

# THE FIVE AREAS IN LIMA – THEIR DIFFERENCES AND SIMILARITIES BASED ON A DATA SCIENCE MARKET RESEARCH

Frank Salvador Ygnacio Rosas

February 02, 2020

## 1. Introduction

### 1.1. Background:

Lima is the capital of Peru, and this city is mainly divided into five zones: north, mid, modern, south and east. Each of these areas has different characteristics, whether in the economic, social and cultural aspects. For instance, while the mid-town Lima is the most traditional part of the city (where the main government buildings are located, such as the Presidential Palace), the center of business and the corporate buildings are in the modern part of Lima. Hence, considering only the recreational aspect, it would be very useful to know the most popular types of venues within each of these five areas of the city. Moreover, it might be relevant to include an assessing of how their similarities and differences could be connected. This will be convenient to generate information for the business sector based on a market research as any market study should have as starting point an overview of the features that are the base of the similarities and differences over a target; in this case, over a city. Therefore, the two sort of questions that this study tries to solve could be “is a hamburger restaurant better in the north zone rather than in the east of the city?”, and “is the south to the north more similar than the center to the east?”.

### 1.2. Problem:

This study will list and visualize the different districts that make up each of the five areas of Lima. With this information, we will determine which are the most frequent venues in each area and their level of acceptance. Similarly, we will assess the similarities to one area of the city with another. Finally, we are going to summarize this information in a sort of general matrix that shows the relationships among the areas of the city of Lima.

## 2. Data Acquisition and Cleaning

### 2.1. We are going to use these five links of Wikipedia than summarize the information of the five areas of Lima:

General link: [https://es.wikipedia.org/wiki/Anexo:Distritos\\_de\\_Lima](https://es.wikipedia.org/wiki/Anexo:Distritos_de_Lima)

North of Lima: [https://es.wikipedia.org/wiki/Cono\\_Norte\\_\(Lima\)](https://es.wikipedia.org/wiki/Cono_Norte_(Lima))

Mid-town Lima: [https://es.wikipedia.org/wiki/Centro\\_hist%C3%B3rico\\_de\\_Lima](https://es.wikipedia.org/wiki/Centro_hist%C3%B3rico_de_Lima)

South of Lima: [https://es.wikipedia.org/wiki/Cono\\_Sur\\_\(Lima\)](https://es.wikipedia.org/wiki/Cono_Sur_(Lima))

East of Lima: [https://es.wikipedia.org/wiki/Cono\\_Este\\_\(Lima\)](https://es.wikipedia.org/wiki/Cono_Este_(Lima))

We will scraping the webpage to extract the relevant information, such as postal codes, districts, population, etc. Complementary, we are going to use the Foursquare data to get the main insights about the venues and their acceptances, and to analyze the possible similarities and differences among the five areas of the city.