THE BATTLE OF NEIGHBORHOODS
DATA SCIENCE
BY IBM-COURSERA

TOPIC: TOP WARDS IN TOKYO TO OPEN A RESTAURANT

INTRODUCTION

Tokyo is the most populous metropolitan area in the world. Currently ranked 3rd in the global economic power index, Tokyo is definitely one of the best places to start up a new business. During the daytime, especially in the morning and lunch hours, office areas provide huge opportunities for restaurants. Shops are usually always full during the lunch hours . Thus opening a breakfast cum lunch restaurant in highly dense office places can be of great interest to generate a good profit. Usually the profit margin for a decent restaurant lie within 15–20% range but, it can even go high enough to 35%.

TECHNOLOGIES USED

MACHINE LEARNING

PYTHON

IMPLEMENTATION

Our project can be implemented using various ways but for simplicity we have here used the python concepts and libraries defined for machine learning algorithms such as clustering, classification, etc.

Here we have used python as it is an easy to code language and mostly used for data science projects to work with because of its versatile nature.

We have used unsupervised machine learning technique - Clustering to implement the project and the algorithm used is the famous k-means clustering algorithm.

Clustering Algorithm Used:

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. There are various algorithms to implement clustering but the most commonly used is k-means and the same has been used in the project work as well.

k-means Algorithm Used:

k-Means is probably the most well-known clustering algorithm used.

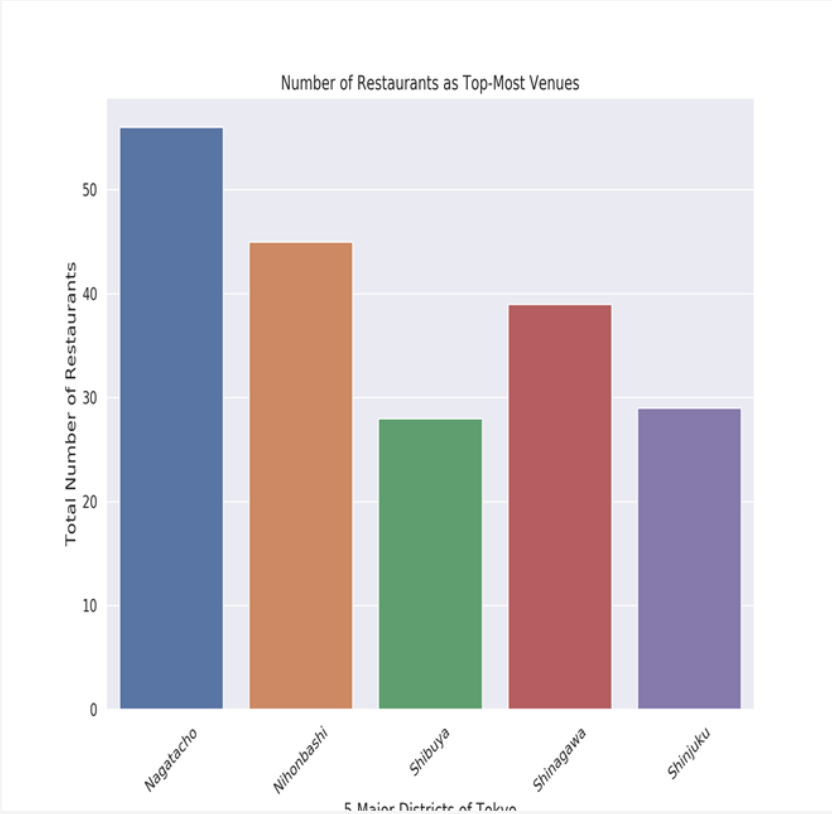
To begin, we first select a number of classes/groups to use and randomly initialize their respective center points.

Each data point is classified by computing the distance between that point and each group center, and then classifying the point to be in the group whose center is closest to it.

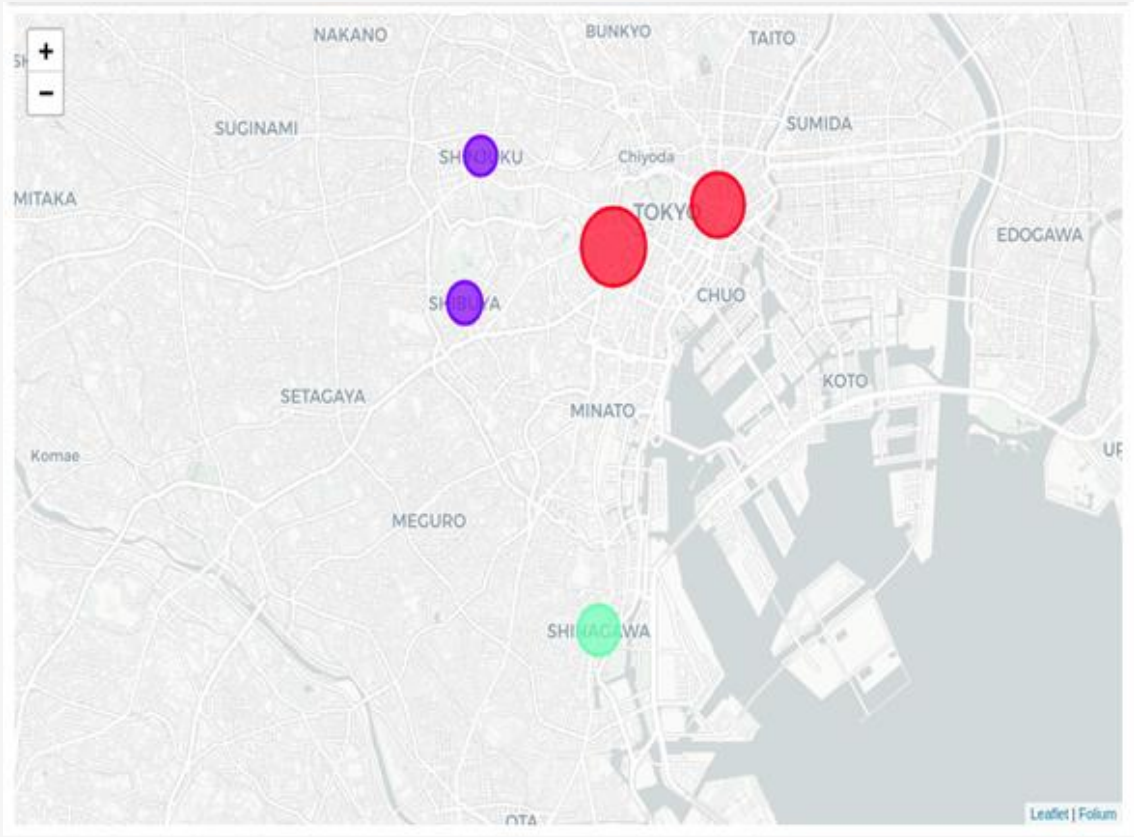
We repeat the process for a set number of iterations or until the group centers don’t change much between iterations.

