Multivariate Analysis

It involves the observation and analysis of more than one statistical outcome variable at a time. In simple terms, it deals with the analysis of more than one variable to understand the effect and relationship between them.

Multivariate analysis is used for:

- Understanding Relationships: It helps in understanding the relationships between different variables in the dataset.
- Pattern Recognition: It identifies patterns and structures in data which are not apparent in univariate or bivariate analysis.
- Predictive Modeling: It is used in predictive modeling to predict the outcome of one variable based on multiple other variables.
- Dimensionality Reduction: Techniques like PCA (Principal Component Analysis) are used to reduce the dimensionality of the data while retaining as much variance as possible.

Features of Multivariate Analysis:

- Correlation Matrix: Shows the correlation coefficients between pairs of variables.
- Scatter Plot Matrix: A grid of scatter plots that show relationships between pairs of variables.
- Heatmap: Visual representation of the correlation matrix.
- Principal Component Analysis (PCA): Reduces the dimensionality of the data while retaining most of the variance.
- Pair Plot: Pairwise relationships in a dataset.

```
# install imp. libraries
%pip install pandas
%pip install matplotlib
%pip install seaborn
%pip install sklearn
```

```
Requirement already satisfied: pandas in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas) (2.0.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas) (2024.1)
Requirement already satisfied: six>=1.5 in
```

```
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: matplotlib in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (3.9.0)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (4.53.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (1.4.5)
Requirement already satisfied: numpy>=1.23 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (2.0.0)
Requirement already satisfied: packaging>=20.0 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (24.1)
Requirement already satisfied: pillow>=8 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib) (2.9.0)
Requirement already satisfied: six>=1.5 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: seaborn in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
seaborn) (2.0.0)
Requirement already satisfied: pandas>=1.2 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
seaborn) (2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
seaborn) (3.9.0)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
```

```
matplotlib!=3.6.1,>=3.4->seaborn) (1.2.1)
Requirement already satisfied: cycler>=0.10 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (4.53.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (24.1)
Requirement already satisfied: pillow>=8 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: six>=1.5 in
/Users/anushkajain/micromamba/lib/python3.12/site-packages (from
python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Collecting sklearn
  Downloading sklearn-0.0.post12.tar.gz (2.6 kB)
  Preparing metadata (setup.py) ... error: subprocess-exited-with-
error

  × python setup.py egg_info did not run successfully.
  │ exit code: 1
  ╰─> [15 lines of output]
      The 'sklearn' PyPI package is deprecated, use 'scikit-learn'
      rather than 'sklearn' for pip commands.

      Here is how to fix this error in the main use cases:
      - use 'pip install scikit-learn' rather than 'pip install
sklearn'
      - replace 'sklearn' by 'scikit-learn' in your pip requirements
files
        (requirements.txt, setup.py, setup.cfg, Pipfile, etc ...)
```

```
      - if the 'sklearn' package is used by one of your dependencies,
        it would be great if you take some time to track which package
uses
        'sklearn' instead of 'scikit-learn' and report it to their
issue tracker
      - as a last resort, set the environment variable
        SKLEARN_ALLOW_DEPRECATED_SKLEARN_PACKAGE_INSTALL=True to avoid
this error

      More information is available at
      https://github.com/scikit-learn/sklearn-pypi-package
      [end of output]

  note: This error originates from a subprocess, and is likely not a
problem with pip.
error: metadata-generation-failed

× Encountered error while generating package metadata.
╰─> See above for output.

note: This is an issue with the package mentioned above, not pip.
hint: See above for details.
Note: you may need to restart the kernel to use updated packages.
```

```python
#importing libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

# Load the data
file_path = 'processed_loan_data.csv'
data = pd.read_csv(file_path)

# Display the first few rows of the data
print("First few rows of the dataset:")
print(data.head())
```

```
First few rows of the dataset:
  Customer ID              Name Gender  Age  Income (USD) Income
Stability  \
0     C-36995  Frederica Shealy      F   56       1933.05
Low
1     C-23855  Nathalie Olivier      M   43       2361.56
Low
2     C-24944    Barbie Goetsch      M   18       1546.17
Low
3     C-40801      Laree Staton      M   18       2416.86
Low
4     C-30073        Brinda Vaz      F   48        777.25
```

```
Low

      Profession Type of Employment    Location  Loan Amount Request
(USD)  \
0        Working        Sales staff  Semi-Urban
72809.58
1        Working           Laborers  Semi-Urban
152561.34
2        Working           Laborers       Rural
42091.29
3  State servant         Core staff  Semi-Urban
25765.72
4        Working           Laborers  Semi-Urban
96080.60

   ...  Credit Score No. of Defaults Has Active Credit Card  Property
ID  \
0  ...        809.44               0                 Active
746
1  ...        637.29               0            Unpossessed
227
2  ...        613.24               0            Unpossessed
883
3  ...        652.41               0                 Active
325
4  ...        764.11               0                 Active
678

   Property Age  Property Type Property Location  Co-Applicant  \
0       1933.05              4             Rural             1
1       2361.56              1        Semi-Urban             1
2       1546.17              2             Urban             1
3       2416.86              2             Rural             1
4        777.25              1        Semi-Urban             1

   Property Price  Loan Sanction Amount (USD)
0       119933.46                    54607.18
1       221050.80                        0.00
2        67993.43                        0.00
3        32423.71                    16747.72
4       146073.26                    67256.42

[5 rows x 24 columns]

#  Select numerical columns
numerical_columns = data.select_dtypes(include=['float64',
'int64']).columns

# Correlation Matrix
correlation_matrix = data[numerical_columns].corr()
```

```
print("\nCorrelation Matrix:")
print(correlation_matrix)
```

