Udemy Data Analysis

By Surabhi Srivastava

“Hello! My name is Surabhi Srivastava. In this project I have utilised Python, Pandas Queries to solve questions that were related to Udemy Dataset.”

Questions to solve

- Display top 10 rows of the Dataset.
- Check last 5 rows of the Dataset.
- Find shape of our Dataset(no. of rows and columns).
- Getting information about our Dataset like total rows, columns, datatype of each column & memory requirement.
- Check all Null Values in the Dataset.
- Check the Duplicate data & Drop them.
- Find Out Number of Courses Per Subjects.
- For which Levels,Udemy providing the Courses.
- Display the Count of Paid & Free Courses.
- Display 10 Most Popular Courses as per num. of Subscribers.

Display Top 10 Rows of Dataset.

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Dataset
print(data)
sr=data.head(10)
print(sr)
```

```
[10 rows x 11 columns]


Process finished with exit code 0
```

Check Last 5 Rows of the Dataset.

```python
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/DS/Ude
print(data)
sr=data.tail(5)
print(sr)
```

```
[5 rows x 11 columns]
```

Find Shape of our Dataset.

```python
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/DS/Ude
print(data)
sr=data.shape
print("shape of dataset:",data.shape)
print("Number of rows:",data.shape[0])
print("Number of columns:",data.shape[1])
```

```
[3682 rows x 11 columns]
shape of dataset: (3682, 11)
Number of rows: 3682
Number of columns: 11
```

Getting information about Dataset

```
import pandas as pd                                                    ⚠ 3
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/D
print(data)
sr=data.info()
print(sr)
```

```
dtypes: bool(1), int64(4), object(6)
memory usage: 291.4+ KB
None
```

Check Null Values in Dataset.

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/DS/Ude
print(data)
sr=data.isnull().values.any()
print("any missing value?",sr)
```

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/DS/Ud
print(data)
sr=data.isnull().sum()
print("total null values",sr)
```

```
[3682 rows x 11 columns]
any missing value? False
```

```
price                0
num_subscribers      0
num_reviews          0
num_lectures         0
level                0
content_duration     0
published_timestamp  0
subject              0
```

Check the Duplicate data & Drop them.

```python
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/mai
print(data)
sr=data.drop_duplicates().any()
print("Is there any duplicate data?",sr)
```

```
is_paid                 True
price                   True
num_subscribers         True
num_reviews             True
num_lectures            True
level                   True
content_duration        True
```

No. of Courses per Subject.

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinker
print(data)
sr=data["subject"].value_counts()
print(sr)
```

```
[3682 rows x 11 columns]
subject
Web Development          1200
Business Finance         1199
Musical Instruments       680
Graphic Design            603
Name: count, dtype: int64
```

Display the count of paid and Free Courses

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/he
print(data)
sr=data["is_paid"].value_counts()
print(sr)
```

```
is_paid
True      3372
False      310
Name: count, dtype: int64
```

10 most Popular Courses as per no. of Subscribers

```
import pandas as pd
data=pd.read_csv("https://raw.githubusercontent.com/DataThinkers/Datasets/refs/
print(data)
sr=data.sort_values(by="num_subscribers",ascending=False).head(10)
print(sr)
```

```
3316      764164  ...       Web Development
1388       19421  ...  Musical Instruments
3556      473160  ...       Web Development
2233       94430  ...       Web Development
2886      130064  ...       Web Development
2034      364426  ...       Web Development

[10 rows x 11 columns]
```