Capstone Project - The Battle of the Neighbourhoods (Week 2)

Exploration and analysis of San Jose and its neighbourhoods

Problem and Solution approach

- Problem
  - Identify the possibility to establish a financial advisory institution in San Jose

- Approach to the solution
  - Analyse the economic indicators of the population across neighbourhoods
  - Explore the venues in each neighbourhood of San Jose
  - Merge both of the data to provide a comprehensive view
  - Cluster the neighbourhoods using K-means
  - Provide visuals for Neighbourhood, clusters

Data acquisition and preparation

- Base data retrieved from Spatial Data Repository of NYU (U.S. Neighbourhoods greenness measures and social variables)
- Base data contains economic indicators and other data of many cities
- Data wrangling applied to arrive at economic indicators for San Jose
- Venues of San Jose neighbourhoods explored using Foursquare
- Top ten venues for each neighbourhood identified
- Venues data and economic indicators data merged for a comprehensive view
- Neighbourhood data clustered using K-means

Data input– Economic indicators

- Critical attributes for analysis
  - Average High income, median high income, percentage of owners, percentage of rent, median age

	City	Neighborhood	Latitude	Longitude	popdensity	avg_hinc	pct_own	pct_rent	medage_cy	medhinc_cy	city_parks
448	San Jose	Almaden Valley	37.2211	-121.869849	5612.600000	176918	0.734866	0.237213	41.210811	140164	0.068
449	San Jose	Alum Rock-East Foothills	37.3772	-121.825214	10173.289474	113307	0.674909	0.300745	35.323684	89187	0.068
450	San Jose	Berryessa	37.4016	-121.856383	9173.466667	130216	0.744864	0.236287	37.942424	108157	0.068
452	San Jose	Blossom Valley	37.2549	-121.843005	9977.140984	115350	0.657751	0.317510	36.403279	98611	0.068
453	San Jose	Buena Vista	37.3212	-121.916699	13386.700000	79466	0.275472	0.667481	32.340000	58664	0.068
454	San Jose	Burbank	37.3213	-121.930542	10040.090909	91458	0.370584	0.587205	35.945455	69339	0.068
455	San Jose	Cambrian Park	37.2598	-121.913999	7439.178261	132315	0.674165	0.302950	40.337681	103343	0.068
457	San Jose	Downtown	37.3405	-121.890340	12526.985714	86554	0.343653	0.603385	32.780000	63735	0.068
458	San Jose	East San Jose	37.3347	-121.825193	13127.814474	102843	0.676735	0.302644	32.343421	85678	0.068
459	San Jose	Edenvale-Seven Trees	37.2797	-121.817703	12941.859375	114389	0.674291	0.299457	32.309375	92459	0.068
460	San Jose	Rose Garden	37.3296	-121.931553	9117.513333	104442	0.447511	0.517272	38.103333	78196	0.068
462	San Jose	Evergreen	37.2981	-121.770502	8227.902326	140063	0.799158	0.178919	36.944186	116423	0.068
463	San Jose	Fairgrounds	37.3020	-121.858670	13640.700000	83621	0.480870	0.485012	31.651163	65232	0.068
464	San Jose	North San Jose	37.3831	-121.931127	5048.100000	107008	0.454169	0.474309	33.514286	84430	0.068
465	San Jose	North Valley	37.3763	-121.874664	12035.338000	101165	0.557130	0.406239	33.468000	84418	0.068
466	San Jose	Santa Teresa	37.2366	-121.793081	7900.364706	136960	0.728919	0.244759	37.592157	112249	0.068
467	San Jose	West San Jose	37.3010	-121.983818	9151.487097	129873	0.580150	0.392746	40.151613	105168	0.068
468	San Jose	Willow Glen	37.2963	-121.901595	8251.063889	117085	0.571972	0.395732	39.108333	90779	0.068

