Capstone Project - The Battle of Neighborhoods

Bo Xue

Introduction/Business Problem

- Asian cuisine is also very popular in the US, especially cities full of Asian immigrants, like the Silicon Valley in California.
- Bubble tea is growing more and more popular, bubble tea stores often have a long queue in some hot business areas
- This project aims to find the best location to open a business in a city in Silicon Valley, San Jose, CA
- It’s important to know if it is profitable prior to launching any business, so this report will try to gather data about similar businesses in the neighborhood and potential customer distribution, competitors, etc.
- The data can be used as part of a business plan to support making decision.

Data Descritpion

- I will use Foursquare API to find venues:
- For competitor study
  - Existing Bubble Tea shops around the neighborhoods in San Jose
  - Similar businesses around the neighborhoods in San Jose
- For potential customer distribution study
  - Universities around the neighborhoods in San Jose
  - Companies around the neighborhoods in San Jose
- Data visualization: I will use folium the map rendering library to visualize the venues

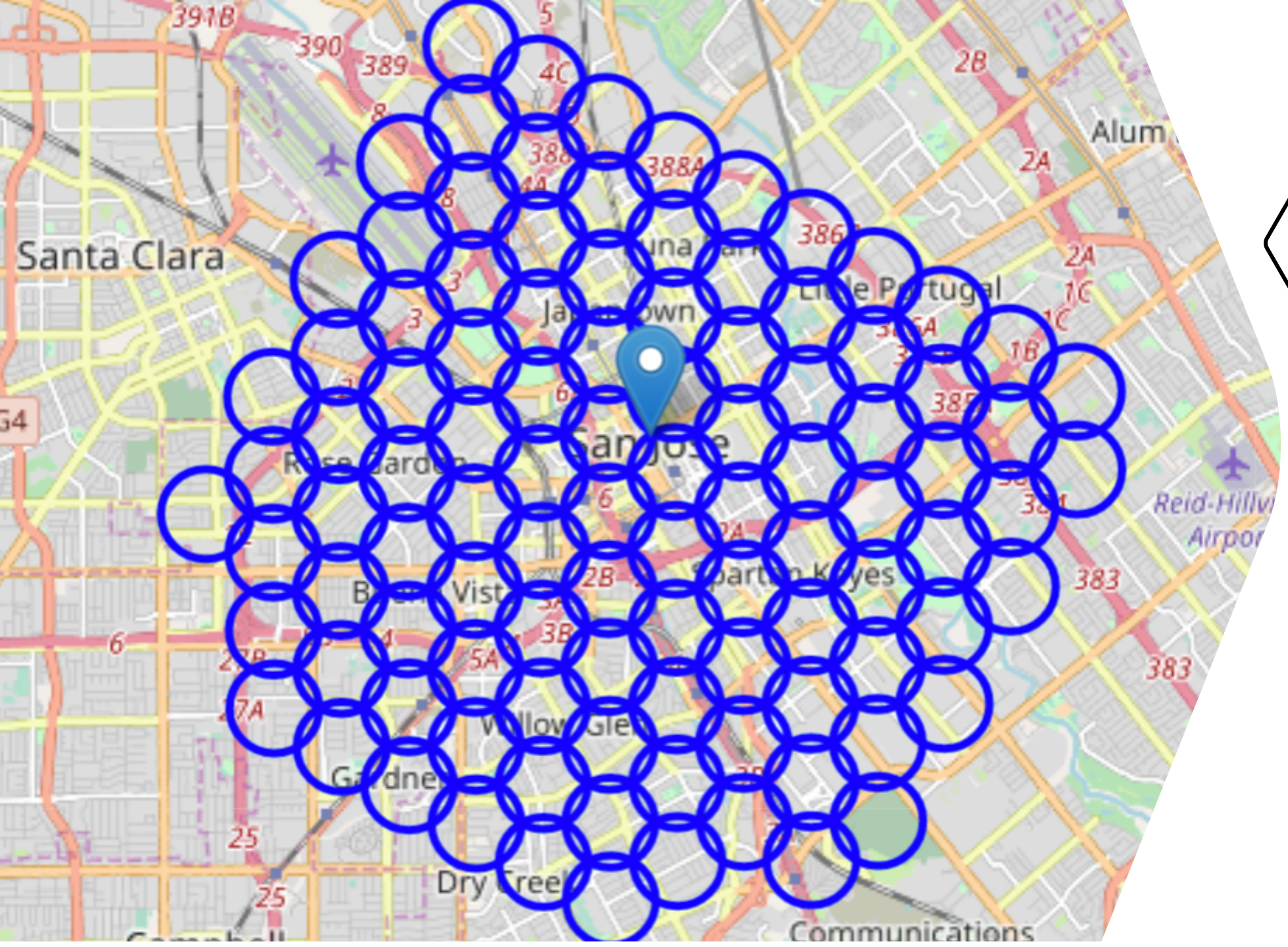
Methodology

1. City partitioning

   Create a hexagonal grid of cells: offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all it's neighbors.

   All following data manipulations are operated against each of the cells

2. Get all bubble tea shops and coffee shops/Cafe in each area cell using the foursquare API (category approach)
3. Find potential customer spots in each area cell with foursquare API category approach (universities and companies)
4. Then we can calculate average number of customers (Entity of universities/companies)
5. Candidate place can be found among the areas that have highest customer/existing-biz ratio
6. Use heatmap to visualize the outstanding spots



