IBM Coursera Data Science Capstone Project

Q: Where should I stay in Fort Lauderdale, Florida, USA?

Damien

2019

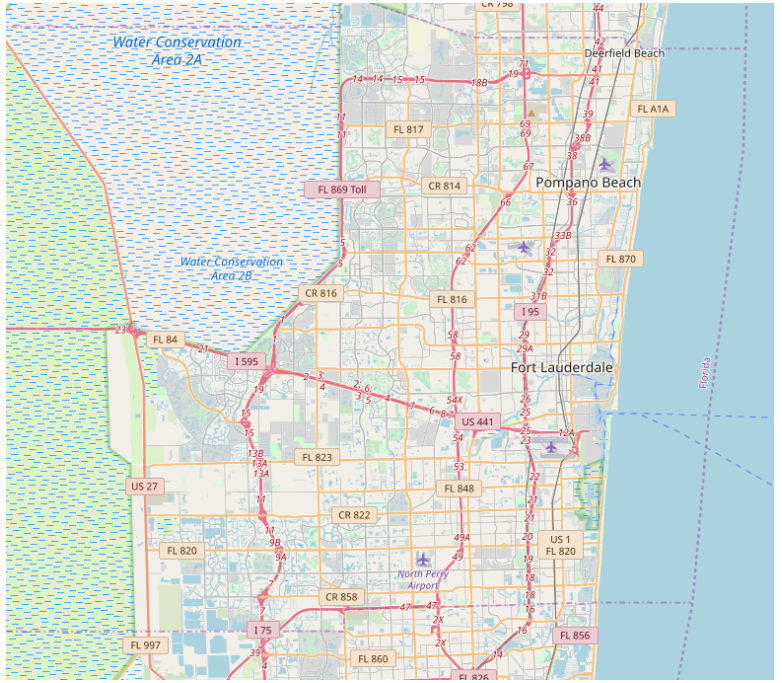
Table of Contents

1. Description of Problem / Key Question

Q: Where should I stay when I travel to Fort Lauderdale, Florida, USA when on vacation, so that I am close to the best shopping malls, restaurants and coffee shops ?

Background:

I like to travel to different cities to experience the local culture, food, atmosphere and of course the best shopping. Bringing all of this data together visually is difficult and I would like to be able to see it altogether in a way that is personalized so that I , my friends and family will also be able to use it.



2. Description of Data

Data sets have been sourced from multiple sources to bring together a complete view.

Zipcode data from https://www.zip-codes.com/city/fl-fort-lauderdale.asp for the city of Fort Lauderdale, Florida, USA will be used and combined with Lat-Long data from

https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/

Only zip codes with population greater than zero will be used to find populated areas for our data set.

The top 10 shopping malls data comes from
https://www.10best.com/destinations/florida/fort-lauderdale/shopping/shopping-centers-districts/

and the associated Lat-Long data from https://www.latlong.net/ for each shopping mall will be acquired and collated.

Foursquare API data is used extensively with **Folium** mapping data to plot and overlay of top 10 best shopping malls data on the foursquare clustered venue data for all the zip code locations in the final output. New analysis may be possible once initial findings are discovered. The data sets are all on different websites in different formats. A lot of data scraping and wrangling is required to create a clean data set.

A new clustered data set will be created and an overlay of shopping mall data will highlight best zip code(s) to stay in when I travel to Fort Lauderdale, Florida, USA.

3. Methodology :
→ Download, Explore and Clean the Dataset

Download and import all USA Zip data with Lat/Long

	Zip	Latitude	Longitude
0	71937	34.398483	-94.39398
1	72044	35.624351	-92.16056
2	56171	43.660847	-94.74357
3	49430	43.010337	-85.89754
4	52585	41.194129	-91.98027

