

# Data Intake Report

Name: G2Mstudycase

Report date: 9/3/2021

Internship Batch: LISUM02

Version:<1.0>

Data intake by: Amna Naeem

Data intake reviewer: Amna Naeem

Data storage location: [DataGlacier/DataSets \(github.com\)](https://github.com/DataGlacier/DataSets)

## Tabular data details:

Cab\_data:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	22.088 MB

City\_data:

<b>Total number of observations</b>	20
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1KB

Customer\_ID:

<b>Total number of observations</b>	49171
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1027KB

Transaction\_ID:

<b>Total number of observations</b>	440098
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8788KB

**Proposed Approach:**

- Hypothesis testing, Exploratory Data Analysis and Time prediction using ARIMA

**Hypothesis statement: Both Cabs yellow and pink are equal in terms of preference and profit.**