Where Is the Next Whole Foods in Los Angeles County?

Spring 2023 - ECON 584 - Group 9

Amanda Rago, Mohammed Alshlash, Micky Sun, Wanlin Chen, Jianian Hua, Peiyao Li

Background and Purpose

- Whole Foods is a chain of grocery stores that focuses on selling natural, organic, and 'healthy' food
- Questions:
  - What factors go into Whole Foods' decision when opening a new store?
  - What neighborhood would be best suited for a new Whole Foods?
- Utilized a logistic regression model to understand what factors were correlated with Whole Foods existence in a location
  - Took these results and forecasted future locations

Study Area

- Originally analyzed at the zip code level in Los Angeles County
- Pivoted to major neighborhoods and cities
  - Sherman Oaks, Studio City, Santa Monica, etc.
- Analyzed 116 major neighborhoods and cities; 29 Whole Foods total
- Average number of Whole Foods is 0.25
- Minimum is 0, maximum is 9, standard deviation is 0.98
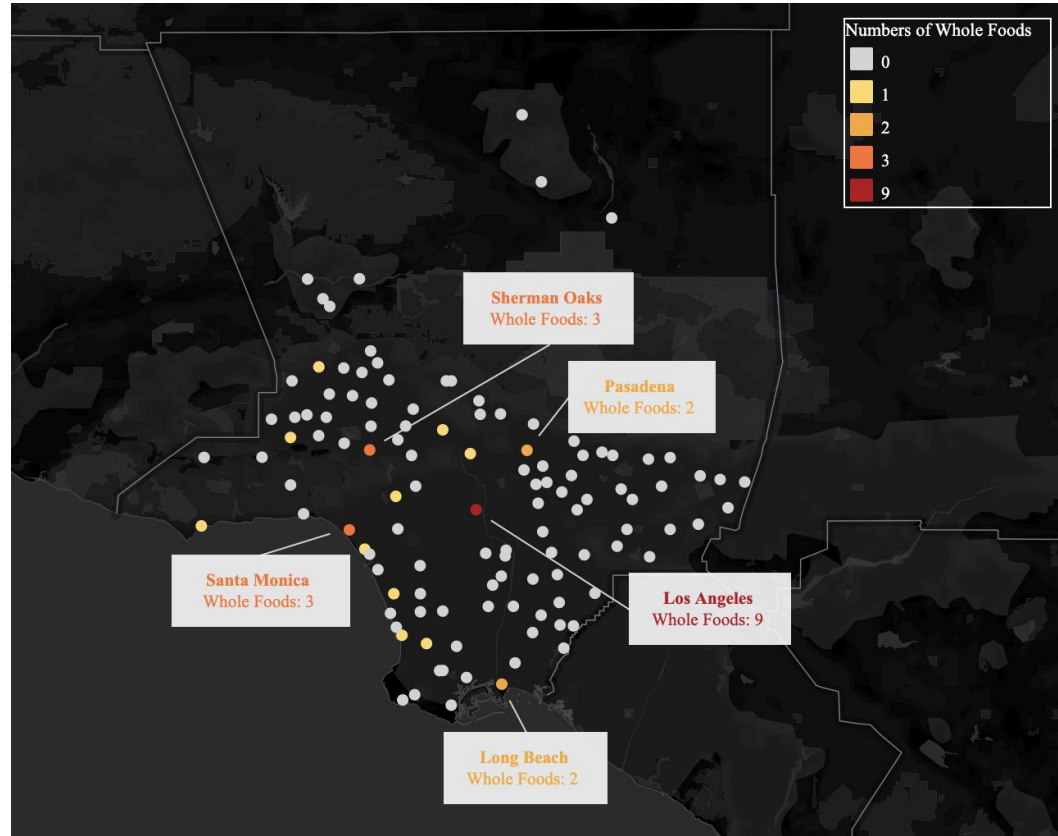
Literature Review

Theme 1: Demographic factors are significant for retail stores' success

- Income level and population size are important in marketing analyzation (Assessment of business potential at retail sites: empirical findings from a US supermarket chain)

- Food deserts has high correlation with ethnicity, worsening the health problems of residents in South Dallas.
  (Demand Metric for Supermarket Site Selection: A Case Study)

Literature Review

Theme 2: Spatial factors affect retailers' decision making of opening new grocery stores

- Perception of location deeply affects retailers to open retail stores
  (Demand Metric for Supermarket Site Selection: A Case Study)

- Spatial factors like land price affect the decision making
  (Geomarketing models in supermarket location strategies)

Literature Review

Theme 3: Competitors' characteristics play two-side role in the assessment process of the possible implementation of grocery stores

- Accessibility and size of competitive stores affect consumer purchasing decisions
  (Assessment of business potential at retail sites: empirical findings from a US supermarket chain)

- Lack of competitors intensifies "food deserts"
 (Demand Metric for Supermarket Site Selection: A Case Study)

Data

- Dependent variable: Whole Foods dummy

- Four categories are comprised of 12 independent variables: competitors level, demographic information, socioeconomic condition, and community safety and accessibility.

