

VENUES AND LIVING IN SAN JOSE

Capstone Project - The Battle of Neighborhoods

IBM Data Science Professional Certificate

Agenda

- ▶ Introduction/Business Problem
- ▶ Data
- ▶ Methodology
- ▶ Results
- ▶ Discussion
- ▶ Conclusion
- ▶ References

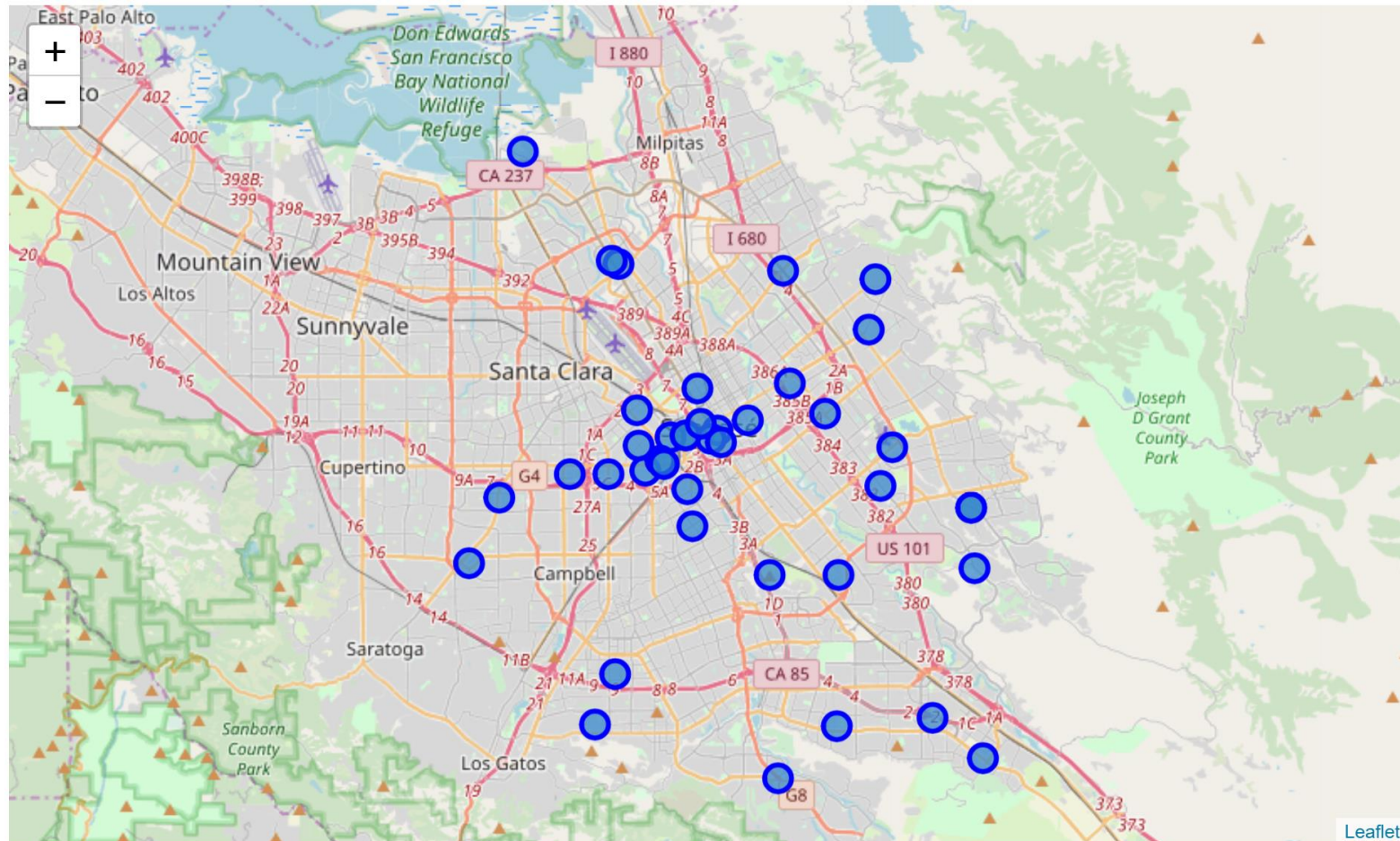
Introduction

- ▶ Every year, more than **7 million** Americans move to a different state within the US [1].
- ▶ One of the **major indicators** when deciding where to move is the **availability of venues** that fit their needs and lifestyle.
- ▶ Couples moving to a new city tend to prefer **lively areas** with **variety of venues** nearby.
- ▶ Specific Business problem:
 - ▶ A couple is moving to **San Jose** and interested in finding a neighborhood with an abundance of venues nearby; especially **coffee shops** and **sandwich joints**.

Data

- ▶ San Jose neighborhood list[2]:
 - ▶ Neighborhood name
 - ▶ Latitude and longitude data
- ▶ Foursquare API:
 - ▶ Nearby venues name
 - ▶ Venue category
 - ▶ Latitude and longitude data ... etc

