EntryLevel: Data Analyst 2 Portfolio Project

Using SQL to Analyze Data in Tables

Abstract

The goal of this project is to explore data from the nonprofit, Education for All to understand their current base of donors and to see how the organization might increase donations.

Table of Contents

Professional Background

I am a writer and have produced work for a variety of online businesses over the years. This focus on content writing and copywriting rose out of my virtual assisting business that I started in 2012 through which I have offered services such as WordPress support, social media marketing, email marketing, and design services.

While building my business I have studied and practiced tech skills such as coding, keeping up with the latest software such as Microsoft Office, Adobe Creative Cloud, Google Suite, Canva, etc. I looked for affordable options to start and complete a coding course and decided to focus on being self-taught and seeking out a mentoring arrangement.

While Full Stack Web Development seems to be the common course of study, I found myself drawn to data analysis. I enjoyed working with data, producing reports, and explaining how to interpret the results. My work as a content writer will complement that skill.

Throughout my professional history, I have served different companies in the areas of data entry, data management, and business analysis. Well maintained data has the power to inform decision makers, keep departments on track with goals, and maintain an accurate record of inventory and financial records. Working with data has been the most exciting aspect of my writing business, my work history, and my personal record keeping.

Introduction

For this exercise, I am to take on the role of a Data Analyst working for a charity called Education for All. In this hypothetical scenario, the Head of Fundraising tasks me with presenting the data on donor insights and donation rates. A fundraising strategy meeting is to take place in two weeks' time in preparation for the following year. During this meeting, I am to present my insights from the donation data to inform Education for All fundraising strategy and increase donations.

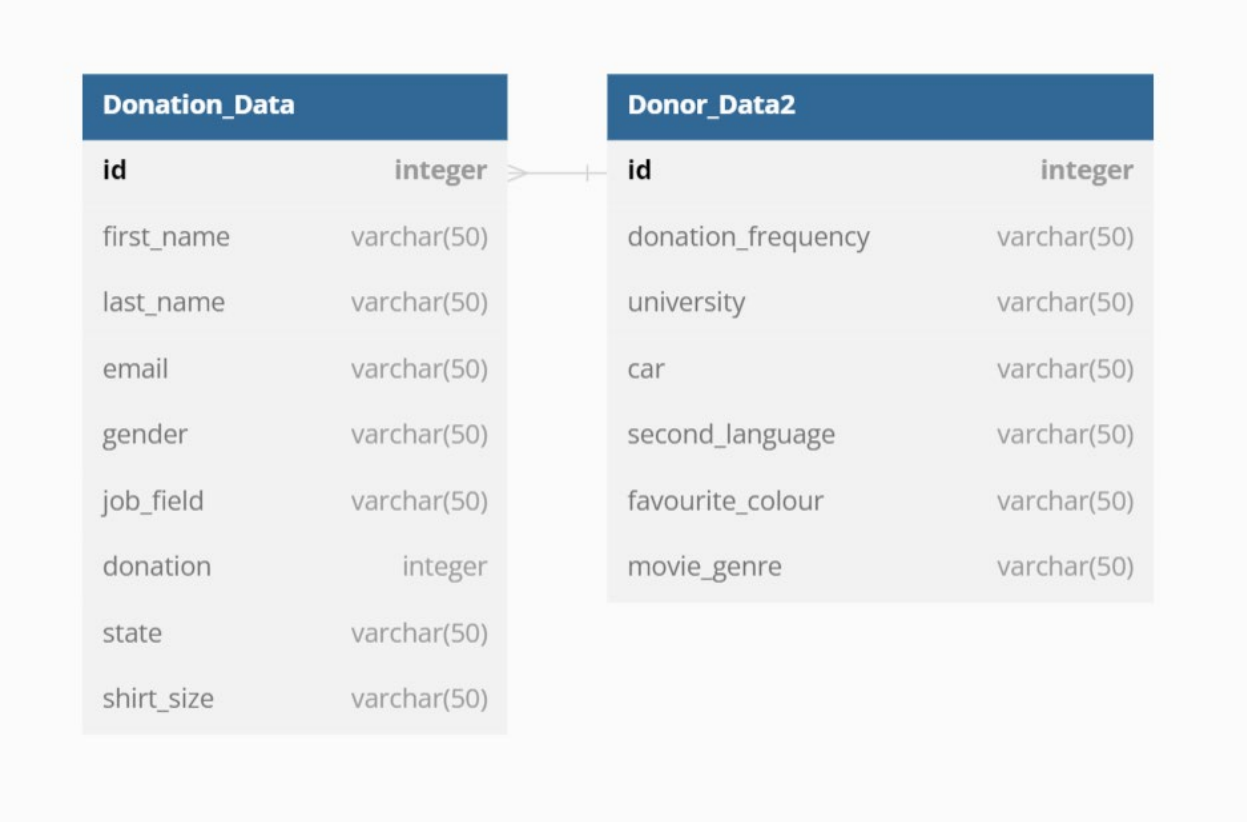
My objectives are:

- Increase the number of donors.
- Increase donation frequency of donors.
- Increase the value of donations.

