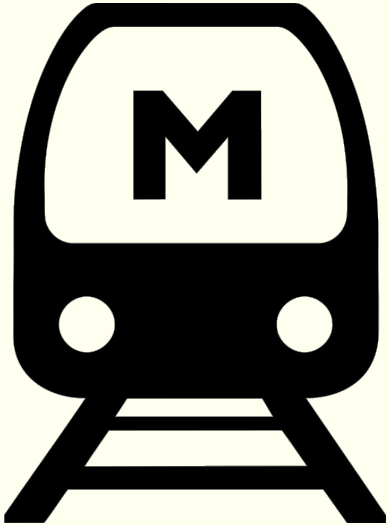
Evaluating the Relationship Between Metro Station Access and Commute Times in Los Angeles County

Vanessa Garcia
DSCI 510 Final Project



PROJECT INTRODUCTION

This project analyzes commute times at the zip code level in Los Angeles County using 2022 data. The goal is to evaluate whether ZIP codes with Metro rail stations experience shorter average commute times than those without a metro rail station.

To conduct this analysis, the project integrates multiple datasets:

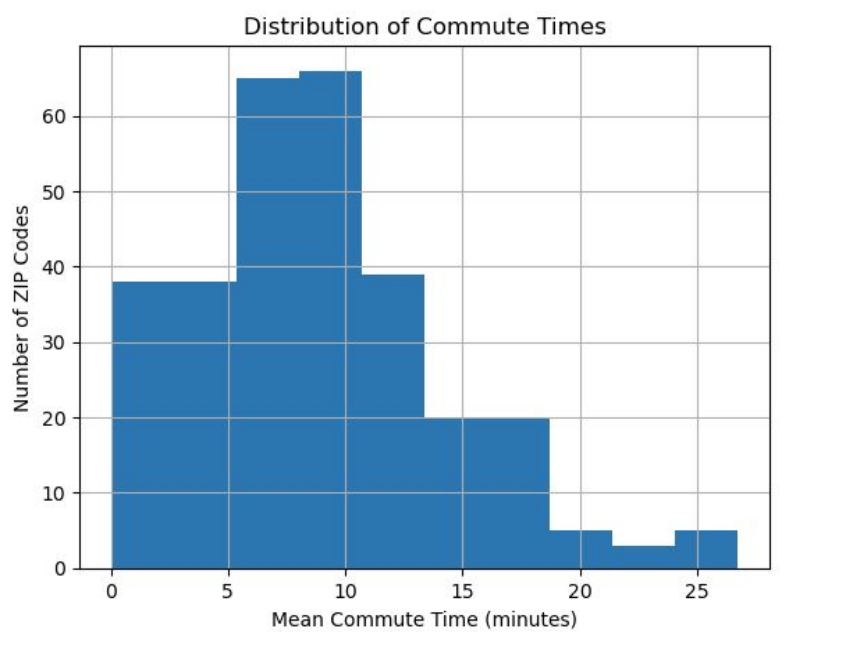
- LA Metro rail station shapefiles
- TIGER/Line ZCTA boundary shapefiles
- The Census ZCTA-to-County crosswalk
- American Community Survey 2022 API data
    a. Median household income
    b. Mean commute time

