Selecting optimum locations for mobile towers/ WiFi hotspots using k-means

Vineet Puri

Dec 28, 2020

1. Introduction

1.1 Background

Network planning is an important job in any telecommunication company. We need to provide good coverage at all the required locations at minimum cost. As the number of towers increase, the cost of installation and maintenance also rises. So, we need to install towers in such a way that our target of coverage is achieved with minimum number of towers. In this project, I have tried to achieve this by using data science tools.

1.2 Business Problem

A huge sports event is going to be organised in the city of Toronto. As a telecommunication company we wish demonstrate our 5G technology during this event. It has been decided by the management that priority will be to cover all the public places or venues in the area and this needs to be done by installing minimum possible number of towers/antennae. So, our main target is to provide coverage at the venues even if there is no coverage at places outside the venues.

Also, the range of our 5G towers is 400 to 500 meters. So, any venue which is more than 400 meters away from the tower will be considered out of range. As a data scientist we need to calculate the best locations to install these towers so that all important places are covered with minimum number of towers.

1.3 Interest

This kind of implementation of data science can be of interest to telecommunication companies who wish to plan their future networks. Although, I have taken the example of mobile towers in this project but the strategy can be used to plan WiFi hotspots and other radio networks too.

2. Data description

2.1 Data sources

I have used Foursquare API to get the list of venues in the target area. We need to provide coverage at all these venues in this list. We need latitude, longitude, venue name from Foursquare around our target location. We selected a target location on map and got its latitude and longitude from google maps. We fetched all the venues in an area of approximately 7X7 Kms centred at our target location.

I am using a free service of Foursquare which gives me a limited number of venues around a location at a time. So, to overcome that I created a square grid (5X5) of 25 locations centred at the target location and then fetched the list of venues for all these locations within 1 Km radius from Foursquare.

Venue	Venue Latitude	Venue Longitude	Venue Category
Bobbette & Belle	43.731339	-79.403769	Bakery
For The Win Cafe	43.728636	-79.403255	Bubble Tea Shop
T-buds	43.731247	-79.403640	Tea Room
Mastermind Toys	43.732046	-79.404141	Toy / Game Store
Menchies Frozen Yogurt	43.728336	-79.403173	Ice Cream Shop
The Belly Buster Submarines	43.733743	-79.404390	Sandwich Place
Shinobu by Maki Sushi	43.732562	-79.404147	Japanese Restaurant
The Burger Cellar	43.732362	-79.403894	Burger Joint
STACK	43.729311	-79.403241	BBQ Joint
The Rolling Pin	43.733315	-79.404318	Bakery

2.2 Data cleaning
We will obviously get duplicate venues by this method which we need to correct this by deleting all duplicate venues.

