Data Intake Report

Name: <Go to Market Strategy>

Report date: <7/09/2021>

Internship Batch:<LISUM01>

Version:<1.0>

Data intake by:<Anye Igwacho>

Data intake reviewer:<Data Glacier>

Data storage location: <https://github.com/DataGlacier/DataSets>

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | <~250000> |
| **Total number of files** | <5> |
| **Total number of features** | <19> |
| **Base format of the file** | <.csv> |
| **Size of the data** | <30MB> |

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

**Proposed Approach:**

* Mention approach of dedup validation (identification)

Dedup in MATLAB, python, rows with NAN and empty cells were deleted

Normalized features, format strings