- Position strategy:
  - Whole Foods and Trader Joe's both offer a unique shopping experience that caters to health-conscious consumers.
  - Ralphs and Trader Joe's both offer lower prices than Whole Foods, which appeal to budget-conscious consumers.

- Cross-sectional data is examined spatially. (116 observations: city/neighborhood)

Variable	Description	Minimum	Mean	StDev	Maximum
Dependent Variable					
Whole Foods	Binary variable: whether a city/neighborhood has a Whole Foods or not, 2023	0	0.25	0.98	9
Competitors Level					
Ralphs	Number of Ralphs in a city/neighborhood, 2023	0	0.61	2.15	22
Trader Joes	Number of Trader Joes in a city/neighborhood, 2023	0	0.41	1.29	12
Demographic Information					
Population	Total population in a city/neighborhood, 2021	5,114	85,579	226,351	2,420,705
Age	Median age in a city/neighborhood	31	40	4	51
Education	Percentage of population with bachelor's degrees or higher, 2020	4.20%	35.09%	19.71%	78.54%
Socioeconomic Condition					
Households Income	Households median income in a city/neighborhood, 2021	$49,554	$93,458	$31,766	$212,115
Unemployment Rate	Unemployment rate in a city/neighborhood	2.85	6.48	1.47	9.92
Housing Price	Average housing price in a city/neighborhood	$89,705	$1,175,492	$769,734	$4,623,281
Apartment Price	Average apartment price in a city/neighborhood	$1,942	$2,439	$439	$3,244
Community Safety and Accessibility					
Violent Crimes	Number of violent crimes in a city/neighborhood	14	28,838	151,968	1,632,191
Traffic Volume	Average annual daily traffic of freeway/highway segment in a city/neighborhood, 2016	28,000	2,157,306	8,037,348	86,666,500
Public Transit Ridership	Percentage of population using public transit service in a city/neighborhood, 2019	0.40%	3.71%	2.68%	13%

