Capstone Project: IBM Data Science

Solomin Oleg

May 2020

Moving from Toronto to New York

- The contractor is deciding to move from Toronto to New York and to start own business.
- He sold his barber shop in financial district of Toronto and is going to make a fortune in Manhattan.
- He expects that in New York it would be tough environment because of rivals pressure and high-demanding public.
- Anyway the contractor has a positive experience and full of bright ideas. The only one thing is beyond his scope is Big Data.

Logo vector created by freepik

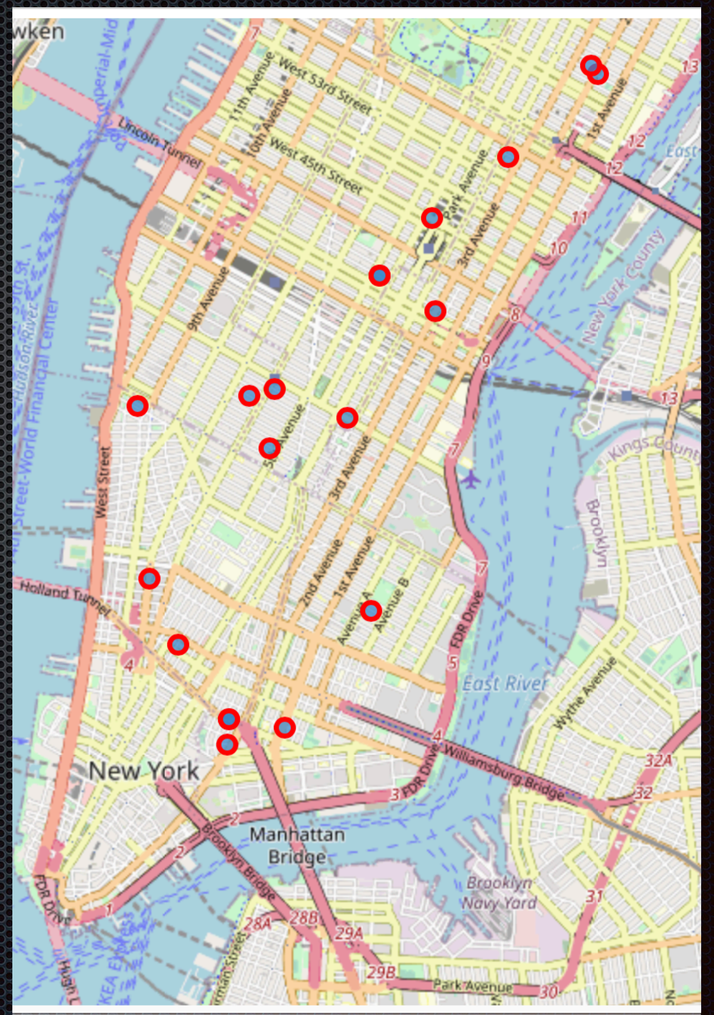
Tough challenge

Even without a thorough investigation it is obvious that we will face up a challenge:

- location in some distance from known venues with an idea that `**less rivals-more clients**`
- location in close distance to known venues in a hope that `**more venues-more clients**`

Barbershops in Manhattan

Using **Folium** and **Foursqare API** we can locate targeted venues on a map



Segmentation

Now we can make a segmentation say in three clusters using simple but efficient **k-means** method.

The KMeans class has many parameters that can be used, but we will use these three:

- **init**: Initialization method of the centroids. Value will be: '**k-means++**'. k-means++ selects initial cluster centers for k-means clustering in a smart way to speed up convergence.
- **n_clusters**: The number of clusters to form as well as the number of centroids to generate. Value will be: **3** (say we have 3 centers)
- **n_init**: Number of times the k-means algorithm will be run with different centroid seeds. The final results will be the best output of n_init consecutive runs in terms of inertia. Value will be: **12**