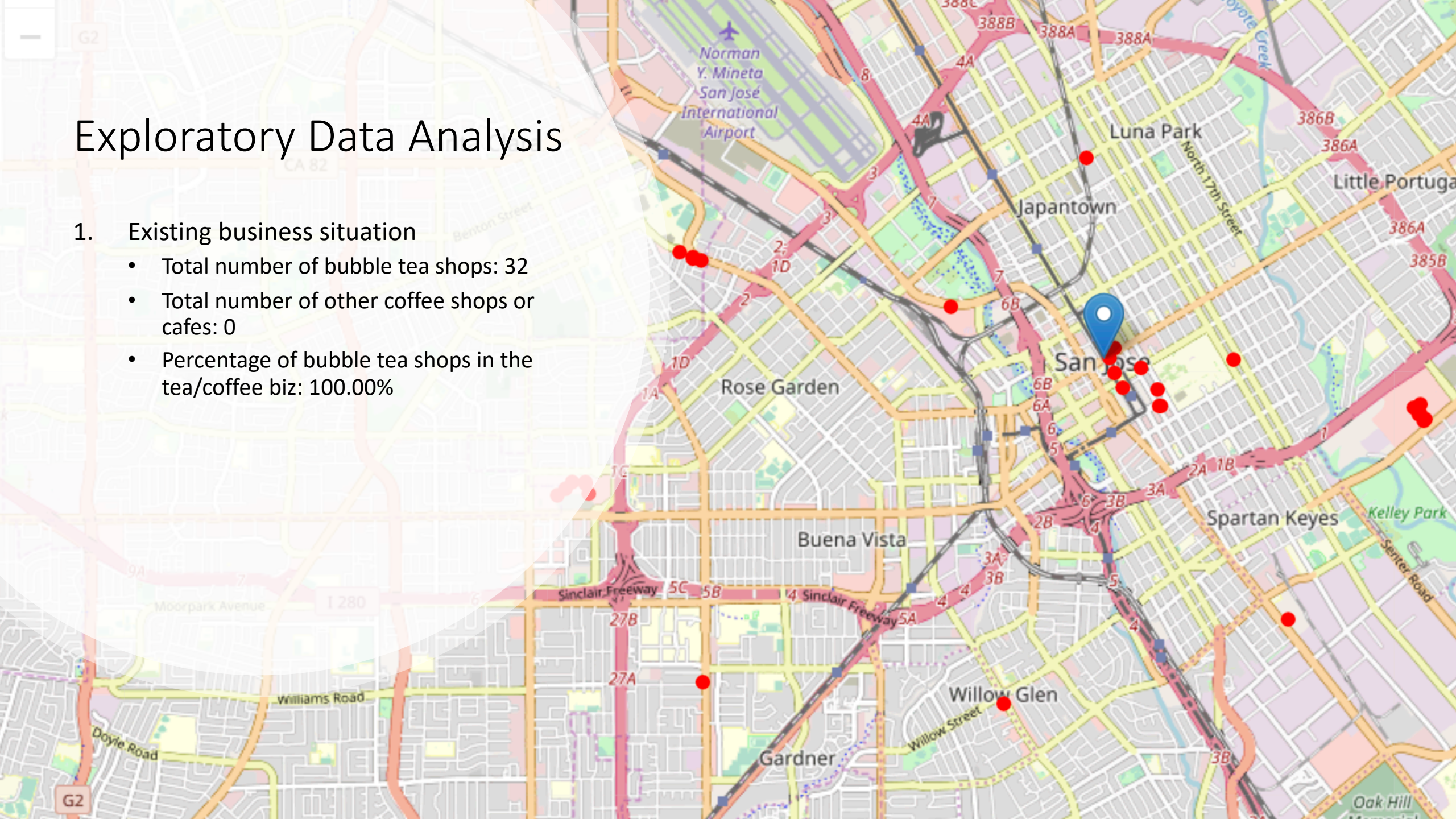
City partitioning

Exploratory Data Analysis

1. City partitioning to carve the city into small cells
2. Existing business situation: current bubble tea shops and other competitors (café, coffee shops)
3. Potential customer cluster situation (universities, companies)
4. Candidate areas that have profit potential
5. Visualization with Heatmap
6. Use Kmeans cluster approach to zoom to get the result set of location

