



# IBM Coursera Data Science Capstone

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

[Abstract](#)

Foursquare, Folium, K-means, Python, Dataframes, Analysis

Damien  
2019

## Table of Contents

1. Introduction: Description of Problem / Key Question .....	2
2. Description of Data .....	2
3. Methodology: Download, Explore and Clean the Dataset.....	3
4. Methodology: Explore Neighbourhoods.....	3
5. Methodology: Analyse Each Zip Code.....	4
6. Methodology: Cluster Neighbourhoods (Zip Codes) .....	4
7. Results: Examine Clusters .....	5
8. Discussion: Observations and recommendations.....	6
9. Conclusion¶ .....	6

## 1. Introduction: 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

ZIP code 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

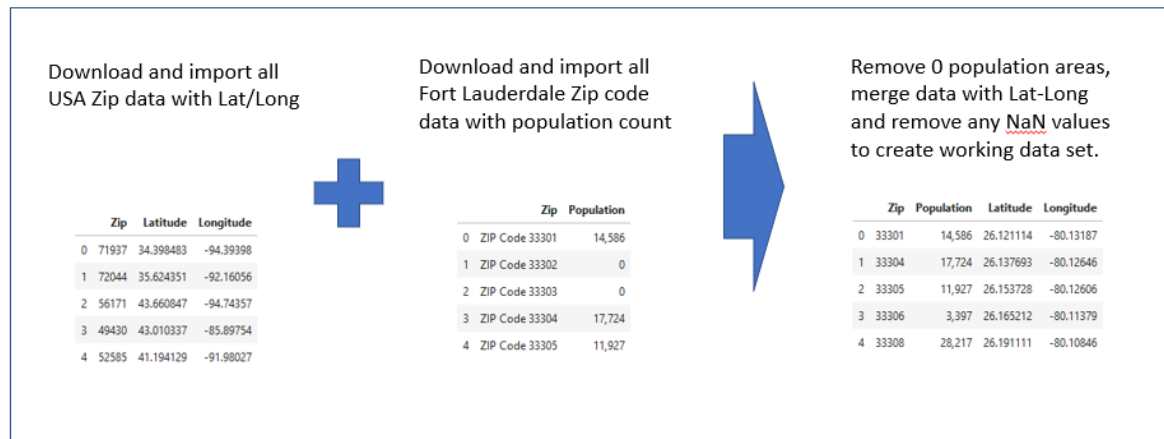
<https://www.10best.com/destinations/florida/fort-lauderdale/shopping/shopping-centers-districts/>

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.

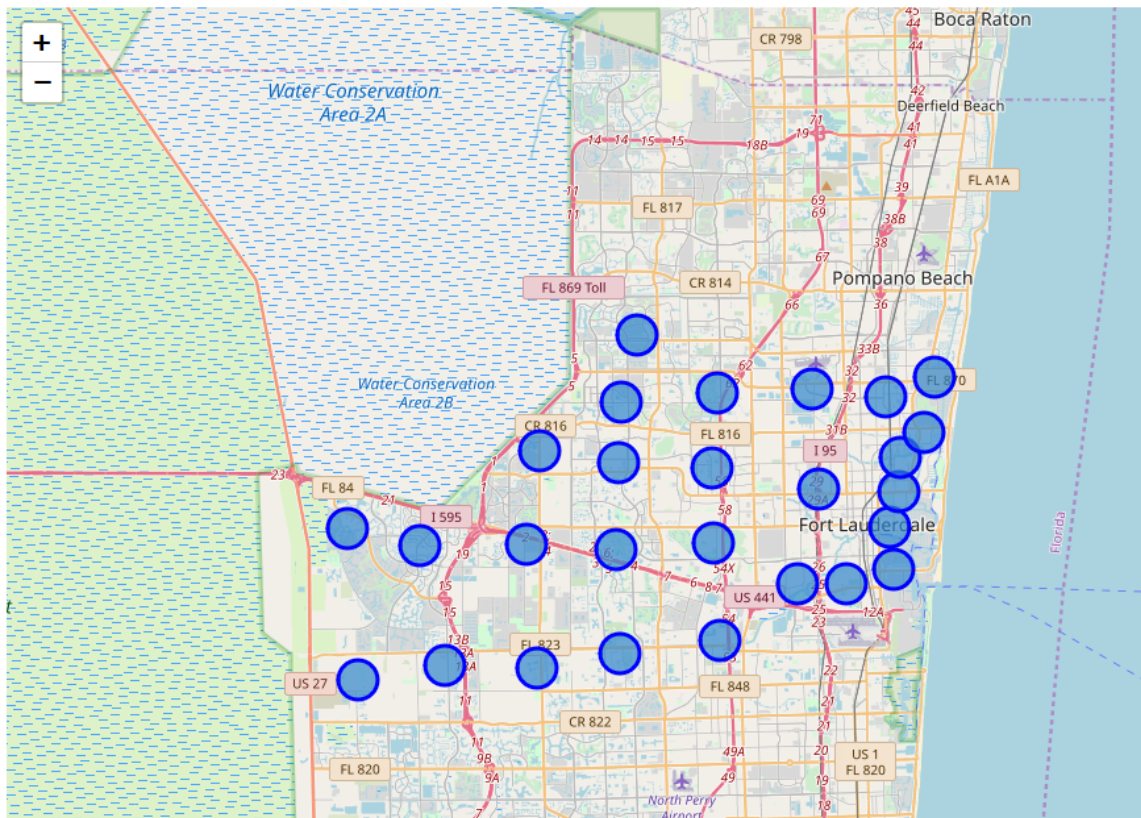
A new clustered data set will be created and an overlay of shopping mall data will highlight best zip code(s) to stay in when I travel to Fort Lauderdale, Florida, USA.

