**INFM 600-Database Assignment**

As part of the Database Assignment, I have been hired by a candidate, let us say Mr. X who is interested in contesting for the elections. My role would be to help the candidate identify opportunities that would support him/her for the campaigning. I have chosen to work on the following opportunities:

1. Candidate donors
2. Affinities between and among different committees or individuals

For the purpose of the assignment, I would be analyzing the ElectionData dataset and using the MySQL software to gain insights into areas such as candidates’ association/memberships with the committees, the donations by the citizens of Maryland, link between the individuals, committees and the candidates.

Areas of Improvement:

1. **Candidate Donors**:

Monetary donations play a crucial role in helping the committees and in turn the candidates to raise money for campaigning activities such as candidate promotions, advertisements, using social media and other platforms to reach out to potential voters to communicate their message and to convince them to vote for them. After analysis of the dataset, I would like to improve the chances for the candidate to select committees that receive the most funding if the candidate is interested in joining a committee. The benefit from joining this committee is that the candidate will have greater chances/probability of getting a higher funding to support his campaigning activities. Also, it is possible for a candidate to join multiple committees. In order to find the committees that receive the most funding, I am looking at the table “MarylandContributions” and retrieving the committees that receive the highest funding:

I am using the below query in MySQL to get the top 30 entries for the maximum funding, the names of the donors and the committee IDs they are making donations to (The committee IDs are unique for each committee and will help identify the committee):

**SELECT CMTE\_ID, NAME, TRANSACTION\_AMT, EMPLOYER, CITY**

**FROM MarylandContributions**

**GROUP BY TRANSACTION\_AMT DESC**

**LIMIT 30;**

The output for the above query is as under:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CMTE\_ID** | **NAME** | **TRANSACTION\_AMT** | **EMPLOYER** | **OCCUPATION** | **CITY** |
| C00581777 | EPSTEIN, MARK | 300000 | QUALCOMM, INC. | SENIOR VICE PRESIDENT | POTOMAC |
| C00492595 | NATIONAL NURSES UNITED | 200000 |  |  | SILVER SPRING |
| C00492595 | NATIONAL NURSES UNITED | 127000 |  |  | SILVER SPRING |
| C00578724 | RETAIL SYSTEMS AND SERVICE, INC. | 120000 |  |  | POTOMAC |
| C00576108 | ST. JOHN, EDWARD MR. | 100000 | ST. JOHN PROPERTIES | CHAIRMAN | BALTIMORE |
| C00576108 | MCMAHON, JOHN | 80000 |  |  | BETHESDA |
| C00508416 | DELANEY, JOHN K. | 67500 | US HOUSE OF REPRESENTATIVES | MEMBER OF CONGRESS | POTOMAC |
| C00010603 | WALLACE, HENRY SCOTT | 66800 | SELF-EMPLOYED | NON PROFIT, DIRECTOR | BETHESDA |
| C00573386 | PLASTER, MARK LEROY DR. | 50000 | U OF MD EMERGENY MEDICINE | PHYSICIAN | ANNAPOLIS |
| C00003418 | MOSIER, FRANK JR. | 36575 | KAZIMIR PARTNERS | INVESTMENT BANKER | SAINT MICHAELS |
| C00010603 | MEYERHOFF, ROBERT E | 33400 | HENDERSEN-WEBB, INC. | EXECUTIVE | PHOENIX |
| C00010603 | GABRIEL, RONALD L TRUST | 32400 | DECEASED | DECEASED | BETHESDA |
| C00027466 | SMITH, JANE MRS | 30000 | HOMEMAKER | HOMEMAKER | COCKEYSVILLE |
| C00042366 | SMITH, MICHAEL D | 28000 | CORNERSTONE GOVERNMENT AFFAIRS | GOVERNMENT RELATIONS | CHEVY CHASE |
| C00488742 | MRIS | 25000 |  |  | ROCKVILLE |
| C00550004 | CHRISTOPHER, WARREN | 20715 | RETIRED |  | UPPER MARLBORO |
| C00010603 | JONES, ADAM | 20000 | BALTIMORE ORIOLES | PROFESSIONAL ATHLETE | LUTHERVILLE |
| C00010603 | POTARAZU, SREEDHAR | 18400 | VITALSPRING | CEO | POTOMAC |
| C00010603 | POTARAZU, SREEDHAR | 18000 | VITALSPRING | CEO | POTOMAC |
| C00576108 | GOULD, MARY T. | 17500 |  |  | LAUREL |
| C00010603 | POTARAZU, SREEDHAR | 17400 | VITALSPRING | CEO | POTOMAC |
| C00586537 | ISLAM, FRANK | 17200 | FI INVESTMENT GROUP | CEO | POTOMAC |
| C00010603 | SMITH, JUNE B. | 17000 | SELF EMPLOYED | IT | SPARKS |
| C00586537 | MARCHICK, DAVID | 16500 | CARLYLE GROUP | MANAGING DIRECTOR | CHEVY CHASE |
| C00010603 | ISLAM, FRANK F | 16200 | FI INVESTMENT GROUP | CEO/CHAIRMAN | POTOMAC |
| C00010603 | POTARAZU, SREEDHAR | 16000 | VITALSPRING | CEO | POTOMAC |
| C00010603 | POTARAZU, SREEDHAR | 15000 | VITALSPRING | CEO | POTOMAC |
| C00010603 | KAPLAN, CAROLYN J | 13400 | SELF | MEDIATOR | POTOMAC |
| C00229377 | T. ROWE PRICE | 12858 |  |  | BALTIMORE |
| C00010603 | GABRIEL, RONALD L TRUST | 12843 | DECEASED | DECEASED | BETHESDA |

