

Find the Location for eSports & Amusement arcade

Pradeep Kumar Vasudeva

April 12, 2019

1. Introduction

1.1. Background

Esports (also known as electronic sports, e-sports, or eSports) is a form of competition using video games. Most commonly, esports takes the form of organized, multiplayer video game competitions, particularly between professional players, individually or as teams. Although organized online and offline competitions have long been a part of video game culture, these were largely between amateurs until the late 2000s, when participation by professional gamers and spectatorship in these events through live streaming saw a large surge in popularity. By the 2010s, esports was a significant factor in the video game industry.

An amusement arcade (often referred to as "video arcade" or simply "arcade") is a venue where people play arcade games such as video games, pinball machines, electro-mechanical games, redemption games, merchandisers (such as claw cranes), or coin-operated billiards or air hockey tables. In some countries, some types of arcades are also legally permitted to provide gambling machines such as slot machines or pachinko machines. Games are usually housed in cabinets.

A startup having plans to setup eSports and Amusement arcades chain in India. As pilot project they are looking for the best location in India to start their first eSports and Amusement arcade. This information can be used to find the location of interest to setup eSports and Amusement arcades.

1.2. Problem

Data that might contribute to find the place may include cities in Indian states, details of local restaurant, pub, café, theater, and many more popular places etc. The stakeholders of startup organization are having trouble to identify the city and location based on population, per capita income of state and nearby venues to select best possible location for pilot project. Also, to use the same information with data to go for future rounds of funding. This

project aims to find the best city in India and best place in the city to setup eSports and Amusement arcades.

1.3. Interest

Obviously, Startup firm would be very interested to find the most suitable city and location to have the first eSports and Amusement setup for competitive advantage, growth, returns and business values. Others who are interested in eSports or Amusement arcades may also be interested.

2. Data Acquisition and Cleaning

2.1. Data Source

The city and state or union territory list can be found in Wikipedia page [List of million-plus urban agglomerations in India](#). The per-capita income for the state or union territory list can be found in Wikipedia page [list of Indian states and union territories by GDP per capita](#). The location data and geographical location coordinates are obtained by leveraging foursquare location data to compare, visualize, explore cities, places of interest, venues and other relevant information.

https://en.wikipedia.org/wiki/List_of_million-plus_urban_agglomerations_in_India

https://en.wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_GDP_per_capita

2.2. Data Cleaning

Data scraped from multiple Wikipedia pages. There several problems encountered during scraping pages, First the column name of city was mentioned as UA. This was cleaned by naming the column with heading ‘City’.

Second The population figures of cities in Kerala are inflated as the definition of urban agglomeration was revised in the 2011 census. The population of urban agglomerations in Kerala is not comparable with other cities in the country. However, for Kerala state cities the available information on the Wiki page is used for this project.

Third there were two population columns with Column names as Population (2011) and Population (2001). By verifying more information, it is found that the population figures of 2011 were latest and are retrieved n 14 December 2014. Hence 2011 Population is considered for the project. The details for the same can be found below. ["India: Major Agglomerations"](#). citypopulation.de. [Archived](#) from the original on 17 December 2014. Retrieved 14 December 2014

After fixing these problems it is found that the cities Hyderabad and Visakhapatnam had super imposed numbers which was addressed by dropping those cities from consideration to avoid program error during calculations.

Luckily there were no missing values, but there were unnecessary data such as per capita PPP, Comparable country by GDP per capita which are of no use for the project hence those were cleaned by dropping the respective columns.

2.3. Feature Selection

After data cleaning. There were 51 cities and 24 states or union territories. After examining the data, the rank, City, Population and State per capita income was considered for the project. Using the details, the Venues are obtained using foursquare location data and the weights for unique Venues were randomly assigned for the purpose of calculations.

