DATA ANALYST: SQL PORTFOLIO

Professional Background

I am an experienced and driven data professional with expertise in information technology, strategic management, analysis and data storytelling. I am skilled in collaborating with all members of the organization to achieve business and strategic objectives. My Technical Profitability includes; MS Excel, SQL, R, Power BI, Tableau, SAP, Mcode and Python.

I have had an experience working as a Content Analyst for an organization called Play Communications in Nairobi, Kenya. My work entailed analyzing music content which were uploaded to different Network Providing Company for Caller tunes, my analysis entailed figuring out the most used Music content across the various Network providing Companies and communicating this findings to Management at the Network Providing Companies.

I have also worked as an IT Support Specialist which most of my work involved manipulating the Companies database using Microsoft Server, ensure stable internet network across the organization and collaborate with team mates on projects like installation of Microsoft365 for every employee within the organization.

I have a Bachelor's degree in Management Information Systems and a Master's degree in Business Administration and Management. My love for Statistics and Numbers made me interested in Data Analytics and I have been trying to improve my knowledge in it this year by enrolling in Certification courses which has really helped me in this Career path of Data Analysis.

Portfolio Outline

Introduction

I was given a task as a Data Analyst for a Charity Organization that is in charge of Education for ALL As the Head of Fundraising, I have been asked to present data on donor insight and donation rate. My objectives within the team includes;

·Increase the number of donors in our database.

·Increase the donation frequency of our donors.

·Increase the value of donations in our database.

I used the data sets EFO_Donation_Data and EFO_Donor_Data to answer the business problem. I applied SQL commands to analyse data: JOIN, ORDER BY, WHERE, BETWEEN, AND, OR, SUM(), COUNT(), AVG(), GROUP BY.

In two weeks' time, I will be presenting findings from the donation data to my team and communicate my strategy which is to improve donations the following year.

Root Cause Analysis

The business problem is that there are not enough donors that donate regularly with high value donation which makes it hard to grow funding.

Therefore, more donors who can donate regularly and will increase the value of their donation needs to be searched for.

To understand the problem, I needed to analyze existing data bases of Donors and Donations. Also, I should present crucial numbers and visualizations of our data sets. I decided to ask some questions to dig the problem deeper:

·How many donors we have in existing database?

- · What is the amount of their donations?
- ·What is the frequency of the donations?
- ·What is the ratio of male donors to female donors?
- · Is the donation amount depending on the university status of donors?

I applied Root Cause Analysis to ask:

1. Why do we not have as many donors as we need?

2. Why in our database we cannot see when people joined our organisation and how long they stay with us?

3. Why we do not add their date of birth, Marital Status to our dataset to research what is the age group and marital status of our donors?

Insights

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

- ·Id
- · First name
- ·Last Name
- ·Email
- ·Gender
- ·Job field
- ·Donation
- ·State
- ·Shirt Size

Donor Dataset includes such data:

- ·Id
- ·Donation frequency
- ·University
- ·Car
- ·Second language
- · Favourite colour
- ·Movie genre

Both data sets were imported into SQLite.

To find the total number of donors with the COUNT() Function, I used a command like this:

```
-- Count Number of Donors
SELECT COUNT(DISTINCT email) AS Number_of_Donor
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id; -- 1000
```

Insights

The INNER JOIN clause allows us to join different columns from multiple tables together.
To find the total sum of donations I used SUM():

```
-- Sum of all Donations
SELECT SUM(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id; -- 249085
```

To find to maximum and minimum amount of donation I used MAX() and MIN():

```
-- Highest to Lowest Donation Amount
SELECT first_name, last_name, gender, job_field,donation,donation_frequency,state,university,car
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
ORDER BY donation DESC;
```

To find the count and sum of each donation frequency I used COUNT() and SUM(), I also used INNER JOIN to figure out the amount of donations Once, Weekly, Monthly, Yearly. WHERE was used to filter the result of a set to include only rows where a specified condition is true:

