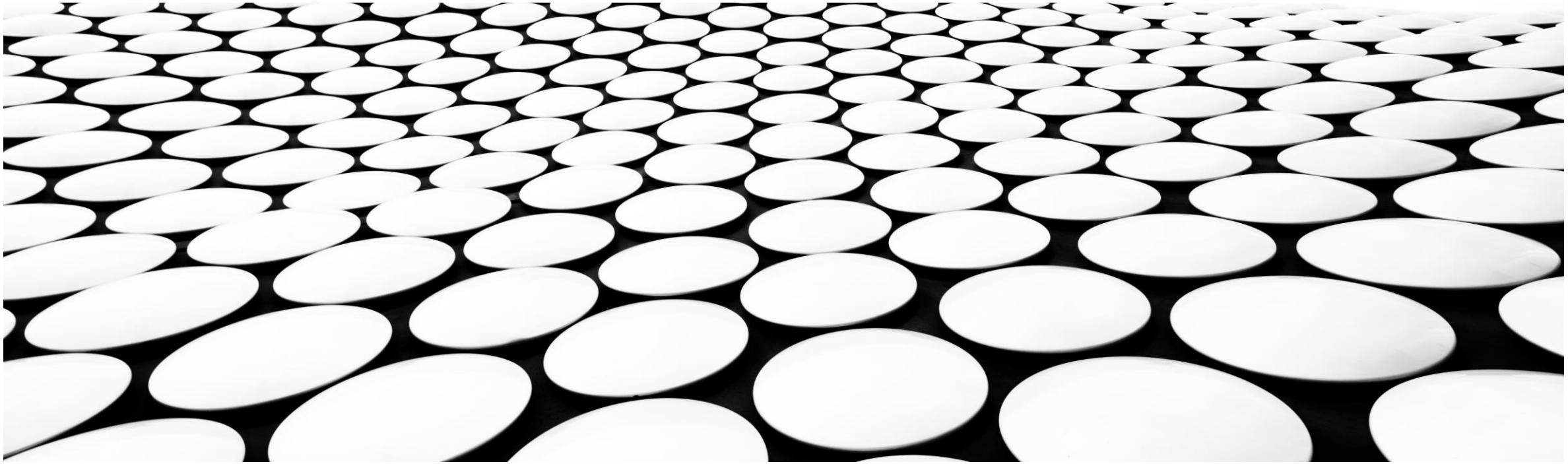
Capstone Project - The Battle Of The Neighborhoods

