**Database Assignment**

**Dataset Description:**

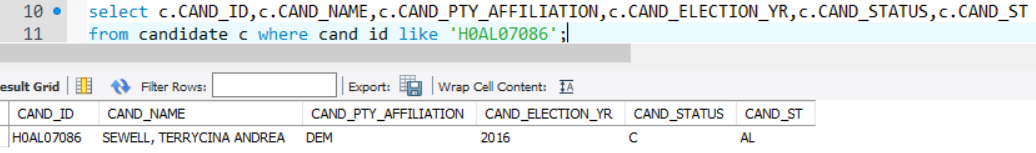
This dataset is based on Election Campaigning. It consists data about candidates who have taken part in the election, committees that have funded the candidates and individuals (specifically, contributors from the state of Maryland who support certain committees and candidates.) The tables in this dataset represent sets of transactions between individuals and political committees and between political committees and candidates. These records can provide insight about the relationships among individuals, the committees, and the candidates they support.

**Objective:**

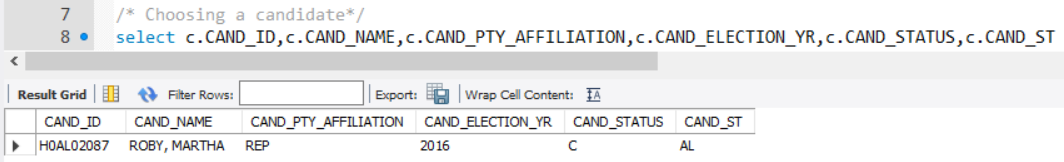
Concentrating on two areas that would help improve election campaigning for the chosen candidate (Terri Sewell) for the year 2018 for House elections.

**Approach:**

Chosen candidate: Terri Sewell representing the Democratic Party for contesting in the House elections.

****

***Figure 1. MySql code to extract Candidate 2: Terri Sewell***

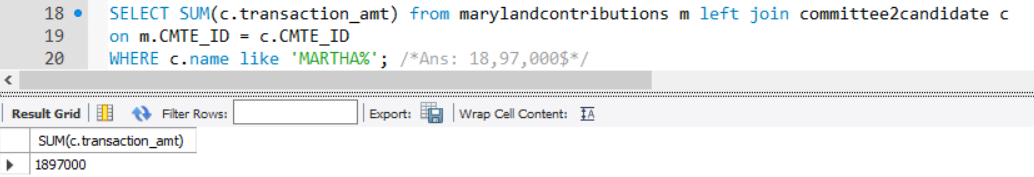
****

***Figure 2. MySql code to extract Candidate 1: Martha Roby***

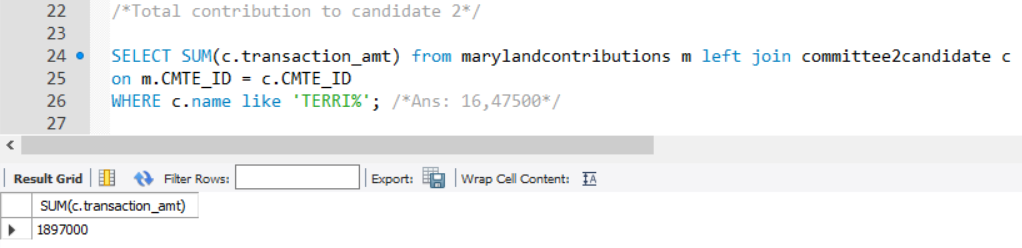
**Area of interest 1: Funding for Candidate**

The factor that primarily helps in making an election campaign successful is gathering proficient funds. Based on the total contributions collected by a candidate from the different committees, it is an essential factor that influences the candidate’s reach.

