Team ID: PNT2022TMID11450

DATA COLLECTION

Data visualization

Data visualization is where a given dataset is presented in a graphical format. It helps the detection of patterns, trends and correlations that might go undetected in text-based data. Understanding your data and the relationship present within it is just as important as any algorithm used to train your machine learning model. Machine learning models will perform poorly on data that wasn't visualized and understood properly.

To visualize the dataset, we need libraries called **Matplotlib and Seaborn**. The **Seaborn and Matplotlib** libraries is a Python 2D plotting library that allows you to generate plots, scatter plots, histograms, bar charts etc.

There are three types variate analysis done in the given dataset by Seaborn library are

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

1.Univariate Analysis:

- The univariate analysis used to visualize the single variable of column data in the given dataset.
- Kinds of visualizing graph are
  - Histogram
  - Distplot
  - Boxplot

2.Bivariate Analysis:

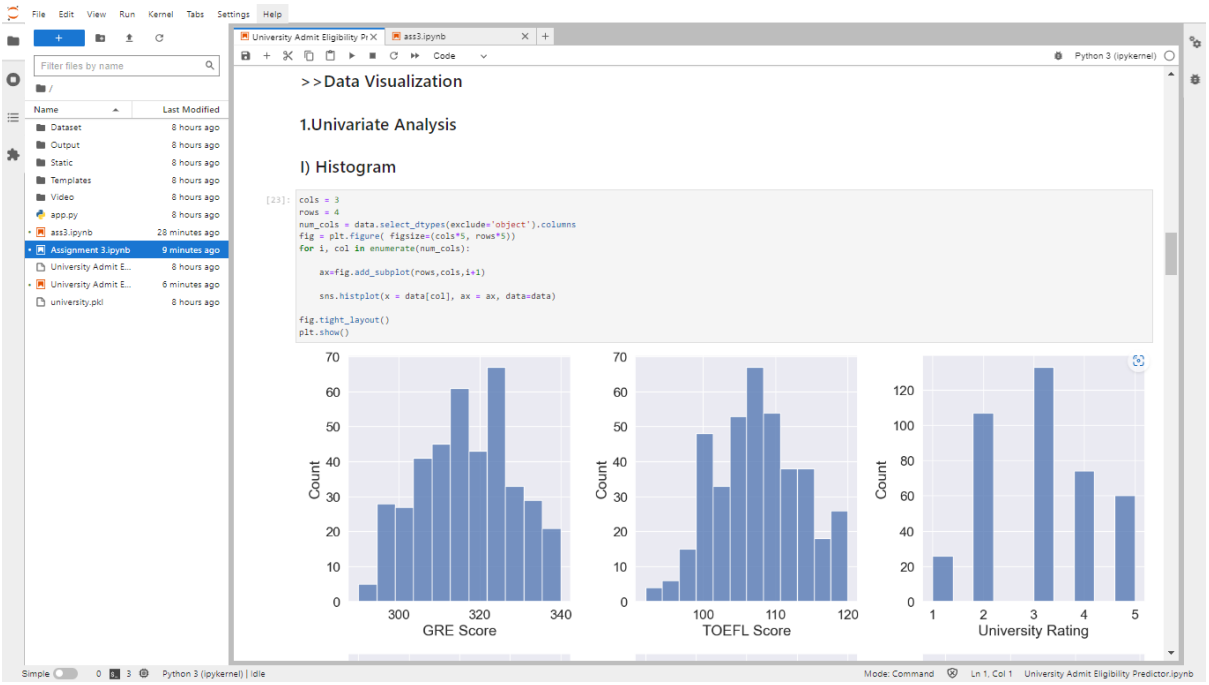
- The bivariate analysis used to visualize the double variables of column data in the given dataset.
- Kinds of visualizing graph are
  - Scatterplot
  - Catplot

