## **Business Understanding**

### **Business Overview**

An election for president of the United States happens every four years on the first Tuesday after the first Monday in November. However, the President and the vice president are not elected directly by the citizens, instead, they’re chosen by ‘’electors’’ through a process called **Electoral College**. When people cast their votes, they are actually voting for a group of people called electors.

The number of electors each state gets is equal to its total number of Senators and Representatives in Congress. A total of 538 electors form the Electoral College. Each elector casts one vote following the general election. The candidate who gets 270 votes or more wins.

### **Business Objective**

The main objective of this report is to identify the states that should be prioritized in order to win the election.

### **Business Success Criteria**

To compile a list of states that will increase the return on investment of the campaign resulting in winning the election.

### **Assessing the Situation**

1. **Resource Inventory**
   1. Datasets:
      1. Grand Electors by State[[Link](https://drive.google.com/a/moringaschool.com/file/d/1AuPlNXKjmEdRTLGqKbP-OP1XZ5UQPIPw/view?usp=sharing)]
      2. Population by State.[[Link](https://drive.google.com/file/d/1VKt_hF2pRqPxcNb1DKotkVXWNd2HX_KL/view)]
   2. Software( Github, Google Collaboratory, SQLite)
2. **Assumptions**
   1. The data provided is correct and up to date
3. **Constraints**
   1. There are no constraints

### **Data Mining Goals**

Our data mining goals for this project are as follows:

* Calculate the grand electors per capita i.e calculate the ratio of grand electors to the population in each state
* Order the states in order of decreasing grand electors per capita.
* Get the running sum of grand electors of the top states.

**Data Mining Success Criteria**

Our success criteria will be measured by the following criteria;

* We target the states that have the running sum that is more than half of all the grand electors in the country.

## **Data Understanding**

### **Data Understanding Overview**

For this project, we are using the availed dataset by the company. These datasets are

* Grand Electors by State - This dataset gives the number of electors per state.
* Population by State - This dataset show the population of each state

### **Data Description**

We have two datasets available for this project. A detailed description of the datasets is provided as follows:

* **Grand Electors by State dataset -** This dataset contains the number of electors per state. It consists of two columns; ***State*** and ***GrandElectors.*** These columns outline all the states in the country and their number of electors respectively. Since there are 52 states in the United States of America, it goes to reason that the dataset has 52 entries.
* **Population by State -** This dataset, on the other hand, focuses on the population of each state. It contains the population for each of the 52 states.

### **Verifying Data Quality**

None of the two datasets had any missing values. There were also no known data errors in the datasets.

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the datasets from the CSV and then created an SQLite database from them.

#### **Cleaning Data**

While doing data exploration, we noticed that in Grand Electors by State dataset the values in State column were in lowercase and in the other dataset the values corresponding to the same column were in uppercase. Therefore, we decided to convert the State values in Grand Electors by State to uppercase so that it corresponds with the values in the other dataset. As a result, this would make the merging of the two datasets easier.

Furthermore, we shortened the name of “District of Columbia” state to “DC”. This was recommended by the company.

#### **Merging of the Datasets**

After cleaning the data, it was time to merge the two datasets.

#### **Deriving New Attributes**

Once merging was complete, we created a new column named “Ratio”. This column was populated by dividing the grand electors in each state with the corresponding population of the said state. Afterwards, we ordered the table in descending order of the ratios.

Next, we created yet another column called “RunningSum”. In this column, we calculated the cumulative sum of the grand electors in the top states.

## **Analysis**

During our analysis, we were able to single out the following states;

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

The above analysis was done using SQLite. The full analysis can be found in the following notebook.[[Link](https://colab.research.google.com/drive/1fKTEuy0yBaprrU2UlxFpNDW2g8PTcQ8Z)]

## **Recommendations**

From our analysis, we would recommend that a candidate would prioritize the above-listed states. Our main reason behind this recommendation would be that the states listed above have the highest descending ratio and also the running total for the grand electors in these states amounts to more than half of all the grand electors in the country. Therefore, prioritizing these states would not only increase the chances of winning the election but also would increase the return on investment in the entire campaign.