Applied Data Science Capstone by IBM/Coursera

HARSH V SINGH



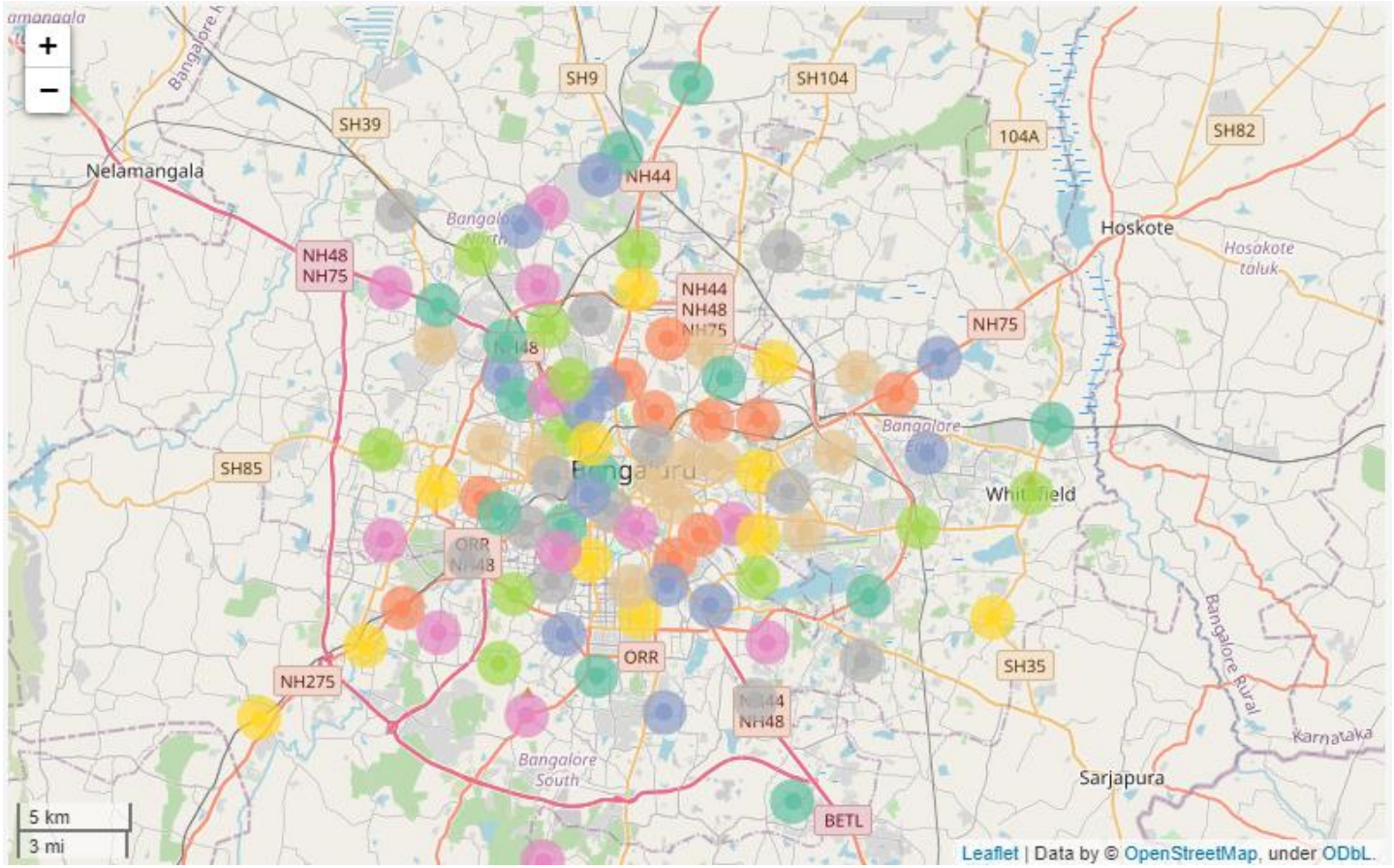
Introduction - Business Problem

- In this project, we will undertake the task of **identifying the best location to open a new Pizza Place in Bangalore, India.**
- The rapid growth of the city provides a great number of lucrative business opportunities. While the city already has a vast number of restaurants spread across different localities and neighbourhoods, there are many new neighbourhoods coming up as a result of the expanding city limits. At the same time, there are new residential and commercial development projects across various locations. Both these factors combined give rise to unique business opportunities for entrepreneurs in the **hospitality and food services industry.**
- In order to identify the most suitable location for opening a new restaurant, specifically a Pizzeria, we will follow these steps -
  1. Collect data and analyze the various neighbourhoods across Bangalore
  2. Identify similar neighbourhoods by categorizing them into clusters
  3. Identify the clusters which typically have a high number of Pizza Places
  4. Among those clusters, identify the neighbourhoods which have a scarcity of Pizza Places as compared to their respective clusters