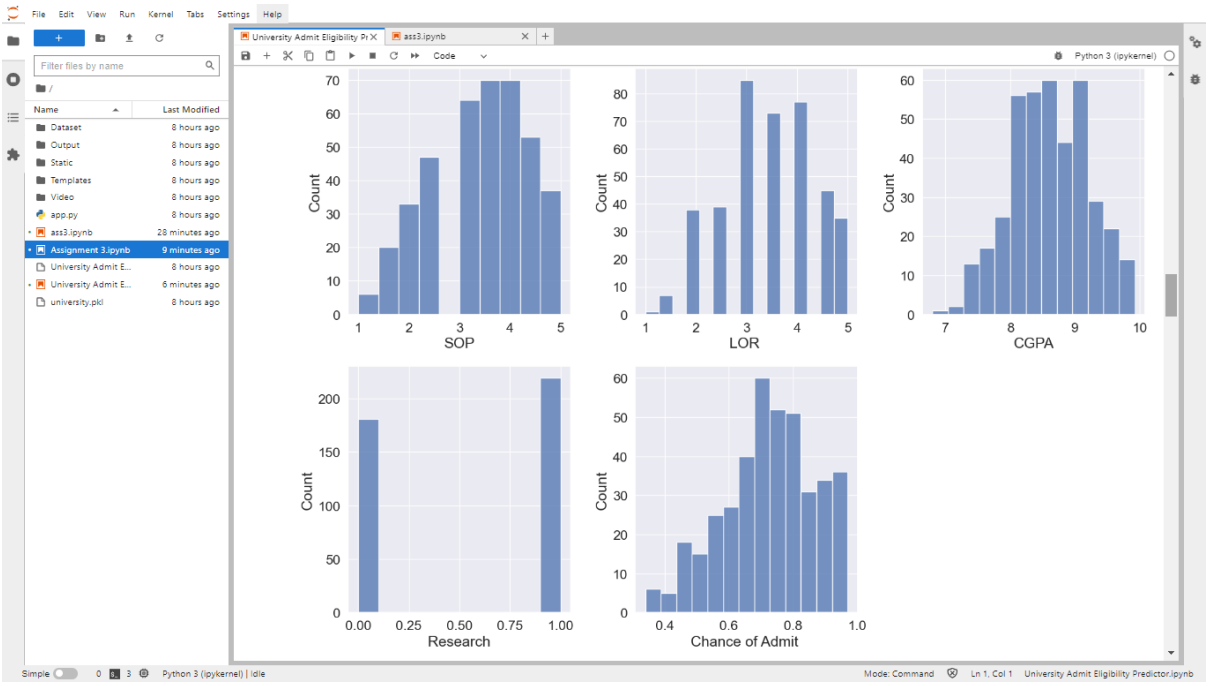
3.Multivariate Analysis:

- The multivariate analysis used to visualize the three or more variables of h data in the given dataset.
- Visualization of Graph is done by Pairplot.

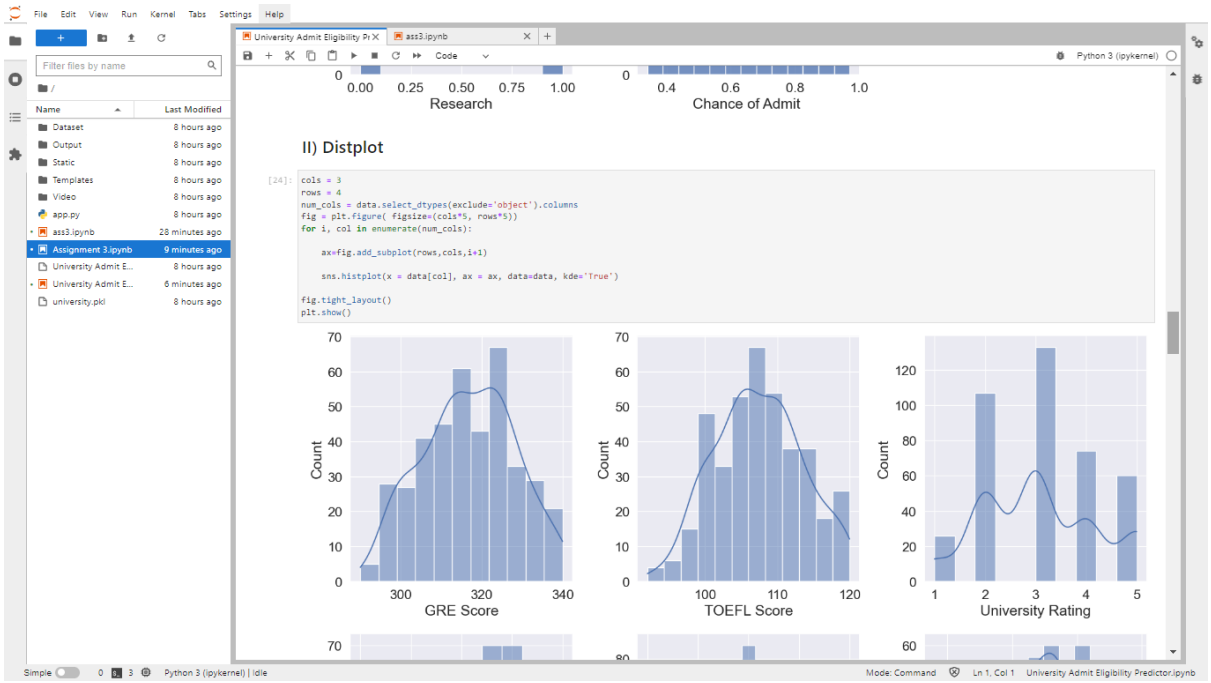
Step-1:

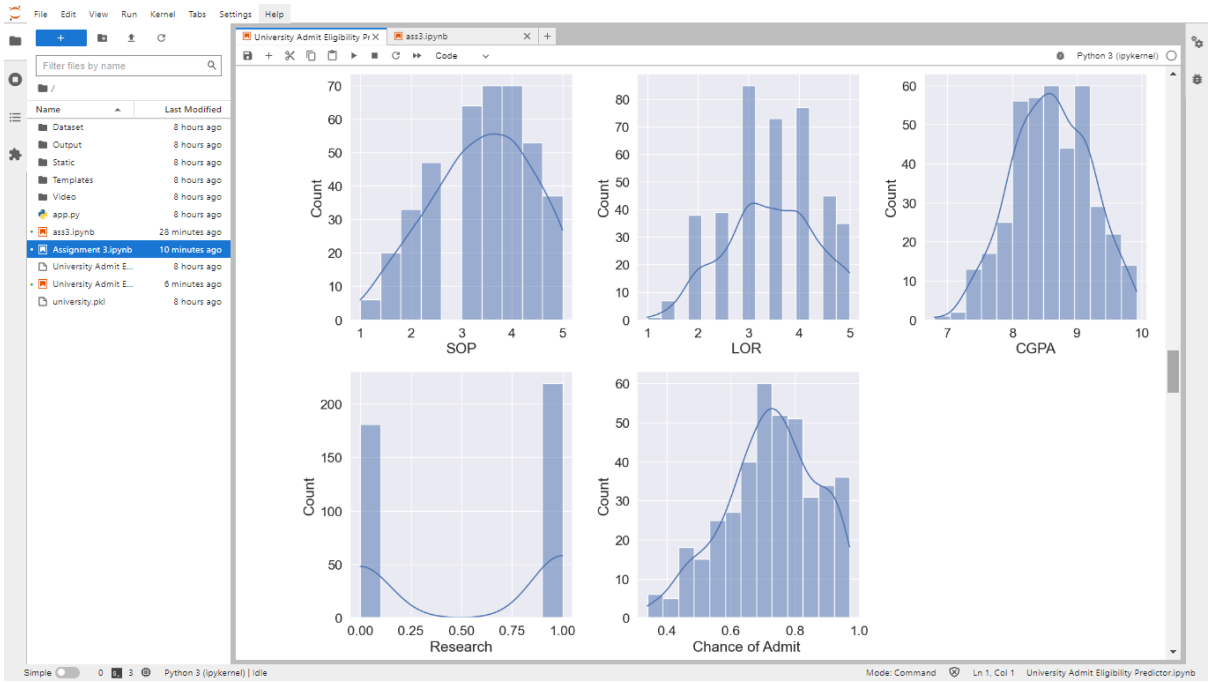
- Firstly, we need to perform the data visualization using Seaborn library.
- Then, Perform the Univariate analysis by Histogram, Distplot, Boxplot.
- **Histogram:**



