Predicting the Success of a Café in Los Angeles, California

ANALYZING NEIGHBORHOOD DATA CAN INFORM STAKEHOLDERS

- The County of Los Angeles is the most populous county in the United States of America with over 10 million residents as of 2010.
- The demographics of the population are as diverse as the geography of the county – Beach communities, mountain towns, sprawling suburbs and a vibrant downtown city.
- Certain cities in the county are saturated with cafés, while others have little to none, though that can be changing due to shifts in societal trends and gentrification.
- Stakeholders can make informed decisions on where to open a café by analyzing data readily found on the web.

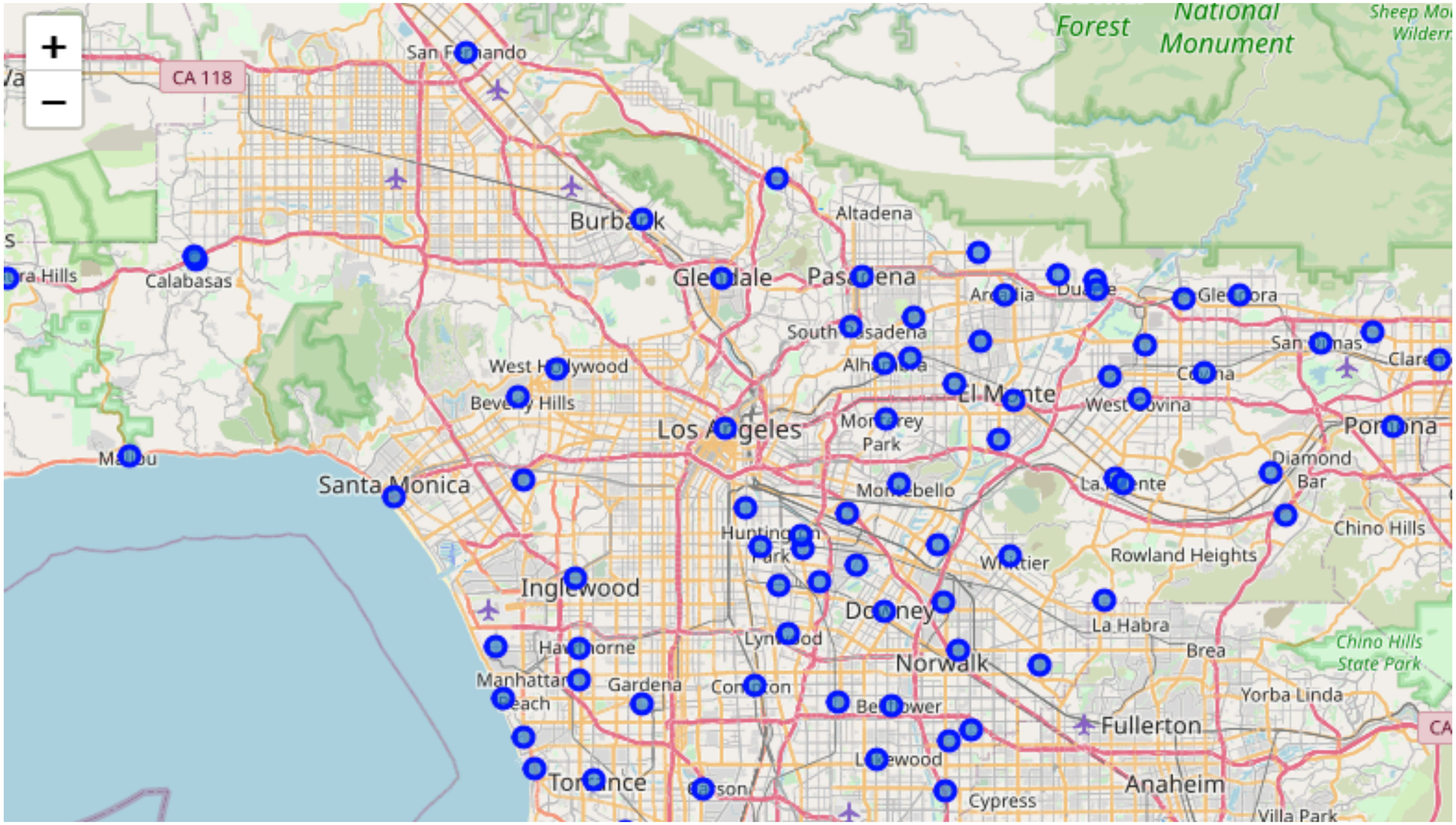
DATA ACQUISITION AND CLEANING

- Venue name, category and geographical coordinates for cities in Los Angeles can be found using Foursquare's API.
- City geographical coordinates and population data can be scraped from Wikipedia.
- In total, 7,458 rows and 7 features were returned in Foursquare's venue dataset.
- The dataset was grouped by unique category, where "Café" was the attribute. This data was then merged with the cities geographical data to create our dataframe.

	City	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Agoura Hills	34.14611	-118.77812	Future Track Running Center	34.145819	-118.779251	Sporting Goods Shop
1	Agoura Hills	34.14611	-118.77812	Twisted Oak Tavern	34.145308	-118.778679	Gastropub
2	Agoura Hills	34.14611	-118.77812	Cafe Bizou	34.148410	-118.782587	French Restaurant
3	Agoura Hills	34.14611	-118.77812	Grissini Ristorante	34.145815	-118.778534	Italian Restaurant
4	Agoura Hills	34.14611	-118.77812	Pizza Nosh	34.148311	-118.782181	Pizza Place
5	Agoura Hills	34.14611	-118.77812	Forest Cove Park	34.152290	-118.774749	Park
6	Agoura Hills	34.14611	-118.77812	Target	34.148045	-118.793964	Big Box Store
7	Agoura Hills	34.14611	-118.77812	Hook Burger	34.147716	-118.796260	Burger Joint

