ANALYZING NEIGHBOURHOODS OF KUALA LUMPUR FOR STARTING A NEW BAKERY

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT BY
N. VIGNESWARAN

DECEMBER 2020

INTRODUCTION

Kuala Lumpur, or KL, is the federal capital and most populous city of Malaysia. Kuala Lumpur is the fastest growing metropolitan area of the country. It's the economic, financial and cultural city of the country and one of three Malaysian Federal Territories. Greater Kuala Lumpur, or the Klang Valley, is a fundamental urban aggregation with an approximation population of 7.2 million in 2016 with a population density that is nearly equal to that of the city proper. It's 2020 population is now estimated at 7,996,830.

The city is mainly a mix of Malays, Chinese, and Indians, although there are many cultures in the city such as Eurasians, Kadazans, Ibans and indigenous peoples from East Malaysia and Peninsular Malaysia. The diversity among the population create a path for variety of foods. A report notes that the Malaysian market has seen positive growth of baked goods as more products are consumed at meal times in place of rice and noodles. So, in this project, we will study the neighbourhood and make recommendation accordingly.

BUSINESS PROBLEM.

The main objective of this project is to analyze the neighbourhood of Kuala Lumpur to select a best location for a opening of bakery. Bakeries can categorize as conventional stand-alone bakeries, commercial bakeries, bakery counters, and home-based bakers providing a handmade breads and cakes. So, we would like to extract and analyse a various data about neighbourhood of Kuala Lumpur using data science methodology and machine learning techniques such as clustering to provide a right solution.

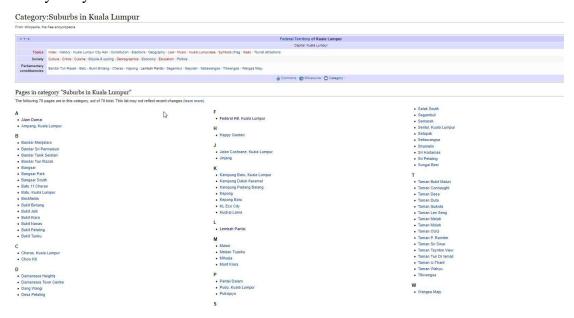
DATA

To solve the problem, we will need the following data:

- List of Kuala Lumpur neighbourhood
- Geographical coordinates of neighbourhood
- Venue data from Foursquare

Neighbourhood data

The data extracted from Wikipedia page using web scrapping with BeautifulSoup library in Python.



Geographical coordinates of neighbourhood

Using Geopy geocoders, the geographical coordinates of various suburbs in Kuala Lumpur were extracted. Additionally, two columns were included into data frame for latitude and longitude.



70 rows × 3 columns

Venue data from Foursquare

Foursquare is a database of more than 105 million places worldwide and an API that enables location data for Apple, Samsung, Microsoft, Tencent, Snapchat, Twitter, Uber, and others. The venue data extracted using Foursquare API to study the various neighbourhood in Kuala Lumpur. The Data provides various locations of existing bakeries around the neighbourhood and its vital to analyse and draw conclusion of this project.