Data Intake Report

Name: <EDA CabData>

Report date: <13 April 2022>

Internship Batch:<LISUM08::30 >

Version:<1.0>

Data intake by:<Apurva Shah>

Data intake reviewer:<intern who reviewed the report>

Data storage location: < https://github.com/lazarulian/EDA-CabCompanyGrowth>

**Tabular data details: Cab Data**

|  |  |
| --- | --- |
| **Total number of observations** | <359391 > |
| **Total number of files** | <1> |
| **Total number of features** | <7> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <24.7mb> |
|  |  |

**Tabular data details: 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** | <608bytes> |
|  |  |

**Tabular data details: CustomerID**

|  |  |
| --- | --- |
| **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** | <1.5mb> |
|  |  |

**Tabular data details: Cab 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** | <10.1mb> |
|  |  |

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

**Proposed Approach:**

* Initially, I am going to use the EDA guidelines in the google pdf to identify particular outliers in the data set
* Using this data, I will clean and merge it together to make one master data file
* I will go through the data and create graphs to identify correlations between variables and compare the two cab companies.
* My Interests are documented in my notebook