|  |
| --- |
| **Tractor Dealer Finder and Guide for Pune Location** |
|  |

|  |  |
| --- | --- |
| 24/08/2019 | IBM Capstone project |

**Table of contents**

|  |  |
| --- | --- |
| **Topic** | **Page number** |
| Introduction section | 2 |
| Data section | 4 |
| Methodology section | 7 |
| Result section | 10 |
| Discussion section | 11 |
| Conclusion section | 12 |

1. **Introduction:**

**Problem background:**

Currently, I am working with a agriculture machinery manufacturing company.

Using the Foursquare data to track the dealer’s locations and to analyses the customer spend pattern and dealer satisfactory rating by customers would be helpful in understanding any relation between specific regions and frequency and intensity of complaints received from customers.

The location data updated at every customer's end would help the customer too in understanding dealer locations near to him and to understand other things like tips/ratings of each dealer.

Typical problems addressed in this exercise would be:

How many dealers are there based on proximity to each target audience?

Which is the dealer with best rating?

How many dealers from same company are in specific areas?

What unique dealers are in Pune and open for business?

**Target Audience:**

Poeple purchasing the ahriculture machinery equipment who can explore dealers based on the proximity to their location and can can also explore the ratings of each dealer.

1. **Data Description:**

**Data requirements:**

Data requirements for this project require data specific Pune city. The data will be used from Wikipedia page with link given below.

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyses such large dataset and produce analytics to find a solution.

Few details of the data required are:

To access location of a dealer, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the dealers with its labels respectively.

Population of a neighborhood is very important factor in determining a dealer's growth and number of customers who turn up to service. Logically, the more the population of a neighborhood, the more people will be interested to walk openly into a dealer and less the population, a smaller number of people frequently visit a dealer. Also, if more people visit, better the dealer is rated because it is accessed by different people with different manufacturing companies. Hence is very important factor.

Income of a neighborhood is also very important factor as population was. Income is directly proportional to richness of a neighborhood. If people in a neighborhood earns more than an average income, then it is very much possible that they will spend more however not always true with very less probability.

**Data collection:**

1. Collecting geographical coordinates is not difficult but after googling for more than 2 days, it was not available on open source data websites such as Wikipedia, India gov website, census report websites etc. So I decided to use Google maps API to fetch latitude and longitude but google API has limited number of calls that I could make with my free account. So, it would take around 15 - 20 days to fetch location of all the neighborhoods.



2. Population by neighborhood is again easy to find out given that it’s readily available. I was able to find population data for few cities. Few neighborhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. The data frame for neighborhood population looks like:

3. Foursquare API:

Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (in my case: 500mts) and also corresponding coordinates, venue location and names. After calling, the following data frame is created:

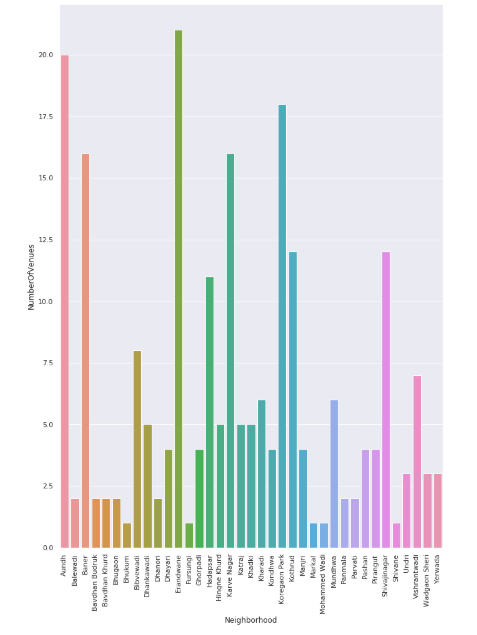


1. **Methodology:**

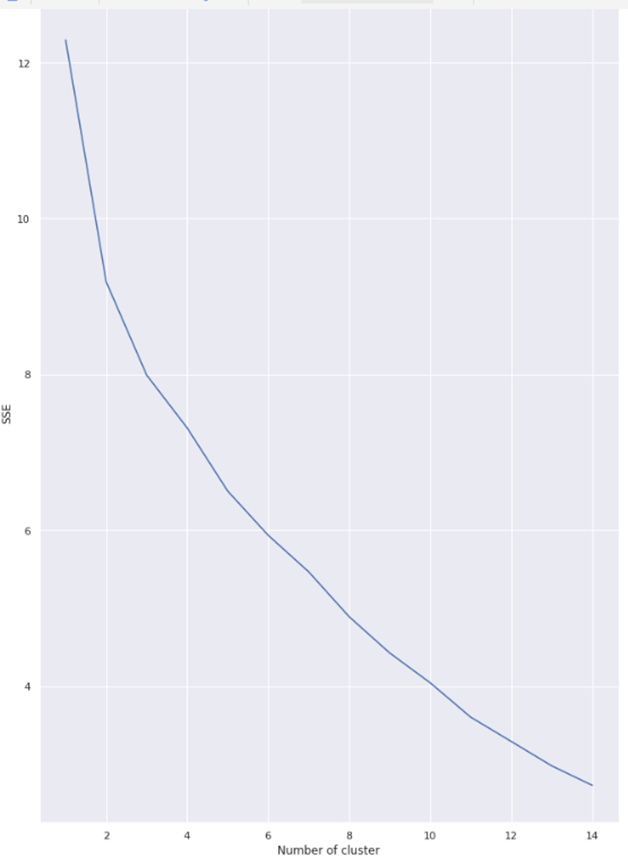
**Exploratory analysis:**

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.