The following query helps analyze the total contribution by committees collected by our candidate-Terri Sewell with respect to our opponent candidate- Martha Roby:



***Figure 3. Total contribution to Martha***

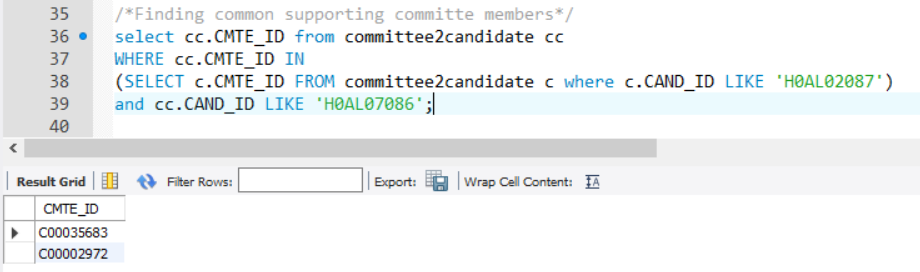


***Figure 4. Total contribution to Martha***

Based on the results generated from the query, we observe that our opponent candidate has collected 2,49,500$ more than our candidate! This query result allows us to understand that raising enough money for campaigning would have an impact in winning elections.

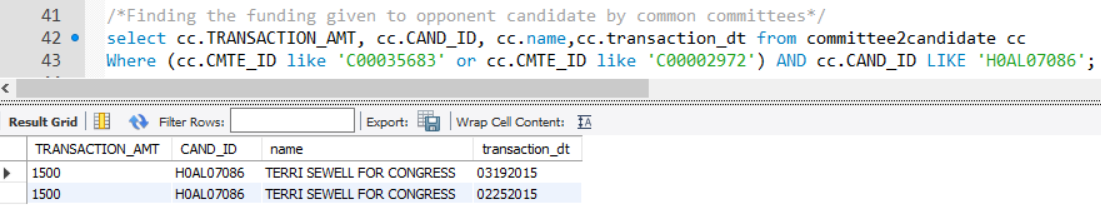
**Strategies to improve based on Result:**

* Having enough funds would imply that the candidate can have better advertising strategies. Advertising is an important factor as it helps reach a target audience. The wider the scope of reaching a target audience, the better the chances of winning in the election.
* In order to raise funds, it is important to understand the sources that could help a candidate monetarily. Therefore, we shall select our second area of interest as the committees that have helped in funding. Using the following query, we found the common committees that support both our candidates:

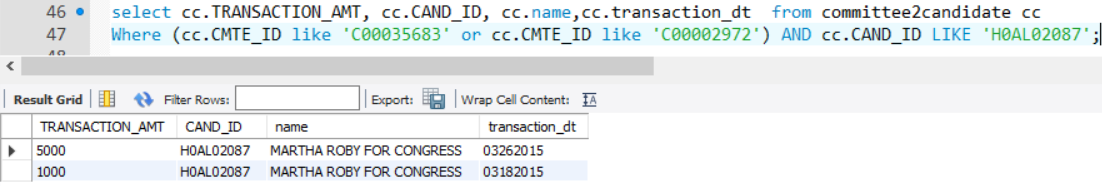


***Figure 6. Finding common committees that support both candidates***

We try to analyze the funds donated by these common committees and these are the results:



***Figure 7. Funding to Terri by common committees***

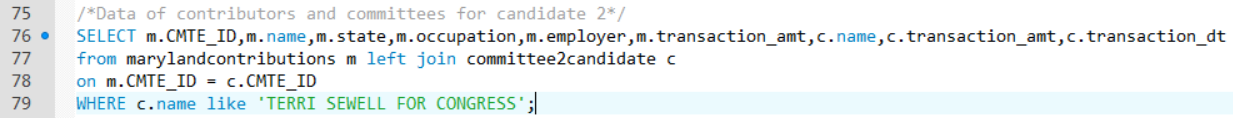


***Figure 8. Funding to Martha by common committees***

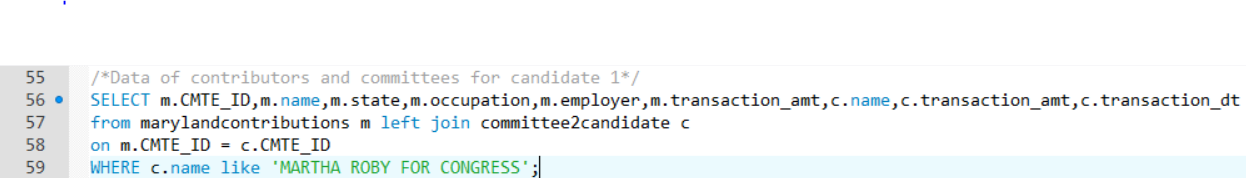
The total funds generated by the common committees for our candidate are 3,000$ whereas the funds generated by the opponent candidate are 6,000$! Thus, a good strategy would be to build and improve relations with common committees along with extending relations with other committees that could support monetarily.

