FINDING OPTIMUM LOCATIONS FOR NEW BUSINESSES

COURSERA CAPSTONE

ROHITH RAMAKRISHNAN

INTRODUCTION

- For achieving success in any business , the location is the prime factor that determines its success or failure. Be it a grocery store , shopping mall , or even an ice cream parlor, the accessibility and the land mark close by play a vital role in customer outreach . Many at times, locations near reputed malls often draw crowds and could be a great opportunity to endorse the products or service.

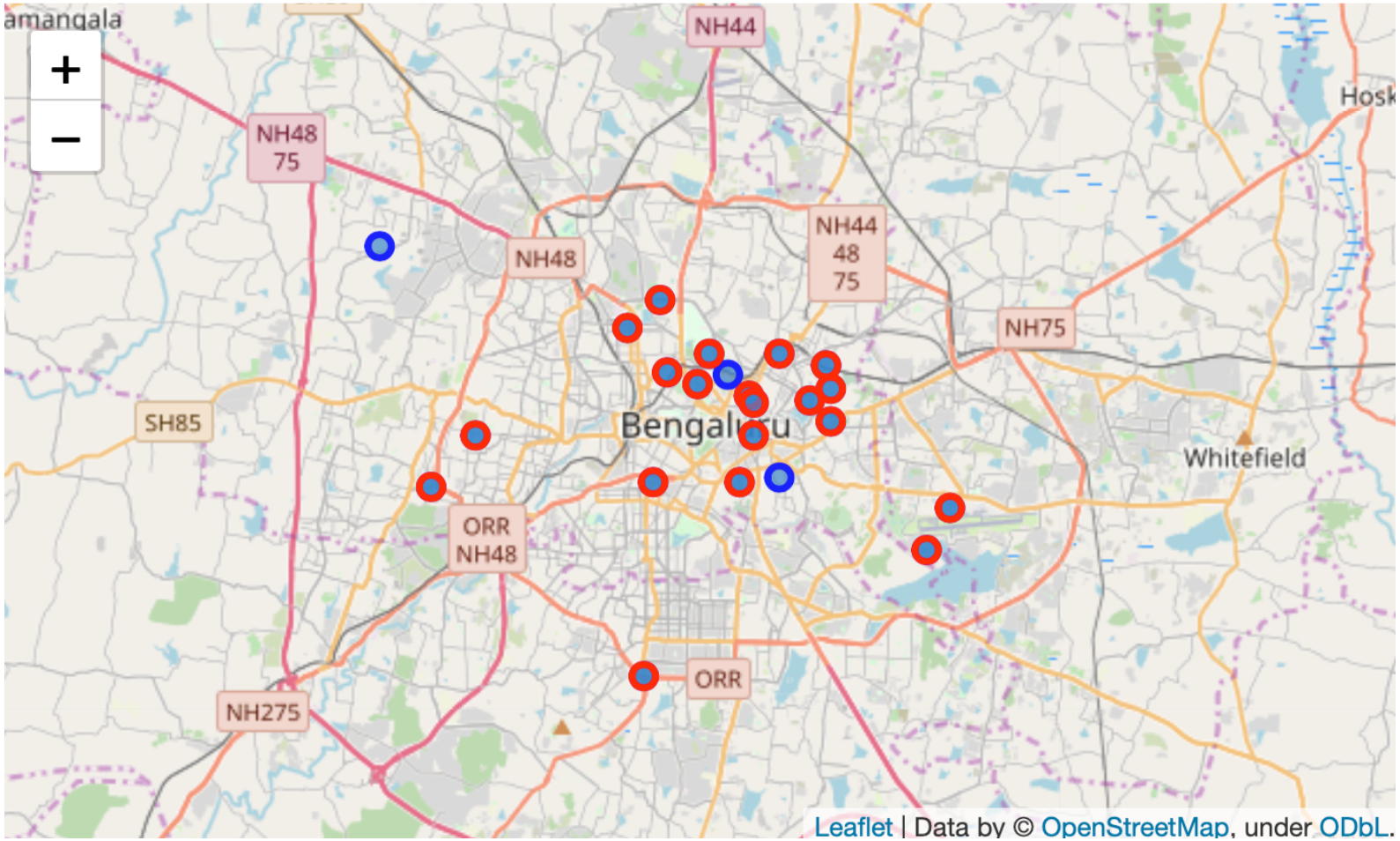
BUSINESS PROBLEM

- The objective of this capstone project is to analyze and select the optimum locations in the city of Bangalore , India to open a new business proposal. Using data science methodology and machine learning techniques like clustering, the aim is to provide solutions to answer the business question: In the city of Bangalore , India if a developer is looking to open a new service where would you recommend that they open it?

APPORACH TO SOLVE THE QUESTION

- We require the following data:
- List of neighborhoods in Bangalore.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to huge public attractions like shopping malls. We will use this data to perform clustering on the neighborhoods.
- Sources of data and methods to extract them This Wikipedia page(https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighborhoods.

- We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages.
- Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods.
- This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.



Result

For convenience , the data selected were from the Central Region of Bangalore.

After clustering all the areas of the city, the area highlighted red are the most suitable locations (Majority of the Central Region is heavily populated with tourists attraction and shopping malls) and the blues are the least preferred.