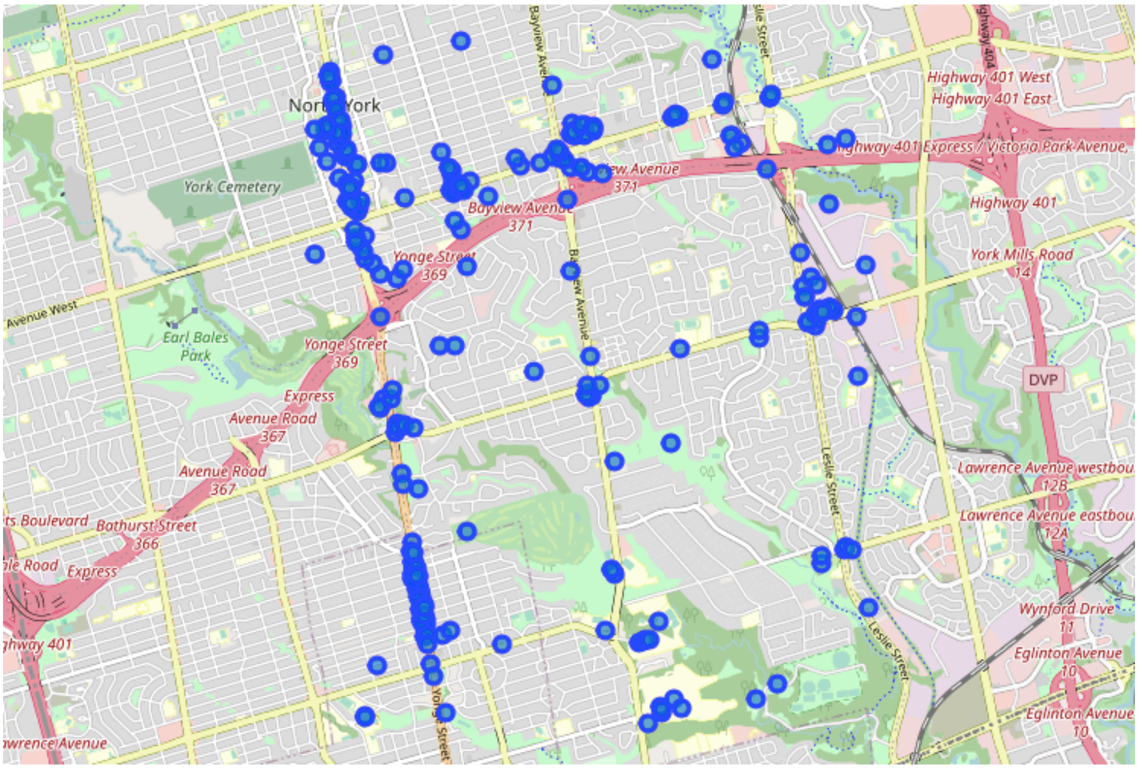
3. Methodology

We created a DataFrame of the data fetched from Foursquare. The list obtained from Foursquare had 771 venues. After removing the duplicate venues we were left with 281 venues.

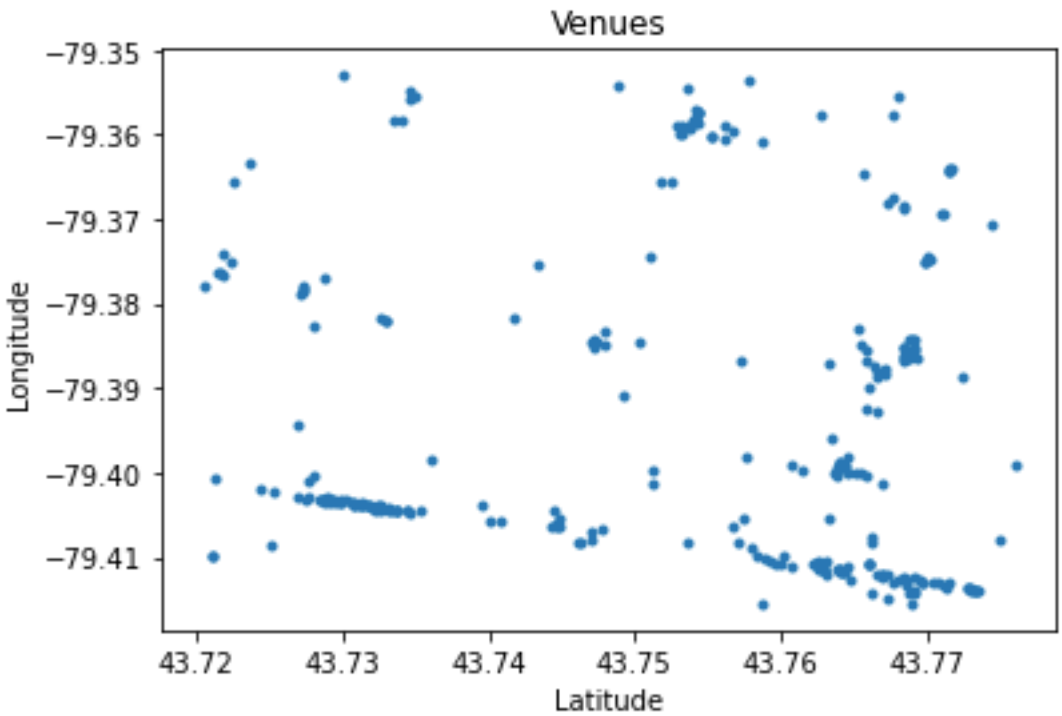
Our venues DataFrame looks like as follows-

