	Urban	1	236444.5		
1	C-23567	Manavendra Dha	F	37	1154.84	Low	Working	Sales staff	Rural	104771.50	463.76	Y	Y	2.0	833.31	0	Unimproved	22	1164.84	1	Rural	1	142327.3		
2	C-34000	Ritesh Gadhay	F	32	1266.27	Low	Working	Self	Urban	176668.91	463.15	N	Y	3.0	827.44	0	Unimproved	1	1266.27	1	Urban	1	308991.24		
3	C-10600	Rishi Sankhala	M	40	1366.72	High	Pensioner	Self	Rural	97000.16	446.15	N	Y	2.0	833.30	0	Inactive	730	1366.72	1	Self	Urban	0	125412.1	
4	C-12106	Sanjay Sankhala	F	40	1456.23	High	Pensioner	Self	Urban	100000.00	376.10	N	N	NaN	NaN	0	Inactive	156	1456.23	4	Self	Urban	1	180000.0	
10000	C-9076	Tanisha Dha	F	38	1346.60	Low	Commercial associate	Self	Urban	156760.07	500.59	Y	Y	4.0	684.22	0	Inactive	681	1346.60	4	Self	Urban	1	212778.0	
10006	C-17587	Karlyn Madhavan	M	32	2019.76	Low	Working	Core staff	Urban	47624.80	343.62	Y	Y	NaN	786.14	0	Inactive	211	2019.76	4	Urban	1	88916.95		
10007	C-46479	Karlyn Madhavan	M	38	2252.03	Low	Working	Core staff	Self	Urban	18670.86	101.80	Y	N	1.0	636.46	0	Inactive	270	2252.03	2	Rural	0	21588.27	
10008	C-3000	Manavendra Dha	F	21	1846.35	Low	Working	Self	Urban	95400.73	273.21	N	Y	2.0	863.46	0	Unimproved	489	1846.35	1	Self	Urban	1	130081.17	
10009	C-42462	Manavendra Dha	M	21	1233.16	Low	Working	Laborman	Self	Urban	107960.03	963.43	Y	Y	3.0	827.16	1	Active	368	1233.16	4	Self	Urban	0	133425.43

20000 rows × 23 columns

Default stage
For train data
Train, test = train, default stage
Train, test = train, default stage
For Test data
Test, train = train, default stage
Test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

Python 3.8.5

Home < > Close of Screen

Loading test Data

FileEditSelectionViewGo↔↔

Handling Missing Data
In train and test

FileEditSelectionViewGo

Loan Sanction Amount Prediction Data With ML

20000 rows × 23 columns

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

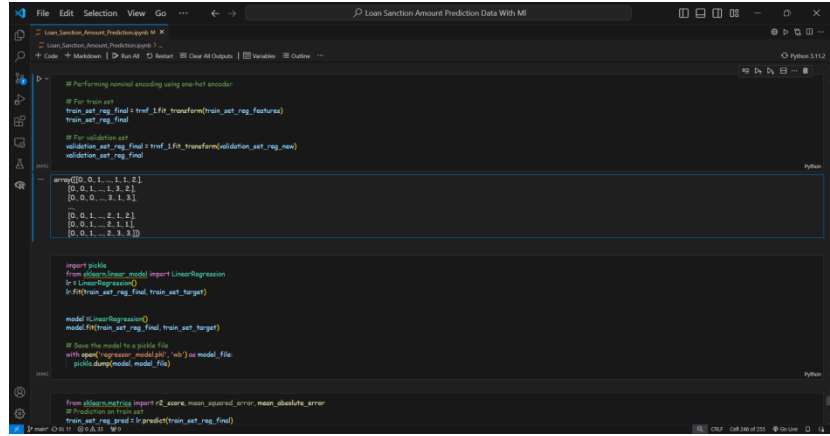
At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data
Train, test = train, default stage
Train, test = train, default stage
At For Test data
At test, train = train, default stage
At test, test = test, default stage
print("There are train, test, train and test, valid columns in train dataset")

20000 rows × 23 columns

At default stage
At For Train data

Data Transformation	 <pre> # Performing manual encoding using one-hot encoder # For train set train_set_reg_final = tfidf_transformer.transform(train_set_reg_features) train_set_reg_final # For validation set validation_set_reg_final = tfidf_transformer.transform(validation_set_reg_new) validation_set_reg_final array([[0, 1, ..., 1, 1, 2], [0, 0, 1, ..., 1, 3, 0], [0, 0, 0, ..., 3, 1, 1], [0, 0, 1, ..., 2, 1, 0], [0, 0, 1, ..., 2, 1, 1], [0, 0, 1, ..., 2, 3, 3]]) import pickle from sklearn.linear_model import LinearRegression lr = LinearRegression() lr.fit(train_set_reg_final, train_set_reg_target) model = LinearRegression() model.fit(train_set_reg_final, train_set_reg_target) # Save this model to a pickle file with open("regressor_model.pkl", "wb") as model_file: pickle.dump(model, model_file) from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error # Prediction on train set train_set_reg_pred = lr.predict(train_set_reg_final) </pre>
Feature Engineering	-
Save Processed Data	-