Data (Competitors Level)
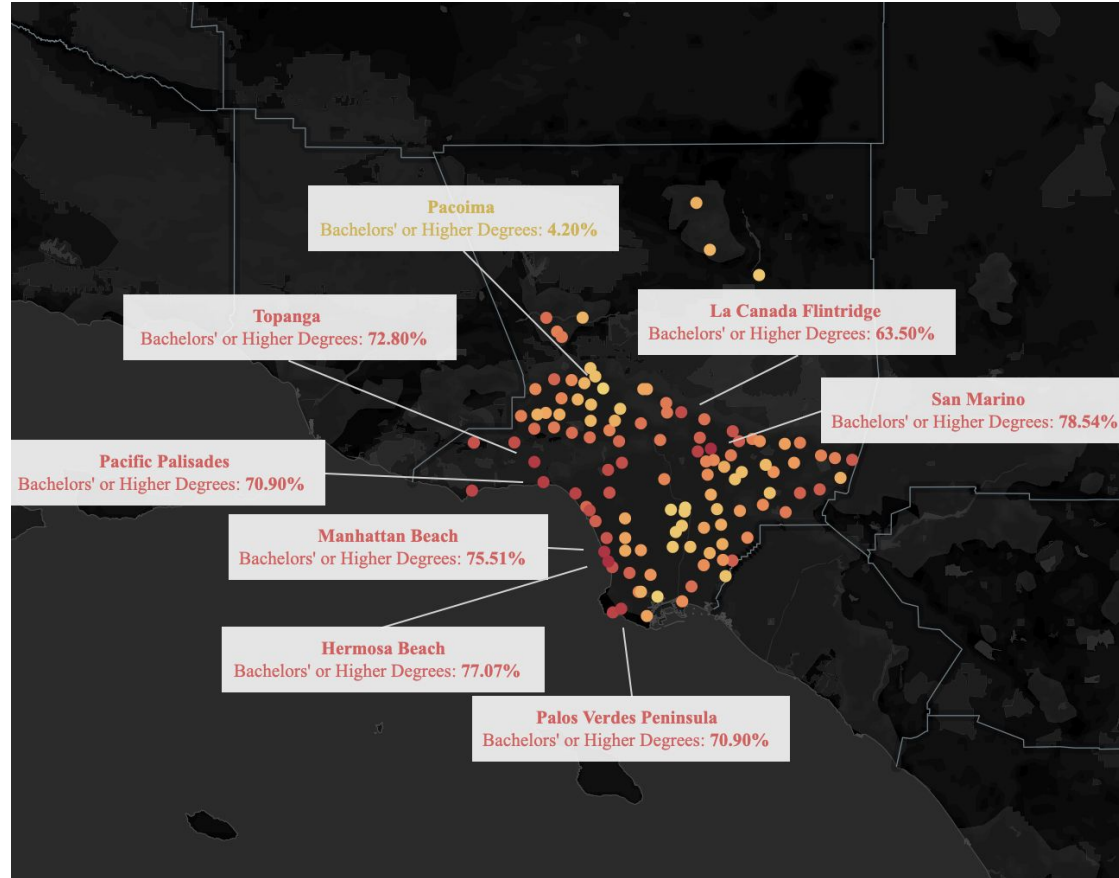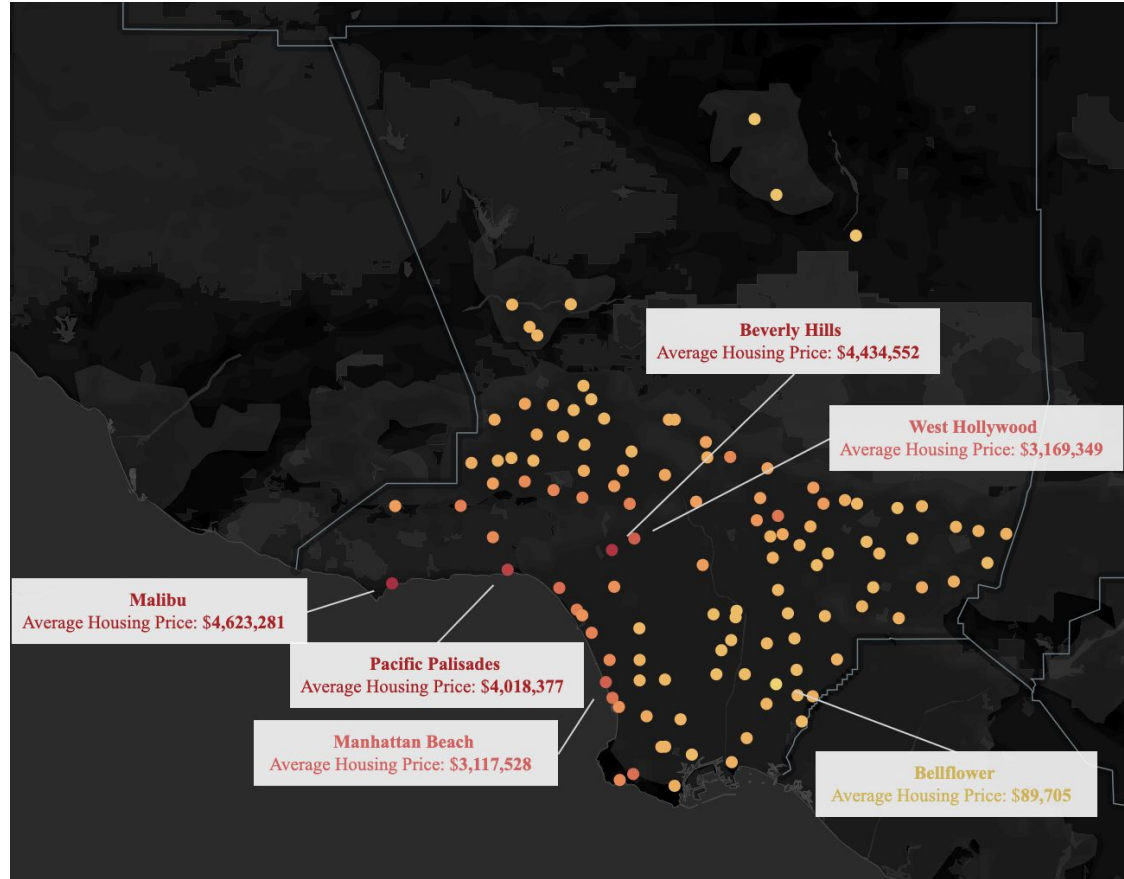
Trader Joes (48)
