Final Portfolio Report

Data Analyst Track

TABLE OF CONTENTS

INTRODUCTION

A Data Analyst plays a pivotal role in helping a business to understand its data in other to develop strategies that would yield positive outcomes. Part of the responsibilities of a data analyst is to sieve through an organisations data (both historic and newly collected) to identify trends and patterns and use these to extract actionable insights that will help inform how a business should approach certain problems or opportunities encountered.

For this project, the focus is on helping a charity organisation to analyze its historical data and come up with recommendations to achieve the following objectives;

1. Increase the number of donors
2. Increase the donation frequency of donors
3. Increase the value of donations received from donors

Given a timeline of two weeks, a comprehensive report with insights from the analysis performed is to be presented to the Head of Fundraising for the Education for All charity. It should be noted that the overall objective for performing this analysis is to inform a fundraising strategy ahead of the coming year.

PROBLEM STATEMENT

As a first step to tackling this problem, it was necessary to formulate a problem statement to summarise the specific problem that needs to be solved for. The problem statement is thus framed as follows;

> *How can the Education for All charity attract an even greater flow of donations to support its charitable efforts?*

Haven put the problem into perspective by framing a problem statement, the data required to perform this analysis is then considered along with asking further questions to better understand the kind of data being used.

DATASET AND DESCRIPTION

Two datasets were used to perform this analysis.

Dataset 1: EFO_Donation_Data with total of 1000 records consists of the following attributes;

id = Donor ID

first_name = Donor first name

last_name = Donor last name

email = Donor email address

gender = Donor gender

job_field = Donor job field

donation = Donation amount

state = Donor state of residence (US)

shirt_size = Donor t-shirt size

Dataset 2: EFO_Donor_Data with total of 500 records consists of the following attibutes;

id = Donor ID

donation_frequency = Frequency of donation

university = Donor University attended

car = Donor car make

second_language = Donor second language

favourite_colour = Donor favourite colour

movie_genre = Donor favourite movie genre

To ensure that a comprehensive analysis was done, a combined form of each dataset was achieved using a ***left join*** function is some of the SQL queries that were written to extract insights.

QUESTIONS & ROOT CAUSE ANALYSIS PERFORMED

The dataset alone may not always offer a complete source of information to better understand the kind of problem needed to be solved. As in this case, the Subject Matter Expert (SME) or any other stakeholder within the organization needs to be engaged to gather more relevant information to aid the analysis. For instance, answers to the following questions were sort after;

1. What was the period over which the donation data was collected?
2. What was the previous donation canvasing strategies?
3. Were the ages of the donor also collected and stored somewhere in the database?

The root cause analysis was performed using the five why methodology where ***why*** was asked five times to the root cause of the organization not getting as many valuable donations. Answers to these whys are hypothesis that would either be proved right or wrong via the analysis that follows.

Why 1: why is the organization not getting enough donations?
Ans 1: Because there are not a lot of people donating.

Why 2: why are there not a lot of people donating?
Ans 2: Because not many people know that the organization exists.

Why 3: why don't many people know that the organization exists?
Ans 3: Because people do not know the impact it is making in the society.

Why 4: why don't people know about the impact it is making in the society?
Ans 4: Because it is not talking about it enough through the right channels.

Why 5: why is it not talking about it through the right channels?
Ans 5: Because it does not have a good publicity team.

Therefore, the root cause of why Education for All may not be getting as many donations could be because it doesn't have a good publicity team to help push out the great works that its doing and so, many people who would have potentially contributed generously to the cause are not getting to see the impact that they have in the society.

INSIGHTS BACKED BY DATA

After loading both datasets into SQLite, descriptive and diagnostic analysis were then performed to extract several insights to aid better understanding of the data. Some of these analyses along with the codes used to generate them are briefly discussed in this section.

