

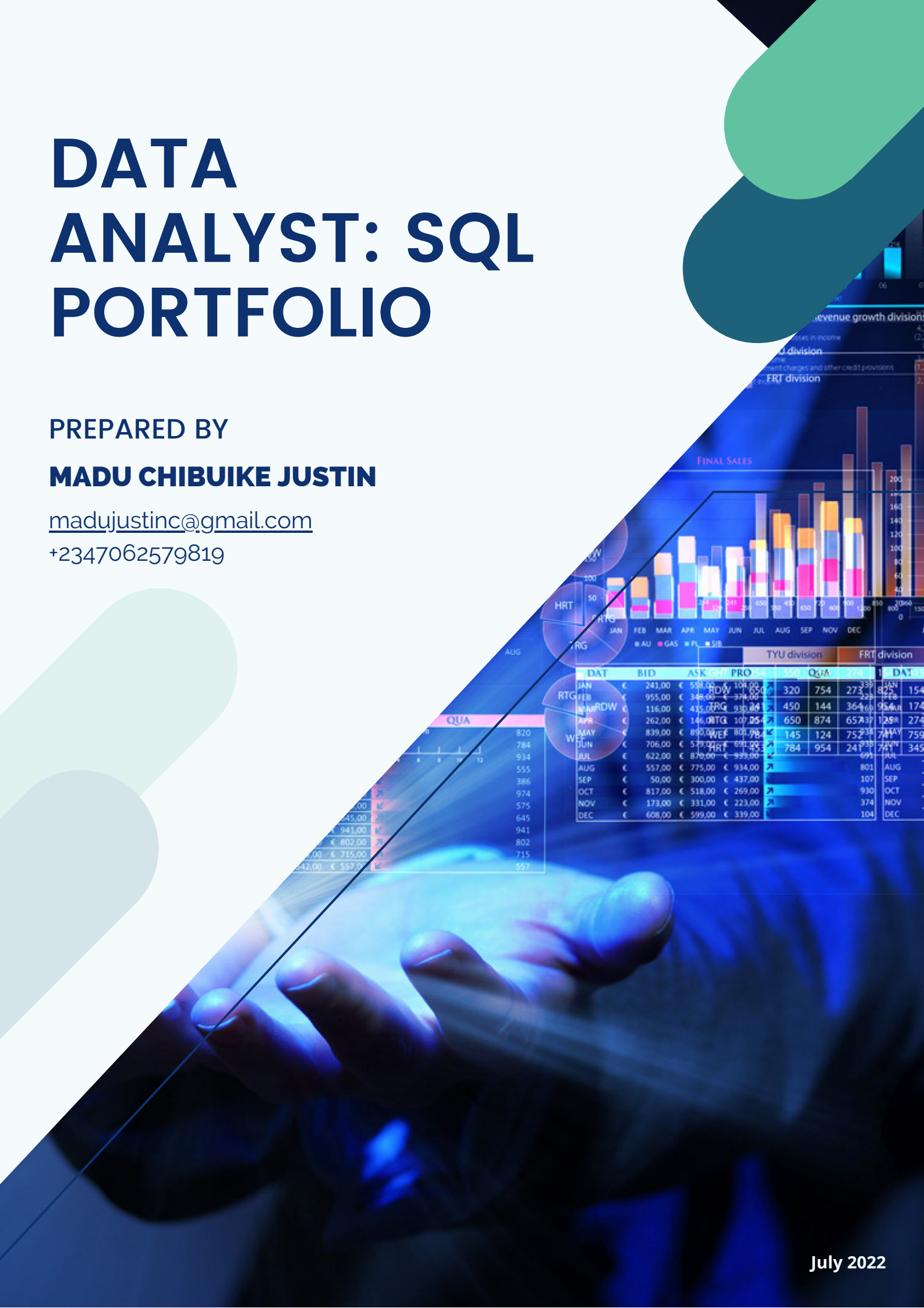
DATA ANALYST: SQL PORTFOLIO

PREPARED BY

MADU CHIBUIKE JUSTIN

madujustinc@gmail.com

+2347062579819



Professional Background

Passionate, motivated, teamwork-oriented, and enthusiastic data analyst who strives to pose, answer questions and solve business problems with quantitative-driven insights. Through the development of personal projects, valuable training, and work experience.

I have learned the importance of having an iterative, hypothesis-oriented approach to analysis. With significant experience in data visualization, transformation, creating reports, creating dashboards, and sharing insights.

Highly educated, possessing a bachelor's degree in Chemical Engineering and professional certification in;

- Data Analysis - Power BI/SQL/Azure/Excel/Alteryx by UDEMY,
- Data Analytics with Excel by Simplilearn
- Data Analyst 1 and 2 by EntryLevel
- SQL - MYSQL for Data Analytics and Business Intelligence by UDEMY and a
- Certified Jobberman Certificate of Honor on Soft Skills.