	City Points	City Points Latitude	City Points Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Toronto 11	43.72864909867068	-79.40463791281924	Bobbette & Belle	43.731339	-79.403769	Bakery
1	Toronto 11	43.72864909867068	-79.40463791281924	For The Win Cafe	43.728636	-79.403255	Bubble Tea Shop
2	Toronto 11	43.72864909867068	-79.40463791281924	T-buds	43.731247	-79.403640	Tea Room
3	Toronto 11	43.72864909867068	-79.40463791281924	Mastermind Toys	43.732046	-79.404141	Toy / Game Store
4	Toronto 11	43.72864909867068	-79.40463791281924	Menchies Frozen Yogurt	43.728336	-79.403173	Ice Cream Shop
5	Toronto 11	43.72864909867068	-79.40463791281924	The Belly Buster Submarines	43.733743	-79.404390	Sandwich Place
6	Toronto 11	43.72864909867068	-79.40463791281924	Shinobu by Maki Sushi	43.732562	-79.404147	Japanese Restaurant
7	Toronto 11	43.72864909867068	-79.40463791281924	The Burger Cellar	43.732362	-79.403894	Burger Joint
8	Toronto 11	43.72864909867068	-79.40463791281924	STACK	43.729311	-79.403241	BBQ Joint
9	Toronto 11	43.72864909867068	-79.40463791281924	The Rolling Pin	43.733315	-79.404318	Bakery

We visualised these venues on a map of Toronto. I have used Folium library to create maps.



I also created a scatter plot to see the spread of venues in the city as it gives us a better idea about possible clustering.



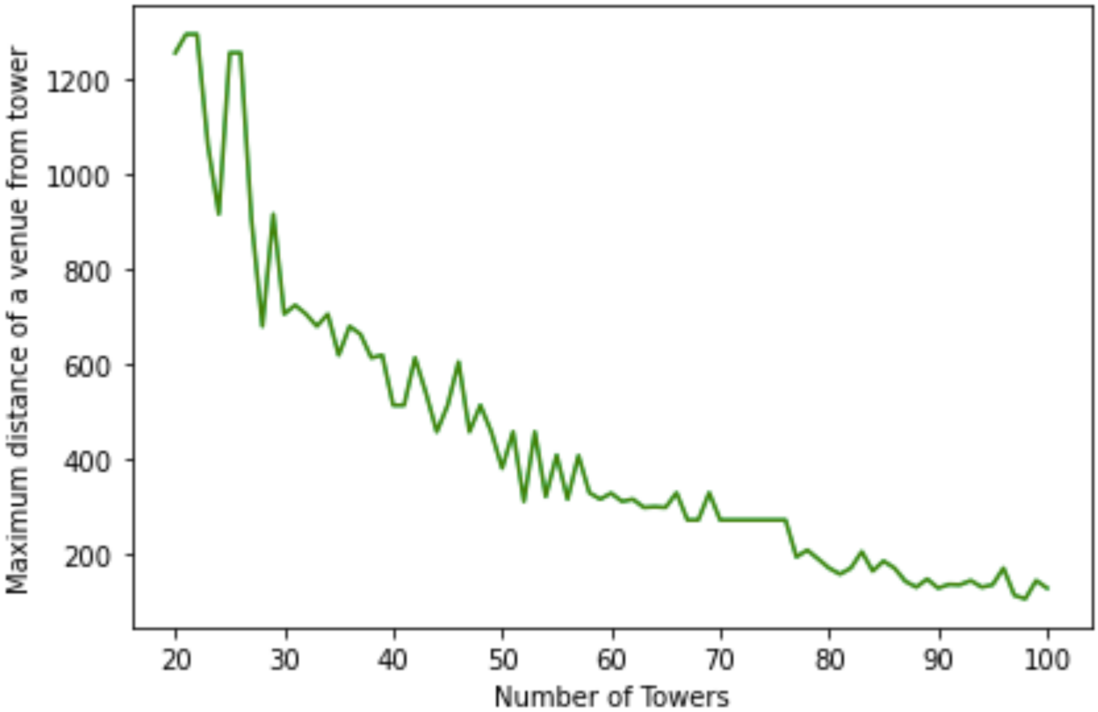
Then I applied k-means algorithm to form clusters. It is clustering algorithm for unsupervised learning.Each cluster will be served by a single tower located at the cluster centre. But we need to form clusters in such a way that no venue is out of range from its centre.

