



OPENING A MODERN MEXICAN RESTAURANT IN LOS ANGELES



Pictures from Phoenix Magazine, Thrillist, LA weekly, The Independent

- **What is Modern Mexican food?**

- **Modern Mexican food focuses more on innovation and fine ingredients, it evolved out of the traditional Mexican food and not so much into the authenticity, which is quite distinguishable at the first time you see the dishes.**

- **It is healthier, fancier and distinctive, that's also the reason it is particularly attractive to younger generation who would like to try out new things and can afford the "higher than usual" price.**

The screenshot shows a table with columns: ZIP Code, Postal City 1, and Location. The data includes:

ZIP Code	Postal City 1	Location
0	90713	Lakewood 90713(33.84871142900005, -118.11357922799999)
1	91306	Winnetka 91306(34.208404020000046, -118.57593995299999)
2	90002	Los Angeles 90002(33.94895070600006, -118.24697958699994)
3	90506	Torrance 90506(33.88535286100006, -118.32659746799999)
4	90069	West Hollywood 90069(34.08940300900008, -118.37978902499998)
...

**COUNTY OF LOS ANGELES
OPEN DATA**

Home Catalog GIS Catalog How to Videos Developer Tools Your Ideas

HOW MANY ZIP CODES IN THE LOS ANGELES COUNTY?

- Los Angeles is a major metropolitan district of US, it hosts more than 10 million residents, including 3.5 million immigrants, most of which have arrived after 1980. Given the enormous population, business and density in the area, there are more than 300 zip codes assigned to the county.
- The zip code coordinates data was provided by county of Los Angeles website which was originally from USPS
- In total there are 370 zip codes in the data frame

USING ONE HOT ENCODING TO CATEGORIZE THE ZIP CODES IN LA AREA

- Data Acquisition:
- The data will be acquired from Foursquare API, including the venues names, coordinates and their categories in each zip codes.
- The Radius was set to be 1000, total 18462 venues were received.

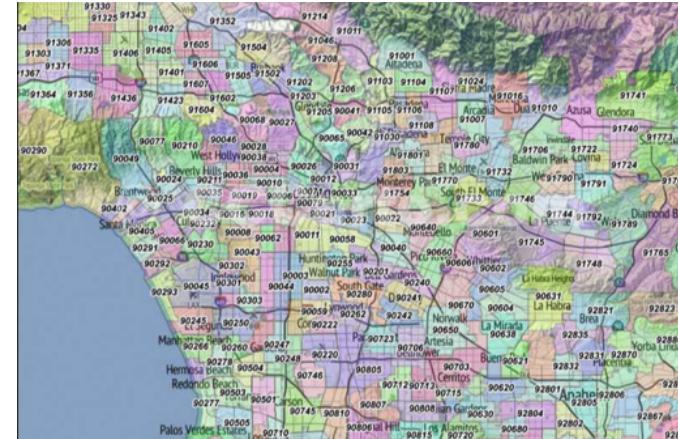
Results Snippet:

the total number of venues are **18462***

PostCode	PostCode	Latitude	PostCode	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	90713	33.848711	-118.113579	California Sushi & Teriyaki	33.846888	-118.115684		Japanese Restaurant
1	90713	33.848711	-118.113579	Starbucks	33.846731	-118.116127		Coffee Shop
2	90713	33.848711	-118.113579	99 Cents Only Stores	33.847387	-118.117002		Discount Store
3	90713	33.848711	-118.113579	Daiso Japan	33.856183	-118.116483		Discount Store
4	90713	33.848711	-118.113579	Starbucks	33.856889	-118.115436		Coffee Shop



