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**](https://github.com/sreedharsiddhu/Data-Glacier/tree/main/week%202)

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 342 |
| **Total number of files** | 1 |
| **Total number of features** | 6 |
| **Base format of the file** | csv |
| **Size of the data** | 15.33 KB |

|  |  |
| --- | --- |
| **Proposed Approach** | **Assumptions** |
| Ensured correct alignment of holiday dates across years. | Assumed that holidays had a consistent impact on travel patterns. |
| Added a binary flag to indicate whether each date was a holiday or not. | Assumed that the holidays included were significant enough to influence cab usage. |
| Merged holiday data with the master dataset to analyze its effect on cab usage. | Assumed that holiday data was comprehensive and included all relevant holidays. |
| Verified and corrected date formats to ensure consistency. | Assumed that holidays occurred on the same dates each year. |
| Holiday data was obtained from here: [https://www.kaggle.com/datasets/donnetew/us-](https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021) [holiday-dates-2004-2021](https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021) |  |
| Ensured data were in the range of the timelines provided in the problem overview |  |
| Performed feature engineering |  |