1. Sector_score LOCATION_ID PARA_A SCORE_A PARA_B SCORE_B ... District Loss LOSS_SCORE History Score Risk

count 776.000000 776.000000 776.000000 776.000000 776.000000 776.000000 ... 776.000000 776.000000 776.000000 776.000000 776.000000 776.000000

mean 20.184536 272.543814 131.313326 3.512887 139.579718 3.131443 ... 2.505155 0.029639 2.061856 0.104381 2.702577 0.626289

std 24.319017 5072.659778 3589.671373 1.740549 3589.718604 1.698042 ... 1.228678 0.184280 0.375080 0.531031 0.858923 0.484100

min 1.850000 1.000000 0.000000 2.000000 0.000000 2.000000 ... 2.000000 0.000000 2.000000 0.000000 2.000000 0.000000

25% 2.370000 8.000000 0.210000 2.000000 0.000000 2.000000 ... 2.000000 0.000000 2.000000 0.000000 2.000000 0.000000

50% 3.890000 13.000000 0.875000 2.000000 0.405000 2.000000 ... 2.000000 0.000000 2.000000 0.000000 2.400000 1.000000

75% 55.570000 19.000000 2.482500 6.000000 4.160000 4.000000 ... 2.000000 0.000000 2.000000 0.000000 3.250000 1.000000

max 59.850000 99999.000000 99999.000000 6.000000 99999.000000 6.000000 ... 6.000000 2.000000 6.000000 9.000000 5.200000 1.000000

[8 rows x 17 columns]

2.

Row: 33, Column: 1

Row: 88, Column: 4

Row: 142, Column: 9

Row: 147, Column: 1

Row: 147, Column: 2

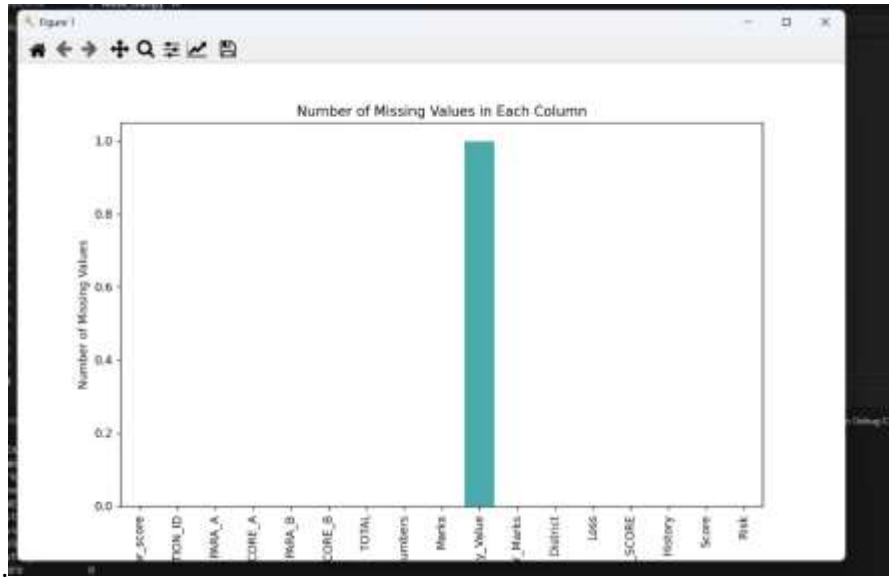
Row: 765, Column: 7

Row: 773, Column: 7

Row: 774, Column: 8

3.

3.

3. 

