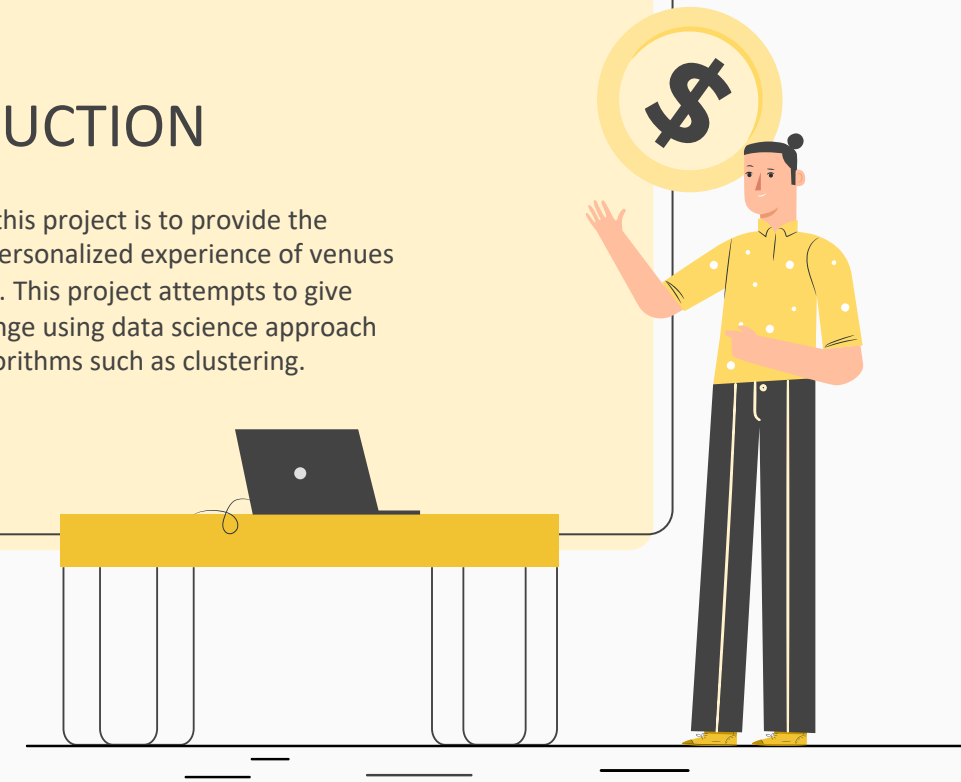
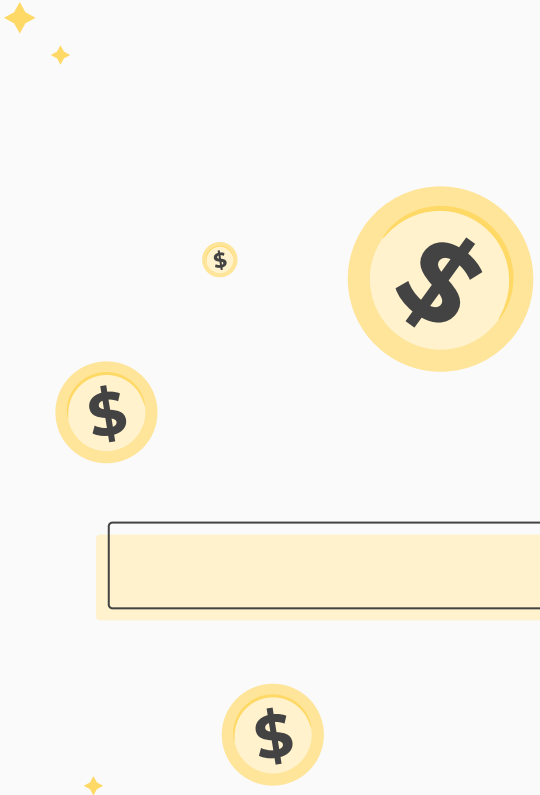
IBM DATA SCIENCE CAPSTONE

By;
Roshan
2021

INTRODUCTION

The problem identified for this project is to provide the customers/users with the most personalized experience of venues around them for them to visit. This project attempts to give solutions to the business challenge using data science approach and machine learning algorithms such as clustering.





“This prompted me to create a recommendation system model that might propose local restaurants their location based on other people's restaurant ratings, to potentially enhance the recommendation suggestions that you can see on popular food delivery services.
”

STUDY OBJECTIVES

1.

Recommendations to tourists based on their preferences and ranked accordingly

2.

Popular restaurants around the area

DATA REQUIREMENTS

Geographic coordinates

Area Population

Area’s Average Income

DATA REQUIREMENTS

01

We need to know its Latitude and Longitude

02

Neighborhood's population is a significant element in influencing a restaurant's growth and the number of customers that come in to dine.

