ISYS 613 - Data Sourcing and Quality - Course\_Project\_Fall2019

Dr. Jeffrey Roberts, Duquesne University - School of Business

**TEAM Participant - MONIKA PAWAR**

**Project Overview**: This Project is designed to assemble data to answer the following research question:

For a given 2020 Democratic Party presidential candidate, how would you characterize their support from

(a) a gender perspective and

(b) an ethnicity perspective?

Analysis - Proof of Concept

**Presidential candidate – ‘Booker, Cory’**

Que1: How many male and female contributor supported the candidate?

To find out the total number of male and female contributors for the candidate, I have used **Group by** to the ‘Gender’ column of my final data frame and **counted** the ['CMTE\_ID'].

After analysis I found out that there are a greater number of female contributors as compared to the male contributors.

|  |  |
| --- | --- |
| GENDER | TOTAL COUNT |
| Female | 4626 |
| Male | 3775 |

Que2: How much transaction amount each ethnicity category has raised for the candidate?

To know which ethnic group has raised the higher total transaction amount for the candidate, I applied **Group by** to the ('Category') Column and used **sum** function to ['TRANSACTION\_AMT'] of the data frame.

White population has contributed the most and Hispanic has given the least contribution and there is NO black population contribution to the candidate.

|  |  |
| --- | --- |
| CATEGORY | TOTAL TRANSACTION AMOUNT |
| White | 1767204.0 |
| Unknown | 95364.0 |
| Asian | 64617.0 |
| Hispanic | 56.0 |

Que3: Which states are the Top 10 highest contributors for the candidate?

In my analysis I wanted to find out which all state has participated the most to raise the transaction amount for the candidate. Here using **Group by** to the State, **sum** to the transaction amount and **sort.values** by='Total\_Transaction\_amount’ so that can limit the output data to top 10 by using head(10) to the data frame.

California has made the highest contributions.

|  |  |
| --- | --- |
| STATE | TOTAL TRANSACTION AMOUNT |
| CA | 456037.0 |
| NJ | 342428.0 |
| NY | 308812.0 |
| PA | 143027.0 |
| CT | 80814.0 |
| MA | 78717.0 |
| TX | 59918.0 |
| FL | 47724.0 |
| MD | 43355.0 |
| IL | 42834.0 |

Que4: Which Occupation type are the Top 5 highest contributors for the candidate?

To know which occupation type people has contributed the maximum amount to the candidate. Here I have used **Group by** OCCUPATION and **sum** to the ['TRANSACTION\_AMT'] and sorted the data using **sort\_values** to get the only top 5 results.

It is quite surprising that not employed people has raised the highest total transaction amount in favor of candidate and all other type of occupation are given lesser amount.

|  |
| --- |
| OCCUPATION TOTAL TRANSACTION AMOUNT |
| NOT EMPLOYED 300992.0 |
| ATTORNEY 273068.0 |
| nan 108399.0 |
| RETIRED 102184.0 |
| EXECUTIVE 70897.0 |

Ques5: What is the count for each ethnicity category according to the Gender?

Lastly, to check total count of contributors according to gender, which is subdivided into ethnicity category. This shows that only female Hispanic contributors are supporting the candidate. And rest other ethnicity category is present in both male and female groups. Here used **Group by** to both the columns Gender and ethnicity and counted to the CMTE\_ID.

|  |  |  |
| --- | --- | --- |
| GENDER | CATEGORY | TOTAL\_COUNT |
| FEMALE | Asian | 78 |
| Hispanic | 6 |
| Unknown | 224 |
| White | 4318 |
| MALE | Asian | 70 |
| Unknown | 181 |
| White | 3524 |