Row: 33, Column: 1

Row: 88, Column: 4

Row: 142, Column: 9

Row: 147, Column: 1

Row: 147, Column: 2

Row: 765, Column: 7

Row: 773, Column: 7

Row: 774, Column: 8

PS C:\Users\tebib\OneDrive\Desktop\J2\Topics in intelligent systems\Midterm> c:; cd 'c:\Users\tebib\OneDrive\Desktop\J2\Topics in intelligent systems\Midterm'; & 'C:\Users\tebib\j2anaconda\python.exe' 'c:\Users\tebib\.vscode\extensions\ms-python.python-2023.22.1\pythonFiles\lib\python\debugpy\adapter/../..\debugpy\launcher' '51648' '--' 'c:\Users\tebib\OneDrive\Desktop\J2\Topics in intelligent systems\Midterm\Audit_trial.py'

Missing values in each column:

Sector_score 0

LOCATION_ID 0

PARA_A 0

SCORE_A 0

PARA_B 0

SCORE_B 0

TOTAL 0

numbers 0

Marks 0

Money_Value 1

MONEY_Marks 0

District 0

Loss 0

LOSS_SCORE 0

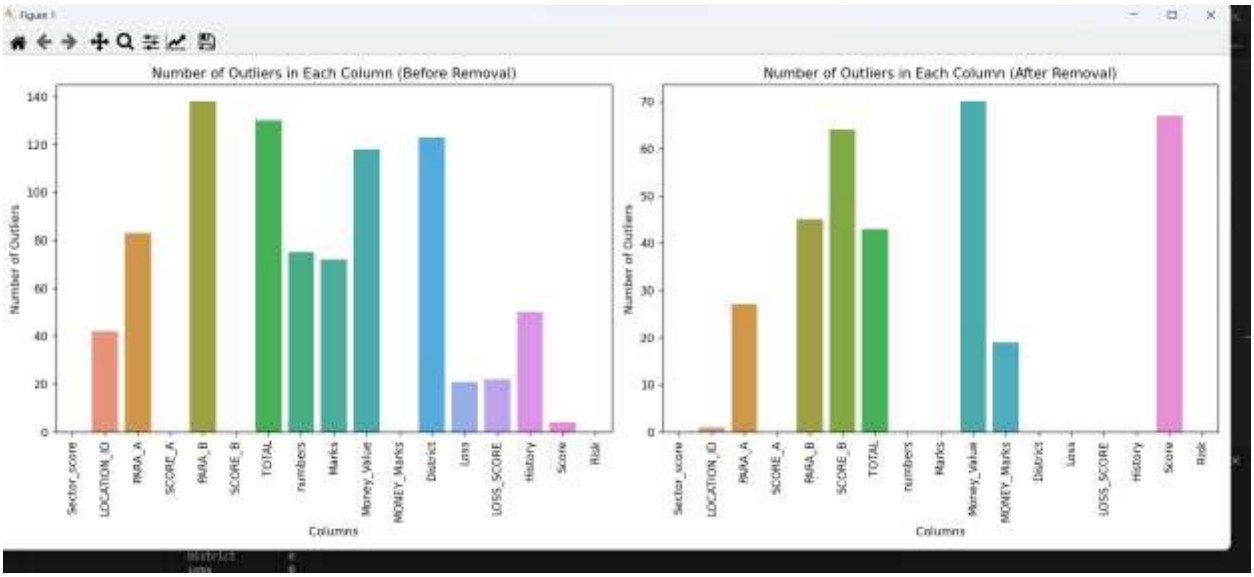
History 0

Score 0

Risk 0

dtype: int64

4.



Outliers in each column after removal:

