The Battle of the Neighborhoods

Finding best location to open an ATM in Bangalore, India

IBM Coursera Data Science Capstone Project

Authored By: Taranpreet Kaur

CONTENTS

Introduction

- Bangalore is one among the fastest growing cities in the world.
- It is referred to as the Silicon Valley of India because of its role as the nation’s leading Information Technology exporter.
- It has a population of over ten million, making it a megacity and third most populous city and fifth most populous urban agglomeration in India.
- Being a demographically diverse city, the needs of the residents are also increasing rapidly.
- Hence, any new organization or an existing one should keep up with their pace in supplying the needs of the customers.

Business Problem

- Our customer is ABC Bank, which is an International Bank and also a market leader.
- ABC Bank has received ample amounts of complaints from residents of Bangalore that there aren't sufficient amount of ATM's.
- Given the extremely large population and the population of the city, our customer wants to identify the best neighborhood area to open more ATM covering the majority of the population.
- The problem statement will be: **Which neighborhood is most densely populated and has lesser number of ATM's?**

Data Description

- The data to be used in this project is not readily available. Hence, the data has been obtained from various sources such as
- Foursquare, which is a local search-and-discovery mobile app which provides search results for its users.
- Wikipedia, which has the details about the neighborhoods in Bangalore. https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore
- The geographic coordinates of each location have been obtained through Geopy
- The population data about each neighbourhood has been obtained from: https://www.ichangemycity.com/assembly-constituencies/mahalakshmi-layout

https://www.census2011.co.in/census/district/242-bangalore.html

Data Features

- As we have to explore and identify the neighborhoods in the city of Bangalore, the Bangalore neighbourhood data is the crucial data for this project.
- The data about each neighbourhood is not readily available, hence we have to scrape the Wikipedia page and obtain the data.
- In order to obtain the coordinates, we make use of geopy library in Python.
- We also need information about each neighbourhood which is obtained through FourSquare API.
- The population about each neighbourhood will let us know which neighbourhood is more preferable.

Methodology

Methodology involves the following stages:

- Data Preprocessing
- Exploratory Data Analysis
- Clustering

Data Preprocessing

- Scraping from the wikipedia page
- Obtain coordinates for each location using geopy library
- Finding columns with null values and replacing them
- Merge the neighborhood dataset with population dataset

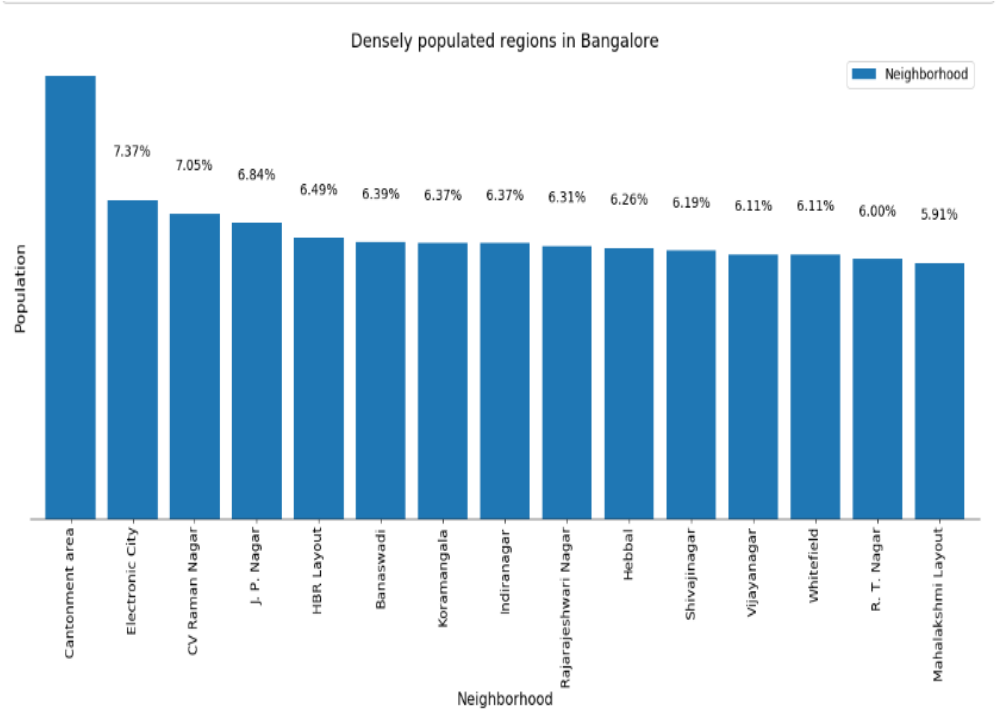
Exploratory Data Analysis

- Begin by exploring the city and finding the count of neighborhoods and regions in the city.
- Find the unique venues in each neighborhood.
- visualize the neighborhoods using a folium map.
- Obtain top 5 venues at each neighborhood, which will let us know which neighborhoods lacks in ATM Machine’s.
- By placing more ATM machine’s in densely populated regions, more customers will be satisfied and the complaints will gradually be reduced. Hence, we sort top 15 neighborhoods based on maximum population.