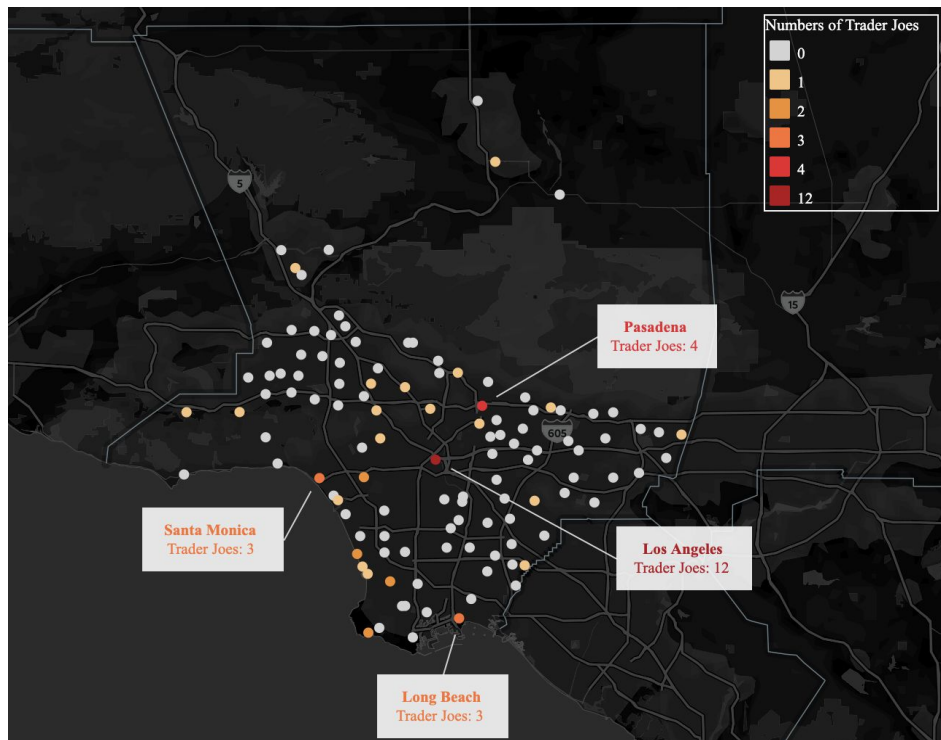
Ralphs (71)
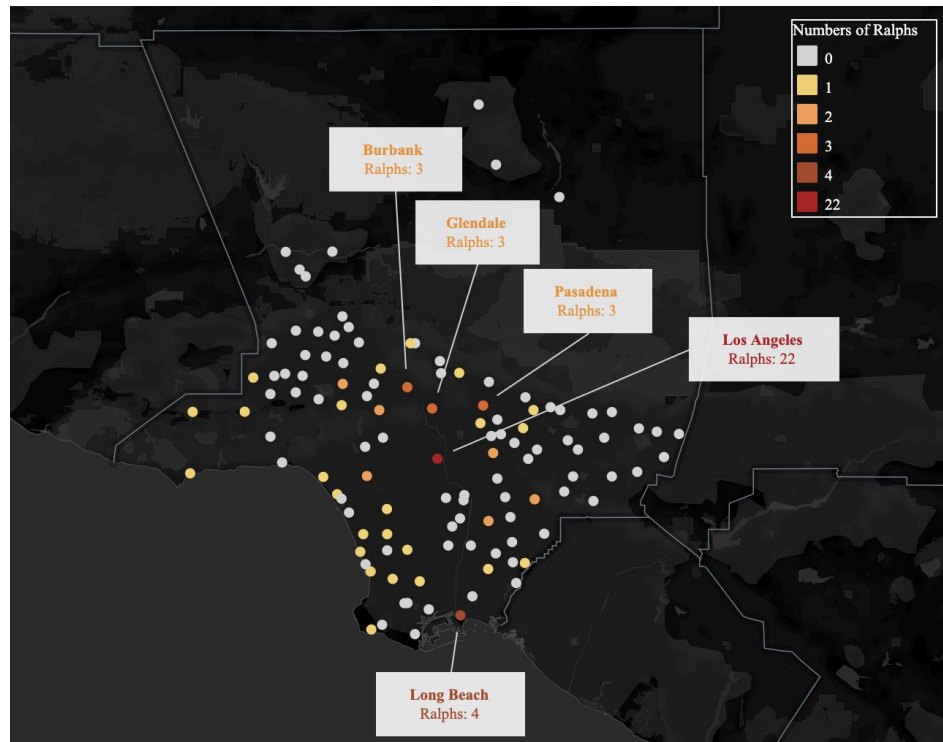
Data (Demographic Information)