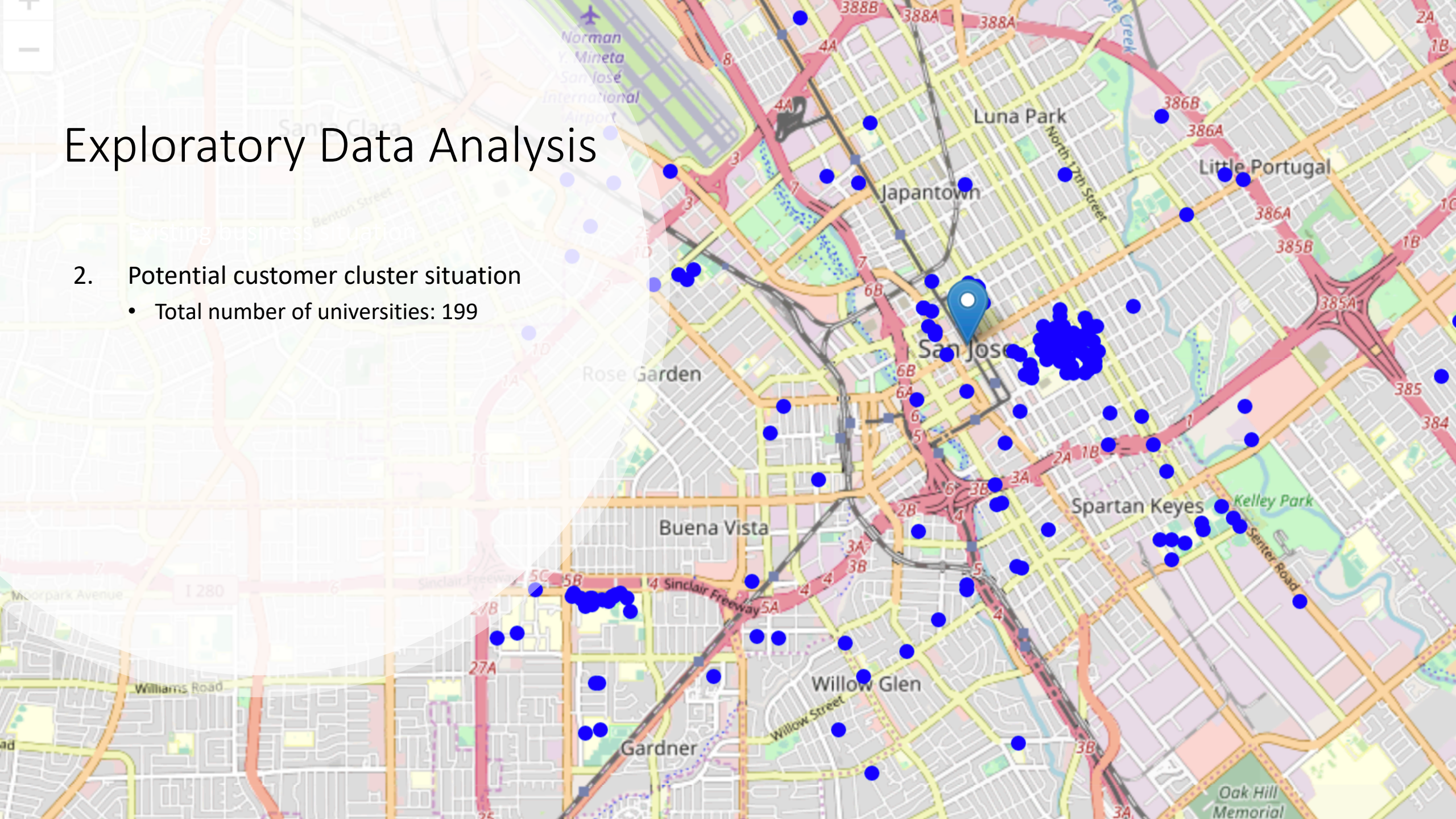
Exploratory Data Analysis

1. Existing business situation
   - Total number of bubble tea shops: 32
   - Total number of other coffee shops or cafes: 0
   - Percentage of bubble tea shops in the tea/coffee biz: 100.00%