Table 1. Simple feature selection during data cleaning

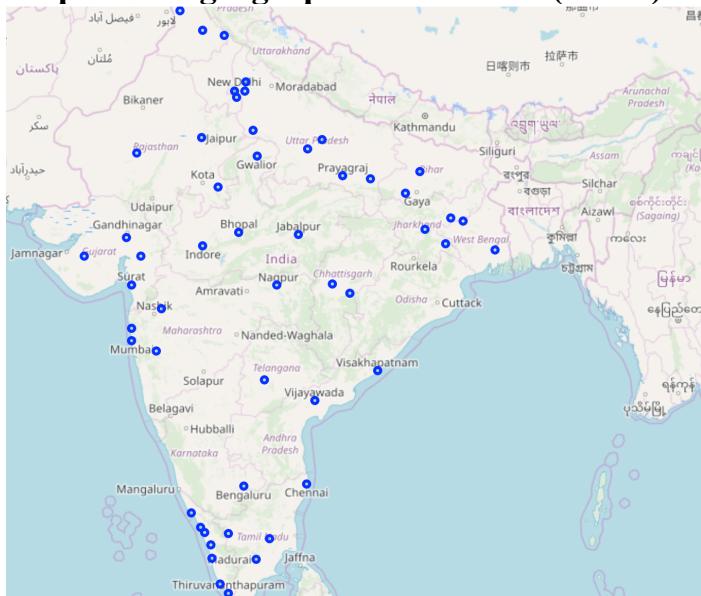
Kept features	Dropped features	Reason for dropping features
State union Territory	Data year	Data year is the year on which the per capita income was calculated, it is of no use
Population (2011)	Per capita PPP	This is IMF PPP Conversion rate hence it is dropped
Population (2001)	Comparable country by DGP per capita	We were interested in knowing the states of India per capita income hence the comparable country was dropped
Rank		
Per capita income		
Per capita income State or union territory		

3. Methodology & Exploratory Data Analysis

3.1. Finding the Location

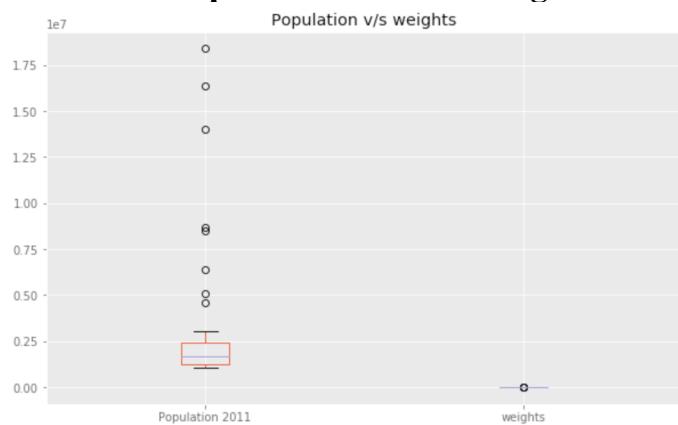
Finding location among the 53 top India cities with available data was not straight forward, with the scraped data all the Indian cities were visualized for geographical locations for possible visual information. Find below the Map of India.

Map of India geographical locations (Cities)



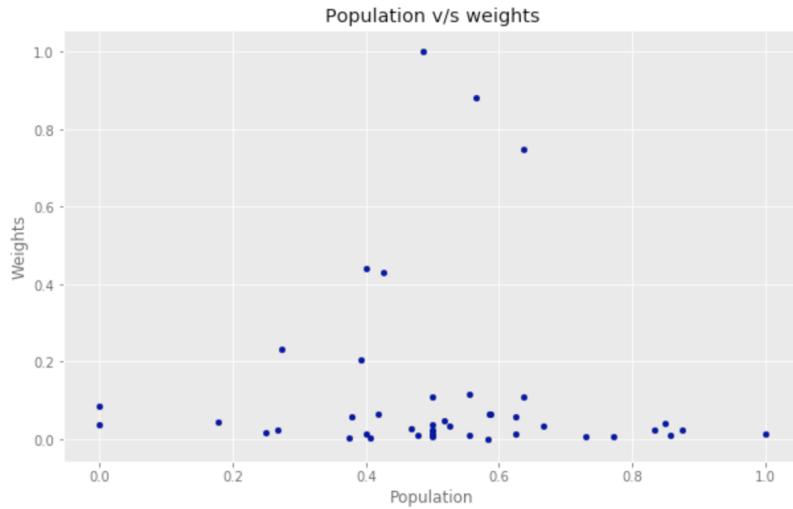
Population and weights were visualized to find some interesting information such that there almost 8-10 cities with huge population. One of these may a prospect for the project and some may be considered for future. Find below the box plot.