- % Population with bachelor's degree or higher:
 - Higher in Coastal cities

Data (Socioeconomic Condition)

- Average housing prices:
  - Coastal cities are more expensive

Beverly Hills
Average Housing Price: $4,434,552

West Hollywood
Average Housing Price: $3,169,349

Malibu
Average Housing Price: $4,623,281

Pacific Palisades
Average Housing Price: $4,018,377

Manhattan Beach
Average Housing Price: $3,117,528

Bellflower
Average Housing Price: $89,705

Data (Community Safety and Accessibility)

- Incidence of violent crimes:
  - Cluster in the northwestern part of Los Angeles county
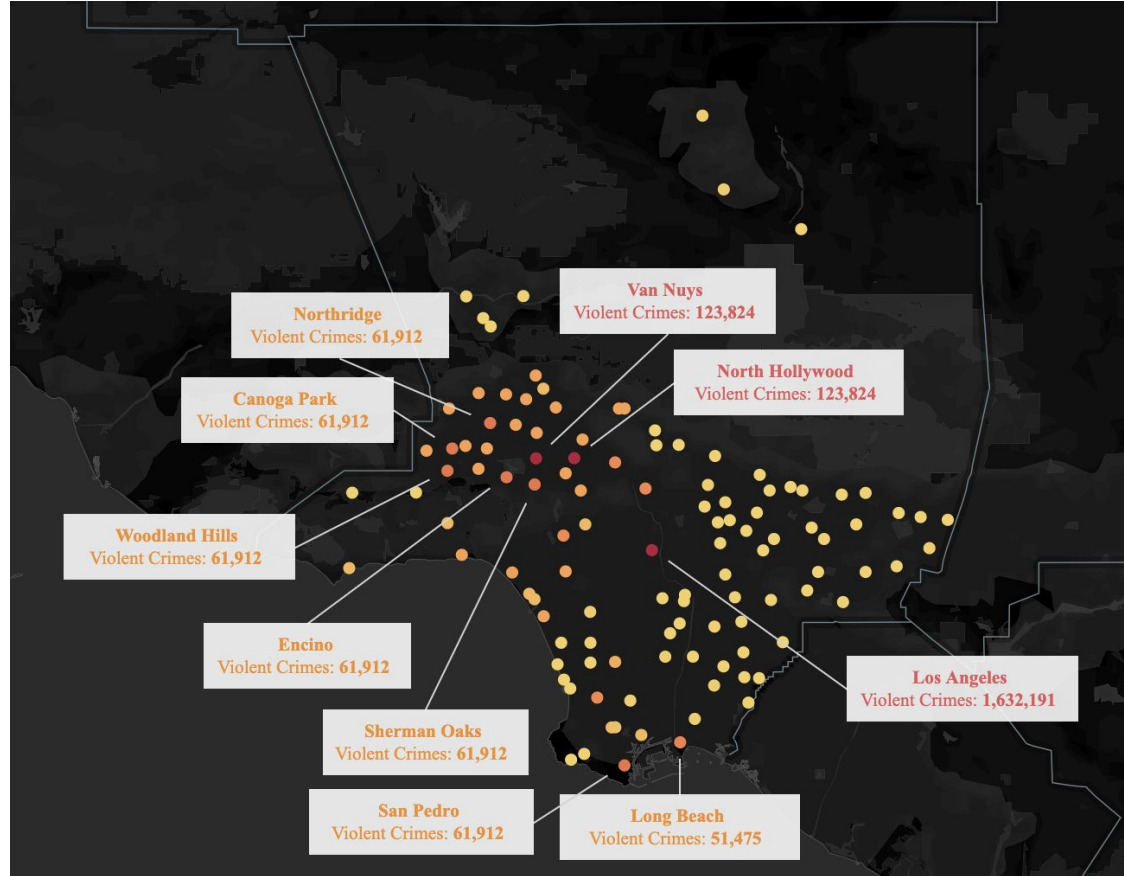
The Model

Logistic regression model in zip code level:

Whole Foods = β X+ e

❖ **Significant Coefficients:**
- Log of Number of Households **
- Four Years Population Growth **
- Ralphs Presence *
- Log of Average Housing Price *

