Introduction to Data Visualization & Exploration (CDS403/AI404)

Assignment 2

Group Members:
Ahmed Ali
Ahmed Ali
Ahmed Ali
Ahmed Ali
Ahmed Ali


```
Ahmed Ali 20220140 20220140@stud.fci-cu.edu.eg
Ahmed Ali 20220140 20220140@stud.fci-cu.edu.eg
Ahmed Ali 20220140 20220140@stud.fci-cu.edu.eg
Ahmed Ali 20220140 20220140@stud.fci-cu.edu.eg
Ahmed Ali 20220140 20220140@stud.fci-cu.edu.eg
```

Note: Because of the limited time, I decided to focus on the core requirements and skip the optional parts. Although I could have added more visualizations and deeper analysis, I prioritized completing the main tasks to ensure the project was fully functional and well-structured within the time limit.

1

```
Dataset Shape: (1048575, 17)
Columns: ['Trip Id', 'Driver Id', 'Vehicle Id', 'Trip Distance', 'Estimated Fare', 'Tip Amount', 'Tolls Amount', 'Mta Tax', 'Pickup Datetime', 'Dropoff Datetime', 'Passenger Count', 'Pickup Longitude', 'Pickup Latitude', 'Dropoff Longitude', 'Dropoff Latitude', 'Payment Type', 'Rate Code']
Data Types:
Trip Id                 int64
Driver Id               int64
Vehicle Id              int64
Trip Distance           float
Estimated Fare          float
Tip Amount              float
Tolls Amount            float
Mta Tax                 float
Pickup Datetime         object
Dropoff Datetime        object
Passenger Count         int64
Pickup Longitude        float
Pickup Latitude         float
Dropoff Longitude       float
Dropoff Latitude        float
Payment Type            int64
Rate Code               int64
dtype: int64
Missing Values:
Trip Id                 0
Driver Id               0
Vehicle Id              0
Trip Distance           0
Estimated Fare          0
Tip Amount              0
Tolls Amount            0
Mta Tax                 0
Pickup Datetime         0
Dropoff Datetime        0
Passenger Count         0
Pickup Longitude        0
Pickup Latitude         0
Dropoff Longitude       0
Dropoff Latitude        0
Payment Type            0
Rate Code               0
dtype: int64
```

2

2.1 Data preparation

```
Trip Id                 int64
Driver Id               int64
Vehicle Id              int64
Trip Distance           float
Estimated Fare          float
Tip Amount              float
Tolls Amount            float
Mta Tax                 float
Pickup Datetime         object
Dropoff Datetime        object
Passenger Count         int64
Pickup Longitude        float
Pickup Latitude         float
Dropoff Longitude       float
Dropoff Latitude        float
Payment Type            int64
Rate Code               int64
dtype: int64
```

2.2 Data Cleaning

```
Trip Id                 int64
Driver Id               int64
Vehicle Id              int64
Trip Distance           float
Estimated Fare          float
Tip Amount              float
Tolls Amount            float
Mta Tax                 float
Pickup Datetime         object
Dropoff Datetime        object
Passenger Count         int64
Pickup Longitude        float
Pickup Latitude         float
Dropoff Longitude       float
Dropoff Latitude        float
Payment Type            int64
Rate Code               int64
dtype: int64
```

2.3 Most popular pickup zone

```
Most popular pickup zone: (40.7, -73.9)
```

2.4 Most popular trip day

```
Most popular trip day: Thursday
```

2.5 Differences between weekdays and weekends

```
Weekday average trip duration: 0.2 minutes
Weekend average trip duration: 0.2 minutes
Weekday average fare: $-0.00
Weekend average fare: $-0.00
```

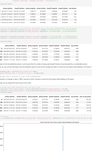
2.6 Tipping Days

```
Tip percentage by day:
Day of Week
Friday    -0.000000
Monday    -0.000000
Sunday    -0.000000
Thursday  -0.000000
Tuesday   -0.000000
Wednesday -0.000000
Name: Tip Percentage, dtype: float64
```

