script and all data can be seen in the ipynb file

1.Identifying the business opportunity (introduction):

Southern Manhattan is the business and financial center on the world, its where wall street, with all its banks and financial companies are located.

This area brings ten of thousands of people to work every day, with their needs of eating or venues for business lunches.

The objective of this project is to answer a business entrepreneur question about a business opportunity that arise for him to open a sushi/Japanese restaurant in that area.

Because wall street area is one of the most expensive and prime location in all of USA, and New York city in particular, this entrepreneur is looking for data based answer about the worthwhileness of this opportunity.

2.Data Overview:

The data will be acquired with a combination of available json file of the new York neighborhoods and boroughs. The data about the venues and businesses and the location of the above will be acquired using Foursquare API and google maps geolocator package accordingly.

3.Methodolgy:

Firstly, we acquired the map of Manhattan with all its neighborhoods as seen in fig.1
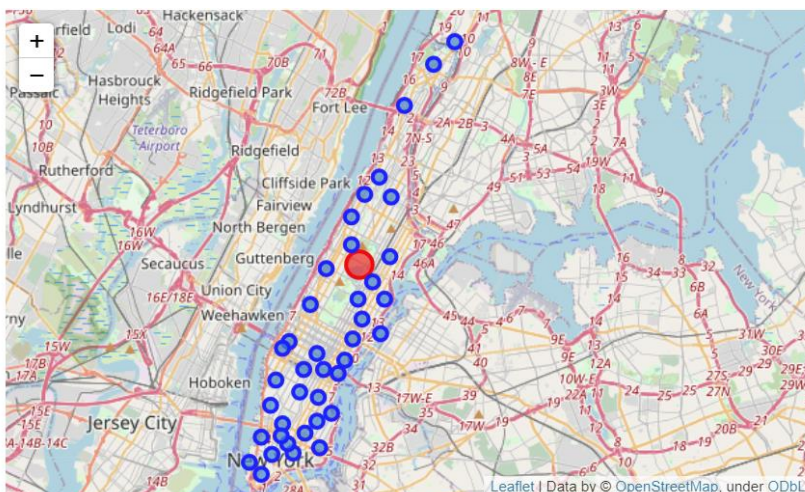

Figure 1: map of Manhattan neighborhoods (the red dots indicates a coordinates of Manhattan as the center of the beginning of our inquiry).

Then, by using foursquare API we will search for all the Sushi\Japanese restaurants in Manhattan island with the intention to check the spread, or density of this venues in Manhattan. Fig.2 shows the DataFrame we get after the acquiring and cleaning the raw data.

	latitude	longtitude	Borough	street	Restaurant catagorie	Restaurant name	id
0	40.709963	-74.010282	Manhattan	Cortlandt Street	Sushi Restaurant	Bento Sushi	5bca0ff80ff4f9002cd191b4
1	40.712742	-74.016065	Manhattan	South End Avenue	Sushi Restaurant	Blue Ribbon Sushi Bar	53d3f607498edfb90a15baf5
2	40.714952	-74.006678	Manhattan	Reade Street	Sushi Restaurant	Okami Fusion Sushi	4bb38ae6715eef3bcc7b86bb
3	40.706070	-74.007433	Manhattan	Water Street	Japanese Restaurant	Shinju Sushi	4a5e00d9f964a520e2bd1fe3
4	40.715057	-74.007607	Manhattan	Church Street	Sushi Restaurant	Kuki Sushi	4e4e4c36bd4101d0d7a71fab

figure 2: Venues data frame, with all the related data obtained via foursquare API.