Insight 1: What was the total amount donated?

```
SELECT sum(donation) AS TotalAmountDonated
FROM Donation_Data;
```

TotalAmountDonated
249085

Insight 2: Is there a clear pattern observed between the total sum donated by each gender?

```
SELECT gender, sum(donation) AS Total_amount_donated
FROM Donation_Data
GROUP BY gender
ORDER BY Total_amount_donated DESC;
```

gender	Total_amount_donated
Male	127628
Female	121457

Insight 3: What is the maximum and minimum amount donated?

```
1 SELECT max(donation) AS Highest_donated_amount,
2 (SELECT min(donation) FROM Donation_Data) AS Lowest_donated_amount
3 FROM Donation_Data;
```

Highest_donated_amount	Lowest_donated_amount
500	5

Insight 4: What are the top 5 job fields with the most donations?

```
1 SELECT job_field, sum(donation) AS contribution
2 FROM Donation_Data d
3 GROUP BY job_field
4 ORDER BY contribution DESC
5 LIMIT 5;
```

job_field	contribution
Human Resources	23060
Research and Development	22862
Product Management	22798
Business Development	22266
Engineering	21968

Insight 5: What is the maximum and minimum amounts donated per gender?

```
SELECT gender, max(donation) AS max_donation,
    (SELECT min(donation) FROM Donation_Data) AS min_donation
FROM Donation_Data
WHERE gender IN (
    SELECT gender
    FROM Donation_Data)
GROUP BY gender;
```

gender	max_donation	min_donation
Female	499	5
Male	500	5

Insight 6: What is the donation distribution per donation frequency?

```
SELECT donation_frequency, COUNT(donation_frequency)
FROM Donation_Data d
LEFT JOIN Donor_Data2 don
ON d.id = don.id
GROUP BY donation_frequency
ORDER BY COUNT(donation_frequency) DESC;
```

donation_frequency	count(donation_frequency)
Once	264
Yearly	259
Weekly	245
Monthly	232

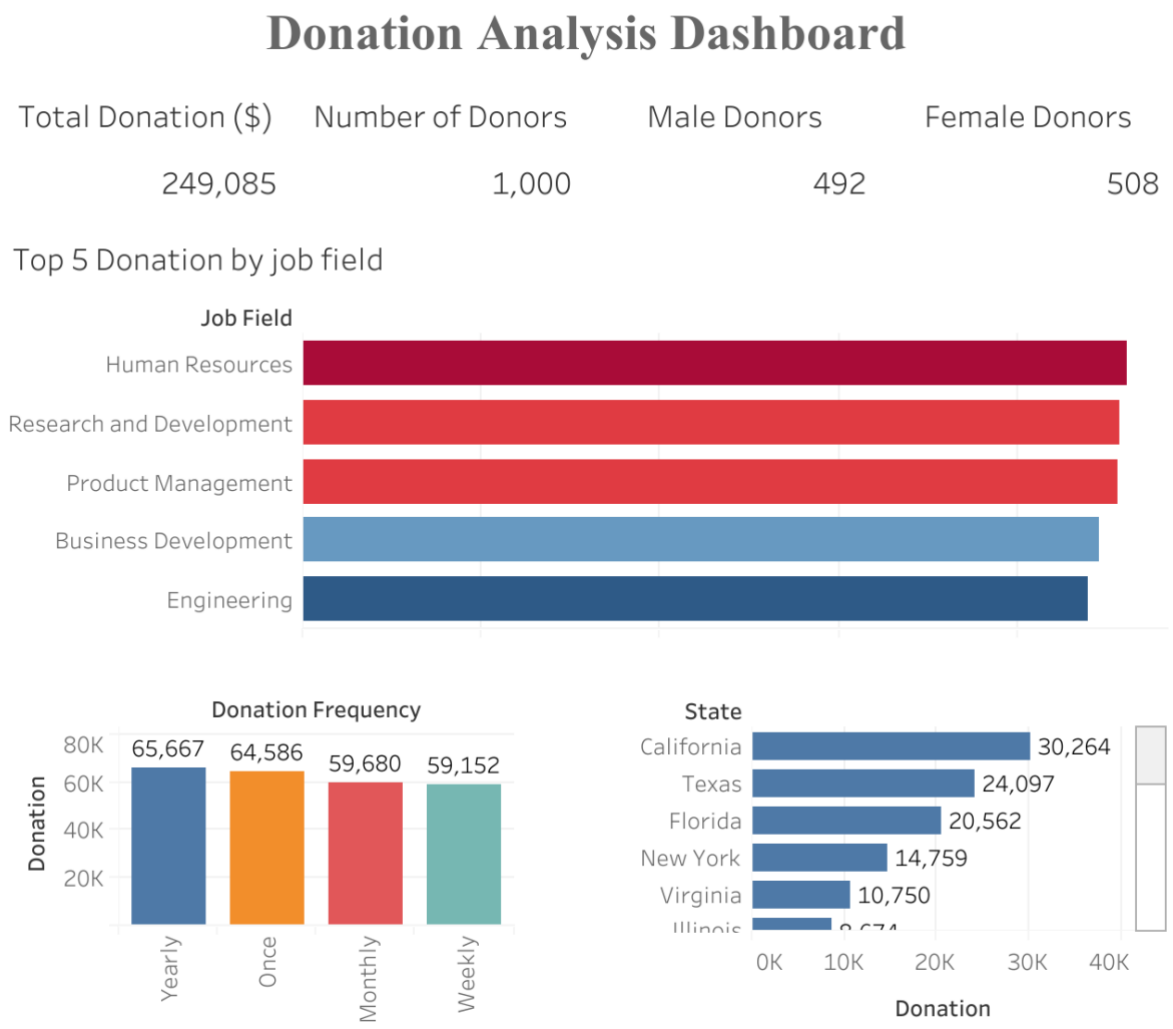
Insight 7: What is the total number of donors split by gender?

```
SELECT
CASE
    WHEN d.donation >= 400 THEN '400 Above'
    WHEN d.donation >= 200 AND d.donation < 400 THEN 'Between 200 and 400'
    WHEN d.donation >= 100 AND d.donation < 200 THEN 'Between 100 and 200'
    ELSE 'Below 100'
END AS Donor_category, COUNT(*) AS Donor_Count
FROM Donation_Data d
LEFT JOIN Donor_Data2 don
    ON d.id = don.id
GROUP BY Donor_category
ORDER BY Donor_Count DESC;
```