#### 4. Methodology: Explore Neighbourhoods



#### 4. Methodology: Explore Neighbourhoods

- Using Foursquare and Folium plot all ZIP Codes areas in Fort Lauderdale to view distribution and confirm data is valid.
- The areas beside the sea are higher density with 5 zip codes in a small area.
- Mostly, there is an even distribution of ZIP codes across the city and they bigger as they move away from the coastline.



## 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 API
- 123 unique categories were returned and is a large sample set to use for analysis.

## 6. Methodology: Cluster Neighbourhoods (Zip Codes)

2 methods were used and the results compared. A **K-means** clustering algorithm was used to find similarities between zip code areas (large circles). The Small black circles(dots) are Top 10 Shopping areas. To compare results a distance calculation was made using Lat-Long data.

### Distance

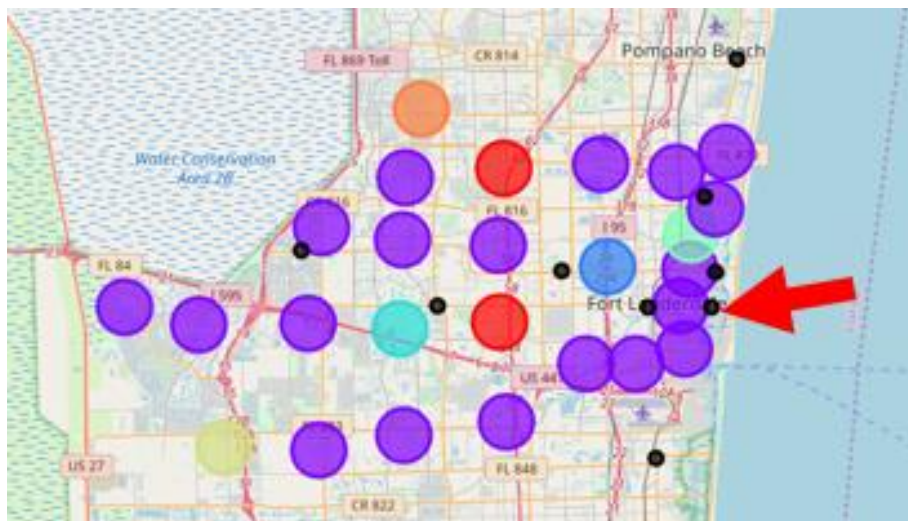
This uses the 'haversine' formula to calculate the great-circle distance between two points – that is, the shortest distance over the earth's surface – giving an 'as-the-crow-flies' distance between the points (ignoring any hills they fly over, of course!).

Haversine formula:

$$a = \sin^2(\Delta\phi/2) + \cos \phi_1 \cdot \cos \phi_2 \cdot \sin^2(\Delta\lambda/2)$$

$\phi/\lambda$  for latitude/longitude in radians

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

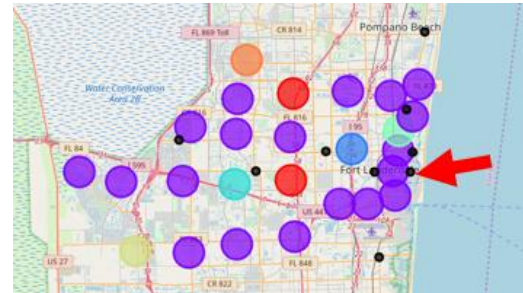
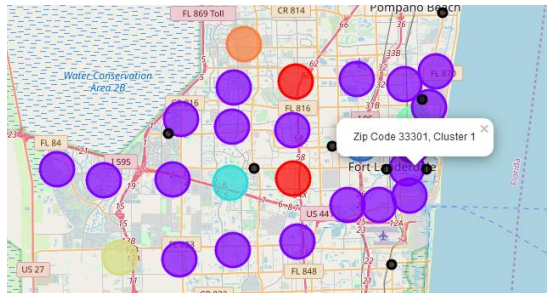


(red arrow points at the preferred zip code in the cluster)

## 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

	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

Calculated “as the crow flies” distance between all Zip codes and shopping areas for comparison using haversine’ formula (Km) (sample data shown above)

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

## 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 restaurants. 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 to gather venue data from Foursquare to assess the areas with most amenities and use k-means clustering to show the areas most suitable. Top 10 Shopping area data was taken from bestTop10.com and lat-long data from latlong.net to overlay the best shopping experience with the best restaurants and venues.

Visually, the Zip code area 33301 was the best to stay in close proximity to 3 of the top 10 shopping areas.

Further analysis was carried out to calculate the shortest total sum of distances between zip code and lat-long centres to be able to predict the best zip code to stay in to be located as close as possible to all shopping areas. The result returned agreed with the clustering exercise.

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.