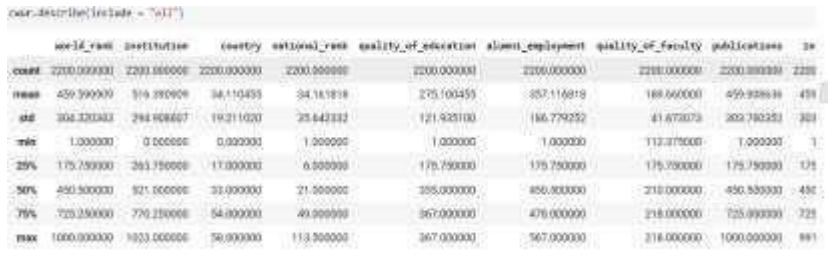
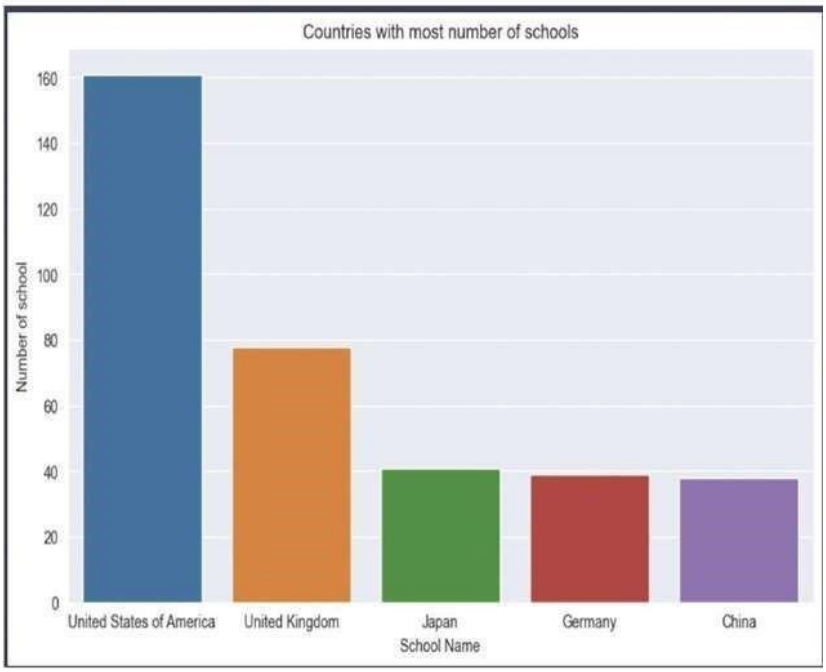
Data Collection and Preprocessing Phase

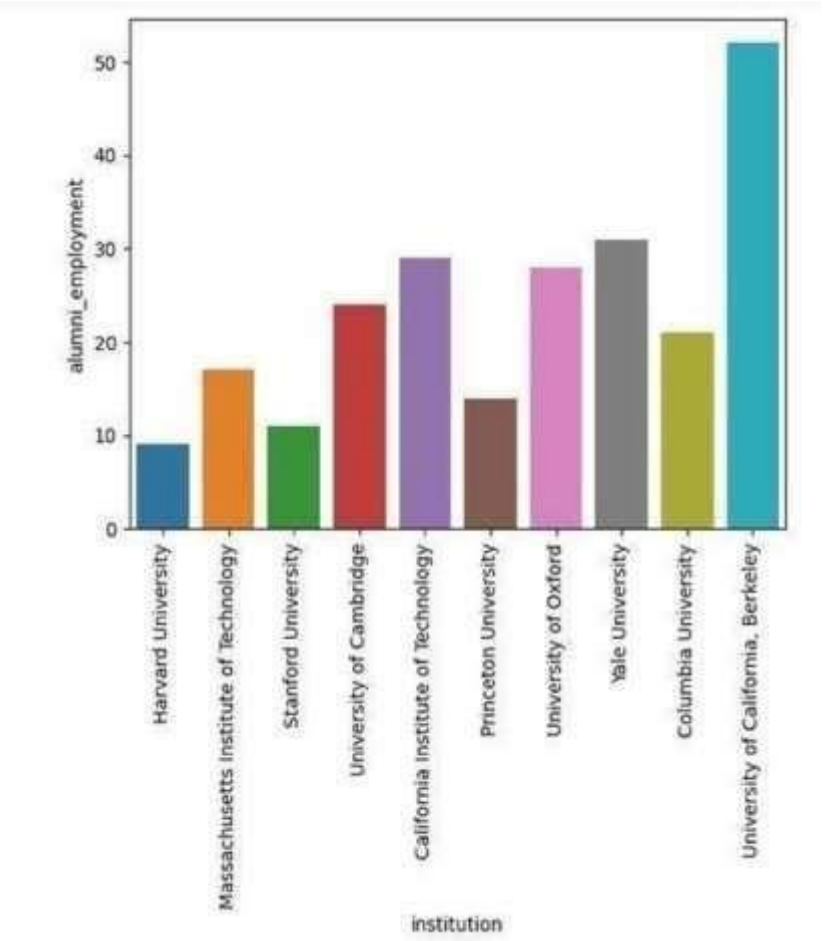
Date	06 July 2024
Team ID	739899
Project Title	SmartLender – Envisioning Success: Predicting University Scores With Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

