Progress Report

Final Project Progress Report
Due: Thursday November 13, 2025

Progress report document structure:
Project scope update: *describe if anything changed since you started working on the project*

The overall project goal remains the same — to explore whether proximity to metro stations is associated with shorter average commute times in Los Angeles County. The main change since the initial proposal is a refinement in spatial resolution: instead of comparing data by county, the analysis will focus on ZIP code–level data, as this provides finer granularity and better reflects local differences in metro accessibility. Additionally, I plan to include median household income as a control variable in the regression model to account for socioeconomic factors that may influence commute behavior. This is now being pulled from the same data source, if this is an issue (now that I only have two sources), I can come up with another data source.

Data sources: Describe data obtained in your python code & describe API you used

- API Data: Accessed the U.S. Census Bureau's American Community Survey (ACS) API using Python. The current script retrieves:
    - *Mean Travel Time to Work (B08303_001E)*
    - *Median Household Income (B19013_001E)*
    - *Geographic Name (NAME)*
        - Data is collected for all ZIP Code Tabulation Areas (ZCTAs) in Los Angeles County.
    - API endpoint: https://api.census.gov/data/2022/acs/acs5
- LA Metro Stop File Data:
    - Metro stop locations shapefile from the Los Angeles Metro GIS Data Portal. This will be used to count the number of metro stations within each ZIP code boundary once integrated.

Issues/difficulties: *Describe any issues you faced so far or what potential issues you expect.*
Overall, API access is functioning. The next steps include integrating the shapefile data, performing spatial joins, and starting exploratory visualizations.

- Geographic Matching: The main upcoming challenge is spatially joining the metro station shapefile with ZIP code boundaries to count nearby stations per area. This will require using additional libraries.
- Data Consistency: ACS and Metro GIS data use different geographic units (ZCTA vs. city limits), so I will need to ensure consistent boundaries for accurate comparison.