Florida Democratic Party: Identifying Sub-Ethnicities of the Hispanic Population

Liana Parker, Ian Mungovan, Rich Beredo
Embry-Riddle Aeronautical University

Issue

The Democratic performance for Hispanics in the state of Florida decreased between 2016 to 2018, with a majority of the loss attributed to South Florida. The Florida Democratic Party wanted to analyze what caused the loss in voter performance by identifying Hispanics sub-ethnicities. This allows a greater aim for different hispanic voting pools and increase future performance.

Process

1. The user inputs a registered voters last name and zip code into the python code. The code is linked to an excel database that has information on Florida Zip Code data and a Last Name Database.
2. The algorithm finds all sub ethnicity data on the given zip code and information on the last name given.
3. The python code pulls all relevant information needed for the *Addition Rule for Probability* and calculates the mutually exclusive values of probability.
4. *Addition Rule for Probability* is given the calculated values and the sub ethnicity percentage is printed out.

Addition Rule for Probability

P(Y or Z) = P(Y) + P (Z) - P(Y and Z)

Zip code and Surname Data → Addition Rule for Probability In Python → Outputs a person's sub-ethnicity

22.5%

Percentage of Hispanics in Florida's total population in 2018

With a population of approximately 18.8 million people, only 4.2 million are of Hispanic Ethnicity

The three most common sub-ethnicities:
- Cubans : 28.691% of Hispanics
- Puerto Ricans: 20.012% of Hispanics
- Mexicans: 14.91% of Hispanics

6.3%

Of Active registered voters are Hispanic Democrats

94.5% of registered voters are active in the state of Florida

- Florida Democrats total : 37.06%
- Active voters of Hispanic Origin: 17%

Results and Analysis

The method utilized two given factors: surname and zip code
- This proved to be the most accurate for use in determining target demographics in certain areas
- There were multiple factors that could have been used in order to determine accuracy such as the following: First name, gender, age, occupation. Further research and testing proved that the use of these factors weren't as accurate and reliable as Zip Code and Surname
- Results are dependent on the combination of the following:
  - Frequency of unique names given for each sub-ethnicity
  - Diversity of a given zip code

Testing Accuracy and Validity:
- The product was tested on accurate voter information, provided by the Florida Democratic party, in order to test accuracy and validity.
- 7 different cases were found from the given initial data.
- The algorithm produced 7 accurately predicted primary sub-ethnicities.

This information proves the algorithm is close to 100% accurate in guessing a person's Primary sub-ethnicity, however, only a limited number of groups were tested. More data sets are needed to accurately assure the results are correct with close to zero error.

Sub-ethnicities Chosen

- The following 8 sub-ethnicities were used in the databases: Columbian, Cuban, Dominican, Guatemalan, Hondoran, Mexican, Puerto Rican, and Venezuelan.
  - These ethnicities were chosen because they make up an average of 85% of Hispanics in Florida

Lessons Learned

- Multiple methods were discussed on how to produce the results
  - Instead of having an A.I. system learn how to predict ethnicities, using probability proved to be the most effective with the data and resources available
- Finding reliable data with specific sub-ethnicities and needed information proved to be a lengthy process to find reliable information

Future Improvements

1. Increase data capacity to yield more accurate results and have a more conclusive answer.
2. Python code should be improved to print batch results
3. Include more detailed data categories to provide a more precise answer such as occupation, gender, and first name

References

Hispanic Latino Population by Specific Origin by ZIP Code. (n.d.). Retrieved from http://proximityone.com/hispanic_origin_zip.htm

Jain, A. (2017, May 13). avikj/Name-Ethnicity-Classifier. Retrieved from https://github.com/avikj/Name-Ethnicity-Classifier

Most Common Surnames - Genological Records. (2014). Retrieved from https://forebears.io/