Data Intake Report

Name: Cab Investment Analysis

Report date: 15 June 2021

Internship Batch: LISUM01

Version: 0.1

Data intake by: Runa Veigas

Data intake reviewer: Runa Veigas

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

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 4 |
| **Total number of features** | 19 |
| **Base format of the file** | csv |
| **Size of the data** | 44,5894KB |

**Proposed Approach:**

* As this a dataset that contains information of cab users, it can be possible that one user will take the cab more than one time thus the Customer ID will be repeated. A Transaction ID though should be unique as every transaction will generate a unique ID. Thus duplicates were checked for Transaction ID and there were no duplicates found.
* **Assumptions**:

1. Data provided is a random sample that truly represents the population.
2. There are only two cab service providers across the country.
3. There is no other expense from the cab company other than the Cost Price given in the data set, which is used to compute the profit.