

Q1

Python Console × +

```
... # Display a message on the console and fill out the blank: [6pts]
...
... df = sns.load_dataset('taxi')
... print(f"There are {df.shape[0]} observations inside the raw dataset.")
... print(f"There are {df.shape[1]} features[columns] inside the raw dataset.")
>>> ...
There are 6433 observations inside the raw dataset.
There are 14 features[columns] inside the raw dataset.
>>>
```

Q2

```
Python Console x +
...
... print(f"The cleaned dataset has {df.shape[0]} # of observations.")
... print(f"The cleaned dataset has {df.shape[1]} # of features.")
... print(f"The list of removed features are {drop_count}")
...
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 0
>>> 44
>>> 26
>>> 45
>>> 26
>>> 45
The cleaned dataset has 6433 # of observations.
The cleaned dataset has 9 # of features.
The list of removed features are 5

>>>
```

Q3

```
>>> print(df.isnull().sum())
... print(df.isna().sum())
...
pickup          0
dropoff          0
passengers       0
distance         0
fare             0
tip              0
tolls            0
total            0
color            0
dtype: int64
pickup          0
dropoff          0
passengers       0
distance         0
fare             0
tip              0
tolls            0
total            0
color            0
dtype: int64

>>> |
```

Q4

```
Python Console × +
fare      0
tip       0
tolls     0
total     0
color     0
dtype: int64
pickup    0
dropoff   0
passengers 0
distance  0
fare      0
tip       0
tolls     0
total     0
color     0
dtype: int64
>>> print(f"The mean of the total is {round(df['total'].mean(), 2)}$")
... print(f"The mean of the tip is {round(df['tip'].mean(), 2)}$")
... print(f"The variance of the total is {round(df['total'].var(), 2)}$")
... print(f"The variance of the tip is {round(df['tip'].var(), 2)}$")
...
The mean of the total is 18.52$
The mean of the tip is 1.98$
The variance of the total is 190.87$
The variance of the tip is 6.0$

>>>
```

Q5

```
the variance of the tip is 0.0$
>>> df['tip_percentage'] = df['tip'] / df['total']
... print(df.tail(n=5))
...
... print("1-", df[df['tip_percentage'] == 0].shape[0])
... # print("2-", df[df['tip_percentage'] >= 10 and df[df['tip_percentage'] < 15]].shape[0])
... # print("3-", df[df['tip_percentage'] >= 15 and df[df['tip_percentage'] < 20]].shape[0])
... print("4-", df[df['tip_percentage'] >= 0.2].shape[0])
...

      pickup      dropoff ... color tip_percentage
6428 2019-03-31 09:51:53 2019-03-31 09:55:27 ... green          0.17
6429 2019-03-31 17:38:00 2019-03-31 18:34:23 ... green          0.00
6430 2019-03-23 22:55:18 2019-03-23 23:14:25 ... green          0.00
6431 2019-03-04 10:09:25 2019-03-04 10:14:29 ... green          0.00
6432 2019-03-13 19:31:22 2019-03-13 19:48:02 ... green          0.17

[5 rows x 10 columns]
1- 2311
4- 480

>>>
```