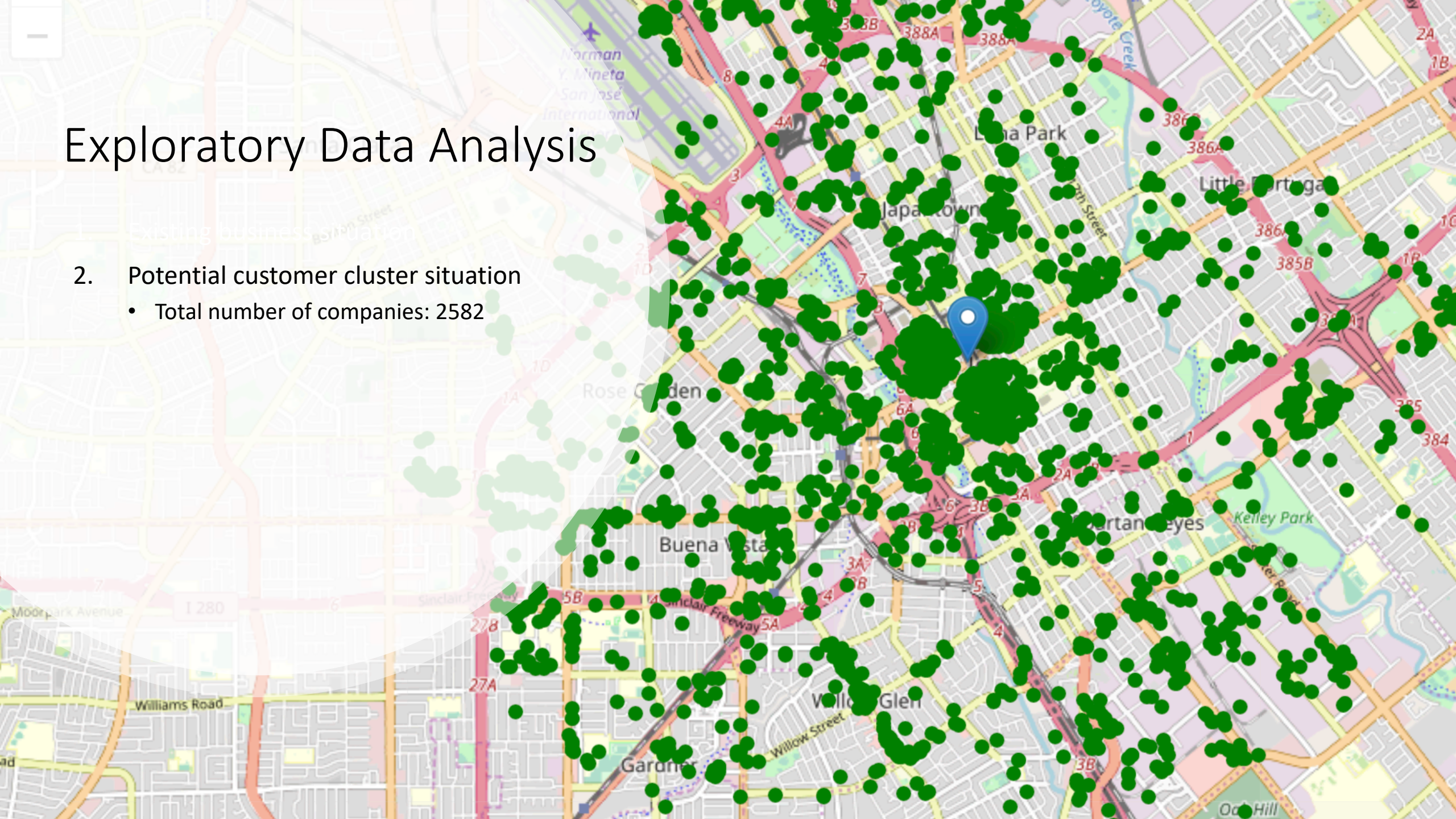
Exploratory Data Analysis

2. Potential customer cluster situation
   - Total number of universities: 199



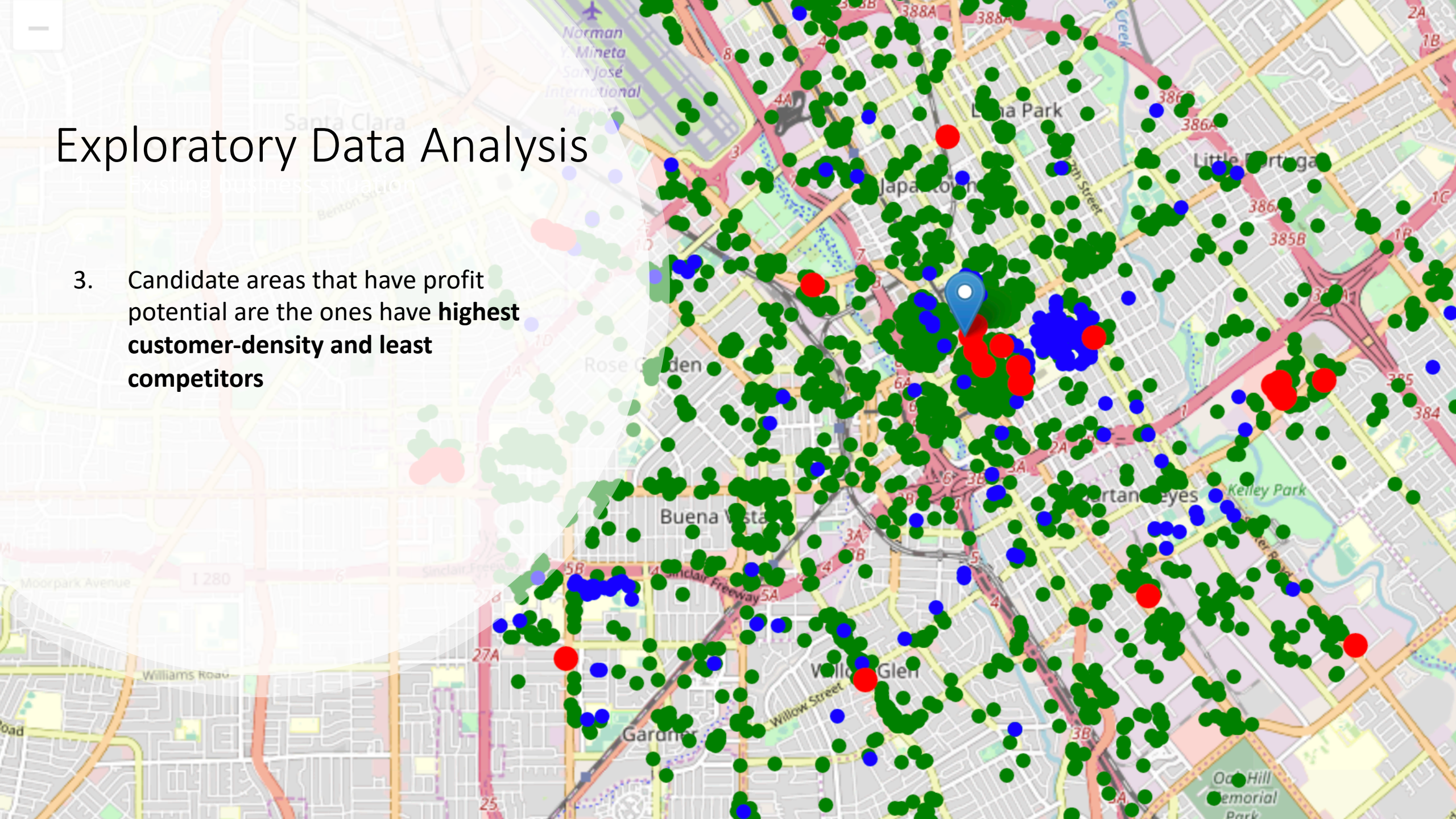
Exploratory Data Analysis

2. Potential customer cluster situation
   - Total number of companies: 2582



Exploratory Data Analysis

3. Candidate areas that have profit potential are the ones have **highest customer-density and least competitors**



Exploratory Data Analysis

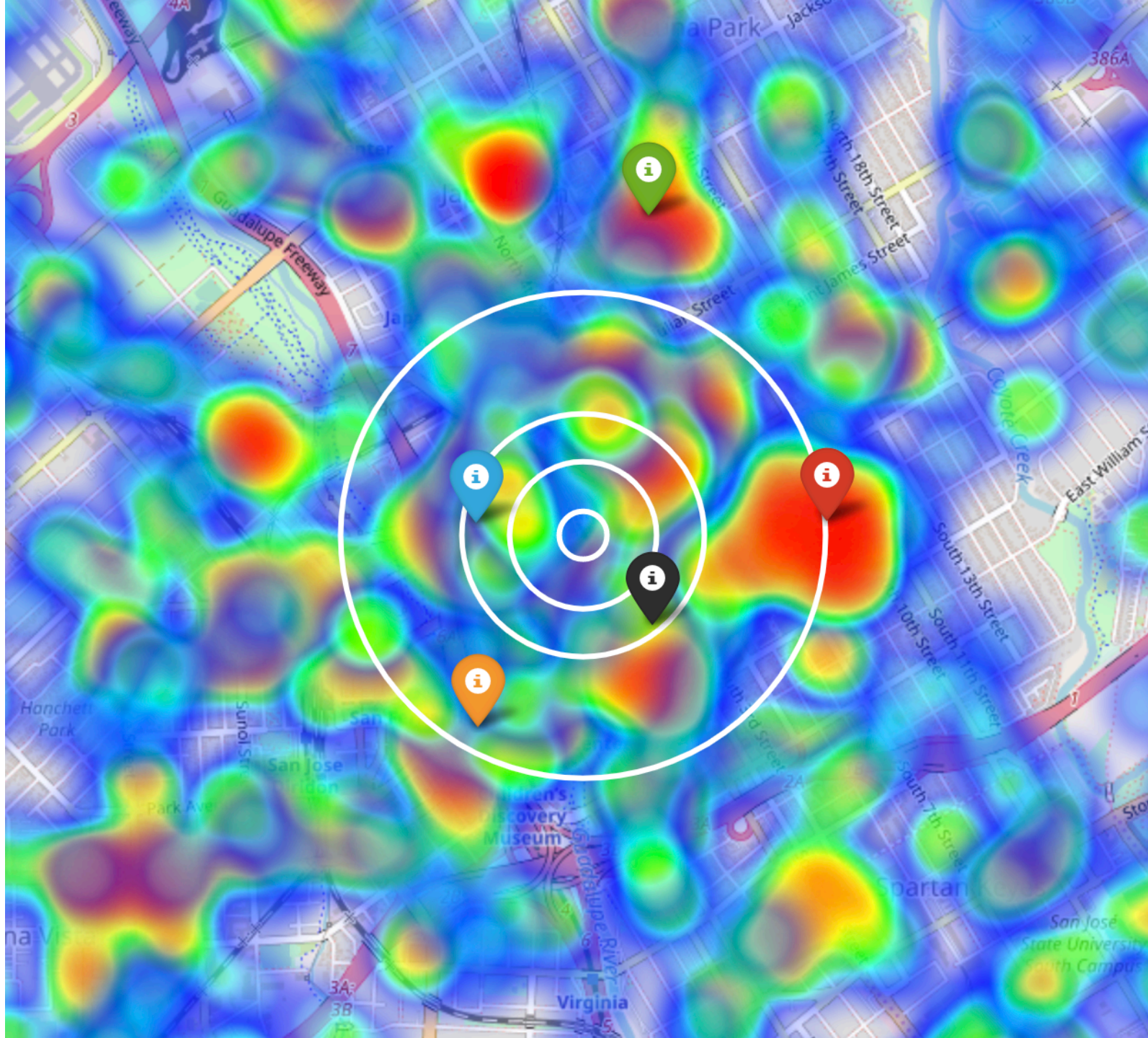
Then we list the top 10 customer dense areas

```
#10 first customer dense areas
df_locations.sort_values(by='Customers in area', ascending=False).head(10)
```

