Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 8/14/2024

Internship Batch: LISUM36

Version:<1.0>

Data intake by: Shuyi Yu

Data intake reviewer: Shuyi Yu

Data storage location: <location URL eg: github, cloud>

**Tabular data details: Cab\_Data.csv**

**Total number of observations**

359392

**Total number of files**

1

**Total number of features**

7

**Base format of the file**

csv

**Size of the data**

21.2MB

**Tabular data details: Customer\_ID.csv**

**Total number of observations**

49171

**Total number of files**

1

**Total number of features**

4

**Base format of the file**

csv

**Size of the data**

1.1MB

**Tabular data details:Transaction\_ID.csv**

**Total number of observations**

<number of rows>

**Total number of files**

1

**Total number of features**

<Number of columns>

**Base format of the file**

<.csv,.txt etc>

**Size of the data**

9MB

**Tabular data details: City.csv**

**Total number of observations**

20

**Total number of files**

1

**Total number of features**

3

**Base format of the file**

csv

**Size of the data**

4KB

**Tabular data details: us\_holiday\_dates.csv**

**Total number of observations**

342

**Total number of files**

6

**Total number of features**

3

**Base format of the file**

csv

**Size of the data**

16KB

**Proposed Approach:**

• Mention approach of dedup validation (identification)

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