Capstone Project Report Zonification Of Restaurants

Applied Data Science

Agenda

- Introduction to problem statement
- Data
- Methodology
- Results
- Recommendations

Introduction to Business Problem Statement

- The problem statement here has been borrowed from online food delivery services (the likes of Swiggy and Zomato
- Quick view into the business model –
  - Facilitate an interaction between two parties (restaurants and consumer) with the help of a third party (delivery boy)
  - A typical transaction starts with the customer going on the online platform (web/app) and placing an order and ends with the customer receiving the order.
- Problem statement
  - During the transaction, there are two legs of travel involved for the delivery boy.
    - The first leg – Location where delivery boy was assigned to restaurant location
    - The second – Restaurant to customer location to deliver the food
  - In most of the cases, the customer does not really want to pay for the first leg of travel and it ends up being the company's responsibility to optimize it.
  - Here we will try to optimize the same through clustering the restaurants.
- The stakeholders for this problem statement will be the Operations team of these companies.
- With the number of transactions ranging to close to ~15-20 lacs per day, a saving of even a single minute per transaction will lead to immensely huge savings (assuming the delivery boys are paid at least 1 Rs per min).
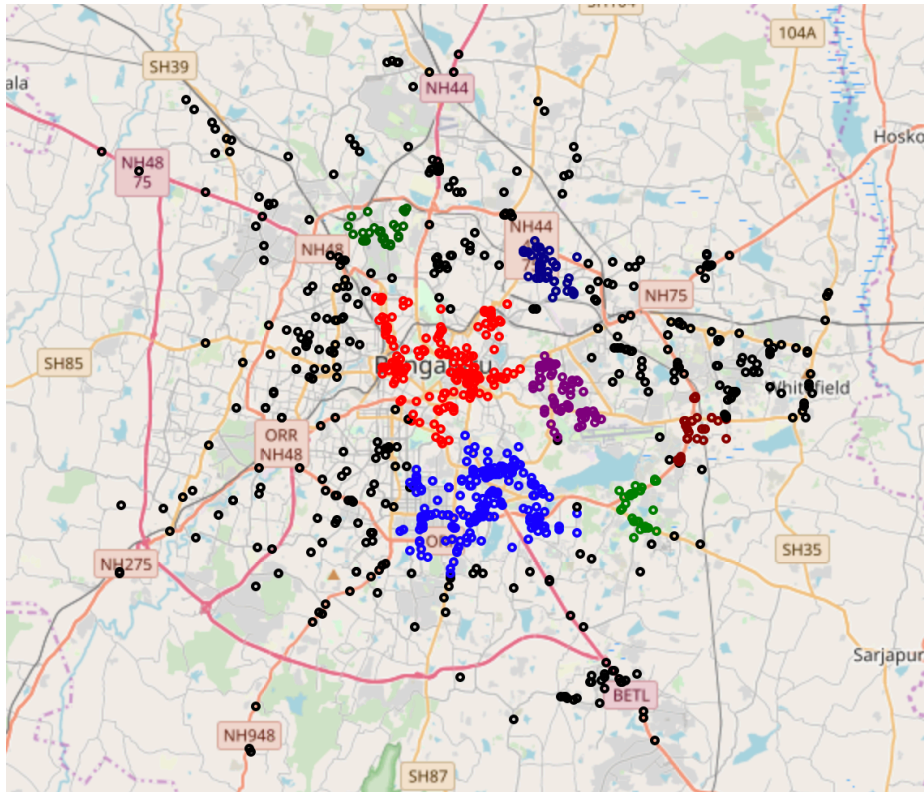
Data

- we relied heavily on the geographical location of restaurants (Bangalore)

- We used Foursquare APIs to get this data

- To avoid the limits of 1000 calls per day and 50 responses pre call on Foursquare APIs, we used the below workarounds

  - Range (the upper and lower limits of the loop) – In this case, we will run the loop over a range of ~15 Kms as a aerial radius from the centre of the city.

  - Step (the step in which the loop is incremented after each run) – In this case, we need to keep it not too small (to avoid hitting the limit) as well as not too large (so that all the restaurants for the given loop and not covered in any other loop don't exceed the limit of responses i.e. 50 rest). For the purpose of this project, we will keep this increment to 1km in each step.