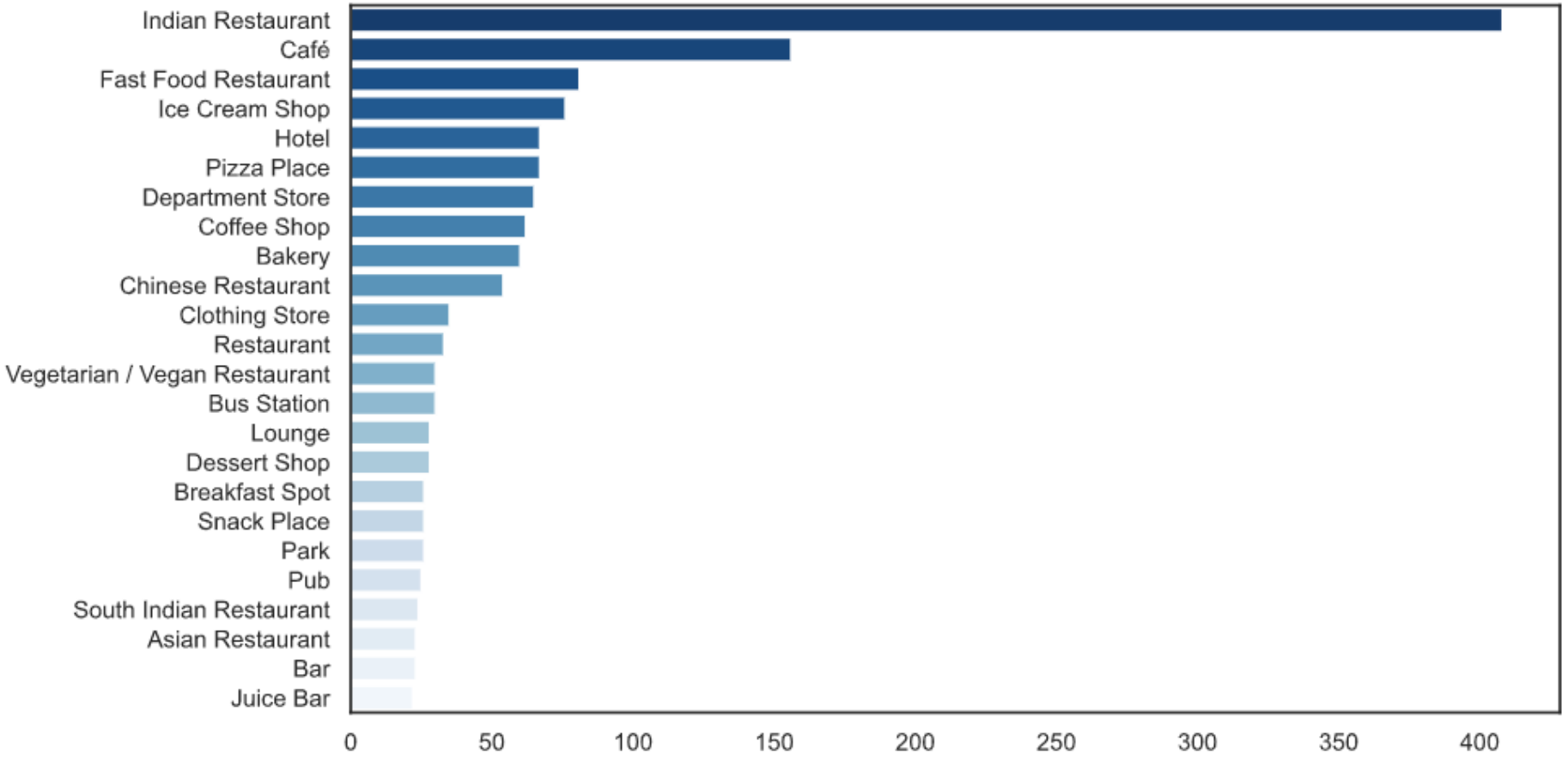
Data Collection And Exploration

- In the first stage of this project, we will collect the necessary data from various sources.
  1. We can get data on pincode-wise neighbourhoods in Bangalore from this website - https://finkode.com/ka/bangalore.html. We can use the **BeautifulSoup** library to extract data from the webpage and store it in a pandas dataframe.
  2. Next, we can do some preliminary clean up of the data - filtering the relevant pincodes, cleaning neighbourhood names, removing/ combining duplicate pincodes into a single row, etc.
  3. Once we have a clean set of neighbourhoods, we will use the **Google Maps API** to fetch coordinates for each of the neighbourhoods.
  4. We can then use the **Folium** library to plot the map of Bangalore and visualize the neighbourhoods.
  5. We will use **Foursquare API** to get the list of venues for each of these neighbourhoods