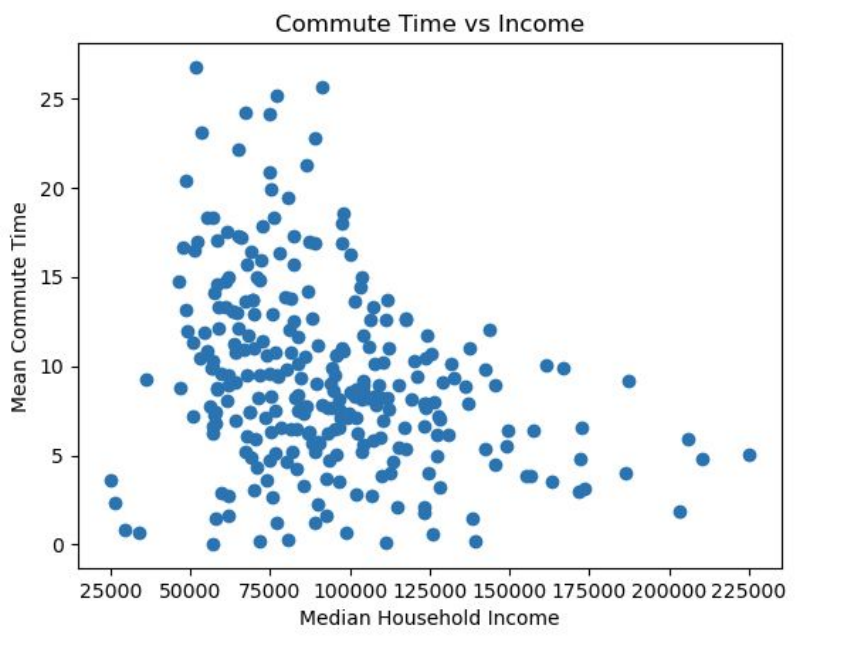
Data Sources

Data Source	Description	Processing & Data Cleaning Approach
LA Metro - Metro Rail Stations Shapefile	Shapefile that provides the latitude and longitude of all Metro Stations in Los Angeles County.	Loaded using GeoPandas; added lat/lon fields; reprojected to match the TIGER ZCTA CRS; spatially joined to assign each station to a ZIP Code Tabulation Area (ZCTA).
American Community Survey (ACS) 2022, 5-Year Estimates	Provides ZIP-code–level socioeconomic and transportation data, including: • Median Household Income (B19013_001E) • Mean Travel Time to Work (S0801_C02_001E)	Retrieved using the Census API and a personal API key. Cleaned numeric fields, fixed scaling, and merged with the ZIP-level station dataset.
US Census Bureau – ZCTA-to-County Crosswalk (CSV)	Identifies what ZIP Code Tabulation Areas belong to which county.	Downloaded directly. Filtered to STATE = “06” (CA) and COUNTY = “037” (Los Angeles). Produced a complete list of LA County ZIP codes before pulling ACS data.
US Census Bureau TIGER/Line 2020 Shapefiles	Shapefile containing the geographic boundaries of all U.S. ZIP Code Tabulation Areas.	Loaded into GeoPandas. Used as polygons for spatial joins. Ensured CRS alignment with Metro station points to assign each station to the correct ZCTA.

Descriptive Statistics

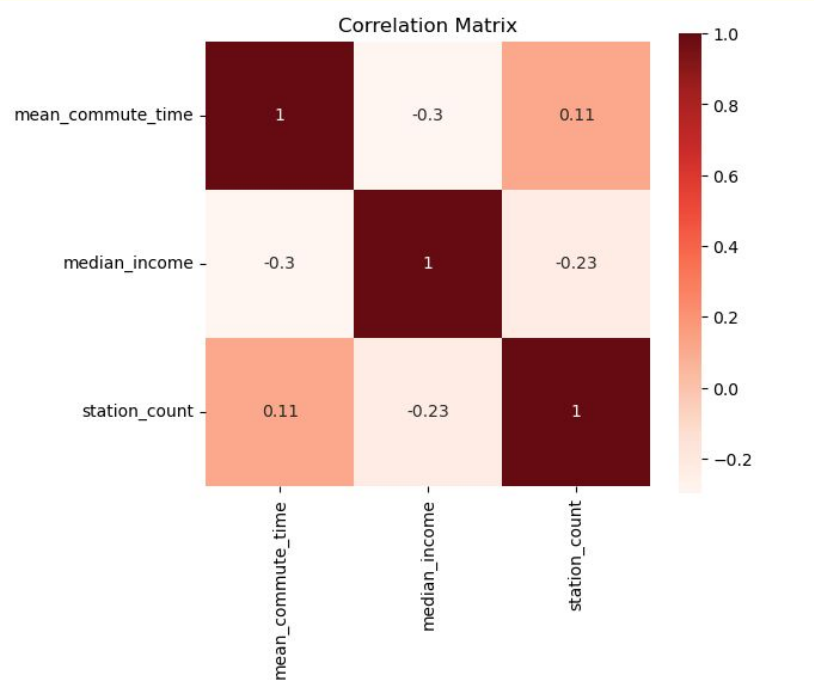
- ~300 ZIP codes included across Los Angeles County
- Includes commute time, median income, and station count per ZIP code
- Average commute time: ~ 9 minutes
- Mean median household income: ~ $92K



