Capstone Project: The Battle of the Neighborhoods (Week 2)

Author: Maximiliano Rivas

Content

- Introduction: Business Problem
- Data
- Methodology and Analysis
- Results and Discussion
- Conclusions
- References

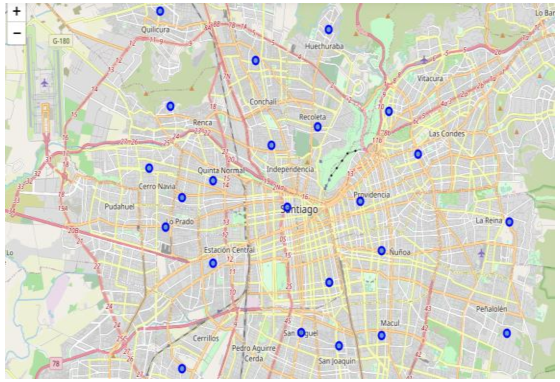
Introduction: Business Problem

- The current project has as objective to discover the best possible location in Santiago de Chile for **starting a new bar entrepreneurship focused on selling premium craft beers to customers**
- Santiago de Chile has around 7MM people living in an area of 15.400 km2 partitioned in 52 different boroughs. **These boroughs have a population, number of houses, area and a large array of different amenities given**
- Our mission is to **leverage different data science tools that helps to determine which boroughs are the most promising and venues inside these ones to locate the stakeholders' bar**
- The criteria used for the analysis are:
  - **Boroughs with high density of people** would be prefer in order to assure a flow of customers into the bar
  - **Boroughs with low number of people per house** will be prefer due to singles and young couples used to visit more bars than families with kids
  - We will prefer **locations with many restaurants, cinemas, theaters, discos and bars** in order to gain exposure for the targeted customers but **avoiding proximity to bars delivering our same services**

Data

- Data about the boroughs and their location

	CUT	Borough	Latitude	Longitude
294	13101	Santiago	-33.437222	-70.657222
295	13102	Cerrillos	-33.500000	-70.716667
296	13103	Cerro Navia	-33.422000	-70.735000
297	13104	Conchalí	-33.380000	-70.675000
298	13105	El Bosque	-33.567000	-70.675000