Framing the Problem

The business problem is that the charity Education for All needs to raise more money. My first thoughts about this problem include questions like, "Who is donating and why?". What is the profile of the typical donor? Are donors more often male or female? What is their education level? Where do they live? What spurred them to donate? Was it an event, a mailing, a television spot, news article, or radio announcement? Was there an incentive to give?

Unfortunately, I don't have all the information that I would need to answer all my questions but with the two data tables given I can answer some of them. The two data tables supplied by the organization are Donation_Data and Donor_Data2. The database schema, therefore, looks like the figure below:



I can analyze the data to find:

- Donor Profile
  - How many donors are male and how many are female?
  - How many donors are in each state?
  - Did the donors go to college?
  - How many donors are in each job field or industry?
- Donations
  - Count and dollar amount of donations received by State.
  - Count and dollar amount of donations received by donors' gender.
  - Count and dollar amount of donations received by job field or industry.

Root Cause Analysis

The 5 "Whys"

1. Why are we limited to $249,085 in donations contributed by 1000 donors across 48 states and the District of Columbia?
    a. We may be missing marketing opportunities in the two missing states (Rhode Island and Vermont) plus the states that have fewer than 50 donors as well as not pushing for more frequent donations amongst current donors.
2. Why are we missing these marketing opportunities?
    a. We may not be making the same effort in other states and possibly need to consider more targeted marketing to working, college educated adults in other states. More frequent communication with donors may encourage them to increase the frequency and amount of their donations and also to share their commitment with others who are not yet donors.
3. Why have we not been able to drive funding in greater numbers in so many states?
    a. Since we don't have information on when donors first contributed and what spurred them to do so, we don't have enough information about what kinds of efforts work. That's information that we should start to gather. Partnering with large and medium sized employers and alumni associations may help to reach this demographic.
4. Why don't we have more relevant information on donors and potential donors to help maximize our marketing?
    a. Keeping better records of events, donation drives, telemarketing calls, all communications and advertising on television, radio, internet, etc. should help pinpoint the most successful campaigns. Also interviewing, and polling donors and potential donors will provide insight as well. Detailed records of dates of events and dates of responses, donor suggestions and ideas should prove very useful in the future.

5. Why don't we communicate with our donors and partners more regularly to find out this information and keep them informed of what our organization is doing.
    a. We may not have good contact information from our donors such as address and phone number. However, we do have their email addresses. Email newsletters, invitations to polls and surveys, and announcements about future drives, incentives and special offers can do a lot to drive more donations and spread the word with minimal effort.

Insights from the Analysis

Describing the Database

For this project I was given two files to import into an SQL database. They are 2.4-EFO_Donation_Data.sql and 2.6-EFO_Donor_Data.sql. The tables contain the following:

Donation_Data

- "id" – this is the unique id given to each donor in the database.
- "first_name" – the donor's first name.
- "last_name" - donor's last name.
- "email" – donor's email.
- "gender" – the gender of the donor in each record.
- "job_field" – the job field or industry the donor works in.
- "donation" – the amount that the donor has donated.
- "state" – the donor's state of residence.
- "shirt_size" – the donor's choice of shirt size out of 7 sizes (XS, S, M, L, XL, 2XL and 3XL).

Donor_Data2

- "id" – This is the unique id given to each donor.
- "donation_frequency" – how often the donor donated (once, weekly, monthly, or yearly).
- "university" – the university the donor attended.
- "car" – The make of the donor's car.
- "second_language" – The donor's second language.
- "favourite_colour" – the donor's favorite color.
- "movie_genre" – movie genre's the donor likes.

I imported both files into SQLite Online and began my analysis. I was able to get a look at each dataset by entering a basic select statement to show all columns of all records of each table.

```
SELECT *
FROM Donation_Data;
```

```
SELECT *
FROM Donor_Data;
```

Each of these select statements returned 1,000 records. Each record contained information about donors with id numbers 1 – 1000. I can use the id number column in each table as the Primary key for a 1 to 1 relationship between the two tables.

