# Data Intake Report

## Name

Exploratory Data Analysis

## Report date

February 8, 2022

## Internship Batch

LISUM06

## Version

<1.0>

## Data intake by

Abdur Rehman Shahid

## Data intake reviewer

<intern who reviewed the report>

## Data storage location

https://github.com/razashahid107/Data-Glacier-Internship

### 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** | 20663 KB |

### Tabular data details: city.csv

|  |  |
| --- | --- |
| **Total number of observations** | 19 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 1 KB |

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

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 1027 KB |

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

|  |  |
| --- | --- |
| **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** | 8788 KB |

### Tabular data details: new\_csv.csv

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 1 |
| **Total number of features** | 8 |
| **Base format of the file** | .csv |
| **Size of the data** | 24345 KB |

### Tabular data details: new2\_csv.csv

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 1 |
| **Total number of features** | 12 |
| **Base format of the file** | .csv |
| **Size of the data** | 29449 KB |

**Note: Replicate same table with file name if you have more than one file.**

## Proposed Approach:

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

The Data provided was in different files and I was not aware of methods via which I could’ve accessed and linked data from different excel files. I therefore used Data Frames approach where I created new excel files named new\_csv.csv and new2\_csv.csv. I tried combining data in these files and then assigned it to df\_3 in my EDA. Via this I tried creating graphs which indicated things like Profit according to age, Profit according to city, Average age of users and numerous other graphs mentioned in my EDA.

**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.**