Donor_category	Donor_Count
Between 200 and 400	384
Between 100 and 200	207
400 Above	205
Below 100	204

VISUALISATION (DASHBOARD)

The format in which the insights have been presented in the previous section is not very easy to comprehend. To improve comprehension, a visually stimulating report dashboard was designed using a visualization tool known as tableau. This dashboard is highly interactive and presents the key insights earlier discussed. The online interactive version of this report dashboard can be found by clicking on this link.



IMPORTANT FINDINGS AND RECOMMENDATION

The key findings from this analysis can be given as follows:

1. The total amount in donations received is $249, 085.00
2. Male and female contribute nearly equally
3. There are slightly more female donors than there are male donors
4. Certain individuals in some job fields contribute more than those in other job fields. Therefore, the top five job fields donating the most come from:
    - Human Resources
    - Research and Development
    - Product Management
    - Business Development
    - Engineering
5. In terms of donation frequency, it seems like there are more donations coming through yearly and once compared to monthly, and weekly.
6. There are more people donating between $200 and $400 (384) compared to people donating $400 and above (205). Quite a significant number of people also donate below $100 (204)

For Education for All to achieve its objective of increasing the number of donors in the coming year, it must ensure that it puts together a formidable team to adequately publicize the work it does using a diversity of publicity media available to it. In addition, the fundraising manager could consider implementing the following strategies to increase value of donations and donation frequency;

1. Target more people in the top 5 job fields identified
2. Encourage these target group to donate generous amount between the region of $400 and above yearly or better still they could spread the payment across weekly or monthly, depending on their preference. Make them patrons of the charity if need be
3. Males are slightly more likely to contribute more than females, therefore it would be beneficial to equally target more males in the top 5 job fields

CONCLUSION

Descriptive and diagnostic analysis was performed using two datasets provided by Education for All. The purpose of which is the properly understand the core business problem related to improving the number of donors and value of donations received by the organization. From the analysis performed, key insights became apparent. For instance, given the historical data, the total amount in donations received was realized to be $249,085. The maximum and minimum donated amounts were $500 and $5 respectively.

The total amount donated by males compared to that donated by females resulted in a difference of $6,171, indicating that males are slightly more likely to donor to charity causes than females. On a whole, more people tended to prefer donating yearly and once as opposed to donating at a more frequent rate of weekly and monthly.

The analysis performed could have been further improved if more valuable data was included in the dataset. For instance;

- Knowing the ages of the donors would have presented another dimension to consider if there are more donors of a particular age range.
- If the dates when donations were received was included, this would have made it possible to perform a trend analysis to determine if there were more donations at a particular period in the past.
- The channel through which donations are received is not known, including this would have made it possible to determine the most preferred channel for donation by donors. This would have enhanced future fundraising strategy by ensuring to this donation channel more accessible to donors.

Some data points were not relevant for this analysis and therefore should be excluded in future analysis. For example; second language, favourite color and movie genre from dataset 2.