FOURSQUARE



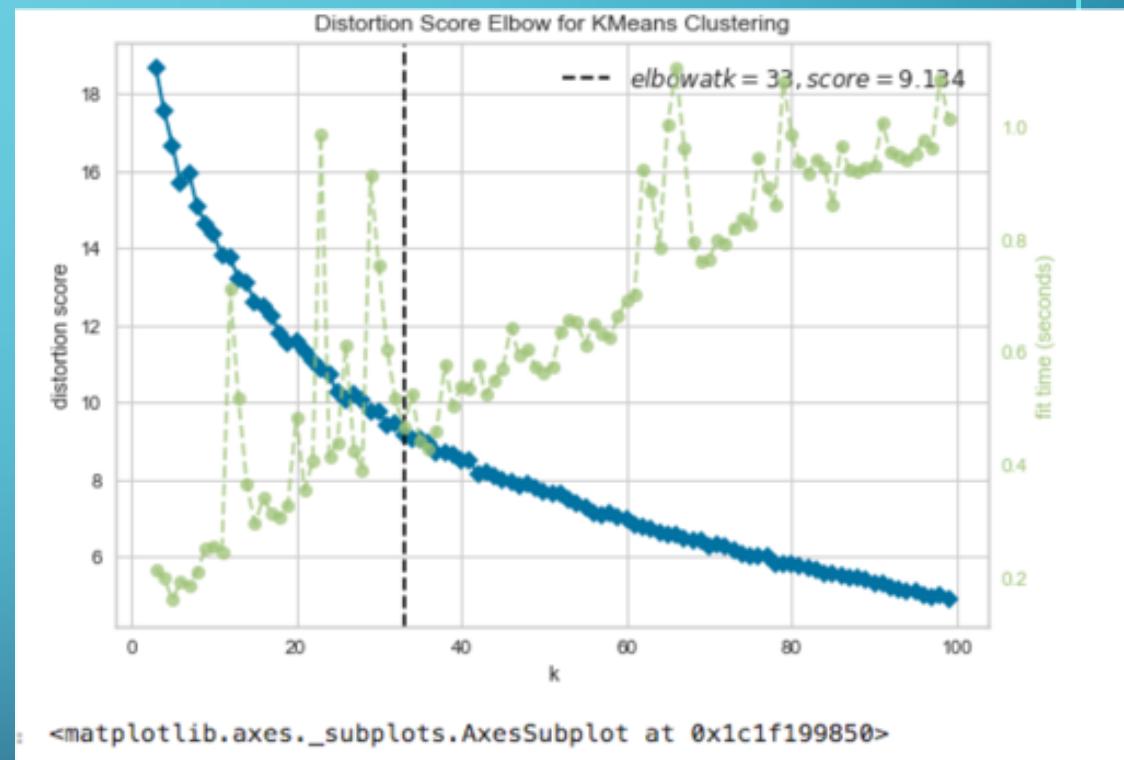
What is the best k value for KMeans

By using Yellow Brick package, I visualized the relationship between distortion score and k values, in the search for the 'elbow' point. However, even with the range upto 100, there is still no obvious elbow can be found. Therefore, I chose k=10 to start with, reasonably big but not too big that that data got segregated too much.

