Data Intake Report

Name: G2M Cab Investment firm
Report date: 22 December 2023
Internship Batch: LISUM28
Version:<1.0>
Data intake by: Vinaniek Ayu Damayanti
Data intake reviewer: https://github.com/viedamayanti/Data-Glacier-Data-Analyst-Intern/tree/main/Week2
Data storage location:

Tabular data details:

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	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

Holiday US data

Total number of observations	346
Total number of files	1

Total number of features	6
Base format of the file	csv
Size of the data	16 KB

Proposed Approach:
- The dataset contains no duplicate entries or null values.
- I have identified outliers in the 'Price Charged' column, and these outliers are also affecting the 'Margin' profit column.