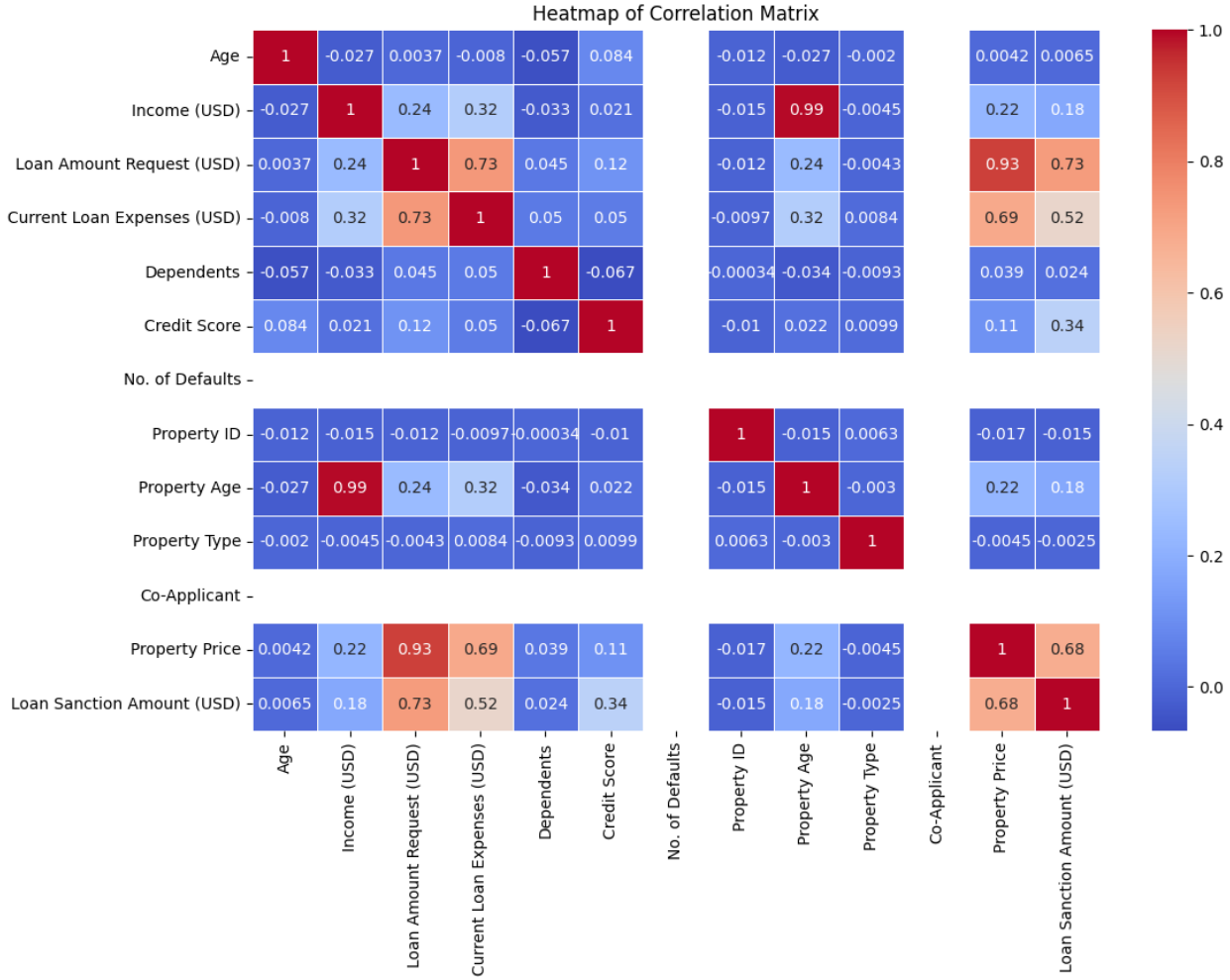
```
Correlation Matrix:
                                    Age  Income (USD)  \
Age                            1.000000     -0.026527
Income (USD)                  -0.026527      1.000000
Loan Amount Request (USD)      0.003691      0.240544
Current Loan Expenses (USD) -0.007977      0.319590
Dependents                    -0.056981     -0.033185
Credit Score                   0.084496      0.021051
No. of Defaults                     NaN           NaN
Property ID                   -0.012239     -0.014644
Property Age                  -0.026711      0.993999
Property Type                 -0.001978     -0.004515
Co-Applicant                        NaN           NaN
Property Price                 0.004209      0.224053
Loan Sanction Amount (USD)     0.006529      0.176932

                               Loan Amount Request (USD)  \
Age                                             0.003691
Income (USD)                                    0.240544
Loan Amount Request (USD)                       1.000000
Current Loan Expenses (USD)                     0.734165
Dependents                                      0.044845
Credit Score                                    0.115891
No. of Defaults                                      NaN
Property ID                                    -0.012065
Property Age                                    0.238503
Property Type                                  -0.004314
Co-Applicant                                         NaN
Property Price                                  0.927475
Loan Sanction Amount (USD)                      0.728908

                               Current Loan Expenses (USD)
Dependents  \
Age                                              -0.007977   -0.056981

Income (USD)                                      0.319590   -0.033185

Loan Amount Request (USD)                         0.734165    0.044845

Current Loan Expenses (USD)                       1.000000    0.049681

Dependents                                        0.049681    1.000000

Credit Score                                      0.050066   -0.066986

No. of Defaults                                        NaN         NaN
```

```
Property ID                                -0.009692   -0.000338

Property Age                                0.317149   -0.034174

Property Type                               0.008420   -0.009304

Co-Applicant                                     NaN         NaN

Property Price                              0.689104    0.038639

Loan Sanction Amount (USD)                  0.518257    0.024086


                            Credit Score  No. of Defaults  Property
ID  \
Age                             0.084496              NaN    -
0.012239
Income (USD)                    0.021051              NaN    -
0.014644
Loan Amount Request (USD)       0.115891              NaN    -
0.012065
Current Loan Expenses (USD)     0.050066              NaN    -
0.009692
Dependents                     -0.066986              NaN    -
0.000338
Credit Score                    1.000000              NaN    -
0.010432
No. of Defaults                      NaN              NaN
NaN
Property ID                    -0.010432              NaN
1.000000
Property Age                    0.021820              NaN    -
0.015065
Property Type                   0.009948              NaN
0.006282
Co-Applicant                         NaN              NaN
NaN
Property Price                  0.109297              NaN    -
0.016678
Loan Sanction Amount (USD)      0.341492              NaN    -
0.014856

                            Property Age  Property Type  Co-Applicant
\
Age                            -0.026711      -0.001978           NaN

Income (USD)                    0.993999      -0.004515           NaN

Loan Amount Request (USD)       0.238503      -0.004314           NaN
```

