

Activity No. 6.2

Course Code: CPE 310

Program: BSIE

Course Title: Fundamentals of Data Science

Date Performed: September 26 2025

Section: IE22S1

Date Submitted: September 26 2025

Name: James Bernard Lozada

Instructor: Roman Richards

1. Discussion

The activity is about Getting Started with Pandas

2. Materials and Equipment

The materials I used in completion of the activity is a laptop and a Google Colab as instructed to the activity given by my instructor

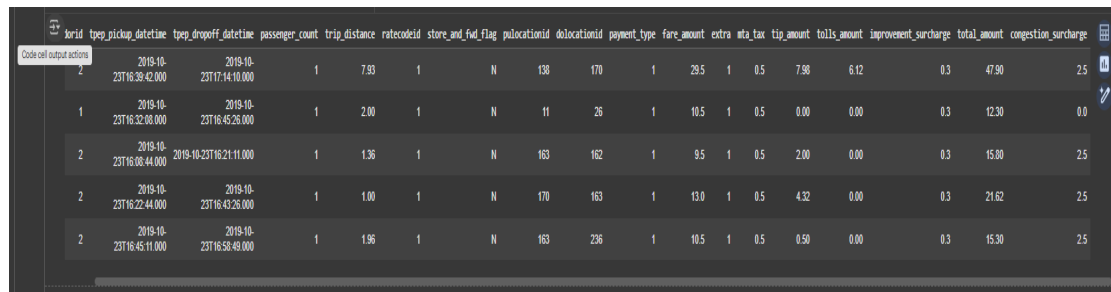
3. Procedure

The procedure in undergoing the completion of the hands on activity is to introduce DataFrames, Series, and Index on Pandas

4. Output

Exercise 1.1

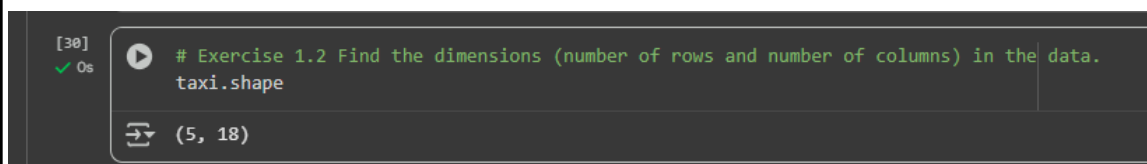
```
# Exercise 1.1 Create a DataFrame by reading in the  
2019_Yellow_Taxi_Trip_Data.csv file. Examine the first 5 rows.
```



trip_id	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	ratecodeid	store_and_fwd_flag	pulocationid	dolocationid	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	congestion_surcharge
2	2019-10-23T16:39:42.000	2019-10-23T17:14:10.000	1	7.93	1	N	138	170	1	29.5	1	0.5	7.98	6.12	0.3	47.90	2.5
1	2019-10-23T16:32:00.000	2019-10-23T16:49:26.000	1	2.00	1	N	11	26	1	10.5	1	0.5	0.00	0.00	0.3	12.30	0.0
2	2019-10-23T16:08:44.000	2019-10-23T16:21:11.000	1	1.36	1	N	163	162	1	9.5	1	0.5	2.00	0.00	0.3	15.00	2.5
2	2019-10-23T16:22:44.000	2019-10-23T16:43:26.000	1	1.00	1	N	170	163	1	13.0	1	0.5	4.32	0.00	0.3	21.62	2.5
2	2019-10-23T16:45:11.000	2019-10-23T16:58:49.000	1	1.96	1	N	163	236	1	10.5	1	0.5	0.50	0.00	0.3	15.30	2.5

Exercise 1.2

```
# Exercise 1.2 Find the dimensions (number of rows and number of  
columns) in the data.
```



```
[30]  
✓ Os # Exercise 1.2 Find the dimensions (number of rows and number of columns) in the data.  
taxi.shape  
↵ (5, 18)
```

