

DATA ANALYST: SQL PORTFOLIO

PREPARED BY

Geet Tandon



Professional Background

I have completed my 10th standard at one school and my 12th standard in commerce with maths at another. I'm a Bachelors of Commerce Honors graduate from Delhi University. I completed my 15 years formal education in Delhi, India in 2018. After that, I found my interests in Data Analysis and Data Science.

Due to my interests, I completed courses like Data Analysis in Python from Dataquest. I'm certified as Associated Data Analyst from DataCamp. I have completed training in Excel, Python, SQL, Tableau, and PowerBI.

Since 2019, I have been working as an Independent Contractor on Appen completing projects related to Data Annotation, Data collection, etc.

Parallel to working, I have completed internships as a Data Science intern at AINE AI, Data Analyst at EntryLevel, Data Collection intern at Central Drug Research Institute.

Portfolio Outline



Professional Background	1
Table of Contents	2
Introduction	3
Root Cause Analysis	4
Insights	5
Findings and Recommendations	16
Conclusion	17



Introduction

A charity "Education for All" has employed me as a Data Analyst and I have been asked by the Head of Fundraising to present the data on donor insights and donation rates.

I had two weeks of time to complete this report. Our team is having a fundraising strategy meeting for the following year and I need to present this report there to inform the team about the insights and ways to increase donations. Business problem identification is the first process in Data Analysis. Our fundraising team wants to:

1. Increase the number of donors in your database.
2. Increase the donation frequency of your donors.
3. Increase the value of donations in your database.

This business problem can be narrowed down to "getting fewer donations" and we want to ways of increasing the donations so that we can help more.

Data is from 1000 donors living in the USA. Data provided gives information about donations from donors, where they live, their job field, their donation frequency, their likings, etc. I have also imported USA population data from external sources.

I will be using SQL to import the data and analyze it to find insights. I will use Tableau to present those insights visually.

Root Cause Analysis

Root Cause Analysis is a method of problem-solving used for identifying the root causes of a business problem. Root Cause Analysis can be done by asking '**WHY**' 5 times. Asking why 5 times will get us to the root cause or much closer to it.

The process I will follow:

- I will first use the SQL to find the initial insights from the data such as who donates more, states wise donations, donor demographics, etc.
- Doing this initial analysis will help me to get to know the data better and frame the Root Cause Analysis questions.
- Initial insights that I found with SQL code and visualizations are included in the insights tab.
- SQL code and visualizations of 5 '**WHYs**' will also be included in the insights section.

The business problem is: '**Education for All charity is getting fewer donations.**'

Based on the initial insights, I found two reasons for the first 'Why' and I asked these 5 'Why' questions to both of the reasons to get to the root cause of our business problem. Below are just the questions. The answer I found for them with SQL code and Tableau Visualizations will be included in the insights section.

1. Why are our charity is getting fewer donations?

2.1 Why number of donations affect donations?

2.2 Why does donation frequency affect donations?

3.1 Why do some states have fewer donors?

3.2 Why do most donors donate on a 'Once' or 'Yearly' frequency?

4.1 Why fewer donors from low-populated state affect donation?

4.2 Why do most donors from some job fields donate on a 'Once' or 'Yearly' frequency?

5.1 Why fewer donation values from low-populated states.

5.2 Why do more donors from this demographic donate less?

Insights

Before starting with insights, I'm going to create a **VIEW** and name it 'data' in SQL to combine both the datasets provided into one.

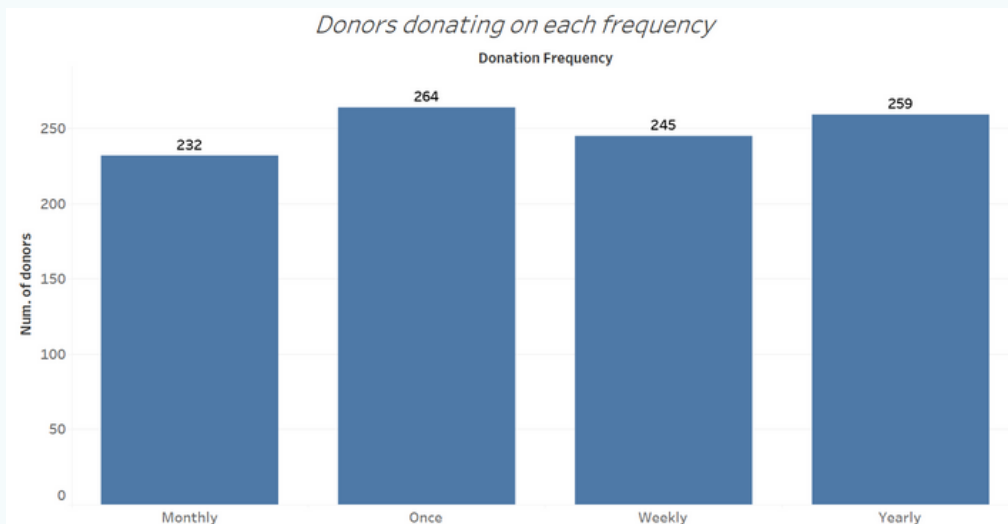
VIEW in SQL is a virtual table based on the result set of an SQL statement. SQL query for this is:

```
CREATE VIEW data AS
SELECT *
FROM Donation_Data
INNER JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id;
```

I am going to start with some initial insights to get to know the data better. SQL query and Tableau visualization will be accompanying each insight.