Possess demonstrated skills in Engineering Safety, Digital marketing, graphic designs, Administration Assistant, etc

Portfolio Outline

| | |
|------------------------------|----|
| Professional Background | 1 |
| Table of Contents | 2 |
| Introduction | 3 |
| Root Cause Analysis | 5 |
| Insights | 6 |
| Findings and Recommendations | 14 |
| Conclusion | 19 |
| Appendix | 20 |





Introduction

Situation: In a hypothetical situation as a Data Analyst working for a charity organization, EDUCATION FOR ALL. I have been asked by the Head of Fundraising to present data on donor insights and donation rates.

Task: The purpose of this report is to increase the number of donors in the database, Increase the donation frequency of the donors and Increase the value of donations in the database. Data given to me were datasets of EFO_Donation_Data and EFO_Donor_Data.

Action: I uploaded the datasets in SQLite Online for analysis. I applied SQL commands like JOIN, ORDER BY, WHERE, BETWEEN, AND, OR, SUM(), COUNT(), AVG(), GROUP BY, HAVING to analyze data. I used the data sets FOR_Donation_Data and FOR_Donor_Data to answer the business problem and finally created charts and dashboards using Tableau Public for visualizations. I have two weeks to work on this report.

Result: In my report, I used the 5 WHY's (Root Cause Analysis) to understand the problem and ask the right questions. As a result, I found out crucial insights from provided datasets, prepared visualizations, and reports to my team.

The Problem

- The business problem is that there has been a decline in donations for the charity organization, EDUCATION FOR ALL. It is very hard to grow to fund. Therefore, they must search for more donors, who can donate weekly and monthly and will increase the value of their donations.
- I had to work on this project and present my report within two weeks.
- To understand the problem, existing databases of Donors and Donations need to be analyzed. Furthermore, the data should be presented in tables and visualizations of the datasets in SQLite and Tableau Public.
- In addition to understand the business problem, I decided to ask some questions like
 - How many donors do we have in the existing database?
 - What is the amount of their donations?
 - What is the frequency of the donations?
 - Where are the donors from?
 - What are their genders?
 - Who are the top 10 donors?
 - Are donations dependent on gender, job field, university degree, or car?
- The impact if the problem is not solved is that it will lead to continual decrease in donations which may eventually lead to the organization having only a few donors.

Root Cause Analysis

Root cause analysis is the process of discovering the root causes of problems to identify appropriate solutions. The first goal of root cause analysis is to discover the root cause of the decline in donations by the donors. The second goal is to fully understand how to fix, compensate, or learn from any underlying issues within the root cause. The third goal is to apply what we learn from this analysis to systematically prevent future issues or repeat successes.

In this work, I will use the 5 Whys technique to determine the root cause by repeatedly asking the question “Why” until I arrive at the root cause of why there is a decline in donations.

- Why is there a decline in donations? Because of poor or lack of donors retention.
- Why is there a lack of donors retention? Because there is no sufficient information on donors' details and feedback?

Insights

Data Design

I have been provided with 2 relational databases such as: FO_Donation_Data and EFO_Donor_Data to answer the business problem. The SQLite Database Management System was used to find out the main insights.

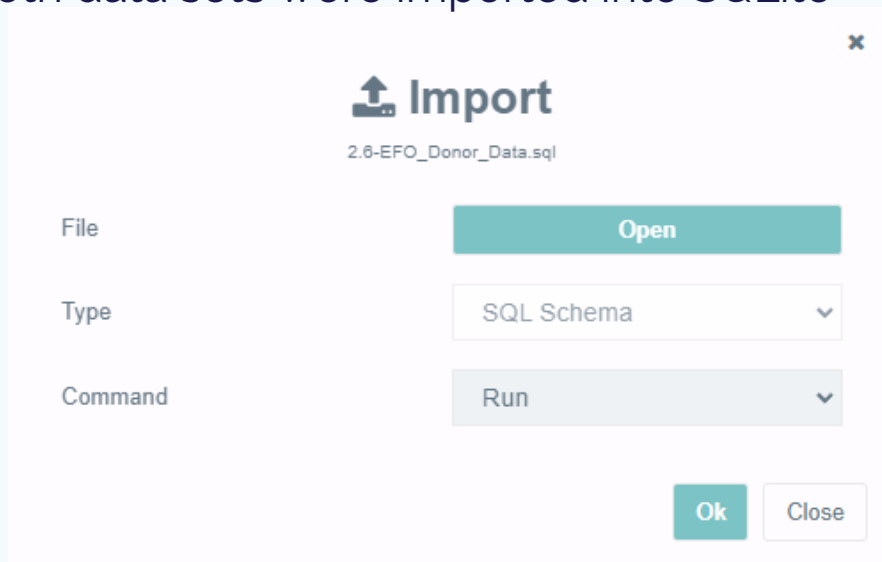
Donation Dataset includes such data:

```
id INT,  
first_name VARCHAR(50),  
last_name VARCHAR(50),  
email VARCHAR(50),  
gender VARCHAR(50),  
job_field VARCHAR(50),  
donation INT,  
state VARCHAR(50),  
shirt_size VARCHAR(50)
```

