# Segmenting and Clustering Neighborhoods in Bali, Indonesia

Fathan Mubina

March 1, 2021

## 1. Introduction

### 1.1 Background

Bali is an administrative province in Indonesia, located in the east of Java island and west of Lombok Province. Situated around the Indian Ocean, the island of Bali is well-known as a popular tourist destination, attracting people from across the world for its breathtaking nature and culture. Covering an area of 2,230 square miles, Bali always has something to offer for the tourists in every part of the island, from south to north. However, the tourism development is primarily focused on the southern part of the island where modern lifestyle is growing rapidly while traditional culture is still being maintained. In the southern development, restaurants of various cuisines from across the world, such as Mexican, Japanese, Italian, Thai, and fusion, can be found across the area. Hence, it is advantageous for restaurant owners to identify which area that becomes the food and beverage (F&B) destination in Bali if they want to open a new restaurant.

### 1.2 Problem

Analysis in this project might be valuable for restaurant owners in identifying areas in Bali that are potential for them to open a new restaurant based on area development, location, nearby tourist destination, and popular venues. The project aims to identify areas in Bali suitable for developing a new restaurant by segmenting and clustering sub-districts or villages from available data.

### 1.3 Interest

Restaurant and business owners that are looking to open new F&B businesses in Bali would be interested in the identification of potential areas that attract the diverse customers visiting the island. Travel businesses and individual travelers might also be interested in the analysis as it might be beneficial for them to prepare their travel itinerary.

## 2. Data Acquisition and Cleaning

### 2.1 Data Sources

The project segments, clusters, and identifies potential areas in Bali by the sub-district or village administrative level. The dataset used in the project is acquired and scraped from a website containing a list of postal codes of sub-districts or villages across the province of Bali that can be found in this link. Additionally, to retrieve the coordinates data of each sub-district or village, the project uses the *geocoder* package in Python. To identify the popular venues in each sub-district or village, the project also uses Foursquare API to gain insights on each area based on user data.

### 2.2 Data Wrangling

The dataset gathered from website scraping contains 714 entries with features including postal code, sub-district/village, district, regency/city, and province. A pandas dataframe is made to contain selected features, such as regency/city, district, sub-district/village, and postal code. After the dataset is imported, the latitude and longitude data of each sub-district or village is retrieved using the *geocoder* package in Python.

However, because the tourism development is more focused on the southern part of the island, the project focuses on identifying potential areas for F&B businesses in southern Bali, particularly in Badung Regency. Situated near the ocean, Badung Regency is home to popular tourist areas where many beaches, beach clubs, restaurants, coffee shops, and shopping malls are located. After filtering the dataset that only includes areas in Badung Regency, there are 62 sub-districts and villages in the new dataset. The new dataset covers district, sub-district/village, latitude, and longitude data.

Then, after the dataset is ready, the project utilizes Foursquare API to retrieve data of the venues within the Badung regency. The venues are then categorized based on their business type and the categories are sorted based on the most occurrences in the area, representing the popular venues in the area. The venue data, containing frequency of occurrence, categories, and sub-district/village, are then used in the machine learning model that segments and clusters the area, suitable for business owners to open new businesses.