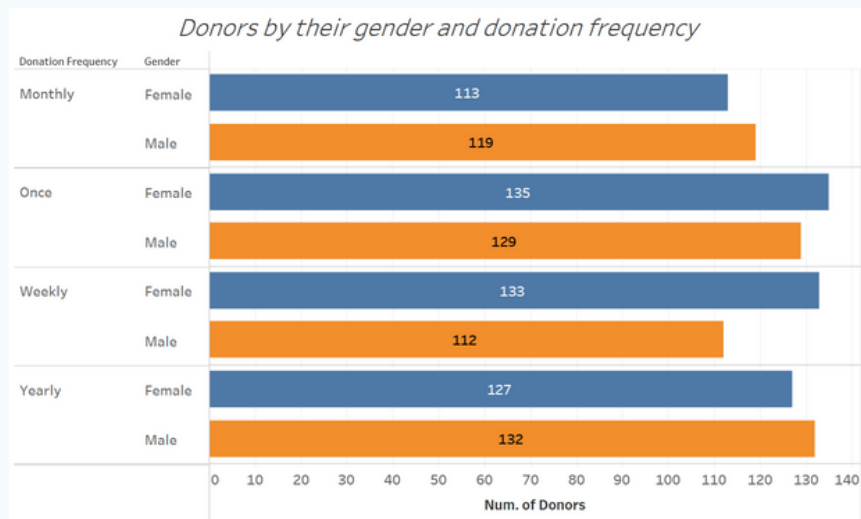
- **Maximum donors donate once, least number of donors donate monthly.**

```
SELECT donation_frequency,
       COUNT(*)
FROM data
GROUP BY donation_frequency;
```



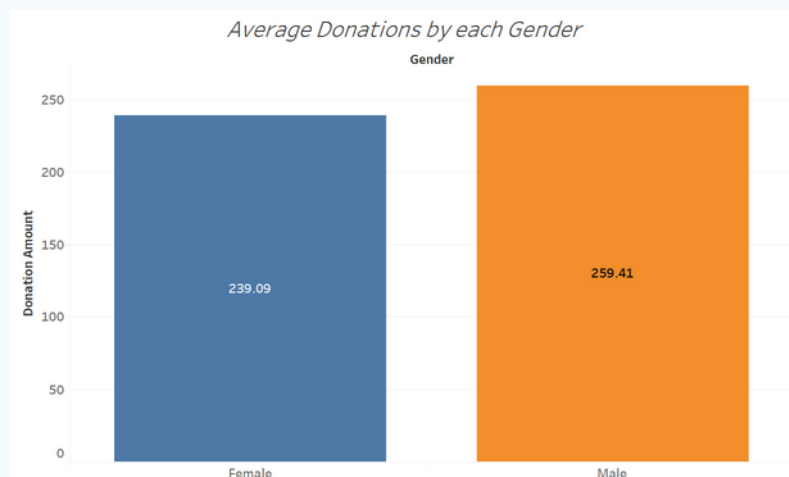
- **Female donors donate more once and weekly, whereas male donors donate more on a monthly and yearly frequency.**

```
SELECT gender,
       donation_frequency,
       COUNT(*)
FROM data
GROUP BY gender, donation_frequency
ORDER BY donation_frequency, gender;
```



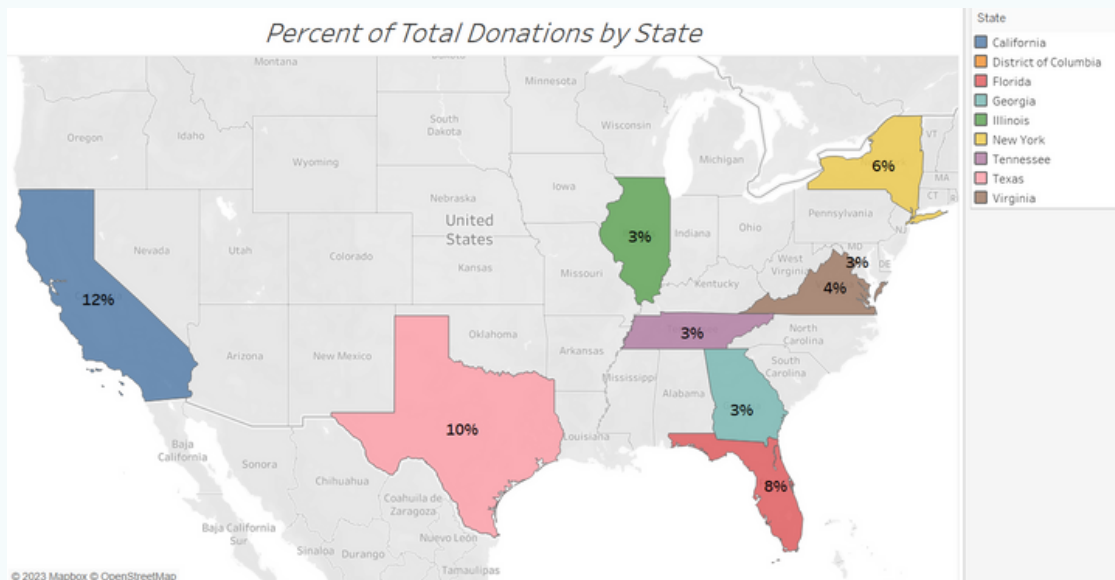
- On average males donate 8% more than females.

```
SELECT gender,
       ROUND(AVG(donation), 2) AS average_donation
FROM data
GROUP BY gender;
```



