

Clustering of Boroughs based on Cultural events on the city of Lima

Freddy Mendoza Ticona

Abril 14, 2019

1. Introduction

1.1. Overview

The goal of this project is to group districts of my city based on cultural activities like concerts, expositions, theatre, festivals, etc., and also on the entertainment venues around it. In order to determinate what is the most fun district to go a have a good time and to choose which district fit more to our preferences.

1.2. Problem

Information about scheduled cultural events is needed, even those that have already passed. That data will help us to determinate which districts have the most cultural activities, which can be used for recommendations systems.

1.3. Interest

People with specific interests can use this information to select which place is more fun to go out. Also for foreign people who want to know more about Lima, and use this classification when they decide to make a trip to visit.

2. Data acquisition

It can be done a clustering with only Foursquare's cultural venues around a district, but for more accurate precision, it is a good idea to include information about past and future events. Fortunately, I found a web page where it can extract it ([here](#)), it contains tables with information like the type of event, address, name, district, and price. For simplicity, I will use only information around March and Abril of 2019.

First, I scraped this web page [List of districts](#), to get the districts of Lima. I used all the tool I learned so far to return coordinates and to search venues with Foursquare, but we need to just filter by cultural venues. The more difficult part is to scrap the cultural agenda web page because it is needed to make a loop to search for every single day around two months.

For the features, I decided to count the numbers of cultural venues that Foursquare will give me, and also the count for every type of cultural event from the cultural agenda web page.