Clustering

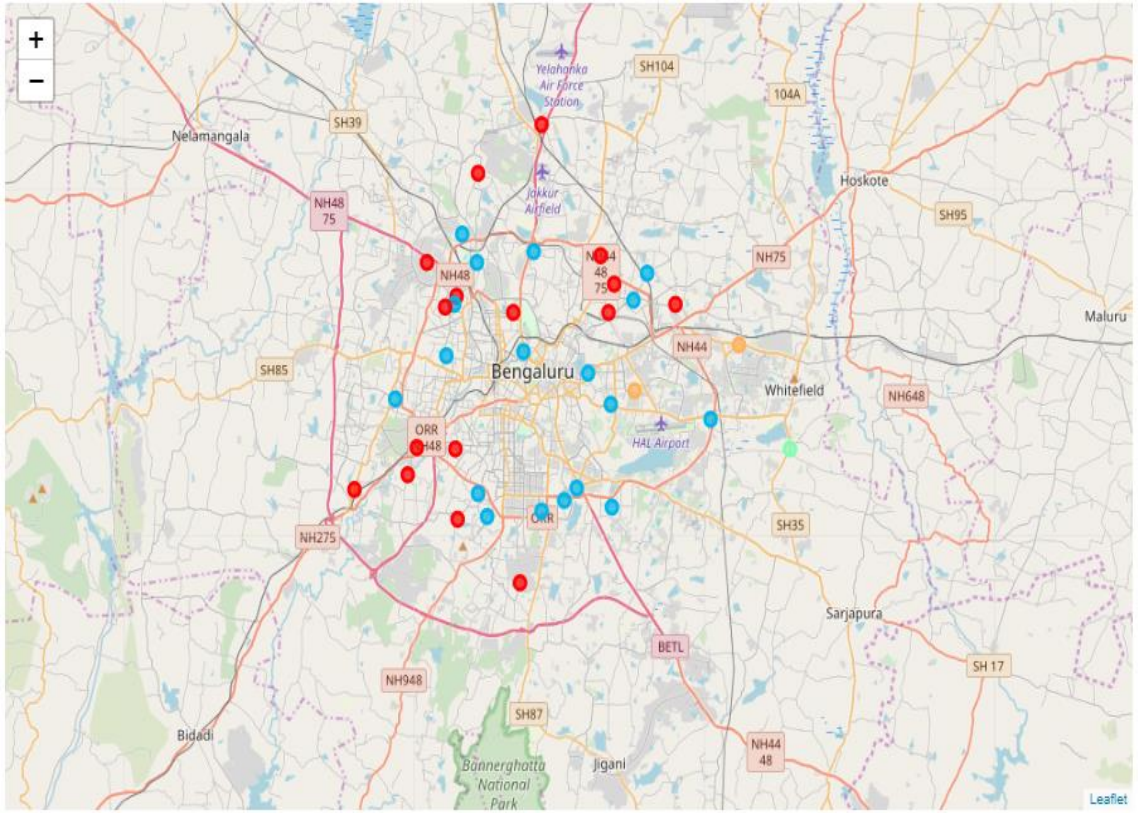
- Sort the neighborhoods into five clusters to obtain better insights about each neighborhood and by that we will know which cluster of neighborhoods have more population.

Results and Discussion



By sorting the neighborhoods based on population and visualizing it in the form of a bar plot, we find that the top 3 locations contribute to the 35% of the total population of the city.

Results and Discussion



By clustering the neighborhoods, we find that most of the densely populated neighborhoods belong to the first cluster.

Conclusion

By the population graph and the cluster map, we see that the first cluster contains most of the populous neighborhoods.

Hence, by placing more ATM machine's in the first cluster's neighborhoods, the needs of the customers will be successfully met and the complaints will reduce drastically.

During this work, some of the machine learning techniques, data wrangling with pandas and data visualization techniques were put to use.