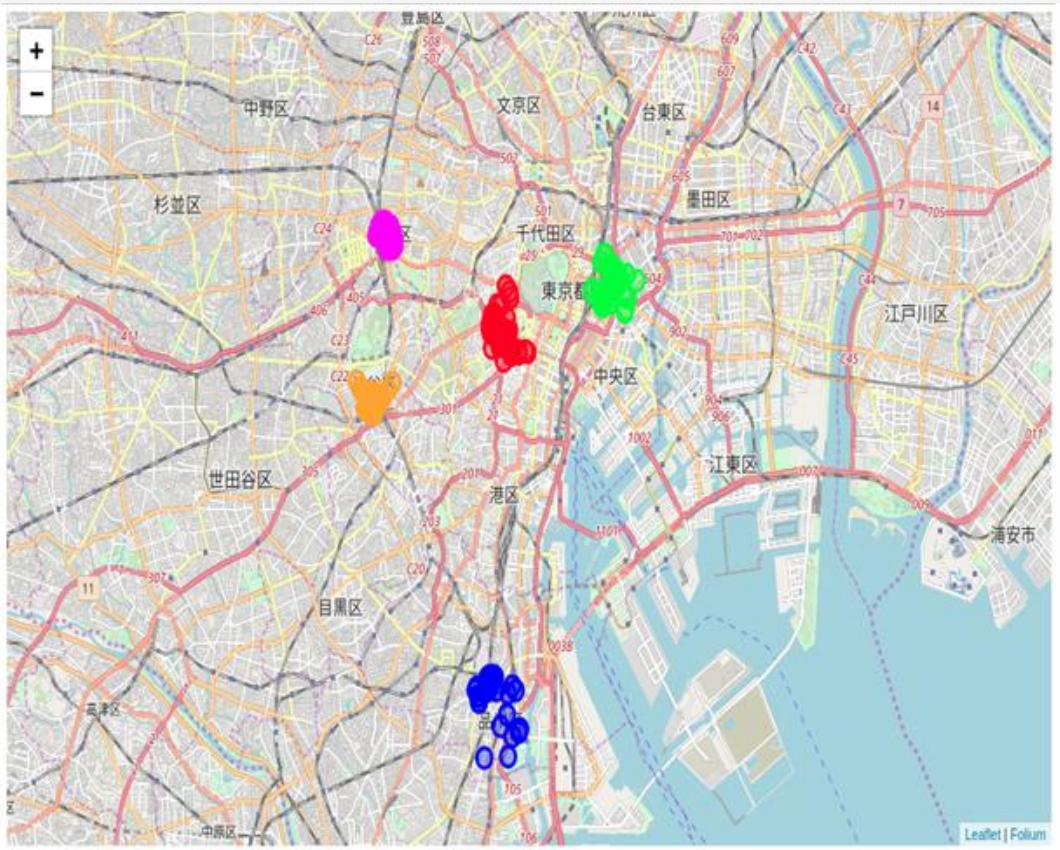
RESULTS

We successfully got the 5 major wards of Tokyo and, as the business problem started with benefits and drawbacks of opening a lunch restaurant in one of the busiest districts, the data exploration in most case gave accurate results. I used data from web resources like Wikipedia, python libraries like Geopy, and Foursquare API, to set up a very realistic data-analysis scenario.



OUTPUTS


NAKANO
BUNKYO
TAITO
SUGINAMI
SUMIDA
Chiyoda
MITAKA
EDOGAWA
CHUO
SETAGAYA
MINATO
KOTO
Komae
MEGURO
Leaflet | Folium


中野区
文京区
台東区
杉並区
墨田区
千代田区
江戸川区
中央区
世田谷区
港区
江東区
目黒区
Leaflet | Folium

CONCLUSION

Finally, to conclude this project, we have got a small glimpse of how real-life data-science projects look like. I've made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map.

With the help of appropriate tools like clustering, k-means and predefined libraries and functionalities it was fun to work on this project and generate the desired outputs.

THANK YOU