Data Intake Report

Name: Cab Case Study
Report date: 12/21/2022
Internship Batch: LISUM16
Version:1.0
Data intake by: Siming Luo
Data intake reviewer:
Data storage location:

Tabular data details:

Total number of observations	359,392
Total number of files	4
Total number of features	14
Base format of the file	.csv
Size of the data	MB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:
- Mention approach of dedup validation (identification)

Before starting analyzing, I plan to check outliers, duplicated values, and null values. Also, it is crucial to do transformation if necessary.

- Mention your assumptions (if you assume any other thing for data quality analysis)

1. Do metropolitans with large population have higher average than other areas?
2. Do cities where people have high income level tend to generate higher profits?
3. Does profits have seasonality?
4. Is the number of customers positively related to the profits?
5. Which payment method people are more likely to use?