

Data Intake Report

Name: Bank Marketing Data Science Project

Report date: 26 April 2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Suvansh Vaid, Ines Perko, Zeynep Başak Eken

Group Name: Bank Marketing DS 01

Data intake reviewer:

Data storage location: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

Tabular data details:

Total number of observations	45211
Total number of files	1
Total number of features	17
Base format of the file	Comma separated values (csv)
Size of the data	4.6 Mb

Proposed Approach:

- Would create a 2 stage pipeline for file ingestion and schema validation
- Would use a yaml configuration file to validate any duplicate values, incorrect column names, etc.