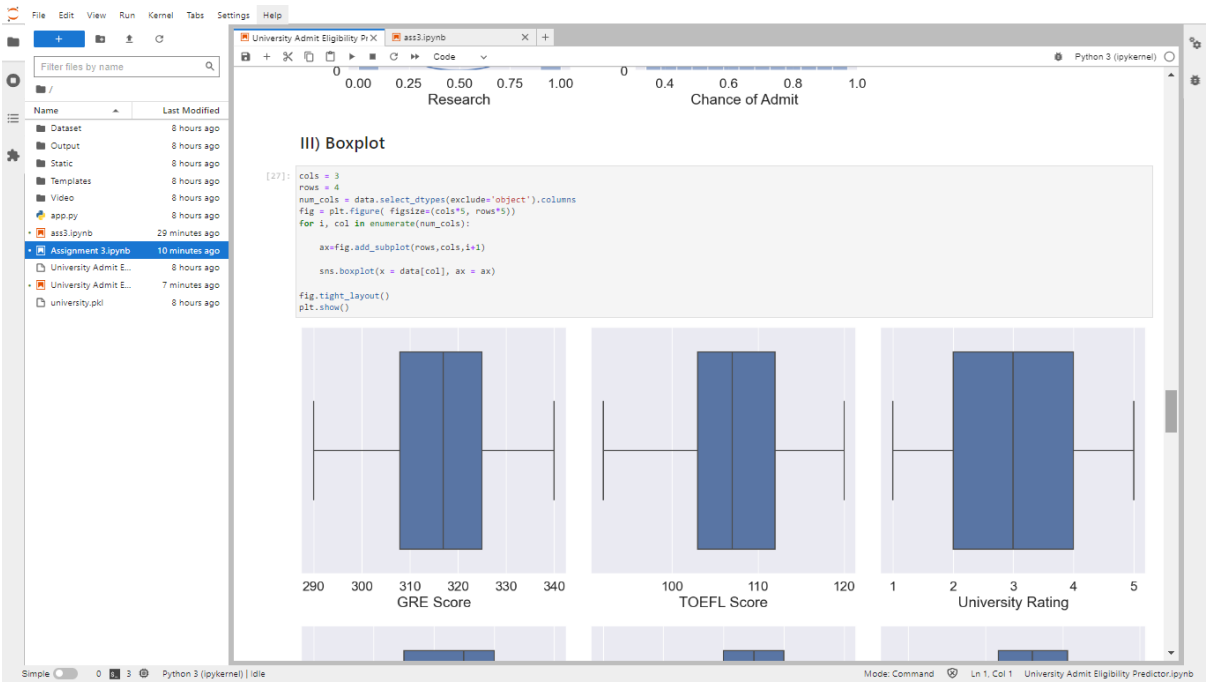


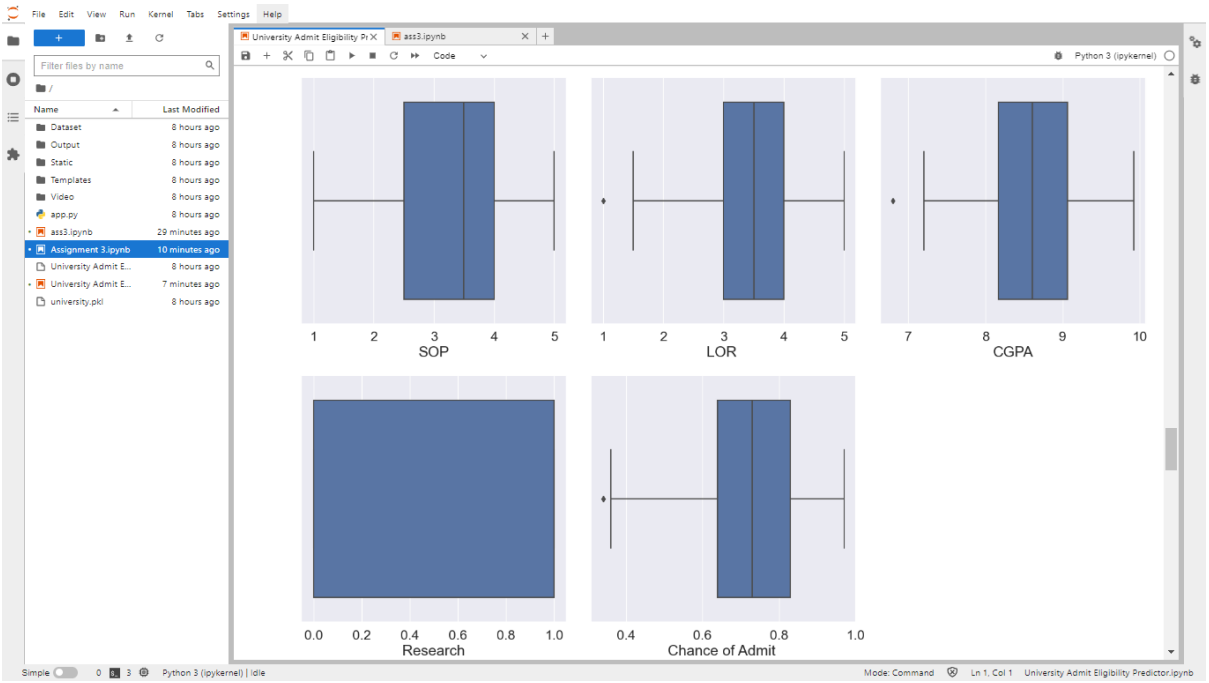
- **Distplot:**


Research
Chance of Admit
II) Distplot
[24]:
cols = 3
rows = 4
num_cols = data.select_dtypes(exclude='object').columns
fig = plt.figure(figsize=(cols*5, rows*5))
for i, col in enumerate(num_cols):
ax=fig.add_subplot(rows,cols,i+1)
sns.histplot(x = data[col], ax = ax, data=data, kde='True')
fig.tight_layout()
plt.show()
GRE Score
TOEFL Score
University Rating
Count