Initially I used the value of k as 20 and then created a scatter plot of clusters formed. Also, the range of our 5G towers is 400 to 500 meters. So, any venue which is more than 400 m away from the centre will be considered out of range and will be plotted as a X. Venues within range of towers will be marked as a dot (.) and towers themselves will be marked as triangles. Also, venues served by same tower have been coloured same. A part of the plot thus obtained is as follows-

As you can see there are many venues which are far more than 400 meters away from their allotted cluster centres and are marked as X in plot. So, these will be out of range from tower which will be installed at the centre. We need to adjust the number of clusters in such a way that all venues are within range.

I had to go through the different values of number of clusters (K) and applied k-means algorithm for each value. Then I selected the minimum value of number of clusters that will cover all the venues. Since, each cluster will be served by one tower, this way we will be able to minimise the number of towers to be installed.

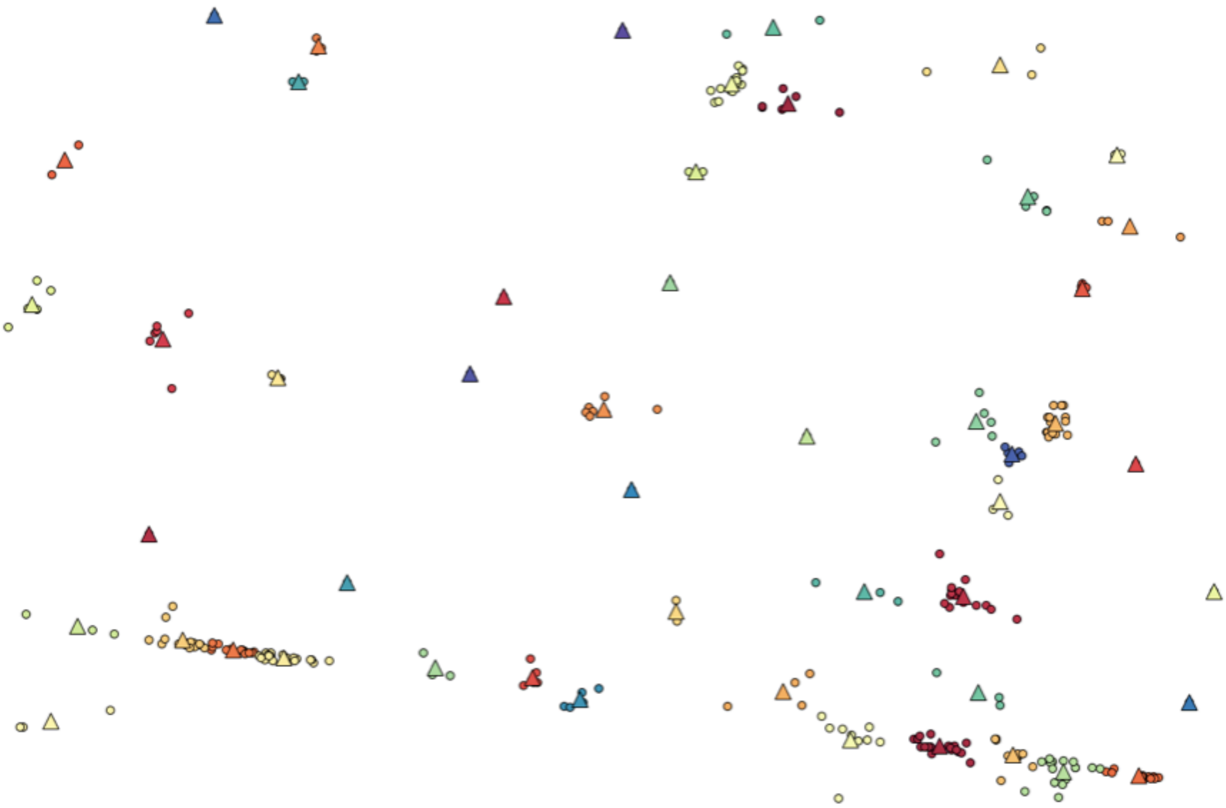
I performed k-means for all the values of K from 20 to 100. I also checked the maximum value from the distances of the venues from their centres and stored these maximum value for all the different values of K i.e. 20 to 100. I created a list of all such maximum distances and then we plotted K or number of towers against these maximum distances.



