

# **DATA ANALYST: SQL PORTFOLIO**

**PREPARED BY**

# Table of Contents

Introduction	3
Root Cause Analysis	4
Insights	5
Findings and Recommendations	10
Conclusion	15



# Introduction

I have been tasked by the Head of the Fundraising at the charity organization “Education for All” to present insights from donation in order to come up with fundraising strategies and increase donations.

My objectives are to:

- Increase the number of donors in our database
- Increase the donation frequency of our donors
- Increase the value of donations in our database.

The datasets EFO\_Donation\_Data and EFO\_Donor\_Data were presented to solve the business problem.

The following SQL commands were applied to analyze the data: SUM(), COUNT(), GROUP BY, WHERE, AND, OR, AVG(), ORDER BY, JOIN.

Root cause analysis was applied to get a deeper understanding of the problem. Insights were drawn from the datasets, data was visualized (using Tableau) and report was made from observations and presented to the team.

# Root Cause Analysis

Business problem: The funds generated through donations are not enough to actualize the objectives of the charity organization.

In the process of my analysis, the following questions helped me dig deeper into the problem based on the datasets provided:

- Are males donating more than females?
- In what states are we getting more donors?
- What do top 10 donors have in common?
- What are the incentives to attract donors?
- The people that donated the least, what do they have in common?
- How do the job field, university influence the amount that is donated?

While going through the datasets, I noticed some trends and symptoms which brought me to apply the “5 WHYs Approach” and led to the realization of the following questions:

1. Why are the funds through donations insufficient?
2. Why are they not getting donors that donate hugely and frequently?
3. Why are we not targeting the right audience?
4. Why are we not investigating who the right audience is?
5. Why do we not have the data on why people donate to our cause?

This approach shows that the root cause may be as a result of the insufficient information we have of our donors. Fields such as age, date of donation would have made our analysis easier but are not present in the datasets.

# Insights

Two datasets “EFO\_Donation\_Data and EFO\_Donor\_Data” were provided for the analysis. The table below shows the data present in each of the datasets.

EFO_Donation_Data	EFO_Donor_Data
id	Id
first_name	donation_frequency
last_name	University
email	Car
gender	second_language
job_field	favorite_color
donation	movie_genre
state	
shirt_size	

Both datasets were imported into SQLite



After going through the fields present in each dataset, I used the COUNT command to determine the total number of donor records we have in our dataset.

```
1 SELECT COUNT(*)
2 FROM Donation_Data;
```

To get the total amount of funds raised, I used the SUM() function.

```
1 SELECT SUM(donation)
2 FROM Donation_Data;
```

I applied the AVG() and ROUND() functions to determine the average amount of donation we receive per donor which was rounded to nearest whole number.

```
1 SELECT ROUND(AVG(donation),0)
2 FROM Donation_Data;
```

The MIN() function was used to determine the minimum amount of donated.

```
1 SELECT MIN(donation)
2 FROM Donation_Data;
```

To find the top 10 states with the highest amount of total donations and donors, COUNT(), SUM(), GROUP BY, ORDER BY, DESC and LIMIT were applied. The GROUP BY method grouped the donations by state and the ORDER BY was used to sort the data in descending order (DESC was used).

```
1 SELECT state, COUNT(*)
2 FROM Donation_Data
3 GROUP BY state
4 ORDER BY COUNT(*) DESC
5 LIMIT 10
```

```
1 SELECT state, SUM(donation)
2 FROM Donation_Data
3 GROUP BY state
4 ORDER BY SUM(donation) DESC
5 LIMIT 10;
```

COUNT(), SUM(), GROUP BY and ORDER BY was used to determine the count and sum of the donations grouped based on the genders present in our dataset.