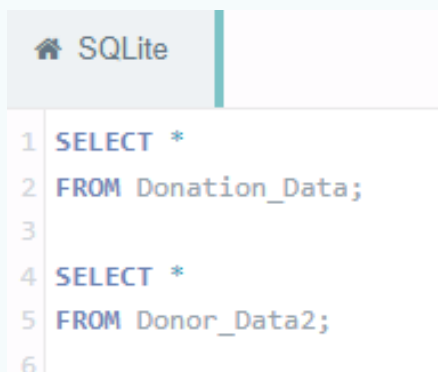
Donor Dataset includes such data:

```
id INT,  
donation_frequency VARCHAR(50),  
university VARCHAR(50),  
car VARCHAR(50),  
second_language VARCHAR(50),  
favourite_colour VARCHAR(50),  
movie_genre VARCHAR(50)
```

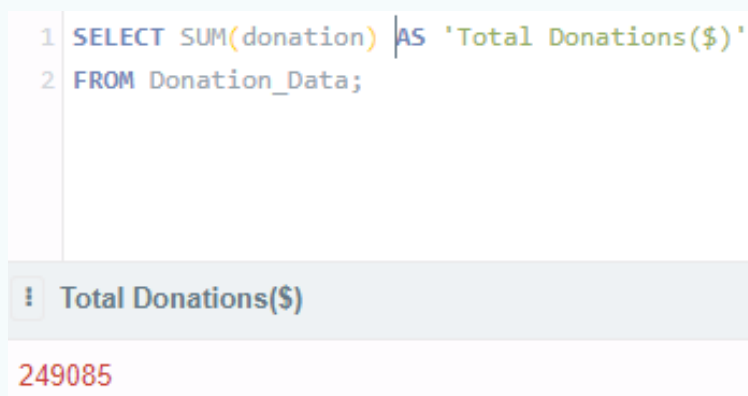
Both data sets were imported into SQLite



After importing the dataset into SQLite, I used the SELECT statement to view the datasets.



I started querying the database to get insights on the datasets. I first checked the total SUM of donations made by donors by using the SELECT statement and aggregate function SUM()



I checked the number of donors in the database using the SELECT statement and aggregate function COUNT()

```
1 SELECT COUNT(donation) AS 'Number of Donors'
2 FROM Donation_Data;
```

| ! Number of Donors |
|--------------------|
| 1000 |

I now checked the maximum and minimum donations by our donors by using the SELECT and aggregate functions MAX() and MIN()

```
1 SELECT MAX(donation) AS 'Highest Donation',
2 MIN(donation) AS 'Least Donation'
3 FROM Donation_Data;
```

| ! Highest Donation | Least Donation |
|--------------------|----------------|
| 500 | 5 |

I also checked the AVERAGE Donations made by our donors by using the AVG() function

```
1 SELECT AVG(donation) AS 'AVG Donation'
2 FROM Donation_Data;
```

| ! AVG Donation |
|----------------|
| 249.085 |

COUNT and SUM of Donors by Donation Frequency

The JOIN function allows us to combine the Donation_Data table and Donor_data table. The ON function specifies how we combine the two tables. I used the below queries to find out how many donors and the sum of donors in each donation_frequency.

```
1 SELECT donation_frequency, COUNT(donation) AS 'Number of Donors'
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 GROUP BY donation_frequency;
6
8 SELECT donation_frequency, SUM(donation) AS 'Donation Amount in $'
9 FROM Donation_Data
10 JOIN Donor_Data2
11 ON Donation_Data.id = Donor_Data2.id
12 GROUP BY donation_frequency;
13
```

Donors with University Education that donates weekly and above 400

In addition to the JOIN and ON function, WHERE, AND and ORDER BY were used in this query. WHERE is a clause that filters the result of a set to include only rows where a specified condition is true. AND allows you to combine multiple conditions in a WHERE clause to make the result set more specific and useful. ORDER BY is used for sorting the result set by a particular column either alphabetically or numerically.

I used the below queries to find out the records of donors with university education who donates weekly and donates above \$400. This is very important in the analysis to target high quality donors during campaigns.

```
1 SELECT Donation_Data.first_name, Donation_Data.last_name,
2 Donation_Data.state, Donation_Data.donation,
3 Donor_Data2.donation_frequency, Donor_Data2.university
4 FROM Donation_Data
5 JOIN Donor_Data2
6 ON Donation_Data.id = Donor_Data2.id
7 WHERE donation_frequency = 'Weekly'
8 AND donation >= 400
9 AND university != 'null'
10 ORDER BY donation;
```

Top 10 Male and Female Donors who Donated Between \$400 and above (\$500)

In addition, to JOIN, ON, WHERE, AND, ORDER BY, GROUP BY, the HAVING clause, BETWEEN, and LIMIT were used in this query. The HAVING clause is like a WHERE clause, but applies only to groups as a whole, and filters these groups based on the specified condition. The BETWEEN command is used to select values within a given range and the LIMIT is a clause that lets you specify the maximum number of rows the result set will have.