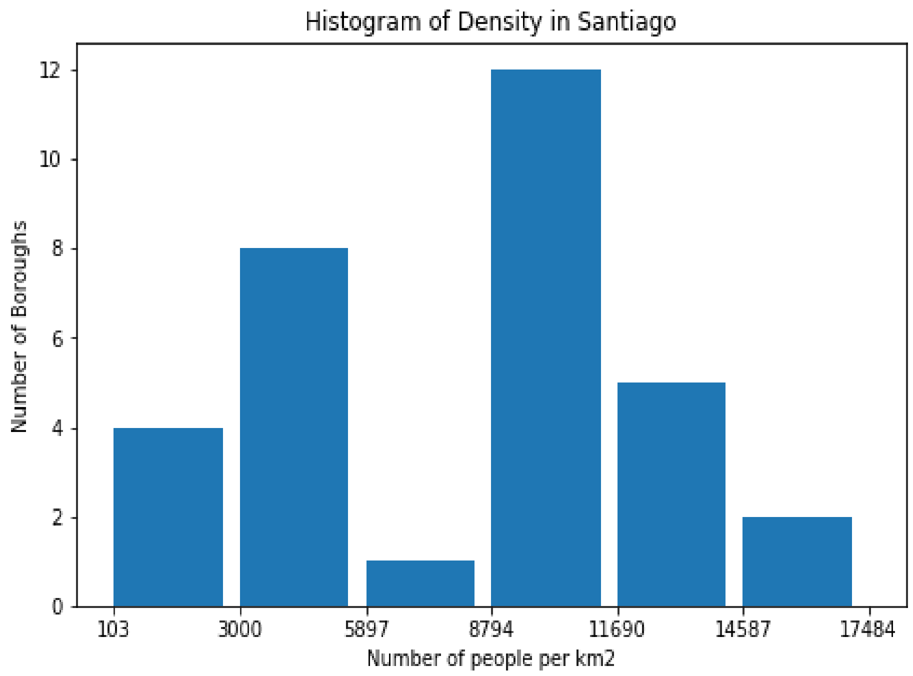


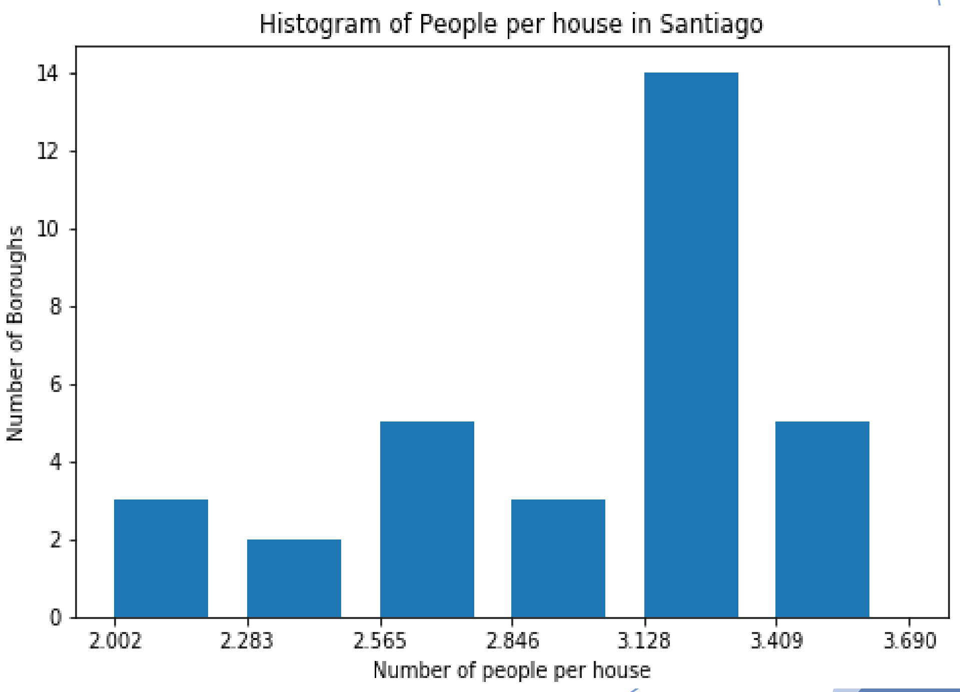
- Data about population, number of houses and area of boroughs

	CUT	Population	Houses	Area (km2)	Density
274	13102	80832	24547	16.779650	4817.263672
275	13131	82900	23855	6.277112	13206.710938
276	13132	85384	31777	28.417034	3004.676758
278	13109	90119	31480	9.979139	9030.739258
283	13113	92787	29801	23.438091	3958.812012