After acquiring the venues data we will plot this on the New York map, when this time we are looking at wall street as the central point at a radius of 3KM.
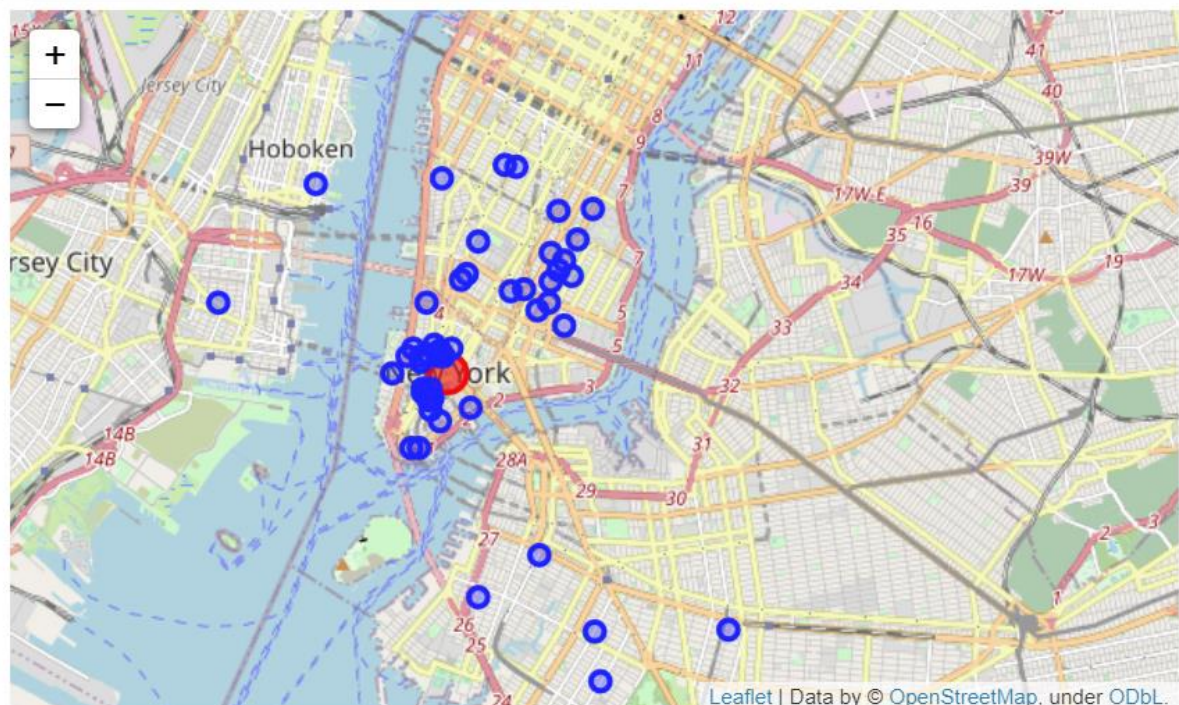

Figure 3: map of south Manhattan with wall street as central point(red circle) with blue circle as the venues.

Then, at the next step we thought about how to approach all the data we acquired, that to this point is pretty much standard procedure, we tried different failed approach of getting visitors ratings, tips

and number of visitor in each venue, but this information is largely missing and leaving us with not enough data points to do any kind of the data science analysis technique (mainly K Mean Clustering that is relevant to this problem). After much research and some consultation, a different approach was suggested – Weber's Theory of the location of industries (1909). Shortly weber's theory suggests that the location of industries is determined among other factors mainly in the lateral distribution of people and economic activities, so that in the end we get clusters of similar industries clustered together, this is why we can see cobbler streets, money changer bunched close together, and so forth. This theory encourage us to examine the number of venues as a function of distance from our desire location area.

4.1 Modelling:

In order to make a recommendation with accordance to our Theory of the location of industries approach we will use K Mean Clustering in order to cluster the venues in groups according to their geographical distance from wall street area. This can be seen in fig 4.

	latitude	longtitude	Borough	street	Restaurant catagorie	Restaurant name	distance	id
0	40.709963	-74.010282	Manhattan	Cortlandt Street	Sushi Restaurant	Bento Sushi	0.473130	5bca0ff80ff4f9002cd191b4
1	40.712742	-74.016065	Manhattan	South End Avenue	Sushi Restaurant	Blue Ribbon Sushi Bar	0.847029	53d3f607498edfb90a15baf5
2	40.714952	-74.006678	Manhattan	Reade Street	Sushi Restaurant	Okami Fusion Sushi	0.253518	4bb38ae6715eef3bcc7b86bb
3	40.706070	-74.007433	Manhattan	Water Street	Japanese Restaurant	Shinju Sushi	0.749880	4a5e00d9f964a520e2bd1fe3
4	40.715057	-74.007607	Manhattan	Church Street	Sushi Restaurant	Kuki Sushi	0.291669	4e4e4c36bd4101d0d7a71fab

figure 4: DataFrame of venues data with distance from center location.

