

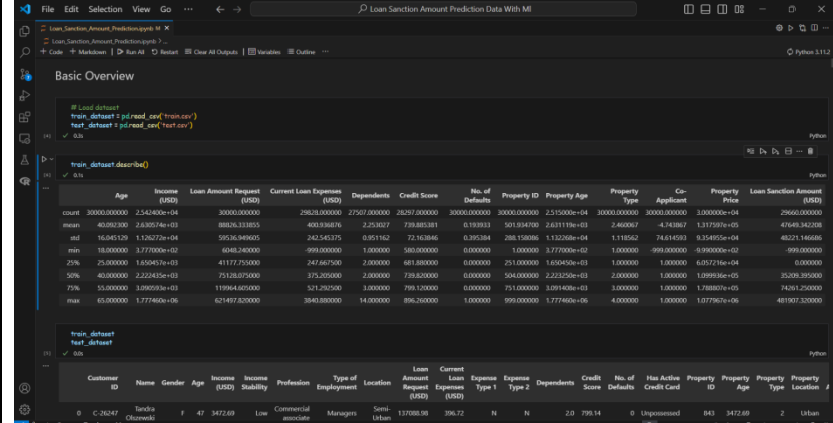
Data Collection and Preprocessing Phase

Date	15 June 2024
Team ID	739732
Project Title	Loan Sanction Amount Prediction Data With ML
Maximum Marks	6 Marks

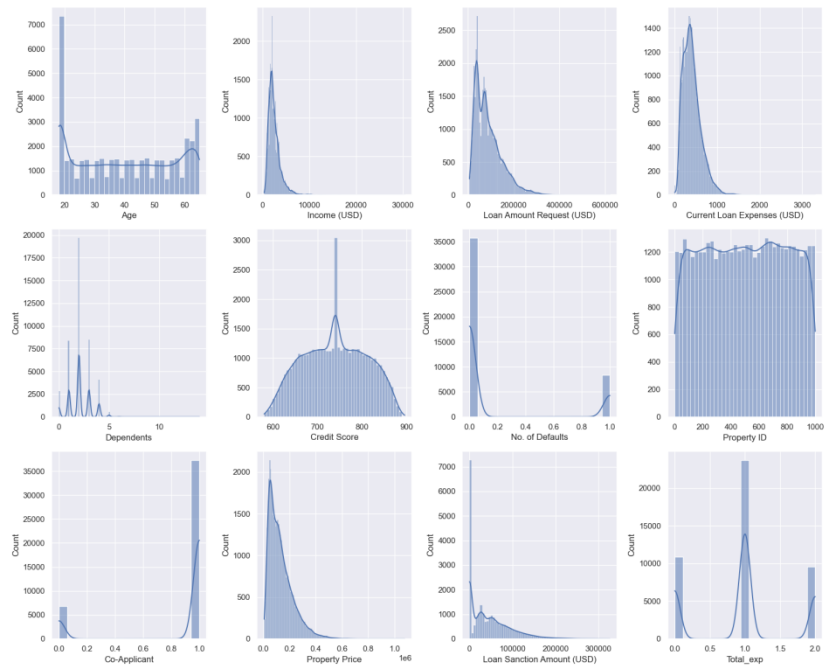
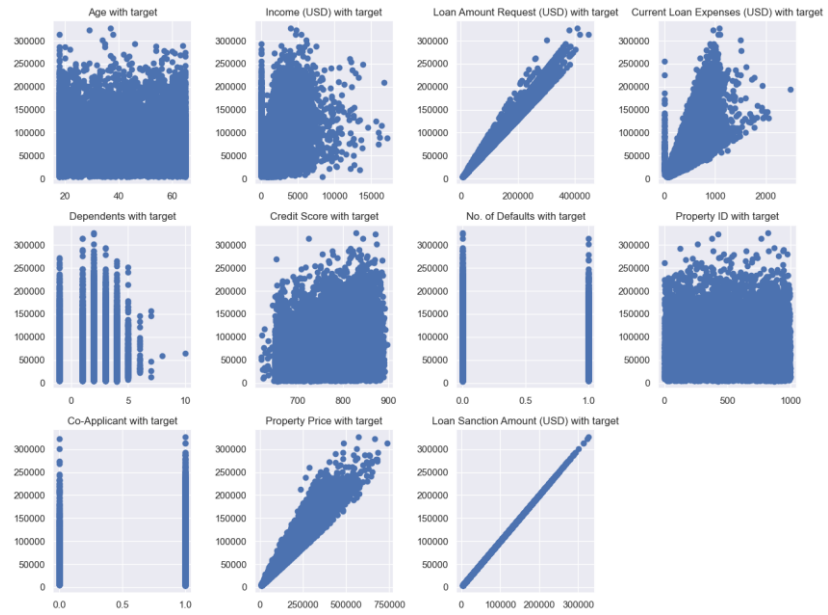
Data Exploration and Preprocessing Template

