

IBM Applied Data Science Capstone Project

*Venues Data Analysis of the 20 most populated cities in
Ecuador*

Daniela Sánchez

1. Introduction

1.1. Background

Tourism is one of the most important economic and cultural activities for any country. This industry promotes the economic reactivation of the site and generates sources of employment. Tourism has a favorable impact on sectors such as construction, hotels, gastronomy, and transport. Tourism activities creates demand across diverse industries.

Ecuador is a small country with an estimated population of 17 million, but it is also one of seventeen megadiverse countries in the world according to Conservation International. The country is divided into four regions: the Andes, the Amazon, the Pacific Coast and the Galapagos. The diversity of its four regions has resulted in hundreds of thousands of species of flora and fauna. From beaches to volcanoes, you can expect to find almost any kind of environment here [1]. The three most populated cities (Quito, Guayaquil and Cuenca) are attractive destinations for vacation or retirement.

1.2. Problem

The objective of this project is to analyze the most common venues of the 20 most populated cities in Ecuador. Due to the diversity of places to visit, determine if there is a relation between the cities and the categories of the venues. This analysis could contribute to identify a business opportunity and recommend the best location to open a business.

1.3. Interest

People who are interested to open a business that has relation with the tourism industry. Others who are interested in this analysis could be travelers or tourists who want to know the most common venues that they can enjoy in each analyzed city.

2. Data

To achieve the mentioned problem the following data was considered:

- From the National Institute of Statistics and Censuses of Ecuador (INEC), I obtained the data of the Ecuadorian population distributed by provinces, cantons (local government area) and cities. This dataset was obtained from the last census carried out in Ecuador [2]. I cleaned the data and reduced the data to the 20 most populated cities.
- The latitude and longitude coordinates of the cities was obtained with the python's package geocoder.
- I used **Forsquare API** to get the most common venues of the selected cities.

References

1. [Ecuador – Wikipedia](#)
2. [INEC](#)