Visualizing Neighbourhoods In Bangalore



Top 25 Categories Of Venues In Bangalore Neighbourhoods



*Source: Foursquare API

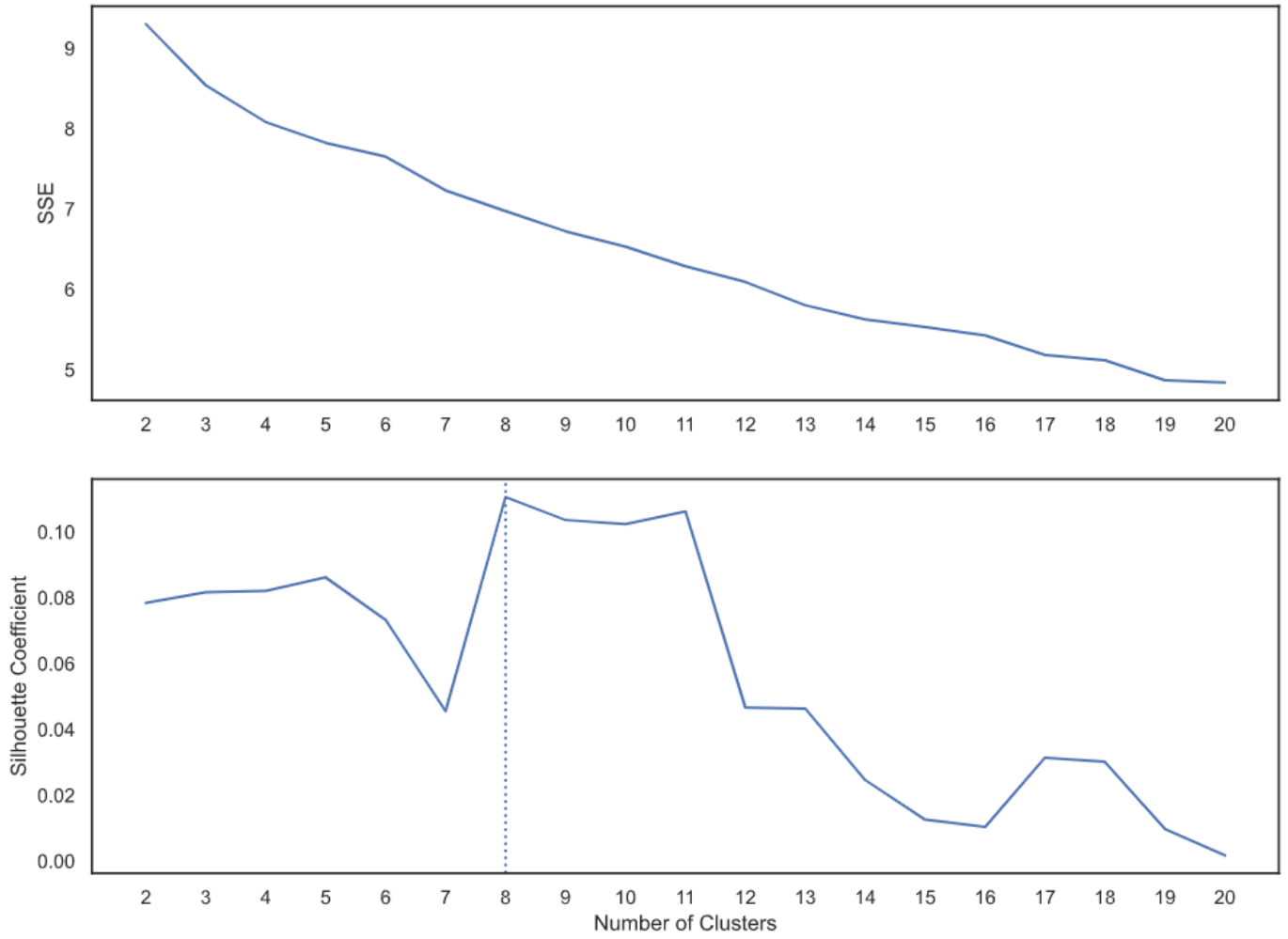
Sample Data - Most Common Categories Across Neighbourhoods

