

# Data Intake Report

Name: week 7 Group Project advance

Report date: 02/05/2022

Internship Batch: LISUM05

Version: 1.0

Data intake by: Christopher Irvin Ballon Peralta

Data intake reviewer: Data Glacier

Data storage location: <https://drive.google.com/file/d/16-nzZR91-ijrfjUcI2PniTpOgrvFAykA/view>

## Tabular data details:

Train.csv

<b>Total number of observations</b>	13647309
<b>Total number of files</b>	13647311
<b>Total number of features</b>	48
<b>Base format of the file</b>	csv
<b>Size of the data</b>	2.239 GB

## Proposed Approach:

- It's necessary to calculate Date field Time from last transaction.
- Trim fields, convert to capital, drop NA values, sort string fields.
- Recode products to reduce dimensionality

Test.csv

<b>Total number of observations</b>	929615
<b>Total number of files</b>	929617
<b>Total number of features</b>	48
<b>Base format of the file</b>	csv
<b>Size of the data</b>	107.7 MB

## Proposed Approach:

- It's necessary to calculate Date field Time from last transaction.
- Trim fields, convert to capital, drop NA values, sort string fields.
- Recode products to reduce dimensionality