Sector_score 0

LOCATION_ID 1

PARA_A 27

SCORE_A 0

PARA_B 45

SCORE_B 64

TOTAL 43

numbers 0

Marks 0

Money_Value 70

MONEY_Marks 19

District 0

Loss 0

LOSS_SCORE 0

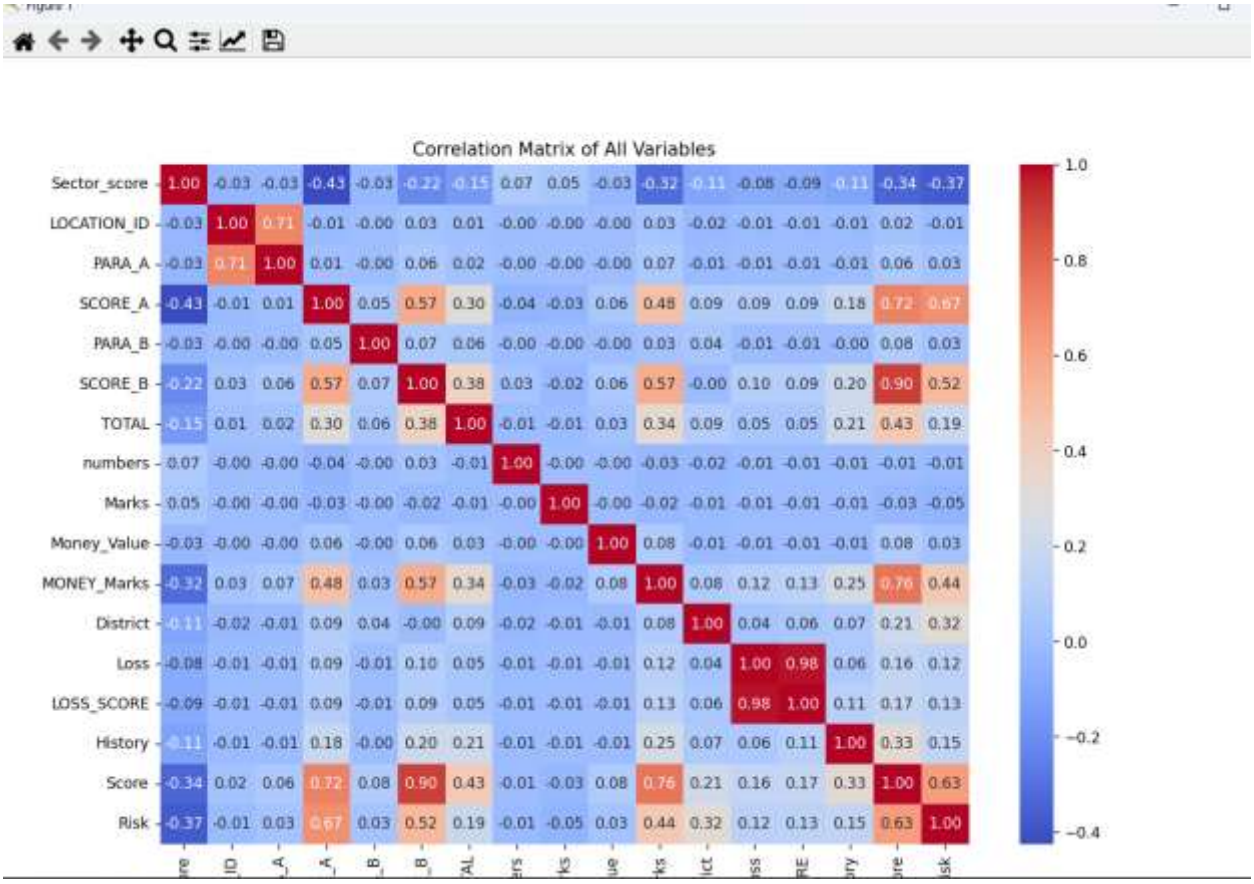
History 0

Score 67

Risk 0

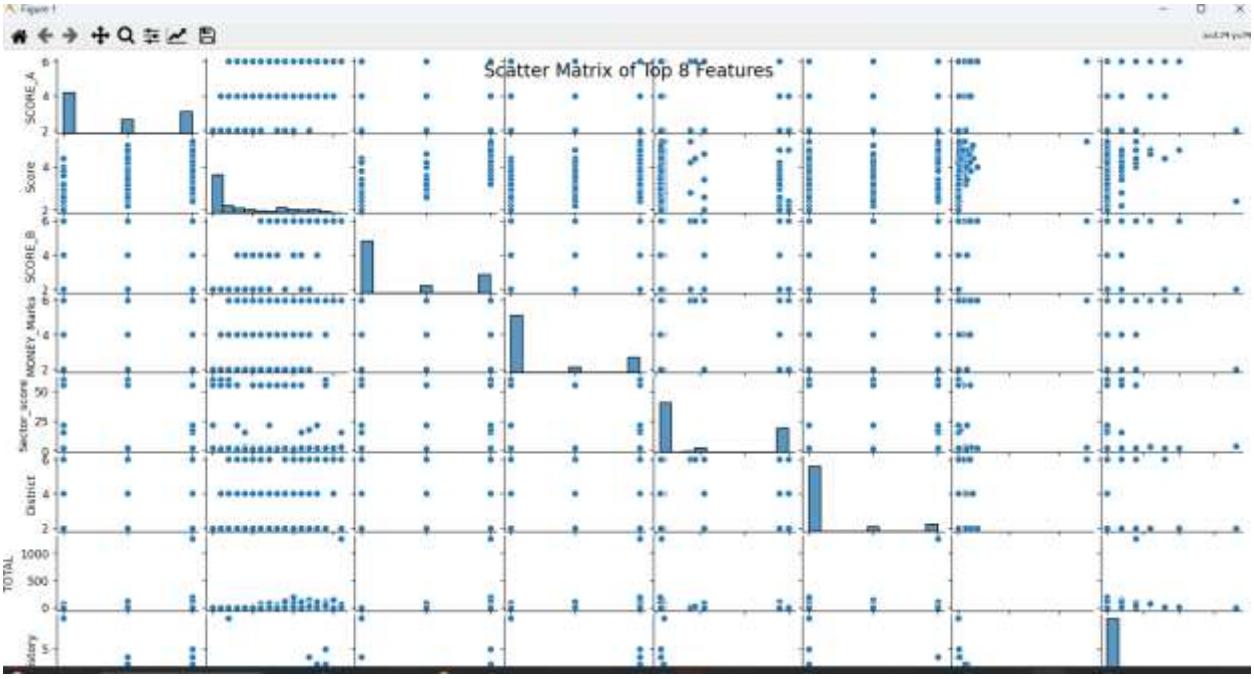
dtype: int64

5.



1. Calculate the correlation matrix using Pandas.
2. Use Seaborn or Matplotlib to visualize the correlation matrix.
3. Rank the feature columns based on their correlation values with the target variable

6.



The figure layout has changed to tight

self._figure.tight_layout(*args, **kwargs)

7.

	SVM Accuracies 10-fold cross validation mean average	Standard deviation of accuracies Change (increase or Decrease) in comparison to the Initial accuracy	Change (increase or Decrease) in comparison to the Initial accuracy
Cleaned Audit_Dataset (removing rows)	0.9429403929403929	0.08967486434797127	This is Initial accuracy
Cleaned Audit_Dataset (replacing missing values)	0.6262737262737262	0.0023654746100803268	Decrease of 33.57%
Normalized Audit_Dataset	Decrease of 33.57%	0.01189563000157557	Increase of 4.93%
Reduced feature Dataset to only half i.e 8	0.9922244422244422	0.01038140221113776	

attributes, pick your best to show highest accuracy.			Increase of 0.049283649284049 3 or 4.93%
Apply PCA on the Dataset and use only 4 features	0.9429403929403929	0.08967486434797127	N.A

8.

PCA Explained Variance Ratio: [0.23000898 0.11934669 0.10633977 0.06756796 0.0645947 0.06280755

0.062595 0.05989377 0.05721948 0.05078734 0.04497741 0.03121728

0.0222949 0.01822434 0.0012181 0.00090673]

9.

Confusion Matrix:

[[65 0]

[0 91]]

10.

Performance Metric	Score
Precision	1.0
Recall	1.0
False Alarm Rate	0.0
Kappa-Accuracy	1.0
Specificity	1.0
AUC score	1.0