

Data Intake Report

Name: G2M insight for Cab Investment firm (Must for all Specialization)

Report Date: 05/13/2022

Internship Batch: LISUM09

Version: <1.0>

Data intake by: Data Glacier Virtual Internship 2022

Data intake reviewer:

Data storage location:

Cab_Data.csv details:

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	20,663 KB

City.csv details:

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	1KB

Customer_ID.csv details:

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	1,027KB

Transaction_ID.csv details:

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	8,788KB

Proposed Approach:

After merging three different dataframe, the first data of each user is kept, and the new dataframe has a total of 359392 rows.

Hypothesis:

1. Middle-aged and seniors may need more cab service

From the dataframe and the chart, we will see that younger people are more likely to use cab services.

2. Populous cities will have a more extensive customer base

From the chart, we can notice that most of the user base occurs in cities with high population density.

3. Allowing customers to pay in cash will be more profitable

The results show that Credit card payments are more prone to negative profits than cash payments, which means that letting users pay in cash will be a better choice.

4. Women are more likely to use this service than men

More than half of the clients are Male.

5. Is the profit increasing year by year?

From 2016 to 2018, we can see that profits are decreasing year by year.