Data: Insights

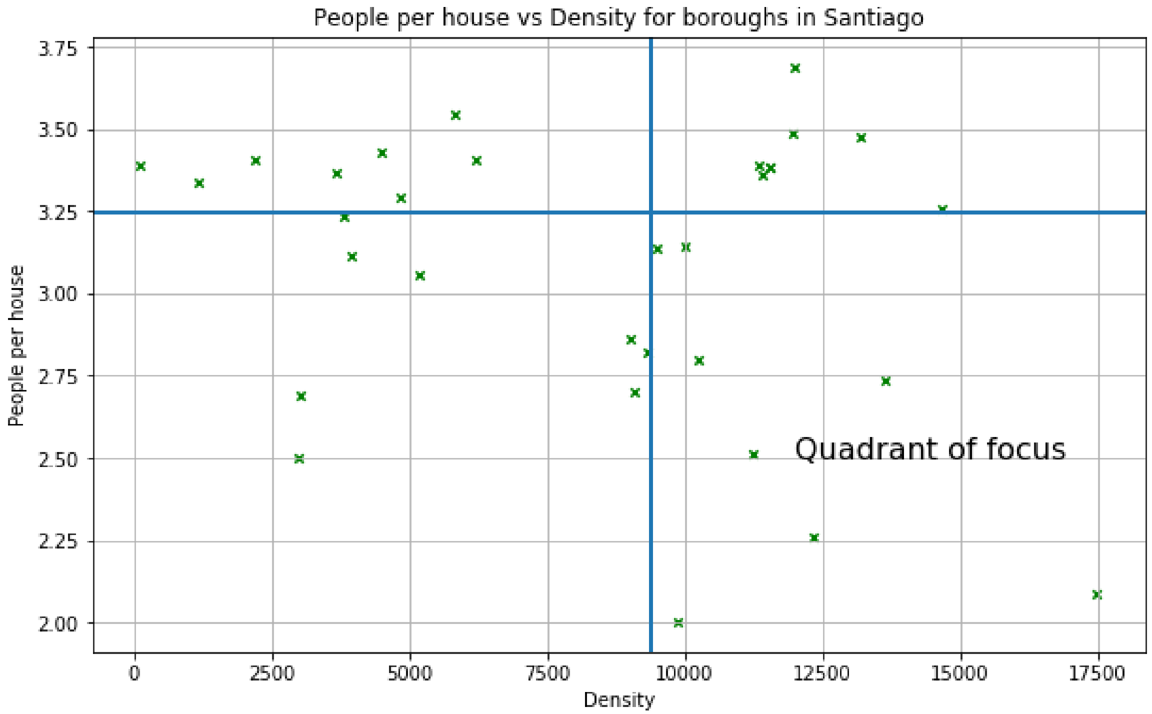
- Around half of the boroughs (15) have low density of people (<8.8k people/km2) and the other half high density (>8.8k people/km2). We must focus on this half
- Around one third of the boroughs (10) have a low number of people per houses (<2.85 people/house). We must focus on this third.





Data: Insights

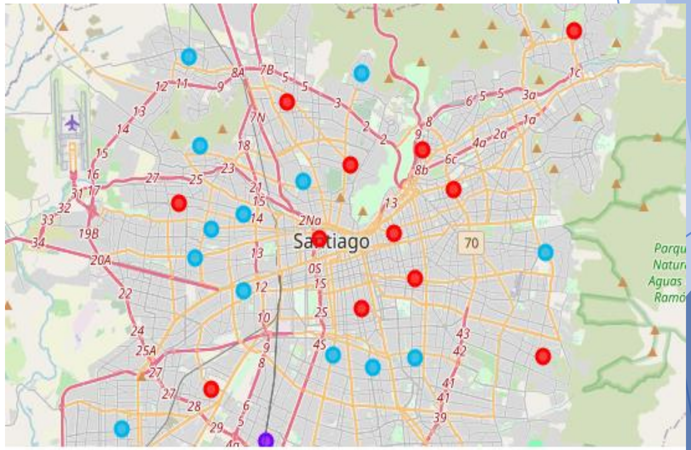
- Finally, with the 2 previous pieces of information we have a quadrant of interest to put focus on



Methodology and Analysis

- **Clustering the boroughs with k-means:**
  - We applied the machine learning clustering technique called K means with a k parameter of 5
  - It is possible to notice that the k-means model creates the cluster 0 and 2 with a group of boroughs while the clusters 1, 3 and 4 only have 1 or 2 boroughs. We will consider these last clusters as outliers for our analysis