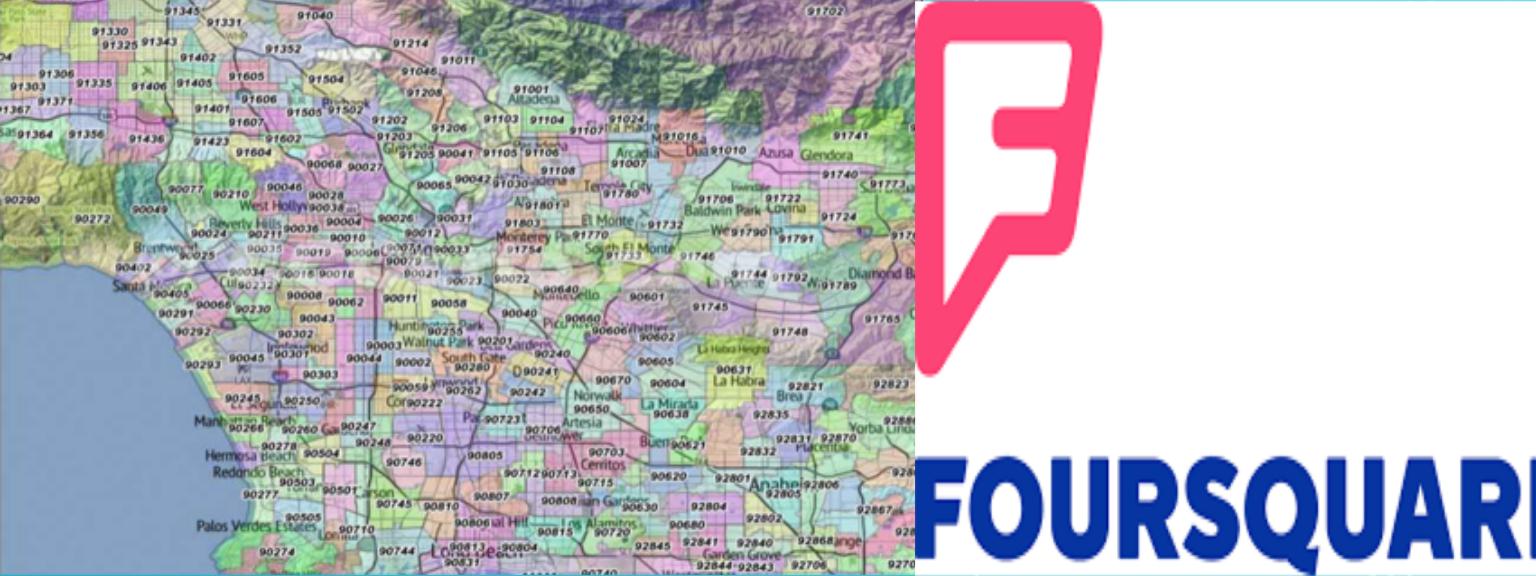
Code Snippet

```
from yellowbrick.cluster import KElbowVisualizer ##### Yellow Brick

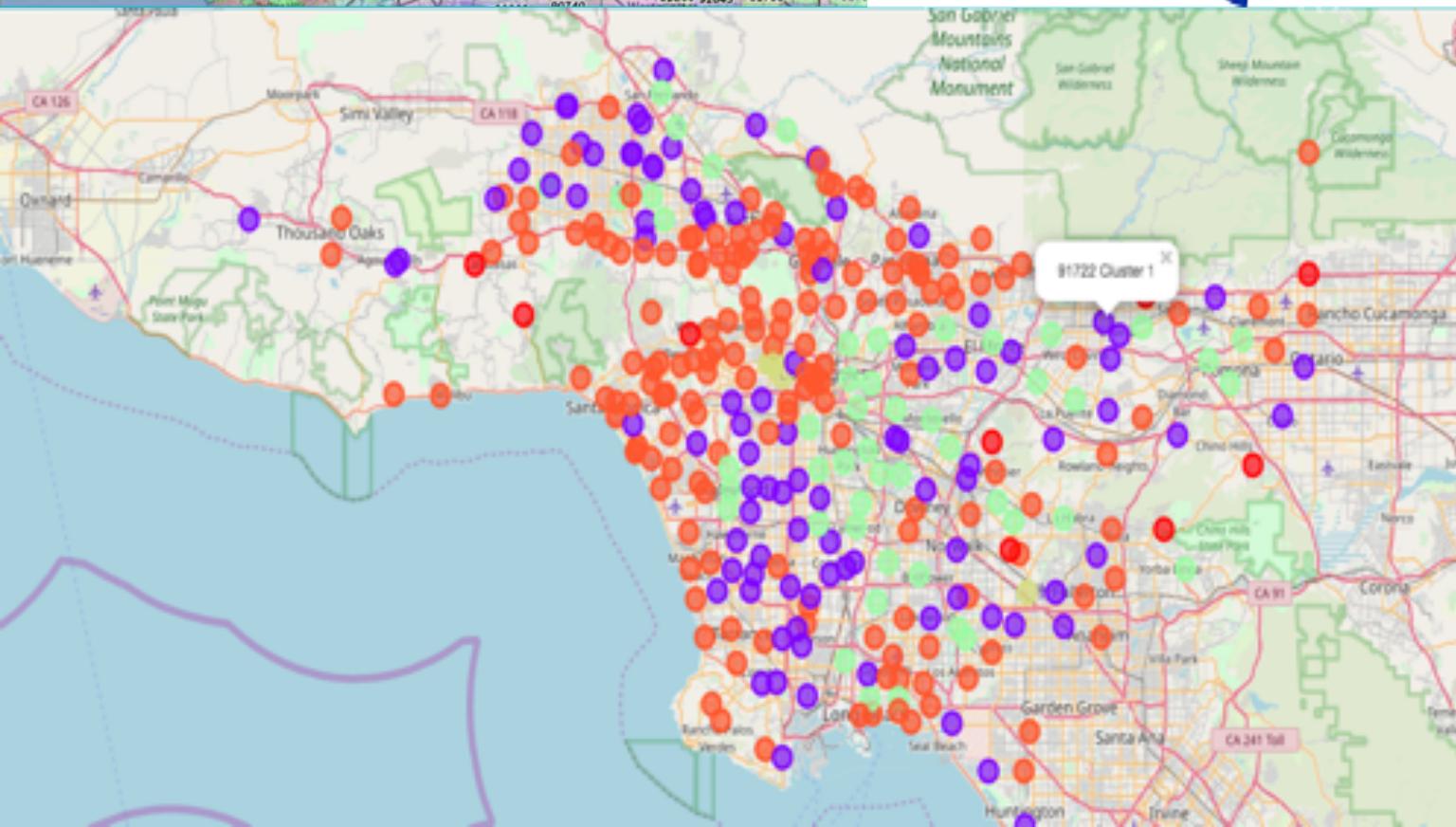
model=KMeans()
visualizer= KElbowVisualizer(model, k=(3,100))
visualizer.fit(grouped_drop)
visualizer.show()
```