We can get the following insights from the output:

* The committees receiving the maximum donations
* The names of the donors in the highest monetary bracket with maximum donations
* The occupations and employers of the donors that can help Mr. X get a clue about the social bracket of donors making the maximum donations.
* The geographical distribution of the donors that can help Mr. X decide where the highest donors are placed in case he wants to campaign more in those areas. The top cities with highest donors are Potomac, Silver Spring, and Bethesda. The geographical distribution can be seen in the following visualization:

**Figure 1: The above graph shows the distribution of the donation amount based**

**on cities.**

1. **Affiliations of the committees:**

As part of any election campaign, it is necessary to have information about the party affiliations of other candidates (particularly opponents) and committees and how they vary based on states and cities. The party affiliation of a committee or a candidate in a particular region can also provide some hint about the affiliation of the residents of the region. This gives some clue about the preferences of the masses, and how affiliations vary with respect to states and cities. Assuming that the candidate who has hired me is from Maryland and is running for office in the same state, it is important that Mr. X has all details about which party affiliation is dominant in which region of the state and which committee is affiliated to which party.

Using the ElectionData dataset, I would like to provide Mr. X with details of party affiliations of the various committees in the cities in the state of Maryland where he is interested in contesting elections. I would like to retrieve only the information relevant to Maryland from the complete dataset. I hope that this information enables Mr. X to decide if:

* He wants to campaign more for himself in areas where his party affiliations are NOT prevalent and would like to promote his ideals there.
* He wants to campaign in areas that have the same party affiliations as him and would make it comparatively easier for him to gain support.
* Which party poses a stronger competition in the specific regions.

To retrieve this information, I would check the “Committee” table and select the committee ID, the name, city and party affiliation of the committee only for the state of Maryland. I will use the below query to retrieve the information:

**SELECT CMTE\_ID, CMTE\_NM, CMTE\_PTY\_AFFILIATION, CMTE\_CITY, CMTE\_ST FROM committee**

**WHERE CMTE\_ST='MD'**

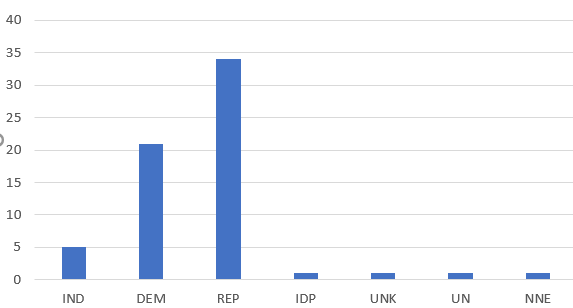
**GROUP BY CMTE\_CITY ASC;**

The output for the above query is as follows (posting a snippet of the output as the original output has 131 rows):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CMTE\_ID** | **CMTE\_NM** | **CMTE\_PTY\_AFFILIATION** | **CMTE\_CITY** | **CMTE\_ST** |
| C00577494 | DAVIS FOR CONGRESS | DEM | BEL AIR | MD |
| C00556761 | THE PEOPLE FOR DAVID BANACH | REP | BELCAMP | MD |
| C00584698 | FRIENDS OF RICH DOUGLAS | REP | BLADENSBURG | MD |
| C00063024 | SERVICE STATION DEALERS OF AMERICA, INC., SSDA-PAC (FORMERLY NCPR-PAC) | DEM | BOWIE | MD |
| C00601021 | ROBIN FICKER FOR CONGRESS | REP | BOYDS | MD |
| C00511360 | KATHY AFZALI FOR CONGRESS | REP | BRADDOCK HEIGHTS | MD |
| C00577882 | VOGT FOR MARYLAND | REP | BRUNSWICK | MD |
| C00255190 | BARTLETT FOR CONGRESS COMMITTEE | REP | BUCKEYSTOWN | MD |
| C00438861 | WALTER FOR CONGRESS COMMITTEE | REP | CENTREVILLE | MD |
| C00516930 | FRIENDS OF JEREMY STINSON FOR CONGRESS | IND | CHELTENHAM | MD |
| C00507335 | LAFERLA FOR CONGRESS | DEM | CHESTERTOWN | MD |
| C00578583 | MINORITY UNITY PAC; THE | DEM | CHEVERLY | MD |
| C00509281 | BROADUS FOR CONGRESS | REP | CLINTON | MD |
| C00507608 | DAVE LOCKWOOD FOR CONGRESS | DEM | COLESVILLE | MD |
| C00475558 | ANDREW FOR CONGRESS | DEM | COLLEGE PARK | MD |
| C00551713 | WALLACE FOR AMERICA | REP | CROFTON | MD |
| C00585570 | FRIENDS OF SHEWBERT FOR PRESIDENT | REP | DAMERON | MD |
| C00511535 | MUSE FOR MARYLAND | DEM | DAVIDSONVILLE | MD |