Box Plot – Population 2011 v/s weights



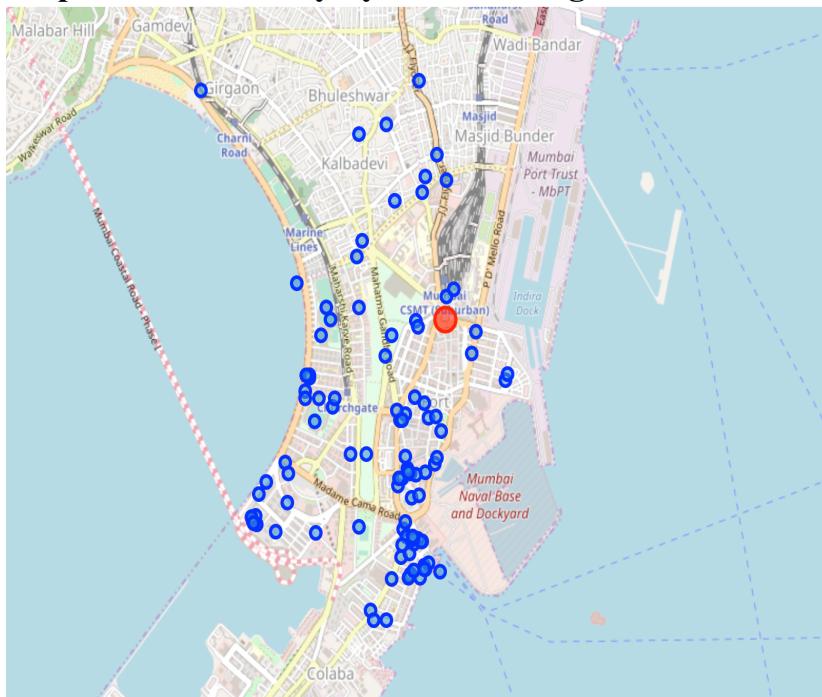
The location weight is at high with average city population and the weight of the location is really low when the city population is either very high or low. This may be good information for future. Find below scatter plot.

Scatter plot Population v/s weights



By visualization the Model selected City shows that the area covered by venues is large and almost all parts of Mumbai has good number of popular venues which makes it crucial to analyze the data and find best suitable place for eSports and Amusement Arcade

Map of Mumbai City by Venue Categories



4. Predictive Modeling

There are many methods and algorithms to predict or find required results based on the problem to be solved, Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields.

For this project considering the problem statement and desired output clustering technique will be used to solve the problem.

4.1. Applying Algorithm

There are many models for clustering out there. In this project, we will be presenting the model that is considered the one of the simplest models among them. Despite its simplicity, k-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data.

Some real-world applications of k-means include:

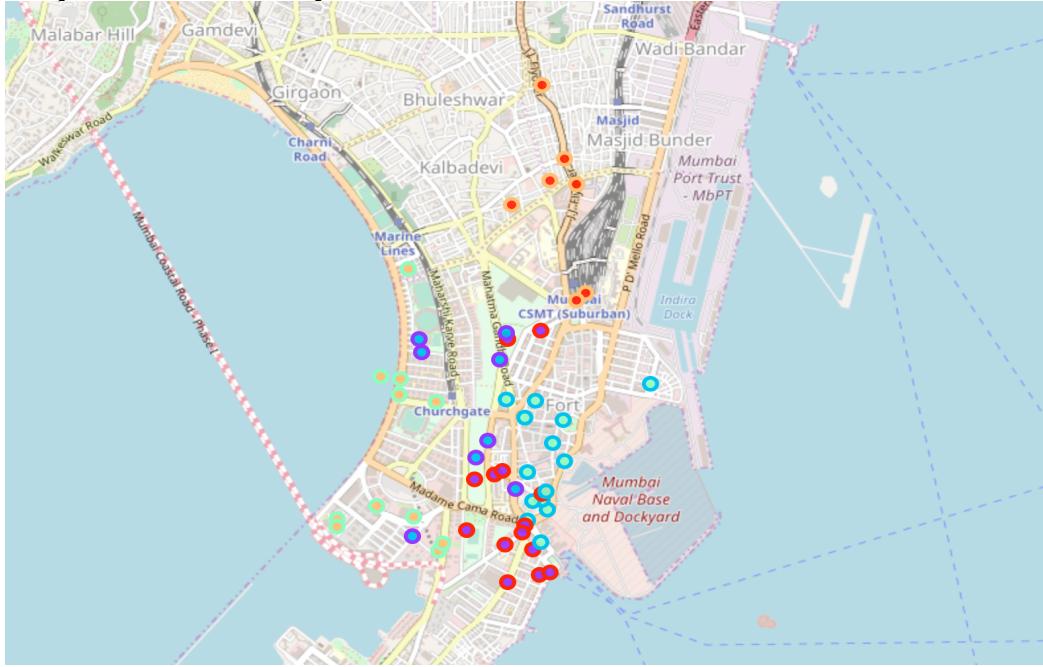
- customer segmentation,
- understand what the visitors of a website are trying to accomplish,
- pattern recognition, and,
- data compression.