03

Also, the higher the number of visitors, the higher the restaurant's rating because it is visited by individuals of various tastes. As a result, it is a crucial element.

04

A neighborhood's income is just as significant as its population. The wealth of a neighborhood is directly related to income.

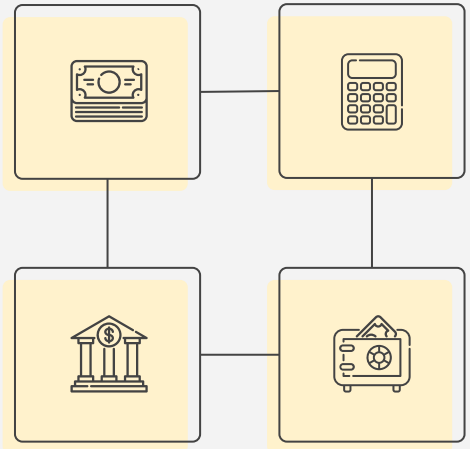
DATA COLLECTION

STEP 01

Gathering location

STEP 02

Google Maps API to get latitude and longitude



STEP 04

The Foursquare API

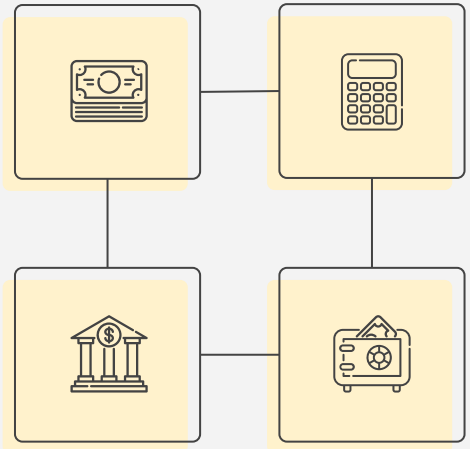
STEP 03

Pandas dataframe using the beautifulsoup4 package

METHODOLOGY

STEP 01

Get a list of the city of Bengaluru's neighborhoods



STEP 02

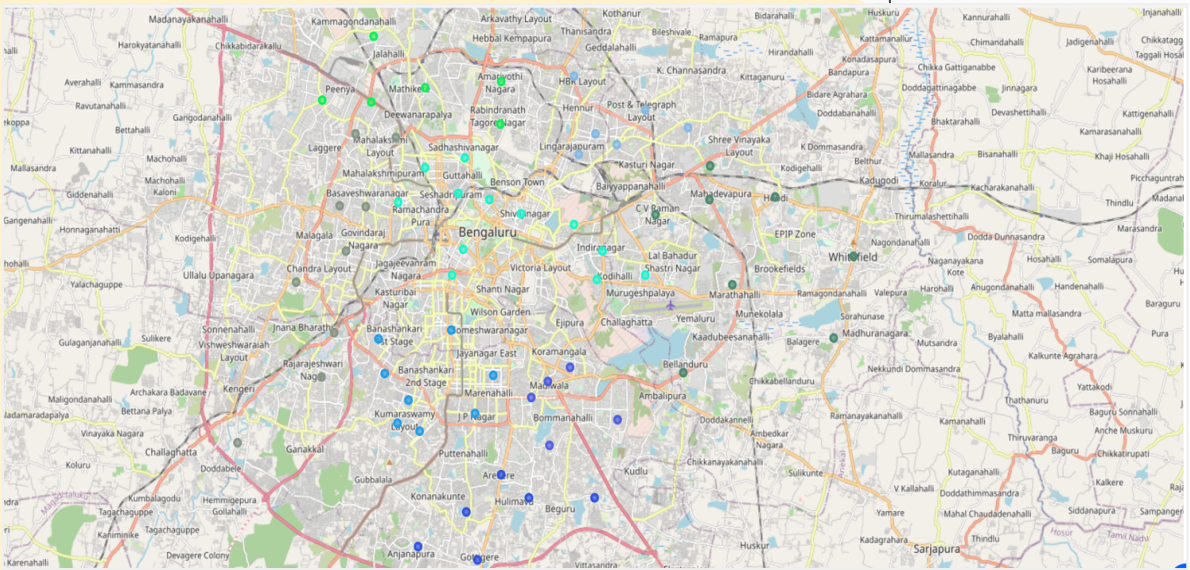
We'll collect the data, put it into pandas DataFrame,

