Data Intake Report

Name: G2M Insight for Cab Industry Project Data Intake Report
Report date: 6/03/2024
Internship Batch: LISUM34
Version: N/A
Data intake reviewer: Carmelo R. Casiraro
Data storage location:

Tabular data details: Cab_data.csv

Total number of observations	359,392 rows
Total number of files	1
Total number of features	7 columns
Base format of the file	.csv
Size of the data	20,663 KB

Tabular data details: Customer_ID.csv

Total number of observations	49,171 rows
Total number of files	1
Total number of features	4 columns
Base format of the file	.csv
Size of the data	1,027 KB

Tabular data details: Transaction_ID.csv

Total number of observations	440,098 rows
Total number of files	1
Total number of features	3 columns
Base format of the file	.csv
Size of the data	8,788 KB

Tabular data details: City.csv

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

Tabular data details: US_Holiday_Dates.csv

Total number of observations	342 rows
Total number of files	1
Total number of features	6 columns
Base format of the file	.csv
Size of the data	16 KB

Approach:

Unique Row Identification:
- All 5 data sets were merged together to create a better understanding and to perform the exploratory data analysis.
- All rows were identified by Transaction ID so I could analyze the different transactions from the passengers from each cab company.

Duplicate Rows, N/A Values:
- Any duplicate rows and N/A values were eliminated using different snippet codes.

Data Set Understanding:
- The main focus from the data collected from these data sets will be from: profits, locations, age groups, passenger retention, number of rides by cab company, payment methods, and distance of rides given to passengers by cab companies, future speculation of profits and rides, rides given during the seasonal holidays.

Assumptions:
- Data analysis is done with data sets shared and provided to me.
- Data analysis will be performed with data from the years 2016-2018.
- Data will be drawn from different sections to perform charts and cross analysis.
- The different methods of payments by passengers will be cash or credit card.
- Calculating profits are conducting by retrieving the payments paid by the passengers
- Rides and profits will increase during the seasonal holidays.
- Different cities will perform better than the others.