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** | 20 |
| **Total number of files** | 4 |
| **Total number of features** | 7 |
| **Base format of the file** | csv |
| **Size of the data** | 32.22 KB |

|  |  |
| --- | --- |
| **Proposed Approach** | **Assumptions** |
| Handled missing population data with interpolation or average growth rate estimates. | Assumed that population growth was steady unless otherwise indicated. |
| Standardized city names across all datasets. | Assumed that missing data was minimal and could be accurately estimated. |
| Calculated average growth rates for cities over multiple years. | Assumed that yearly population estimates were accurate for calculating growth rates. |
| Records for each cities was manually searched for on google | Data provided by google is accurate or was approved before being published for public consumption |
| Performed feature engineering |  |