❖ RMSE = 0.26
❖ Pseudo R-squared = 0.345

Logit Regression Results

	Dependent variable:
	Whole Foods presence
const	-61.7317
	(-24.23)
Ralphs	0.6993*
	(0.421)
Trader Joes	-0.2578
	(0.524)
Bristol Farms	0.4608
	(0.879)
Log(Total population)	-2.4703
	(1.812)
Log(Median Age)	0.6537
	(2.978)
Log(Number of Households)	4.0652**
	(1.947)
Log(Households Median income)	0.5616
	(1.58)
Unemployment Rate	-0.1689
	(0.178)
Log(Violent Crime)	0.1192
	(0.14)
Log(Average Apartment Rent)	2.7122
	(3.144)
Log(Average Housing Price)	1.1887*
	(0.707)
Four Years Population Growth	0.0939**
	(0.046)
Four Years Income Growth	-0.0046
	(0.027)
Pseudo R-squared	0.3454
Observations	263
Log Likelihood	-56.969
Note	*P<0.1; **P<0.05; ***P<0.01

The Model

Logistic regression model in city level:

Whole Foods = β X+ e

❖ **Significant Coefficients:**

- Log of Total Population **
- Ralphs Presence *
- Trader Joes Presence *
- Log of Total Violent Crimes *
- Education *
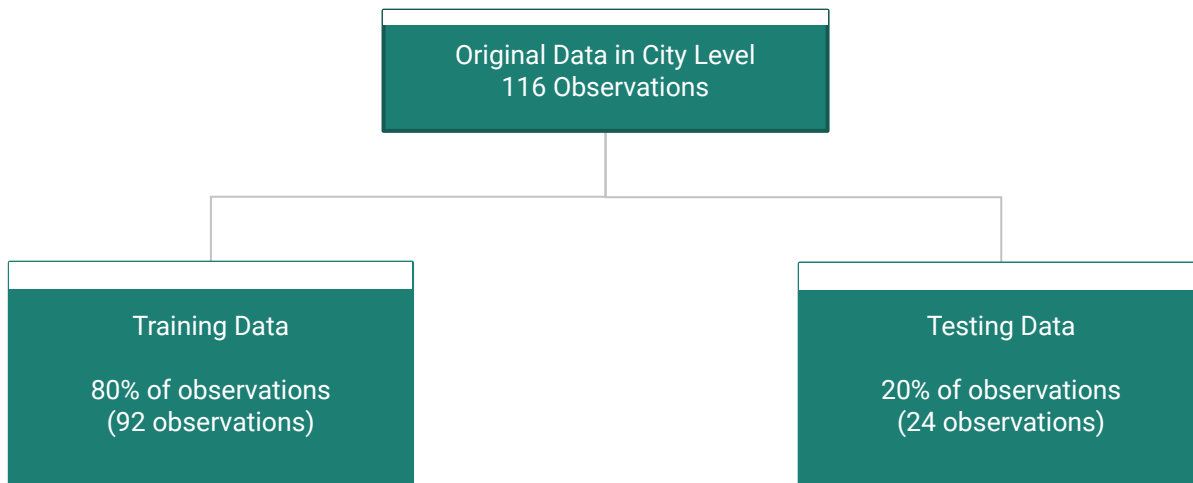
❖ Pseudo R-squared = 0.697
❖ RMSE= 0.19

Logit Regression Results	
	Dependent variable:
	Whole Foods presence
const	-22.917
	(74.811)
ralphs	2.296*
	(1.21)
Trader Joes	-2.5036*
	(1.318)
Bristol Farms	0.9176
	(2.195)
Log(Total population)	3.0199**
	(1.543)
Log(Median age)	-12.3676
	(10.841)
Log(Households Median income)	-4.3858
	(5.352)
Unemployment Rate	-0.0772
	(0.604)
Log(Violent Crime)	0.5955*
	(0.321)
Log(Average apt price)	6.7631
	(7.654)
Log(Average housing price)	1.0547
	(2.544)
Education	0.2536*
	(0.13)
Public transportation traffic	-0.5517
	(0.511)
Total traffic volume	5.93E-07
	(4.48E-07)
Pseudo R-squared	0.6971
Observations	116
Log Likelihood	-13.53
Note	*P<0.1; **P<0.05; ***P<0.01

Model Validation

Final Model to predict potential locations:

$$Whole\ Foods = \beta_0 + \beta_1\ log(Total\ Population) + \beta_2\ Ralphs + \beta_3\ Tader\ Joes$$
$$+ \beta_4\ log(violent\ crime) + \beta_5\ Education$$

Original Data in City Level
116 Observations

Training Data

80% of observations
(92 observations)

Testing Data

20% of observations
(24 observations)