	Address	Latitude	Longitude	X	Y	Distance from center	Customers in area
34	NO ADDRESS	37.336823	-121.879235	-3.388419e+06	1.486822e+07	1200.961894	19
45	NO ADDRESS	37.336739	-121.895642	-3.386919e+06	1.486909e+07	538.874459	15
25	NO ADDRESS	37.348127	-121.887529	-3.386919e+06	1.486736e+07	1610.662367	10
44	NO ADDRESS	37.333000	-121.887408	-3.387919e+06	1.486909e+07	538.874459	8
55	NO ADDRESS	37.329175	-121.895579	-3.387419e+06	1.486995e+07	1066.987298	5
54	NO ADDRESS	37.325436	-121.887346	-3.388419e+06	1.486995e+07	1462.348076	5
35	NO ADDRESS	37.340563	-121.887468	-3.387419e+06	1.486822e+07	665.063509	5
26	NO ADDRESS	37.351867	-121.895764	-3.385919e+06	1.486736e+07	2143.416259	5
48	NO ADDRESS	37.347955	-121.920353	-3.383919e+06	1.486909e+07	3505.764636	4
59	NO ADDRESS	37.344127	-121.928527	-3.383419e+06	1.486995e+07	4139.862545	4

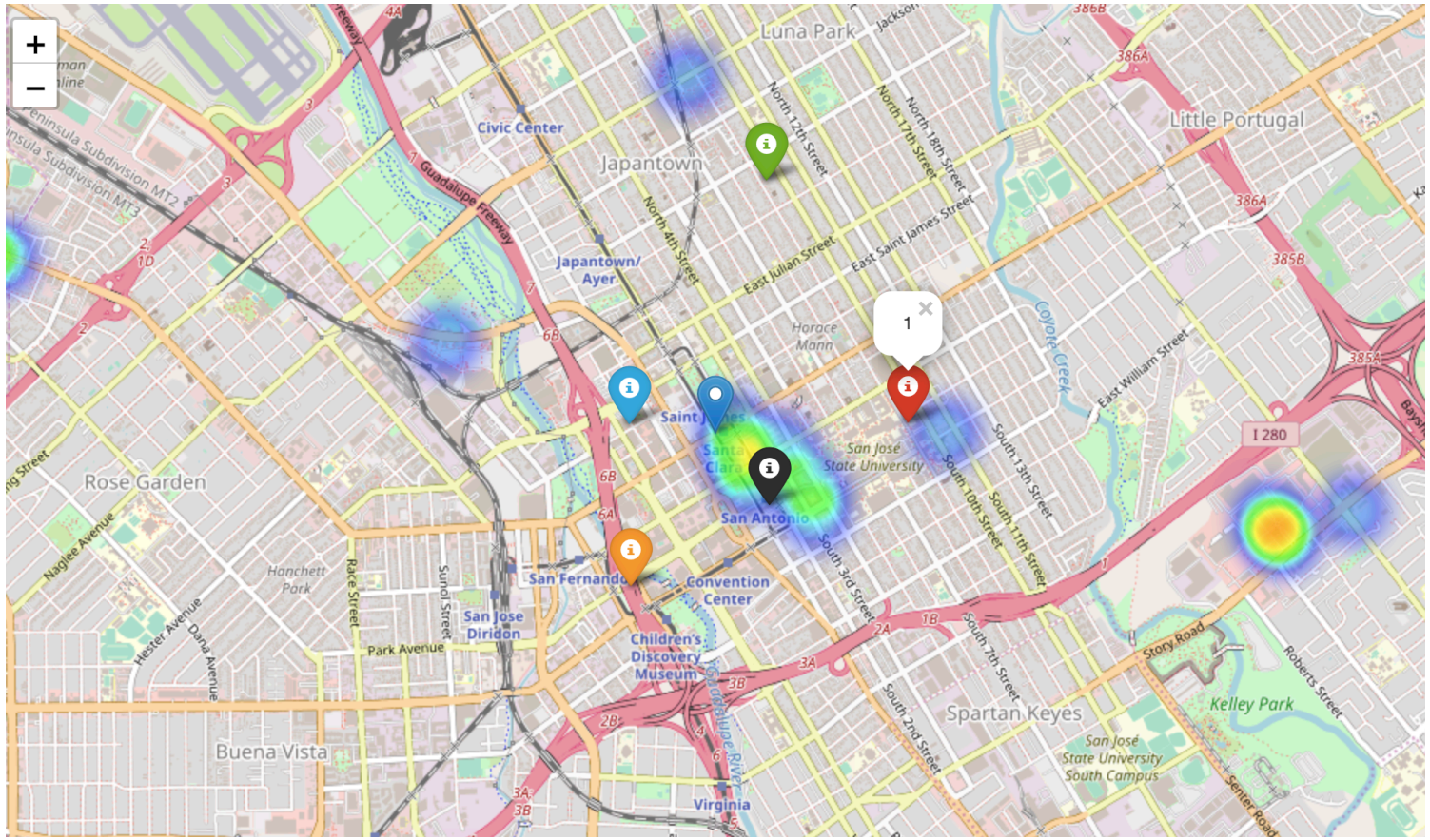
Exploratory Data Analysis

We can see that the customer-dense areas are all around the center of city San Jose



