Unsupervised Learning for Vacation Planning

Aakarsh Yadav, Bachelors of Technology in Electrical Engineering

Manipal University Jaipur, Rajasthan

Author Note

Abstract

Planning a vacation is an energy and effort cumbersome task, it can be tiring and sometimes overwhelming. Apart from deciding the location, it is hard to decide what venues and in what order it should be visited. In this project, unsupervised machine learning method is used with KMeans clustering algorithm to plan and organize the destinations of a vacation.

*Keywords*:  Unsupervised machine learning, KMeans

## INTRODUCTION

In any part of the world, when someone finalizes the destination of their upcoming vacation it gives a lot of joy and excitement but brings along number of questions which requires a lot of effort to answer them. First would be the venues to be visited in that location, which can be easy if there are not much, but exhausting to look if there are a lot of them like in the city of Paris, New Delhi, New York, Hawaii and many other vacation hot spots around the word. One can be easily confused while looking for the venues and when to visit and in what order to visit. The background of this project is to organize the famous venues of the city of choice.

**PROBLEM**

**My friend is planning to go for a 5 days vacations in New Delhi, help her to find the venues she can visit and inn what order.**

**DATA**

Data required for this project is basically the data of the famous venues in New Delhi which can be acquired using Foursquare API. The data will be a single dataframe with the coordinates and category of the venues. New Delhi's coordinate has to be provided to the Foursquare API which can be acquired by the help of 'Geopy' library.

**DATA EXTRACTING**

Once we have the coordinates of the New Delhi, it can be provided to the Foursquare API along

with the Foursquare developer credentials and the data can be requested. This will provide all the

details about the nearby venues but our main focus is on the famous locations like monuments

and landmarks. This can be done by providing 'categoryID' of the monuments/landmarks found

in the 'Venue Category' section of the Foursquare website to the categoryID parameter while

requesting to the URL.

**DATA CLEANING**

The Foursquare API collects a large set of data but we will extract only the name and category of

the venue along with the coordinates. Example of the data is shown in table 1. As the project

aims at the basic solution, we can remove the datapoints which have missing coordinates.

**METHODOLOGY**

Folium library was used to map New Delhi with the help of its coordinates. Now a new dataframe X has to made which will only have the Latitude and Longitude of all the nearby venues. This dataframe will then be provided to the KMeans algorithm as it only takes numerical values.

KMeans clustering algorithm, on the basis of Euclidean's distance formula and calculates the distance between each datapoint from their randomly generated centers and updates them once they have their own clusters. The value of 'K', which is the number of clusters to be generated is the same number as the number of days of the vacation which is 5 in our case. Hence the venues are clustered in 5 groups.
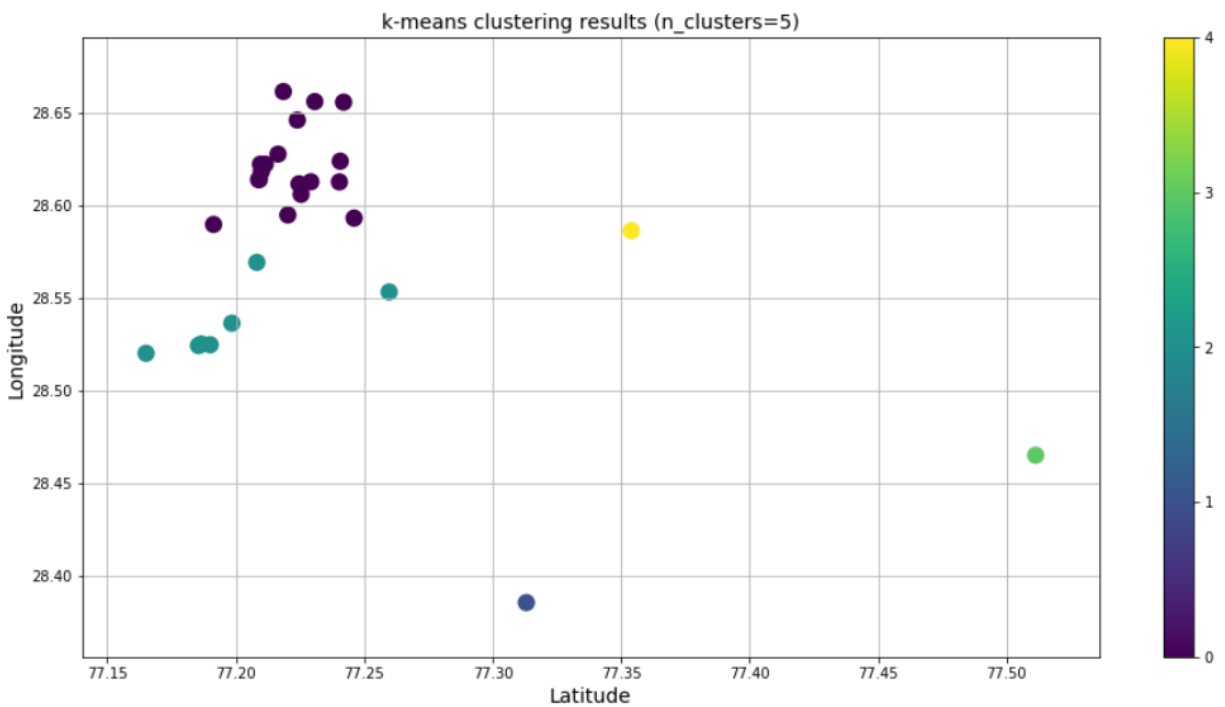
**RESULTS**