Download and import all Fort Lauderdale Zip code data with population count

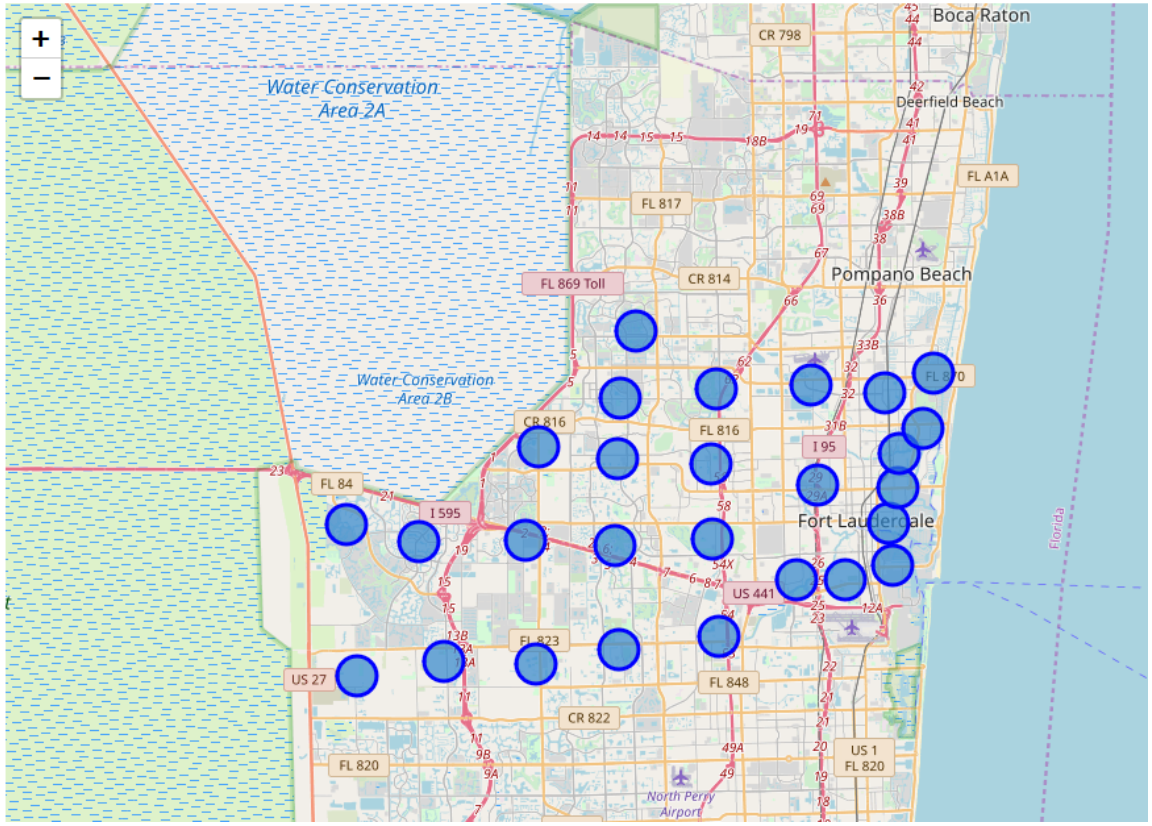
	Zip	Population
0	ZIP Code 33301	14,586
1	ZIP Code 33302	0
2	ZIP Code 33303	0
3	ZIP Code 33304	17,724
4	ZIP Code 33305	11,927

Remove 0 population areas, merge data with Lat-Long and remove any NaN values to create working data set.

	Zip	Population	Latitude	Longitude
0	33301	14,586	26.121114	-80.13187
1	33304	17,724	26.137693	-80.12646
2	33305	11,927	26.153728	-80.12606
3	33306	3,397	26.165212	-80.11379
4	33308	28,217	26.191111	-80.10846

4. Methodology : Explore Neighbourhoods

- Using Foursquare and Folium plot all ZIP Codes areas in Fort Lauderdale.
- The areas beside the sea are higher density
- Mostly, there is a even distribution of ZIP codes across the city



5. Methodology: Analyse Each Zip Code

- Utilize the Foursquare API to explore the neighbourhoods, venues e.g. restaurants, coffee shops etc… and segment them.
- 303 Venues were returned by Foursquare
- 123 unique categories