I used the below queries to find out the records of top 10 Males and top 10 Females that donated between \$400 and \$500

```
1 SELECT first_name, last_name, gender, donation, state, donation_frequency
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 WHERE gender = 'Male'
6 GROUP BY donation
7 HAVING donation BETWEEN 400 AND 500
8 ORDER BY donation DESC
9 LIMIT 10;
```

```
SELECT first_name, last_name, gender, donation, state, donation_frequency
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
WHERE gender = 'Female'
GROUP BY donation
HAVING donation BETWEEN 400 AND 500
ORDER BY donation DESC
LIMIT 10;
```

Top 10 States with Highest Donors and Top 10 States with the Lowest Donors

I used the COUNT and DESC in the ORDER BY clause to know the top 10 States that have the highest number of donors. I also duplicated the query but using the default ORDER BY i.e. ASC to get the States with the 10 lowest number of donors.

```

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

```

```

SELECT state, COUNT(*)
FROM Donation_Data
GROUP BY state
ORDER BY COUNT(*)
LIMIT 10;

```

COUNT and SUM of Donations by Shirt Size

Here, I used the aggregate functions COUNT and SUM to know the donors shirt size that has the highest number and total sum of donations.

```

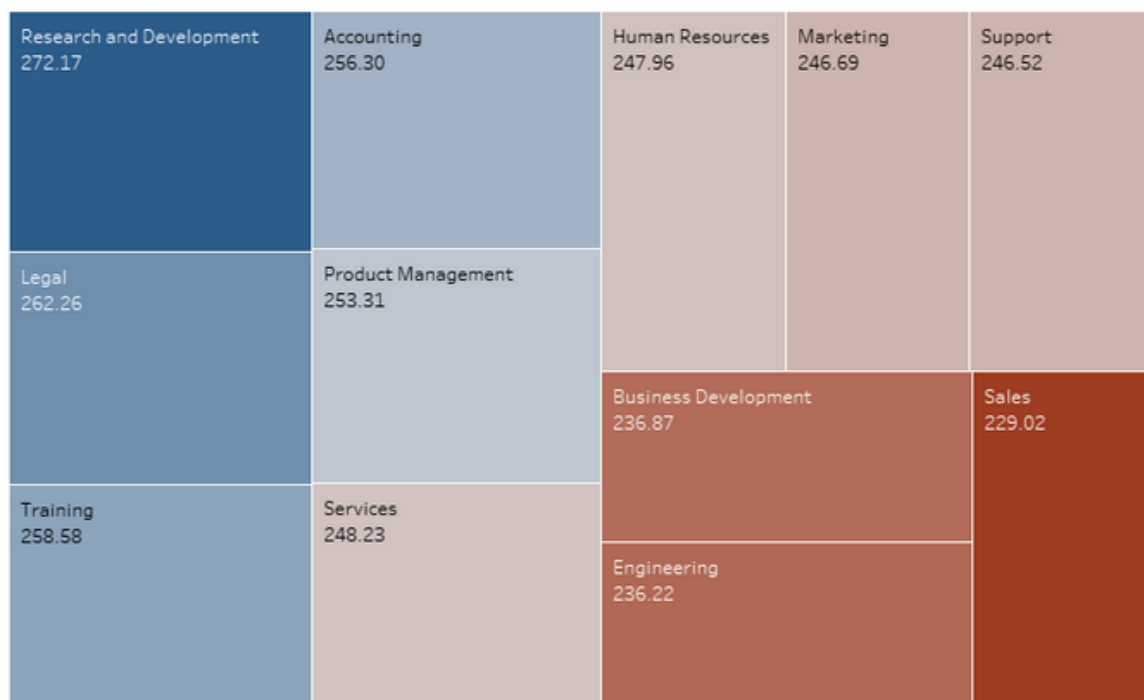
1 SELECT shirt_size, COUNT(donation), SUM(donation)
2 FROM Donation_Data
3 JOIN Donor_Data2
4 ON Donation_Data.id = Donor_Data2.id
5 GROUP BY shirt_size;