Data input – Venues for each neighbourhood

- Venues data for each neighbourhood triggering economic activity

|	Neighborhood	Airport Terminal	American Restaurant	Arts & Crafts Store	Asian Restaurant	Auto Garage	BBQ Joint	Bakery	Bank	Bar	Beer Garden	Big Box Store	Bookstore	Breakfast Spot	Bubble Tea Shop	
0	Almaden Valley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
1	Alum Rock-East Foothills	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
2	Berryessa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
3	Blossom Valley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
4	Buena Vista	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.02439	0.000000	0.00000	0.024390	0.000000	0.02439	0.000000	0
5	Burbank	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.055556	0.00000	0.000000	0.055556	0.00000	0.000000	0
6	Cambrian Park	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.142857	0.000000	0.00000	0.000000	0
7	Downtown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023810	0.00000	0.000000	0.02381	0.000000	0.000000	0.00000	0.000000	0
8	East San Jose	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
9	Edenvale-Seven Trees	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
10	Evergreen	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0

Data – Economic indicators and venues merged

- Sample

	City	Neighborhood	Latitude	Longitude	popdensity	avg_hinc	pct_own	pct_rent	medage_cy	medhinc_cy	city_parks	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	San Jose	Almaden Valley	37.2211	-121.869849	5612.600000	176918	0.734866	0.237213	41.210811	140164	0.068	2	Playground	Gym / Fitness Center	Restaurant	Yoga Studio
1	San Jose	Alum Rock-East Foothills	37.3772	-121.825214	10173.289474	113307	0.674909	0.300745	35.323684	89187	0.068	1	Martial Arts Dojo	Pizza Place	Video Store	Deli / Bodega
2	San Jose	Berryessa	37.4016	-121.856383	9173.466667	130216	0.744864	0.236287	37.942424	108157	0.068	3	Construction & Landscaping	Fish & Chips Shop	Dance Studio	Deli / Bodega
3	San Jose	Blossom Valley	37.2549	-121.843005	9977.140984	115350	0.657751	0.317510	36.403279	98611	0.068	1	Food Truck	Mexican Restaurant	Convenience Store	Farmers Market
4	San Jose	Buena Vista	37.3212	-121.916699	13386.700000	79466	0.275472	0.667481	32.340000	58664	0.068	1	Motorcycle Shop	Dessert Shop	Grocery Store	Pharmacy
5	San Jose	Burbank	37.3213	-121.930542	10040.090909	91458	0.370584	0.587205	35.945455	69339	0.068	1	Convenience Store	Discount Store	Mexican Restaurant	Karaoke Bar