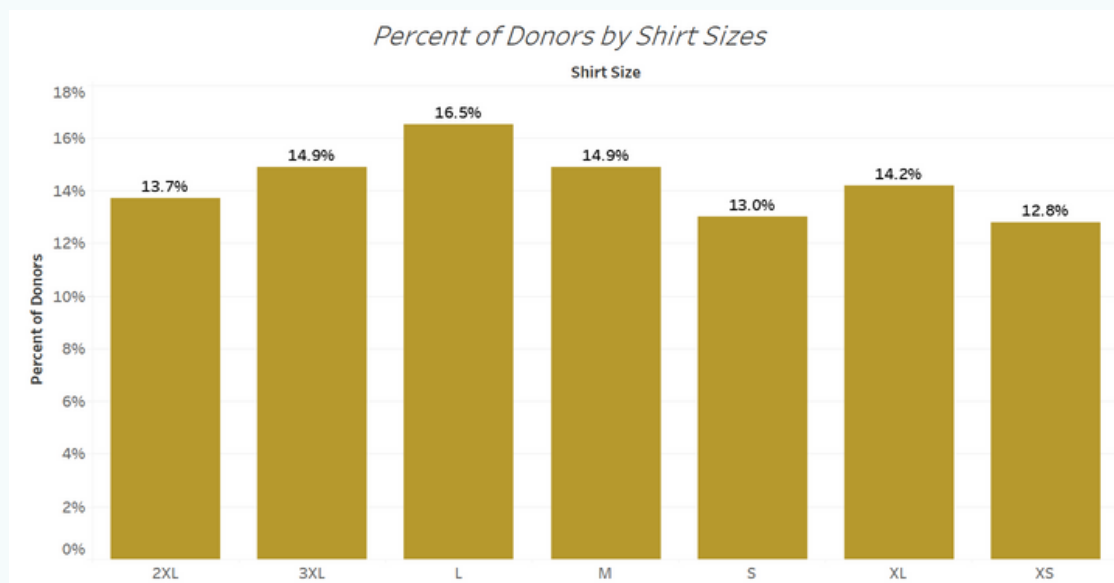
- Maximum donations come from California State. California, Texas, Florida, and New York constitute 36% of the total donations received.

```
SELECT state,
       ROUND(CAST(SUM(donation) AS FLOAT) /
             (SELECT SUM(donation) FROM data) * 100, 2) AS percent_of_total_donation
FROM data
GROUP BY state
HAVING percent_of_total_donation > 3.0
ORDER BY percent_of_total_donation DESC;
```

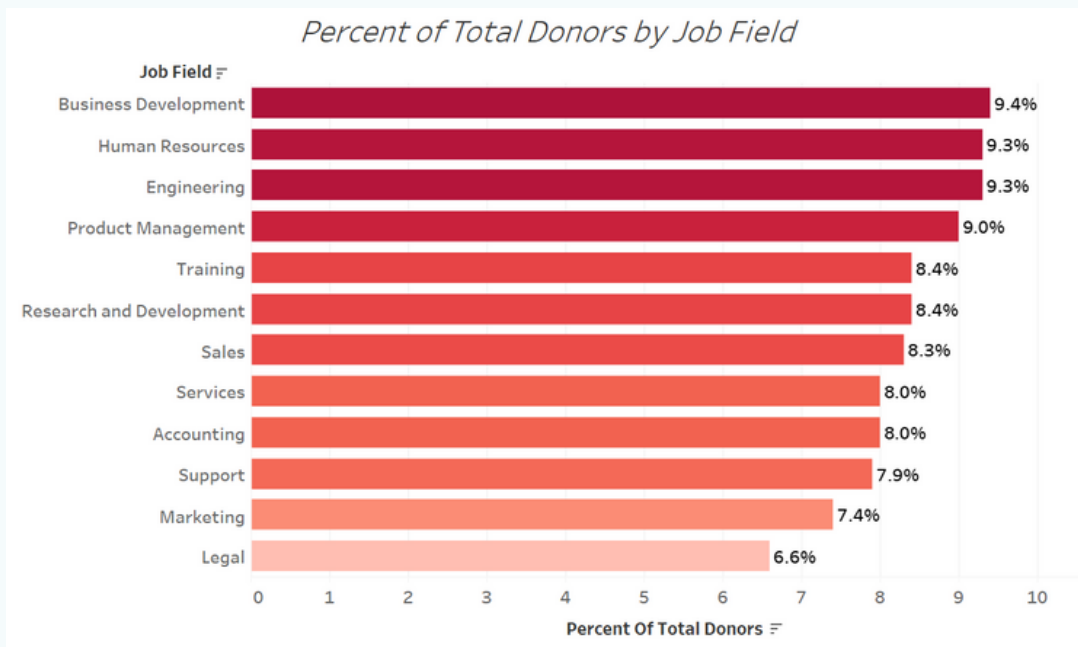
- **16.5% of the total donors are having shirt size 'L'.**

```
SELECT shirt_size,
       ROUND(CAST(COUNT(*) AS FLOAT) /
             (SELECT COUNT(*) FROM data) * 100, 2) AS percent_of_total_donors
FROM data
GROUP BY shirt_size;
```



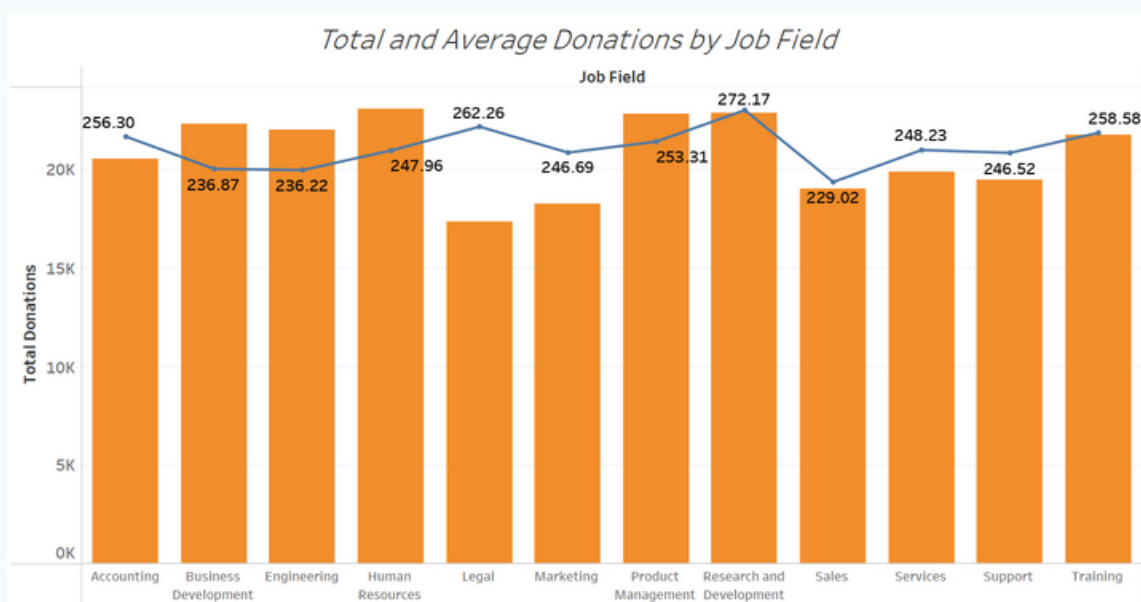
- **Business Development, Human Resources, Engineering, and Product Management job fields constitute 37% of the total donors in our dataset.**

```
SELECT job_field,
       ROUND(CAST(COUNT(*) AS FLOAT) /
             (SELECT COUNT(*) FROM data) * 100, 2) AS percent_of_total_donors
FROM data
GROUP BY job_field
ORDER BY percent_of_total_donors DESC;
```

