

Data Intake Report

Name: Cross Selling Data Analysis

Report date: October 30th, 2022

Internship Batch: LISUM12

Version:<1.0>

Data intake by: Richard Coltenback

Data intake reviewer: None

Data storage location: <https://github.com/RColtenback/Cross-Selling-Data-Analysis.git>

Tabular data details:

Train.csv

Total number of observations	13647309
Total number of files	1
Total number of features	48
Base format of the file	.csv
Size of the data	2.29 GB

Proposed Approach:

- Column names will have to be changed to make the variables easier to understand.
- There will be lots of good data points since there are over 13 million rows of data. I might need to use dask or modin to analyze the dataset since it may take too long with pandas.