Data Intake Report

Name: G2M_Cab_Insights
Report date: 13 January 2023
Internship Batch:LISUM17
Version:<1.0>
Data intake by:Surbhi Sharma
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/SurbhiS19/G2M-cab-insights

Tabular data details:
Cab Data

Total number of observations	3593927
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

City Data

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 bytes

Customer Data

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	1.1 MB

Transaction Data

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	9 MB

Proposed Approach:
- **Mention approach of dedup validation (identification):**
  After merging the data, duplicates removed based on Transaction ID. Then null values are checked, but didn't find any null values

- **Mention your assumptions (if you assume any other thing for data quality analysis):**
  The assumption for removing duplicates, that there could be only one unique id for a transaction has been considered.

Hypothesis and Visualizations has been done on the following points:
1. Data Correlation for both the cabs
2. Month-wise profit distribution of both cabs
3. Year-wise profit distribution of both cabs
4. City-wise profit distribution of both cabs
5. Annual profit rate
6. Price vs Distance and Cost vs Distance comparison
7. Gender and Year wise cab usage
8. Age and year wise can usage
9. Income wise Cab usage