Insights

```sql
SELECT COUNT(*) AS One_time_donor, SUM(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
WHERE donation_frequency = 'Once'; -- 264 HIGHEST COUNT, Sum = 64586x

--Sum of Weekly Donation and Count of Weekly Donors
SELECT COUNT(*) AS Weekly_donor, SUM(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
WHERE donation_frequency = 'Weekly'; --245, Sum = 59152

-- Sum of Monthly Donations and Count of Monthly Donors
SELECT COUNT(*) AS Monthly_donor, SUM(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
WHERE donation_frequency = 'Monthly'; --232 LOWEST COUNT, Sum = 59680

-- Sum of Yearly Donation and Count of Yearly Donors
SELECT COUNT(*) AS Yearly_donor, SUM(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
WHERE donation_frequency = 'Yearly'; --259, Sum = 65667
```

To find the top 10 donors I also used INNER JOIN to join the column from multiple tables together. WHERE was used to filter the result of a set to include only rows where a specified condition is true. . ORDER BY function with DESC order was used to order the donor data
by the highest to lowest donations:

```sql
-- Top 10 Donors
SELECT first_name, last_name, gender, job_field,donation,donation_frequency,state
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
ORDER BY donation DESC
LIMIT 10;
```

Insights

To find the ratio of male to female donors I used COUNT():

```sql
-- Count of Male Donors
SELECT COUNT(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Male'; --492

--Count of Female Donors
SELECT COUNT(donation)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Female'; --508
```

I used INNER JOIN command and COUNT() function on the data to understand how many females and males with university education donated more than $350:

```sql
-- Female donors with University Education and donation > 350
SELECT first_name, last_name, gender, donation, university,donation_frequency
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Female' AND university NOTNULL AND donation > 350
ORDER BY donation DESC;

SELECT COUNT(*)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Female' AND university NOTNULL AND donation > 350 -- 101
```

Insights

```
-- Male donors with University Education and donation > 350
SELECT first_name, last_name, gender, donation, university,donation_frequency
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Male' AND university NOTNULL AND donation > 350
ORDER BY donation DESC;

SELECT COUNT(*)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Male' AND university NOTNULL AND donation > 350 -- 117
```

I also found out how many females and males without education donated more than $350:

```
-- Female donors without University Education and donation > 350
SELECT first_name, last_name, gender, donation, university,donation_frequency
FROM Donation_Data
LEFT JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Female' AND university IS NULL AND donation > 350
ORDER BY donation DESC;

SELECT COUNT(*)
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Female' AND university IS NULL AND donation > 350 -- 40


-- Male donors without University Education and donation > 350
SELECT first_name, last_name, gender, donation, university, donation_frequency
FROM Donation_Data
LEFT JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Male' AND university IS NULL AND donation > 350
ORDER BY donation DESC;

SELECT COUNT(*)
FROM Donation_Data
LEFT JOIN Donor_Data2
ON Donation_Data.id =  Donor_Data2.id
WHERE gender = 'Male' AND university IS NULL AND donation > 350 -- 39
```
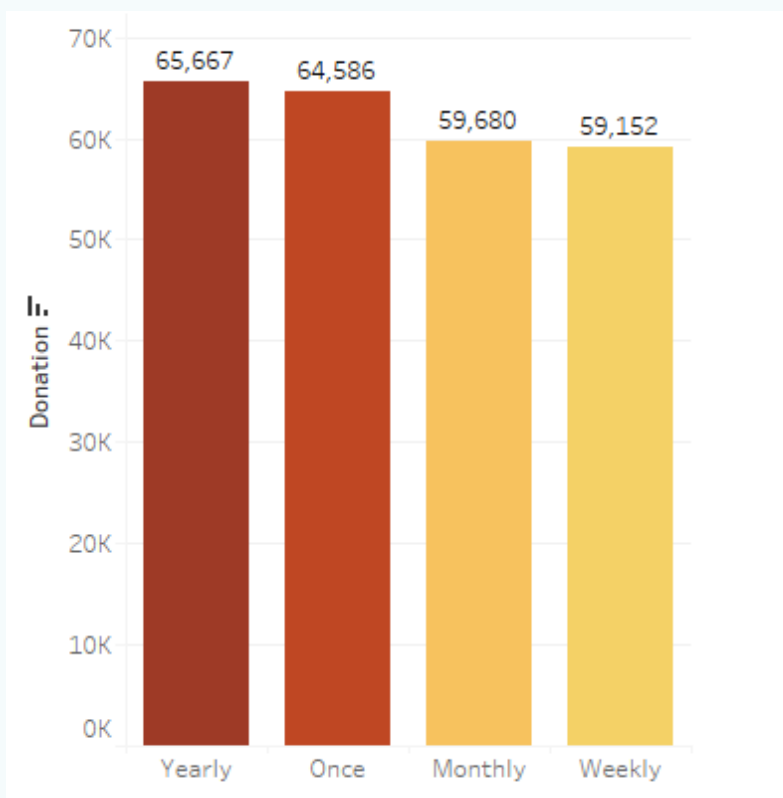
Insights

