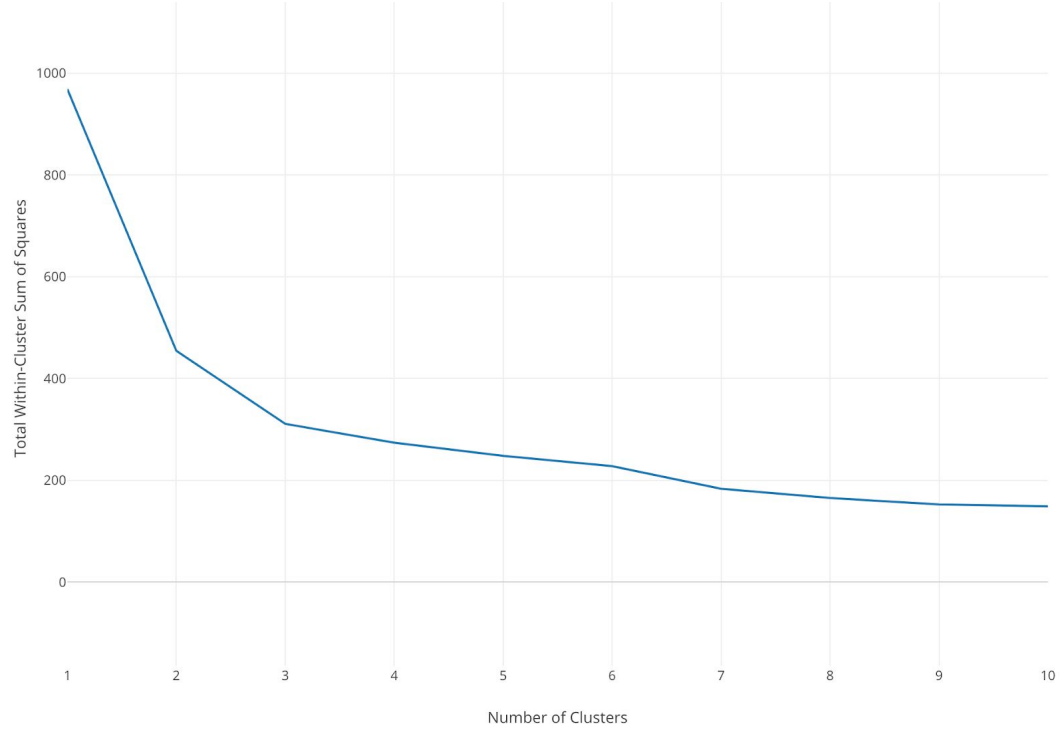
CISC 5950 Big Data Programming Project 1

Weifeng Yan, Alma Marko, Jiarui Zhang

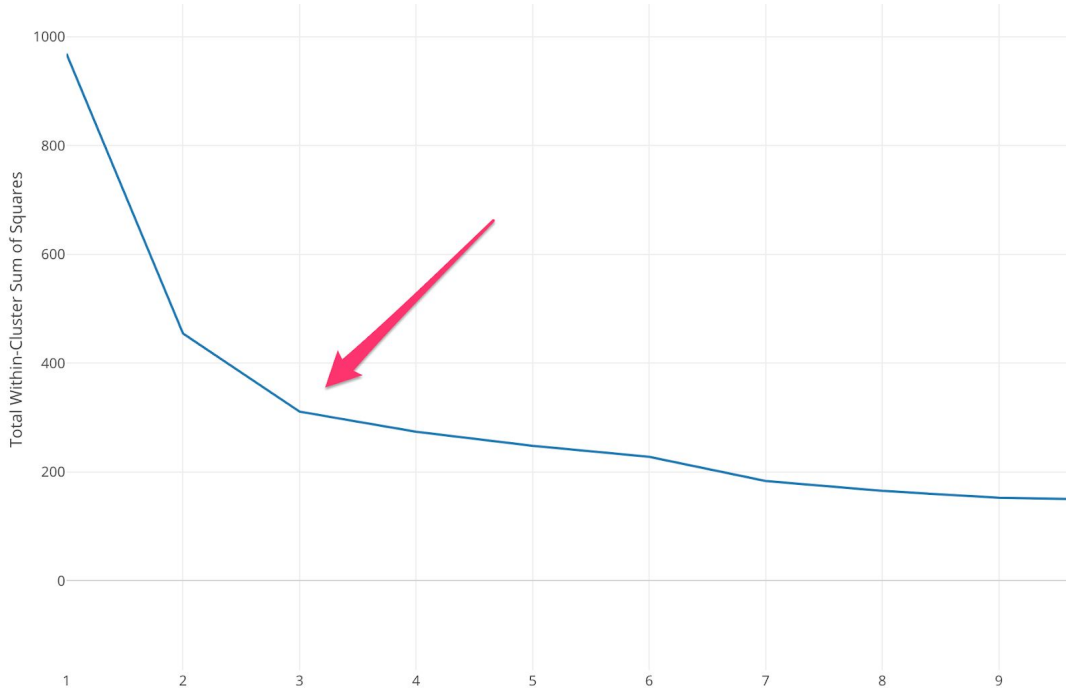
Part 3: Bonus Question

1. How to decide how many clusters to create

We iteratively build the K-Means Clustering models as we increase the number of clusters starting from 1 to, let's say, 10. Then we can calculate the distance between all the members (in our example they are the counties) that belong to each cluster and the center of each cluster every time we build a new model.
This will give us a series of models with different distance values. Finally, we draw a line that connects all the distances of the clusters like the below.



