

# Data Intake Report

Name: G2M insight for Cab Investment firm  
Report date: 14.03.2023  
Internship Batch: LISUM19  
Version:<1.0>  
Data intake by:Emirhan SARP  
Data intake reviewer:  
Data storage location: github.com/data\_glacier

## Tabular data details:

### Cab Data

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	

### Customer Data

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	

### Transaction Data

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	

### City Data

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	

**Proposed Approach:**

- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)

1. Yellow Cab makes more profit
2. Cab industry earns the most in new York
3. Men travel by taxi more than women
4. People pay by card rather than cash
5. Men pay more for taxis than women

**Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.**

**Please do not forget to remove this section while converting the file into pdf.**