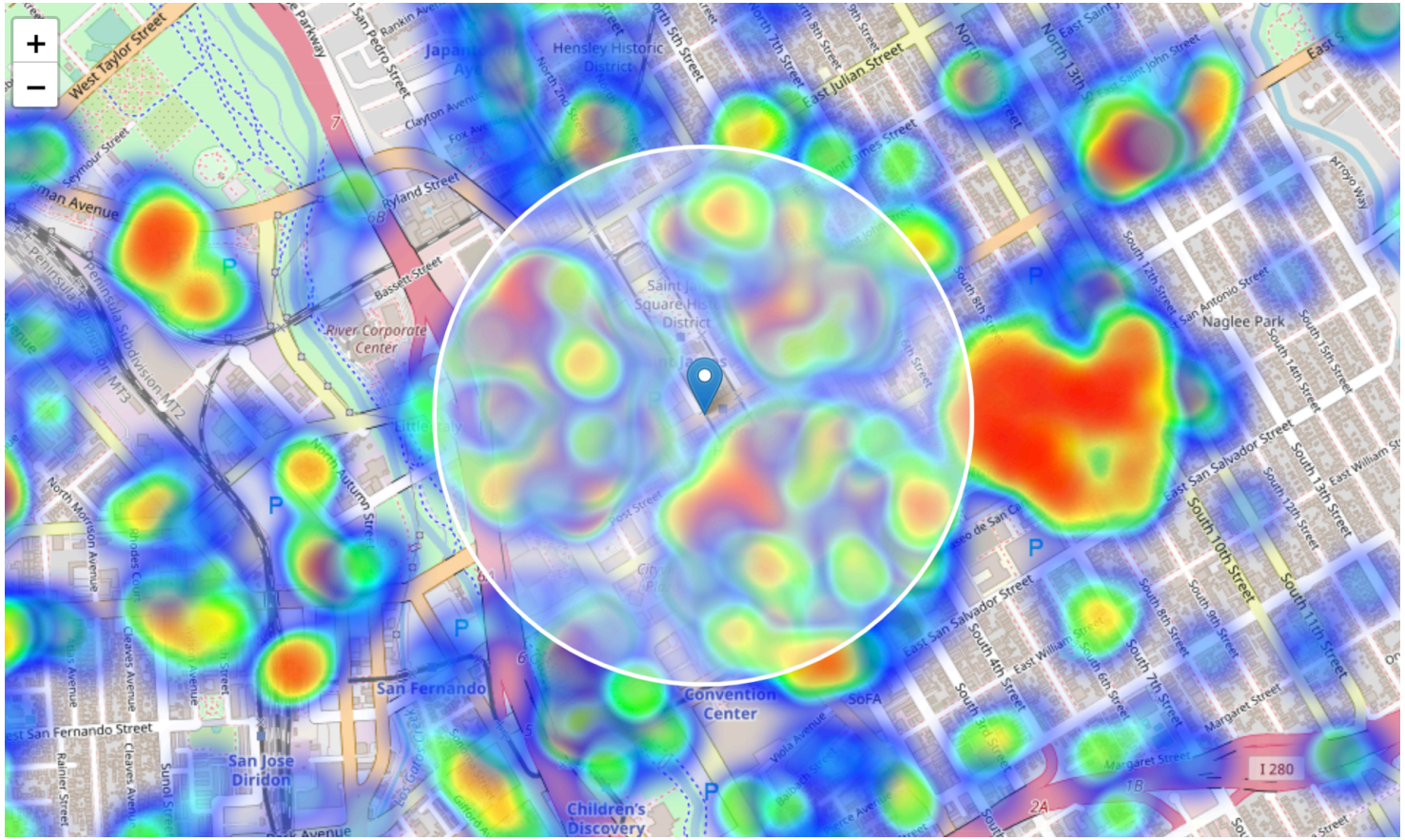
Exploratory Data Analysis

Visualize density of bubble tea shops with heatmap and use info-sign to mark the customer-dense locations



Result and Discussion

- From prior exploration we can see that candidate location should still be focus around the center of the city
- We zoom in to calculate all the data for each venue of the neighborhood
- **563** candidate neighborhood centers generated



Result and Discussion

Then among the 563 location candidates we calculate:

1. Number of bubble tea shops in the vicinity (radius = 150 meters)
2. Number of customers

Result and Discussion

Here is a glance at the candidate locations data frame

Sort with most nearby customers on top

	Latitude	Longitude	X	Y	Bubble Tea Shops nearby	Customers
214	37.338178	-121.894823	-3.386900e+06	1.486888e+07	0	84
189	37.338677	-121.894261	-3.386919e+06	1.486879e+07	0	82
238	37.337164	-121.894249	-3.387019e+06	1.486897e+07	0	80
213	37.337804	-121.894000	-3.387000e+06	1.486888e+07	0	75
239	37.337538	-121.895072	-3.386919e+06	1.486897e+07	0	72
188	37.338303	-121.893438	-3.387019e+06	1.486879e+07	0	68
212	37.337431	-121.893176	-3.387100e+06	1.486888e+07	0	68
237	37.336790	-121.893426	-3.387119e+06	1.486897e+07	0	54
257	37.334422	-121.889870	-3.387600e+06	1.486905e+07	1	48
232	37.334920	-121.889308	-3.387619e+06	1.486897e+07	1	47

