

# IBM APPLIED DATA SCIENCE CAPSTONE FINAL PROJECT

Recommending Advertisement Board Placement Locations for a  
Local Business

## THE FIRST STEP WAS TO IDENTIFY THE PRECISE BUSINESS NEED

- What is the exact purpose of this project?
  - The business is a Dessert Shop
  - They want to place (3) advertisement boards, to attract the attention of dining patrons within a 1.5 kilometer radius around their location.
  - They want to maximize the number of dining patrons who will be able to see their advertisement boards
  - The Ads will be targeted specifically for generating foot traffic, by appealing to those who are dining (stop by for dessert afterwards), and undecided patrons who are browsing the eatery districts

## CONCEPT AND LOGIC OF THE SOLUTION

- In order to be visible to the maximum number of dining patrons, we need to place the advertisement boards in proximity to the maximum number of restaurants
- The Ad locations will be within 1.5 kilometers, because dining patrons will only travel a short distance on impulse.
- So, we need to determine (3) locations within the prescribed radius, which have visibility to the maximum number of restaurant venues

# DATA

- What kind of data is required to generate a valid solution?
- We need geographic coordinates of all restaurant venues within the prescribed radius
- We include the restaurant venue names as a unique ID, because some restaurants may be in suites, within a single geo location. So, the dataset looks like the table shown here: →

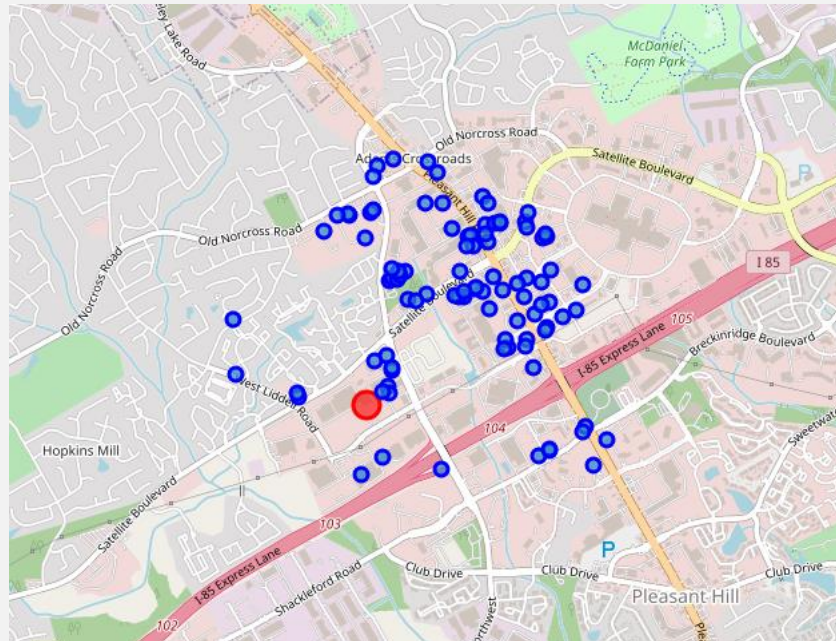
VenueName	Latitude	Longitude
Haru Ichiban	33.956893	-84.136399
Choong Man Chicken	33.953473	-84.142153
.	.	.
.	.	.

## GATHERING AND CLEANING THE DATA

- Our data, including the (3) stated fields, comes from Foursquare
- We retrieve Name and Geo Coordinates of all restaurant type venues within the prescribed area, using a URI to the Foursquare API
- Of course, the Get Request also returns a lot of unneeded information as well
- We extract only the (3) required fields, and put them into a new dataframe

# VISUALIZING THE RAW DATA

- Here is a snapshot of all restaurant venues in the area, as represented on a Folium Map. There are 96 restaurants (blue dots) and our client's dessert shop (Red dot)



## FINDING THE (3) OPTIMUM LOCATIONS

- To find the (3) areas with the most restaurant venues, we use K-Means Clustering
- Logically, we begin with a K-value of (4), because that would be three clusters in addition to the one where our clients Dessert Shop is located
- We can increment the K-value upwards from there, and evaluate results with each pass, until an optimum result is reached.

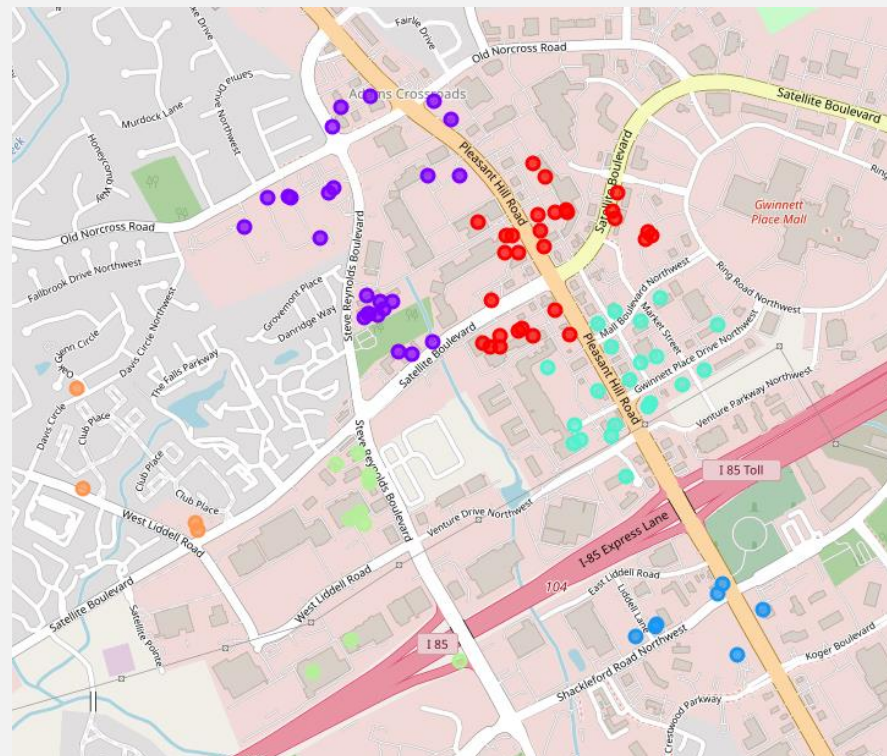
## SO, HOW WELL DID IT GO?

