PLACING A TACO STAND

Using Foursquare API and K-means clustering algorithm

RODRIGO SALAZAR

2 AUGUST 2020

Introduction

Are you going to open a restaurant, but you do not know where to place it?

Locating a restaurant has never been an easy task. A few years ago, owners would have to collect data by driving around the targeted area and be asking people's opinions to gather data as much as possible to ensure a successful business.

Now days we have Foursquare!! This API lets us gather venue data for any specific regions, using this API you have data in minutes about nearby restaurants, gyms, all types of restaurants, bars, etc.

So, in this document we will utilize Foursquare to help clients define where would be based to locate a taco stand and ensure them a successful business.

Data & References

First, we need to define the area where we would like to locate the taco stand.

Using google maps [1] we selected longitude and latitude was around Los Angeles, CA.

Also, we defined some boundaries. (Image 1)

In order to retrieve information from every city and postal code we used the web page Open data soft [2] and we got all the postal codes for the entire California.

	city	zip	dst	geopoint	longitude	state	latitude	timezone
0	Calexico	92232	1	[33.026203, -115.284581]	-115.284581	CA	33.026203	-8
1	Goshen	93227	1	[36.357151, -119.425371]	-119.425371	CA	36.357151	-8
2	Huron	93234	1	[36.209815, -120.0847]	-120.084700	CA	36.209815	-8
3	June Lake	93529	1	[37.765218, -119.07769]	-119.077690	CA	37.765218	-8
4	Fresno	93761	1	[36.746375, -119.639658]	-119.639658	CA	36.746375	-8

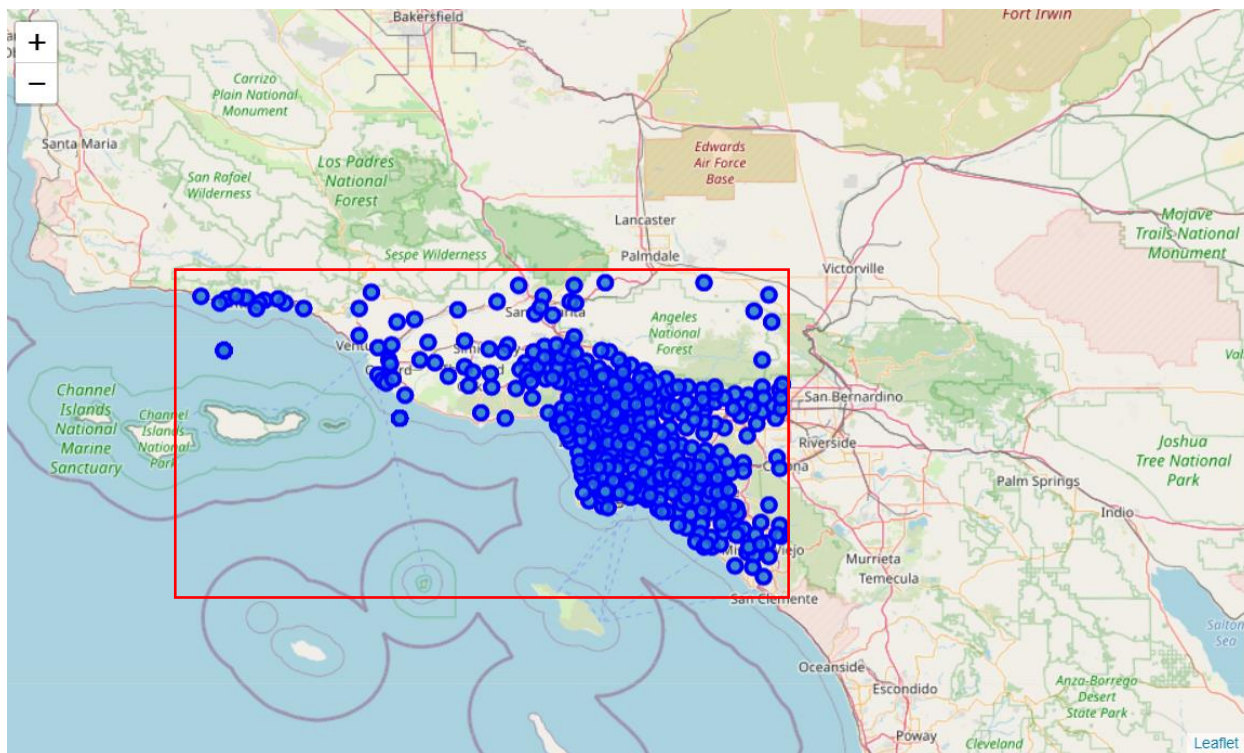Also, we analyzed some information about the income of the people living in the surrounding areas. [3]

	code	name	state	statefips	agi_stub	zipcode	number	amount	year	agi_group
2	02650	Total income	CA	6	6	90504	790	231326	2017	$100,000 or more
8	02650	Total income	CA	6	6	90706	370	113438	2017	$100,000 or more
13	02650	Total income	CA	6	6	90011	20	9774	2017	$100,000 or more
15	02650	Total income	CA	6	6	90018	260	107299	2017	$100,000 or more
21	02650	Total income	CA	6	6	90031	180	68242	2017	$100,000 or more

Finally, for the heuristics of how to select a good locations we used this guide made by restaurant engine [4]

References

[1] "Google maps," [Online]. Available: https://www.google.com/maps/.