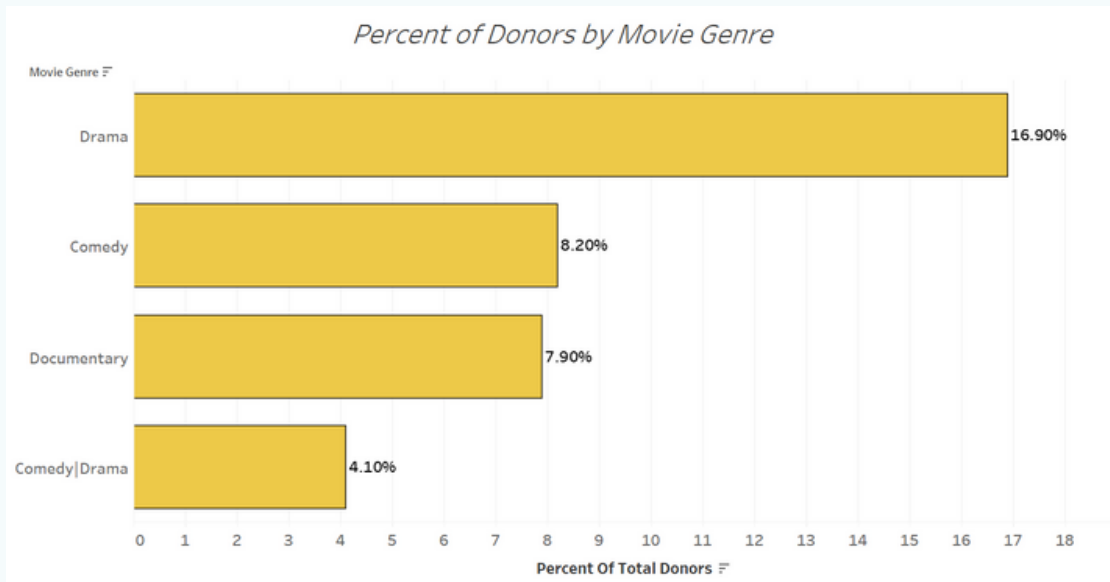
- **Maximum total donations come from the Human Resources job field, whereas on average donors from Research & Development job field donate the highest.**

```
SELECT job_field,
       SUM(donation) AS total_donations,
       AVG(donation) AS avg_donations
FROM data
GROUP BY job_field
ORDER BY total_donations DESC;
```



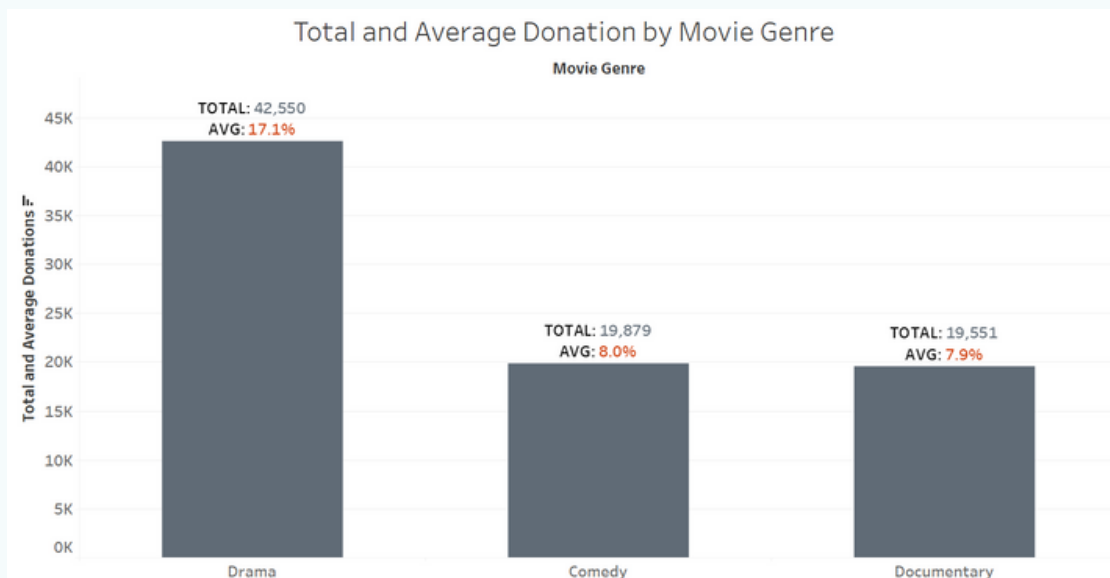
- **Approximately 17% of the total donors like the Drama movie genre.**

```
SELECT movie_genre,
       ROUND(CAST(COUNT(*) AS FLOAT) /
             (SELECT COUNT(*) FROM data) * 100, 2) AS percent_of_total_donors
FROM data
GROUP BY movie_genre
HAVING percent_of_total_donors > 4.0;
```



- **Maximum donors liking the Drama movie genre are the highest donors as well. Donors like Drama, Comedy, and Documentary movie genre donates 33% of the total donations.**

```
SELECT movie_genre,
       SUM(donation),
       ROUND(CAST(SUM(donation) AS FLOAT) /
             (SELECT SUM(donation) FROM data) * 100, 2) AS percent_of_total_donation
FROM data
GROUP BY movie_genre
HAVING percent_of_total_donation > 4.0;
```



These were the initial insights I found during my analysis. These helped me with finding Root Cause as well of our Business Problem.