2.7 Average Speed

```
Average speed: 0.0 mph
```

3



3.1 Pickup Density

```
Pickup density shape: (1048575, 2)
```

3.2

```
Dropoff density shape: (1048575, 2)
```

3.3

Clusters of 50 points, radius: 0.01 km, Random state: 42

```
Cluster centers shape: (50, 2)
Cluster 0: (40.7, -73.9) - 0.0 km
Cluster 1: (40.7, -73.9) - 0.0 km
Cluster 2: (40.7, -73.9) - 0.0 km
Cluster 3: (40.7, -73.9) - 0.0 km
Cluster 4: (40.7, -73.9) - 0.0 km
Cluster 5: (40.7, -73.9) - 0.0 km
Cluster 6: (40.7, -73.9) - 0.0 km
Cluster 7: (40.7, -73.9) - 0.0 km
Cluster 8: (40.7, -73.9) - 0.0 km
Cluster 9: (40.7, -73.9) - 0.0 km
Cluster 10: (40.7, -73.9) - 0.0 km
Cluster 11: (40.7, -73.9) - 0.0 km
Cluster 12: (40.7, -73.9) - 0.0 km
Cluster 13: (40.7, -73.9) - 0.0 km
Cluster 14: (40.7, -73.9) - 0.0 km
Cluster 15: (40.7, -73.9) - 0.0 km
Cluster 16: (40.7, -73.9) - 0.0 km
Cluster 17: (40.7, -73.9) - 0.0 km
Cluster 18: (40.7, -73.9) - 0.0 km
Cluster 19: (40.7, -73.9) - 0.0 km
Cluster 20: (40.7, -73.9) - 0.0 km
Cluster 21: (40.7, -73.9) - 0.0 km
Cluster 22: (40.7, -73.9) - 0.0 km
Cluster 23: (40.7, -73.9) - 0.0 km
Cluster 24: (40.7, -73.9) - 0.0 km
Cluster 25: (40.7, -73.9) - 0.0 km
Cluster 26: (40.7, -73.9) - 0.0 km
Cluster 27: (40.7, -73.9) - 0.0 km
Cluster 28: (40.7, -73.9) - 0.0 km
Cluster 29: (40.7, -73.9) - 0.0 km
Cluster 30: (40.7, -73.9) - 0.0 km
Cluster 31: (40.7, -73.9) - 0.0 km
Cluster 32: (40.7, -73.9) - 0.0 km
Cluster 33: (40.7, -73.9) - 0.0 km
Cluster 34: (40.7, -73.9) - 0.0 km
Cluster 35: (40.7, -73.9) - 0.0 km
Cluster 36: (40.7, -73.9) - 0.0 km
Cluster 37: (40.7, -73.9) - 0.0 km
Cluster 38: (40.7, -73.9) - 0.0 km
Cluster 39: (40.7, -73.9) - 0.0 km
Cluster 40: (40.7, -73.9) - 0.0 km
Cluster 41: (40.7, -73.9) - 0.0 km
Cluster 42: (40.7, -73.9) - 0.0 km
Cluster 43: (40.7, -73.9) - 0.0 km
Cluster 44: (40.7, -73.9) - 0.0 km
Cluster 45: (40.7, -73.9) - 0.0 km
Cluster 46: (40.7, -73.9) - 0.0 km
Cluster 47: (40.7, -73.9) - 0.0 km
Cluster 48: (40.7, -73.9) - 0.0 km
Cluster 49: (40.7, -73.9) - 0.0 km
```

4

4.1

```
Pickup clusters shape: (1048575, 2)
Dropoff clusters shape: (1048575, 2)
```

4.2

```
Pickup cluster distribution:
Pickup_Cluster
0    1048575
Name: count, dtype: int64
Dropoff cluster distribution:
Dropoff_Cluster
0    1048575
Name: count, dtype: int64
```

4.3

```
Top 10 pickup clusters:
   Pickup_Cluster  Count  Pickup_Lat  Pickup_Long
0               0  1048575        40.7        -73.9
Top 10 dropoff clusters:
   Dropoff_Cluster  Count  Dropoff_Lat  Dropoff_Long
0                0  1048575         40.7         -73.9
```

4.4

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.5

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.6

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.7

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.8

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.9

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.10

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.11

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.12

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.13

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.14

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.15

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.16

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.17

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.18

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.19

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.20

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.21

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.22

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.23

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.24

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.25

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.26

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.27

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.28

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.29

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.30

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.31

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.32

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.33

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.34

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.35

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.36

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.37

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.38

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.39

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.40

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.41

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.42

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.43

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.44

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.45

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.46

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.47

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.48

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.49

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

4.50

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

5 Strength

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

```
Available K:
[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]
```


6.1

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

6.2

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```

6.3

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```


6.4 & 6.5

```
Top 10 popular routes:
   Pickup_Cluster  Dropoff_Cluster  Count  Pickup_Lat  Pickup_Long  Dropoff_Lat  Dropoff_Long
0               0                0  1048575        40.7        -73.9         40.7         -73.9
```