

25-10-2024

# Training Day – 27

October 25, Friday\*

## - \*Topic:\* Descriptive Analysis with Pandas

- Summarized data using `.describe()`, `.mean()`, and `.sum()`.
- Example: Analyzed a dataset's central tendencies and spread.

Today I made my own data set and convert it into DataFrame and apply some functions of **pandas library**. Like - `.info()`, `.set_index()`, `.reset_index()`, adding an column name 'Roll no.', `.drop()`.

- Also done 'Indexing' and 'Slicing' on large dataset for reading the data according to given rows and column detail, using `.iloc` function.

- Also practice for feeling NAN values by identifying outliers, with the help of mean, median, mode.

The screenshot shows a Jupyter Notebook interface with a dark theme. At the top, there are tabs for '+ Code' and '+ Text'. The first code cell contains `DF=DF.reset_index()` and `DF`. Below the code, a table view of the DataFrame is displayed with columns 'Name', 'age', and 'Marks'. The table has three rows with indices 0, 1, and 2. Below the table, there are three buttons: 'Generate code with DF', 'View recommended plots', and 'New Interactive sheet'. The second code cell contains `L=[1,2,3]` and `print(L)`. Below the code, the output `[1, 2, 3]` is shown. The third code cell contains `df['Roll no.']=L` and `df`. Below the code, a table view of the DataFrame is displayed with columns 'a', 'b', 'c', 'd', 'message', and 'Roll no.'. The table has three rows with indices 0, 1, and 2.

```
DF=DF.reset_index()
DF
```

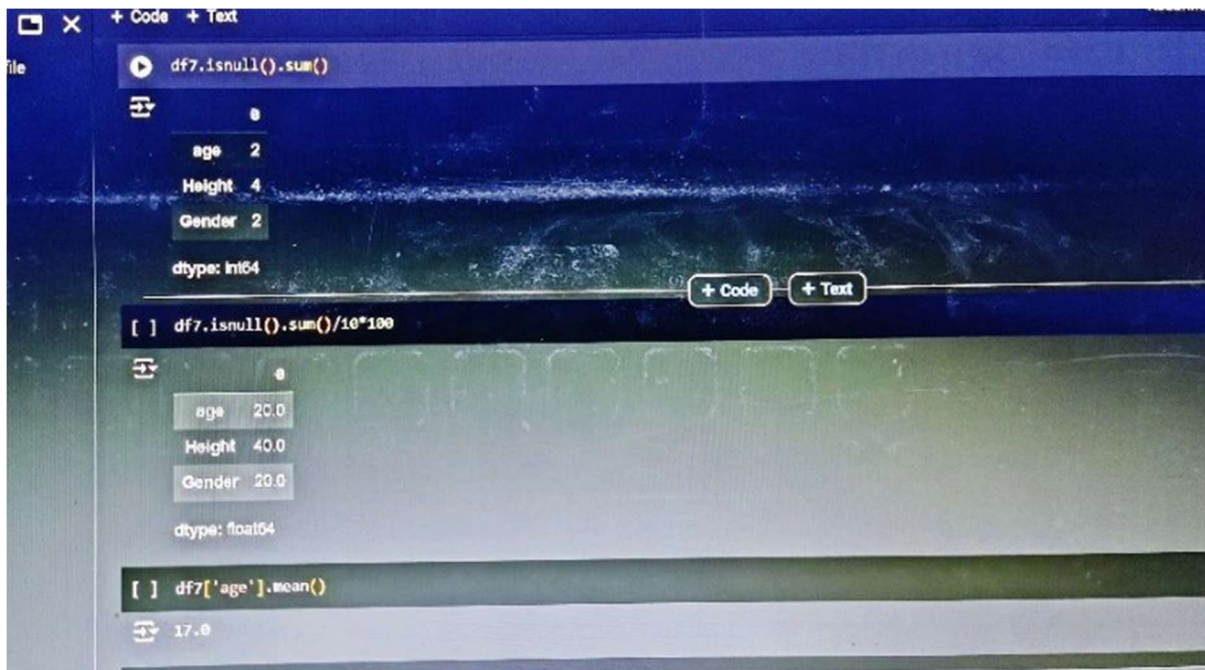
	Name	age	Marks
0	a	10	56.7
1	b	20	19.0
2	c	30	45.0

Next steps: [Generate code with DF](#) [View recommended plots](#) [New Interactive sheet](#)

```
[ ] L=[1,2,3]
    print(L)
```

```
[ ] df['Roll no.']=L
    df
```

	a	b	c	d	message	Roll no.
0	1	2	3	4	hello	1
1	5	6	7	8	world	2
2	9	10	11	12	foo	3



The screenshot shows a Jupyter Notebook interface with three code cells. The first cell contains `df7.isnull().sum()` and its output is a Series with values 0 for age, 4 for Height, and 2 for Gender, with a dtype of int64. The second cell contains `df7.isnull().sum()/10*100` and its output is a Series with values 20.0 for age, 40.0 for Height, and 20.0 for Gender, with a dtype of float64. The third cell contains `df7['age'].mean()` and its output is 17.0.

```
df7.isnull().sum()
age      0
Height   4
Gender    2
dtype: int64
```

```
[ ] df7.isnull().sum()/10*100
age      20.0
Height   40.0
Gender    20.0
dtype: float64
```

```
[ ] df7['age'].mean()
17.0
```

```
df = pd.DataFrame({"Values": [10, 20, 30, 40, 50]})
print(df.describe())
```