Exploring Bengaluru Neighbourhoods

Analysis towards selection of a neighbourhood suitable to start a Sports Complex

Capstone Project for the Applied Data Science Specialization by IBM on Coursera

INTRODUCTION

DISCUSSION OF BUSINESS IDEA

Idea Statement:

Sports centres play an important role in a community. The benefits of sports not only help the body, but they also create a less stressed mind. People who exercise regularly are more than twice likely to consider themselves happy.

Houses that are located closely to newly built sport centres notice increase in property values. Studies have shown that people not only love going to recreation centres, but they are also willing to pay to live closer to one.

Having a sports centre improves the quality of life and empowers a community in general.

Bengaluru is the centre of high-tech industry in India with pleasant weather for most part of the year, hence attracting a huge crowd of people into the city. Given that it is a IT hub, there is an emerging problem of people spending most of their working time sitting. In the recent years, there has been a surge in fitness awareness.

Running a sport centre in the IT hub trends to be profitable, given the nature of the community.

Target Audience:

1. Anyone who is looking to open a sport centre in Bangalore. This analysis will explore Bangalore neighbourhoods and provide probable fits for opening a sport facility along with an analysis of what kind of facility to open
2. Government officials who are looking to uplift a community.
3. Anyone who wants to explore similar Bangalore neighbourhoods (based on the different venues in the area)

Data that will be used in the analysis:

1. A dataset which will give the latitudes and longitudes of all the different wards in the city of Bengaluru This data will be used to mark the boundaries of the different wards. The latitude and the longitude data will help in plotting a choropleth map.

2. A dataset to provide the area and population of each of the wards. This data will help in analysing the neighbourhoods in terms of population density and explore the target areas for setting up a sporting centre. The population density can be segmented into 3 categories and a bubble plot can be used to represent the different categories. Higher the population density, better the footfall.

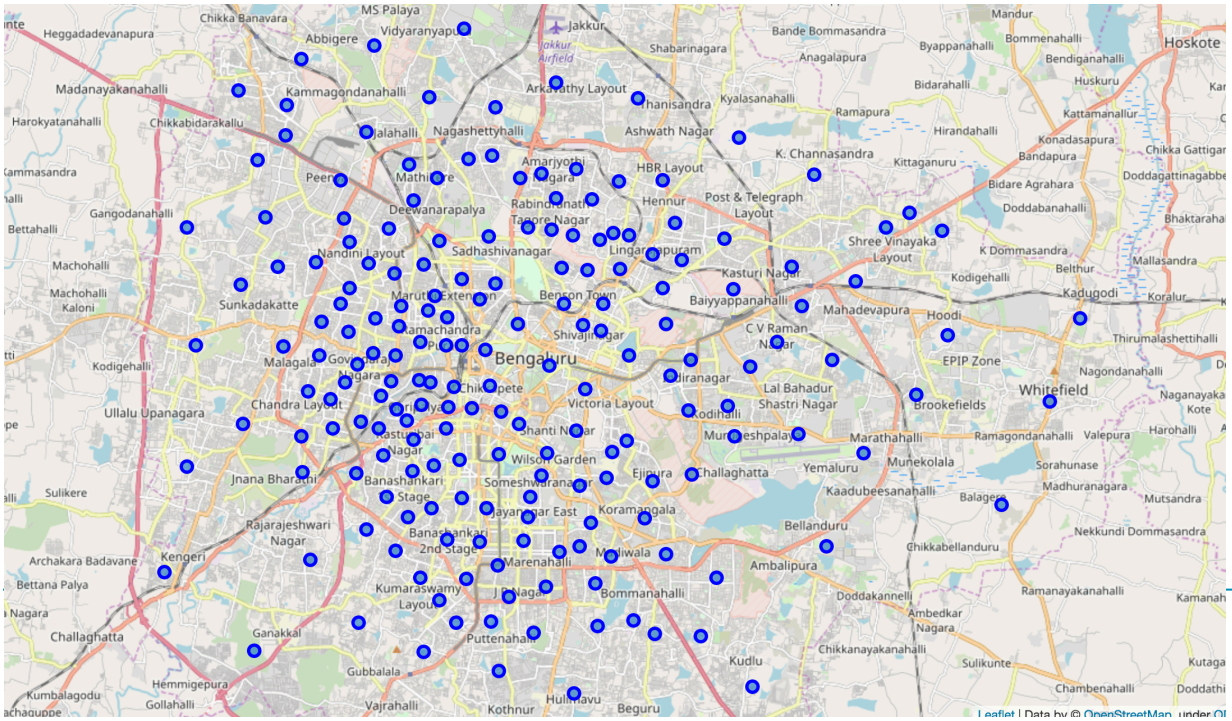
3. A dataset that gives the different venues in every ward. Foursquare location data will be used for this. The venue details and the different categories will help in clustering the areas to find neighbourhoods that are similar to each other. The neighbourhoods will be clustered based on non-sport venues, with the assumption that such characteristics define the development of sports facilities. We will also derive the density of sporting facilities in each neighbourhood. Once we have the clusters defined, a neighbourhood can be chosen such that there is least competition and also one which is similar to the neighbourhood with the higher number of sporting facilities.