Along with '5 Whys' I wrote SQL code and made visualizations to go deeper into the Root Cause of our Business Problem. We will look into them now.

5 WHY ANALYSIS

1. Why our charity is getting fewer donations?

- Because of the number of donors.
- Because of donation frequency.

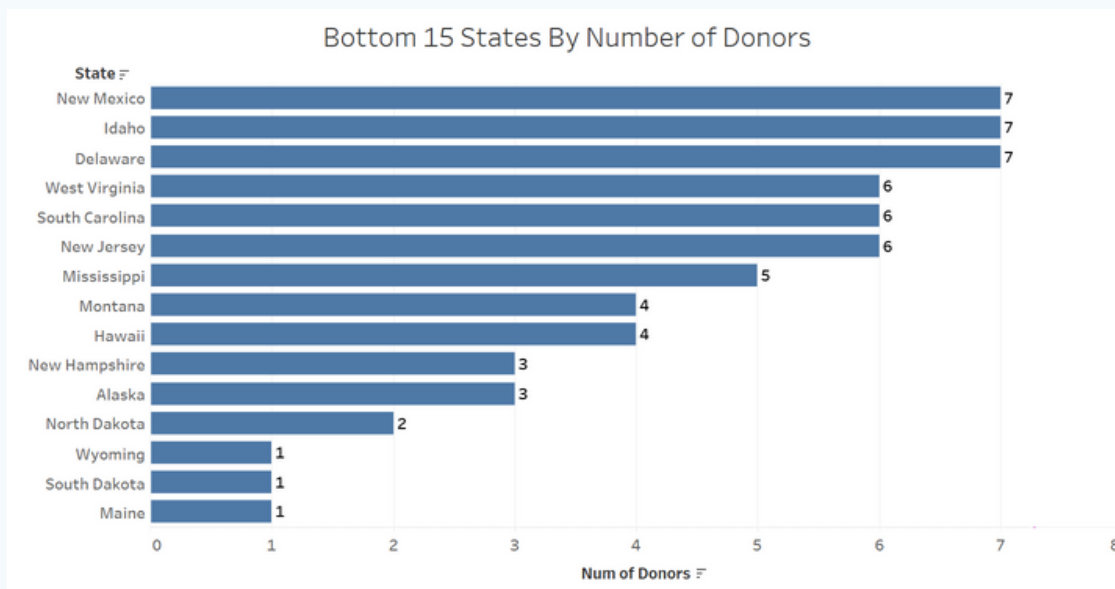
For the first '**WHY**', I got two reasons. I will go deeper into them one by one and try to find the root cause for both of these reasons.

Both of these reasons will get their separate '**4 WHYS**'. I will start with the first reason.

2.1. Why number of donors affect donations?

Ans. Because some states have fewer donors.

```
SELECT state,  
       COUNT(*) AS num_of_donors  
FROM data  
GROUP BY state  
ORDER BY num_of_donors ASC  
LIMIT 15;
```

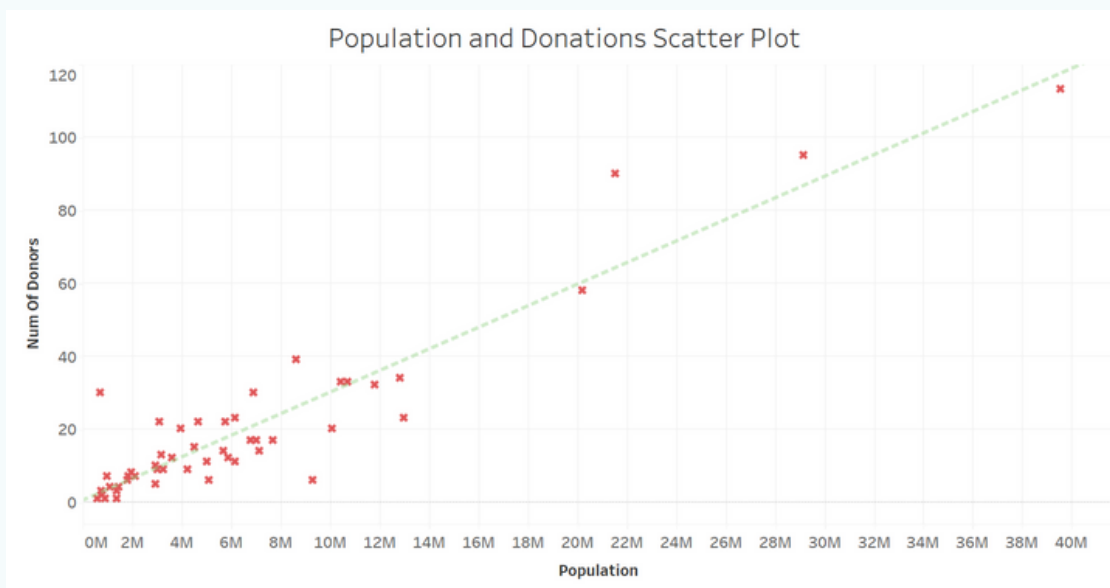


3.1. Why do some states have fewer donors?

Ans. Because the number of donors from the state and the population of the state is correlated. Thus fewer donors from low-populated states. To show this I'm going to the US_state_pop data imported externally that contains state-wise population.

```
WITH donors_count AS (
    SELECT state,
           COUNT(*) AS num_of_donors
    FROM data
    GROUP BY state
)
```

```
SELECT US_state_pop.State,
       population,
       num_of_donors
FROM US_state_pop
INNER JOIN donors_count
ON US_state_pop.state = donors_count.state;
```