[2] "Open data soft," [Online]. Available: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=CA.

[3] "USA IRS Zipcode data," [Online]. Available: https://public.opendatasoft.com/explore/dataset/usa-irs-zipcode-data/table/?disjunctive.zipcode&disjunctive.agi_group&disjunctive.name.

[4] "Restaurantengine," [Online]. Available: https://restaurantengine.com/great-restaurant-location/#:~:text=Consider%20the%20downtown%20location%2C%20or,usually%20good%20spots%20for%20restaurants..

Methodology

(Image 1): All zip codes used.

For the analysis of the data I utilized the Foursquare API to explore every postal code from the selected area and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each postal code from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

|	name	categories	lat	lng
0	EDGE Performing Arts Center	Dance Studio	34.088245	-118.329057
1	Gold's Gym	Gym	34.089513	-118.329800
2	Cactus Taqueria	Taco Place	34.088149	-118.326515
3	Project Angel Food	Non-Profit	34.087750	-118.326441
4	The Pack Theater	Indie Theater	34.090662	-118.330680

Then, we defined some categories that would influence our business that we considered important to attract people to eat tacos. The venue categories were: **Bars, Nightclubs, Mexican Restaurants**, and another **Taco stands**.
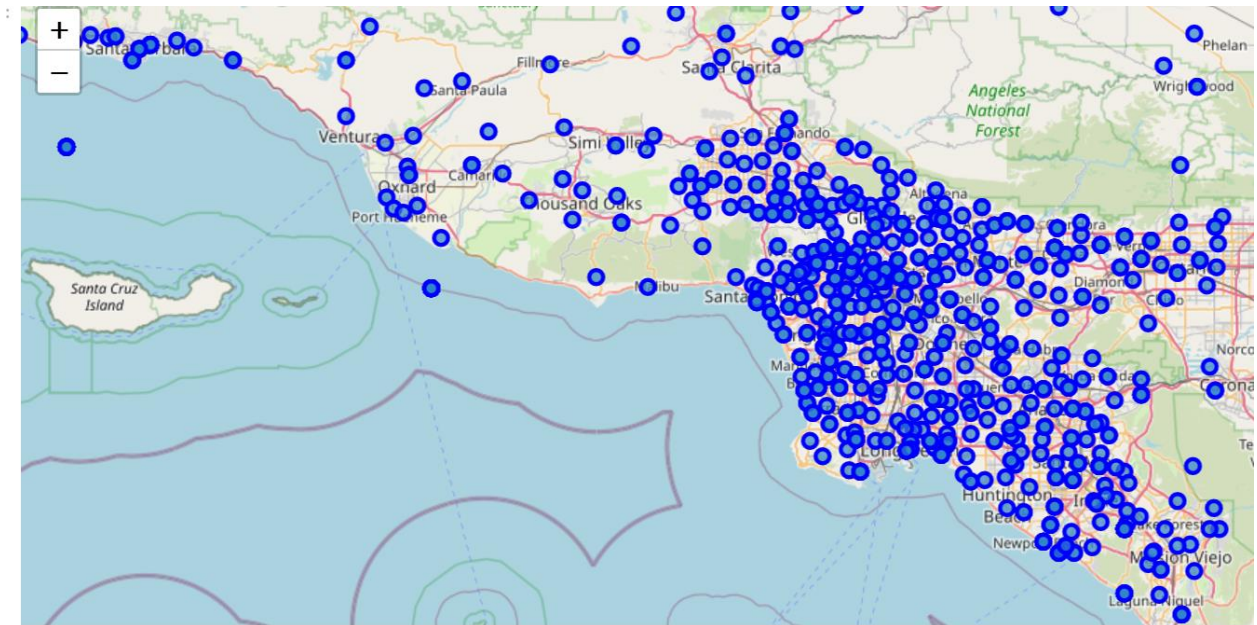
(Image 2): Filtered venues: **Bars, Nightclubs, Mexican Restaurants**, and **Taco stands**.