The data for Bengaluru wards will be taken from the open city forum.

METHODOLOGY

The GeoJSON data for Bengaluru has the information related to neighbourhoods, wards, their areas in square feet and population of each ward. The data can be used to visualise the geographic details of the wards in Bengaluru, using the **Folium** library.

Bengaluru has 27 neighbourhoods with 198 wards. We get a visual of the wards as below:



In order to explore the wards, we use the Foursquare API, to see all the venues that are spread across a radius of 1KM around the longitude and latitude of each ward. From the data, we see that there are 252 unique venues across the 198 wards. Since our aim is to choose a ward that will be suitable for opening a sporting complex, we need to select a ward:

a. that has less competition
b. that has a high population density
c. that is similar in characteristics to a ward with high number of sports facilities

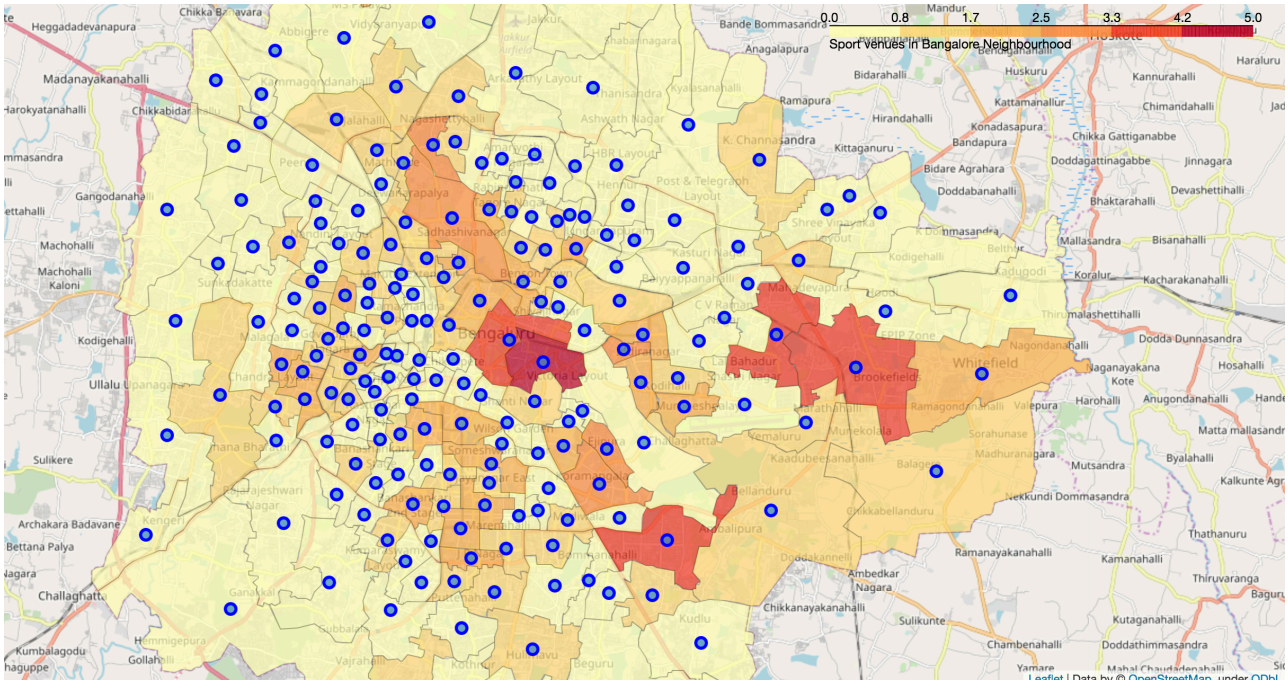
Wards that have less competition:

In order to find the venues that would prove to be less competition to our new sports complex, we need to determine the number of sporting venues in each of the wards, from the venues details from Foursquare. On cleaning and categorising the data, we see that there are 111 wards that have few sporting venues! Here are a few of them.