4.2. Solution & the problems

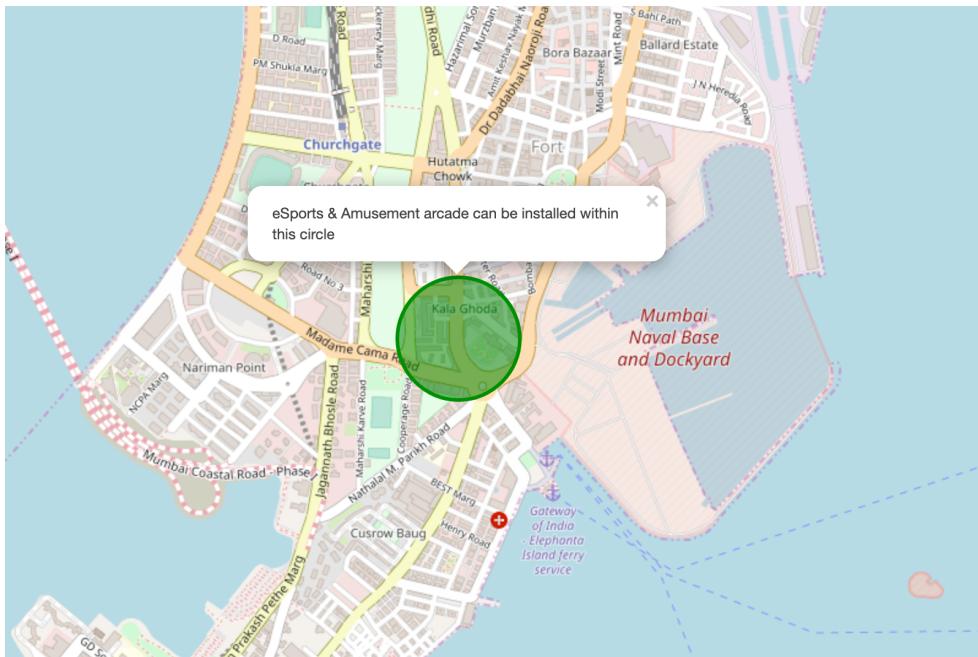
K-Means is probably the most well known simplest clustering algorithm. It's easy to understand and implement. To begin with clustering for the data, which we discussed earlier in the sections Data source and Data cleaning, first selected 3 as number of clusters and later total number of clusters are selected as 5. The center points are randomly initialized, and 5 different groups were visualized on Mumbai city Map to check and understand the distinct groupings. Each Venue Category is classified by computing the distance between that point and each cluster center, and then classifying the

point to be in the cluster whose center is closest to it. To visualize each cluster was given different colors. The only downside of this approach is to select the clusters upfront however for this project this may not be trivial because we are looking for best location and have plans to expand the business to other locations.

Map of Mumbai City – Clusters



Map of Suggested Location in the Mumbai city for eSports and Amusement Arcade



4.3. Result

The best location to setup eSports & Amusement arcade is found based on the cluster's weights.

The cluster weights are found by taking the average of all Venue categories' weight in the respective clusters.

The new eSports & Amusement arcade can be setup within 2 kilometers radius for geo coordinates (18.9268,72.8314) location in Mumbai City

5. Conclusion

In this study I analyzed various cities in India and several states,

the features like city population, per capita income for each state, geographical locations, Venues categories people most visit are considered along with randomly assigned weight for each venue category.

the K means clustering machine learning technique is used to find the final location in the Mumbai city

Based on the problem statement Mumbai city has been found as the best city to setup eSports & Amusement Arcade, a location within Mumbai city is found using the clusters and average weight of clusters.

The K means clustering technique is very useful in identifying locations similarly for various other cities

6. Discussion & Future Directions

The K means clustering model works mainly based on the features selected and the number of clusters used.

It is possible to use the foursquare API radius with bigger radius search to find more venues and categories.

Also, more features can be added based on the data and its available source.

In the Four-Square API, we have queried the Venues of a locality by specifying the LIMIT and Radius of our choice.

We have chosen less LIMIT as the number of API calls that can be done using a free account in Four Square is limited.

We can increase the limit for more accurate results.