Data Exploration and Preprocessing Template

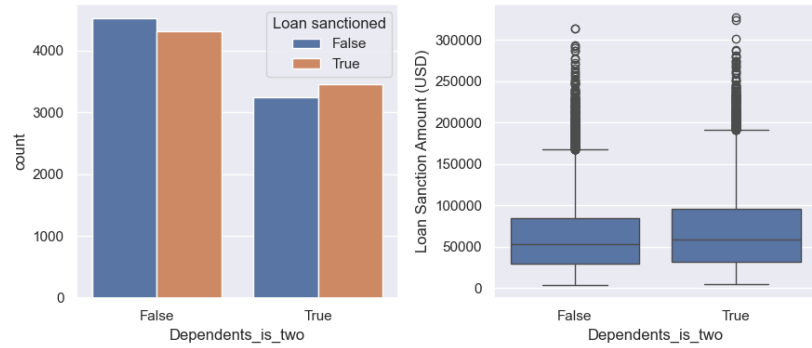
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

Section	Description
Data Overview	<p>Dimension: 8 rows x 131 columns</p> <p>Descriptive statistics:</p> 

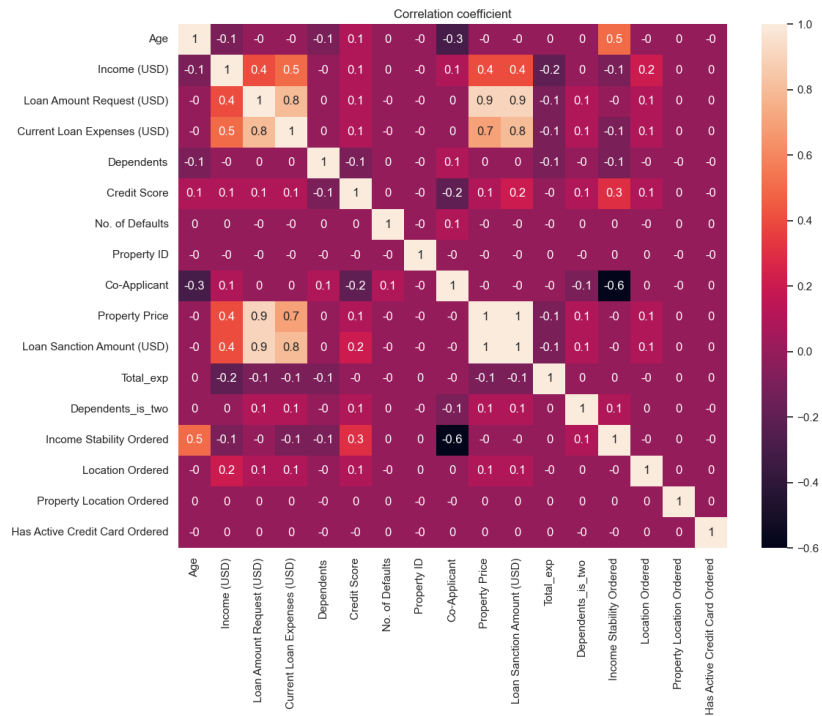
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

