

Maanya Medapa
DSCI 510
Professor Tregubov
November 13, 2025

Final Project Progress Report

Project scope update

The primary objective of this project has not changed. We will examine the relationship between public transit accessibility, neighborhood-level characteristics, and environmental and social equity outcomes in Los Angeles County. The primary change involves building a clean data pipeline before integrating all the datasets.

For this project submission, the project intentionally focuses on implementing a working API connection, establishing a processing workflow, creating the full project folder structure, and writing automated tests

Data sources

Source: Los Angeles County using the U.S. Census American Community Survey (ACS) 2023 dataset

Variables: Median household income, Total population, Households without a vehicle, Census tract GEOID

URL: <https://api.census.gov/data/2023/acs/acss5> . No API key was required, as the database is fully public.

Implemented in `src/load.py` , it returns JSON data that is converted into a pandas DataFrame, after querying all census tracts in Los Angeles County.

Issues/difficulties

The first difficulty was accessing GitHub via my terminal with a personal token, as password authentication was not working.

I also struggled with creating accidental extra files in my repository. I expect challenges when integrating LA Metro GTFS data, since it requires parsing transit stop times and routes. I also might need additional Python packages to assess the tract-to-stop distance using spatial analysis.