SQLite

```
1 SELECT gender, COUNT(*)
2 FROM Donation_Data
3 GROUP BY gender;
4
```

gender	count(*)
Female	508
Male	492

```
1 SELECT gender, SUM(donation)
2 FROM Donation_Data
3 GROUP BY gender
4 ORDER BY sum(donation);
```

The query below was written to find the top 10 donors. The state, donation and gender were viewed in order to draw insights.

```
1 SELECT state, donation, gender
2 FROM Donation_Data
3 GROUP BY state
4 ORDER BY MAX(donation) DESC
5 LIMIT 10;
```

The donation frequency and their count were determined using the query below:

```
1 SELECT donation_frequency, COUNT(*)
2 FROM Donor_Data2
3 GROUP BY donation_frequency
4 ORDER BY COUNT(*) DESC;
```

JOIN was used to join the different columns of interest from the two datasets. In this case, the donation field from Donation\_Data, donation\_frequency from Donor\_Data2 were used to determine the donation frequency of the donors and the total amount donated grouped by the frequency.

```
1 SELECT Donation_Data.donation, Donor_Data2.donation_frequency, SUM(Donation_Data.donation)
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 GROUP BY donation_frequency
6 ORDER BY donation DESC;
```

The bottom 10 donors were determined using the query below. A condition was set using WHERE.

```
1 SELECT Donation_Data.gender, Donation_Data.donation, Donation_Data.state, Donor_Data2.donation_frequency
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE donation BETWEEN 5 AND 20
6 ORDER BY donation DESC
7 LIMIT 10;
```

WHERE and AND were used to filter and combine different conditions. In this case, female donors with university degrees (the first query) and without degrees (the second query) that donated above \$400.

```
1 SELECT Donation_Data.id, Donation_Data.gender, Donation_Data.donation, Donor_Data2.university
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE gender = "Female"
6 AND university != "NULL"
7 AND donation > 400
8 ORDER BY donation DESC;
```

```
1 SELECT Donation_Data.id, Donation_Data.gender, Donation_Data.donation, Donor_Data2.university
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE gender = "Female"
6 AND university IS NULL
7 AND donation > 400
8 ORDER BY donation DESC;
```

Query for male donors that donated above 400 with and without university degrees.

```
1 SELECT Donation_Data.id, Donation_Data.gender, Donation_Data.donation, Donor_Data2.university
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE gender = "Male"
6 AND university IS NULL
7 AND donation > 400
8 ORDER BY donation DESC;
```

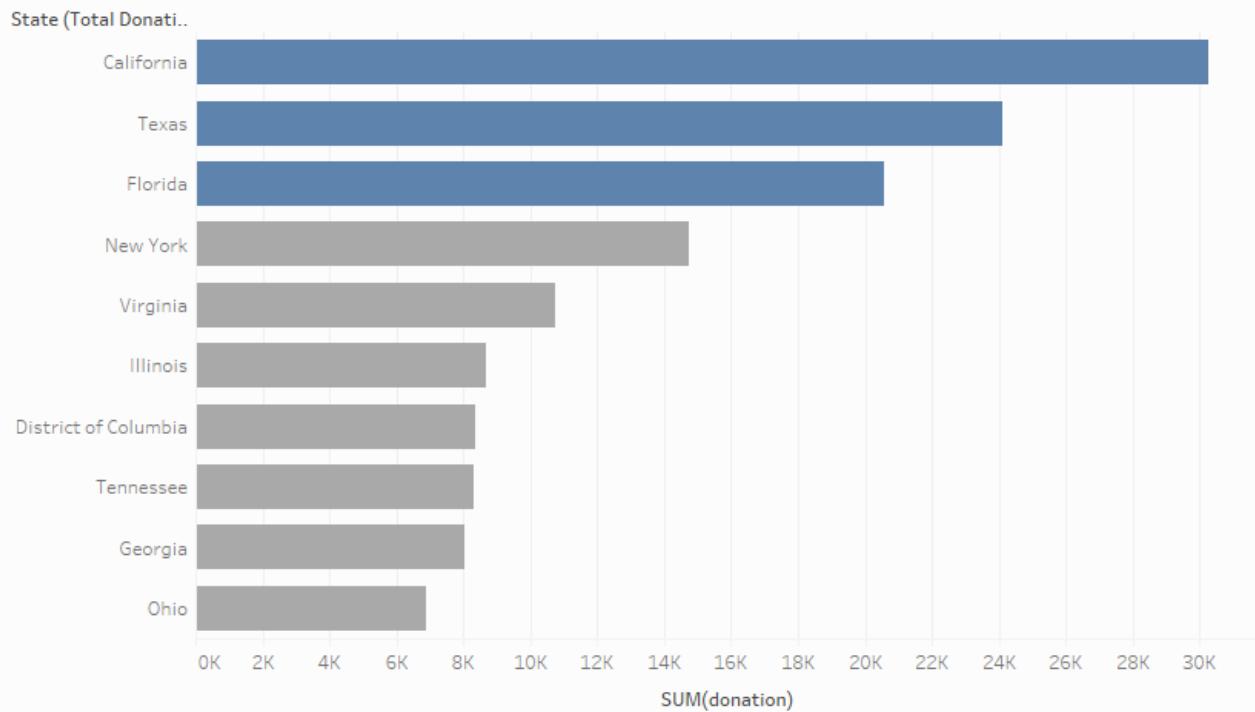
```
1 SELECT Donation_Data.id, Donation_Data.gender, Donation_Data.donation, Donor_Data2.university
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE gender = "Male"
6 AND university != "NULL"
7 AND donation > 400
8 ORDER BY donation DESC;
```