Training data results

- R-squared = 0.66

- RMSE = 0.28

$\beta_{logPop} = 3.92$

$e^{\beta_{Ralphs}} = 5.94$

$e^{\beta_{TraderJoes}} = 0.12$

$\beta_{logcrime} = 0.42$

$e^{\beta_{Education}} = 1.26$

Log total population (+)	1%	3.92% ↑
Ralphs (+)	1 unit	494% ↑
Trader Joes (-)	1 unit	88% ↓
Log violent crime (+)	1%	0.42% ↑
Education rate (+)	1 unit	26% ↑

Logit Regression Results

	Dependent variable:
	Whole Foods presence
const	-59.1578
	(22.048)
Log(Total population)	3.9247**
	(1.683)
Ralphs	1.7822*
	(1.00)
Trader Joes	-2.0878*
	(1.133)
Log(Violent Crime)	0.4249*
	(0.219)
Education	0.2327***
	(0.083)
R-squared	0.6566
Observations	116
Log Likelihood	-13.53
Note	*P<0.1; **P<0.05; ***P<0.01

Model Accuracy

|		
True positive rate (TPR) (Sensitivity)	$$\frac{TP}{\text{Actual positive}}$$	0.67
True negative rate (TNR) (Specificity)	$$\frac{TN}{\text{Actual negative}}$$	0.90
Accuracy	$$\frac{TN + TP}{\text{Total}}$$	0.88
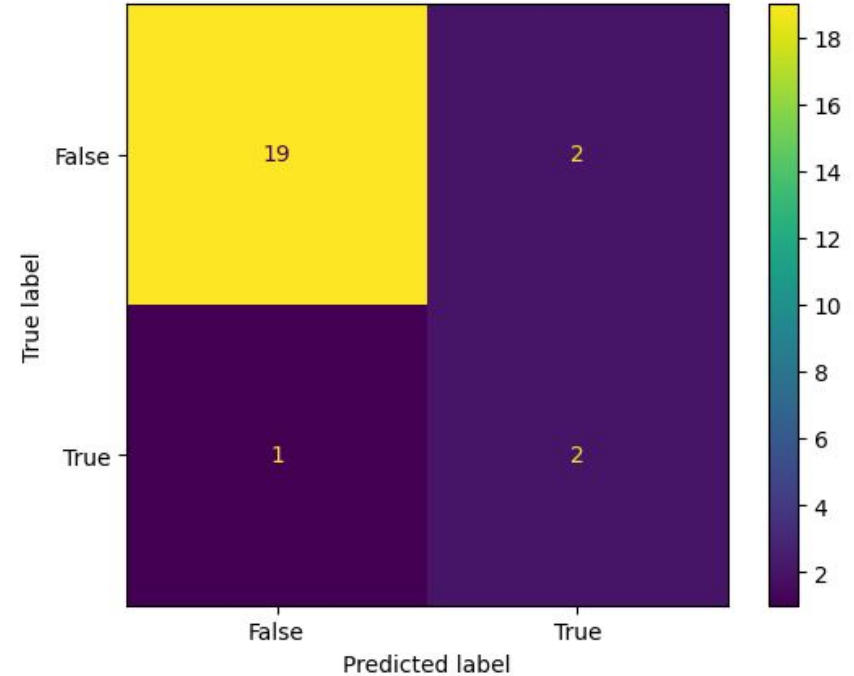
Prediction evaluation

Strategy Comparison

	Whole Foods	Trader Joe's	Ralphs
Constant	-45.4359 (.002)	-36.7858 (.001)	-11.9961 (.052)
Log of Total Population	2.6472 (.013)	2.6540 (.003)	0.8609 (.096)
Log of Violent Crime Number	0.5823 (.003)	0.0012 (.993)	-0.0013 (.990)
Percentage of Bachelor's Degree or Higher	0.2028 (.001)	0.1492 (.000)	0.0314 (.151)
Number of Whole Foods		-.1.2180 (.128)	1.5884 (.073)
Number of Trader Joe's	-1.4932 (.058)		1.3735 (.024)
Number of Ralphs	1.5521 (.051)	0.9473 (.035)	