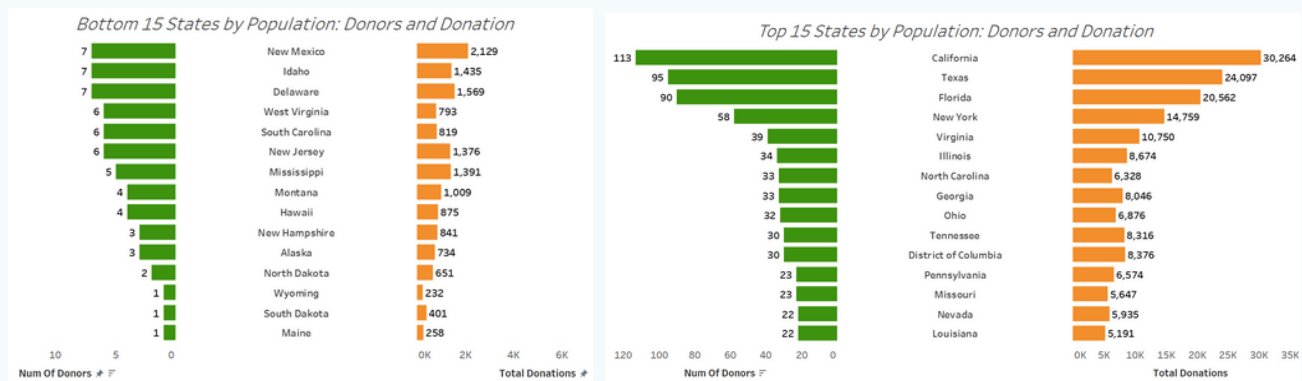


4.1. Why do fewer donors from low-populated states affect donations?

Ans. Because fewer donors lead to fewer donation values. We can see this in the visualization below. I have created two visualizations, one with states with fewer donors and one with states with more donors.

```
SELECT state,
       COUNT(*) AS num_of_donors,
       SUM(donation) AS total_donations
FROM data
GROUP BY state
ORDER BY num_of_donors ASC
LIMIT 15;
```

```
SELECT state,
       COUNT(*) AS num_of_donors,
       SUM(donation) AS total_donations
FROM data
GROUP BY state
ORDER BY num_of_donors DESC
LIMIT 15;
```

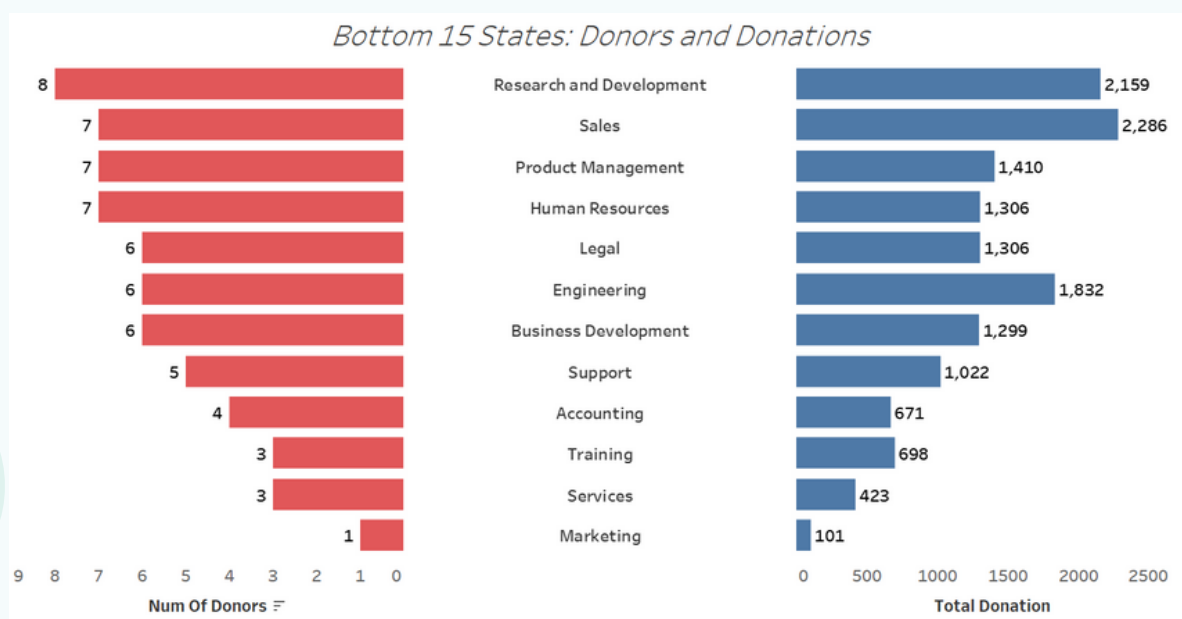


5.1. Why fewer donation values from low-populated states?

Ans. Because donors from job fields: Research & Development and Sales are donating maximum in less populated states. Comparing this to our initial insights, we see that these job fields are not even in the top 5 maximum donating job fields. This shows that there are less high donating donors in these states.

```
WITH bottom_states AS (
    SELECT state,
           COUNT(*) AS num_of_donors
    FROM data
    GROUP BY state
    ORDER BY num_of_donors ASC
    LIMIT 15)
```

```
SELECT job_field,
       COUNT(*) AS num_of_donors,
       SUM(donation) AS total_donation
FROM data
WHERE state IN (SELECT state FROM bottom_states)
GROUP BY job_field
ORDER BY num_of_donors DESC;
```