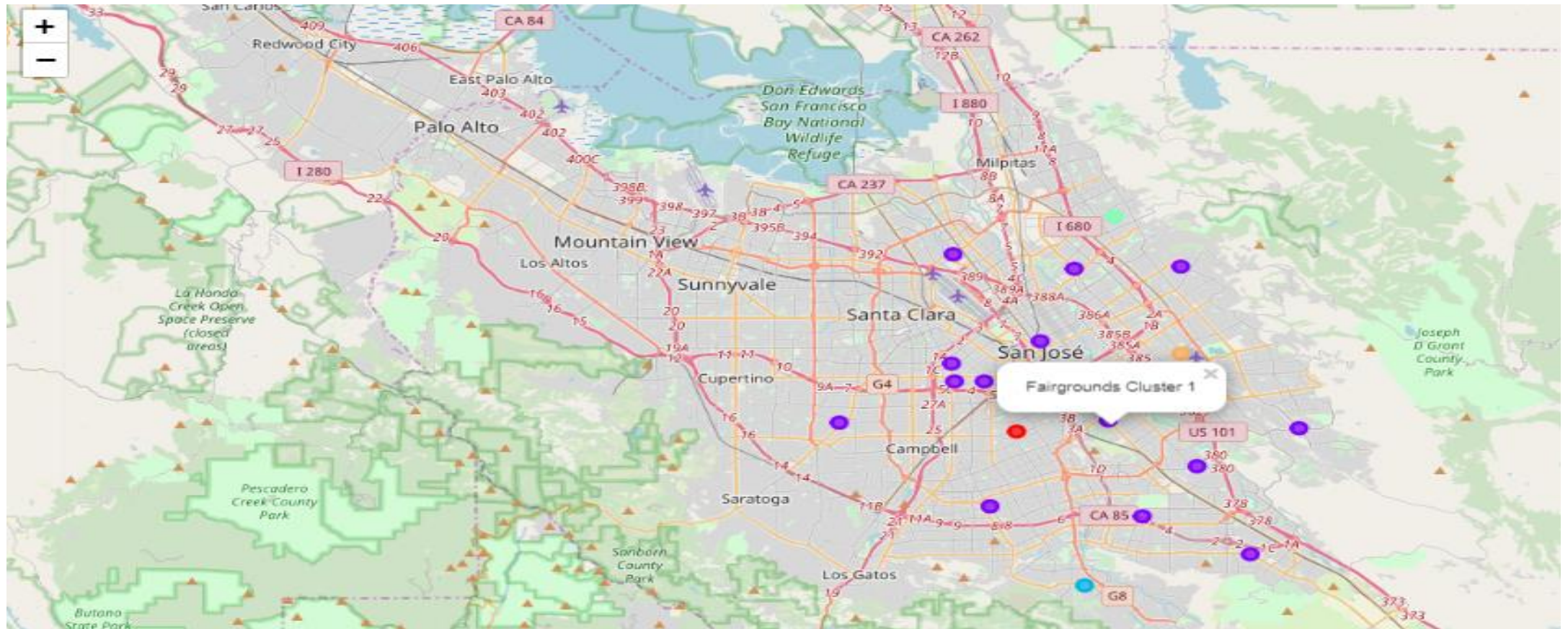# San Jose map with neighbourhood having highest average high income
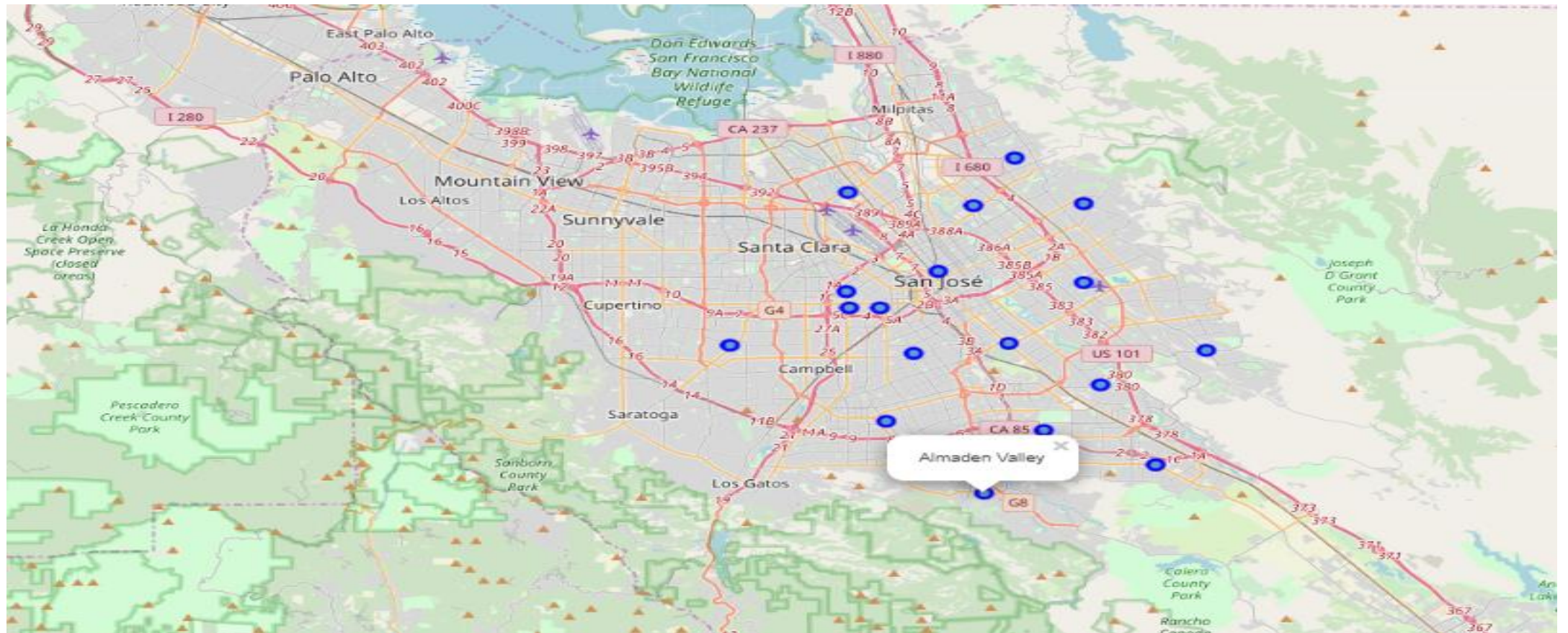
Statistical analysis – Neighbourhoods and Venues

