Relationships between the level of provincial Human Development Index and the type of venues of a province

Armando Medina

March 20, 2020

1. Introduction

Human Development Index (HDI) is a composite indicator that shows the level of development according to human life standards. The index is calculated by three main factors: income, education and health. While the Foursquare API allows you to explore the type of venues along with its respective ranking that exists in a certain location. Our objective is to find a relationship between the type of venues and their ranking with the HDI.

My main objective is to find out if there are relationships between the type of places in a province and its HDI.

This information may be used by public policymakers and private entrepreneurs or investors.

2. Business Problem

The potential problem identified is finding if there is a relationship between the types of places and their development indices to have information that helps better urban planning and infrastructure investment. In particular, the current planning system does not have an adequate way of measuring the impact on the HDI of a DR province. Our goal is to find a relationship between the type of places and development rates to help politicians and investors make better decisions.

3. Data

Based on definition of our problem, the data that we are going to use will be:

- **HDI**: The Human Development Index by province developed by UNDP for the Dominican Republic. Source: http://bit.ly/3bb3EQL
- **Wikipedia**: From WIkipedia we are going to obtain the population and density of each province. Source: http://bit.ly/3bbV1oL
- **GeoJson**: Source: https://gadm.org/download_country_v3.html
- **Google Maps**: With the Google Maps library we will obtain the lat & lng information for each province.
- **Foursquare**: With the Foursquare API we will obtain the information of the places and their category. Source: http://bit.ly/2UIDTq3

4. Methodology

In this project, we will direct our efforts to detect relationships between the Human Development Index of the provinces of the Dominican Republic and the type of venues they have.

First, we collected the HDI data, the geographical coordinates of each province, its population, area, population density, and we have made a query by province to obtain the places from the Foursquare API, ranking them by popularity within each province.

Second, we made the calculation and exploration of '**Density of places of interest**' in the different provinces of which are: Parks, Hospitals, Schools, Restaurants and Department Stores.

Last, we use the k-means clustering algorithm to explore possible relationships or groupings that may yield the information.

5. Results and Discussion

We use a Machine Learning algorithm to cluster by type of places with the information obtained from Foursquare on equal terms for all provinces.

We specified the algorithm to generate three groups. The provinces in each group are similar to each other in terms of the characteristics included in the data set.

Then, we created a profile for each group with their common characteristics:

- **Cluster 0:** These are the provinces with the lowest economic activity in the country and its reflected in an HDI and Income Index.
- **Cluster 1:** These are provinces that have higher incomes than the rest of the country, among them we found La Altagracia, which has the most important tourist pole in the Caribbean, San Pedro de Macorís with its sugar mill, and the National District with the largest population density in the country and it concentrates more than 70% of economic activity.
- **Cluster 2:** These are provinces with an HDI similar to those of group 1 but, in their composition, they have a better balance between the Health Index, the Income Index and the Education Index.

The following issues should be highlighted

- The information on Foursquare may be biased, but we think it is a good approximation.
- Outside the Distrito Nacional (capital of the country) whose population density is much higher than that of the rest of the provinces, we did not find a relationship between the density and its HDI.
- There is a poor distribution in the national territory, with the provinces with the highest population density having the least geographical area but perhaps the highest industrial concentration.
- In our first group using the entire category of place curiously, cluster 2 has an almost perfect coincidence with the main tourist centers of the country.

6. Conclusion

We have concluded that there is a correlation between the density of the type of business and its corresponding index. In other words, in provinces with higher income, there is a greater number of shops (or vice-versa, we have an endogeneity issue). The ones with the higher level of education, there are a greater number of educational centers. Lastly, in those provinces with higher value for the health index, we encounter more hospitals or health centers.