STEP 03

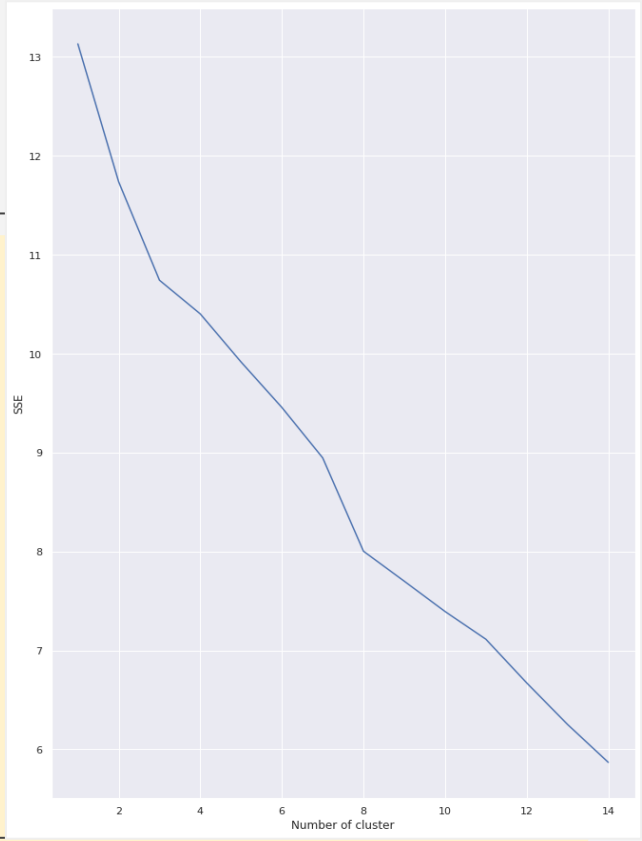
Folium package to display the neighborhoods on a map

STEP 04

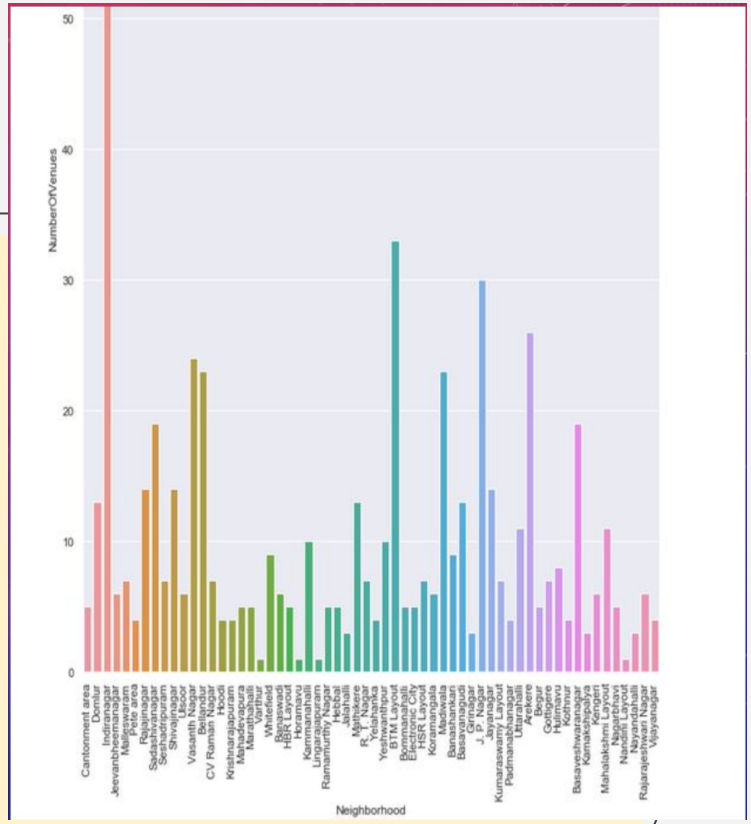
Then, by grouping the rows by neighborhoods and calculating the mean of the frequency of occurrence of each venue type



Clusters formed



K-Means Cluster Graph



EXPLORATORY DATA ANALYSIS

DISCUSSIONS

01

Population and Income were the most significant variables in developing the recommender system

02

When the population of a neighborhood grows, it does not always indicate that the community's average income grows as well.

03

Similar neighborhoods must also be thrown into the appropriate cluster during clustering.

04

Another point to consider is that the number of clusters used might yield a wide range of results

RESULTS ANALYSIS

Out[103]:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Arekere	Venue Category_Indian Restaurant	Venue Category_Sporting Goods Shop	Venue Category_Department Store	[0.2295988840700646]
1	Begur	Venue Category_Supermarket	Venue Category_Café	Venue Category_Mobile Phone Shop	[0.6361321887351776]
2	Bommanahalli	Venue Category_Department Store	Venue Category_Indian Restaurant	Venue Category_South Indian Restaurant	[0.4365669702740494]

FINAL RESULT

LIMITATIONS AND CONCLUSION

1

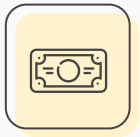
Additional Factors to be considered

2

Sandbox Tier Account

3

Future system - Best places to build additional infrastructure



1

Foursquare API to find local venues

2

It will assist consumers in quenching their hunger