Predicting electoral participation using neighborhood characteristics

Elitza Maneva

Introduction

An important indicator of the quality of a democracy is the level of participation of its population in its electoral processes. Government officials everywhere are interested in incentivizing voting. The question we want to ask is the following: What kinds of neighborhoods, and local business should we target, specifically, in order to have the biggest possible effect on the total participation in an election?

Geographic data about voter participation is available after every election. It is often studied with respect to demographic characteristics of the different regions such as average age, average education level, ethnic and racial characteristics, population density and economic status. We believe that there are other indicators related to the lifestyle of the population, which can be deduced from the environment in which the citizens live, such as the characteristics of their neighbourhoods. This is the type of data about businesses and other local organizations that is readily available on Foursquare. Such indicators can be used to help us target particular economic sectors and communities that may not be so easily identifiable from census data.

Data

We will use three types of data:

In the first place, we can use the participation data for the Parliamentary Elections in Catalonia that took place on February 14, 2021. The vote count, and census data is available for each polling station on the open data portal of the Catalan Government: https://analisi.transparenciacatalunya.cat/ca/Societat-benestar/Eleccions-al-Parlament-de-Catalunya-2021-Recompte-/ix2p-vyw4

Second, we will use the geographical coordinates of the area serviced by each polling station. This information is available on the website of the Catalan Institute for Cartography and Geology: https://www.icgc.cat/Administracio-i-empresa/Descarregues/Capes-de-geoinformacio/Seccions-censals

The data is in map format so we will use QGIS to extract the centroids of all the polygons and convert them to geographic coordinates.

Finally, we will use data from Foursquare's API by querying areas within a fixed radius from the coordinates of the centroid of the area of a polling station for different types of venues. A variety of characteristics of these venues may provide useful information for our problem: the category of the venue, its ratings, hours during which it is popular, popularity score, etc.