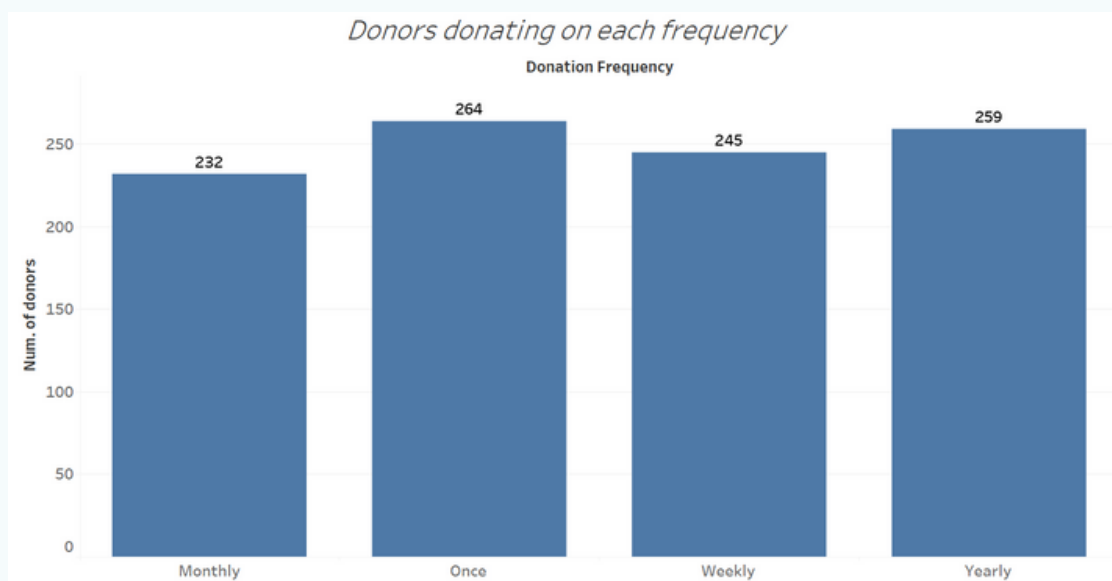


We dig deeper into our problem and found the first reason's root cause. Now, I'm going to start with the second reason **'4 WHYS'**.

2.2. Why does donation frequency affect donation values?

Ans. Because most donors donate on a 'Once' or 'Yearly' frequency.

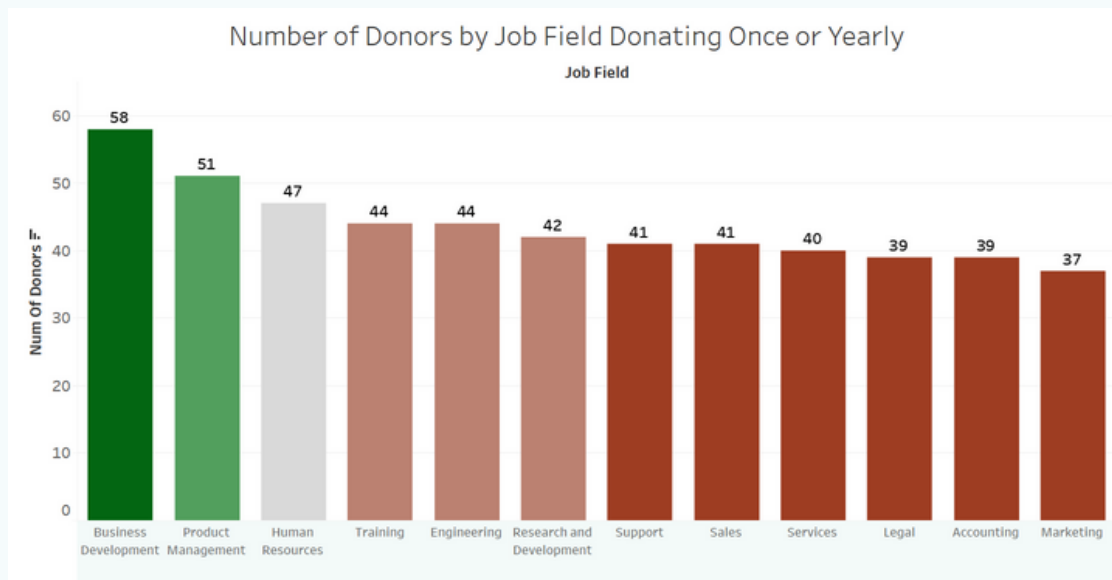
```
SELECT donation_frequency,  
       COUNT(*)  
FROM data  
GROUP BY donation_frequency;
```



3.2 Why do most donors donate on a 'Once' or 'Yearly' frequency?

Ans. Because donors from job fields: Business Development, Product Management, and Human Resources donated more 'Once' or 'Yearly'. These are 3 out of the top 5 maximum donating job fields as we saw in initial insights.

```
SELECT job_field,  
       COUNT(*) AS num_of_donors  
FROM data  
WHERE donation_frequency IN ('Once', 'Yearly')  
GROUP BY job_field  
ORDER BY num_of_donors DESC;
```