Exercise 1.3

```
Exercise 1.3 Using the data in the 2019_Yellow_Taxi_Trip_Data.csv  
file, calculate summary statistics for the fare_amount, tip_amount,  
tolls_amount, and total_amount columns.
```

```
# Exercise 1.3 Using the data in taxi
taxi.tip_amount.value_counts()
```

tip_amount	count
7.98	1
0.00	1
2.00	1
4.32	1
0.50	1

dtype: int64

```
# Exercise 1.3 Using the data in taxi
taxi.tip_amount.value_counts()
taxi.tolls_amount.value_counts()
```

tolls_amount	count
0.00	4
6.12	1

dtype: int64

```
taxi.fare_amount.value_counts()
```

fare_amount	count
10.5	2
29.5	1
9.5	1
13.0	1

dtype: int64

```
taxi.total_amount.value_counts()
```

total_amount	count
47.90	1
12.30	1
15.80	1
21.62	1
15.30	1

Exercise 1.4

```
# Exercise 1.4 Isolate the fare_amount, tip_amount, tolls_amount,
and total_amount for the longest trip by distance (trip_distance).
```

<code>taxi.loc[taxi['trip_distance'].idxmax()]</code>	
	0
output actions	2
vendorid	2
tpep_pickup_datetime	2019-10-23T16:39:42.000
tpep_dropoff_datetime	2019-10-23T17:14:10.000
passenger_count	1
trip_distance	7.93
ratecodeid	1
store_and_fwd_flag	N
pulocationid	138
dolocationid	170
payment_type	1
fare_amount	29.5
extra	1
mta_tax	0.5
tip_amount	7.98
tolls_amount	6.12
improvement_surcharge	0.3
total_amount	47.9
congestion_surcharge	2.5
dtype: object	
5. Supplementary Activity	
1. Discuss your chosen dataset from the previous discussion post.	

```
X.loc[X['age'].idxmax()]
```

	161
age	77.0
sex	1.0
cp	4.0
trestbps	125.0
chol	304.0
fbs	0.0
restecg	2.0
thalach	162.0
exang	1.0
oldpeak	0.0
slope	1.0
ca	3.0
thal	3.0

dtype: float64

```
X.shape
```

```
(303, 13)
```

Maximum Index

Dimension

```

{ 'uci_id': 45, 'name': 'Heart Disease', 'repository_url': 'https://archive.ics.uci.edu/dataset/45/heart+disease', 'data_
  name      role      type demographic \
0      age  Feature    Integer          Age
1      sex  Feature    Categorical        Sex
2      cp   Feature    Categorical        None
3  trestbps Feature    Integer          None
4      chol Feature    Integer          None
5      fbs  Feature    Categorical        None
6  restecg Feature    Categorical        None
7  thalach Feature    Integer          None
8  exang   Feature    Categorical        None
9  oldpeak Feature    Integer          None
10     slope Feature    Categorical        None
11      ca  Feature    Integer          None
12     thal Feature    Categorical        None
13     num  Target     Integer          None

description  units  missing_values
0            None  years              no
1            None  None               no
2            None  None               no
3  resting blood pressure (on admission to the ho... mm Hg              no
4            serum cholestoral mg/dl              no
5      fasting blood sugar > 120 mg/dl None              no
6            None  None               no
7      maximum heart rate achieved None              no
8      exercise induced angina None              no
9  ST depression induced by exercise relative to ... None              no
10           None  None               no
11 number of major vessels (0-3) colored by flour... None              yes
12           None  None              yes
13      diagnosis of heart disease None              no

```

DATA set of Heart Disease

```
X.cp.value_counts()
X.sex.value_counts()
X.age.value_counts()
```

64	10
63	9
67	9
53	8
61	8
43	8
45	8
55	8
65	8
42	8
46	7
66	7
48	7
50	7
49	5
47	5
70	4
39	4
68	4
35	4
40	3
69	3
71	3
37	2
34	2
38	2
29	1
77	1
74	1
76	1

df.type: int64

Summary Statistics of Heart Disease