Neighborhood	Sporting_Venues_Count
A Narayanapura	0
Agaram	0
Azad Nagar	0
Bagalakunte	0
Banasavadi	0
Banashankari Temple ward	0
Basavanagudi	0
Basavanapura	0
Begur	0
Benniganahalli	0

All these 111 wards seem to be wards that would give less competition to our new sports complex.

Superimposing the density of the sporting venues over the Bangalore wards map would give us a clear visual picture of the density across wards. After some basic cleanup of data, we can visualise the densities as below:

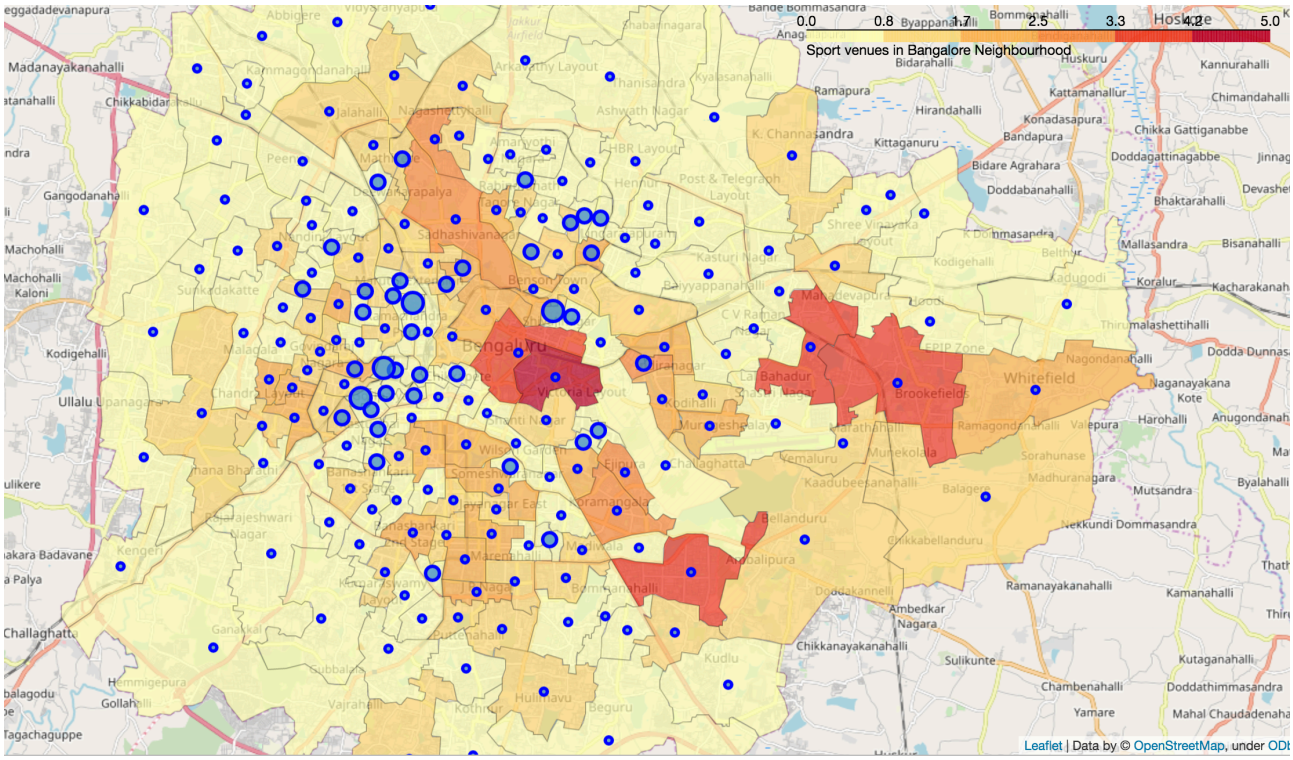


The darker the hue, higher the number of sport venues in the ward.

Wards that have high population density:

Our next criteria would be to compare the population density for each ward. The higher the density, more the footfall. We use the population detail at a ward level, to derive the population density of each ward. We then segregate the different population densities into 3 bins - Low, Medium and High and assign a category to each ward.