	Borough	Neighborhood	Latitude	Longitude
0	Fort Lauderdale	33301	26.121114	-80.13187
1	Fort Lauderdale	33304	26.137693	-80.12646
2	Fort Lauderdale	33305	26.153728	-80.12606
3	Fort Lauderdale	33306	26.165212	-80.11379
4	Fort Lauderdale	33308	26.191111	-80.10846
5	Fort Lauderdale	33309	26.185461	-80.17218
6	Fort Lauderdale	33311	26.138830	-80.16865
7	Fort Lauderdale	33312	26.094665	-80.17987
8	Fort Lauderdale	33313	26.148613	-80.22446
9	Fort Lauderdale	33314	26.067966	-80.22016
10	Fort Lauderdale	33315	26.094385	-80.15422
11	Fort Lauderdale	33316	26.101114	-80.12931
12	Fort Lauderdale	33317	26.113664	-80.22376
13	Fort Lauderdale	33319	26.183628	-80.22122
14	Fort Lauderdale	33321	26.211122	-80.26310
15	Fort Lauderdale	33322	26.150863	-80.27284
16	Fort Lauderdale	33323	26.156854	-80.31408
17	Fort Lauderdale	33324	26.110631	-80.27469
18	Fort Lauderdale	33325	26.112924	-80.32097
19	Fort Lauderdale	33326	26.112366	-80.37697
20	Fort Lauderdale	33327	26.120134	-80.41441
21	Fort Lauderdale	33328	26.062016	-80.27202
22	Fort Lauderdale	33330	26.054790	-80.31581
23	Fort Lauderdale	33331	26.056111	-80.36373
24	Fort Lauderdale	33332	26.049367	-80.40885
25	Fort Lauderdale	33334	26.182161	-80.13341
26	Fort Lauderdale	33351	26.179495	-80.27183

6. Methodology: Cluster Neighbourhoods(Zip Codes)

K- means clustering algorithm was used to find similarities between zip code areas (large circles) Small black circles(dots) ● are Top 10 Shopping areas.

Distance

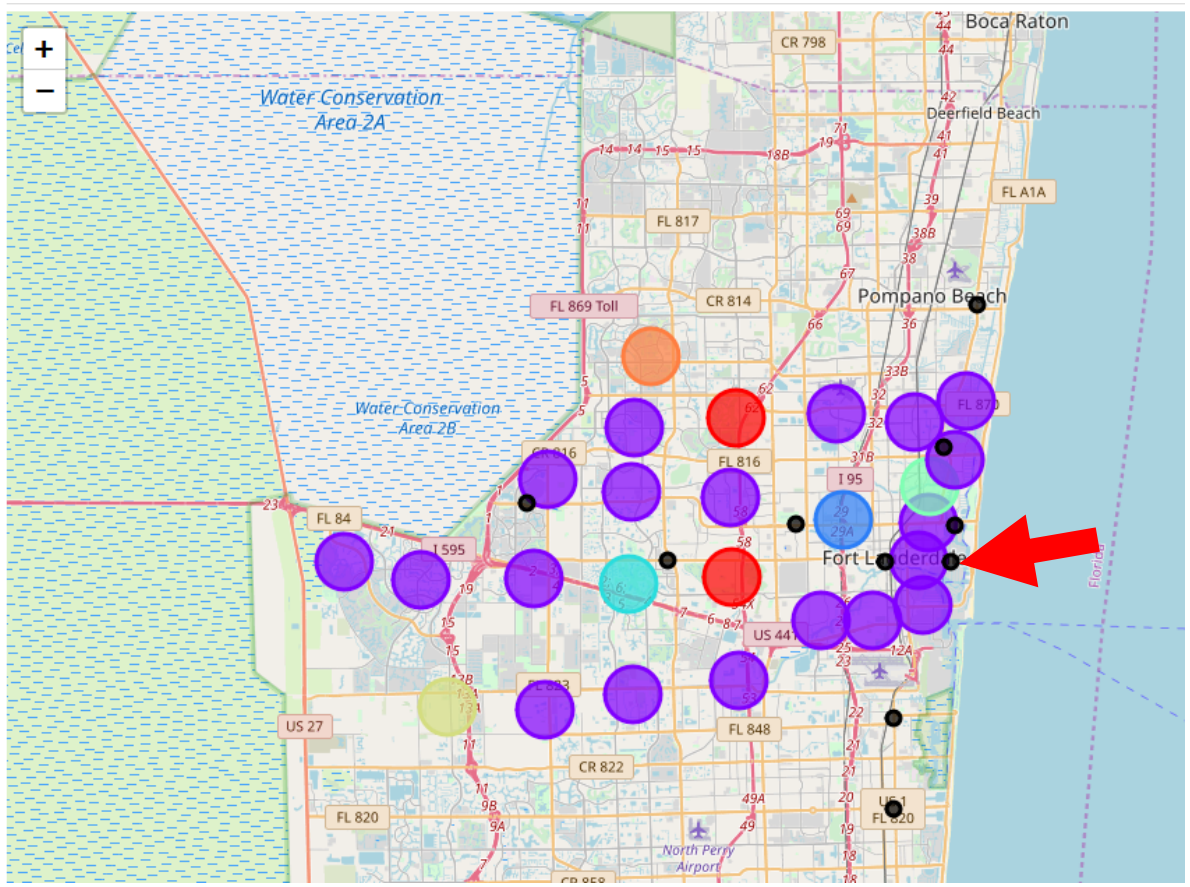
This uses the '**haversine**' **formula** to calculate the great-circle distance between two points – that is, the shortest distance over the earth's surface – giving an 'as-the-crow-flies' distance between the points (ignoring any hills they fly over, of course!).