No clear elbow point indicates the venues across the zip codes are not that distinct



FOURSQUARE



USING ONE HOT ENCODING TO CATEGORIZE THE ZIP CODES IN LA AREA

Visualization of the zip codes categorization by folium map.

From the figure we can see that there are 3 major groups: cluster 1, cluster 6 and cluster 9

```
la_cluster[la_cluster.labels==6].head()
```

	zipcode	city	latitude	longitude	labels	1st most frequent venue	2nd most frequent venue	3th most frequent venue	4th most frequent venue	5th most frequent venue	6th most frequent venue	7
1	91306	Winnetka	34.208404	-118.575940	6	Pizza Place	Mexican Restaurant	Convenience Store	Video Store	Grocery Store	Italian Restaurant	
8	91702	Azusa	34.127122	-117.906209	6	Mexican Restaurant	Convenience Store	Fast Food Restaurant	Bank	Park	Coffee Shop	
23	91724	Covina	34.091232	-117.858336	6	Park	Pizza Place	Liquor Store	Coffee Shop	Tree	Mexican Restaurant	
35	90033	Los Angeles	34.048564	-118.211520	6	Mexican Restaurant	Fast Food Restaurant	Bakery	Seafood Restaurant	Burger Joint	Bookstore	
40	91008	Duarte	34.148373	-117.965966	6	Bar	Business Service	Sushi Restaurant	American Restaurant	Mexican Restaurant	Convenience Store	E

From the output above we can see that cluster6 has the most Mexican restaurants, therefore we pick cluster6 to continue

WHICH CLUSTER IS THE ONE WE WANT?

Investigating the population in the Los Angeles area



To find out the population characteristics in LA area, I went to US census website, downloaded published data with their online tools.