Based on which bin the ward falls into, we can then visualise the population density of the wards using a Bubble plot, with the bubbles varying in sizes as per the bin.



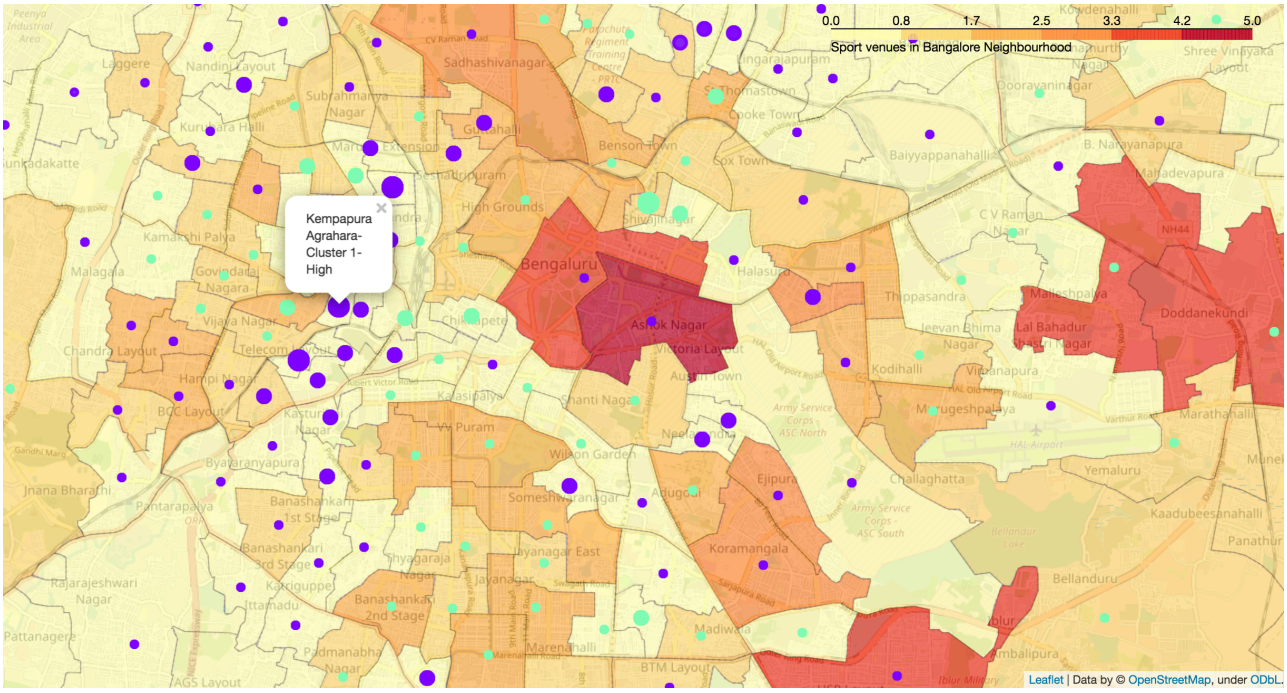
Wards that are similar in characteristics

From the above visualisation, we see that there is one ward with the maximum number of sports venues (Shantala Nagar). If we cluster all these wards on the basis on their non-sporting venues, it would mean that, a ward in the same cluster as that of Shantala Nagar would be a good fit for our new sports complex.

In order to cluster the wards on the basis of non-sporting venues, we use the k-mean algorithm, which is the most common method of unsupervised learning.

We run the k-mean algorithm to segregate the wards into 3 clusters, which seems to be an optimum k value. Running the algorithm, we derive the different clusters that each ward fits into.

Since we need to choose a ward that suffices all the 3 criteria (competition, population and characteristic), visualising all the 3 parameters on a single map gives the best picture of our analysis:



RESULT

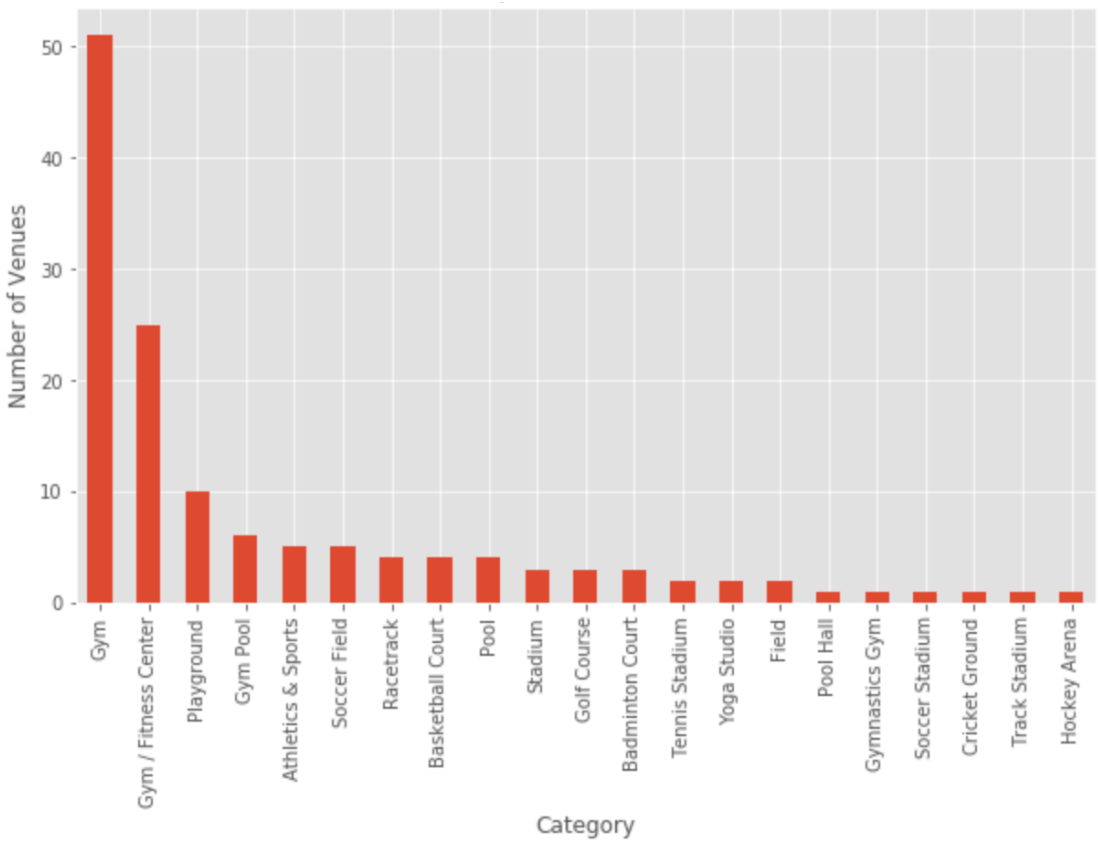
As can been seen clearly from the previous visualisation, there seem to be multiple wards that suffice our requirement. One of them being Kempapura Agrahara.

From the colour of the choropleth area of the ward (light yellow), we understand that the ward has few sporting venues (zero) and hence will be a no-competition ward.

From the size of the bubble, we know that the population density of the ward is high.

We also see that Ashok Nagar has a very high number of sport venues (the maximum of all the wards). And we also see from the bubble colour, that Kempapura Agrahara falls in the same cluster as that of Ashok Nagar. Hence, we know that the 2 wards have similar characteristics, which we assume would help drive traffic to our new sports complex.

Now that we have a choice of wards to set up our sports complex, from the data that we have, we can see what are the popular facilities that our sports complex should house. For this we group the sporting venue categories and see what are the most populous of the lot, across all of Bengaluru. We can visualise this analysis as below:



This shows us that our new sporting complex should accommodate a Gym, a playground and a Soccer field, since they seem to be that most popular amongst the crowd in Bangalore.

DISCUSSION

Bengaluru, with its expanding IT population, is in need of robust sporting and fitness infrastructure. The city has numerous wards, each having distinct characteristics. With limited venues data from Foursquare, we have been able to get a set of wards, that we assume will drive more footfall to our new sports complex.

We took into consideration the population density of each ward and also the number of sports venues already present in each ward, which will showcase the competition for our new sports complex.

In order to cluster the 198 wards, we use the k-means algorithm over the non-sporting venues in each ward.

In order to improve our choice of a target ward, we can definitely consider more parameters, such as the running rates of plots in each ward and the age group of the population in the wards.

CONCLUSION

As a result, with the data in hand, our final map gives a picture of the wards that can be chosen for our new setup. Setting up the sports complex in these neighbourhoods would definitely boost the quality of life of the people in the neighbourhood as well as improve the standard of the neighbourhood in itself.