SOP
LOR
CGPA
Research
Chance of Admit
Count

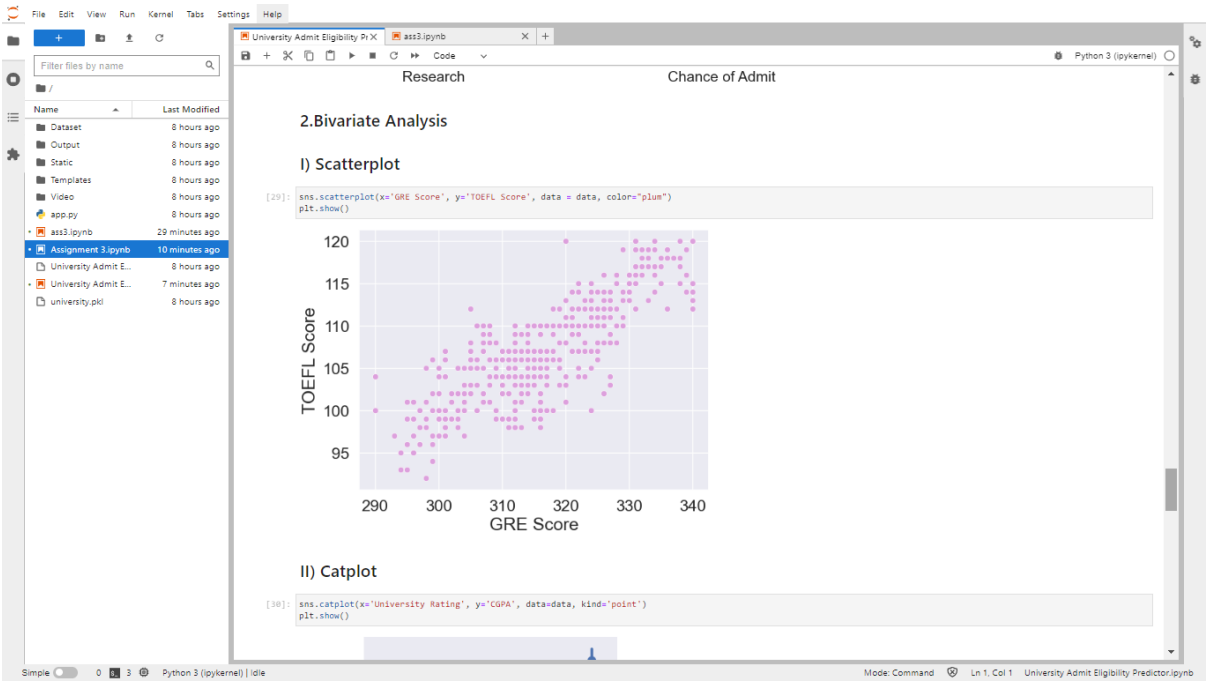
- **Boxplot:**



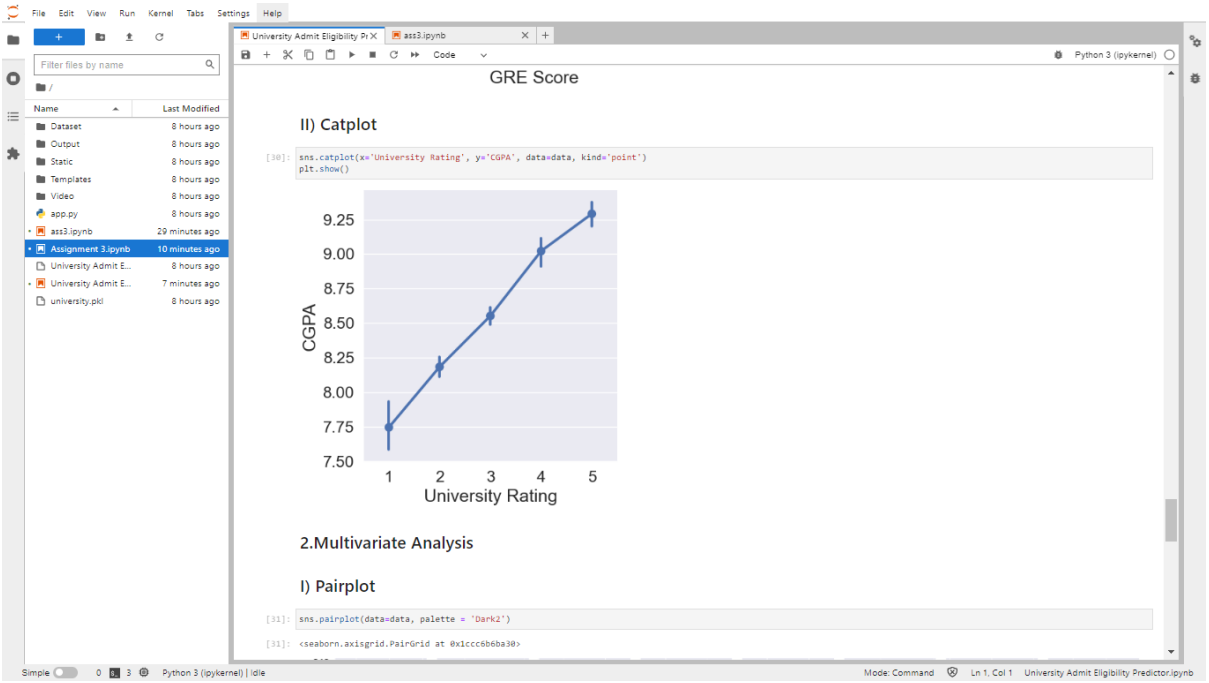


Step-2:

- Perform the Bivariate analysis by Scatterplot, Catplot.
- **Scatterplot:**


Research
Chance of Admit
2.Bivariate Analysis
I) Scatterplot
[29]: sns.scatterplot(x='GRE Score', y='TOEFL Score', data = data, color="plum")
plt.show()
TOEFL Score
GRE Score
II) Catplot
[30]: sns.catplot(x='University Rating', y='CGPA', data=data, kind='point')
plt.show()

- **Catplot:**


GRE Score
II) Catplot
[30]: sns.catplot(x='University Rating', y='CGPA', data=data, kind='point')
plt.show()
CGPA
University Rating
2.Multivariate Analysis
I) Pairplot
[31]: sns.pairplot(data=data, palette = 'Dark2')
[31]: <seaborn.axisgrid.PairGrid at 0x1ccc6b6ba30>

Step-3:

- Perform the Multivariate analysis by Pairplot.
- **Pairplot:**