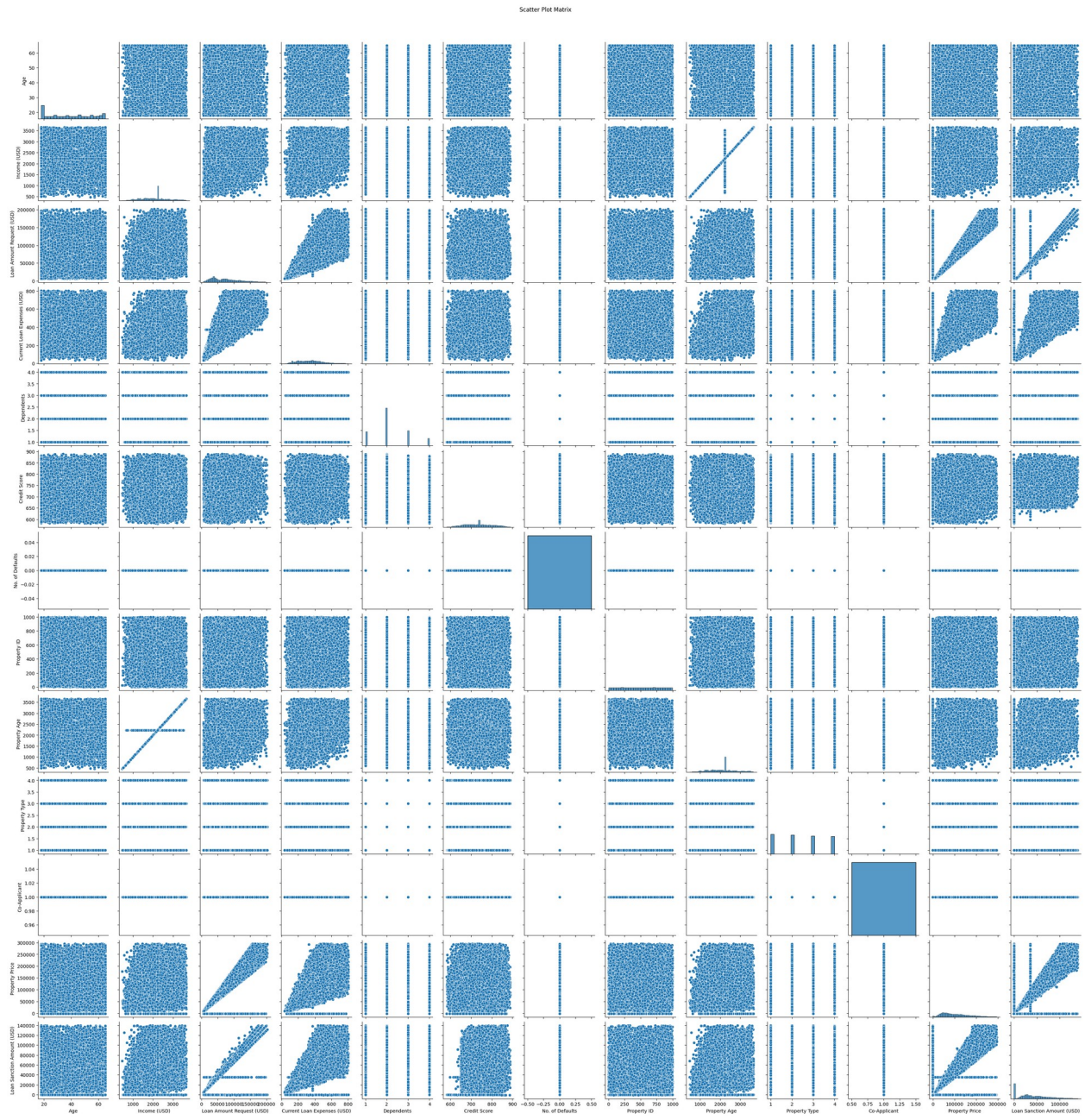
Current Loan Expenses (USD)	0.317149	0.008420	NaN
Dependents	-0.034174	-0.009304	NaN
Credit Score	0.021820	0.009948	NaN
No. of Defaults	NaN	NaN	NaN
Property ID	-0.015065	0.006282	NaN
Property Age	1.000000	-0.003009	NaN
Property Type	-0.003009	1.000000	NaN
Co-Applicant	NaN	NaN	NaN
Property Price	0.221754	-0.004485	NaN
Loan Sanction Amount (USD)	0.176245	-0.002520	NaN

