Capstone Project - The Battle of Neighborhoods

Finding the most populated US cities with highest number and density of Peruvian restaurants

1-Description of the problem and discussion of the background

Scenario

A businessman in the importing and delivery industry has noticed how difficult it is for international cuisine restaurant owners to get specific ingredients, that are only available in their native countries; those groceries and fresh produce are essential to provide customers with authentic recipes and dishes.

Business Problem

The businessman is a Peruvian food aficionado and he is interested in importing those special ingredients to Peruvian restaurants in US; however, he needs some help in deciding what city is the best option for his business.

He is interested in high density of restaurants within the selected city, so he can get the most revenue.

I have been given the exciting task of assisting him to make data-driven decisions on what cities are suitable for his needs. This will be a major part of his decision-making process.

2-A description of the data and how it will be used to solve the problem

Why using data?

Without leveraging data to make decisions about this new enterprise, my customer could spend countless hours walking around spending precious time and efforts and ending choosing a city that is not the best option.

Data will provide better answers and better solutions to this task at hand.

How the data will be used to solve the problem?

I will concentrate in finding the top 5 ranked cities in US by population, and using an API to get the number of Peruvian restaurants in those cities; then analyze that data to get the mean coordinates and the mean distances to mean coordinate(MDMC) for its restaurants, to calculate density and display the findings in map charts.

The best city for the goods delivery business will be the one with a combination of highest number of restaurants and at the same time the lowest mean distances to mean coordinate average of the restaurants.

Data sources and methodology description

Top 5 populated cities in US will be retrieved from Wikipedia:

- Source: https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

Foursquare API will be used to get restaurant data within those cities:

- Source: https://foursquare.com/developers/

Initial Data Wrangling and Cleaning:

- Data will be cleaned and converted to a useable form as a dataframe.

Data Analysis and Location Data:

- Foursquare location data will be leveraged to explore Peruvian restaurants in each city.
- Data will be manipulated and analyzed to derive subsets of the initial data.
- Statistical analysis will be performed to locate mean coordinates and respective restaurant mean distances to mean coordinate calculation.

Data Visualization:

- Analysis and plotting visualizations.
- Data will be visualized by using Folium.

Interested Audience

I believe this is a relevant challenge with valid questions for anyone seeking business opportunities for any region, not only for a food delivery market, but for selecting store locations, place to live or work, as the same methodology can be applied in accordance to demands as applicable. Lastly, this project can also serve as a good practical exercise to develop Data Science skills.

Success criteria

The success criteria of the project will be relying in a combination of data selection and data science skillset to provide the best possible based on the correct data manipulation and findings.

End