Tableau Visualisation
I made use of Tableau to visualize the data after anaylsis
Most donors are one time donors but the donation frequency with
the highest sum of donation are the yearly donors

Donation Frequency	Donation Count	Total Donation
Monthly	232	59,680
Once	264	64,586
Weekly	245	59,152
Yearly	259	65,667

Insights

Also, Majority of our donors live in California, Texas and Floria
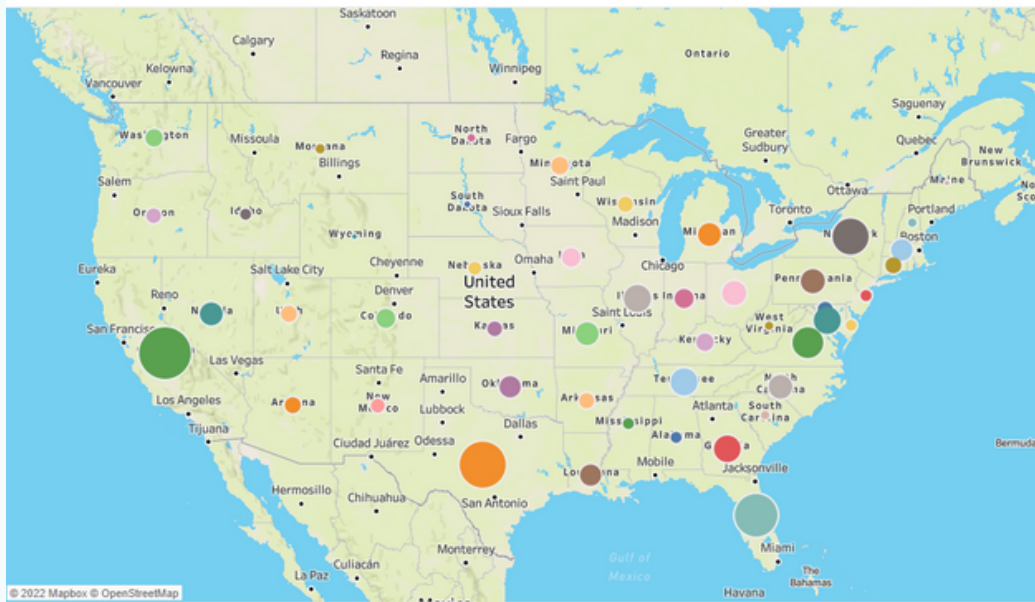
Findings and Recommendations

From my analysis, I found out:

1. The total number of donors we have in the database is 1000.
2. The total sum of donation is $249,085.
3. The highest amount donated is $500.
4. The lowest amount donated is $5.
5. Most of the donors are one time donors.
6. Majority of the donors live in California, Texas and Florida.
7. The people with high donation amount are males and females that attended University. 101 females and 117 males with university educations donated more than $350 whereas, 40 females and 39 males without university education donated more than $350. Among people without university education the number of males and females, who donate more than $350 is almost the same. Therefore, we can see that our target audience is males and females with university education.

Conclusion

I analyzed two datasets i.e. EFO_Donation_Data and EFO_Donor_data to help Education for ALL understand their business problem which involves finding new donors and finding a way to raise higher donation with regular frequency.

Therefore, our donors are very dissimilar people. They live in different states, work in absolutely unsimilar job fields, some of them have university education and some of them no. Amount of donation are also vary between $5 and $500 dollars, and some of them make weekly donation others yearly donation.

I have found out that frequency of Yearly and Once donations is much higher than Monthly.

Weekly frequency is very low. So, we need to work on improvement of donation frequency, especially Weekly donations.

Gender is not crucial thing. Thus, we can notice that people, who are educated themselves are happy to support others to get education. Some people with NULL education also provide big amounts, but not a big number of them.

In conclusion, we must concentrate to improve our databases, add more useful information about our donors such as Age and Marital Status. There is lack of valuable data. We must find useful channels for promoting our charity organisation.