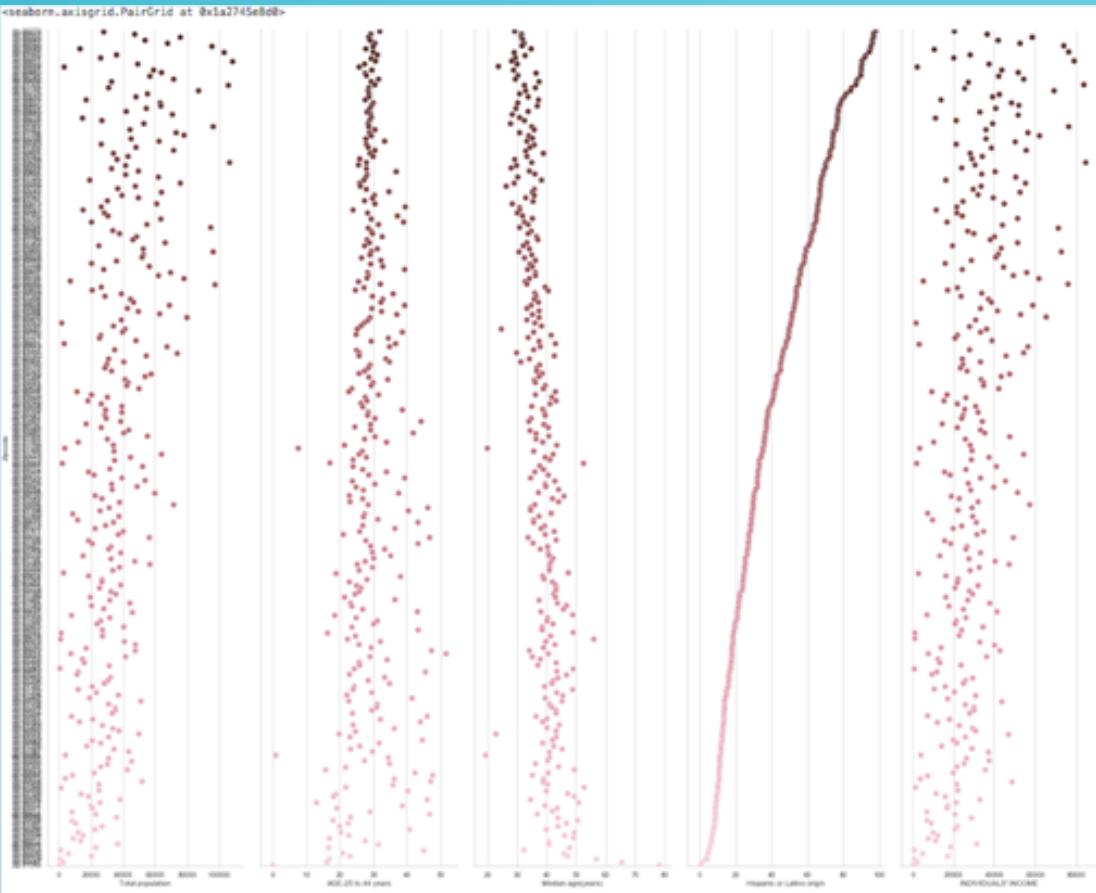
Data from 2011 to 2017 was used

There are more than 400 characteristics listed in the downloaded data. I chose the most relevant 5 for our downstream study, they are:

- 'Zipcode'
- "Total population"
- "AGE-25 to 44 years"
- "Median age(years)"
- "Hispanic or Latino origin"
- "INDIVIDUALS' INCOME"

GEO.id2	HC01_EST_VC01	HC01_EST_VC06	HC01_EST_VC12	HC01_EST_VC28	HC01_EST_VC53
0	Id2 Total; Estimate; Total population	Total; Estimate; AGE - 25 to 44 years	Total; Estimate; Median age (years)	Total; Estimate; Hispanic or Latino origin (of...	Total; Estimate; INDIVIDUALS' INCOME IN THE PA...
1	90001	54760	29.8	27	39668
2	90002	49475	27.6	25.8	34903
3	90003	66200	27.9	26.2	46773
4	90004	62760	34.1	35.8	52273