- To avoid a lot of duplication, we kept a limit of 1600 m on the API call radius

Methodology

1. Problem Definition – Identification and articulation of the problem statement we are trying to solve for

⬇

2. Downloading data – Using Foursquare API in a loop to identify all the restaurants across Bangalore along with their latitudes and longitudes

⬇

3. Cleaning data – Removing duplicates as well as non-required columns from the dataframe

⬇

4. Visualizing data – Plotting the restaurants on a map to get a visual representation of the spread

⬇

5. DBSCAN – Running DBSACN and evaluating the results through a visualization

⬇

6. K Means Clustering – Running K Means and evaluating the results

⬇

7. Interpretation and conclusion
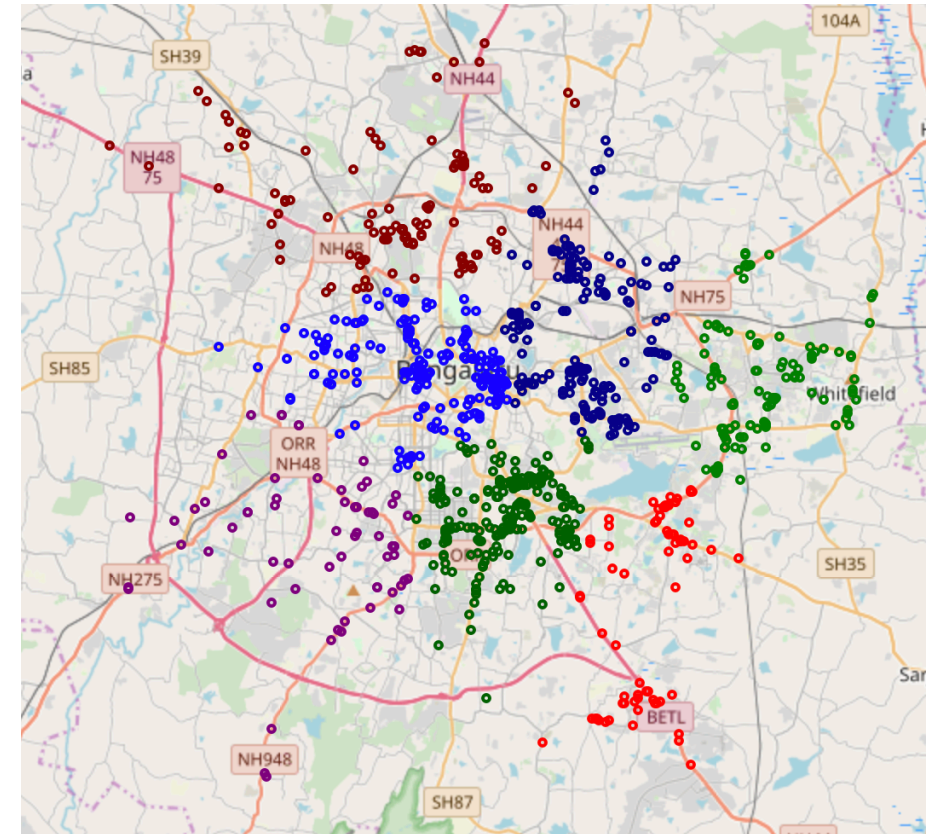
Results

DBSCAN Results

- A large number of restaurants are left out as noise (marked in black)
- Certain clusters (like royal blue) are too large to be operated efficiently.

K Means – Elbow curve
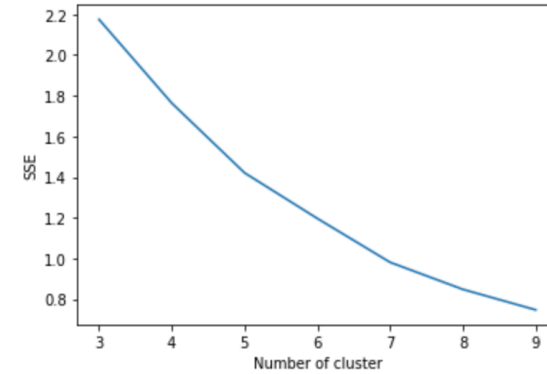
K Means – Results

Recommendations

- Map the restaurants to specific cluster based on the K-means clustering results

- Tshould be specific teams of Deliver boys exclusive for each zone posted in each of the zones/clusters

- This will result in significant savings on the delivery boy travel time/distance front

Thank you