Sort with most nearby bubble tea shops on top

	Latitude	Longitude	X	Y	Bubble Tea Shops nearby	Customers
201	37.333317	-121.884119	-3.388200e+06	1.486888e+07	3	18
227	37.333051	-121.885192	-3.388119e+06	1.486897e+07	3	32
207	37.335561	-121.889059	-3.387600e+06	1.486888e+07	2	20
183	37.336433	-121.889321	-3.387519e+06	1.486879e+07	2	4
276	37.331538	-121.885180	-3.388219e+06	1.486914e+07	2	13
231	37.334547	-121.888485	-3.387719e+06	1.486897e+07	2	40
208	37.335935	-121.889883	-3.387500e+06	1.486888e+07	2	10
233	37.335294	-121.890132	-3.387519e+06	1.486897e+07	2	16
226	37.332677	-121.884368	-3.388219e+06	1.486897e+07	2	19
250	37.331805	-121.884107	-3.388300e+06	1.486905e+07	2	6

Result and Discussion

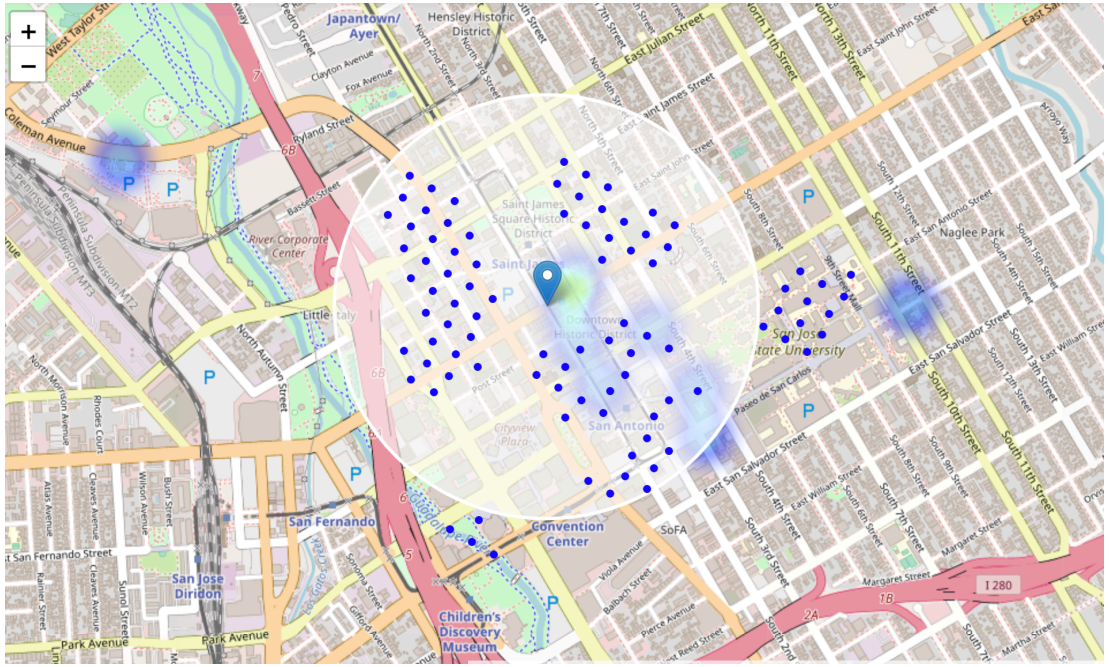
Filter criteria:

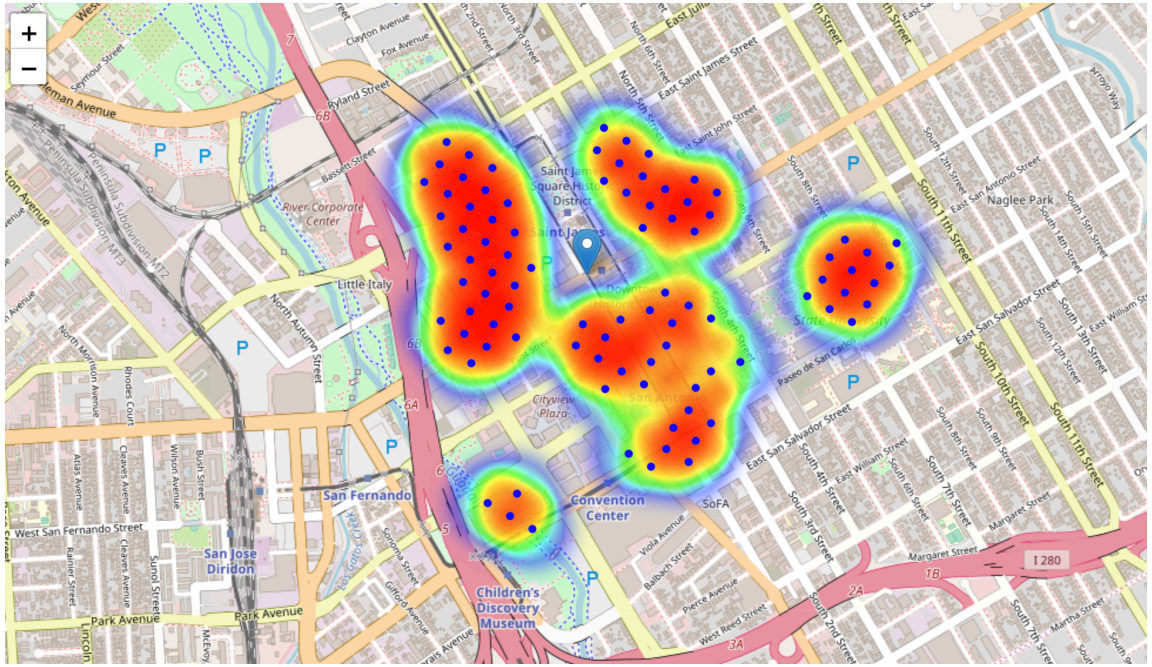
1. Locations have less than 1 (including) bubble tea shops nearby.
2. Locations have more than 20 customers

Select the good ones meet both conditions

```
Locations have no bubble tea shops nearby: 546
Locations have more than 20 customers: 98
Locations with both conditions met: 92
```