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	A F Station Yelahanka,BSF Campus Yelahanka	Indian Restaurant	Hot Spring	Lake	Department Store	Movie Theater
1	Adugodi	Café	Spa	Tea Room	Kerala Restaurant	Nightclub
2	Agara,Koramangala I Block,Koramangala,St. John...	Indian Restaurant	Café	Pub	Coffee Shop	Ice Cream Shop
3	Agram	Pizza Place	Historic Site	Indian Restaurant	Golf Course	Nightclub
4	Amruthahalli,Byatarayanapura,Kodigehalli,Sahak...	Indian Restaurant	Ice Cream Shop	Bus Station	Resort	Fast Food Restaurant
5	Anandnagar (Bangalore),H.A. Farm,Hebbal Kempa...	Bus Station	Ice Cream Shop	Fast Food Restaurant	Lake	Gym
6	Arabic College,Nagawara,Venkateshapura	Indian Restaurant	Department Store	Optical Shop	ATM	North Indian Restaurant
7	Ashoknagar (Bangalore),Banashankari,Dasarahal...	Indian Restaurant	Department Store	Park	Athletics & Sports	Snack Place
8	Attur,CRPF Campus Yelahanka,Jakkur,Rajanakunte...	Café	Indian Restaurant	Train Station	Smoke Shop	Food Truck
9	Austin Town,Viveknagar (Bangalore)	Clothing Store	Café	Park	Spa	Indian Restaurant

