Team ID: PNT2022TMID11450

DATA COLLECTION

Analyze the Dataset

Pandas Head () function:

- **Head () function** used to view the first 5 csv file data.
- By using head () function, we can overview the **number of columns and number of rows** in the dataset.
- Then we can allocate the **size** to view number of columns and rows in **head () function** by **head (10).**
- **Syntax for Head () function:**

  data.head()

Pandas Drop () function:

- **Drop () function** is used to delete the column data in the dataset.
- unwanted column data can be deleted by this command.

Pandas Describe () function:

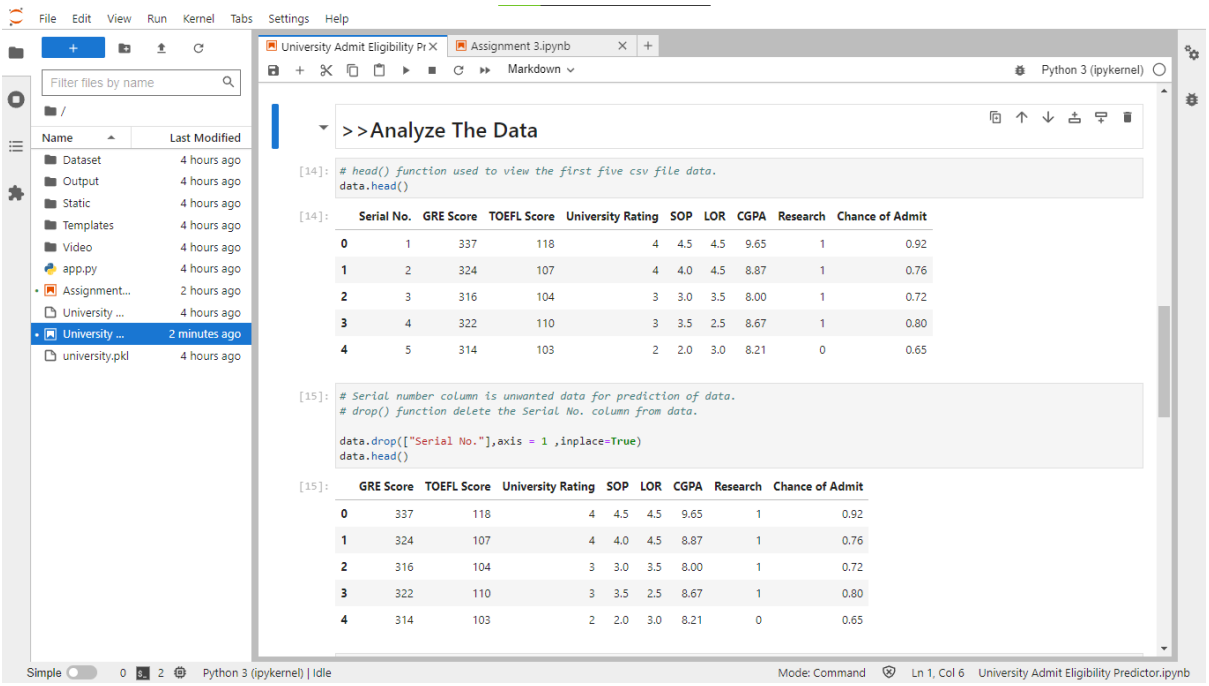
- **Describe () function** computes a summary of statistics like count, mean, standard deviation, min, max, and quartile values.

Pandas Info () function:

- **Info () function** gives information about the dataset like how many Objects, Integer, Float values.
- It also gives the memory size and number of columns.

