

Database Assignment**Question.**

Who are the major contributors in Maryland, do they have any bias for any political party or any agenda, what type of campaigns are run by these committees.

Analysis:

On querying the top contributors from Maryland it was found that major funding contributions came from people listed under retired, unemployed and self-employed category and from organizations like Lockheed Martin and Exelon Corporation which are working directly with government. Thus, it is clear that appeal should be made to the group of people who are looking for jobs by promising them better employment rates or to retired people by being thoughtful about their retirement period and providing them various benefits. Also, it was observed that for contributions coming from large corporations were not directly related to any particular candidate or a committee campaigning for a candidate but rather to committees campaigning for a particular agenda or group of people. Thus, the candidate should be looking to support the causes advocated by these committees and campaign in these particular areas, for instance there were committees which were unions for workers in transportation area, food, rural electricity etc. Thus, campaigning some of the agendas supported by these committees would attract larger funds. Also, it was found that majority of these affiliations were with Republic party. Thus, as a Democrat candidate, we can look for funders who magnanimously funded current democrat incumbents. For Democrats it was found that large number of funds came from organizations like International association of Machinist, USPS, United Airlines etc. Thus, we can campaign by keeping in mind needs for these organizations.

Queries used:

1. create view democrats as
 select ca.CAND_PTY_AFFILIATION,ca.CAND_ID from candidate ca, (
 select ca.CMTE_ID,ca.CAND_ID from committee2candidate ca,
 (select c.CMTE_ID from committee2candidate c,
 (select employer,sum(transaction_amt),count(transaction_tp) as trans_count from
 marylandcontributions group by employer
 order by trans_count DESC limit 10) as md
 where c.employer=md.employer) as cmt where cmt.CMTE_ID=ca.CMTE_ID) as cn where
 cn.cand_id=ca.cand_id and ca.CAND_PTY_AFFILIATION like 'DEM';

Database Assignment

Result:

CAND_PTY_AFFILIATION	CAND_ID
DEM	H6CA46033
DEM	H0CA27085
DEM	H0RI02139
DEM	H4AZ07043
DEM	H4HI01134
DEM	H2CA31125
DEM	H2CA31125

2. select md.employer,count(md.TRANSACTION_AMT) as total from marylandcontributions md,
(select cm.CMTE_ID from committee2candidate cm, candidate c where
c.CAND_ID=cm.CAND_ID and c.CAND_ID in (select Cand_ID from democrats)
) as cd
where cd.CMTE_Id=md.CMTE_ID group by md.employer order by total DESC

Result:

employer	total
INT'L ASSOCIATION OF MACHINISTS	464
INT'L ASSOC. OF MACHINISTS	220
TCU GRAND LODGE	167
RETIRED	88
USPS	86
GD INFORMATION TECHNOLOGY	53
UNITED PARCEL SERVICE INC.	33
NATIONAL ASSOCIATION OF REALTORS	32
GENERAL DYNAMICS CORPORAT	30
UNITED AIRLINES	28

Question.

What are the different types of institutions that are providing funding to the chosen candidate (challenger from College Park for elections contested in year 2016). Also, I wanted to know, who are the major funders and to whom these transactions are made for the candidates contesting from similar geographical area, who had same party affiliation and had incumbent candidate status.

Analysis:

On analyzing the data, it was found that all of the contributions were made to candidate committee. This indicates that the funders are looking for specifically a committee which is vocal for a particular type

Database Assignment

of candidate. Thus, the funding is neither dependent on the popularity or status of a committee or his candidate but rather in the committee's ability to pitch for the candidate. Also, it was observed that none of the transactions were made from the candidate local state or city. Another interesting thing that was observed was that the amount of funding for each of these transactions was not large. Thus in order to gain more funding, the committees need to preach about the candidate outside his local city and state and should also look for small amount transactions, if the funding giants are not ready to contribute.

Queries used:

1. select * from committee2candidate cx where cx.CMTE_ID in (select cm.CMTE_ID from candidate c, committee cm where c.cand_name like 'GALL, ANDREW CHARLES' and c.cand_ID=cm.Cand_ID)

Result:

No rows were returned

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP	ENTITY_TP
---------	-----------	--------	-----------------	-----------	----------------	-----------

2. select * from candidate where CAND_City like 'Washington%' and CAND_PTY_AFFILIATION='DEM' and CAND_ICI='I'

Result:

CAND_ID	CAND_NAME	CAND_PTY_AFFILIATION	CAND_ELECTION_YR	CAND_OFFICE_ST	CAND_OFFICE	CAND_OFFICE_DISTRICT	CAND_ICI
H0CA10149	GARAMENDI, JOHN	DEM	2016	CA	H	03	I
H0DC00058	NORTON, ELEANOR HOLMES	DEM	2016	DC	H	00	I
H2FL23021	HASTINGS, ALCEE L	DEM	2016	FL	H	20	I
H2MD05155	HOYER, STENY HAMILTON	DEM	2016	MD	H	05	I
H4NJ12149	COLEMAN, BONNIE WATSON	DEM	2016	NJ	H	12	I

3. select * from committee2candidate cx where cx.CMTE_ID in (select cm.CMTE_ID from candidate c, committee cm where c.CAND_City like 'Washington%' and c.CAND_PTY_AFFILIATION='DEM' and c.CAND_ICI='I' and c.CAND_ID=cm.CAND_ID)

Result:

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP	ENTITY_TP	NAME	CITY	STATE
C00140715	N	Q1	P	15951145645	24K	CCM	SEAN PATRICK MALONEY FOR CONGRESS	NEWBURGH	NY
C00140715	N	Q1	P	15951145645	24K	CCM	PETE AGUILAR FOR CONGRESS	SAN BERNARDINO	CA
C00140715	N	Q1	P	15951145642	24K	CCM	GRAHAM FOR CONGRESS	TALLAHASSEE	FL
C00140715	N	Q1	P	15951145641	24K	CCM	LOEBSACK FOR CONGRESS	IOWA CITY	IA
C00140715	N	Q1	P	15951145642	24K	CCM	DR. RAUL RUIZ FOR CONGRESS	LA QUINTA	CA

Database Assignment

Potential issues with data:

The data provided was hard to understand for someone who doesn't have domain knowledge of US elections and campaigning. For lot of columns the data stored was an abbreviation of the actual terms thus making it difficult to understand without any metadata. Also for the candidate names in candidates table some of the rows had fake news, which can be easily identified, indicating that the data was corrupted to a certain extent.

	CAND_ID	CAND_NAME	CAND_PTY_AFFILIATION
	H0MD05126	GALL, ANDREW CHARLES	DEM
	H6MD04217	PENA-MELNYK, JOSELINE A.	DEM
	P60007275	LIU, GEOFFREY KUANG HUA	W
	P60010261	PRINCE OF DARKNESS, SATAN LORD OF UNDE	REP
	P60012366	HANNURABI, HEAVEN	IND
	P60017613	HENAO, DARREN MR. BIGDEE	OTH
	P60022886	JESUS CHRIST, YESHUA LORD MIGHTYGOD	KIN

For instance, the last row in the above image shows that candidate name is Jesus Christ, Yeshua Lord Mightygod, which cannot be real name of a candidate. Apart from that lot of employer names where having null value which lead to ambiguity in data, as to whether there were no fixed employer related to that particular transaction or whether there was name registered. All these incorrectness in data, can be handled by properly scrutinizing data and performing data some data cleaning.