# Findings and Recommendations

Below are the insights drawn from our datasets:

- We have a record of 1000 donors in our database.
- The funds donated by the donors' amount to \$249,085
- The average amount donated by the donors is \$249
- The minimum amount donated is \$5
- The maximum amount donated is \$500 of which the donors are from New York and Michigan.

Amount of donations by State



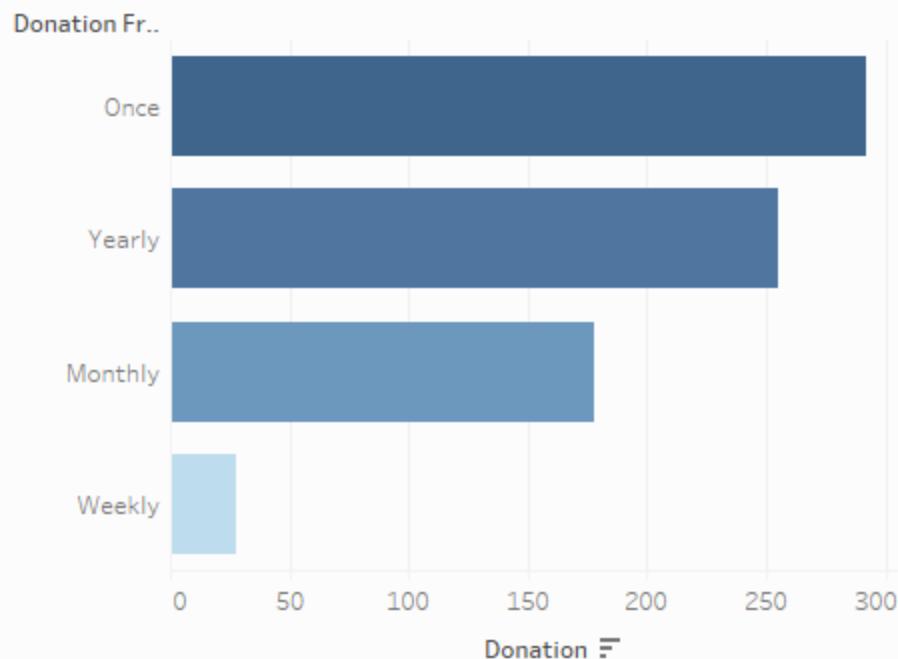
California, Texas, and Florida are the top leading states when it comes to the total amount of funds coming from those regions.