	Cluster Labels	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0	Cerrillos	Fast Food Restaurant	Ice Cream Shop	Sandwich Place	Pharmacy	Burger Joint	Movie Theater	Food & Drink Shop	Park	Café	Department Store
1	0	Cerro Navia	Bus Station	Park	Japanese Restaurant	Pharmacy	Burger Joint	Fried Chicken Joint	Flea Market	Mountain	Grocery Store	Chinese Restaurant
2	0	Conchalí	Furniture / Home Store	Department Store	Ice Cream Shop	Pharmacy	Restaurant	Farmers Market	Gym	Bakery	Sushi Restaurant	Donut Shop
3	0	El Bosque	Flea Market	Pharmacy	Pizza Place	Burger Joint	Supermarket	Gastropub	Sushi Restaurant	Mobile Phone Shop	Chinese Restaurant	Bar
4	2	Estación Central	Pharmacy	Bakery	Fast Food Restaurant	Snack Place	Hot Dog Joint	Chinese Restaurant	Museum	Asian Restaurant	Bar	Sushi Restaurant



Methodology and Analysis

- **Slicing the clusters of boroughs according to density and people per house indexes**
  - Applying the median slicing we get a subset of 4 boroughs, with a 'Density' mean of 12.427 and 'People per house' mean of 2.37 for cluster 0
  - Applying the median slicing in the cluster 2 we get a subset of 4 boroughs as well, but a lower 'Density' mean of 11.152 and a higher 'People per house' mean of 2.79
  - So, we decided to continue working with the cluster 0 filtered and its 4 boroughs

```
12427.0280761725 2.3735781606150073
```

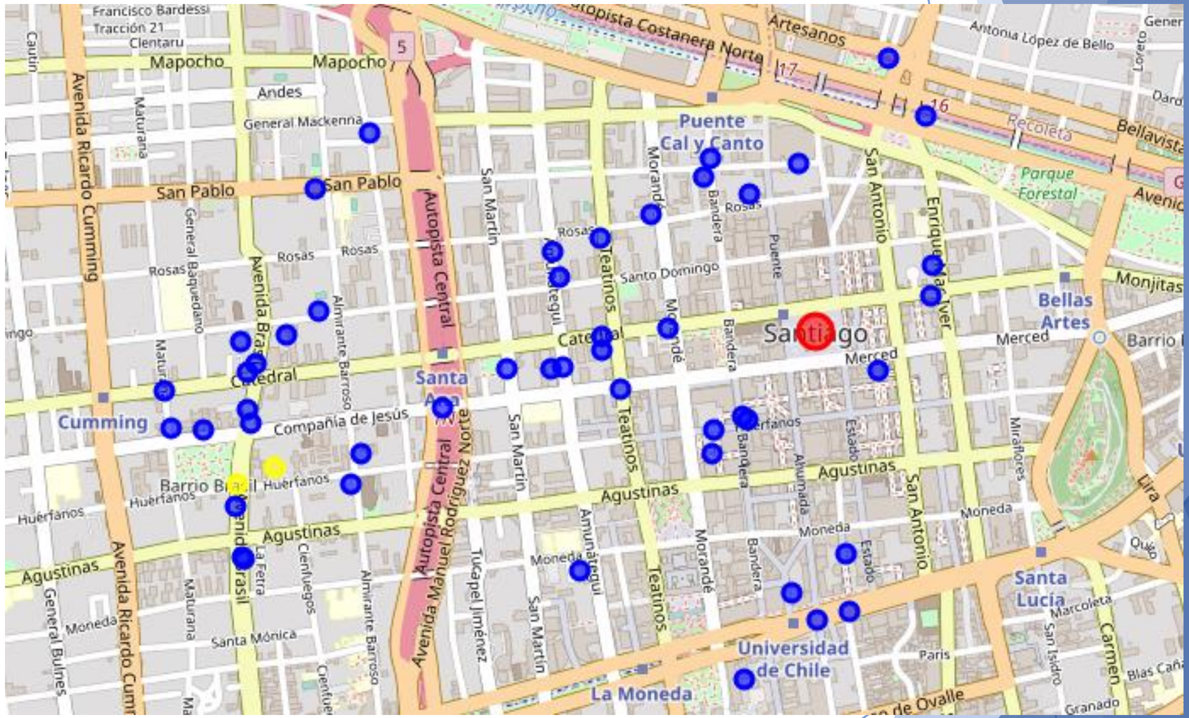
	index	CUT	Borough	Latitude	Longitude	Population	Houses	Area (km2)	Density	People/House	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Mc Comm Ven
0	0	13101	Santiago	-33.437222	-70.657222	404495	193628	23.135237	17483.935547	2.089032	...	Coffee Shop	Art Museum	Tea Room	Bookstc
1	19	13120	Ñuñoa	-33.454000	-70.604000	208237	92248	16.856802	12353.292969	2.257361	...	Coffee Shop	Peruvian Restaurant	Pizza Place	Ca
2	22	13123	Providencia	-33.435000	-70.616000	142079	70965	14.394146	9870.609375	2.002100	...	French Restaurant	Park	Pizza Place	Ho
3	26	13127	Recoleta	-33.406000	-70.640000	157851	50178	15.784667	10000.274414	3.145821	...	Park	Nightclub	Pharmacy	Bake