Clustering

Then we used the k means clustering algorithm to fit 20 clusters to all our venue data.
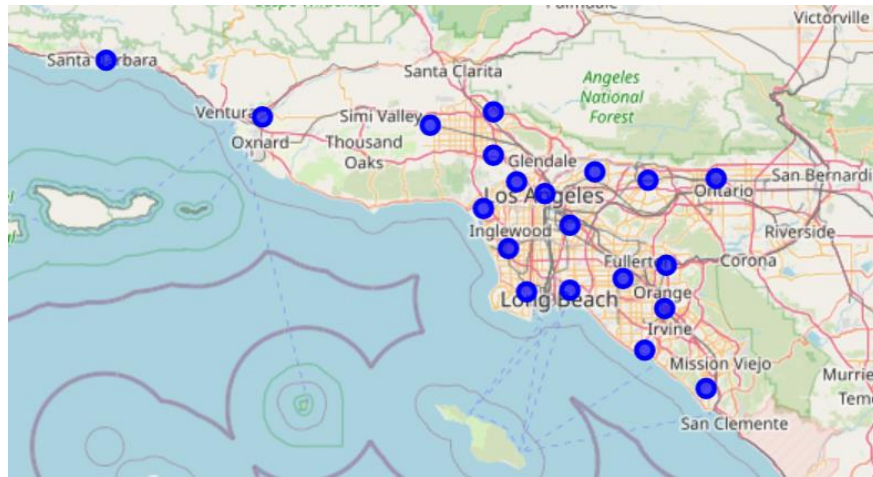

(Image 3): Important venues clustered.

This is to distribute our venues equally as potential geometrical centroids to place our restaurant.

To select which location to pick out of all those 20, we analyzed the **IRS data** [3] and filtered by the zip codes that had the most income.

We used households with yearly earnings of more than $100,000.

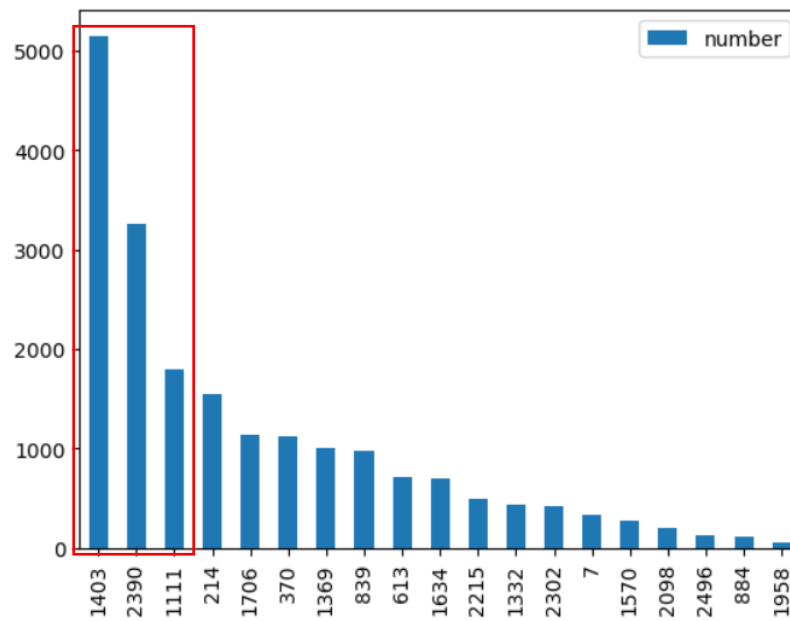With this IRS information we were able to have a score for the 20 selected points.
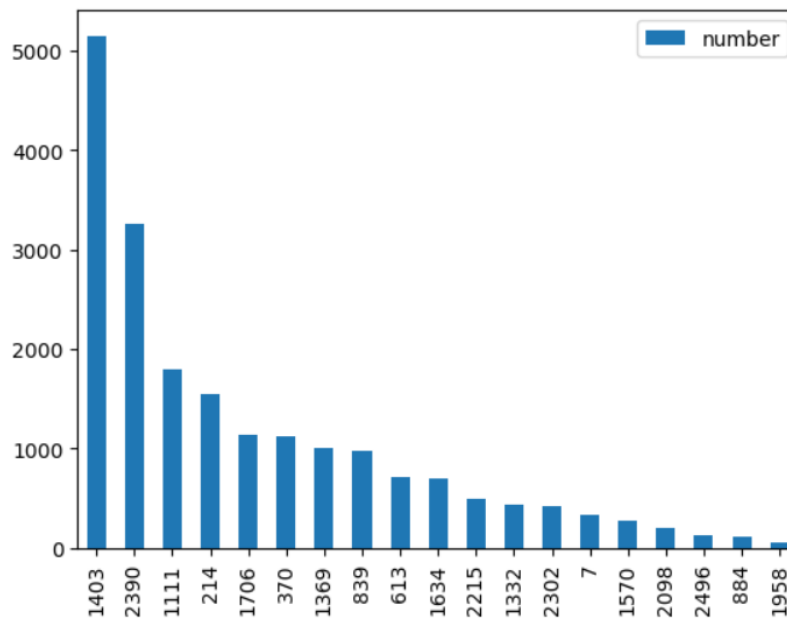
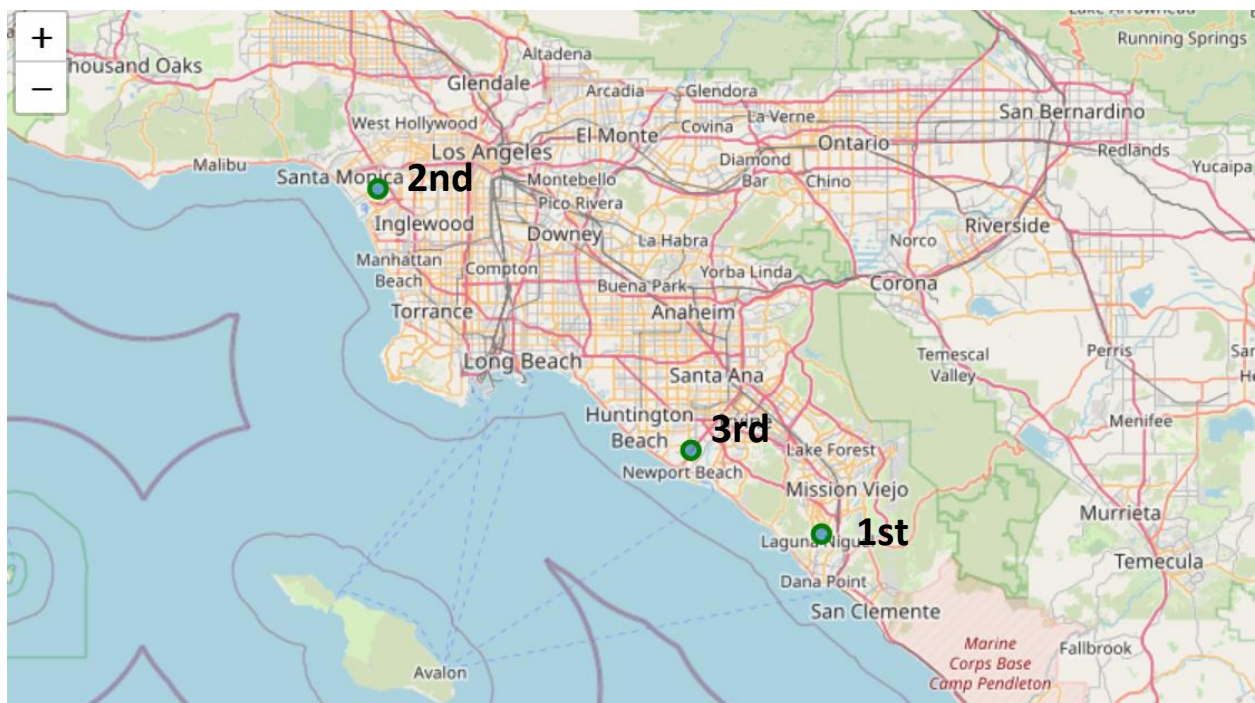(Image 4): Axis y: Number of households nearby with more income than $100,000