Elbow method is not relevant in this case as our focus is to have all venues within 400 m range of their towers. So, optimum value of k will be the point when maximum distance falls below 400. The value was found to be 50 in this case. At this value the maximum distance of a venue was found to be 382.05 meters. A sample of the print of maximum distance for different K values is as follows:

```
Max distance of a venue for 40 number of towers is 514.78 meters
Max distance of a venue for 41 number of towers is 514.78 meters
Max distance of a venue for 42 number of towers is 614.75 meters
Max distance of a venue for 43 number of towers is 541.16 meters
Max distance of a venue for 44 number of towers is 458.83 meters
Max distance of a venue for 45 number of towers is 514.78 meters
Max distance of a venue for 46 number of towers is 606.24 meters
Max distance of a venue for 47 number of towers is 458.83 meters
Max distance of a venue for 48 number of towers is 514.78 meters
Max distance of a venue for 49 number of towers is 458.83 meters
Max distance of a venue for 50 number of towers is 382.05 meters
Max distance of a venue for 51 number of towers is 458.83 meters
Max distance of a venue for 52 number of towers is 312.15 meters
Max distance of a venue for 53 number of towers is 458.83 meters
Max distance of a venue for 54 number of towers is 322.27 meters
Max distance of a venue for 55 number of towers is 409.93 meters
Max distance of a venue for 56 number of towers is 316.84 meters
Max distance of a venue for 57 number of towers is 408.63 meters
Max distance of a venue for 58 number of towers is 330.48 meters
```

I again applied the k-means algorithm taking number of towers as 50, then plotted it and ensured that no venue is out of range from its tower. I had to execute k-means multiple times to achieve this as it gives slightly different cluster formations each time.



As you can see above, now all venues are within range and we got the correct cluster formation. So, we were able to cover 281 venues with 50 towers.

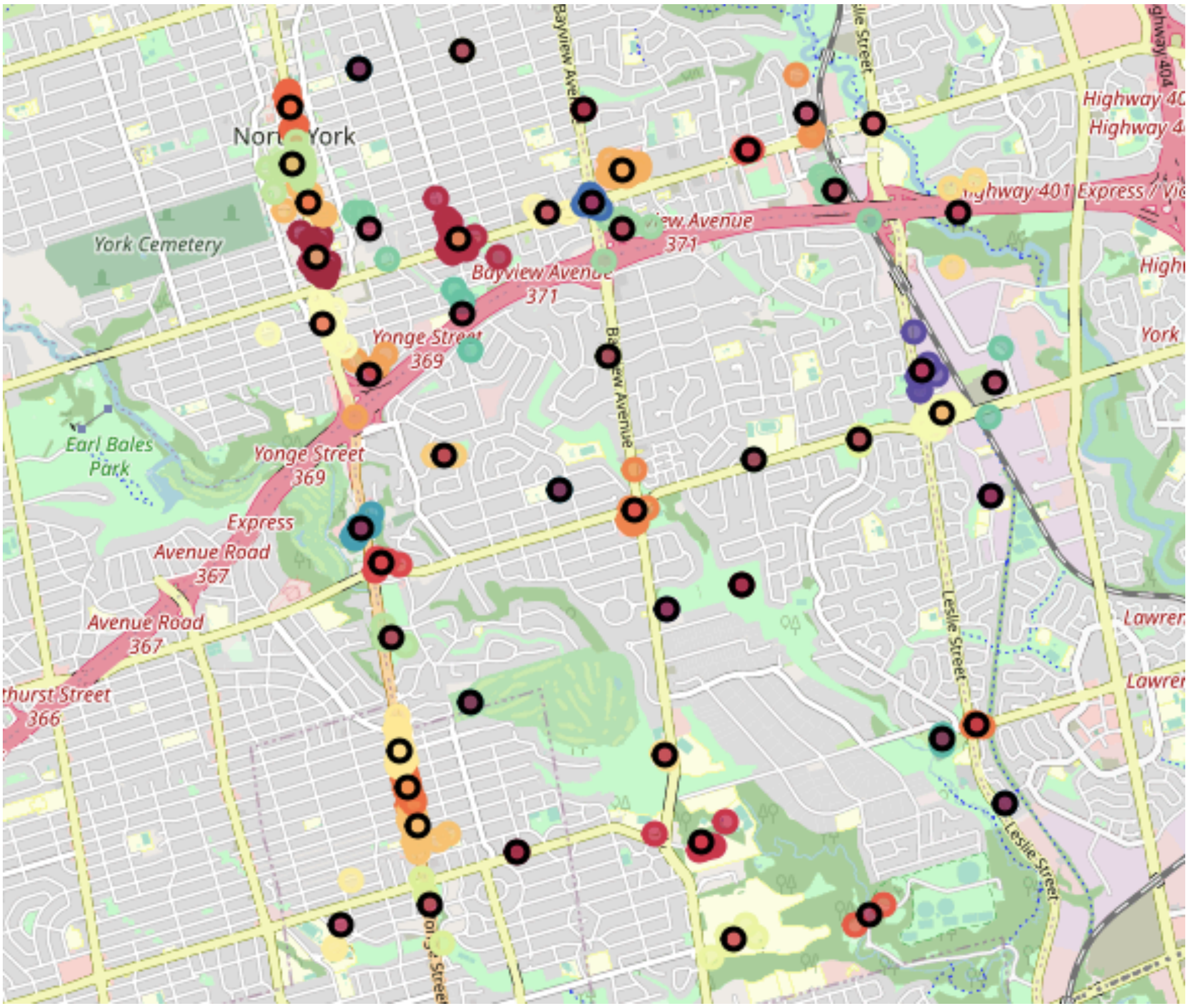
A tower which will serve more venues is expected to carry greater traffic. We will now categorise towers on the basis of the number of venues they will serve. This will help us in allotting resources accordingly. Also, if we face some issue in a tower regarding resource requirement and we solve it by increasing resources, then we can apply the same solution to other towers of the same category. This will help us in rectifying issues before they occur.