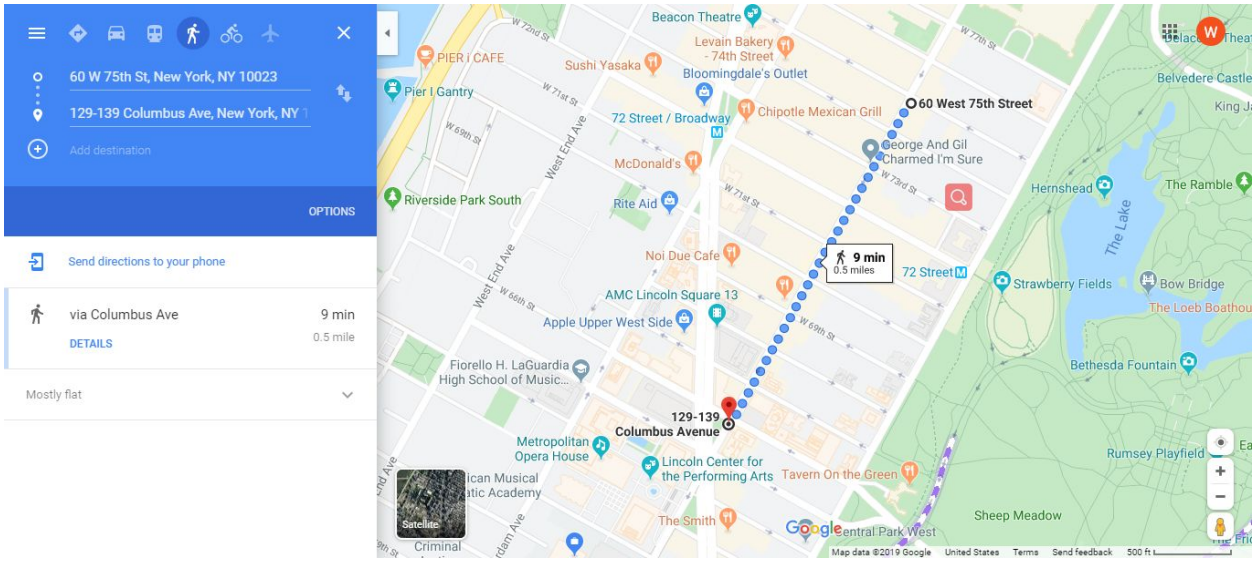
The Y-Axis represents the distance and the X-Axis represents the number of clusters.
And a rapid drop in the curve at point Number of clusters= x means that the performance of having x number of clusters is much better than having x-1 number of clusters. The distance keeps decreasing as you increase the number of clusters, but usually, there is a point where the decline settles.



2. How to create K-means

(a) Select the feature of VEHICLE COLOR and Violation Location to train the model.
kmeans = KMeans(n_clusters=elbow_curve, random_state=0).fit(df_kmeans)
After that, we can predict Black vehicle parking illegally at 34510, 10030, 34050.
Use: kmeans.predict([‘Black’, 34510])
Get which cluster the [‘Black’, 34510] belongs to and this cluster’s tickets probability.
We do the same thing for [‘Black’, 10030], [‘Black’, 34050].

(b) For the second bonus question, we want to find the best place to park at 10AM that is within 0.5 miles of Lincoln Center. Considering that 0.5 miles is equivalent to 10 blocks, we know that we can park within a 10 block radius of Lincoln Center, which is from W 53rd St to W 73rd St. Thus, to solve this problem, we select the feature of STREET NAME, to train the model and then we select the zone with the least violations. For this problem, we are using the rows of data where VIOLATION TIME=10AM and STREET NAME from W 53 St to W 73 St. Then we perform k-means analysis on our data within this range and then create clusters. The areas that are not covered by our clusters will be the safe places to park.


60 W 75th St, New York, NY 10023
129-139 Columbus Ave, New York, NY
Add destination
OPTIONS
Send directions to your phone
via Columbus Ave
9 min
0.5 mile
DETAILS
Mostly flat
Beacon Theatre
Levain Bakery - 74th Street
PIER I CAFE
Sushi Yasaka
Bloomingdale's Outlet
Pier I Gantry
72 Street / Broadway
Chipotle Mexican Grill
60 West 75th Street
George And Gil Charmed I'm Sure
McDonald's
Riverside Park South
Rite Aid
Noi Due Cafe
9 min
0.5 miles
72 Street
Strawberry Fields
Hernshead
The Ramble
The Lake
Bow Bridge
AMC Lincoln Square 13
Apple Upper West Side
Fiorello H. LaGuardia High School of Music...
Bethesda Fountain
129-139 Columbus Avenue
Metropolitan Opera House
Lincoln Center for the Performing Arts
Tavern On the Green
Rumsey Playfield
Satellite
The Smith
Sheep Meadow
Central Park West
Map data ©2019 Google
United States
Terms
Send feedback
500 ft