Challenging Task – 2

Name: Gangireddy Madhurima

Reg Number: 21MIS1155

Question Number: 4

Q4	Convert the fare data into decimal data Split the home.dest with two separated columns, first column is about Home city and second column is destination city and updated its entries, for example "St Louis, MO" St Louis will comes under Home city column and remaining part into destination city. From the above result identify the Pclass = 1 passengers who are belongs to "New York" city.

Aim:

To analyze Titanic dataset:

- Convert the fare column into decimal data.
- Split the home.dest column into two separate columns: one for Home City and one for Destination City.
- Identify the passengers with Pclass = 1 who belong to "New York" city.

Procedure:

1. Load the Titanic Dataset:

    - The dataset contains information about passengers, including their fare, home destination, and passenger class (Pclass).

2. Convert fare to Decimal:

- The fare column may contain non-numeric values. Use the pd.to_numeric() function to convert it into decimal data. Non-numeric values will be replaced with NaN.

3. Split home.dest into Two Columns:

- Split the home.dest column into Home City and Destination City by separating the values based on the / delimiter.

4. Filter for Pclass = 1 and Passengers from New York:

- Extract passengers from the dataset who belong to Pclass = 1 and whose Home City is "New York". Filter the dataset using these conditions.

5. Display the Output:

- Print the resulting filtered data that shows only the passengers who meet the criteria of being in Pclass = 1 and from "New York".

Question Solved:

Convert the fare data into decimal format, split the home.dest column into two separate columns (Home City and Destination City), and identify the passengers with Pclass = 1 who are from "New York" city.

Codes and corresponding Outputs :

1. Installing necessary libraries

```
!pip install pandas numpy matplotlib scikit-learn
```

2. Importing Libararies to Notebook

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
```

3. Converting the Fare Data into Decimal Data – Question 1

```
titanic_data = pd.read_csv('titanic_data.csv')

titanic_data['fare'] = pd.to_numeric(titanic_data['fare'],
errors='coerce')

print(titanic_data[['fare']].head())
```

```
titanic_data.head()
```

titanic_data.head()

	x	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	home.dest
0	1	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	St Louis, MO
1	2	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.5500	C22 C26	S	Montreal, PQ / Chesterville, ON
2	3	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.5500	C22 C26	S	Montreal, PQ / Chesterville, ON
3	4	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.5500	C22 C26	S	Montreal, PQ / Chesterville, ON
4	5	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.5500	C22 C26	S	Montreal, PQ / Chesterville, ON

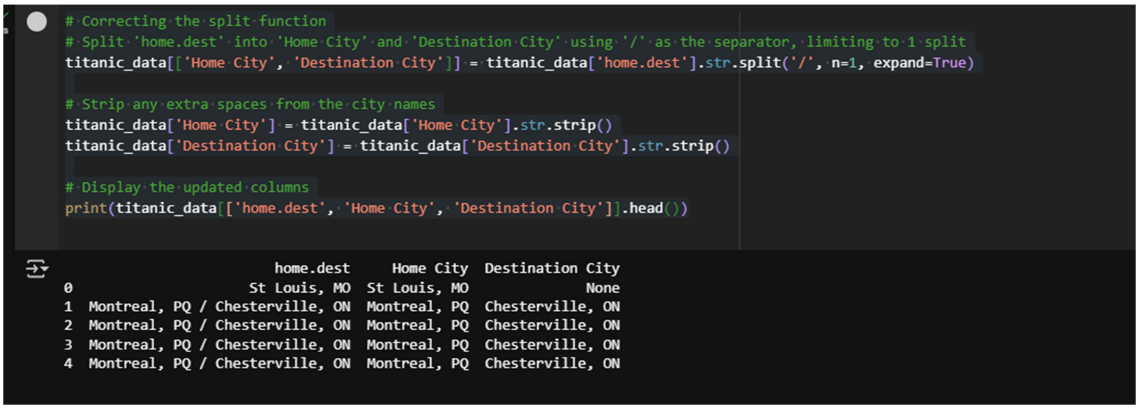
Next steps: Generate code with titanic_data | View recommended plots | New interactive sheet

4. Split home.dest column into Home City and Destination City – Question 2

```
titanic_data[['Home City', 'Destination City']] =
titanic_data['home.dest'].str.split('/', n=1, expand=True)

titanic_data['Home City'] = titanic_data['Home City'].str.strip()
titanic_data['Destination City'] = titanic_data['Destination
City'].str.strip()

print(titanic_data[['home.dest', 'Home City', 'Destination
City']].head())
```



```
# Correcting the split function
# Split 'home.dest' into 'Home City' and 'Destination City' using '/' as the separator, limiting to 1 split
titanic_data[['Home City', 'Destination City']] = titanic_data['home.dest'].str.split('/', n=1, expand=True)

# Strip any extra spaces from the city names
titanic_data['Home City'] = titanic_data['Home City'].str.strip()
titanic_data['Destination City'] = titanic_data['Destination City'].str.strip()

# Display the updated columns
print(titanic_data[['home.dest', 'Home City', 'Destination City']].head())
```

```
                     home.dest     Home City  Destination City
0                  St Louis, MO  St Louis, MO              None
1  Montreal, PQ / Chesterville, ON  Montreal, PQ  Chesterville, ON
2  Montreal, PQ / Chesterville, ON  Montreal, PQ  Chesterville, ON
3  Montreal, PQ / Chesterville, ON  Montreal, PQ  Chesterville, ON
4  Montreal, PQ / Chesterville, ON  Montreal, PQ  Chesterville, ON
```

5. Filter passengers in Pclass = 1 from New York – Question 3

```
pclass_1_new_york = titanic_data[(titanic_data['pclass'] == 1) &
(titanic_data['Home City'].str.contains('New York', na=False))]

print(pclass_1_new_york[['pclass', 'Home City', 'Destination City']])
```