

Predicting the Best Location to Build a Hotel in Bali

1. Introduction

1.1 Background

Bali is one of 34 provinces in Indonesia which is famous for its beautiful beaches. It is not surprising that Bali becomes the most visited destination in Indonesia by tourists, either domestic or foreign tourists. Its beautiful nature, especially the beaches, will attract anyone seeing it. Many tourists stay there for a long time to enjoy the nature. Even, there are some rich people intentionally purchasing a house in Bali to provide their accommodation when they and their families take a holiday in Bali. From the above description, it can be assumed that the most needed facilities in a tourist destination area is a place for staying. This project will explore some of the districts in Bali, specifically Southern Bali which have many tourist attraction areas, to predict the best location to build a new hotel to provide accommodation for tourists.

1.2 Business Problem

As a tourist destination, of course, there are already a lot of hotel in Bali. Therefore, this project will explore districts in Southern Bali and try to predict the optimal location to build a new hotel.

1.3 Audience Target

This project specifically will be targeted for stakeholders interested in opening a hotel in Southern Bali, Indonesia.

2. Data

2.1 Data Source and Description

Data which is be required for this exploration is a list of districts in Bali province, Indonesia. It can be found in Wikipedia from the following link: https://en.wikipedia.org/wiki/Kecamatans_of_Bali. Kecamatan is the Indonesian language means district. Wikipedia articles can be written by anyone, so it could be mistaken happened. Therefore, I use a PDF file from Indonesia's government website for data verification. Since I am an Indonesian native speaker, so it is easy for me to verify the data which comes from the local language. The file for verification can be found in the following link: <https://www.kemendagri.go.id/files/2019-05/Kode&Data%20Wilayah/51.bali.fix.pdf>.

2.2 Data Preparation

First, I do web scraping to acquire a list of districts in Bali using BeautifulSoup package in Python and save it into a CSV file. Once CSV file have been created, the next step is to read that CSV file into pandas dataframe. Data displayed in the dataframe still does not in appropriate form. The name of districts

indexed as column names, so I perform dataframe transposing to change the name of districts as rows and rename that single column as District.

The next step is to split the string of the columns in the dataframe. For example, the original form of string in the first index is *Abang, Karangasem*, which is *Abang* is a name of district and *Karangasem* is a name of regency. I split the district and regency name to make it clear which one is the districts and which one is the regencies. I put the regencies in a new column named Regency.

Next, data verification is performed to make sure that I use the proper data for exploring. I review the generated data from web scraping to be compared to the PDF file data from Indonesia's government website. As a result, I found some mismatches between the two data such as a mismatch between district and regency, a missed district, a should-not-be-there district, and some typo cases. Then, I clean and fix those problems and make sure that the dataset for exploring is already correct and resave it into a new CSV file contained 57 districts of Bali.

Since I am going to explore only the Southern Bali province, I shorten the dataset into 9 rows containing the districts in the Southern Bali province, included 8 districts in Bali island and 1 district in Nusa Penida island. For exploring the districts, I am going to use Foursquare location. Hence, the latitude and longitude coordinates are required for this exploration. To obtain the coordinates, I use Geocoder package in Python and put them into the two new columns named Latitude and Longitude. It causes some new problems. Not all of the generated-coordinates generated the proper values for each district. For example, for Kediri district in Tabanan regency, the coordinates refer to Kediri regency in East Java province, as you can see in the following figure.

```
[19]: geolocator = Nominatim(user_agent="my_explorer")

df['Latitude'] = df['District'].apply(geolocator.geocode, timeout=20).apply(lambda x: x.latitude)
df['Longitude'] = df['District'].apply(geolocator.geocode, timeout=20).apply(lambda x: x.longitude)
df
```

:[19]:

	District	Regency	Latitude	Longitude
0	Denpasar Barat	Denpasar	-8.662244	115.198067
1	Denpasar Selatan	Denpasar	-8.706134	115.225676
2	Denpasar Timur	Denpasar	-8.632914	115.246211
3	Denpasar Utara	Denpasar	-8.624513	115.209009
4	Kediri	Tabanan	-7.805898	112.168507
5	Kuta	Badung	-8.727835	115.173005
6	Kuta Selatan	Badung	-8.808849	115.169528
7	Kuta Utara	Badung	5.067268	97.041101
8	Nusa Penida	Klungkung	-8.745652	115.534745

Figure 1. Codes and outcome of generating Latitude and Longitude coordinates using Geolocator