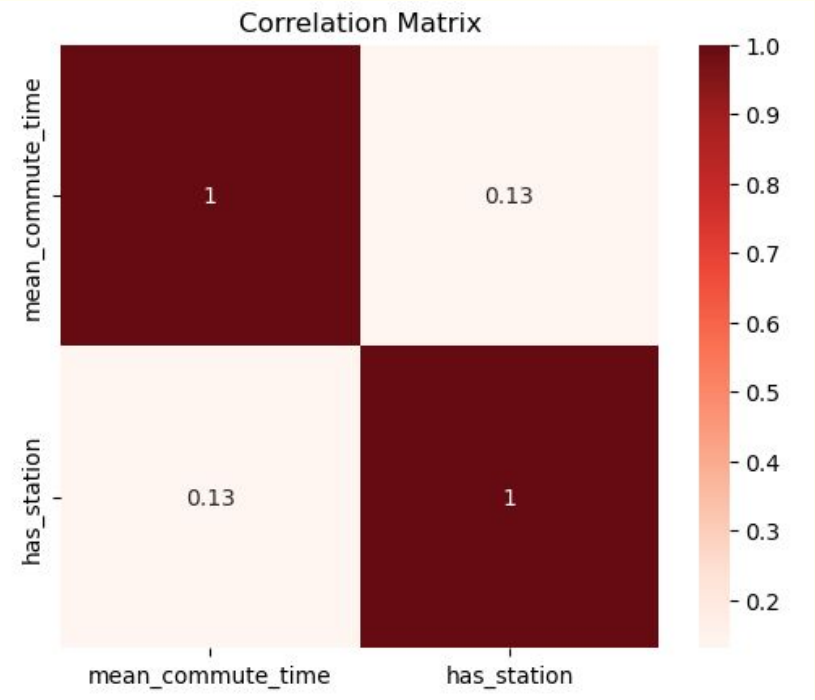


Correlation Analysis

Correlation Between Commute Time, Income, and Metro Station Presence (ZIP-Level)



Correlation Between Metro Access and Commute Times



REGRESSION RESULTS: Do Metro Stations Reduce Commute Time?

OLS Regression Results

Dep. Variable:	mean_commute_time	R-squared:	0.090
Model:	OLS	Adj. R-squared:	0.083
Method:	Least Squares	F-statistic:	13.77
Date:	Mon, 24 Nov 2025	Prob (F-statistic):	1.97e-06
Time:	21:35:48	Log-Likelihood:	-852.72
No. Observations:	283	AIC:	1711.
Df Residuals:	280	BIC:	1722.
Df Model:	2		
Covariance Type:	nonrobust		

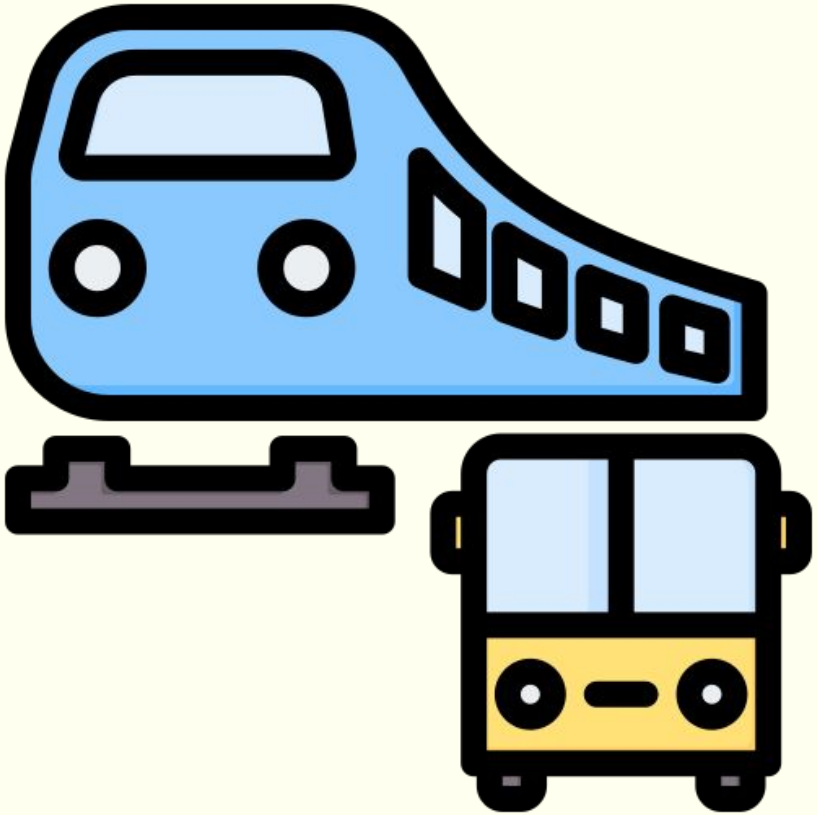
	coef	std err	t	P>|t|	[0.025	0.975]
const	13.5538	0.936	14.480	0.000	11.711	15.396
median_income	-4.621e-05	9.15e-06	-5.049	0.000	-6.42e-05	-2.82e-05
station_count	0.0738	0.348	0.212	0.832	-0.611	0.759

Omnibus:	17.401	Durbin-Watson:	1.683
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20.113
Skew:	0.530	Prob(JB):	4.29e-05
Kurtosis:	3.762	Cond. No.	3.16e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.16e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Key Findings:

- Income is strongly related to commute time.
- Station count has no meaningful effect.
- Overall Model Fit
  - R-squared is 0.09 > the model explains about 9% of ZIP-level commute time differences, which is typical for neighborhood-level social data
  - The significant part of the model is income
  - Station count adds no explanatory power



Conclusion

After accounting for income and station density, the data suggests that Metro station presence has little to no measurable impact on average commute times across LA ZIP codes.

Challenges

Considerations for future research

- Expand geography:
  - Patterns of people moving out East to San Bernardino County, for example, it would be beneficial to expand.
- Include more explanatory variables.

Questions?