Final Report
IBM Data Science Capston Project
Analysing the best location to open a cafe in Bali

MICKAEL GRATIA
JULY 2020

Introduction

Description of the problem

Starting a business in Bali combines two of every entrepreneur's greatest desires: living in paradise while making a steady profit. The number of businesses springing up in Bali over the last few years has turned this island paradise into something more lucrative than just another Southeast Asian holiday destination.

Starting a business in Bali by opening a coffee shop is a good choice because of its simplicity and the handsome profit that will come with it. Plus, who doesn't like coffee? You can do this in major parts of the island. Good news is, opening a coffee shop in Bali will not require a huge start-up capital.

We know many foreigners planning on opening a coffee shop in Bali have no idea how and where to start at first. We will skip the "how" part for thés Capstone project and I propose we focus on the "where".

To help entrepreneurs in their choice of location, we will leverage the Foursquare location data and define which district is worth considering. The main indicator will be competition, going to a place where the market is not already submerged with coffee shops.

Data and its use

Bali island is divided into several districts. Bali can be described as a mix from the bustling streets and white-sand beaches of Kuta, Legian and Seminyak to the upscale resorts of Nusa Dua, the cultural charms of Ubud and the cool highlands of Bedugul. Each district has specific attractions, some are also being developed. So for this project we will need the following data:

- **Bali data on its districts**
  - Data source: Wikipedia https://en.wikipedia.org/wiki/Category:Districts_of_Bali
  - Description: this data set contains the list of districts composing Bali. It will be used to explore the various businesses

- **Coordinates:** Latitude and longitude of each district.
  - Data source: Python geocoded package
  - Description: this data set will be required to be able to plot each district onto a map

- **Venue data**
  - Data source: Foursquare API
  - Description: this data set will be obtained via an API to collect all the venues in each district. It will also be filtered on coffee shops.