Where to expand

The underserve level: Whole Foods presence dummy - model predicted probability
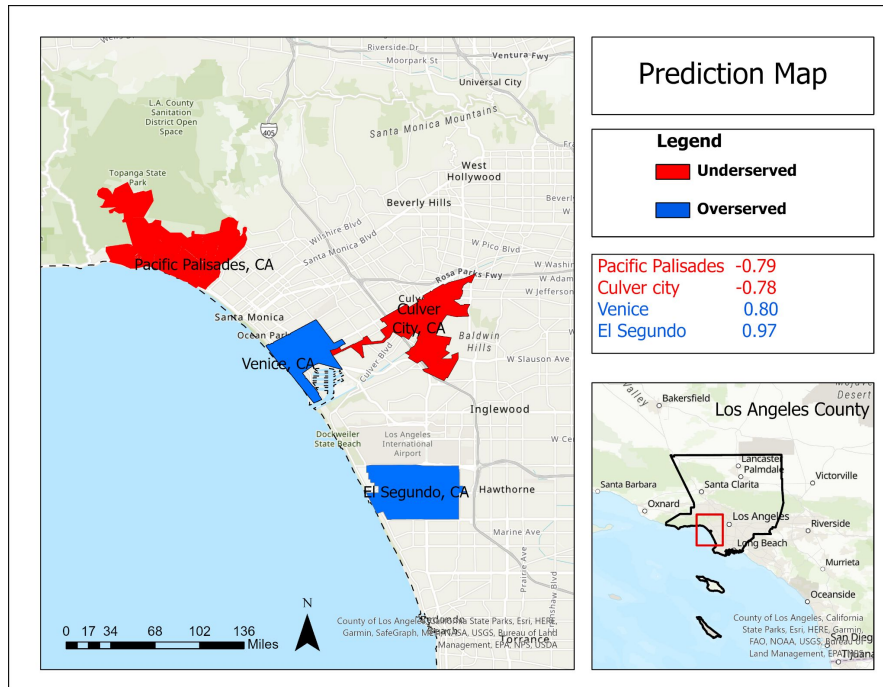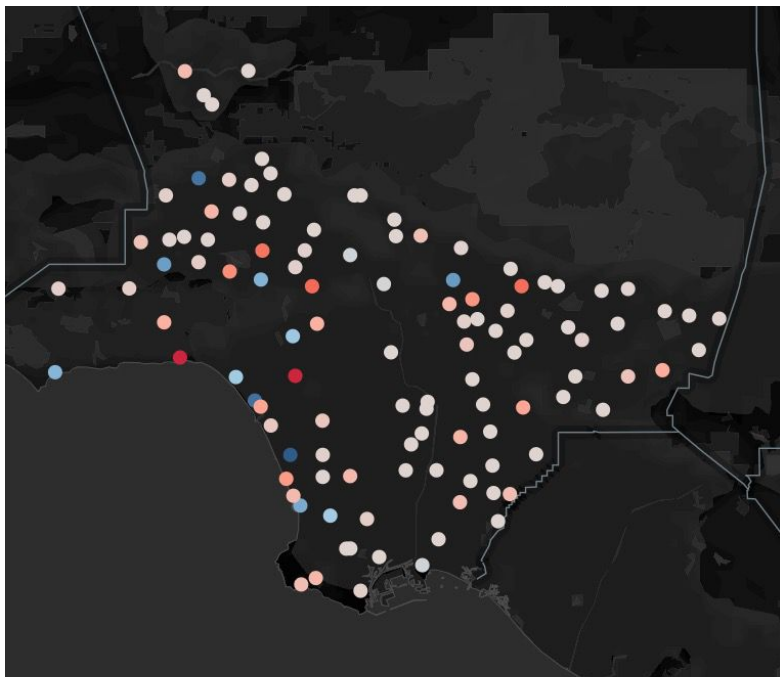
Further Studies and Conclusions

- Two potential locations for a new store: Pacific Palisades and Culver City
- The model predicts the closure the Venice and El Segundo location
  - This decision would be best made with sales data
- Expand study area to West Coast (California, Oregon, Washington) with neighborhood level of aggregation
  - More data
  - Reliable & detailed results
  - More meaningful results for a corporate client
- Include more competitors and commercial rent data

Thank you!

References

Rincón, & Tiwari, C. (2020). Demand Metric for Supermarket Site Selection: A Case Study. Papers in Applied Geography, 6(1), 19–34. https://doi.org/10.1080/23754931.2020.1712555

Smith, & Sanchez, S. (2003). Assessment of business potential at retail sites: empirical findings from a US supermarket chain. The International Review of Retail, Distribution and Consumer Research, 13(1), 37–58. https://doi.org/10.1080/0959396032000051684

Baviera-Puig, Buitrago-Vera, J., & Escriba-Perez, C. (2016). Geomarketing models in supermarket location strategies. Journal of Business Economics and Management, 17(6), 1205–1221. https://doi.org/10.3846/16111699.2015.1113198

The economics of supermarket and grocery store location - USDA. (n.d.). Retrieved April 19, 2023, from https://www.ers.usda.gov/webdocs/publications/42711/12705_ap036f_1_.pdf?v=0