Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: Jun 20, 2021

Internship Batch: LISUM01

Version:1.0

Data intake by: Jingxi Liao

Data intake reviewer: Jingxi Liao

Data storage location: https://github.com/BananaPea/VC.git

**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** | 21.2 MB, 759 B, 1.05 MB, 9 MB |

**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)
* Mainly use EDA to have data visualization and statistical hypothesis test for data analysis
* Used seaborn, matplotlib, pandas, numpy, datetime library
* Data date time period is from 02/01/2016 to 31/12/2018 instead of 31/01/2016 to 31/12/2018 because of 1094 days duration.

**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.**