

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 13 August 2023

Internship Batch: LISUM24

Version: 1.0

Data intake by: Riwaj Neupane

Data intake reviewer: Data Glacier

Data storage location:

Cab_data 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	59.9 MB

City_data 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	4.0 KB

customer_data 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	4.0 MB

Transaction_data 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	32.3 MB

Combined_data details:

Total number of observations	359392
Total number of files	1
Total number of features	20
Base format of the file	csv
Size of the data	152.4 MB

Proposed Approach:

- No null values in the individual dataset
- Changing data type of population from str to int by removing symbol ‘,’
- Combining transaction and customer data on column ‘Customer ID’ and merging it with customer data on ‘Customer ID’
- Then combine it with cab data on column ‘City’
- In combined dataset created a new column users per meaning dividing the users by total population

Description of final dataset:

#	Column	Non-Null Count	Dtype
0	Transaction ID	359392 non-null	int64
1	Customer ID	359392 non-null	int64
2	Payment_Mode	359392 non-null	object
3	Gender	359392 non-null	object
4	Age	359392 non-null	int64
5	Income (USD/Month)	359392 non-null	int64
6	Date of Travel	359392 non-null	object
7	Company	359392 non-null	object
8	City	359392 non-null	object
9	KM Travelled	359392 non-null	float64
10	Price Charged	359392 non-null	float64
11	Cost of Trip	359392 non-null	float64
12	Population	359392 non-null	int64

13	Users	359392	non-null	int64
14	Year	359392	non-null	int64
15	Month	359392	non-null	int64
16	Day	359392	non-null	int64
17	Profit	359392	non-null	float64
18	users per	359392	non-null	float64