- Each neighbourhood was analysed and the data was grouped by neighbourhood by taking the mean of the frequency of occurrence of each category as shown in the below sample:

|	Neighborhood	Airport Terminal	American Restaurant	Arts & Crafts Store	Asian Restaurant	Auto Garage	BBQ Joint	Bakery	Bank	Bar	Beer Garden	Big Box Store	Bookstore	Breakfast Spot	Bubble Tea Shop	
0	Almaden Valley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
1	Alum Rock-East Foothills	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
2	Berryessa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
3	Blossom Valley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
4	Buena Vista	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.02439	0.000000	0.00000	0.024390	0.000000	0.02439	0.000000	0
5	Burbank	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.055556	0.00000	0.000000	0.055556	0.00000	0.000000	0
6	Cambrian Park	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.142857	0.000000	0.00000	0.000000	0
7	Downtown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023810	0.00000	0.000000	0.02381	0.000000	0.000000	0.00000	0.000000	0
8	East San Jose	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
9	Edenvale-Seven Trees	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0
10	Evergreen	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0

San Jose map with neighbourhood clusters (sample) using K-means Clustering

Statistical analysis and results – Neighbourhoods and Venues

Average high income indicators in each neighborhood
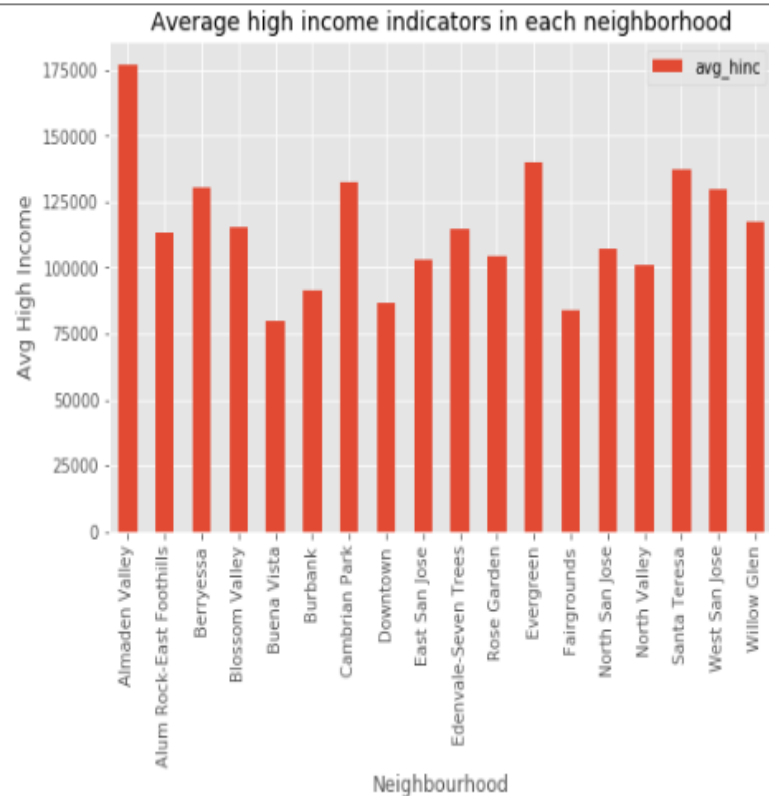
- 14 out of 18 neighbourhoods (nearly 78% of neighbourhoods) have an average high income value of more than 100'000 USD
- 9 out of 18 neighbourhoods which is 50 percent of the neighbourhoods have a median high income of above 90'000 USD
- In 12 out of the 18 neighbourhoods (nearly 67% of neighbourhoods) the percentage of owners is more than 50 percent in the population. In the rest of the neighbourhoods, the percentage of rent is higher than then percentage of owners
- The attribute 'medage_cy' shows that San Jose is a city of middle age group combined with average high income
- The merged data shows a lot of venues (shops, construction firms, restaurants, Pharmacies) which trigger a lot of economic activity along with their employees

Conclusions

- Overall, the economic indicators, venue data and the visual maps indicate the market potential which will enable an investor to establish a financial advisory institution

- Analysis can be extended to other cities also for identifying their economic potential

- Base data contains additional attributes which can be analysed for different purposes like to set up institutions for other purposes, for example education institutions, child care centres, hospitals.