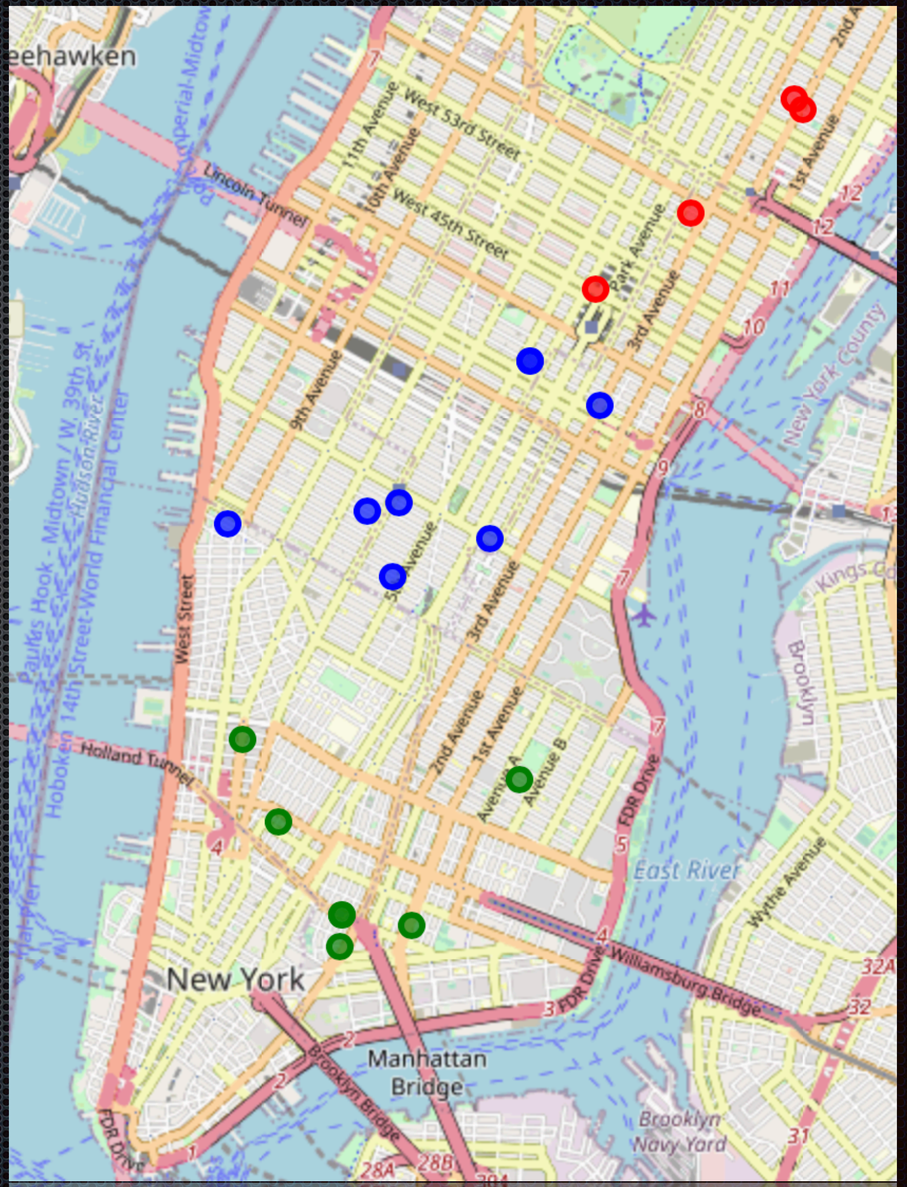
Initialize KMeans with these parameters, where the output parameter is called **k_means**.

Segmentation

Fine! We can locate three areas with targeted venues. But for better understanding it's appropriate to add more venues "for men". For example we add following keywords: barber, gym, steak, sport, health, cycle.

```
In [149]: k_means = KMeans(init="k-means++", n_clusters=3, n_init=12)

In [150]: X=barberM.values[:,4:6]

In [151]: k_means.fit(X)
          labels = k_means.labels_

In [152]: barberM["Labels"] = labels
```

```
In [154]: # create map of Manhattan Barbershops with Clusters using latitude and longitude values
          map_manhattan_venue_cluster = folium.Map(location=[latitudeM+0.025, longitudeM+0.03], zoom_start=13)

          # add markers to map
          for lat, lng, label,cluster in zip(barberM['Venue Latitude'], barberM['Venue Longitude'], barberM['Ve
              label = folium.Popup(label, parse_html=True)
              folium.CircleMarker(
                  [lat, lng],
                  radius=5,
                  popup=label,
                  color=color(cluster),
                  fill=True,
                  fill_color=color(cluster),
                  fill_opacity=0.7,
                  parse_html=False).add_to(map_manhattan_venue_cluster)

          display(map_manhattan_venue_cluster)
```



Segmentation

Finally we get more accurate segmentation in three different areas. And we have a valuable information on an environment. Not only rivals but males venues as well.

So it’s starting point to more precise locate place. We later could add areas profile:

- average check
- stars and so on