Axis x: identifying number for the cluster location.

After reviewing this plot, we selected the top 3 candidates to be considered as the best options to place the taco place.

Results

(Image 5): Resulting top 3 locations to place the taco stand.

The three resulting locations match with people with high household income.

We assume that selecting one of this location will ensure the following:

- Overall trade area is growing or stable
- have a minimum traffic count of 24,000, with a small-town population of 7,000 -10,000 within a 3-mile area.
- They can afford the restaurant
- There is enough people in the vicinity
- Minimum of 36 parking stalls is desired.
- Good accessibility into location
- Retail support ideally is within one mile of site.
- Lower crime rate
- Unemployment rate is optimally less than 10%.

Observations: As a further analysis we could increase the search radius from the Foursquare API, this might change the results a certain percentage.

Also, the selection of the venues: Bars, Nightclubs, Mexican Restaurants, and Taco stands was empiric. Which may not be the best options to choose from. This were selected as a targeted audience. We think that audience have a positive correlation between buying tacos and attending those type of venues.

Conclusions

The Foursquare API provides enormous of valuable information that could benefit businesses define store locations. When this information is paired with other datasets like IRS or data from the department of justice could offer a very good initial analysis to help clients decide where to do an investment.

At first, I thought that the best option to place a taco stand would be near Los Angeles downtown. But, based on this information, I would advise the clients to go invest at the proposed locations.

Obviously, I could do further analysis of how other venues impact the location of the taco stands and try to target our audience better.

If the taco stand is positioned at the spotted areas, the client just needs to focus on delivering high quality tacos. That will make a good reputation around the area and will continue to have business in the future years to come.