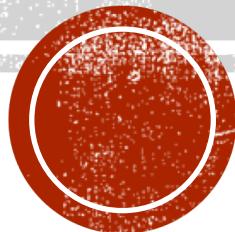


# **GENERATING MAPS WITH LOCATION HOTSPOTS AND LEVERAGING FOURSQUARE FOR TOURISM**

**Justin Zheng 21.8.2019**

**Applied Data Science Capstone**



# INTRODUCTION AND BUSINESS PROBLEM

- The purpose of this data project is to use the Foursquare location and venue data in order to generate maps that would be able to show hotspots in Rittenhouse Square Philadelphia, PA.
- Such maps are intended to be printed and handed to tourists for guidance
- To generate relevant maps, I will need to determine the most important categories of venues, for example a shopping mall or a convenience store.
- Distance and radius of location search for the Foursquare data base for venues will need to be individually determined through trial and error in order to generate an appropriately extensive list of venues in each category



# RELEVANCE

- The target audience for this particular data science project is tourists and venues associated with tourism like hotels
  - Such venues often display and allow visitors to take maps with them to help them navigate in the neighborhood
- These maps are meant to offer a simple and individualized navigational guide



# DATA

- A majority of location data was scraped directly from Foursquare
- Data included postal codes, names, longitude and latitude values and exact addresses of different venues
- Example of a cleaned data frame with Foursquare data:

1	The Shops at Liberty Place	Shopping Mall	1625 Chestnut St	39.951919	-75.167833	19103	PA
---	----------------------------	---------------	------------------	-----------	------------	-------	----

2	The Gallery at Market East	Shopping Mall	901 Market St	39.952689	-75.158149	19107	PA
---	----------------------------	---------------	---------------	-----------	------------	-------	----

- Longitude and latitude values that were acquired through Foursquare were used to form maps
- Postal codes acquired through Foursquare were further used to identify specific neighborhoods and to selectively clean data
- Other data bases were also used such as Apple Maps and Google to manually fill missing values



# DATA CLEANING

- I ran a function that extracts venue categories from the original data scraped from Foursquare then filtered each row. The end result was as follows:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	postalCode	state
0	On-Line Shopping Store	Shoe Store	1509 Walnut St	US	Philadelphia	United States		Nan	80 [1509 Walnut St, Philadelphia, PA 19102, United States]	[{"label": "display", "lat": 39.95205126106208, "lon": -75.163867}	39.952051	-75.163867	19102	PA
1	The Shops at Liberty Place	Shopping Mall	1625 Chestnut St	US	Philadelphia	United States	at 16th St	378	[1625 Chestnut St (at 16th St), Philadelphia, PA 19103, United States]	[{"label": "display", "lat": 39.95191851893288, "lon": -75.167833}	39.951919	-75.167833	19103	PA
2	The Gallery at Market East	Shopping Mall	901 Market St	US	Philadelphia	United States	at 9th St	458	[901 Market St (at 9th St), Philadelphia, PA 19107, United States]	[{"label": "display", "lat": 39.95268865552114, "lon": -75.158149}	39.952689	-75.158149	19107	PA
3	Chinatown	Neighborhood	Nan	US	Philadelphia	United States		Nan	[Philadelphia, PA 19107, United States]	[{"label": "display", "lat": 39.95548773443018, "lon": -75.156693}	39.955488	-75.156693	19107	PA
4	Walnut Street	Shopping Mall	Nan	US	Philadelphia	United States		Nan	[Philadelphia, PA 19103, United States]	[{"label": "display", "lat": 39.94973955125363, "lon": -75.167047}	39.949740	-75.167047	19103	PA

- Irrelevant Columns were removed and columns renamed with a loop function

- Filled out missing data



# METHODOLOGY

1. I imported requests for requests, pandas and numpy for data analysis. I also imported Nominatim for latitude and longitude transformation, Image and HTML from IPython for images, json\_normalize for changing a json file into a pandas data frame library and folium for plotting.
2. Establishing geolocator for Philadelphia
3. Parameters such as category and radius were added to a url with my credentials to retrieve data from Foursquare.
4. Sent the GET request and transformed it into a pandas dataframe



## 5. Data cleaning

```
CS_clean_columns = ['name', 'categories'] + [col for col in df_CS.columns if col.startswith('location.')] + ['id']
df_clean_CS = df_CS.loc[:,CS_clean_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list5 = row['categories']
    except:
        categories_list5 = row['venue.categories']

    if len(categories_list5) == 0:
        return None
    else:
        return categories_list5[0]['name']

# filter the category for each row
df_clean_CS['categories'] = df_clean_CS.apply(get_category_type, axis=1)
```