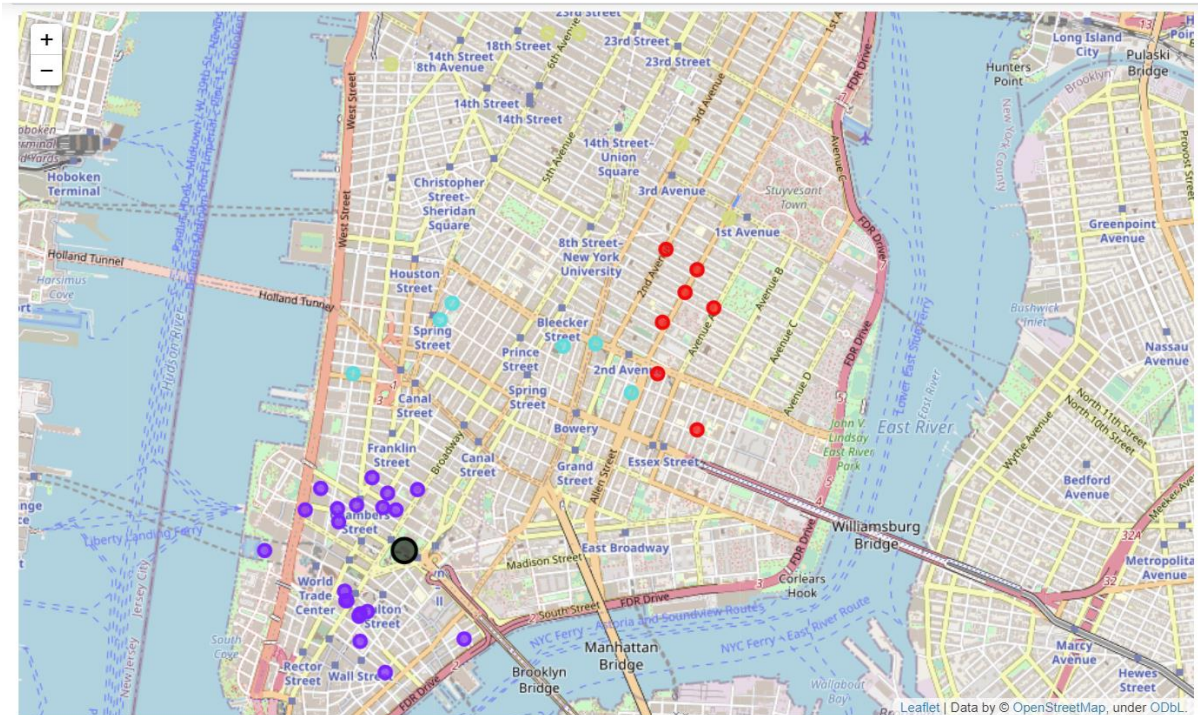A map of this clustering can be seen in fig 5.

Figure 5: map of venues clustering according to distance from the central location in black circle.

A summary of the data in fig 6. Shows that near the center location there are significantly higher number of Sushi/Japanese venues, so according to this data set, and according to the parameters and theory we chose to make our recommendation, we will recommend to the entrepreneur to open his restaurant with distance radius between 500-1000 Km from the central location.

	mean distance	Number of venues in cluster	cluster color
Cluster 0	2.201030	7	red
Cluster 1	0.499132	20	purple
Cluster 2	1.432522	8	pale blue
Cluster 3	3.056143	5	mustard yellow

Figure 6: Summary of Clustering of accumulated data.

Next our final step will be to try to support out recommendation with another set of data. As was mentioned earlier we know wall street to be the central location of finance in Manhattan, so we explore the same area around our chosen central point for this type of venues, the results that strongly support out recommendation can be seen clearly in fig 7.

Figure 7: finance venues near our location in blue empty circles.

Conclusion:

In conclusion, with accordance to our approach of the business query that was discussed, and with the approach in which we chose our data and analysis of this data, we can strongly recommend to our said entrepreneur to open is Sushi\Japanese restaurant in the area around wall street up to radius of approximately 1Km from this location.