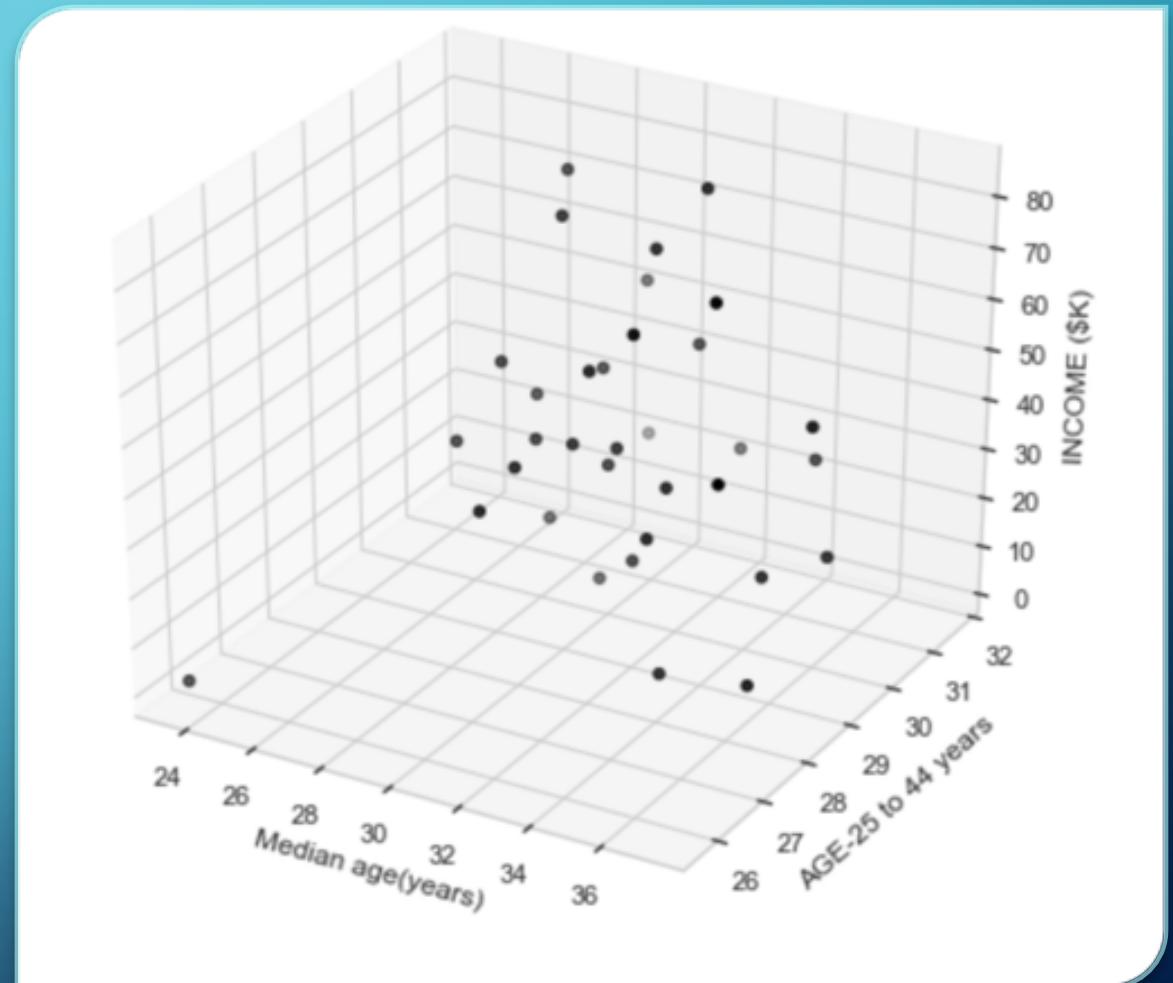
Strip plot for visualization



Based on the notion that more potential customers is positively correlated with potentially more business, I used Seaborn package's stripplot to display the five parameters from year 2017: 'Zipcode', "Total population", "AGE-25 to 44 years", "Median age(years)", "Hispanic or Latino origin", "INDIVIDUALS' INCOME", with sorted "Hispanic or Latino origin" value. The figure shows that the median age is negatively correlated with this value, while others are slightly positively correlated.