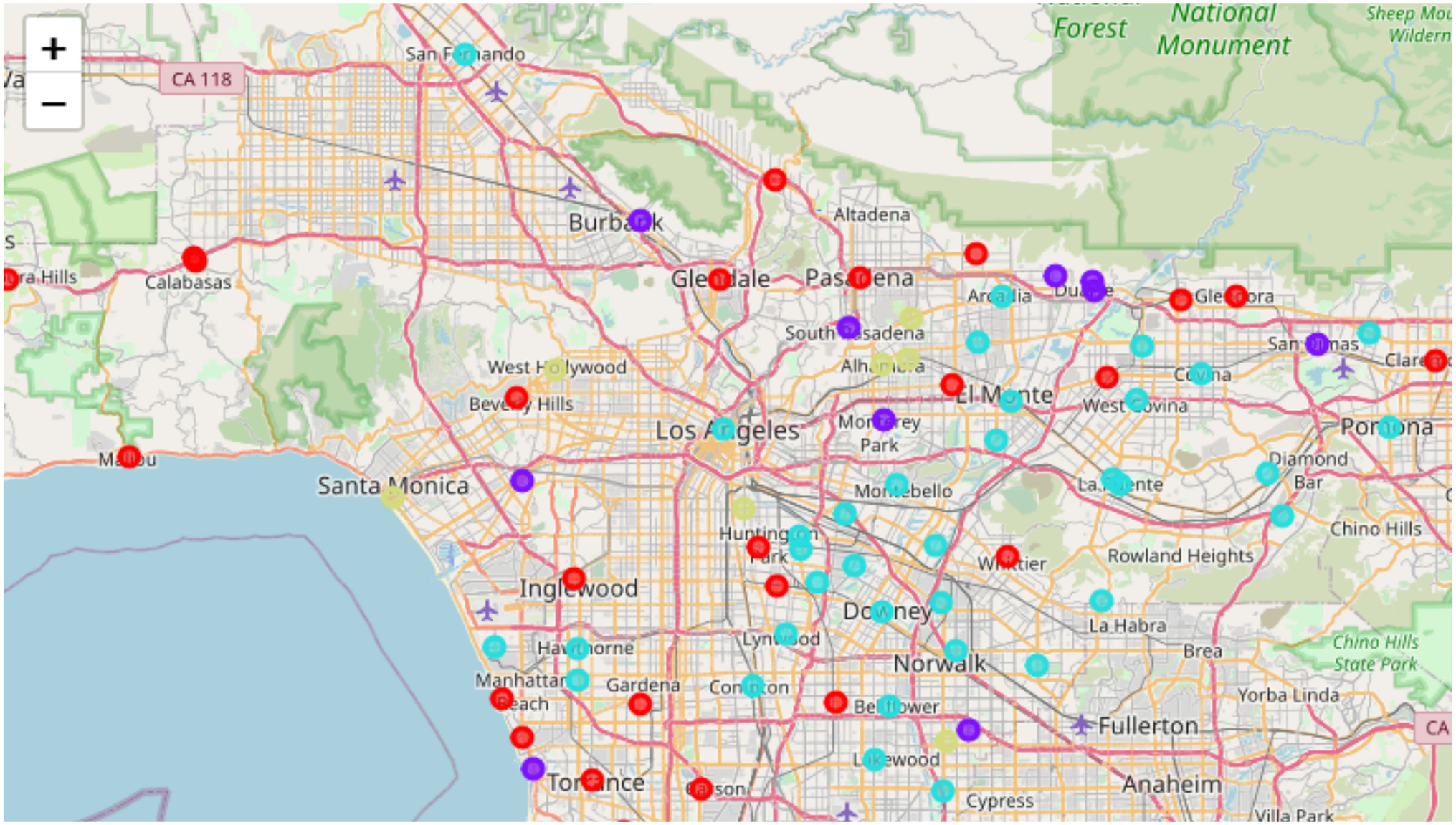
	City	Café	Cluster Labels	Population as of(2010 Census)	Latitude	Longitude
0	Agoura Hills	0.010000	0	20330	34.14611	-118.77812
1	Alhambra	0.040000	3	83653	34.09370	-118.12727
2	Arcadia	0.000000	2	56364	34.13614	-118.03887
3	Artesia	0.040000	3	16522	33.86114	-118.07968
4	Avalon	0.014286	0	3728	33.34411	-118.32139

CITY DATA VISUALIZATIONS



Cities from the Wikipedia dataset were mapped and visualized using Python's Folium attribute.

VENUE DATA VISUALIZATIONS



Venues categorized as “Cafés” were then merged, clustered and visualized.

ANALYZING THE RESULTS

- K-means was used to cluster venue into 4 groups.
- The mean average of cafés per city could easily be interpreted and visualized.
- A recommendation of where to open a café avoiding competition could then be given.

	City	Café	Cluster Labels	Population as of(2010 Census)	Latitude	Longitude
74	Santa Monica	0.050000	3	89736	34.01158	-118.49227
71	San Marino	0.044118	3	13147	34.12208	-118.10521
1	Alhambra	0.040000	3	83653	34.09370	-118.12727
82	Vernon	0.050000	3	112	34.00513	-118.23025
70	San Gabriel	0.040000	3	39718	34.09708	-118.10824
58	Palos Verdes Estates	0.051282	3	13438	33.80003	-118.39187
85	West Hollywood	0.040000	3	34399	34.09069	-118.37053
3	Artesia	0.040000	3	16522	33.86114	-118.07968

CONCLUSION AND FUTURE DIRECTIONS

- Dataset can be expanded to include features like income and demographics in future models.
- Foursquare venue data is constantly updating. The timeliness of running the model and purchasing a venue location is critical.
- Capturing more data features can be used to run linear regression and other helpful prediction models.