After the clustering of the dataframe X, it can be plotted with different colors for each cluster

with the help of matplotlib and then can be plotted on the New Delhi's map with the help of

folium.
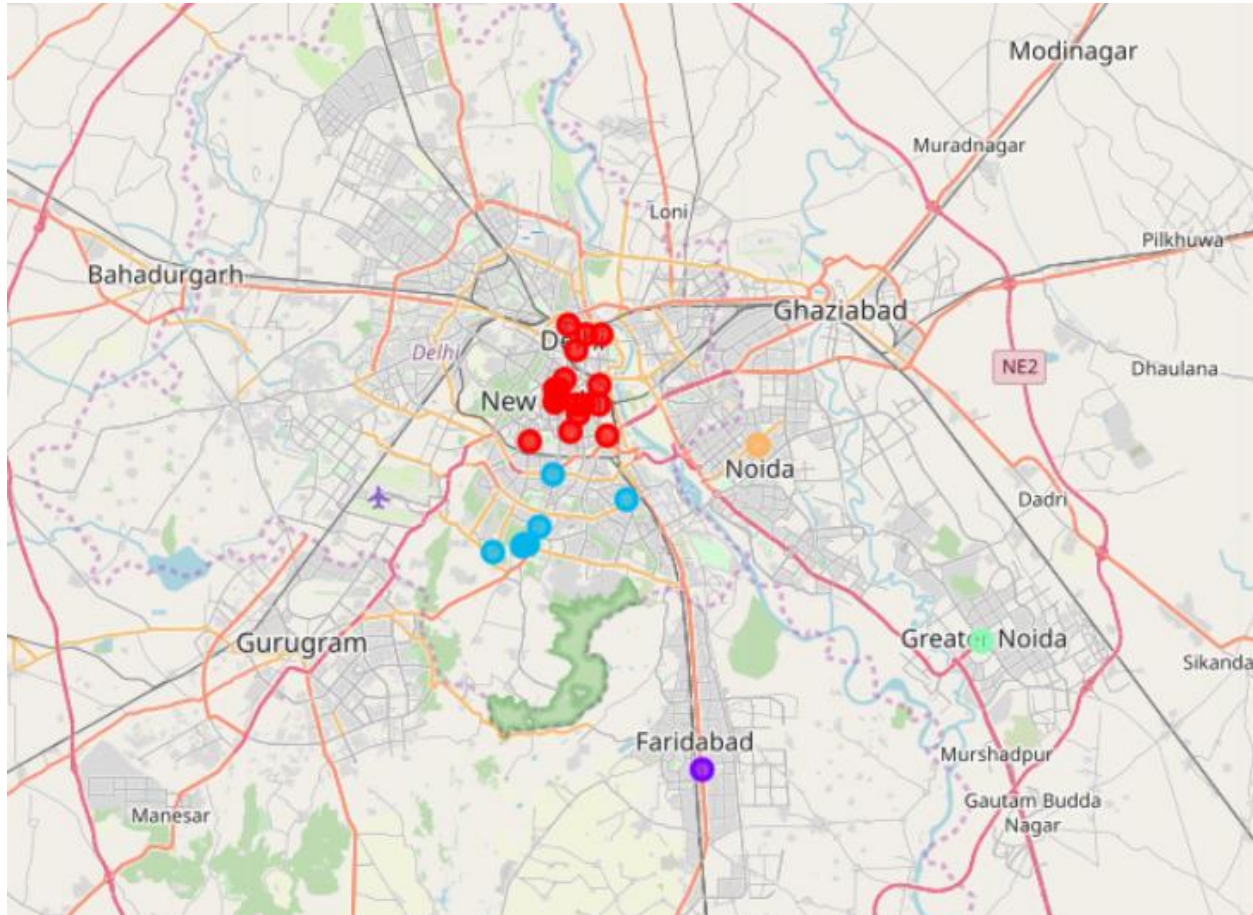


Furthermore, a final dataframe of the venues can be generated with number of days they should

be visited to make the travelling easy. The final dataframe can be seen in table 2.

Table 1

The first few rows of the data used in unsupervised machine learning

| Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|
| Rashtrapati Bhavan \| राष्ट्रपति भवन (Rashtrapa... | 28.614133 | 77.209193 | Monument / Landmark |
| Lok Sabha | 28.613957 | 77.208862 | Monument / Landmark |
| Qutub Minar \| कुतुब मीनार (Qutub Minar) | 28.525265 | 77.186593 | Monument / Landmark |
| Red Fort \| Lal Qila \| लाल क़िला \| لال قلعہ (Re... | 28.655759 | 77.241955 | Monument / Landmark |
| India Gate \| इंडिया गेट (India Gate) | 28.612796 | 77.229207 | Monument / Landmark |
| Lotus Temple (Bahá'í House of Worship) | 28.553349 | 77.259696 | Spiritual Center |
| Chandni Chowk \| चाँदनी चौक (Chandni Chowk) | 28.656076 | 77.230740 | Market |
| Humayun's Tomb \| हुमायूँ का मकबरा (Humayun's T... | 28.593197 | 77.246047 | Monument / Landmark |
| Jantar Mantar | 28.627731 | 77.216436 | Historic Site |

Table 2

The day 1 venues of the final dataframe

| Venue | Venue Category | Day |
|---|---|---|
| Rashtrapati Bhavan \| राष्ट्रपति भवन (Rashtrapa... | Monument / Landmark | 1 |
| Jantar Mantar | Historic Site | 1 |
| Ajmeri Gate | Monument / Landmark | 1 |
| Purana Quila | History Museum | 1 |
| Kota House Naval Officers Mess | Monument / Landmark | 1 |