File Edit Selection View Go Run ··· Search

Diwali_Sales_Analysis.ipynb

C: > Users > Omkar's Laptop > OneDrive > Desktop > Data analytics > Python_Diwali_Sales_Analysis-main > Diwali_Sales_Analysis.ipynb > import numpy as np

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ··· Python 3.12.1

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt # visualizing data
%matplotlib inline
import seaborn as sns
```

[ ] Python

```
df = pd.read_csv('E:\Diwali Sales Data\Diwali Sales Data.csv', encoding= 'unicode_escape')
```

[ ] Python

```
df.shape
```

[15] ✓ 0.0s Python

```
(11251, 15)
```

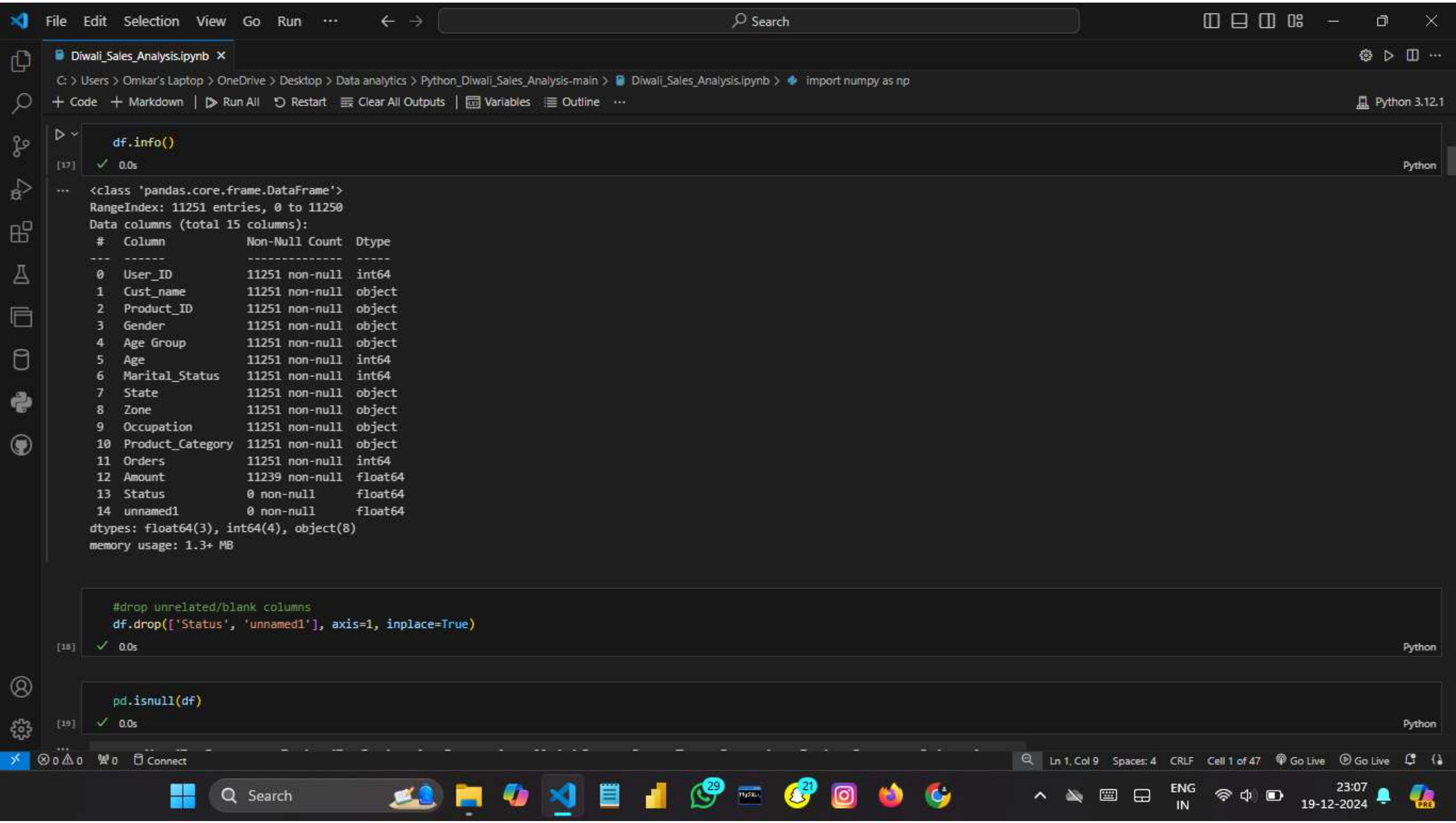
```
df.head(10)
```

[16] ✓ 0.0s Python

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.00	NaN	NaN
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhraï¿½Prade	Southern	Govt	Auto	3	23934.00	NaN	NaN
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.00	NaN	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.00	NaN	NaN
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.00	NaN	NaN
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	Food Processing	Auto	1	23877.00	NaN	NaN
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	Lawyer	Auto	4	23841.00	NaN	NaN
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	IT Sector	Auto	1	NaN	NaN	NaN
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Central	Govt	Auto	2	23809.00	NaN	NaN

⊗ 0 ⚠ 0 0 Connect Ln 1, Col 9 Spaces: 4 CRLF Cell 1 of 47 Go Live Go Live

Search ENG IN 23:07 19-12-2024

