

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 18/6/2021

Internship Batch: LISUM01

Version:<1.0>

Data intake by: Boshra Eisa

Data intake reviewer: Boshra Eisa

Data storage location: <https://github.com/Bushra-Eisa/G2M-Case-Study>

Tabular data details:

Cab_Data.CSV:

Total number of observations	359393
Total number of files	1
Total number of features	7
Base format of the file	CSV
Size of the data	22,418 KB

Customer_ID.CSV:

Total number of observations	49172
Total number of files	1
Total number of features	4
Base format of the file	.CSV
Size of the data	1,027 KB

City.CSV:

Total number of observations	21
Total number of files	1
Total number of features	3
Base format of the file	.CSV
Size of the data	1KB

Transaction_ID.CSV:

Total number of observations	440099
Total number of files	1
Total number of features	3
Base format of the file	.CSV
Size of the data	8788KB

Data.CSV:

Total number of observations	359393
Total number of files	1
Total number of features	18
Base format of the file	.CSV

Size of the data	38740KB
------------------	---------

Proposed Approach:

Firstly, I started off by loading the CSV files provided into a Jupyter notebook and started merging the relevant files together to form a single master data frame to perform the analysis on. I have also added a calculated 'Profit' column by subtracting the Cost of Trip from the Price Charged and changed the data type for the Date of Travel. After checking for Null values, my data frame was in perfect condition to start the analysis.

The data analysis consisted of:

- calculating the profits for both companies, the Yellow Cab and the Pink Cab, annually and monthly.
- Checking which company has more demand per city
- The cost of the trip per KM for each company
- How much percentage the users spend out of their income on the cabs for each company
- The gender of the users for each company
- The payment methods for the users for each company

The hypothesis was done by finding the p-value for the points mentioned above.

In conclusion, the data supported that the Yellow Cab is the better choice to invest in.