4 rows × 21 columns

	index	Borough	Houses	Area (km2)	Density	People/House	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Mc Comm Ven
0	5	Estación Central	52486	14.353261	10244.431641	2.801528	2	Pharmacy	Bakery	Fast Food Restaurant	Snack Place	Hot Dog Joint	Chinese Restaurant	Muse
1	7	Independencia	36666	7.355460	13633.545898	2.734986	2	Restaurant	Sandwich Place	Park	Peruvian Restaurant	Chinese Restaurant	Farmers Market	C
2	28	San Joaquín	30096	9.942262	9504.075195	3.139686	2	Food Truck	Restaurant	Theater	Farmers Market	Sandwich Place	Chinese Restaurant	BBQ Jc
3	29	San Miguel	42947	9.613154	11229.821289	2.513656	2	Sushi Restaurant	Pizza Place	Restaurant	Plaza	Peruvian Restaurant	Park	La Americ Restau

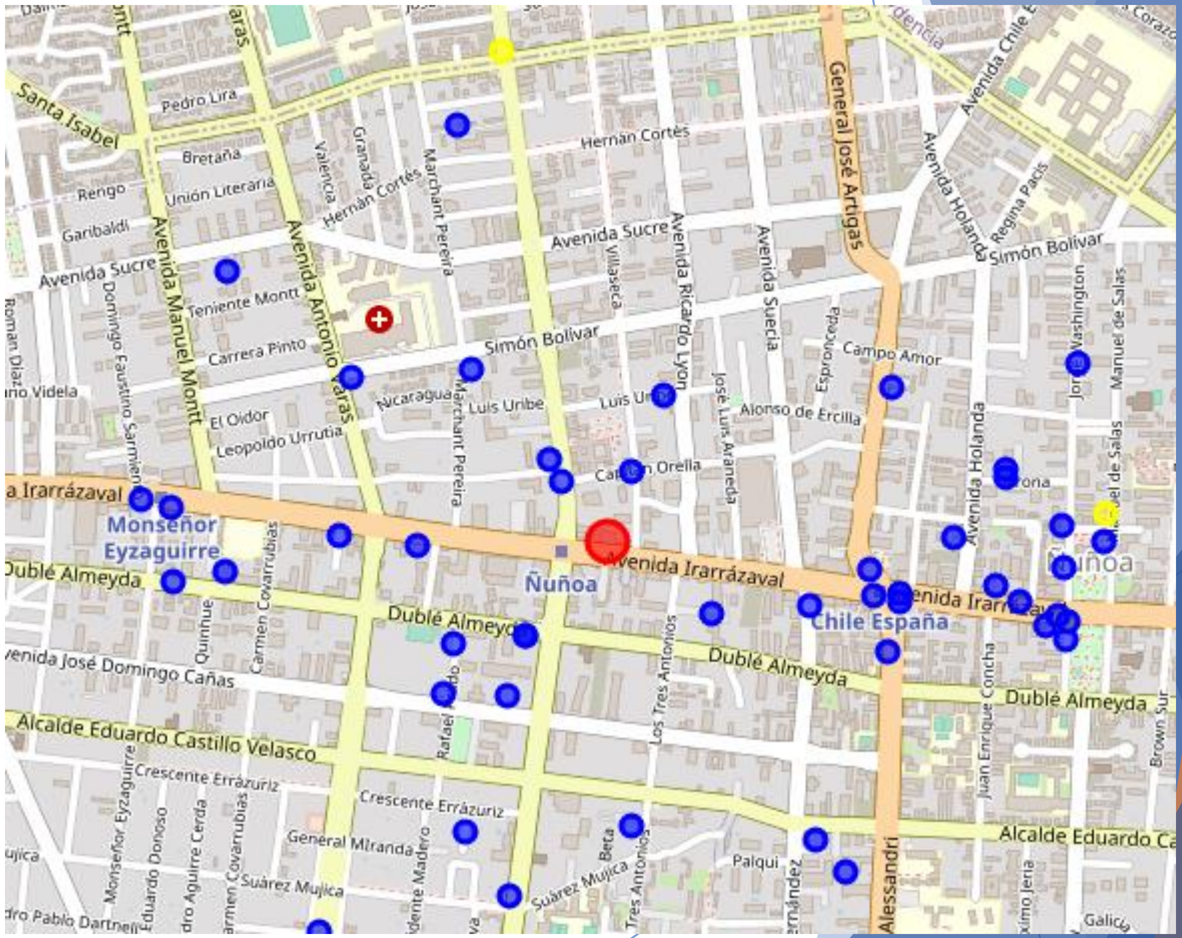
Methodology and Analysis

- Here we extract all the bars within the **Santiago Borough**
- And map them, coloring in yellow those bars specialized in selling beers
- There are plenty of bars in the nearby but only 2 of them specialized in beers so this is a potential neighborhood