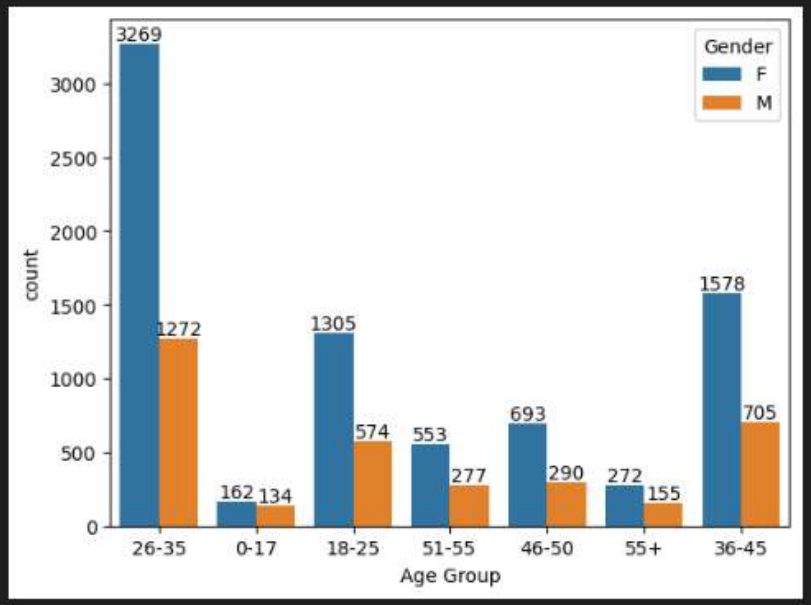

File Edit Selection View Go Run ···
Search
Diwali_Sales_Analysis.ipynb
C: > Users > Omkar's Laptop > OneDrive > Desktop > Data analytics > Python_Diwali_Sales_Analysis-main > Diwali_Sales_Analysis.ipynb > import numpy as np
+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ···
Python 3.12.1
df.info()
[17] ✓ 0.0s Python
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 User_ID 11251 non-null int64
1 Cust_name 11251 non-null object
2 Product_ID 11251 non-null object
3 Gender 11251 non-null object
4 Age Group 11251 non-null object
5 Age 11251 non-null int64
6 Marital_Status 11251 non-null int64
7 State 11251 non-null object
8 Zone 11251 non-null object
9 Occupation 11251 non-null object
10 Product_Category 11251 non-null object
11 Orders 11251 non-null int64
12 Amount 11239 non-null float64
13 Status 0 non-null float64
14 unnamed1 0 non-null float64
dtypes: float64(3), int64(4), object(8)
memory usage: 1.3+ MB
#drop unrelated/blank columns
df.drop(['Status', 'unnamed1'], axis=1, inplace=True)
[18] ✓ 0.0s Python
pd.isnull(df)
[19] ✓ 0.0s Python
Ln 1, Col 9 Spaces: 4 CRLF Cell 1 of 47 Go Live Go Live
Search
ENG IN
23:07 19-12-2024

File Edit Selection View Go Run ··· Search

Diwali_Sales_Analysis.ipynb ●

C: > Users > Omkar's Laptop > OneDrive > Desktop > Data analytics > Python_Diwali_Sales_Analysis-main > Diwali_Sales_Analysis.ipynb > df.head(10)

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ··· Python 3.12.1