**Area of interest 2: Analyzing influential people from Maryland**

As discussed above, targeting an audience plays a vital role in campaigning. Another factor that could help would be support from influential people. The table “*marylandcontributions*” in our dataset gives us an insight into individuals that donate money to committees, their employer, occupation and the transaction amount and date. We shall find the total number of individuals that support our candidate and our opponent candidate with the following query:



***Figure 9. Query to extract individuals who support Terri***



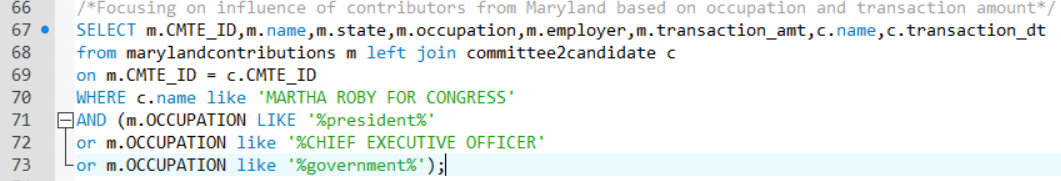
***Figure 10. Query to extract individuals who support Martha***

The results of these queries indicate that:

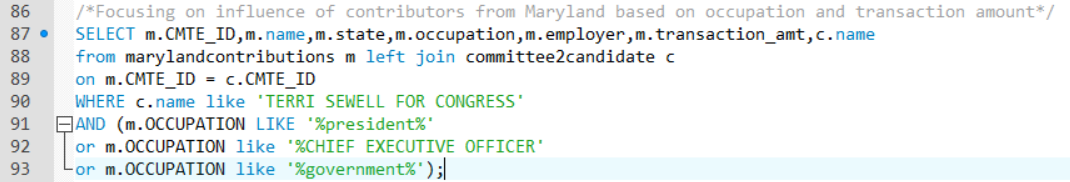
* Our candidate, Terri Sewell has 914 individuals from the state of Maryland who support him.
* Our opponent candidate, Martha Roby has 648 individuals from the state of Maryland who support her.

**Strategies to improve based on Result:**

* Although the number of supporters that our candidate, Terri Sewell has, is profoundly greater than the number of supporters of our opponent candidate, Martha Roby; it is important to understand who amongst these supporters have a greater influence on the audience based on their profession and reach. The following query gives us an insight to answer the areas discussed above:



***Figure 11. Query to extract Martha’s high influencers based on their work profile***



***Figure 12. Query to extract Terri’s high influencers based on their work profile***

The above query generates the following results:

* Our candidate, Terri Sewell, has 57 individuals who work at higher designations in their respective companies/organizations.
* Our opponent candidate, Martha Roby, has 93 individuals who work at higher designations in their respective companies/organizations.
* Thus, we understand that even though the total number of individuals who support our candidate are much greater than the number of people who support our opponent, the number of *influential people* who support Martha are greater than those who support Terri.
* This analysis suggests that it is not only important to have a large number of supporters from the state, but it is equally important to have supporters with greater influence. Networking with the 93 individuals who support Martha, can work as an impacting strategy in gaining higher number of votes. Extending and maintaining relations with the 57 individuals that support Terri should also be focused upon.

**Issues with the dataset:**

* The *transaction\_amt* variable has certain negative values. This is a major issue as it fails to justify why the amount is negative and what the negative amount signifies. It also affects the calculations for generating total funds collected by the candidate.
* Moreover, the data doesn’t seem to have information of all the candidates that have been funded and returns “NULL VALUES” for certain queries. Thus, signifying an incomplete dataset that needs data cleaning.