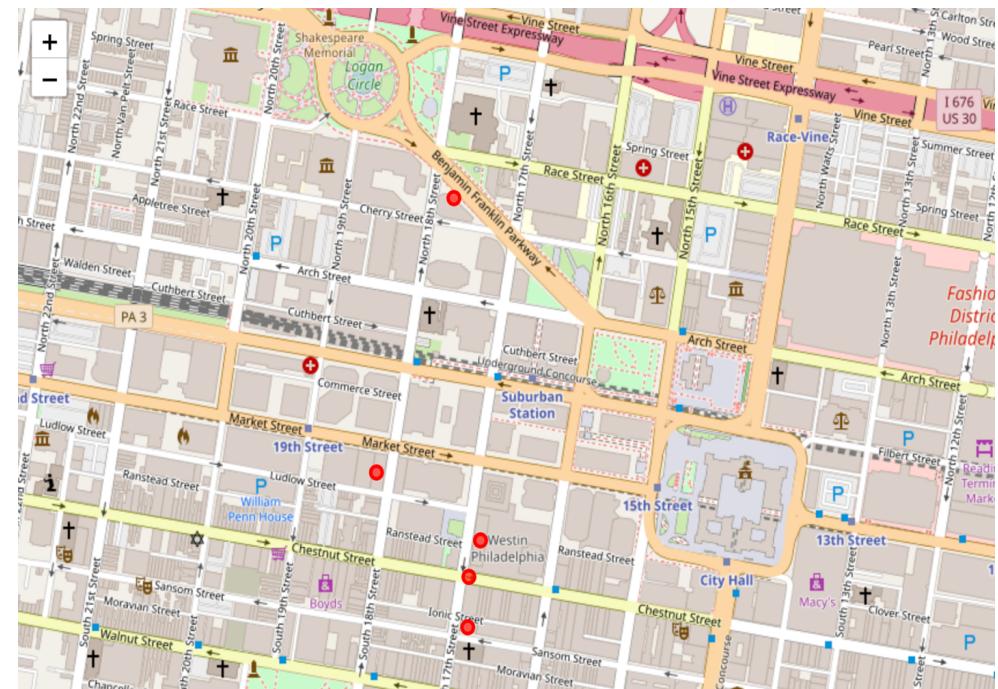
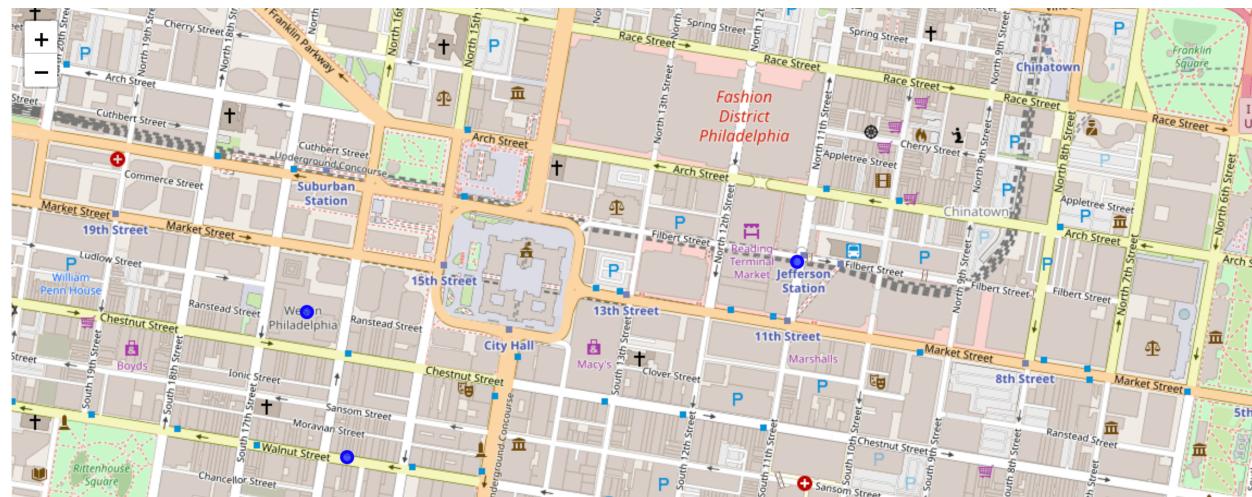
6. Re-running searches with different radius amounts ranging from 500, 1000, 1500 and 2000 to find optimal distance