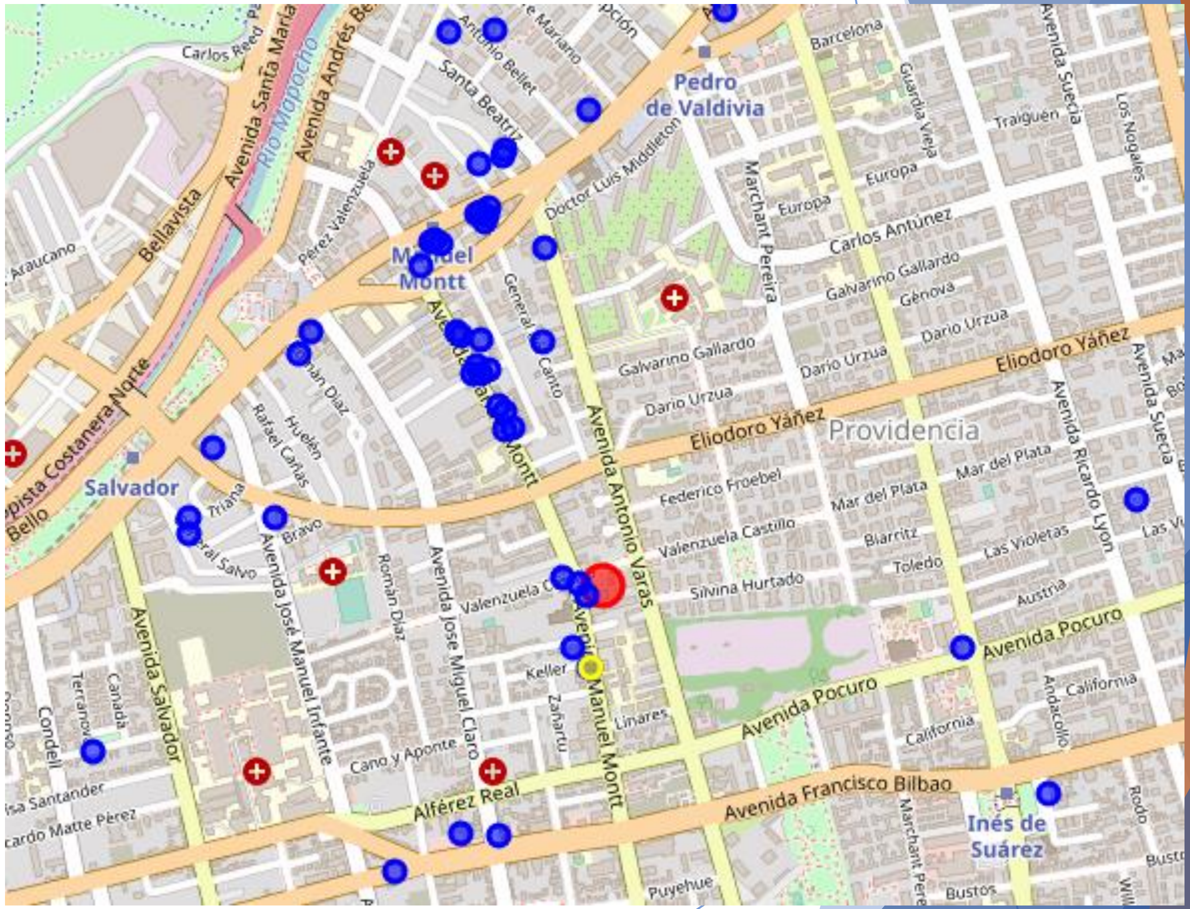
Methodology and Analysis

- Here we extract all the bars within the **Ñuñoa Borough**
- And map them, coloring in yellow those bars specialized in selling beers
- Again, there are plenty of bars in the nearby but only 2 of them specialized in beers, which are far apart each other, so this is a potential neighborhood