Gender	COUNT(donation)	SUM(donation)
Female	508	121457
Male	492	127628

Data shows that there are more female donors than male donors, however, men make bigger donations which may be as a result of higher earning power. Targeting individuals with higher earning power can help increase the amount of donations from both genders.

donation	donation_frequency	SUM(Donation_Data.donation)
292	Once	64586
255	Yearly	65667
178	Monthly	59680
28	Weekly	59152
<b>Total = 753</b>		

## Donations count by donation frequency



Out of our 1000 donors, only 753 donors are active and donate at different frequencies-once, weekly, monthly, and yearly. The amount of funds accrued from donors with yearly frequency is higher than that of the others with different frequencies. We should find ways to retain and increase the frequency of donations.

## Donors with the least donations

gender	donation	state	donation_frequency	car	university
Male	19	Michigan	Weekly	Land Rover	Kincla
Male	19	Missouri	Monthly	Cadillac	null
Female	19	California	Weekly	Ford	Layne
Female	18	California	Weekly	Toyota	Capini
Male	17	California	Yearly	Ford	Poytres
Male	17	Minnesota	Weekly	Buick	Rochewell
Female	17	New York	Once	Chevrolet	null
Female	17	Texas	Monthly	Toyota	Cove
Male	17	Alabama	Yearly	Toyota	Llewellen
Male	16	Kansas	Weekly	Kia	Havers

## Top 10 donors with the highest donations

gender	donation	state	donation_frequency	car	university
Male	500	Michigan	Yearly	Honda	Walasik
Male	500	New York	Monthly	Pontiac	Leithgoe
Female	499	Virginia	Yearly	Ford	Sparhawk
Female	499	Delaware	Yearly	Ford	Antoszewski
Male	498	Wisconsin	Monthly	Volvo	Trotton
Male	497	New York	Weekly	Maserati	Rockcliffe
Male	494	California	Weekly	Kia	Cominetti
Female	494	California	Monthly	Maybach	Coates
Male	494	California	Monthly	Dodge	Baumber
Male	493	Maryland	Monthly	Infiniti	Armatidge

It can be observed that the car that donors drive is not a determinant of the amount of donations they make.

## Donations above \$400

	Female	Male
With university	66	80
Without university	28	26

We have 28 females, 26 males without a university degree and 66 females, 80 males with university degrees that donated above \$400. The university they attended is not a determinant of the amount donated as a variety of universities can be seen amongst our top donors in our data. From our data, we can see that educated people are more likely to donate to our cause which may be because they can identify with it. As a result, we can target more people with university degrees since a vast majority of our top donors have degrees.

# Conclusion

To achieve these three core objectives of the organization “Education for all”: increase the number of donors in our database, increase the donation frequency of the donors and increase the value of donations in our database; I have employed the following tools to analyze the current database presented by the organization: Root cause analysis, Query functions and Tableau; with these tools, I was able to draw insights and came up with recommendations for the organization.

In terms of insights, the following information was retrieved from the database: There are a total of 1000 donors recorded by the organization, a total amount of \$249,085 was donated, the minimum amount donated was \$249, the maximum amount donated was \$500, and the average amount donated was \$249. Of all 1000 donors, only 753 are active. There were more women donors than men, even if men donated a higher amount on average. Other important findings include: there is little to no correlation between the frequency of donation and any other data fields, most donors for the organization are one-time donors, and the top three states from which the highest donations were made are California, Texas, and Florida.

Following the insights gathered, I recommend the following: To increase donation frequency, the organization must find new ways to get donors to donate more frequently, and organize fundraising events in seasons when donations are seen to be higher; To increase the number of donors, the organization should increase awareness in states where there is low turnout of donors; To increase the value of donations, it is important that

the organization create incentives beyond T-shirts that will incentivize donors to make larger donations, it is also important to target people in regions with a higher earning power as they are more capable of making higher donations.