**Overview:**

The election campaign for any candidate is an organized effort of many key organizations involved in the political campaign. The campaign team consists of campaign managers, political consultants, techniques for campaigning such as advertisement, media management, rallies, Campaign finance and contributors, campaign agenda and the representative candidate himself. For any candidate to drive a successful campaign, the team needs to perform analysis on a number of factors such as financial contributions, identify areas of interest, identify target voters, influential factors and strong campaign message, identify ideal party affiliation with respect to political moto and objective, identify affiliations between communities and contributors and identify the current political influencers and potential competitions.

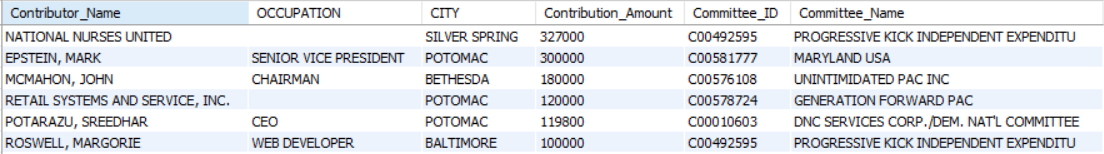
**Resources and Objectives:**

The following study is an analysis of the election data available to identify areas of opportunities and drive the political campaign for a new candidate for the upcoming elections. The election database available consists of four tables namely candidate table, committee, candidate2committee and Maryland contributions.

**Data Set Analysis and Observations:**

In order to drive the political campaign successfully, we need to identify opportunities and focus on the areas of interest. One of the main areas to focus would be the campaign finance. We need to analyze the contribution behavior in the state of Maryland to identify potential donors. The dataset also provides various committees affiliated with the contributors. This will help us identify the campaign contributors that have made significant contributions in the electoral campaigns of their respective candidates. This can be achieved with the help of the following subquery:

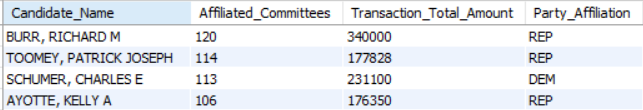
*SELECT M.name as Contributor\_Name, M.OCCUPATION, M.CITY, sum(M.transaction\_amt) as Contribution\_Amount, C.cmte\_id as Committee\_ID, C.CMTE\_NM as Committee\_Name FROM infm600\_hw3.marylandcontributions as M inner join infm600\_hw3.committee as C ON M.CMTE\_ID=C.CMTE\_ID group by M.name, C.cmte\_id order by Contribution\_Amount desc;*



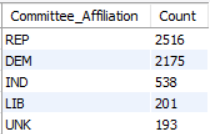
From the above query, we see that the organization National Nurses United is one of the largest contributors with $327,000 followed by Mark Epstein, John McMahon, Retail Systems and Services Inc with contributions of over $100,000. This can give us the target contributors to pitch our campaign message to and involve these contributors to fund for our campaign drive. The resultant dataset also provides us with the committee names associated with each of the major contributors and we can see that certain committees are common among the major contributors such as Progressive kick Independent expenditure, Unintimidated Pac Inc and DNC Services Corp. This will give us a boosted start to pitch our campaign strategy to the right committees who can help fund our drive. On further analysis of the employer column we realize that the major contributors are people who are retired and self-employed. This can help us focus on the right demographic and also try and focus on potential public and private sector employees with the help of campaign advertisements and other media.

The campaigning team has to decide on the party affiliation for the candidate based on the analysis of support for each political group. The team also needs to identify the influential candidates on the basis of various committee affiliations and contributions for these candidates in order to plan the electoral campaign.

*Select CAN.CAND\_NAME as Candidate\_Name, count(CMTE\_ID) as Affiliated\_Committees, sum(TRANSACTION\_AMT) as Transaction\_Total\_Amount, CAN.CAND\_PTY\_AFFILIATION as Party\_Affiliation from infm600\_hw3.committee2candidate as COMM, infm600\_hw3.candidate as CAN where COMM.CAND\_ID=CAN.CAND\_ID group by CAN.CAND\_NAME order by count(CMTE\_ID) desc;*



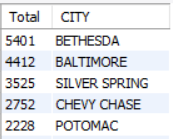
*SELECT CMTE\_PTY\_AFFILIATION as Committee\_Affiliation, count(CMTE\_NM) as Count FROM infm600\_hw3.committee group by CMTE\_PTY\_AFFILIATION order by count desc;*



The above query gives us the most influential candidates for the upcoming election based on the number of affiliated committees supporting these candidates and the party they are affiliated to. We can see that candidate Richard Burr from the republican party has the highest number of affiliated parties and electoral contributions followed by Patrick Toomey (REP), Charles Schumer (DEM) and others. On further analysis of the party affiliations we realize that majority committees are affiliated with the Republican Party followed by the Democratic party and the Independent candidate parties. This gives us an insight on the major preference of the political committees and other contributor groups to make an informed decision on the party affiliation for our candidate.

The team also needs to identify areas for driving political rallies based on the party affiliations, identify active voters based on area and occupation so that we can target the right support group to gather votes for the upcoming election.

*select count(CITY) as Total, CITY from marylandcontributions group by CITY order by Total desc;*



On analyzing the results from the above query, we can see that the city of Bethesda has the most politically active contributors in the state of Maryland with a total count of 5401. Other cities with active political contributors by order include Baltimore, silver spring, Chevy chase, Potomac etc. This will help us plan the areas of interest to drive our political rallies and campaigns for the upcoming election.

**Limitations:**

The Election data set under study has a lot of inconsistencies such as:

* The data set provided consists of number of missing values in table and has unknown values and certain outliers associated with it. For example, Party affiliation field has 11576 missing values, which means there are 11576 political committees which not associated with any political groups.
* Many of these committees have Party affiliation set as UNK which means these are unknown values.
* The Transaction amount column in Maryland Contributions table has negative values associated with some of the contributors which represent illegal values.
* The city column has the same city name spelled in different ways which makes it difficult to analyze the data.

With proper data cleaning process, the data can be presented in a consistent format for better analysis and study.