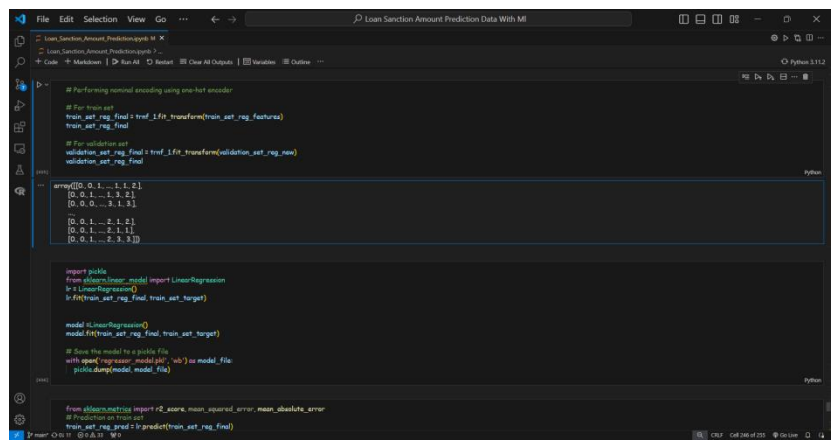
The screenshot displays a Jupyter Notebook interface. The top part shows a data table with 17 columns: Customer ID, Name, Gender, Age, Income (USD), Income Stability, Profession, Type of Employment, Location, Loan Amount Request (USD), Current Loan Exposure (USD), Expense Type 1, Expense Type 2, Dependents, Credit Score, Has Active Credit Card, Property ID, Property Age, Property Type, Property Location, and Asset Collateral. The table contains 10 rows of data. Below the table, there is a plot titled 'Loan Amount Request (USD)' on the y-axis (0 to 100,000) and 'Income Stability' on the x-axis (Low, High). The plot shows data points colored by 'Income Stability' (Low: red, High: blue). A legend at the bottom indicates: 'Income Stability' (Low: red, High: blue), 'Loan Amount Request (USD)' (0 to 100,000), and 'Income Stability' (Low: red, High: blue).

[illegible]

The screenshot displays a Jupyter Notebook environment. The top navigation bar shows standard file operations (File, Edit, Selection, View, Go) and a title bar indicating the project is 'Loan Sanction Amount Prediction Data With ML'. The notebook interface includes a left sidebar with icons for file explorer, search, and other tools. The main workspace contains two code cells. The first cell defines a dictionary named 'features' with 15 keys representing various property and financial attributes. The second cell defines a dictionary named 'target' with a single key 'Loan Sanction Amount'. The code is as follows:

```
features = {
    "Type of Employment": 1000,
    "Property Area": 1000,
    "Property Type": 1000,
    "Current Monthly Income": 1000,
    "Current Monthly Expense": 1000,
    "Property Age": 1000,
    "Property Location": 1000,
    "Type of Property": 1000,
    "Number of Rooms": 1000,
    "Number of Bathrooms": 1000,
    "Number of Kitchens": 1000,
    "Number of Bedrooms": 1000,
    "Number of Living Areas": 1000,
    "Number of Dining Areas": 1000,
    "Number of Garages": 1000,
    "Loan Sanction Amount": 1000
}

target = {
    "Loan Sanction Amount": 1000
}
```

Data Transformation	 <pre> # Performing manual encoding using one-hot encoder # For train set train_set_reg_final = train_set_reg.transform(train_set_reg_features) train_set_reg_final # For validation set validation_set_reg_final = train_set_reg.transform(validation_set_reg_new) validation_set_reg_final array([[0, 0, 1, ..., 1, 1, 2], [0, 0, 1, ..., 1, 3, 2], [0, 0, 0, ..., 0, 1, 3], [0, 0, 1, ..., 0, 1, 2], [0, 0, 1, ..., 0, 1, 1], [0, 0, 1, ..., 0, 3, 3]]) import pickle from sklearn.linear_model import LinearRegression lr = LinearRegression() lr.fit(train_set_reg_final, train_set_target) model = LinearRegression() model.fit(train_set_reg_final, train_set_target) # Save the 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	-