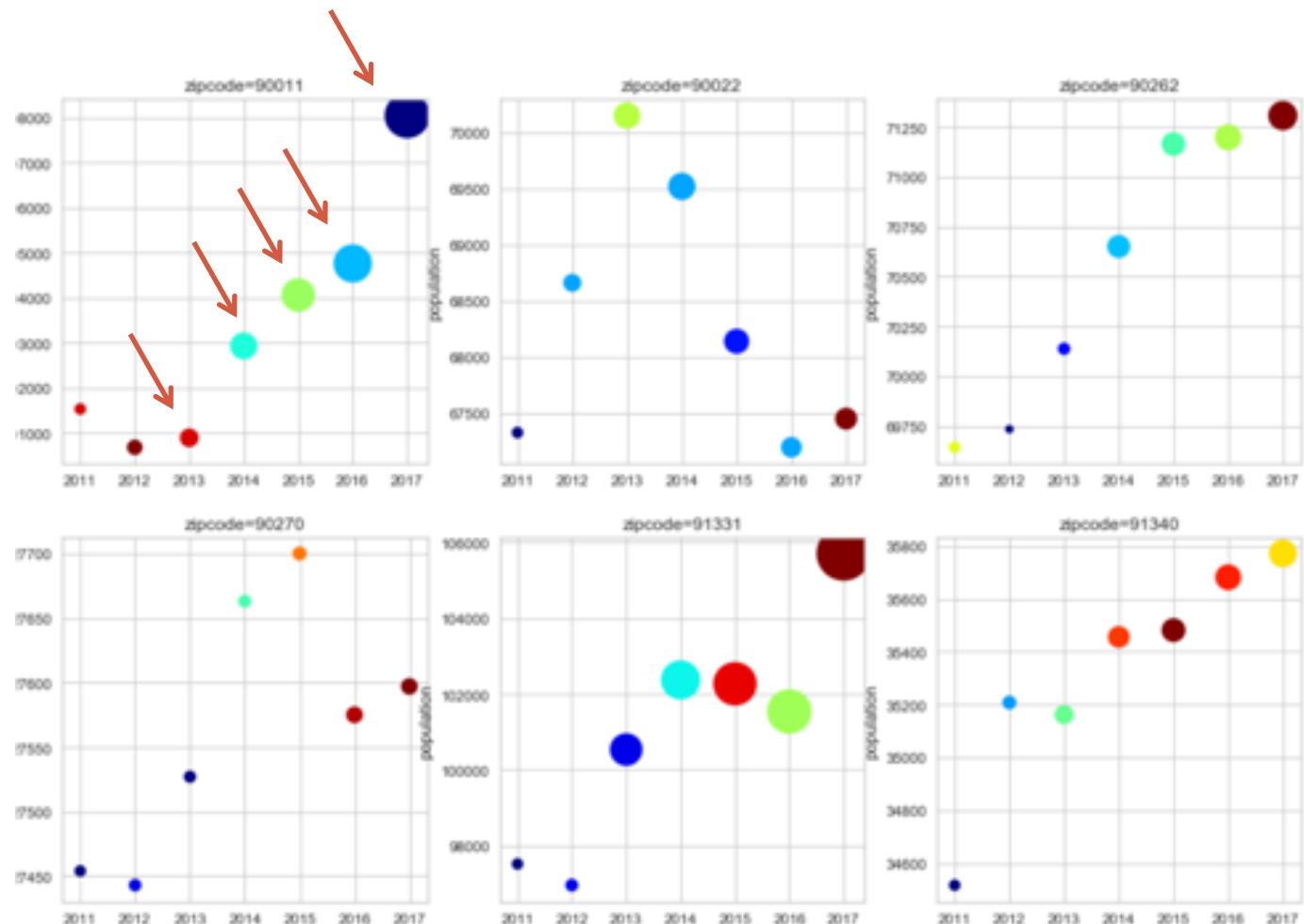
3D plot to find out the most influential factor for 'Income'

- I assigned the data with over 75% "Hispanic or Latino origin" to a new data frame, used matplotlib scatterplot to look at the correlation between "AGE-25 to 44 years", "Median age(years)" and "INDIVIDUALS' INCOME" in a 3D plot. The figures show the percentage of "AGE-25 to 44 years" is a better predictor for the higher income. As you may imagine, higher income customers may find our modern Mexican food more attractive.



- When I merged the top zip codes with "Hispanic or Latino origin" and data from above mentioned group 6 from KMeans analysis, I got 22 zip codes from the 2017 data. With the 22 zip codes, I extracted different data from year 2011 to 2017, including 'Total population', 'AGE-25 to 44 years' and "INDIVIDUALS' INCOME". Then I used scatter plot to visualize the change of the values over the years for top 6 zip codes, with population on the y axis, 'AGE-25 to 44 years' in colors, "INDIVIDUALS' INCOME" by the size of the dots. The idea is to identify a zip code area with growing high income, growing population and growing 'AGE-25 to 44 years' subpopulation. The result demonstrates 90011 is the best choice among the ones we analyzed.

Dot plot to visualize the best zip code for the Modern Mexican restaurant location



THANK YOU!