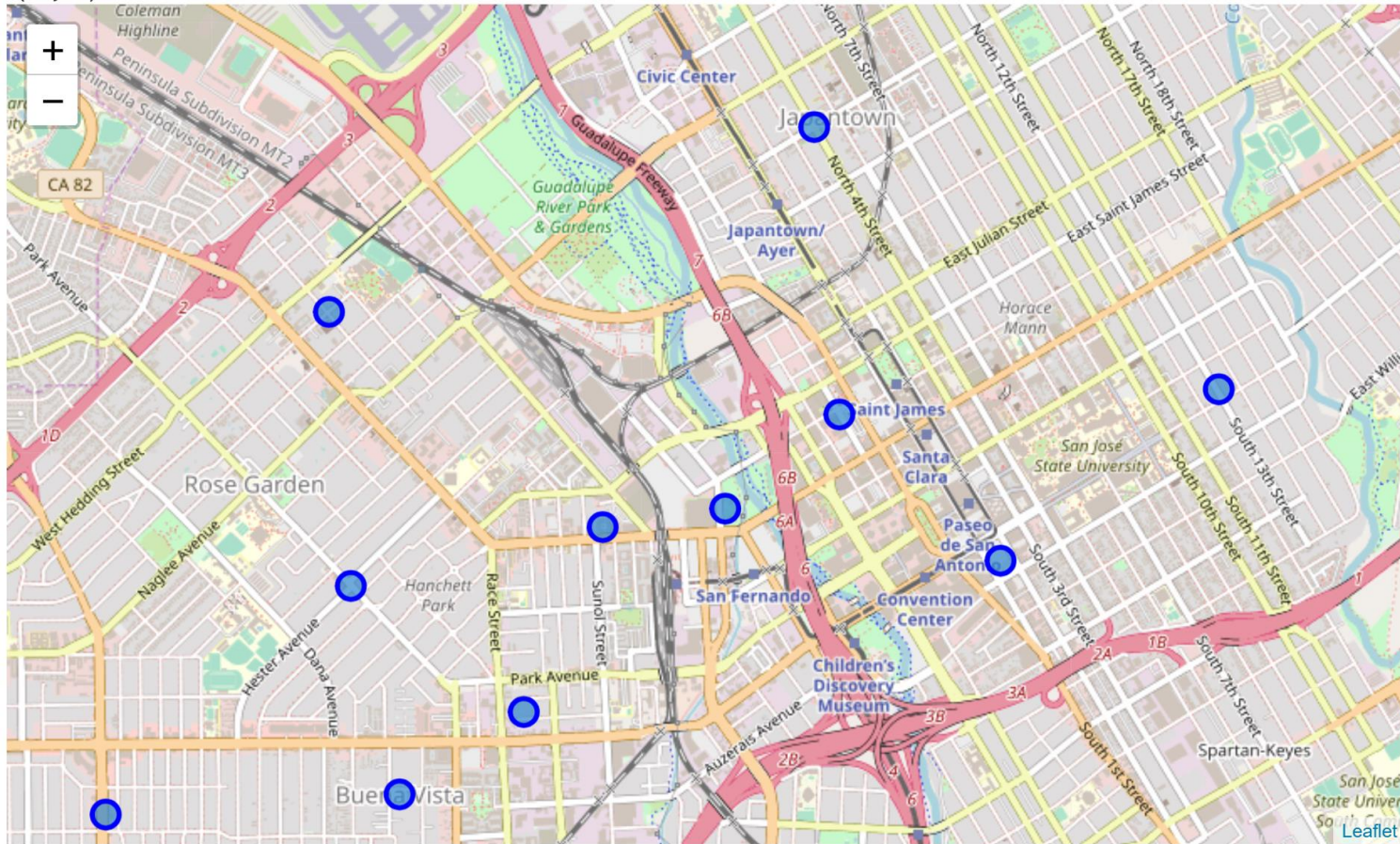
Data



Methodology

- ▶ Retrieve nearby venues for each neighborhood from Foursquare API
- ▶ Join neighborhood venues to create a list of all venues
- ▶ **Problem:**
 - ▶ Redundant venues assigned to multiple neighborhoods at the same time
- ▶ **Reason:**
 - ▶ Neighborhood centers retrieved from Wikipedia are too close
- ▶ **Solution:**
 - ▶ Remove neighborhoods that cause redundancy