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	

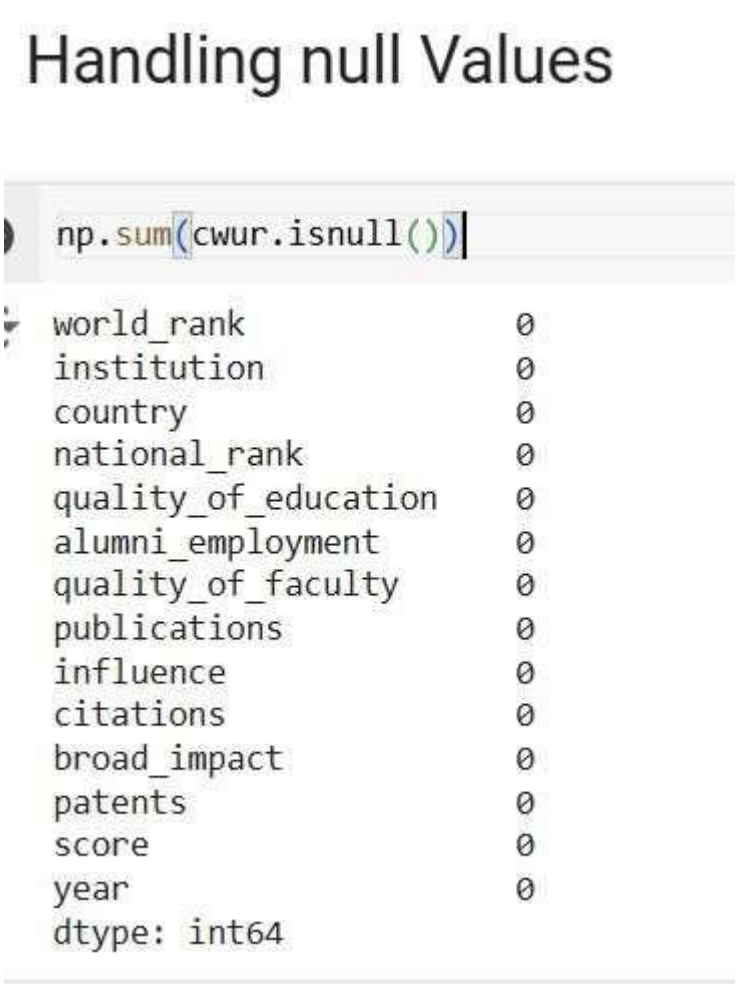
Univariate Analysis	*a.Univariate Analysis* 

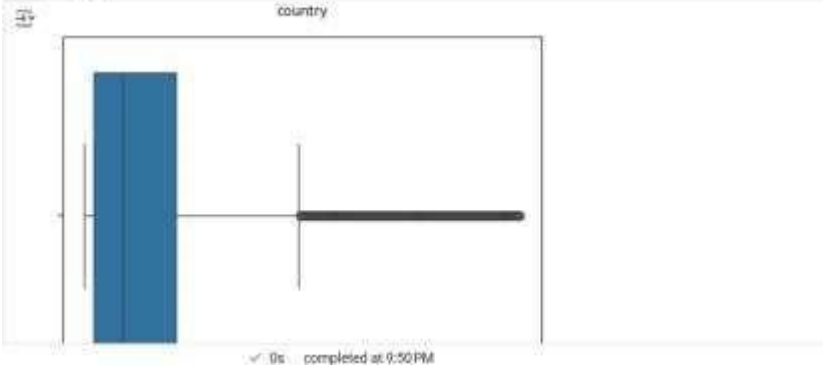
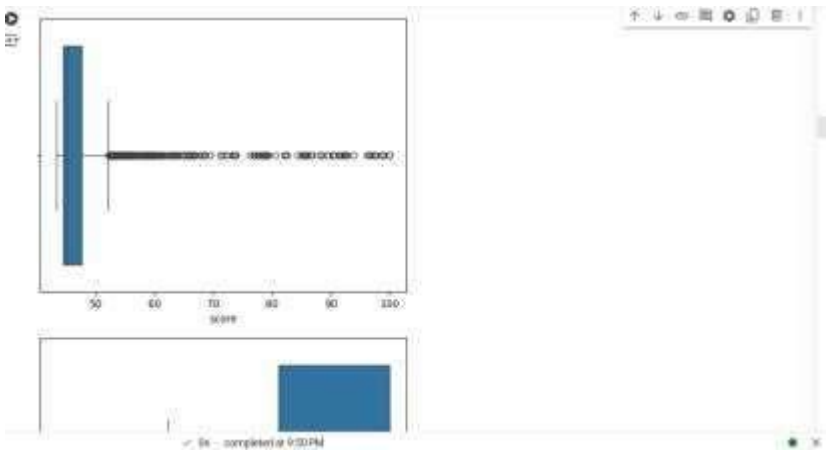
Bivariate Analysis	