```

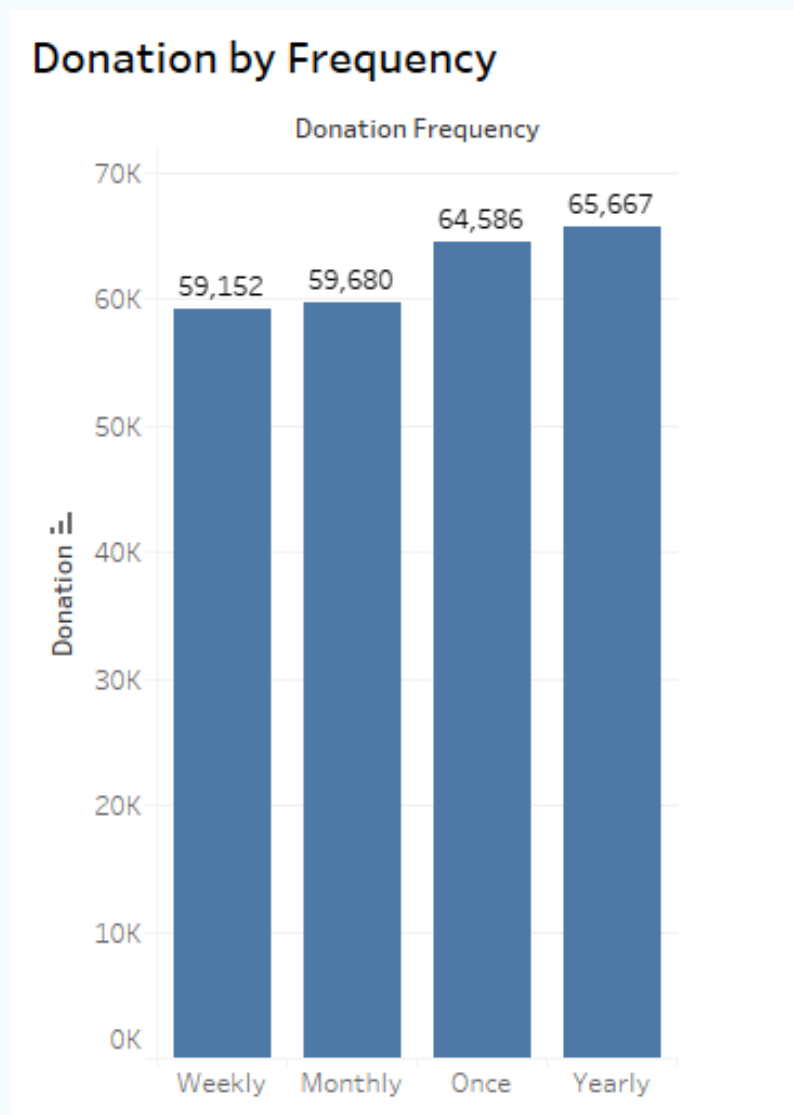
Furthermore, I used Tableau Public for creating visualizations and dashboards not only because of its speed, and scalability, but also, it is one of the best tools to create visual answers to most business questions, from bar charts to more complex visualization.

Here it is seen that donors in the Human Resources Job field have donated more than any other job title in the donors database.

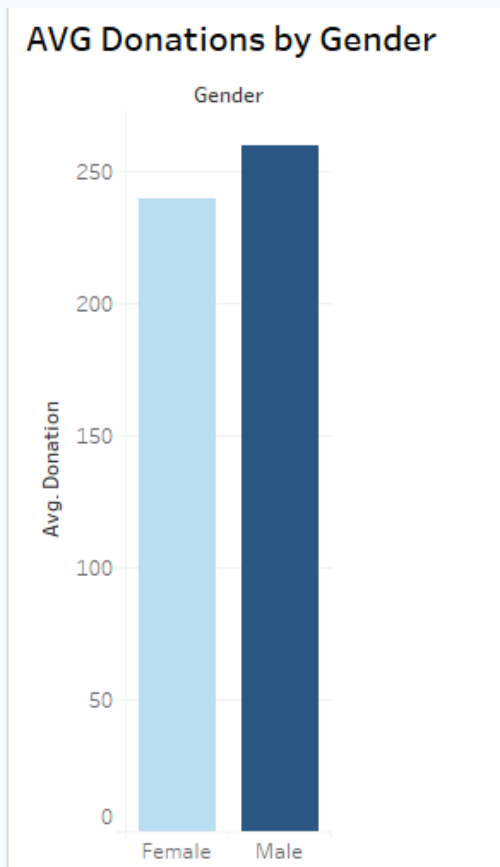
AVG Donations by Job Field



Donors that donate yearly have made more donations than those that donate once, weekly and monthly.

