|  |
| --- |
| Capstone Project-Data science specialization from ibm  Neighborhood exploration of Bangalore City and locality suggestion for opening a vegan restaurant |
| Bibhu Ashish Panda |



Table of Contents

[Executive Summary 3](#_Toc47885797)

[Introduction 3](#_Toc47885798)

[Data sources 4](#_Toc47885799)

[Methodology 4](#_Toc47885800)

[Result 5](#_Toc47885801)

[Conclusion 5](#_Toc47885802)

# Executive Summary

Data Science is the process of using data to understand different things, to understand the world. It is the science of validating a model or hypothesis with the help of data. In other words, Data science is the art of uncovering the insights and trends that are hiding behind data. In a nutshell Data science is the study of data. It is more about data than it is about science. Data science is relevant today because we have tons of data available to us now a days.

Data about location has been especially important for all businesses to derive intelligence that helps to take important strategic decisions. Data about location helps the business to understand the behavior of consumers in terms of their preferences. It gives much more insight than where the customers stay and work. With the widespread availability of Location data service providers like Foursquare and Google, the access to location data has become quite easy.

The Foursquare Places API, that is being used in this project provides location-based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.

This project aims to use the data science tools and techniques to explore the various places of a city in India and recommend a suitable place to start a business in India

# Introduction

Bangalore is considered to be the Silicon Valley of India and has witnessed tremendous growth in the last 10 to 15 years with a lot of Indian and foreign MNCs setting up their IT offices in the city. Even after such an explosive growth, the city is still considered to be a hot bed for innovation in software and IT industry. With inflow of a lot of professionals to this IT capital of India, the city holds tremendous potential for all kind of business opportunities. With a lot of young and health conscious IT crowd, opening a vegan restaurant holds a lot of promise for any entrepreneur who wants to own a piece of the city’s business. With the help of Data science tools and techniques this project aims to suggest a locality to a would-be entrepreneur or business to open a vegan restaurant in Bangalore.

# Data sources

When it comes to analysis of data, the source of the data holds the same importance as the process that is used to analyze the data. Data science inside a company is only going to be as valuable as the data collected. Garbage in, garbage out is a rule in any sort of analysis. A wrong source of data may not be able to give the proper view of the analysis and the analyst will not be able to derive a proper conclusion from the data thereby resulting in wasted effort and time. So, it is always important to choose a source that is reliable and correct. Considering these factors, I have used the below sources of data to analyze the locations of Bangalore and derive my recommendations

1. Wikipedia Bangalore Neighborhood locations

[**https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Bangalore**](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)

1. Geocoder Package for latitude and longitude coordinates
2. Foursquare API venue data

# Methodology

The below methodology was applied to address the requirement

For the problem question at hand, it was assessed that any reliable sources of location data would be helpful. I examined the APIs for genome, Zomato and Wikipedia and found the Wikipedia data to be reliable and useful without any unnecessary data.

The Wikipedia neighborhood data for Bangalore was extracted using an API and then JSON library was used to load the data. The data was cleaned and formatted using various Python libraries. After the data was cleaned it was visually explored to find out if we have all the relevant geographic data. This data was then used to overlap with the location data of FourSquare API.

As I was dealing with a set of unlabeled data and as this requirement dealt with the assessment of human behavior, I used Clustering Association approaches. Clustering which is an unsupervised machine learning algorithm is grouping of data points or objects that are similar and helps in discovering structure, summarization, and anomaly detection of the data sets. There are various types of clustering algorithms such as partitioning, hierarchical, or density-based clustering.

As part of this approach, k-means clustering was used to segregate the localities to various clusters . k-means is a type of partitioning clustering that is, it divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels. This means it is an unsupervised algorithm. Objects within a cluster are similar and objects across different clusters are quite different or dissimilar.

# Result

The analysis shows that there are a good number of vegan restaurants spread across the Bangalore city. Though the number is scattered across Bangalore city, couple of places where the Vegan restaurants can be targeted to open are places where there is one restaurant, but they have a good number of IT folks. These places can be Electronic City, BTM Layout, Indira Nagar. It is also surprising to see that Peenya like place having maximum number of Vegan restaurants.

# Conclusion

Data Science helps us to make sense of data. Machine learning algorithms like supervised (regression, classification) and unsupervised (clustering) are great tools that help business to be proactive in their decision-making process.