	Property Price	Loan Sanction Amount (USD)
Age	0.004209	0.006529
Income (USD)	0.224053	0.176932
Loan Amount Request (USD)	0.927475	0.728908
Current Loan Expenses (USD)	0.689104	0.518257
Dependents	0.038639	0.024086
Credit Score	0.109297	0.341492
No. of Defaults	NaN	NaN
Property ID	-0.016678	-0.014856
Property Age	0.221754	0.176245
Property Type	-0.004485	-0.002520
Co-Applicant	NaN	NaN
Property Price	1.000000	0.675525
Loan Sanction Amount (USD)	0.675525	1.000000

```
# Heatmap of the Correlation Matrix
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title('Heatmap of Correlation Matrix')
plt.show()
```




```
# Scatter Plot Matrix (Pair Plot)
sns.pairplot(data[numerical_columns])
plt.suptitle('Scatter Plot Matrix', y=1.02)
plt.show()
```




```
# Principal Component Analysis (PCA)
pca = PCA(n_components=2)  # Reduce to 2 dimensions for visualization
pca_result = pca.fit_transform(data[numerical_columns].dropna())
pca_df = pd.DataFrame(data=pca_result, columns=['Principal Component
1', 'Principal Component 2'])

# Scatter Plot of PCA Result
plt.figure(figsize=(10, 7))
plt.scatter(pca_df['Principal Component 1'], pca_df['Principal
Component 2'], alpha=0.5)
plt.title('PCA Result')
```

```
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```