

Optimizing Where to Establish a New Hotel on the Most Tourist-Visited Philippine Provinces as Analyzed Through Foursquare API Location Data Using Unsupervised Machine Learning Algorithm

Sandy C. Lauguico

July 2020

Coursera Applied Data Science Capstone Project



Image Source: <https://www.thecrazytourist.com/25-best-islands-visit-philippines/>

Introduction

Having the Department of Tourism's (DOT) slogan, "It's more fun in the Philippines," the country's tourism is one of the growing sectors that significantly contribute in the economy. According to several articles (Erika Hueneke, 2019; Amy Thomas, 2019; Karen Hastings, 2020), the Philippines has some of the most beautiful islands in the world, which attracted majority of the 8,211,535 foreign tourists in different provinces last 2018 as reported by DOT. The Philippines as an archipelago has a lot of geographical and natural settings that can offer attractions, fun activities, leisure, and relaxation to accommodate those who are wandering.

However, having a good vacation is not just about the experiences from those beautiful sceneries. One of the major factors that needs to be considered are the hotels where the guests are going to stay. Hotel experiences can make or break the whole vacation. It is very important that hotels also offer the services which could best accommodate the tourists to enhance and to promote tourism further. Without considering this, a relaxing getaway can turn into a stressful experience which could rate down the country's tourism.

Business Problem

The current construction industry that is working together with the DOT is possibly relying on different inferences on where to best place a new hotel. This may result to potential investment losses and inefficient use of resources. A data-driven analysis would significantly provide a decision that can maximize profit while promoting tourism. The objective of this project is to be able to effectively locate the best provinces where to establish new hotels based on tourism data, existing hotels, and hotel customer-based clustered experiences. This optimal approach may help in decreasing the risks of losses and increasing the possibilities of win-win arrangement for businesses, communities, and customers. The system will also be able to assist the construction industry in determining what type of hotel (from 1- to 7-star hotels) can be established to offer variety in prices and services.

Data

In achieving the objectives, the following data are necessary to accomplish the analysis:

- List of the Philippine provinces
- The provinces' corresponding coordinates in terms of latitude and longitude
- The total number of tourists visiting a province

- Venues within the provinces, specifically the hotel category.

The list of provinces was obtained from a Wikipedia page through scraping using BeautifulSoup (https://en.wikipedia.org/wiki/List_of_Philippine_provinces_by_population). This dataset lists originally each of the provinces and the population residing in the area. It contains all the 81 provinces and the key cities of each province.

To get the coordinates of the provinces in terms of latitude and longitude, the geocoder library was used. The coordinates were listed and merge into a dataframe with the names of the provinces.

A separate dataset was used to get the total number of tourists from domestic, overseas, and international guests visiting a certain province in 2018. The data was from the Philippines' Department of Tourism and it is recorded and presented in a pdf format. This results to manual integrating of the 'number of tourists' column to the dataframe which consists the provinces and coordinates.

Foursquare API was then used to get the venues in all of the provinces to be extracted in the 'Hotel' venue category.