Because of that, to fix this problem, in another cell in my notebook, I try to generate coordinates for Kediri with a clear address. So does for another district incorrectly produces coordinates, Kuta Utara. Then, replace the wrong values in the dataframe by the two new generated coordinates for each district.

```
[22]: df.loc[df['District'] == 'Kediri', ['Latitude', 'Longitude']] = [kediri_latitude, kediri_longitude]
df.loc[df['District'] == 'Kuta Utara', ['Latitude', 'Longitude']] = [kuta_utara_latitude, kuta_utara_longitude]
df
```

```
:[22]:
```

	District	Regency	Latitude	Longitude
0	Denpasar Barat	Denpasar	-8.662244	115.198067
1	Denpasar Selatan	Denpasar	-8.706134	115.225676
2	Denpasar Timur	Denpasar	-8.632914	115.246211
3	Denpasar Utara	Denpasar	-8.624513	115.209009
4	Kediri	Tabanan	-8.564455	115.132072
5	Kuta	Badung	-8.727835	115.173005
6	Kuta Selatan	Badung	-8.808849	115.169528
7	Kuta Utara	Badung	-8.678042	115.169861
8	Nusa Penida	Klungkung	-8.745652	115.534745

Figure 2. Codes and outcome of generating after fixing the Latitude and Longitude coordinates

Data preparation have been done. As you can see above, the coordinates of Kediri and Kuta Utara have changed. To make it easy for viewing the location of the districts, the map of Bali created using Folium package as you can see in the below figures.

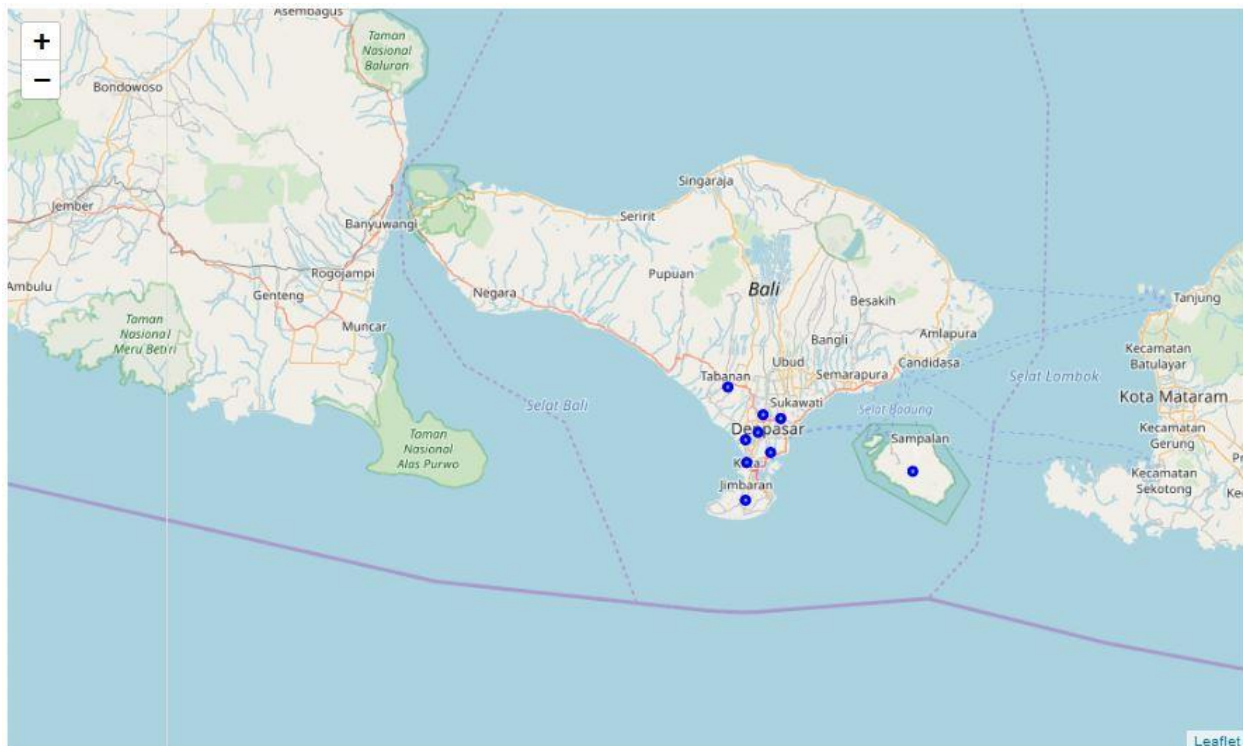


Figure 3. Generated map of Bali with mark of each district created using Folium

The next step is to explore and analyze the dataset using Foursquare as well as clustering the districts to obtain information for solving the problem. I am going to complete this report next week when I will have completed my exploration.