Team ID: PNT2022TMID11450

DATA COLLECTION

Splitting Dependent and Independent Columns

- Splitting the dataset into the matrix of independent variables and the vector or dependent variable.
- Mathematically, Vector is defined as a matrix that has just one column.
- To read the columns, we will use **iloc () function** of pandas (used to fix the indexes for selection) which takes two parameters:
    - Row selection
    - Column selection

Syntax for iloc () function:

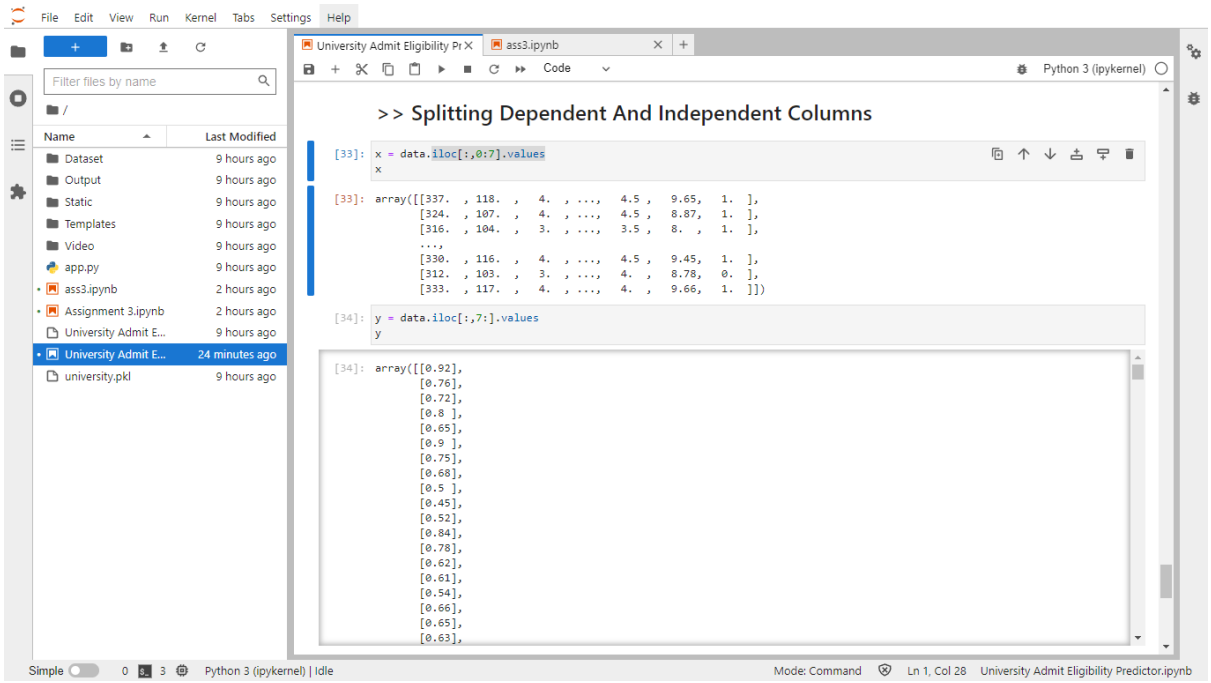
- **X = data.iloc[:,0:7].values :- ":"** indicates that you are considering all the rows in the dataset and **"0:7"** indicates that you are considering columns 0 to 7 as input values and assigning them to variable **x**.
- **Y = data.iloc[:,7:0].values :- ":"** indicates you are considering all the rows and **"7:"** indicates that you are considering only the last column as output value and assigning them to variable **y**.

Syntax for shape () function:

- **X.Shape() :-** 400 rows with 7 columns.
- **Y.Shape() :-** 400 rows with 1 columns.

Step-1:

- Firstly, we need to split the independent variables and dependent variables.
- In pandas, **iloc () function** is used to split data by full columns present in the dataset.
- Then, we are assigning **X** variable for full columns and rows and **Y** variable for 1 column and all rows.



Step-2:

- After splitting the data, we need to check the shape of the splitted variables **X** and **Y.**
- Then, it will show the number of columns and rows.