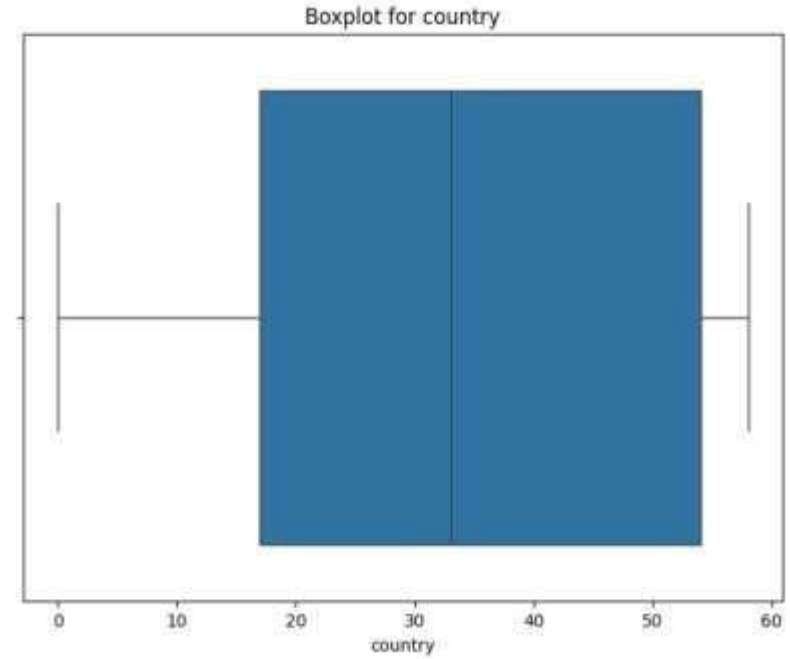
Multivariate Analysis	*c. Multivariate Analysis* 

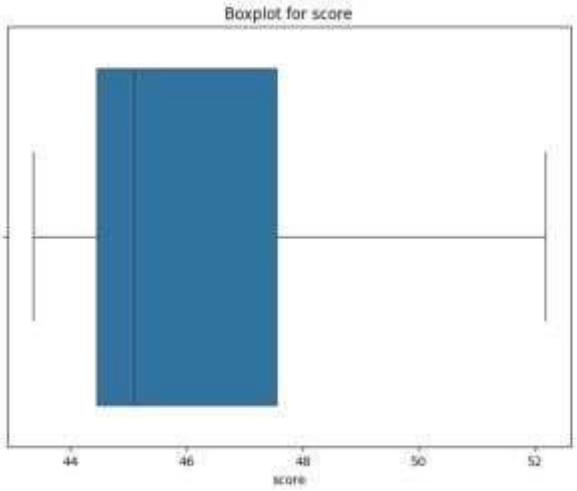
Data Preprocessing Code Screenshots

Loading Data	

Handling Null Values	

Viewing outliers	Handling outliers `def fun(col):` `  sns.boxplot(x=col,data=cwur)` `  plt.show()` `for i in cwur.columns:` `  fun(i)` 
	

Handling outliers	``` # Iterate over each column and plot boxplot for column in cwur.columns: plt.figure(figsize=(8, 6)) # Adjust the figure size as needed sns.boxplot(x=cwur[column]) plt.title(f'Boxplot for {column}') plt.xlabel(column) plt.show() ``` 



Saved Processed Data

```
cwur.shape
(2200, 14)
```