Where should we open a new Café in Gold Coast, Queensland, Australia?

Truc Thanh Thai

Introduction

Gold Coast is a coastal city which is about 66 kilometers from Brisbane, the capital city of Queensland, Australia. The city is the sixth-largest city in Austraia and the largest non-capital city in Australia. Apart from 70 kilometers of coastline with famous beaches, this city is also a major tourist destination with a variety of entertainment sites such as skyline, theme park.

According to Wikipedia, there are 80 suburbs in Gold Coast. In 2016, the population in Gold Coast is 569,997 and there are around 10 million tourists visiting the city every year. However, to the best of our knowledge, there are very few Café shop in Gold Coast. The lack of Café shop may make Gold Coast inconvenient for the local citizen and also for tourists, especially in festival, holiday time during the year.

In this paper, I would like to applied knowledge, skills gained from the 'Applied Data Capstone' course to advise the place where Café shop can be opened. The solution presented here takes into account the density of the population. Therefore, this paper may be of interest of a wide range of readers including local and international investors, tourist company, and those in coffee industry. The code provided in Github repository may be of interest of python, machine learning leaners.

Methods

Data source

1. List of suburbs in Gold Coast and its population from Wikipedia (https://en.wikipedia.org/wiki/List_of_Gold_Coast_suburbs#Suburbs)
2. Coordinates of each suburb using 'geopy' function
3. Background map of Gold Coast in the shape of GeoJSON from Australia Government Website (https://data.gov.au/dataset/ds-dga-6bedcb55-1b1f-457b-b092-58e88952e9f0/distribution/dist-dga-d20d0a54-7680-43c4-8c46-a08e3bc43fa0/details?q=queensland%20suburb)
4. Information about each suburb using FourSquare (https://api.foursquare.com/v2/venues/explore)

Data processing

List of suburbs: because there is lots of information from Wikipedia website, we only select data needed for this paper. From a table with several information about each suburb in Gold Coast including Name, Type, Postcode, Population 2016 cencus, Population 2011 census, Population Change etc…, we will select only the name, postcode and 2016 population for each suburb. All supperscript letters such as [a] from the table should be removed to have clean data. The data from the table will be transformed into dataframe using 'pandas' library.

Coordinates of suburbs: the name of each suburb will be provided to 'Nominatin' library through 'geocode' function. The returned coordinates (i.e. latitude and longitude) will be added to the datframe (i.e. the list of suburbs above). For example, the coordinates of 'Austinville, Queensland, Australia' are -28.1331291, 153.3158014.

Background map: From Australia Government, the geographical data of each suburb is stored in a shapfefile with all suburbs in Queensland. We use a free, open-source QGIS software to load this shapefile and extract only suburbs in Gold Coast based on the name and postal code. The extracted shapefile will then be exported to geoJSON file. For other peers t be able to access this file in case they want to re-run the script, we will upload this geoJSON file into our website and use python script to load.

Information about suburb: After declaring credential to access Foursquare, we pass the coordinates of each suburb to Foursquare to retreive venues around each suburb with a radius of 500 meters.

Data analysis

All data analysis will be carried out using Jupyter Notebook within IBM Watson Studio platform. Data from Wikipedia will be read using the 'read_html' function in 'pandas' library. Coordinates of suburbs will be retrieved using 'geocode' function in 'Nominatim' library. Data retreived from Foursquare will be explored and the mean of the frequency of occurrence of each category will be calculated. The data about Café will then be extracted and used in fitting model. We use k-mean clustering method to classify suburbs in Gold Coast into three clusters based on the occurrence of Café. The result will be presented in folium map with the background indicating population density and markers indicating suburbs' name and postal code.