**Assignment 2**

**Databases and SQL**

The dataset given describes four tables related to election data. A brief description of the tables is as follows:

1. **candidate:** contains information about candidates like candidate ID, party affiliations, election year, status, name, state, city etc.
2. **committee:** contains information about committees like committee name, designation, type, party, treasurers name etc.
3. **committee2candidate:** contains information about contributions made from committees to candidates.
4. **marylandcontributions:** contains information about individuals who have contributed towards their preferred candidates.

**Candidate of choice**

Ruppersberger, C. A. Dutch, candidate ID = H2MD02160, Affiliated Party = Democratic

select sum(TRANSACTION\_AMT) from committee2candidate

where CAND\_ID like "%H2MD02160%" and TRANSACTION\_AMT > 0;

This query returns the total contribution towards my candidate which is 21000.

**Query 1**

When I think of identifying opportunities for support, money (or, funding) is the first parameter to analyse and work with. Funds is a very important factor to consider since it supports your campaigns financially. Hence, the first thing on my list would be to compare the funding received by my candidate and the funding for others within the same political party.

select c2c.CAND\_ID,sum(c2c.TRANSACTION\_AMT) as 'Total Contribution' from committee2candidate c2c

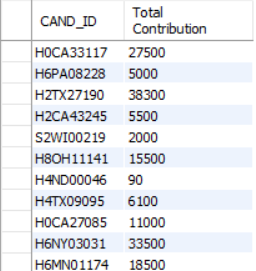
left join candidate c ON c2c.CAND\_ID = c.CAND\_ID

where c.CAND\_PTY\_AFFILIATION LIKE "%DEM%" and c2c.TRANSACTION\_AMT > 0

GROUP BY c2c.CAND\_ID;

The query fetches the total contributions towards all democratic candidates. From the output we can see that there exists a varying range of contribution amount. We are concerned with the candidates who are receiving funding higher than Ruppersberger, i.e. 21000. Using this data, I can identify the competitors within the party, analyse the type of campaigning they are conducting and where and tweak mine accordingly. This could also be an opportunity for Ruppersberger to strengthen his connections with the other powerful candidates of the party and in doing so, could garner more attention towards himself as well, possibly leading to increase in exposure and funding.

**Output Screencap:**



**Query 2**

One of the main components of being a politician is interacting with the citizens and letting them see the leader in you. As a result, your reach should ideally be far and wide to cover all bases; all strata of society. Hence, I have chosen to analyse the type of contributors that are funding Ruppersberger and compare them with the untapped pool.

**2A:**

select distinct OCCUPATION from marylandcontributions

where CMTE\_ID like "%C00227744%"

or CMTE\_ID like "%C00163832%"

or CMTE\_ID like "%C00040998%"

or CMTE\_ID like "%C00096156%"

or CMTE\_ID like "%C00341602%"

or CMTE\_ID like "%C00109017%"

or CMTE\_ID like "%C00078451%"

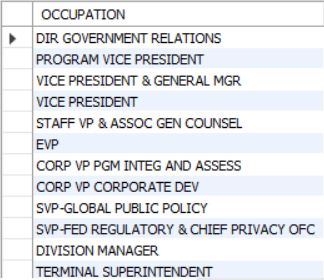
or CMTE\_ID like "%C00141002%"

or CMTE\_ID like "%C00144766%"

or CMTE\_ID like "%C00078451%"

or CMTE\_ID like "%C00325092%";

**Output Screencap:**



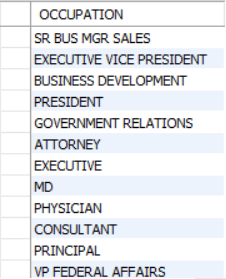
**2B:**

select distinct OCCUPATION from marylandcontributions

where CMTE\_ID NOT IN ('%C00227744%','%C00163832%','%C00040998%','%C00096156%','%C00341602%','%C00109017%',

'%C00078451%','%C00141002%','%C00144766%','%C00078451%','%C00325092%');

**Output Screencap:**



Query 2A returns the occupations of the Maryland individuals who are a part of the committees that fund Ruppersberger while Query 2B returns the ones that aren’t.

It’s clear that Ruppersberger has a quite healthy monetary support from high level professionals. Although this is an added advantage, we must not ignore the required support from the working-class public. If the campaign is pivoted to include them we could gain more influence and support, in turn increasing the candidate reach and funding. With more data, we could drill down to see what the remaining individuals have in common with Ruppersberger’s agenda and design marketed campaigns to attract them.

**Issues**

There exists several **data formatting and consistency** issues within the dataset. Most of them encompass the below specifics:

1. Null values
2. Impossible transaction amounts (negative values)
3. Zip Code format not consistent

The variables also don’t seem to be very self-explanatory and that can lead to some confusion as to what the data represents.