**Identifying the Most Important Problems in Latin American Countries – Based on Citizens’ Perceptions**

Report for the Applied Data Science Capstone project

Author: JUAN MARTINEZ

Date: 19/01/2021

**Purpose of the project**

This project's aim is to explore what are the Latin American citizens' views on what constitutes the most important problems in their countries. Latin American countries, as any other country in the world, have their particular issues that citizens deal with and which affect their development. Because of plenty of characteristics shared by Latin American communities, groups of countries in the region have common problems. By identifying these problems, we can more accurately advice on the adequate policy interventions that would improve the conditions of living in each subregion[[1]](#footnote-1) of the study.

The results and conclusions of this work might be of interest specially to governments, policymakers, and other agents with the power to influence the destiny of each region in the study who are willing to make the contribution.

**Introduction**

Latin America is home to about 642 million citizens, conformed by countries that share plenty of characteristics: from the colonial past and later independence of most of them; the romance languages, mainly Spanish, except for Brazil, French Guiana and Suriname; similar economic practices, political history and religion; but also, the similarities extend to the problems and challenges that obstruct their development.

These shared characteristics make it possible to look at Latin America as a particular region where patterns repeat within its communities, and this study, by analyzing data on the problems people from the region face, attempts to identify the main problems in the region and define profiles for different subregions that will make it easier for governments and policymakers to determine where and how to act to improve the conditions in each subregion.

**Objectives of the study**

The goals of this project can be summarized in the following three objectives:

1. Identify what are the most important problems Latin American citizens perceive in their respective countries.
2. Find patterns and define clusters for Latin American subregions based on the problems they share, which can make analysis of the living conditions in the region easier and allow us to think about specific policy implementations to improve them, following a strategy of “a measure to fix them all” for all the subregions with similar problems.
3. Assess how the problems reported in each subregion relate to indicators that are relevant to measure the quality of governance, such as trust in the government and institutions and satisfaction with life.

**Methodology**

The goals of this project can be met by means of tools of data analysis. Specifically, the study will follow these steps:

1. Cleaning, processing, and analysis of the data to obtain information on what are the main problems people perceive in their countries, both at the country level and by subregions.
2. Develop the appropriate visualizations to make analysis of the data easier.
3. A KMeans algorithm will be employed to define clusters for each subregion in the study based on the problems people perceive as the most important.
4. Comparisons of clusters’ performance in certain indicators will be carried on through ANCOVA analysis and other analysis of the scores.

**Specifications of the data**

**Source of the data:** All the data to work with is obtained from the Latinobarometro survey of 2018, which gathers data on 20,204 Latin American citizens’ values, perceptions, and opinions on a wide range of topics, of which most are relevant to their daily lives (https://www.latinobarometro.org/latContents.jsp).

After removing observations from subregions with less than 30 observations, the data will cover:

* Number of countries: 18
* Number of subregions: 176
* Number of observations: 15825

**Variables:** From the survey, diverse variables will be used for the different parts of the analysis. These parts include visualizations, clustering, geospatial visualization, and ANCOVA analysis. Some of the most important variables are:

* Most important problem in the country: responses to the question "most important problem in your county".

For ANCOVA analysis:

* Trust in the government.
* Trust in the institutions (composite of trust in the government, trust in the police, trust in the congress, and trust in the justice system)
* Satisfaction with life.
* Average education level.
* Average socioeconomic status.

Coordinates for geospatial visualization will be obtained from the Nominatim API.

1. In this study, subregion will refer to subnational divisions such as provinces, states, departments, etc., which are denominated differently in each country. [↑](#footnote-ref-1)