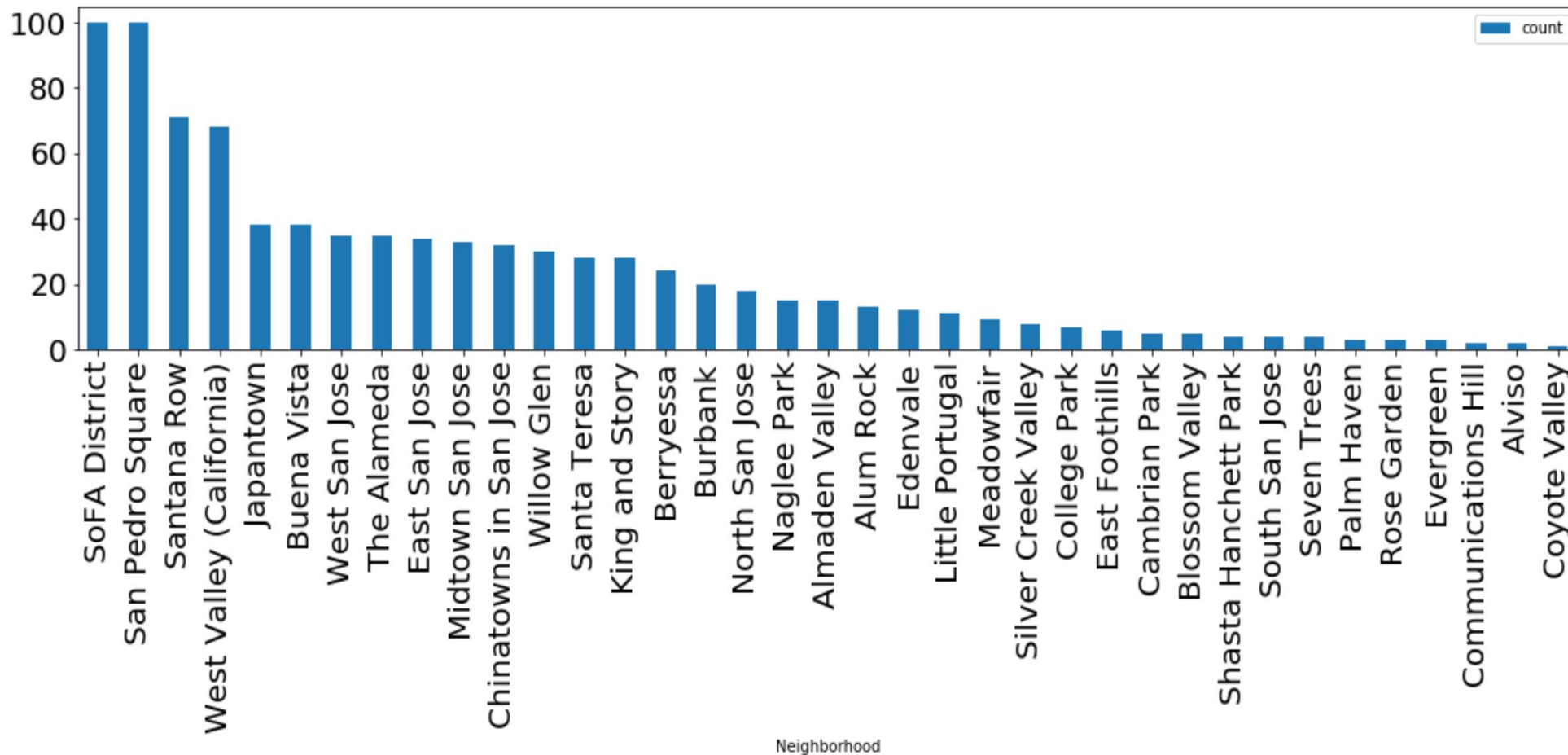
Methodology



After redundant neighborhoods are removed

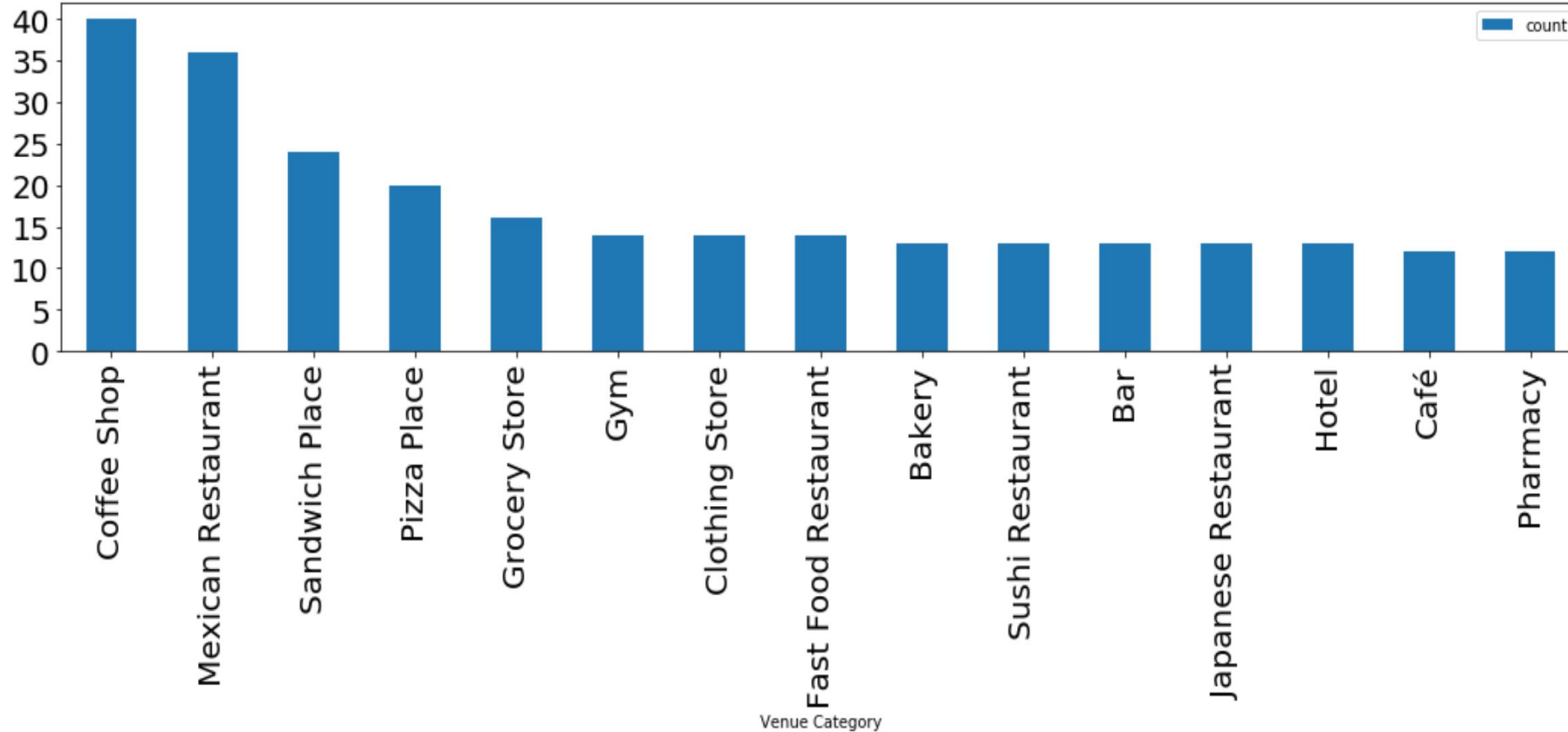
Methodology

- Identify **lively neighborhoods** (neighborhoods with the most nearby venues)



Methodology

- Identify **most frequent venue** categories in San Jose



Methodology

- Identify the most **frequent venues per neighborhood**

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Almaden Valley	Coffee Shop	Bank	Pet Store	Shopping Mall	Sandwich Place	Salon / Barbershop
1	Alum Rock	Pizza Place	Diner	Liquor Store	Thrift / Vintage Store	Bakery	BBQ Joint
2	Alviso	Mexican Restaurant	Playground	Hockey Arena	History Museum	Fish & Chips Shop	Financial or Legal Service
3	Berryessa	Pizza Place	Chinese Restaurant	Convenience Store	Japanese Restaurant	Shipping Store	Bubble Tea Shop
4	Blossom Valley	Baseball Field	Yoga Studio	Ethiopian Restaurant	Flower Shop	Fish & Chips Shop	Financial or Legal Service