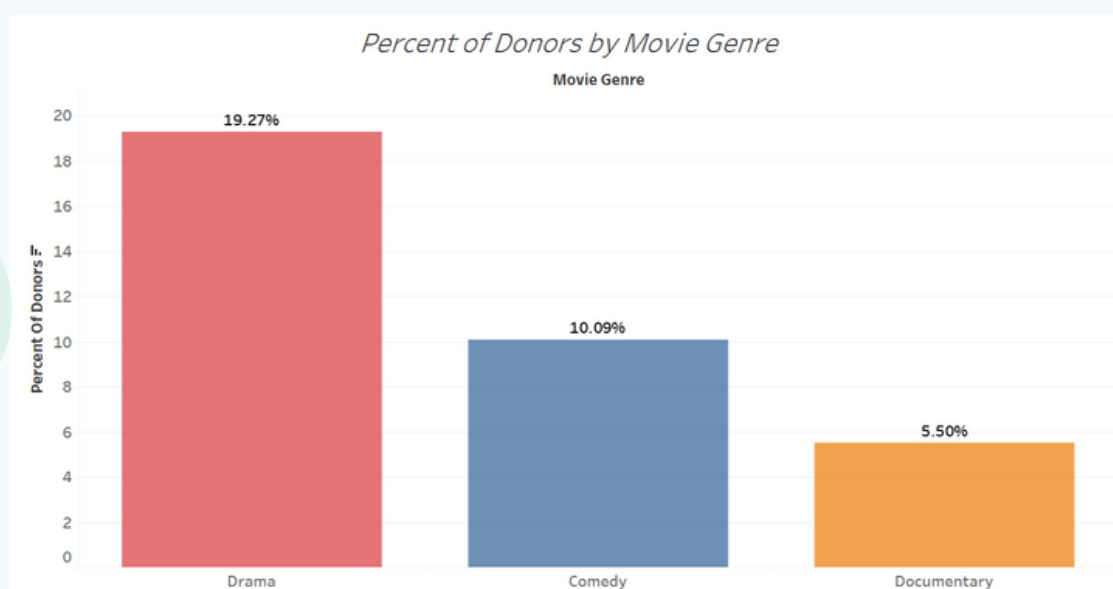


4.2. Why do donors from these job fields donate more on a 'Once' or 'Yearly' frequency?

Ans. Because digging deeper, I found most donors like the Drama movie genre in these job fields and are donating more on a 'Once' or 'Yearly' frequency.

```
WITH donor_granular AS (
    SELECT movie_genre,
           COUNT(*) AS num_of_donors
    FROM data
    WHERE job_field IN ('Human Resource', 'Business Development', 'Product Management')
           AND donation_frequency IN ('Once', 'Yearly')
    GROUP BY movie_genre
)
```

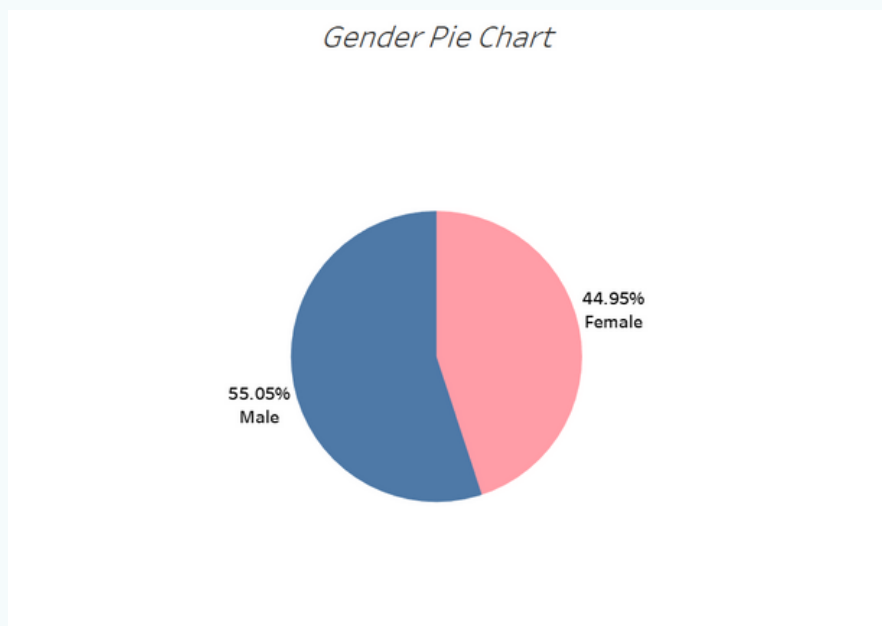
```
SELECT movie_genre,
       ROUND(CAST(num_of_donors AS FLOAT) /
             (SELECT SUM(num_of_donors) FROM donor_granular) * 100, 2) AS percent_of_donors
FROM donor_granular
GROUP BY movie_genre
HAVING percent_of_donors > 4.0;
```



5.2. Why do more donors from this demographic donate less?

Ans. Doing some deeper analysis, I found that most of the donors in this demographic are males who generally donate more 'Yearly'. Females are fewer but they donate more 'Weekly' than males as we saw in our initial insights.

```
WITH gender_donation_frquency AS (  
    SELECT gender  
    FROM data  
    WHERE job_field IN ('Human Resource',  
                        'Business Development',  
                        'Product Management')  
    AND donation_frequency IN ('Once', 'Yearly')  
)  
  
SELECT gender,  
    ROUND(CAST(COUNT(*) AS FLOAT) /  
        (SELECT COUNT(*) FROM gender_donation_frquency) * 100, 2) AS percent_of_donors  
FROM gender_donation_frquency  
GROUP BY gender;
```



These are the '**5 WHYS**' for both of the reasons and they helped me to find the root cause of the Business Problem.

In the Findings and Recommendation section, we are going to talk about this in more detail.

Findings and Recommendations

Doing Root Cause Analysis helped me to find the reasons why our charity 'Education for All' is getting fewer donations.

We can pin the reasons to be:

- Fewer donors from low-populated states lead to fewer donations.
- More donors donate on a 'Once' or 'Yearly' frequency.
- Some donors in a few job fields donate less than others.
- Fewer Females than Male donors but Male donors donate more on average.
- Drama movies are the most liked but donors liking them donate less occasionally.
- Donations are concentrated in a few states only.
- Most donors have a shirt size of 'L'.
- Male donates more 'Yearly' whereas Female donors donate more 'Weekly'.

Based on these findings and Root Cause Analysis, I would recommend certain actions to increase donations:

- Getting more data for deeper analysis such as income, age, dependents of donors, date of donation, etc.
- Focussing more on Female donors as they donate more 'Weekly'.
- Focussing on donors from the 'Engineer' job field as their average donation value is high.
- Focussing more on donors liking the 'Comedy' and 'Documentary' movie genre as they donate more 'Weekly'.
- Focussing on low-populated states as well as I believe that is an 'untapped' resource.
- One last recommendation is to introduce 'Helping Points'. This doesn't provide any monetary benefits but is given to donors whenever they donate which will make them feel nice and that they are doing something for society.

Conclusion

This was the report on a project for our charity '**Education for All**' where I did an analysis on SQL and visualized using Tableau to find the Root Cause of the **Business Problem: Getting fewer donations.**

I found some reasons and have used appropriate methods to analyze and visualize. I also have given my findings and recommended steps to take that can solve our Business Problem.

While doing this project, I got the opportunity to learn more about topics like SQL, Tableau, Report Writing, and how to find the Root Cause of Business Problems.