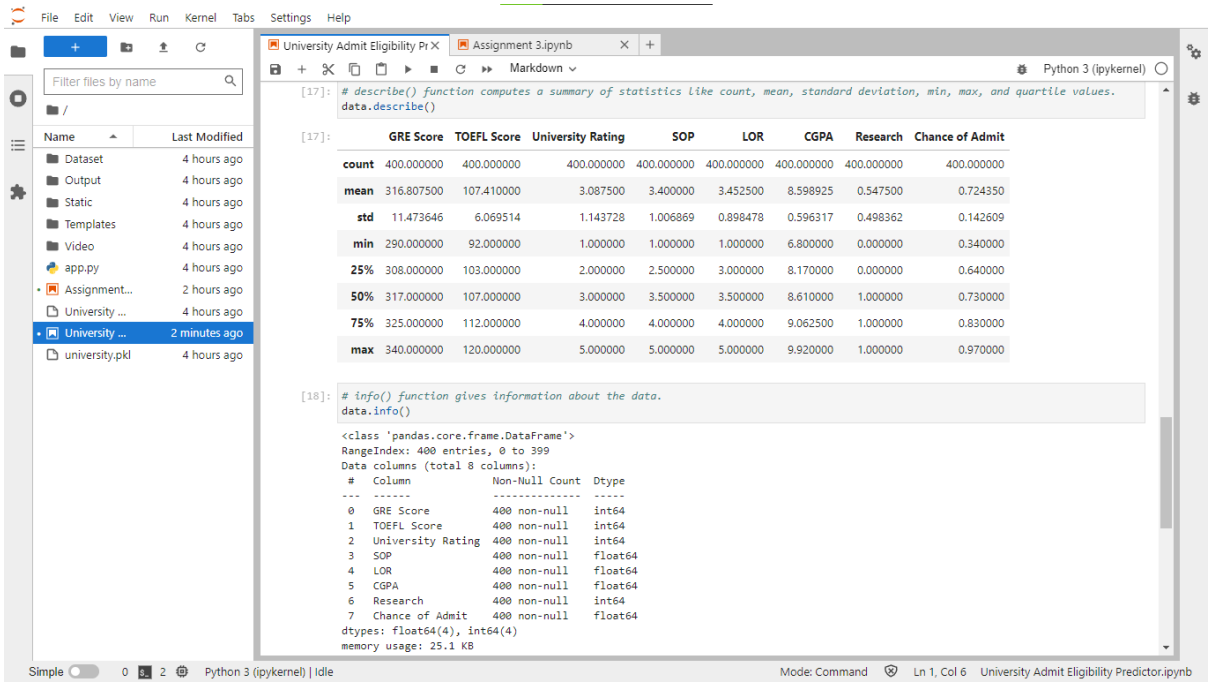
Step-1:

- Firstly, Use the **head () function** to view the first 5 rows of the dataset and
- There has unwanted data named as **"Serial No."** is not required for Prediction of data.
- Use **drop () function** to delete the **"Serial No."** column in the data.



Step-2:

- After deleting the **"Serial No."** column, view the data using **head () function**.
- Use the **describe () function** to view summary of statistics like count, mean, standard deviation, min, max, and quartile values.
- Use the **info () function** to get the information about the dataset like how many Objects, Integer, Float values.
- This function also gives the memory size and number of columns.


File Edit View Run Kernel Tabs Settings Help
Filter files by name
/
Name
Last Modified
Dataset 4 hours ago
Output 4 hours ago
Static 4 hours ago
Templates 4 hours ago
Video 4 hours ago
app.py 4 hours ago
Assignment... 2 hours ago
University ... 4 hours ago
University ... 2 minutes ago
university.pkl 4 hours ago
University Admit Eligibility Pr
Assignment 3.ipynb
Markdown
Python 3 (ipykernel)
[17]: # describe() function computes a summary of statistics like count, mean, standard deviation, min, max, and quartile values.
data.describe()
[17]:
GRE Score TOEFL Score University Rating SOP LOR CGPA Research Chance of Admit
count 400.000000 400.000000 400.000000 400.000000 400.000000 400.000000 400.000000 400.000000
mean 316.807500 107.410000 3.087500 3.400000 3.452500 8.598925 0.547500 0.724350
std 11.473646 6.069514 1.143728 1.006869 0.898478 0.596317 0.498362 0.142609
min 290.000000 92.000000 1.000000 1.000000 1.000000 6.800000 0.000000 0.340000
25% 308.000000 103.000000 2.000000 2.500000 3.000000 8.170000 0.000000 0.640000
50% 317.000000 107.000000 3.000000 3.500000 3.500000 8.610000 1.000000 0.730000
75% 325.000000 112.000000 4.000000 4.000000 4.000000 9.062500 1.000000 0.830000
max 340.000000 120.000000 5.000000 5.000000 5.000000 9.920000 1.000000 0.970000
[18]: # info() function gives information about the data.
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 8 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 GRE Score 400 non-null int64
1 TOEFL Score 400 non-null int64
2 University Rating 400 non-null int64
3 SOP 400 non-null float64
4 LOR 400 non-null float64
5 CGPA 400 non-null float64
6 Research 400 non-null int64
7 Chance of Admit 400 non-null float64
dtypes: float64(4), int64(4)
memory usage: 25.1 KB
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 6
University Admit Eligibility Predictor.ipynb