

**Coursera Capstone**  
**IBM Applied Data Science Specialization**

**Analysis of Opening a New Shopping Mall in Bali, Indonesia**

**Nelson Agus Kesuma**

## Introduction

Bali is a heaven for nature enthusiasts and party goers to enjoy and pamper themselves with the wondrous of natures and beaches. This creates a great place for tourism industry to bloom, including resorts, hotels, restaurants, clubs and shopping malls. For sellers and business owners, this is a great place to start businesses to cater for a high demand. There is a need to build more shopping malls to as traffic increases and poses a good business opportunity to build convenient shopping malls for tourists to enjoy not only nature but also shopping in Bali

## Business Problem

The objective of this capstone project is to analyse and select best locations in Bali, to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to recommend a good location to build a new shopping mall

## Data

The following data is needed:

- List of neighbourhoods in Bali.
- Latitude and Longitude coordinates of these neighbourhoods.
- Venue data related to shopping malls and cluster neighbourhoods

## Target Audience

Targeted for property developers and business owners who are interested in creating new business or expanding portfolios.

## Data Source

Wikipedia page ([https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Bali](https://en.wikipedia.org/wiki/List_of_districts_of_Bali)) will be able to provide a list of neighbourhoods in Bali. Usage of web scraping to extract data from the Wikipedia page with the help of Python library and BeautifulSoup packages are required. Then geographical coordinates of the neighbourhoods using Python Geocoder will provide latitude and longitude coordinates of the neighbourhoods.

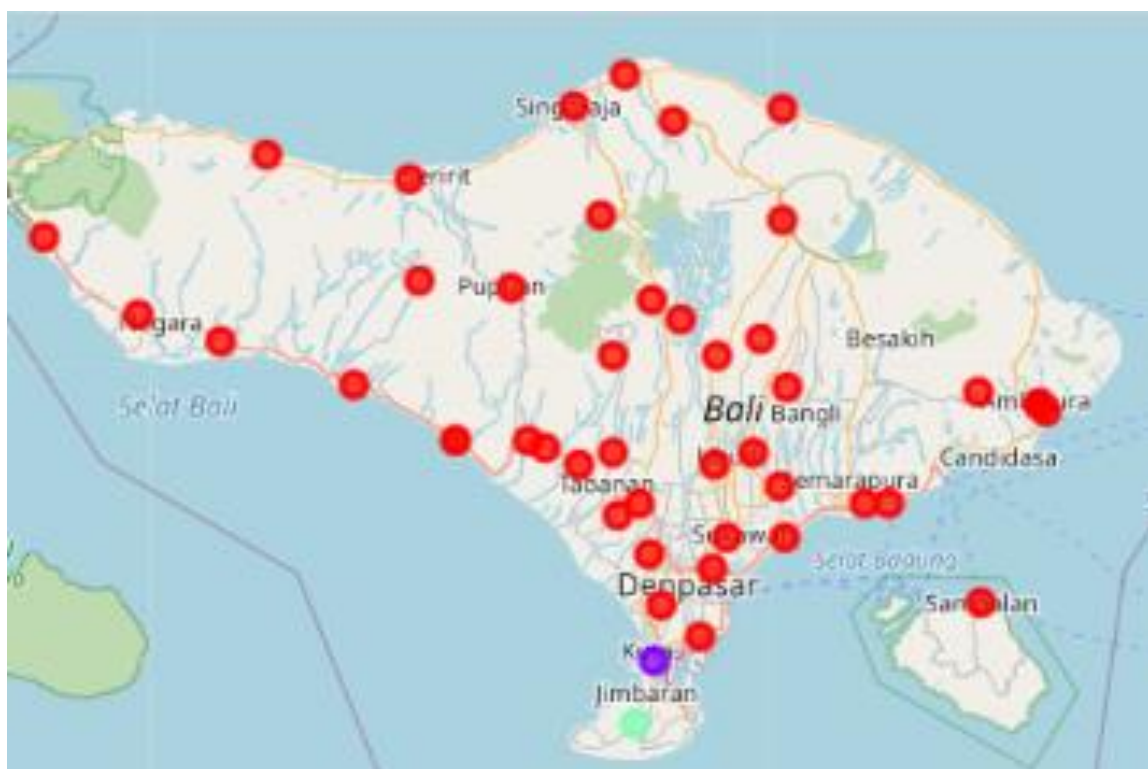
Afterwards, Foursquare API to get the venue data can be used to fetch the real map, particularly in shopping mall category. This will train skills in web scraping, working with API from Foursquare, data cleaning, data wrangling, machine learning (K-clustering) and map visualization (Folium)

## Results

The results from the k-means clustering show that it is possible to categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Shopping mall:”

- Cluster 0: Neighbourhoods with low number to no existence of shopping malls
- Cluster 1: Neighbourhoods with moderate number of shopping malls
- Cluster 2: Neighbourhoods with higher concentration of shopping malls

The results of the clustering maps are visualized in the map below in which cluster 0 is red dots, cluster 1 is purple dot and cluster 2 is green dot.



## Discussions

With the map shown above, observed shopping malls in Bali are highly concentrated on the west side around Kuta and Jimbaran areas in which majority of the tourist attractions are located (marked with purple and green dots) . This has provided convenience for tourists and boosted the Bali economy. However, large areas of Bali marked with red dots still have very little number of shopping malls. Therefore, there is a chance of undersupply to consumers and less competition required if property developers are interested in opening a mall. Property developers with unique selling propositions and features will be able to stand out at cluster 0 with little competitions. With the large populations in Bali which is around 4 million residents, there will be demands from people to improve their quality of lives.

As for cluster 1 and cluster 2, there are way too many shopping malls built catered to tourists. With the coronavirus happening, tourists will be less and building a new shopping mall there is not recommended as there has high competitions, less tourists and lesser demands.

## **Limitations and Suggestions for Future Research**

In this project, there is only one factor considered which is the frequency of shopping malls occurrence. Hence, there is no factor of income level in each region or number of residents in each region. In reality, in order to build a shopping mall, more factors have to be considered (i.e. land availability, sizes, ROI, population, income level, etc). As this project was made with free Sandbox Tier Account of Foursquare API that came with the limitations of API calls, limited amount of data is available but not limited to findings. Future projects could make use of paid account to obtain more results

## **Conclusion**

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing data, performing machine learning by clustering the data into 3 clusters based on their similarities and lastly providing recommendations to relevant stakeholders to open a new shopping mall. The findings of this project will help the relevant stakeholders i.e. property developers and investors to capitalize on opportunities on high Return of Investment (ROI) locations which is cluster 0 with little competition but high demand from the locals and opportunities to stand out.