Data Intake Report

Name: G2M insight for Cab Investment firm Report date: 13/08/2024

Internship Batch: LISUM36 Version:1.0

Data intake by: SREEDHAR RONGALA

Data intake reviewer:

Data storage location: <https://github.com/sreedharsiddhu/Data-Glacier/tree/main/week%202>

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 497913 |
| **Total number of files** | 4 |
| **Total number of features** | 33 |
| **Base format of the file** | csv |
| **Size of the data** | 107.68 MB |

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

|  |  |
| --- | --- |
| **Proposed Approach** | **Assumptions** |
| Created a master dataset by merging all individual datasets, including transaction, customer, weather, holiday, city,  population, and cab data. | Assumed that all datasets were accurately and consistently formatted, allowing for correct  merging. |
| Ensured that all relevant features were included in the master dataset for comprehensive analysis. | Assumed that the combined dataset retained all  essential data without significant loss during the merge process. |
| Conducted feature engineering to create new variables such as weather impact, holiday flag, repeat customer flag, and season identifier. | Assumed that the engineered features added value and correctly captured the intended aspects of the data. |
| Handled missing data by imputing, filling with appropriate values, or excluding as necessary. | Assumed that missing data was minimal and could be reasonably imputed without significant impact on analysis results. |
| Hypothesis were tested for:   1. Market Share and Profitability: 2. Invest in Yellow Cab 3. Price Sensitivity 4. Customer Retention 5. Demographic Targeting 6. Seasonal Strategy 7. Weather Impact | All datasets were cleaned and features were engineered before testing for the hypothesis |