BUSINESS UNDERSTANDING

Business overview

Elections in the United States are held quite differently from other countries. The president and the vice president are elected by electors through the **Electoral College.** The electors in each state are equal to the number of Senators and representatives in congress. There is a total of 538 electors in the Electoral College. Every elector votes once following the general election and the candidate who gets more than 270 votes wins the general election.

BUSINESS OBJECTIVE

The main objective is to find the states that should be prioritised in order for a candidate to win the general elections.

BUSINESS SUCCESS CRITERIA

Compiling a list of states that will increase the return on investment of the campaign leading to the winning of the potential candidate.

ASSESSING THE SITUATION

We worked with two datasets
- ❖ Grand Electors by State
- ❖ Population by State(links in the google collab notebook)

Assumptions
- ● The data provided is with no errors and up to date

Constraints
- ● No constraints have been noted.

Data Mining Goals
- ❖ Calculate the ratio of grand electors to the population in each state
- ❖ Order the states in decreasing order by grand electors per capita
- ❖ Get the sum of grand electors per capita

Data Mining Success Criteria
- ❖ Obtaining the target states that have a running total of more than 270

DATA UNDERSTANDING

Data understanding overview

For this analysis, we were availed two datasets
- ❖ Grand Electors by State - Gives the number of electors per state
- ❖ Population by State - Shows the population of each state

Data description

The following is a detailed description of the two datasets

❖ Grand Electors by State - It contains the number of electors in each state. The data contains two columns **State** and **GrandElectors.** The data has 52 entries since there are 52 states in the United States of America.
❖ Population by State - The dataset contains the population of the 52 states.

Verifying Data Quality
The datasets did not record any missing values. No errors have been recorded so far.

DATA PREPARATION
The following are steps used in the data preparation
❖ **Loading Data**
Loaded data from CSV files and later on created an SQLite database from them

❖ **Cleaning data**
In the Grand Electors by state, I noticed that some values in the state were in lowercase so I had to standardize them and change them into uppercase. This action was taken so that it corresponds with the other dataset. This would make the merging of datasets easier. I shortened the District of Columbia to DC as recommended by the company.

❖ **Merging of the datasets**
After the cleaning datasets, I merged the two datasets.

❖ **Deriving new attributes**
After the merging, we created a new column called ratio which was as a result of dividing the grand electors in each state with the correspondent population in the specific state. The table was later on ordered in descending order and another column called running_sum was created. In this column, the cumulative sum of the grand electors in the top states was calculated.

ANALYSIS
I was able to filter out the following states as the target for the potential targets since they had a running total of over 270. More analysis can be found in this link to my notebook .https://colab.research.google.com/drive/1rTYNIDKFQ5xDpkCszl2xgSa3UEslaYRt?authuser=1#scrollTo=zdRjR8RR3l5O

1. WYOMING 2. VERMONT 3. D.C 4. ALASKA 5. NORTH DAKOTA 6. RHODE ISLAND 7. SOUTH DAKOTA 8. DELAWARE 9. MAINE 10. NEW HAMPSHIRE 11. MONTANA 12. HAWAII

13. WEST VIRGINIA
14. NEBRASKA 15.
IDAHO 16. NEW
MEXICO 17.
NEVADA 18.
KANSAS 19.
ARKANSAS 20.
MISSISSIPPI 21.
UTAH 22.
CONNECTICUT 23.
IOWA 24. ALABAMA
25. SOUTH
CAROLINA 26.
MINNESOTA 27.
KENTUCKY 28.
OKLAHOMA 29.
OREGON 30.
WISCONSIN 31.
LOUISIANA 32.
WASHINGTON 33.
TENNESSEE 34.
MARYLAND 35.
INDIANA 36.
COLORADO 37.
MISSOURI 38.
MASSACHUSETTS
39. MICHIGAN 40.
ARIZONA

RECOMMENDATION

After my analysis, my recommendation is that the aspiring candidates should target the cities listed above. The reason behind this is because the running sum of their grand electors total is more than half of the total grand electors in the country. This would greatly increase the chance of the candidate winning the election and also return the investments of their campaign