Data Intake Report

Name: G2M
Report date: 13/7/2022
Internship Batch: LISUM11: 30 June - 30 Sept 2022
Version: 1.0
Data intake by: Aly Medhat Moslhi
Data intake reviewer:
Data storage location: https://github.com/alymedhat10/G2M.git

Tabular data details:

Name of The File	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	20.1 MB

Name of The File	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	759 byte

Name of The File	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	1MB

Name of The File	nst-est2019-01
Total number of observations	60
Total number of files	
Total number of features	12
Base format of the file	.xlsx
Size of the data	<size in GB,TB,PB,MB etc>

Name of The File	Transaction_ID
Total number of observations	440091
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	8.85 MB

Proposed Approach:

1. **Dedup Validation:**

- There were not any null values in all five data sets.

2. **Assumptions**

- The profit is the difference between the cost of the trip and the price charged
- There is a relation between the state population and the company exposure
- The outliers were kept