Data Intake Report

Name: <G2M>

Report date: <2022-02-06>

Internship Batch:<LISUM06(26)>

Version:<1.0>

Data intake by:<Mengyu Ding>

Data intake reviewer:<Hongjian Yuan>

Data storage location: < https://github.com/Obliviousdd/data-glacier >

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | <359392> |
| **Total number of files** | <4> |
| **Total number of features** | <14> |
| **Base format of the file** | <csv> |
| **Size of the data** | <43.8MB> |

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

**Proposed Approach:**

* Mention approach of dedupe validation (identification)

I used *pandas. duplicated* to find duplicates in customer ID to make validate the demographic information.

* Mention your assumptions (if you assume any other thing for data quality analysis)
* #yellow users have more cabs during specific date
* #users and margin have positive relation
* #men are more likely to take a cab
* #age between 20-40 are the most users
* #km travel ranges are pretty average

**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.**