Haversine formula:

$a = \sin^2(\Delta\phi/2) + \cos\phi1 \cdot \cos\phi2 \cdot \sin^2(\Delta\lambda/2)$

ϕ/λ for latitude/longitude in radians

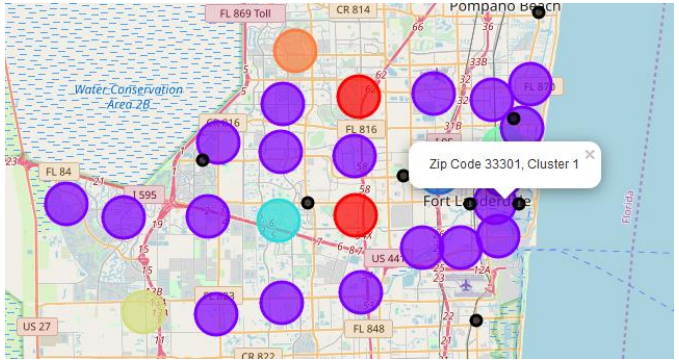
https://www.movable-type.co.uk/scripts/latlong.html

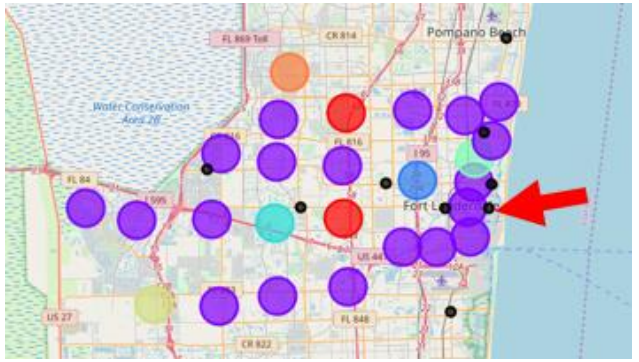


7. Results: Examine Clusters

Red Arrow points at the purple cluster closest to shopping malls (black dots)

The Zip code 33301 is a member of the cluster with the most restaurants and amenities and closest to the black dots.





