

Data Intake Report

Name: G2M-insight-for-Cab-Investment-firm

Report date: 14-03-2023

Internship Batch: LISUM19

Version: 1.0

Data intake by: Gabriel Antunes Soares

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/soaresgabe/G2M-insight-for-Cab-Investment-firm>

Tabular data details: Cab_Data

| | |
|-------------------------------------|---------|
| Total number of observations | 359392 |
| Total number of files | 1 |
| Total number of features | 7 |
| Base format of the file | .csv |
| Size of the data | 20.6 MB |

Tabular data details: City

| | |
|-------------------------------------|------|
| Total number of observations | 20 |
| Total number of files | 1 |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 1 KB |

Tabular data details: Customer_ID

| | |
|-------------------------------------|-------|
| Total number of observations | 49171 |
| Total number of files | 1 |
| Total number of features | 4 |
| Base format of the file | .csv |
| Size of the data | 1 MB |

Tabular data details: Transaction_ID

| | |
|-------------------------------------|---------|
| Total number of observations | 440098 |
| Total number of files | 1 |
| Total number of features | 3 |
| Base format of the file | .csv |
| Size of the data | 8.78 MB |

Proposed Approach:

- Data Cleaning:

The date in the "cab" dataset was in a numeric format, so it was converted to a datetime format.

A new column "profit" was added to the "cab" dataset by subtracting the "Cost of Trip" from the "Price Charged" columns.

- Data Exploration:

The unique values of the "Company" and "City" columns were examined in the "cab" dataset.

The unique values of the "Payment_Mode" column were examined in the "transaction" dataset.

- Data Merging:

The "cab", "city", "customer", and "transaction" datasets were merged using their common columns.

- Data Analysis:

The total number of customers, total revenue, total cost, and total profit were calculated for each company and city.

The average profit per ride and average profit per customer were calculated for each company and city.

The factors affecting the profitability of the cab companies were identified by performing exploratory data analysis and visualizations.

- Conclusion:

Based on the analysis, conclusions were drawn about the profitability of the cab companies and recommendations were made to improve their profitability.

Based on the analysis, the following conclusions can be drawn:

- The Yellow cab is preferred over the Pink cab, for unknown reasons, nothing justifies this preference.
- Although the profit has decreased for the Yellow cab, it is still the most preferred cab service and has the highest average profit per customer.
- The investor should invest in the Yellow cab, as it has the highest average profit per customer and is the most preferred cab service.