**Summary for Master data:**

--------------------------------------------------

Shape: (359392, 14)

Data Types:

1. Transaction ID int64
2. Date of Travel int64
3. Company object
4. City object
5. KM Travelled float64
6. Price Charged float64
7. Cost of Trip float64
8. Customer ID int64
9. Payment\_Mode object
10. Gender object
11. Age int64
12. Income (USD/Month) int64
13. Population object
14. Users object
15. dtype: object

Null Values:

1. Transaction ID 0
2. Date of Travel 0
3. Company 0
4. City 0
5. KM Travelled 0
6. Price Charged 0
7. Cost of Trip 0
8. Customer ID 0
9. Payment\_Mode 0
10. Gender 0
11. Age 0
12. Income (USD/Month) 0
13. Population 0
14. Users 0

Duplicate Rows: 0

**Unique Values in Each Column:**

1. Transaction ID 359392
2. Date of Travel 1095
3. Company 2
4. City 19
5. KM Travelled 874
6. Price Charged 99176
7. Cost of Trip 16291
8. Customer ID 46148
9. Payment\_Mode 2
10. Gender 2
11. Age 48
12. Income (USD/Month) 22725
13. Population 19
14. Users 19

Total Rows: 359392

**Revenue per company**:

Company

Yellow Cab 125,853,887.19

Pink Cab 26,328,251.33

The dates don’t really make sense for now, so date column is quite useless

Yellow cab has more revenue but why ?

A graph with orange and blue bars

Description automatically generated

In New York the yellow company is making so much more money than pink

A graph of blue and orange bars

Description automatically generated

For yellow cab the average price charged per kilometre in different cities varies a lot and it somewhere around 16 to 26 in New York

Whereas for Pink company it is quite constant 12 to 14

A graph showing a number of transactions

Description automatically generated

Yellow company has more repeat customers