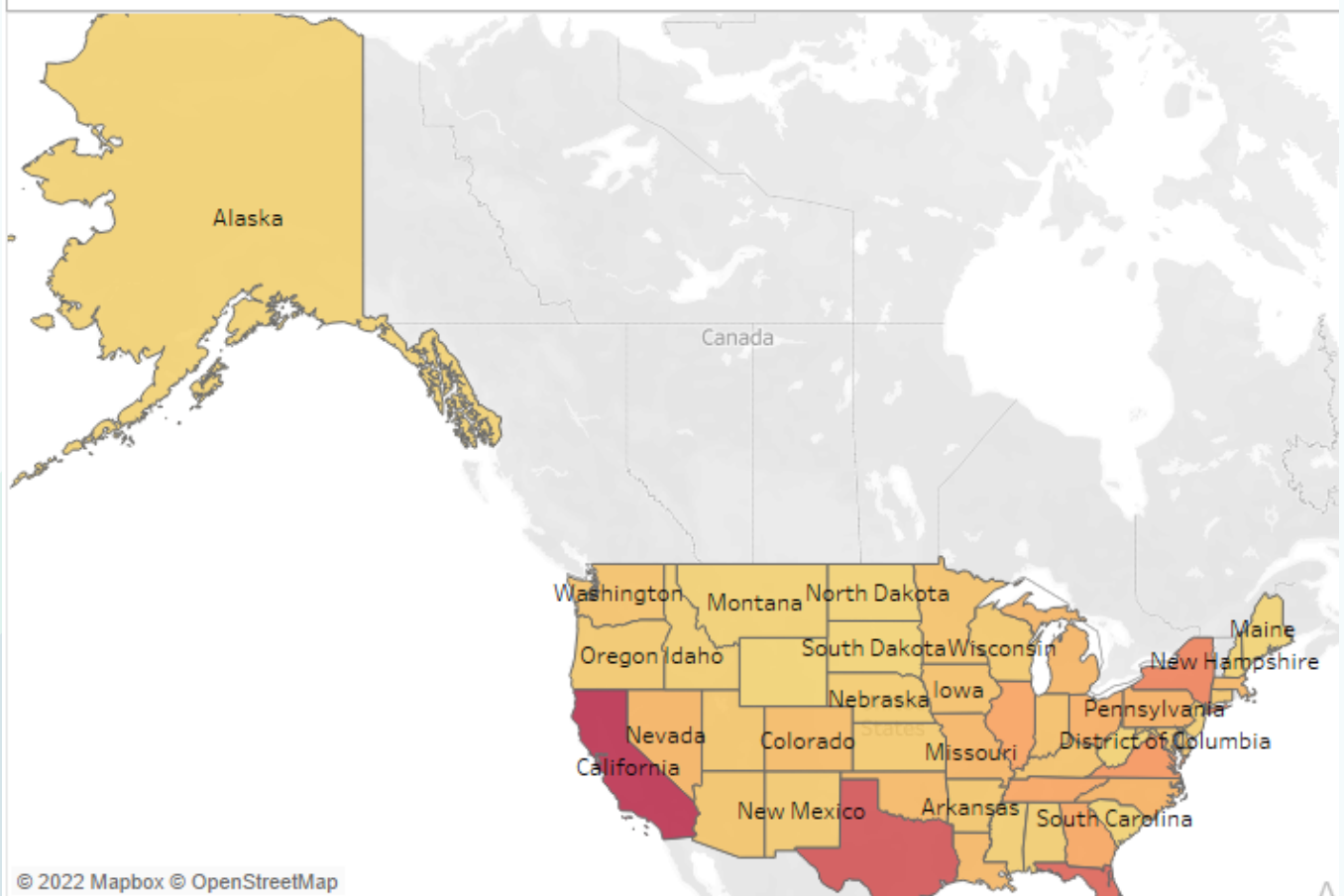


The male donors have made more donations than the female donors in the donation database.



There are 49 states from which the donors are located and majority of the donors come from California.

States of Donors



Findings and Recommendations

From the datasets, Donation Data and Donor Data, I found significant insights which are;

- Total sum of donations is \$249085
- Number of donors is 1000
- Maximum donation made by donors is \$500
- Minimum donation made by donors is \$5
- Average donations made by donors is \$249.085

COUNT and SUM of Donors by Donation Frequency

| donation_frequency | Number of Donors |
|--------------------|------------------|
| Monthly | 232 |
| Once | 264 |
| Weekly | 245 |
| Yearly | 259 |

| donation_frequency | Donation Amount in \$ |
|--------------------|-----------------------|
| Monthly | 59680 |
| Once | 64586 |
| Weekly | 59152 |
| Yearly | 65667 |

The number of donors who donated Once and Yearly is more than any other donors in the donation frequency while donors who donated Yearly and Once had the highest amount of donations. This means that the amount of donations is slightly affected by the number of donors. However, to increase donation frequency, the number of donors that donate Weekly and Monthly should be encouraged. Getting weekly and monthly donations is a good practice.

Donors with University Education that donates weekly and above 400


| first_name | last_name | state | donation | donation_frequency | university | |
|------------|------------|----------------------|----------|--------------------|----------------|--|
| Aile | Buey | Nevada | 400 | Weekly | Windless | |
| Janine | Ferber | District of Columbia | 408 | Weekly | O'Lehane | |
| Sisely | Mathieu | District of Columbia | 409 | Weekly | Flanner | |
| Barbey | Nugent | New York | 409 | Weekly | Dulanty | |
| Maurita | Penn | California | 416 | Weekly | Messent | |
| Kanya | Pre | Oklahoma | 418 | Weekly | Cluatt | |
| Krishna | Pauler | New York | 419 | Weekly | Tranmer | |
| Manny | Twiddy | Washington | 420 | Weekly | Chillingsworth | |
| Angelo | Tansey | New York | 420 | Weekly | Powling | |
| Marissa | Grimes | Idaho | 421 | Weekly | Greenlees | |
| Erinna | Beaty | Minnesota | 421 | Weekly | Finby | |
| Kathi | Sissens | Louisiana | 423 | Weekly | Nolot | |
| Dar | Le Fleming | Montana | 423 | Weekly | Trehearn | |
| Tobey | Piatkow | Indiana | 433 | Weekly | Inger | |
| Reece | Paten | New Hampshire | 434 | Weekly | Mines | |
| Brant | Teaser | New York | 435 | Weekly | Minci | |
| Terrie | Roast | New York | 436 | Weekly | Cobbold | |
| Brynn | Bartoleyn | Michigan | 436 | Weekly | Clowney | |
| Georgina | Bithell | Florida | 439 | Weekly | Turmel | |
| Ediva | Crease | Texas | 447 | Weekly | Meeny | |
| Elisa | Livesley | Georgia | 452 | Weekly | Enevoldsen | |
| Kacy | Cleall | New Jersey | 460 | Weekly | Daughtery | |
| Terrance | Toulch | Georgia | 465 | Weekly | Rubert | |
| Willamina | Chadbourn | Massachusetts | 473 | Weekly | Mundy | |