For instance, in the output snippet above, the candidate can see the two major party affiliations of the committees in some of the cities in the state of Maryland.

We can visualize the above output (ignoring the missing values) as below in a graph with the party affiliations on the x-axis and the number of cities on the y-axis:



**Figure 2: The above figure is a plot of number of cities versus the party affiliations in the cities across the state of Maryland**

It is also possible to extract data about the party affiliations of both committees and other candidates across the nation by running the below queries. This information will be helpful if the candidate wishes to run for a bigger role in the future.

1. **Some major party affiliations of other candidates across the USA:**

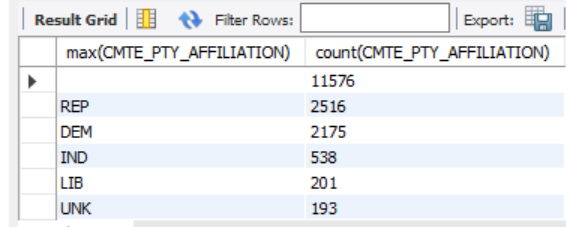
QUERY:

**SELECT MAX(CAND\_PTY\_AFFILIATION), COUNT(CAND\_PTY\_AFFILIATION)**

**FROM CANDIDATE GROUP BY CAND\_PTY\_AFFILIATION**

**ORDER BY COUNT(CAND\_PTY\_AFFILIATION) DESC;**

OUTPUT:



1. **Some major party affiliations of committees across the USA:**

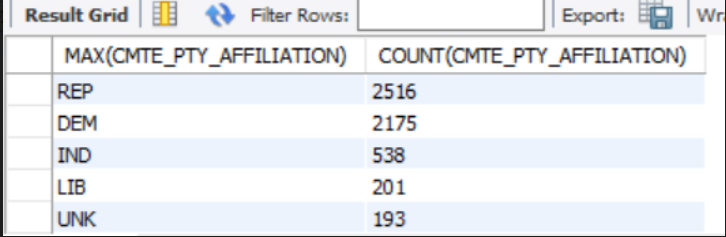
QUERY:

**SELECT MAX(CMTE\_PTY\_AFFILIATION), COUNT(CMTE\_PTY\_AFFILIATION)**

**FROM COMMITTEE GROUP BY CMTE\_PTY\_AFFILIATION**

**ORDER BY COUNT(CMTE\_PTY\_AFFILIATION) DESC;**

OUTPUT:

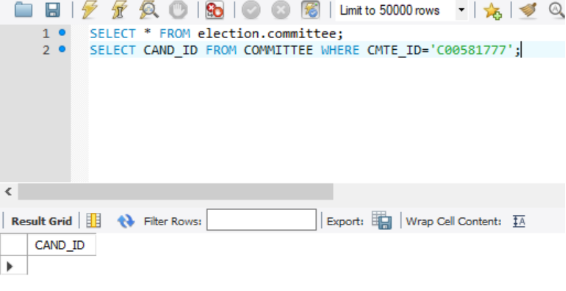


Both the tables above give useful information about dominance with respect to party affiliations with respect to committees and candidates and can help Mr. X assess his own party affiliation relative to others in terms of numbers.

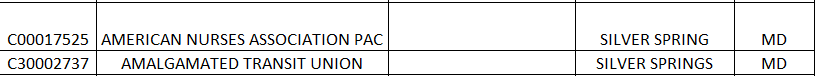
**Potential issue with the data:**

While working with the dataset, when I was trying to find information about the committees that received the maximum funds from donors and shortlisted the top 30 from the Marylandcontributions table, I found that a lot of the CMTE\_IDs present in the MarylandContributions table were absent in the Committee and Committee2Candidates tables. For some committee IDs that were present in the Committee table, there were no candidates associated with those committees. Therefore, it was hard for me to track more information about the respective committees.

As part of my approach, I noted down the committee IDs with maximum donations and when I tried tracking the candidates associated with them, blank values popped up in the output. This made it challenging to identify which candidates associated with those committees. An example of the same can be found as follows:



Also, (in addition of a large number of missing entries) there were some inconsistencies in the dataset with respect to how the data was entered, an example of which can be found below where we can see that the city Silver Spring is entered incorrectly:



Word count: 1597 words