```
# Total Amount vs Age Group
sales_age = df.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)

sns.barplot(x = 'Age Group',y= 'Amount' ,data = sales_age)
```

[24] ✓ 0.1s Python

```
<Axes: xlabel='Age Group', ylabel='Amount'>
```

Acquiring CodeLLDB platform package: 4%

⊗ 0 ⚠ 0 0 Connect Spaces: 4 Cell 4 of 47 Go Live Go Live

Search ENG IN 23:10 19-12-2024

File Edit Selection View Go Run ··· Search

Diwali_Sales_Analysis.ipynb ●

C: > Users > Omkar's Laptop > OneDrive > Desktop > Data analytics > Python_Diwali_Sales_Analysis-main > Diwali_Sales_Analysis.ipynb > df.head(10)

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ··· Python 3.12.1

State

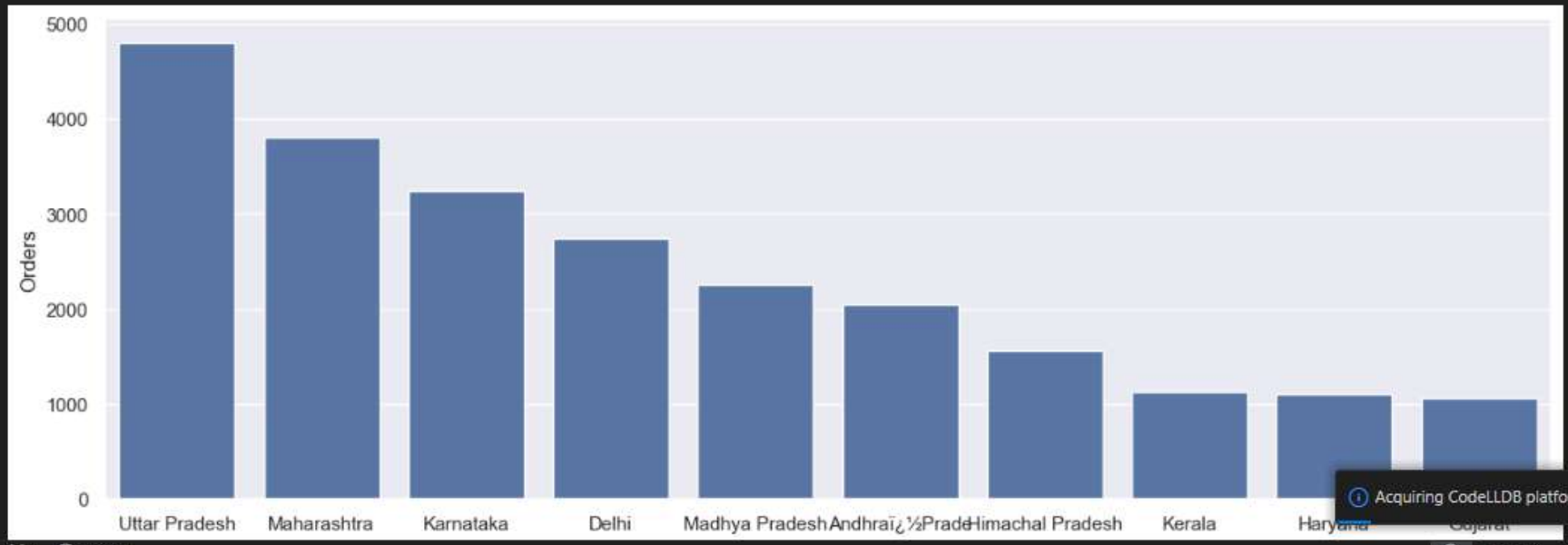
```
# total number of orders from top 10 states

sales_state = df.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x = 'State',y= 'Orders')
```

[25] ✓ 0.2s Python

```
<Axes: xlabel='State', ylabel='Orders'>
```



Acquiring CodeLLDB platform package: 6%

⊗ 0 ⚠ 0 0 Connect Spaces: 4 Cell 4 of 47 Go Live Go Live

Search ENG IN 23:10 19-12-2024