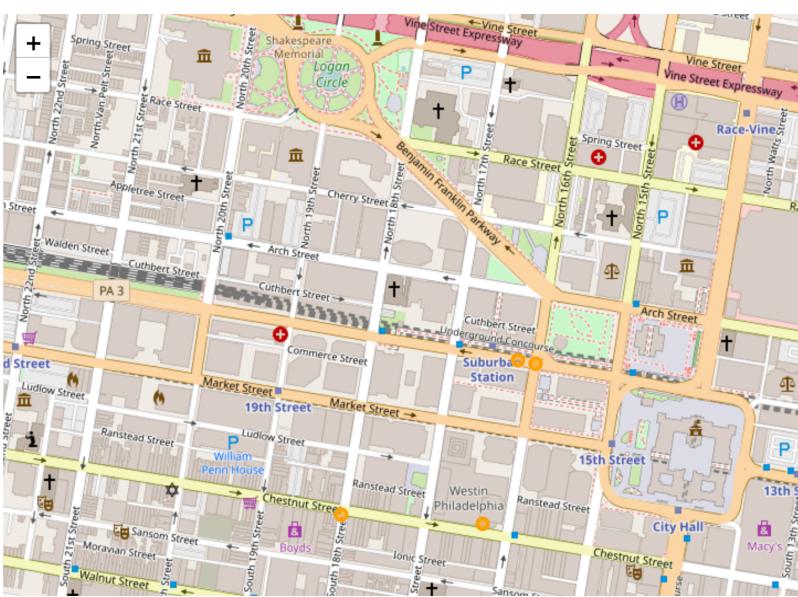
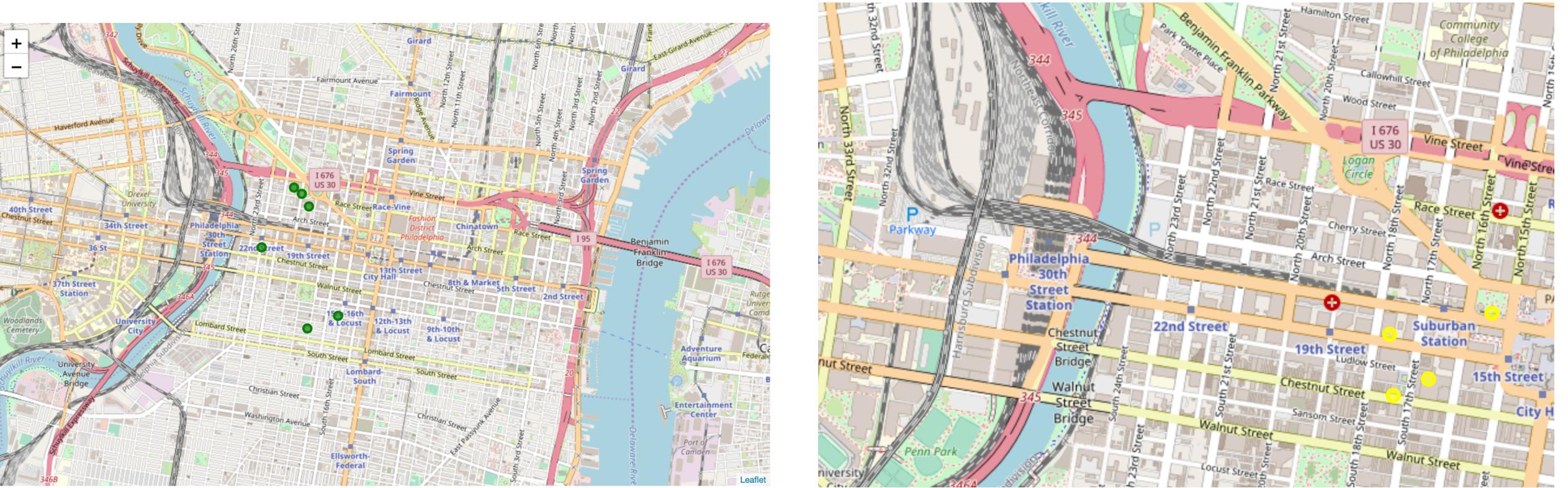
7. This was an iterative process and the final result looked like this

		name	categories	address	lat	lng	postalCode	state
2		Kimpton Hotel Palomar Philadelphia	Hotel	117 S 17th St	39.950809	-75.168642	19103	PA
3		Club Quarters Hotel in Philadelphia	Hotel	1628 Chestnut St	39.951466	-75.168632	19103	PA
21		The Westin Philadelphia	Hotel	99 S 17th St	39.951937	-75.168411	19103	PA
25		Sonesta Philadelphia Rittenhouse Square	Hotel	1800 Market St	39.952834	-75.170345	19103	PA
29		Embassy Suites by Hilton Philadelphia Center City	Hotel	1776 Benjamin Franklin Pkwy	39.956422	-75.168899	19103	PA



# RESULTS AND MAPS





- Venues were marked by a circle with each category designated a different color
- Final result was five clustered maps



# DISCUSSION AND CONCLUSION

- More databases should be used to increase accuracy
  - Other non-location based can also be leveraged such as Yelp or Facebook
- Having individual maps increases clarity as if all venues were to be clustered together it would be messy
- Proximity of areas could be measured with different streets, landmarks or buildings instead of postal codes to improve accuracy
- Foursquare data contained many mistakes hence trial and error was needed. Due to such errors, manual methods of data analysis was needed, hence machine learning is not the most appropriate method to approach this kind of problem
- In this project, I leveraged Foursquare location data in order to create five separate and individual cluster maps of different venues near Rittenhouse Square in Philadelphia