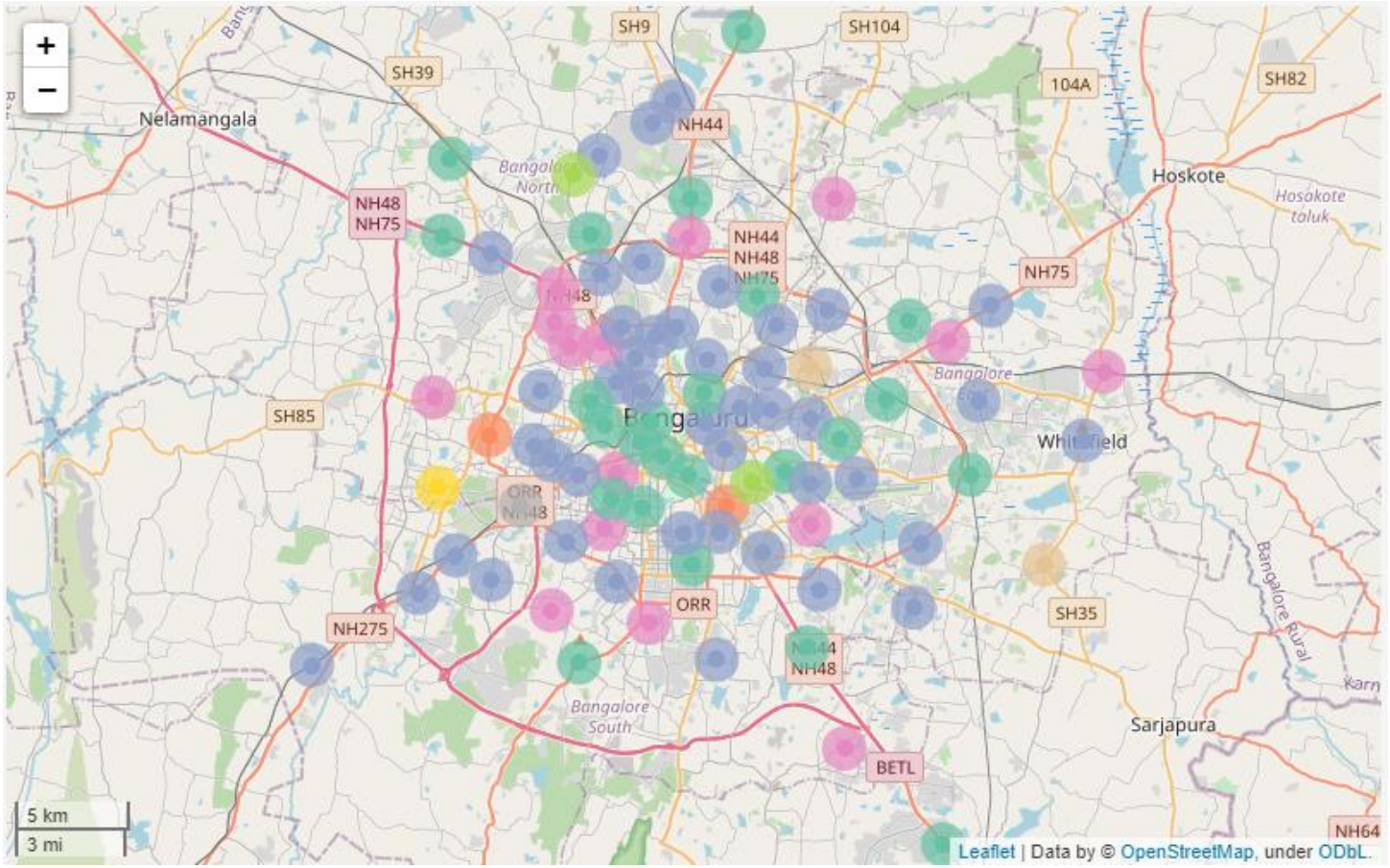
Modelling And Analysis

- Given that we have data containing the features of each neighbourhood in Bangalore, we can now use this data to build our K-means clustering model.
- We will run the algorithm for multiple values of k, and select the one with the highest **Silhouette Coefficient**. We will also plot the sum of squared errors and silhouette coefficients for each value of k.
- For our final analysis, we will add the Cluster Labels to the venue data. We will then analyze which clusters of neighbourhoods typically have a **high ratio of Pizza Places** located in their vicinities.
- Finally, we will try to **shortlist neighbourhoods** within those clusters which don't have a sufficient number of pizza places.