| | | | | | | |
|-----------|------------|------------|-----|--------|------------|--|
| Zak | Giannazzi | Texas | 474 | Weekly | Damant | |
| Clerc | Pietron | Minnesota | 476 | Weekly | Royan | |
| Shana | Farnsworth | Texas | 482 | Weekly | Guerrier | |
| Karena | Andrieu | Texas | 487 | Weekly | Forrington | |
| Maura | Ferrol | California | 488 | Weekly | Turford | |
| Charlotta | Bellison | Florida | 489 | Weekly | Fraser | |
| Tonnie | Stockney | California | 494 | Weekly | Cominetti | |
| Amalea | Knill | New York | 497 | Weekly | Rockcliffe | |

In the dataset, there are 32 donors with university education who donated weekly and above \$400. However, to increase donations, donors in this category should be targeted during promotion and campaigns. Ideally, people with a university education should have a good job; this means they earn higher income and weekly donation is highly recommended for frequency in donations.

Top 10 Male and Female Donors who Donated Between \$400 and above (\$500)

| first_name | last_name | gender | donation | state | donation_fr equency |
|------------|--------------------|--------|----------|------------|------------------------|
| Beverlie | Andriesse | Male | 500 | Michigan | Yearly |
| Worthy | Le feaver | Male | 498 | Wisconsin | Monthly |
| Amalea | Knill | Male | 497 | New York | Weekly |
| Tonnie | Stockney | Male | 494 | California | Weekly |
| Beverlee | Camacke | Male | 493 | Maryland | Monthly |
| Emmit | McKenzie | Male | 491 | Nevada | Once |
| Ludvig | Alexsandro wicz | Male | 489 | California | Monthly |
| Padraig | Trittam | Male | 488 | Illinois | Once |
| Rupert | Hazelgreave | Male | 487 | Minnesota | Yearly |
| Shanie | Judd | Male | 486 | Florida | Once |



| first_name | last_name | gender | donation | state | donation_frequency |
|------------|------------|--------|----------|-------------|--------------------|
| Clevie | Camilletti | Female | 499 | Virginia | Yearly |
| Corbett | Lansdale | Female | 494 | California | Monthly |
| Hurley | Bogey | Female | 492 | Florida | Weekly |
| Babbette | Fyers | Female | 491 | New Mexico | Monthly |
| Karilynn | Ivan | Female | 490 | Kentucky | Monthly |
| Charlotta | Bellison | Female | 489 | Florida | Weekly |
| Maura | Ferrolti | Female | 488 | California | Weekly |
| Karena | Andrieu | Female | 487 | Texas | Weekly |
| Ford | Evins | Female | 486 | Washington | Yearly |
| Emery | Rospars | Female | 483 | Connecticut | Monthly |

Although gender does not have much impact on donations, it's good to know that the top 10 males have donated a total sum of \$4923 while the top 10 females have donated a total sum of \$4899.

Top 10 States with Highest Donors and Top 10 States with the Lowest Donors

Top 10 States

| state | COUNT(*) |
|----------------|----------|
| California | 113 |
| Texas | 95 |
| Florida | 90 |
| New York | 58 |
| Virginia | 39 |
| Illinois | 34 |
| North Carolina | 33 |
| Georgia | 33 |
| Ohio | 32 |
| Tennessee | 30 |

Least 10 States

| state | COUNT(*) |
|---------------|----------|
| Maine | 1 |
| South Dakota | 1 |
| Wyoming | 1 |
| North Dakota | 2 |
| Alaska | 3 |
| New Hampshire | 3 |
| Hawaii | 4 |
| Montana | 4 |
| Mississippi | 5 |
| New Jersey | 6 |

