

# DATA COLLECTION

## Reading the Dataset

### Pandas read\_csv() function:

- Firstly, Import [Admission\\_Predict.csv](#) file in Dataset folder.
- Load a [Admission\\_Predict.csv](#) data file into panda's libraries using `read_csv()` function.
- Need to locate the directory of the CSV file at first.
- If dataset is in some other location, Use the below command in **Jupyter lab**.
  - `Data=pd.read_csv(r" File_location/filename.csv")`
- **Note:**
  - **r** stands for "**raw**" and will cause backslashes in the string to be interpreted as actual backslashes rather than special characters.
- Admission\_Predict Dataset contains following Columns:
  1. Serial No
  2. GRE Score
  3. TOEFL Score
  4. University Rating
  5. SOP
  6. LOR
  7. CGPA
  8. Chance of Admit

## Step-1:

- Import the **Pandas** libraries as object name **pd** and define variable name as **data** to store the dataset.
- If the dataset at different location, then use the command **r" < location of the dataset >"**
- Then run the code by give **data** variable to view the output of the dataset.

The screenshot displays a Jupyter Notebook environment. On the left, a file explorer sidebar shows a directory structure with files like 'Dataset', 'Output', 'Static', 'Templates', 'Video', 'app.py', 'Assignment...', 'University ...', 'University ...', and 'university.pkl'. The main notebook area has a title bar 'University Admit Eligibility Predictor' and a code editor with the following content:

```
>Reading the Dataset

[3]: # read_csv() reads the dataset from the csv file.
data = pd.read_csv(r"Dataset/Admission_Predict.csv")
data
```

The output of the code is a Pandas DataFrame with 400 rows and 9 columns. The columns are: Serial No., GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, Research, and Chance of Admit. The first few rows are displayed, followed by an ellipsis indicating the continuation of the data, and then the last few rows (395 to 399).

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...	...
395	396	324	110	3	3.5	3.5	9.04	1	0.82
396	397	325	107	3	3.0	3.5	9.11	1	0.84
397	398	330	116	4	5.0	4.5	9.45	1	0.91
398	399	312	103	3	3.5	4.0	8.78	0	0.67
399	400	333	117	4	5.0	4.0	9.66	1	0.95

Below the table, it states '400 rows x 9 columns'. At the bottom of the notebook, the status bar shows 'Simple', '0', '2', 'Python 3 (ipykernel) | Idle', 'Mode: Command', 'Ln 1, Col 1', and 'University Admit Eligibility Predictor.ipynb'.