

Foursquare venue's as an Election Predictor

Ankit Gulati

March 21, 2018

1. Introduction

1.1 Background

One of the most anticipated events of modern democracies is election. Citizens cast their valuable votes to elect whom they think will best serve their needs ranging from economic, social, health and many more. Conversations about the result of an election is the highlight of almost every household, news channel, social media and every facet of human communication. So it is imperative to have models using as many indicators to best predict the outcome of an election.

1.2 Problem

There are many data indicators available, which are used to predict and analyse the outcome of an election. The United States is primarily divided into a two party election. This project aims to study the trends and similarities between counties based on the most visited venues and assess how strong would these data points be at predicting the outcome of the 2016 presidential elections. This would provide information on the correlation between political identities of a county with its popular venues.

2. Data

2.1 Data Sources

Data relating to US Presidential election is found in abundance on the Internet and Kaggle. However, the data offered by opendatasoft.com had the most relevant features for my model and can be found [here](#). Since the model would use the venues in each county, I chose to use the latitude and longitude of the centroids of the respective counties, which is also provided in the above linked dataset. Foursquare was used to obtain the relevant popular venues.