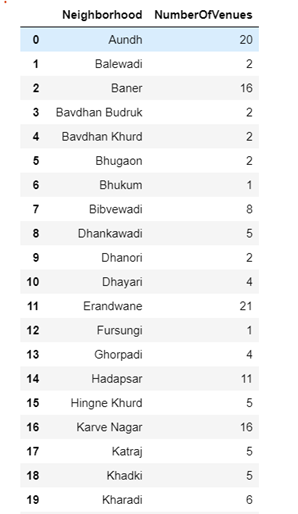
Also, while producing graph for number of clusters, I produced a graph to explore all the values for clusters and then finding the best by exploring the elbow graph.



1. **Result:**

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy.

The following image shows the result:

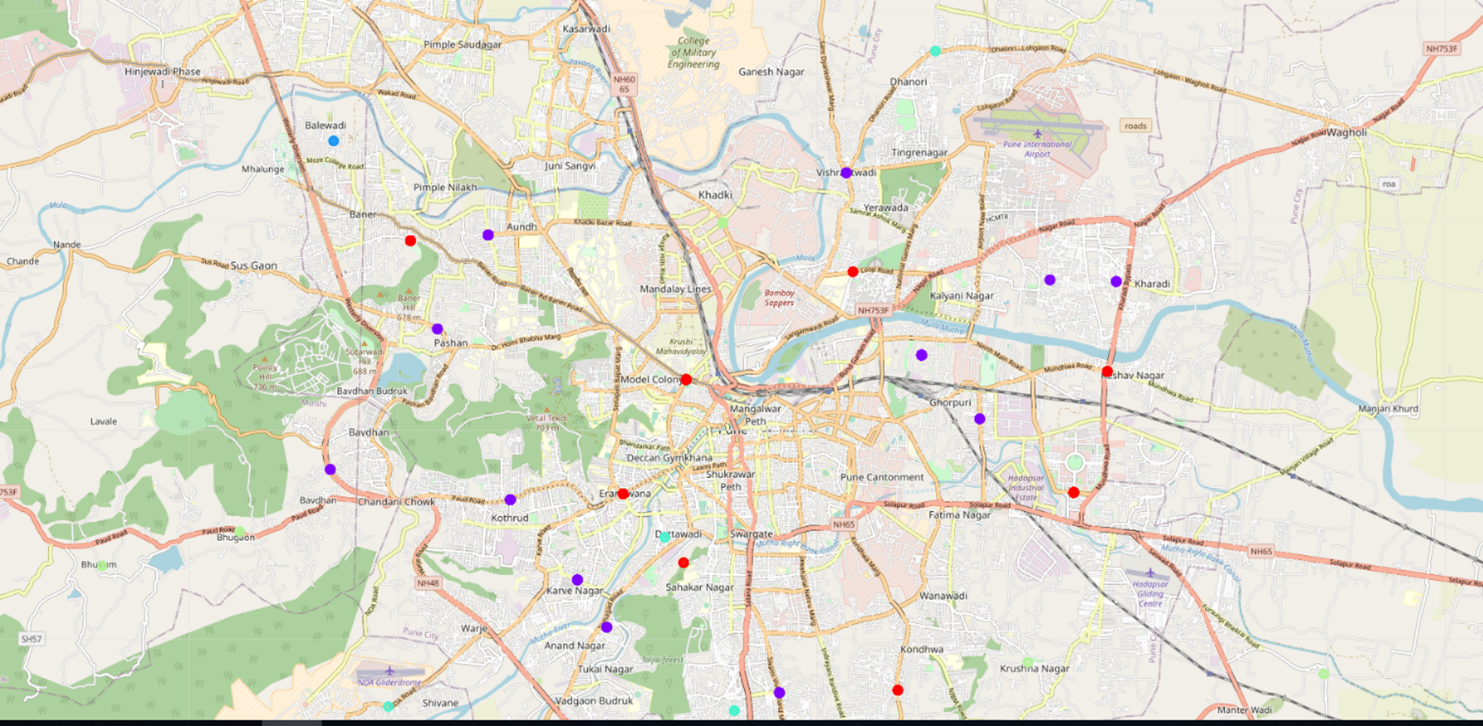




1. **Discussion:**

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also, during clustering, similar neighborhoods must be dumped into the right cluster.

The following graph shows the clusters:



Another observation that we can make is that choosing number of clustering could produce very diverse results. Some may be over fitted or some may be under fitted. Hence analysis of number of clusters must be done. Ref elbow graph in the Methodology section.

1. **Conclusion:**

The created guide approach considers factors such as population, income and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data, but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfil all their needs.

The given model considers finding veg restaurants with high visiting frequency as the dealer data was not readily available. But a similar exercise can be done for specific brand dealers to guide the buyers.