- Values below 6 showed only two relatively dense clusters, and other more dispersed clusters
- Values above 6 began to show a tight grouping of more than three clusters, with a couple of outlier clusters
- Thus,  $K=6$  proved to be the optimal parameter for clustering this group of restaurants by geographic proximity



# LET'S LOOK AT THE RESULT

- Here are the optimal clusters, at  $K=6$



## HOW DOMINANT ARE THE 3 LARGEST CLUSTERS NUMERICALLY?

- The three dominant clusters (#0, #1, #3) represent 75 out of 96 restaurant venues in the local area
- The client's Dessert Shop is in cluster #4, so they have direct visibility to the (10) restaurants in that cluster as well
- So, total coverage is 85 out of 96 area restaurants

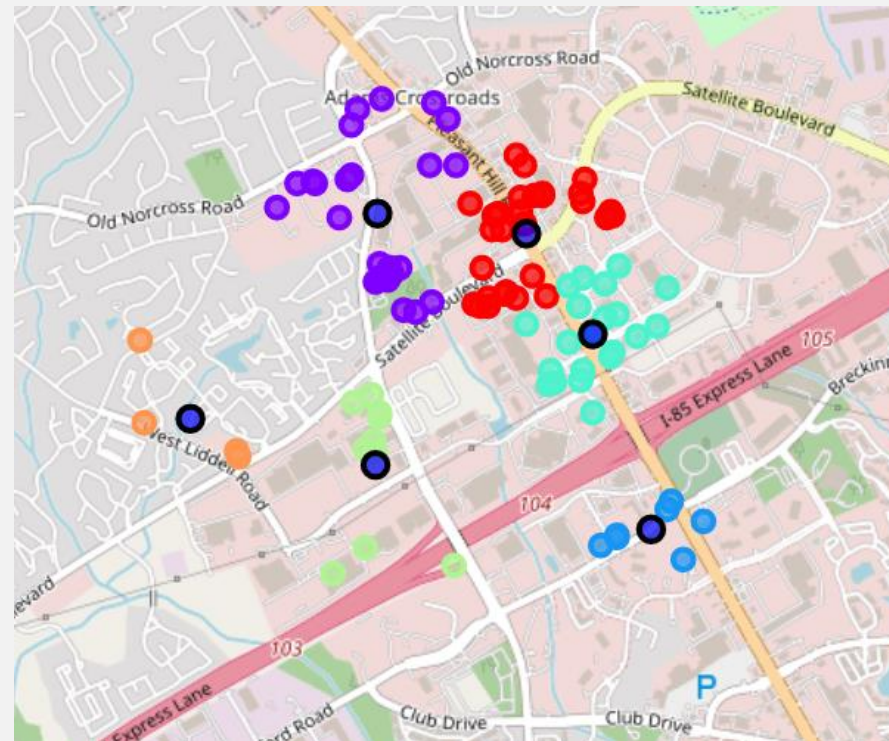
	VenueName	Latitude	Longitude
Clusters			
0	29	29	29
1	25	25	25
2	7	7	7
3	21	21	21
4	10	10	10
5	4	4	4

# DETERMINING THE RECOMMENDATION

- We still needed to determine exact reference points locating the Advertisement Boards
- The recommended locations are the Centroids of the three dominant clusters
- We determine the centroids from the mean of geo coordinates in each cluster

	Latitude	Longitude
Clusters		
0	33.959447	-84.134741
1	33.960172	-84.141112
2	33.948976	-84.129431
3	33.955902	-84.131912
4	33.951253	-84.141207
5	33.952928	-84.149090

## HOW DOES IT LOOK WITH THE CENTROIDS ADDED?



## THE RECOMMENDED LOCATIONS

- So, we have three reference locations for Ad Board placements:
  - Centroid of Cluster 0: [33.959447, -84.134741] “2170 Pleasant Hill Rd”
  - Centroid of Cluster 1: [33.960172, -84.141112] “3093 Steve Reynolds Blvd”
  - Centroid of Cluster 3: [33.955902, -84.131912] “3550 Gwinnett Place Dr NW”

## WILL OUR CLIENT BE SATISFIED WITH THE RESULTS?

- By placing Ad Boards in the three recommended locations, the client will have Ad Exposure, or potential visibility, to 88.5% of all restaurant venues in the prescribed area.
- That is a very strong result, and the client should be pleased

## HOW USEFUL IS THIS APPROACH?

- This approach should work well for similar problems of other businesses
- One key consideration is that, if the geographic area and distribution of target businesses had been more complex, the clustering step might have been a challenge
- For more complex cases, the K-value optimization should probably be coded as a loop through a range, with outcomes measured and recorded at each iteration

## CONCLUSION

- This was a very satisfying project, in that the concept and approach worked well, and a satisfying result was produced
- It was also gratifying to learn a useful approach to solving similar problems in the future