**Data Overview**

Cab\_Data.csv

Customer\_ID.csv

Transaction\_ID.csv

City.csv

From the above datasets we created a master dataset with features:

* Transaction ID
* Date of Travel
* Company
* City
* KM Travelled
* Price Charged
* Cost of Trip
* Customer ID
* Payment Mode
* Gender
* Age
* Income (USD/Month)

**Data Preprocessing**

Reviewed source documentation, understanding field names and data types. Identified relationships across files and performed necessary transformations. Detected, and removed outliers. Created a master data set, explaining relationships, and removed duplicates. Analyzed NA values and outliers.

**Hypothesis Generation and Investigation**

1) Which Cab company have higher number of customers?  
2) Which City have the highest and lowest customers with respect to each company?  
3) Which Cab company have highest profit?  
4) Which company customers are preferring with age group of 40-60?  
5) Average Customer monthly income of both the cab companies?

**Analysis Process and Findings**

Defined the business problem, outlined the data intake report, and explained the steps taken for analysis. Applied analytical techniques justified by the nature of the data. Presented findings for each hypothesis, showcasing patterns and trends.

A graph of a number of people

Description automatically generated with medium confidenceA graph of a company

Description automatically generated with medium confidenceA graph with numbers and a bar

Description automatically generated

A screen shot of a graph

Description automatically generated

New York has the highest number of customers and Pittsburgh has the lowest number of customers for yellow cabs. Compared to all cities pink cabs have higher customers in Los Angeles and lowest in Pittsburgh.

A graph of a number of companies

Description automatically generated