

Data Intake Report

Name: Used Cars Data File Ingestion

Report date: November 12, 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Justin Lee

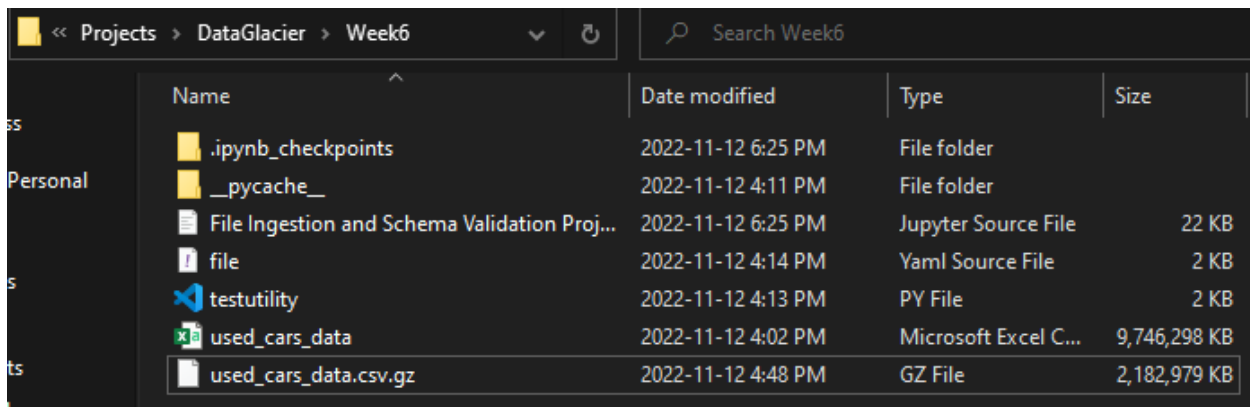
Data intake reviewer:

Data storage location: <https://www.kaggle.com/datasets/ananyamital/us-used-cars-dataset>

Tabular data details:

Total number of observations	3000040
Total number of files	<Number of files received>
Total number of features	66
Base format of the file	.csv
Size of the data	9.76 GB

The File size is too large to upload to GitHub, but the code will prove that I have, as well as screenshots of my file system is provided below.



« Projects » DataGlacier » Week6		Search Week6	
Name	Date modified	Type	Size
.ipynb_checkpoints	2022-11-12 6:25 PM	File folder	
__pycache__	2022-11-12 4:11 PM	File folder	
File Ingestion and Schema Validation Proj...	2022-11-12 6:25 PM	Jupyter Source File	22 KB
file	2022-11-12 4:14 PM	Yaml Source File	2 KB
testutility	2022-11-12 4:13 PM	PY File	2 KB
used_cars_data	2022-11-12 4:02 PM	Microsoft Excel C...	9,746,298 KB
used_cars_data.csv.gz	2022-11-12 4:48 PM	GZ File	2,182,979 KB