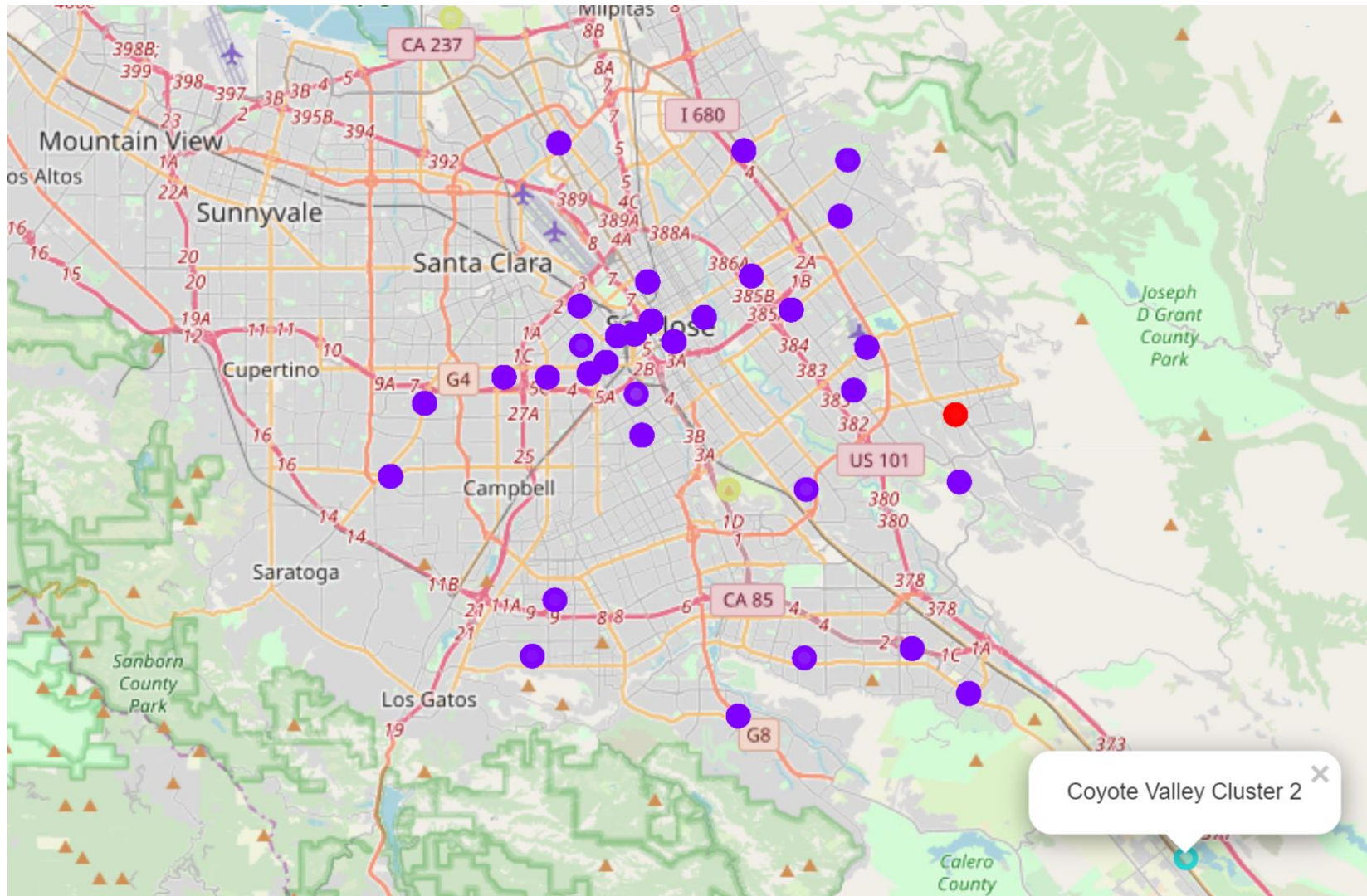
Results

- ▶ **Most lively neighborhoods:**
 - ▶ Since our couple prefer lively neighborhoods, we focus on the top 10 neighborhoods in our recommendation
- ▶ **Most frequent venue category:**
 - ▶ The results line up with what our couple prefer; ranking coffee shops and sandwich places in the top 3 venue categories.
- ▶ **Most frequent venues per neighborhood:**
 - ▶ There are multiple venues that enlist coffee shops and sandwich places as their top frequent venues.

Results

- ▶ One hot technique:
 - ▶ Each row represents a neighborhood
 - ▶ Each column represents a venue category
 - ▶ Values refer to whether a venue category occurs in a neighborhood
- ▶ Clustering:
 - ▶ Why? No labels available nor are needed
 - ▶ K-means with k initialized to 4

Results



Discussion

- ▶ **Most of** San Jose's neighborhoods are **very similar** when it comes to nearby venues.
- ▶ The **two unique areas** detected were the Coyote Valley on the very south end of San Jose and the Rose Garden area in the east.
 - ▶ Can be labeled as **rural** or could be considered **suburbs** of San Jose.

Discussion

- ▶ Recommendations:
 - ▶ Strongly encourage moving in to San Jose.
 - ▶ Frequent and variety of venues
 - ▶ Most frequent venue include coffee shops and sandwich locations
 - ▶ in The Alameda, Naglee Park and San Pedro Square
 - ▶ **San Pedro Square**
 - ▶ Greatest number of venues.
 - ▶ Variety of coffee shops and sandwich places

Conclusion

- ▶ The problem at hand was revolving around recommending the best location for living to a new couple moving to San Jose.
- ▶ The preferences were a lively area where lots of coffee shops and sandwich venues can be found.
- ▶ Data detailing the neighborhoods of San Jose was scrapped off of Wikipedia along with its latitude and longitude information.
- ▶ In addition, Foursquare API was utilized in acquiring close by venues in regard to each neighborhood.
- ▶ We employed multiple techniques in our approach to clean and prepare the data.
- ▶ Furthermore, we took advantage of clustering algorithms in gaining knowledge about San Jose neighborhoods.
- ▶ Finally, we were able to draw great conclusions and recommend the best fit neighborhoods based on the preferences given and the knowledge gained from the process.

References

- [1] New State Residents Statistics, Demographic Data. Retrieved April 4, 2019, from <https://www.governing.com/gov-data/residents-moving-to-new-state-demographics-population-statistics.html>
- [2] Category:Neighborhoods in San Jose, California. (2017, November 15). Retrieved April 4, 2019, from https://en.wikipedia.org/wiki/Category:Neighborhoods_in_San_Jose,_California