

Data Intake Report

Name: File ingestion and schema validation

Report date: 10 February 2023

Internship Batch:LISUM17

Version:<1.0>

Data intake by:Surbhi Sharma

Data intake reviewer:<intern who reviewed the report>

Data storage location: https://www.kaggle.com/datasets/new-york-city/nyc-parking-tickets?select=Parking_Violations_Issued_-_Fiscal_Year_2017.csv

Tabular data details:

Parking violation ticket issued in New York in 2017

Total number of observations	10803028
Total number of files	1
Total number of features	43
Base format of the file	.csv
Size of the data	2.09 GB

Proposed Approach:

- Different approaches of reading the data has been tested
- YML file of the data is created
- GZIP file is created