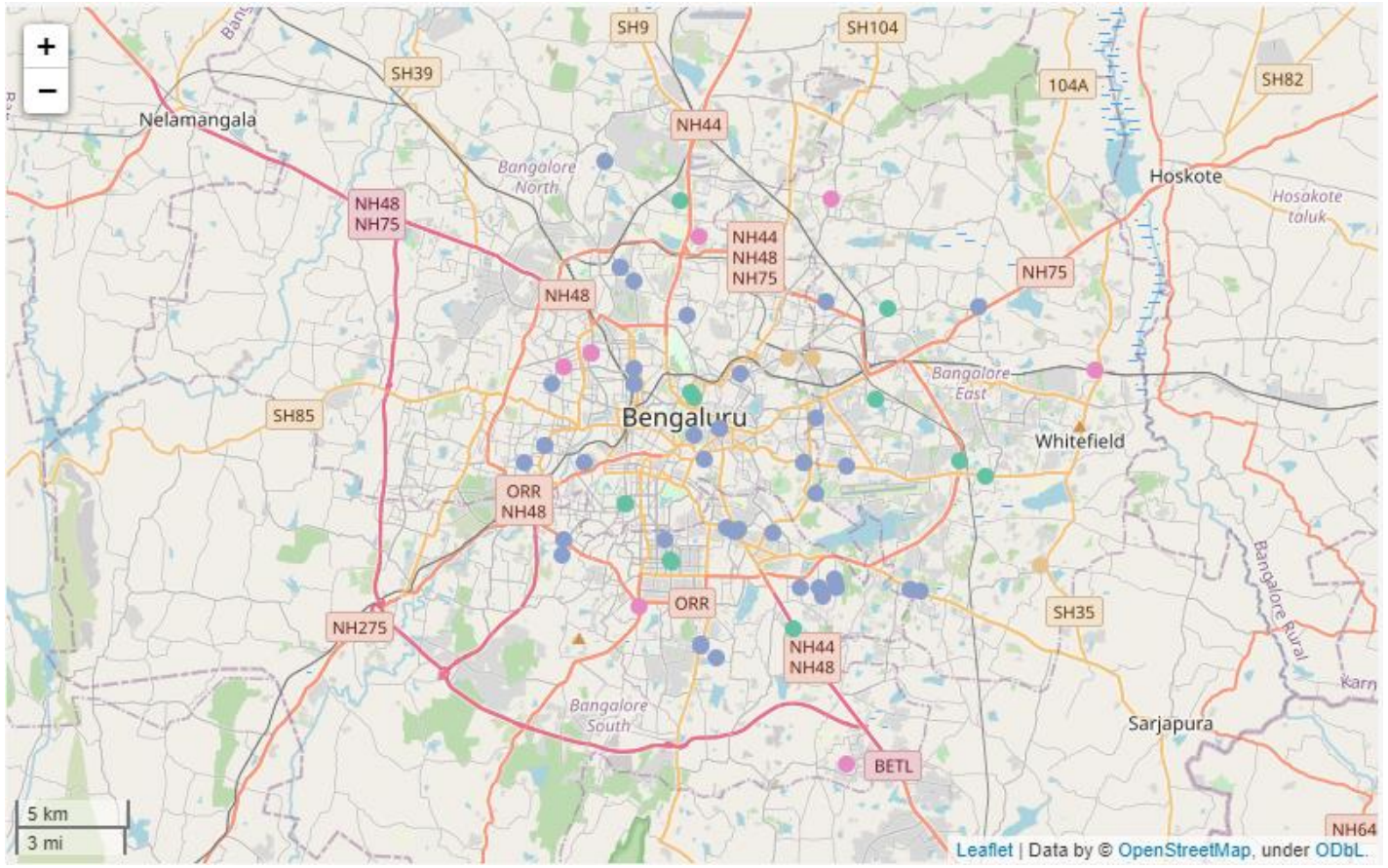
Hyperparameter Tuning For K-means Clustering Model



Neighbourhood Clusters In Bangalore



Locations Of Existing Pizza Places – By Cluster



Suitable Neighbourhoods For Opening New Pizzeria

	Cluster Labels	Neighbourhood	Total Venues	Pizza Venues	Avg. Pizza Venues	Avg. Venues	Exp. Pizza Venues
0	1	Sivan Chetty Gardens	73	0.0	0.028840	33.886364	2.105298
1	7	Jeevabhimanagar,Jeevanbhimanagar,New Thippasandra	47	0.0	0.023466	21.307692	1.102888
2	7	Bangalore Dist Offices Bldg,K. G. Road	38	0.0	0.023466	21.307692	0.891697
3	0	Ashoknagar (Bangalore),Banashankari,Dasarahal...	22	0.0	0.035556	14.062500	0.782222
4	7	Chickpet	28	0.0	0.023466	21.307692	0.657040
5	7	Sampangiramnagar,Shanthinagar,Wilson Garden	25	0.0	0.023466	21.307692	0.586643

Our final result includes a list of 6 neighbourhoods which are most suitable for opening a new pizzeria. We have presented this list in the order of most suitable to least based on the expected number of pizzerias that should *ideally be present in these neighbourhoods*. But given that there are many other factors that may go into deciding the suitability of a location, we believe that it is best to review each of these 6 neighbourhoods in further detail. This may require on-the-ground surveys and collecting other relevant data such as property prices, total population, average age and income demographics, residential vs. commercial developments in the area, etc.