Largest Cluster (1) shows the best Zip code and most common venues in it: 33301

Cluster 1 : Includes the recommended Zip Code 33301

```
FTL_merged.loc[FTL_merged['Cluster Labels'] == 1, FTL_merged.columns[[1] + list(range(5, FTL_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	33301	Italian Restaurant	Bar	Asian Restaurant	Mexican Restaurant	Pizza Place	American Restaurant
1	33304	Wine Shop	Fast Food Restaurant	Intersection	Donut Shop	Clothing Store	Rental Car Location
3	33306	Pizza Place	Restaurant	Italian Restaurant	Breakfast Spot	Coffee Shop	Big Box Store
4	33308	German Restaurant	Seafood Restaurant	Italian Restaurant	Yoga Studio	Pharmacy	Pub
5	33309	Thrift / Vintage Store	Pizza Place	Mexican Restaurant	Grocery Store	Mobile Phone Shop	Donut Shop
7	33312	Park	Boat or Ferry	Garden	Dim Sum Restaurant	Diner	Discount Store

Calculate “as the crow flies” distance between all Zip codes and shopping areas for comparison using haversine’ formula (Km)

Zip code with total shortest distance to all shopping areas is ZIP : 33301 and agrees with clustering on the map !

	Borough	Neighborhood	Latitude	Longitude	Swap Shop	Dania Antique Row	Riverwalk	Pompano Citi Centre	Downtown Hollywood
0	Fort Lauderdale	33301	26.121114	-80.13187	6.27	7.86	1.61	12.83	12.28
1	Fort Lauderdale	33304	26.137693	-80.12646	6.56	9.77	2.91	10.91	14.17
2	Fort Lauderdale	33305	26.153728	-80.12606	6.86	11.53	4.34	9.17	15.95
3	Fort Lauderdale	33306	26.165212	-80.11379	8.43	13.02	6.07	7.67	17.39
4	Fort Lauderdale	33308	26.191111	-80.10846	10.30	15.95	8.83	4.74	20.32

8. Discussion: Observations and recommendations

- The area closest to the coast had a higher density of zip codes and shopping areas giving the 'best' location as zip code 33301 with 3 shopping areas of the Top 10 within close proximity.

- An area further away from the coast at Zip code = 33323 is beside Sawgrass Mills. One of the top shopping areas and staying in this area may be a lower cost area to avail of the same amenities and shopping, however, this would need to be investigated.

- Adding the Top 10 rank to each label of shopping area would provide more info to users/audience.

- An overlay of hotels could also be added to make it easier to find suitable locations on a new map and deepen the analysis and personalisation for the user.

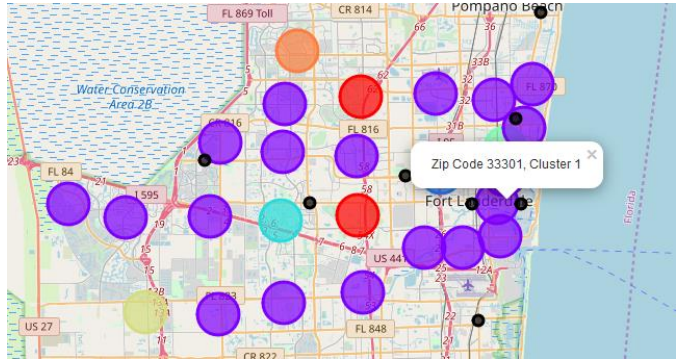
- Using distance calculation also gave the same result and is a quick and easy way to estimate a location/zip code area to consider as a place to stay to be located close to shopping and restuarants. Further analysis of other cities would validate this approach.

- Access to additional data sources such as https://www.zipdatamaps.com/33301 could be used to download additional useful information about this area. Further coding required to scrape this data and align with findings.

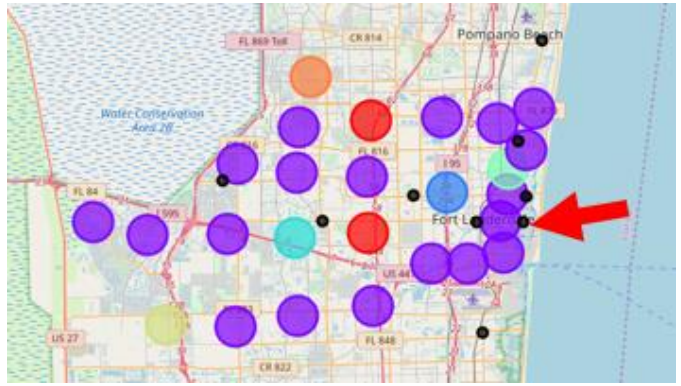
9. Conclusion

Foursquare data was used to answer the question of **"Where should I stay in Fort Lauderdale, Florida when I travel there so I am close to the best shopping with lots of restaurants and amenities ?".**

- Zip code data was used with Lat-Long data
- Venue data from Foursquare
- k-means clustering to show the areas most suitable.
- Top 10 Shopping area data overlaid the best shopping experience
- Visually, based on cluster analysis, the Zip code area 33301 was the winner



Further analysis : **Shortest total sum of distances** between zip code and lat-long centres of shopping areas predicted the same Zip code! 33301



The project has shown how foursquare data and other data sets can be combined to answer a meaningful question for a traveller visiting a new city on where to stay. Zip code 33301.