Data Intake Report

Name: Case Study 1

Report date: 14 July 2022

Internship Batch: LISUM11

Version:1.0

Data intake by: Anshi Mathur

Data intake reviewer: Anshi Mathur

Data storage location: Github (https://github.com/anshimathur0325/CaseStudy1.git)

**Tabular data details:**

Cab\_Data.csv

|  |  |
| --- | --- |
| **Total number of observations** | 359393 |
| **Total number of files** | 1 |
| **Total number of features** | 8 |
| **Base format of the file** | .csv |
| **Size of the data** | 24713 KB |

City.csv

|  |  |
| --- | --- |
| **Total number of observations** | 21 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 1 KB |

Customer\_ID.csv

|  |  |
| --- | --- |
| **Total number of observations** | 49172 |
| **Total number of files** | 1 |
| **Total number of features** | 4 |
| **Base format of the file** | .csv |
| **Size of the data** | 1027 KB |

Transaction\_ID.csv

|  |  |
| --- | --- |
| **Total number of observations** | 440099 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 8788 KB |

**Proposed Approach:**

* From the original data, I first searched for data that had missing values. Then I was able to combine data by sorting each file from the smallest to largest transaction ID and then combining columns for ease of data analysis. Finally, I added a column to account for the profit for each company of each transaction.
* I assumed that the transaction IDs collected throughout these data sets are unique to the user, which allowed me to clean up data and make it look more polished.