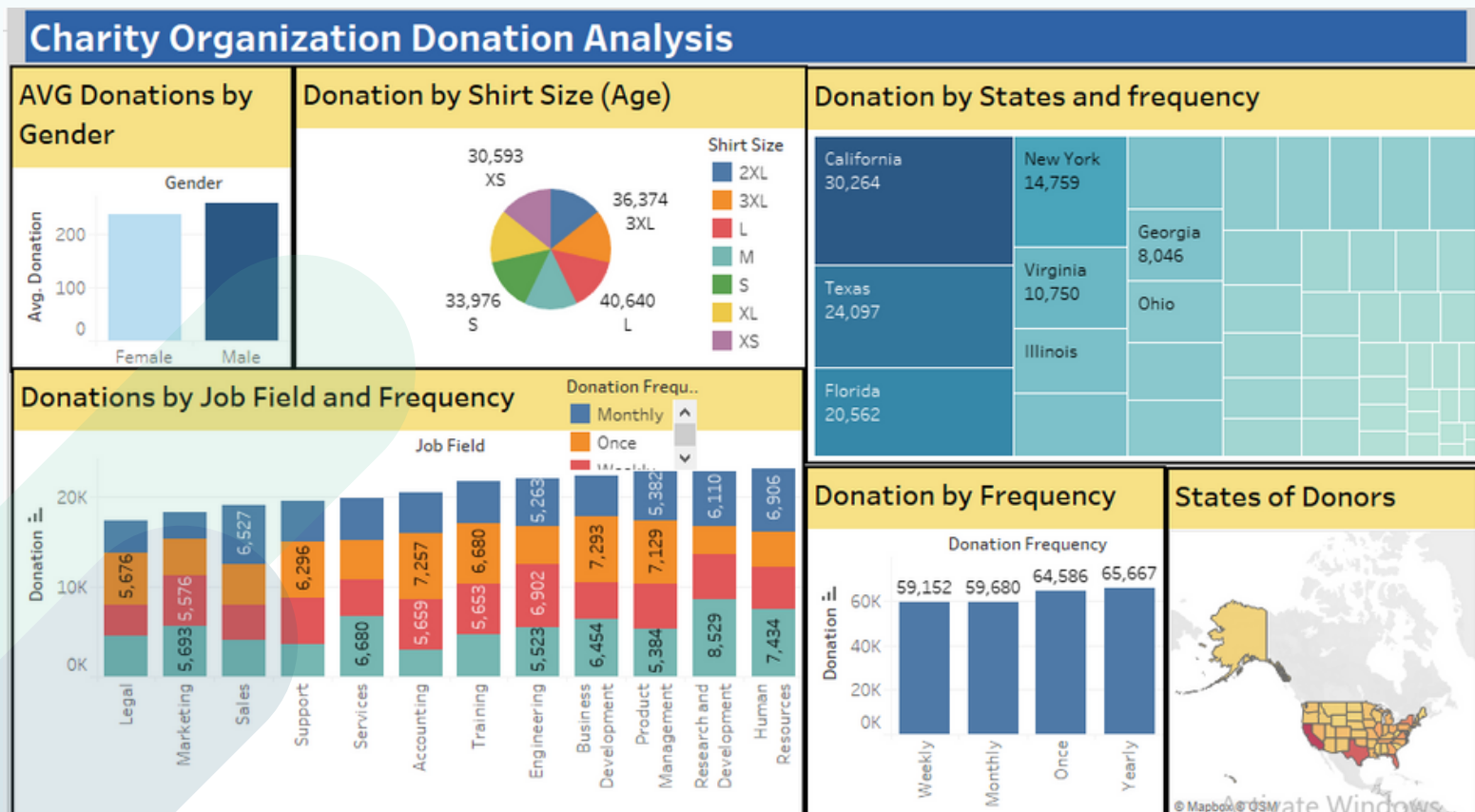
Most of the donors come from California, Texas, Florida, and New York, with 11.3% of donors coming from California.

COUNT and SUM of Donations by Shirt Size

| shirt_size | COUNT(donation) | SUM(donation) |
|------------|-----------------|---------------|
| 2XL | 137 | 33033 |
| 3XL | 149 | 36374 |
| L | 165 | 40640 |
| M | 149 | 37223 |
| S | 130 | 33976 |
| XL | 142 | 37246 |
| XS | 128 | 30593 |

Since the age of donors is not provided, I used the shirt size to hypothetically group the age ranges of donors as shirt size XS, S, and M may represent younger donors while shirt size L, XL, 2XL, and 3XL may represent older donors. Furthermore, the shirt size L has the highest number of donors and donations, which variably means that older donors donate more than the younger donors.

Donation_Donor Dashboard



Conclusion

After analyzing the two datasets, EFO_Donation_Data and EFO_Donor_Data, I found out significant insights that will help solve the Charity Organization's problem which is to increase the number of donors in the database, Increase the donation frequency of your donors and Increase the value of donations in the database.

Getting donors with university education who donates weekly and donates above \$400 is very important to target high quality donors during campaigns.

To increase the number of donors, the organization should utilize their social handles for online presence to be able to get more donors by showing their accomplishments to their online networks. This is because most potential donors may be comfortable engaging with organizations they support via social media.

To increase donation performance, it is recommended to give a first time donor a welcome package and incentives should be given to donors. This will make them feel valuable and maybe introduce friends to the organization

There is a lack of valuable data which are not needed in solving the organizations problem like donor Car, second language, favorite color, and movie genre. However, more actionable data is required to dig deeper to understand donors' behavior and why there is decline in donations. Data for the donor's age, donor complaints, compliments, recommendations, and feedback would help the Charity understand their donors behavior and increase donor retention.

Appendix

Google Sheets Data Set of Donation Data and Donor Data

[Link here](#)

Tableau Data Visualization of Donation and Donor Data

[Link here](#)

GITHUB SQL Codes

[Link here](#)