Result and Discussion

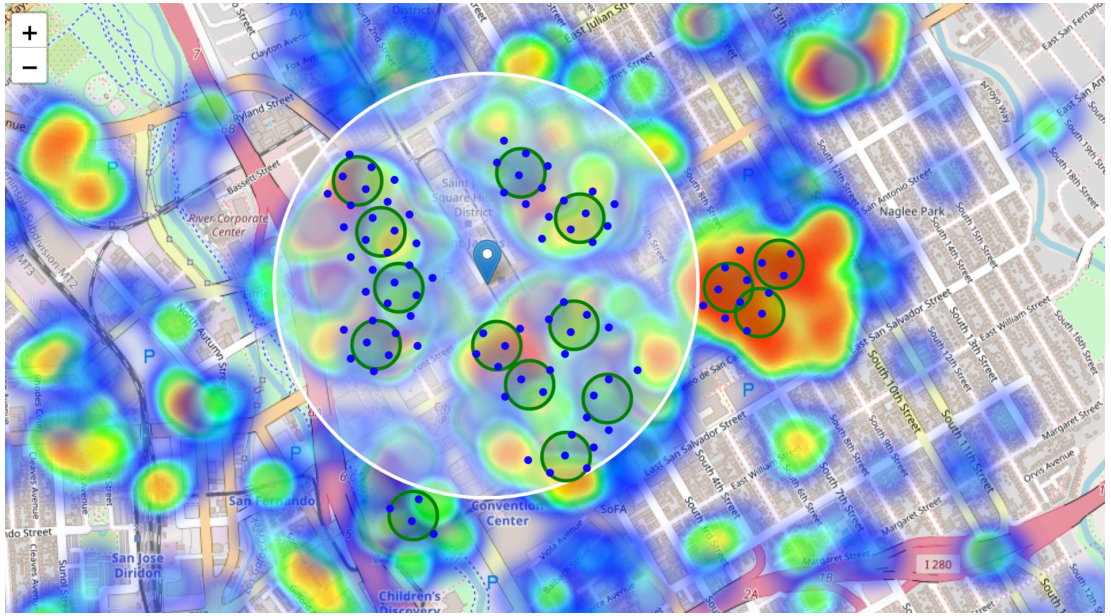




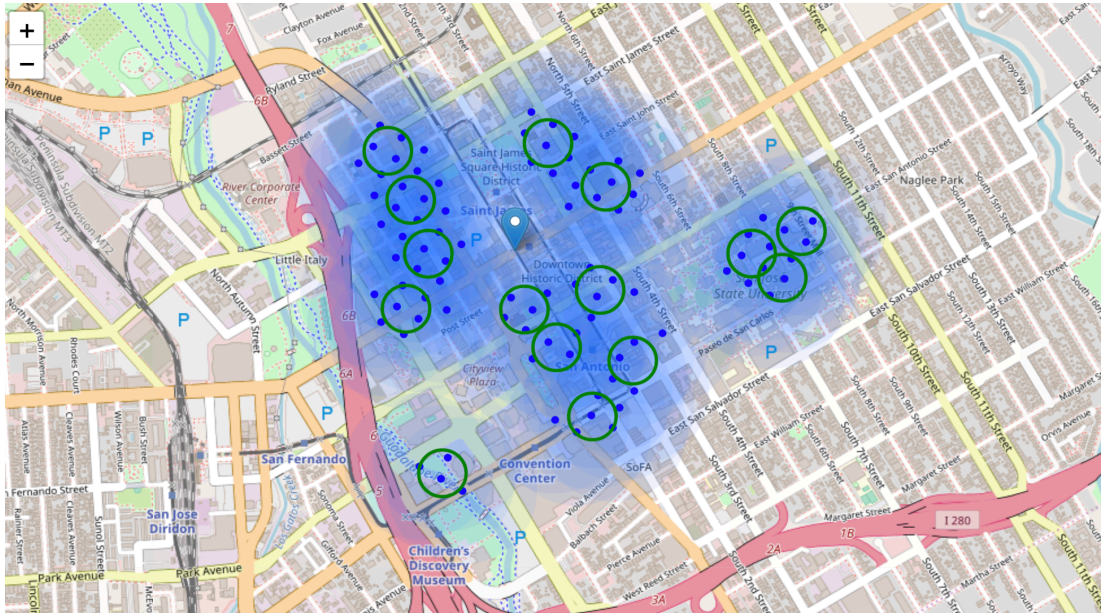
Locations have no bubble tea shops nearby: 546
Locations have more than 20 customers: 98
Locations with both conditions met: 92

Result and Discussion

K means cluster create centers of zones containing good locations.



Heatmap



The shaded areas indicate the clusters

Result and Discussion

On the right is a list of latitude/longit ude of the cluster centers, which is our desired final result set.

User can easily find geocode conve rter on the internet to transform the m into addresses, or use Google cl oud API (It requires credit card infor mation to create billing account, whi ch I didn’t do)

[(-121.89537395222014, 37.339289648214105),
(-121.88731033435064, 37.3350177293404),
(-121.88039257430975, 37.33539209504962),
(-121.88928925207554, 37.33952741630228),
(-121.8933012401226, 37.32938658417022),
(-121.89383752411624, 37.33612939579999),
(-121.89468304830832, 37.33444102449878),
(-121.88760428815883, 37.331132309302276),
(-121.8889627203761 8, 37.33325632393298),
(-121.88709700301852, 37.3381863406691),
(-121.88155550076652, 37.336135699430486),
(-121.89018704724414, 37.33441517945734),
(-121.89450565474245, 37.337782937174474),
(-121.88607274905321, 37.332848052144),
(-121.8796683429025, 37.33682142543998)]

Conclusion and Future Direction

- This analysis can be a good start for business feasibility evaluation, there must be other aspects to consider, the criteria we used for filtering may vary as we understand further about the business environment, which can not be fully covered in this study.
- To be prudent, it is always be good to cross check data with other data sources.
- To be more accurate, it could be possible to give a weight to customers, since younger people who work in a high-tech company have totally different consumer habits from the ones who work in a traditional industry
- This project can be reused for other cities, other lines of businesses. User will need to adjust the parameters, values, criterias to fit their needs.