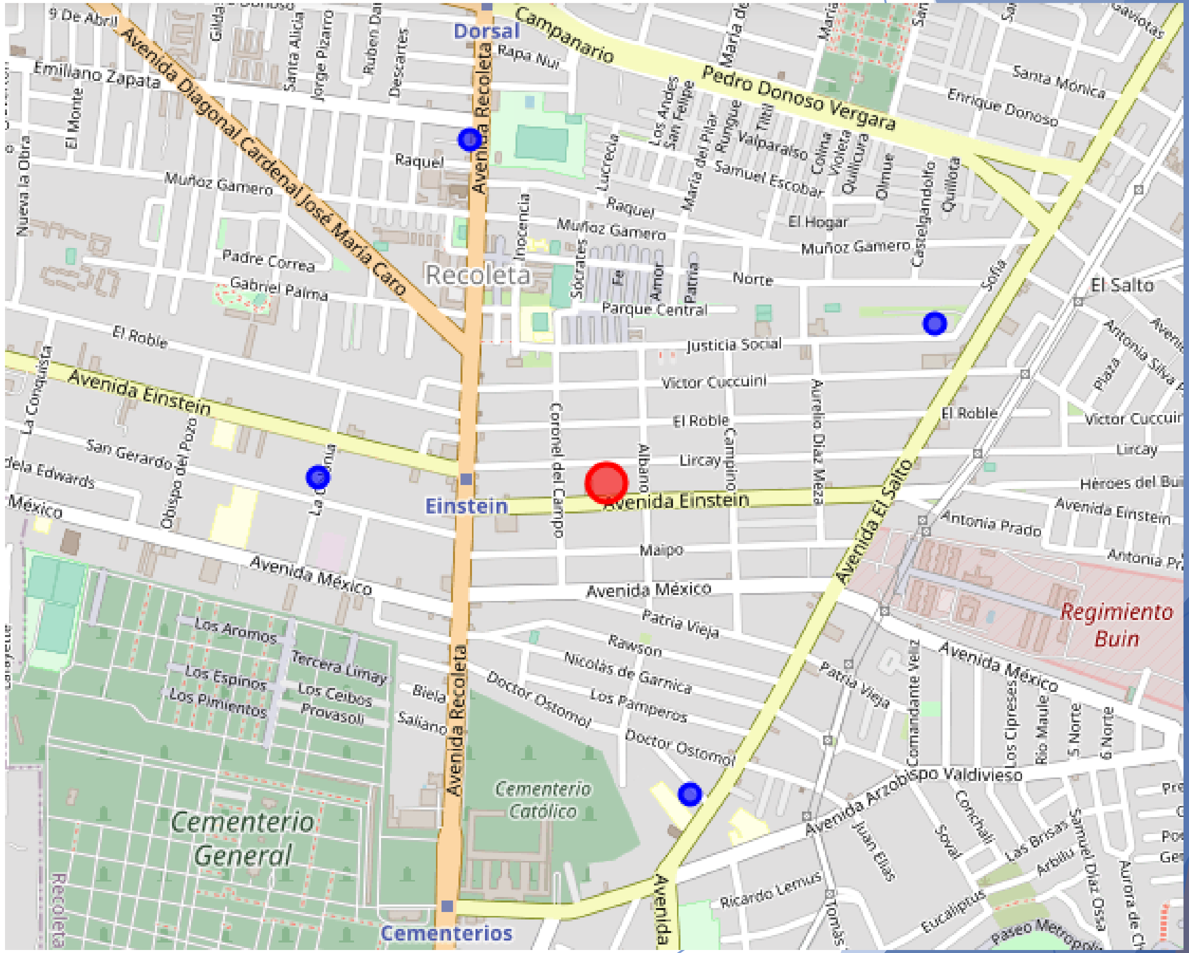
Methodology and Analysis

- Here we extract all the bars within the **Providencia Borough**
- And map them, coloring in yellow those bars specialized in selling beers
- Again, there are plenty of bars in the nearby but only 1 of them specialized in beers and far from the high density of bars, so this is another potential neighborhood



Methodology and Analysis

- Here we extract all the bars within the **Recoleta Borough**
- In this case we map the bars but there are only 4 of them far apart each other
- We consider it is not a neighborhood with potential for our beer bar



Results and Discussion

- We got that there 3 boroughs with neighborhoods within them capable of sustain and assurance a new premium craft beer.:
  - Santiago
  - Ñuñoa
  - Providencia
- Because there are plenty of ‘complementary' amenities in a range of 1km around the 'heart' of each borough which are helpful to:
  - Gain exposure for target customers
  - A constant flow of people
  - A lack of competitive landscape (only 1 or 2 bars specialized in beers), which can be helpful in catch the lovers of beer niche
- Contrasting the results with reality it is highly feasible that a new specialized beer bar could start here, these are well-known zones of restaurants, bars, clubs and nocturnal entertainment services.
- It is necessary though to do:
  - A further analysis in the place to locate this new bar
  - All the economies associated with the entrepreneurship as investment and costs
  - A measure of proximity to workplace and public transport modes.

Conclusions

- Analyzing important business decision as location could be done easily with data science knowledge and a few open sources of information
- It is possible to leverage machine learning techniques to clustering information without the need to build hard mathematical models or having a deep knowledge about the data
- Foursquare API is a powerful tool to gain insights about location of amenities, it is helpful for analyzing potential places for new business as in our case
- Data science analysis need to be done with a common and business sense in order to find shortcuts that allow us to save time and energy, as in our case with the density and people per house analysis rational
- It is the first step for developing the beer bar idea but one of the most important and with huge impact in the revenues and growth of business on the future

References

- Foursquare API
- Wikipedia (comunas de Chile)
- INE (Población comunas de Chile)