Then I added a tower column containing tower id and also a tower category column to venues DataFrame. For tower category I counted the number of venues served by the tower. Higher the number of venues served, higher will be the category of tower.

	City Points	City Points Latitude	City Points Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Tower	Tower category
0	Toronto 11	43.72864909867068	-79.40463791281924	Bobbette & Belle	43.731339	-79.403769	Bakery	11	15
1	Toronto 11	43.72864909867068	-79.40463791281924	For The Win Cafe	43.728636	-79.403255	Bubble Tea Shop	18	17
2	Toronto 11	43.72864909867068	-79.40463791281924	T-buds	43.731247	-79.403640	Tea Room	11	15
3	Toronto 11	43.72864909867068	-79.40463791281924	Mastermind Toys	43.732046	-79.404141	Toy / Game Store	22	20
4	Toronto 11	43.72864909867068	-79.40463791281924	Menchies Frozen Yogurt	43.728336	-79.403173	Ice Cream Shop	18	17

4. Results

Finally I was able to mark our venues and towers on the map of Toronto. Venues being served by same tower are coloured same. Also, towers of same category are coloured same. Towers can be identified by black edged circles.



In summary, I was able to find optimum location for towers in such a way that all venues were covered with minimum possible number of towers. I was also able to categorise towers on the basis of expected load. And finally I was able to visualise the results on a map.

4. Discussion

Network planning is a very complex topic. Many different approaches could have been used. I used k-means algorithm and checked models for different values of K from 20 to 100 and selected minimum value of K which served our purpose of coverage. In this project our target was to provide coverage at only the venues. For WiFi hotspots this strategy can be very useful. But for actual mobile networks we mostly need uniform coverage in entire city or area and may need some other approach.

For categorising mobile towers, I used the parameter of number of venues being served. In future studies, I would like incorporate additional data about venues like footfall or expected visitors too. This will help in better categorisation of towers on the basis of expected load.

In the end I visualised venues and towers on a map. In future studies, work can be done on finding the way to connect these towers to backend network through shortest possible routes.

5. Conclusions

Data science can be effectively used in telecommunication industry too. In this project we used k-means algorithm for network planning. And then we further classified towers too. So, data analysis can be useful in planning, maintenance , upgradation and expansion of telecommunication networks.

6. References

- **Foursquare API**
- **Google maps**
- **K-means clustering**