

Data Intake Report

Name: **G2M insight for Cab Investment firm**
Report date: 07/14/2023
Internship Batch: **LISUM23: 30 June – 30 Sept 2023**
Version:<1.0>
Data intake by: **Mani Shanker Kamarapu**
Data intake reviewer: Mani Shanker Kamarapu
Data storage location: GitHub

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of files	1/4
Total number of features	7
Base format of the file	csv
Size of the data	21.2 MB

City.csv

Total number of observations	20
Total number of files	2/4
Total number of features	3
Base format of the file	csv
Size of the data	4 KB

Customer_ID.csv

Total number of observations	49171
Total number of files	3/4
Total number of features	4
Base format of the file	csv
Size of the data	1.1 MB

Transaction_ID.csv

Total number of observations	440098
Total number of files	4/4
Total number of features	3
Base format of the file	csv
Size of the data	9 MB

Proposed Approach:

- I collected 4 datasets and performed EDA to understand the relations to determine which company is performing better and is a better investment opportunity for XYZ.
- Below are the few hypotheses:
 - The average income of customers using Pink Cab is different from those using Yellow Cab.
 - The number of cab users is influenced by the population of the city.
 - The average price charged by Pink Cab is different from Yellow Cab.
 - The choice of payment mode (Cash or Card) is influenced by the company.
 - The average distance travelled by customers using Pink Cab is different from those using Yellow Cab.
 - The proportion of customers using the cab service is different between males and females.