Understanding the Pool of Donors

To get an overview of the total donors, total donations, the maximum donation amount and the minimum donation amount, and the average, I used the following select statement:

```
SELECT COUNT(id), SUM(donation), MAX(donation), MIN(donation),
AVG(donation)
FROM Donation_Data;
```

The query returned these numbers:

COUNT	SUM	MAX	MIN	AVG
1000	$ 249,085.00	$ 500.00	$ 5.00	$ 249.09

I wanted to see how men and women contributed differently so I started with this query.

```
SELECT gender, COUNT(gender), SUM (donation)
FROM Donation_Data
GROUP BY gender;
```

The results were as follows:

gender	COUNT(gender)	SUM (donation)

Female	508	$ 121,457
Male	492	$ 127,628

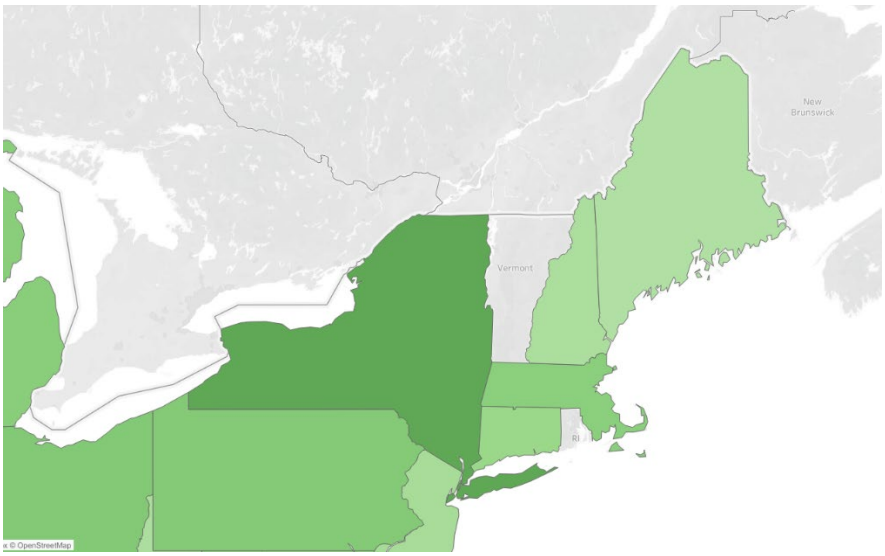
Then I wanted to see how many different job fields were represented among donors and how many donors were in each field.

```
SELECT job_field, COUNT(job_field), SUM(donation)
FROM Donation_Data
group BY job_field
ORDER BY SUM(donation) DESC;
```

The result of this query was a list of 12 job field categories.

```
SELECT state, COUNT(state) --Number of donors by state
FROM Donation_Data
group BY state
ORDER BY COUNT(state) DESC;
```

The query returned 49 rows. The areas with the highest number of donors were California (113 donors), Texas (95), Florida (90), and New York (58). Out of the 49 “states” 48 were the names of states but one of the states listed was the District of Columbia. I knew then that there were two states missing in the list or rather no donors were from those two states. I downloaded the results as a csv file and uploaded it into Tableau to make a map. It looked like every state was highlighted but when I zoomed into the Northeast region of the map, I could see that Rhode Island and Vermont were greyed out with no data.



To see the level of frequency donors, donate and how much they contribute, I used a select statement with an inner join to pull data for the same id numbers from both tables.

```
SELECT Donor_Data2.donation_frequency, COUNT(Donor_Data2.id),
SUM(Donation_Data.donation)
FROM Donor_Data2
JOIN Donation_Data
ON Donor_Data2.id = Donation_Data.id
GROUP BY donation_frequency;
```

The results of the query were as follows:

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

```
SELECT university, COUNT(id)
FROM Donor_Data2
GROUP BY university
ORDER BY COUNT(id) DESC;
```

748 donors attended university and 252 donors did not attend university.

```
SELECT second_language, count(id)
FROM Donor_Data2
GROUP BY second_language
ORDER BY COUNT(second_language) DESC;
```

272 donors speak a second language while 728 do not.

Top Donors

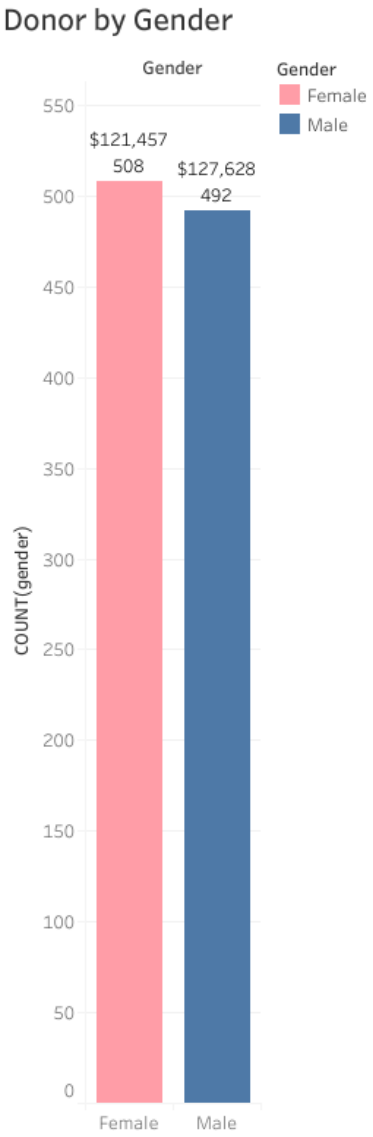
I queried the database to see the top 10 donors who contributed the most.

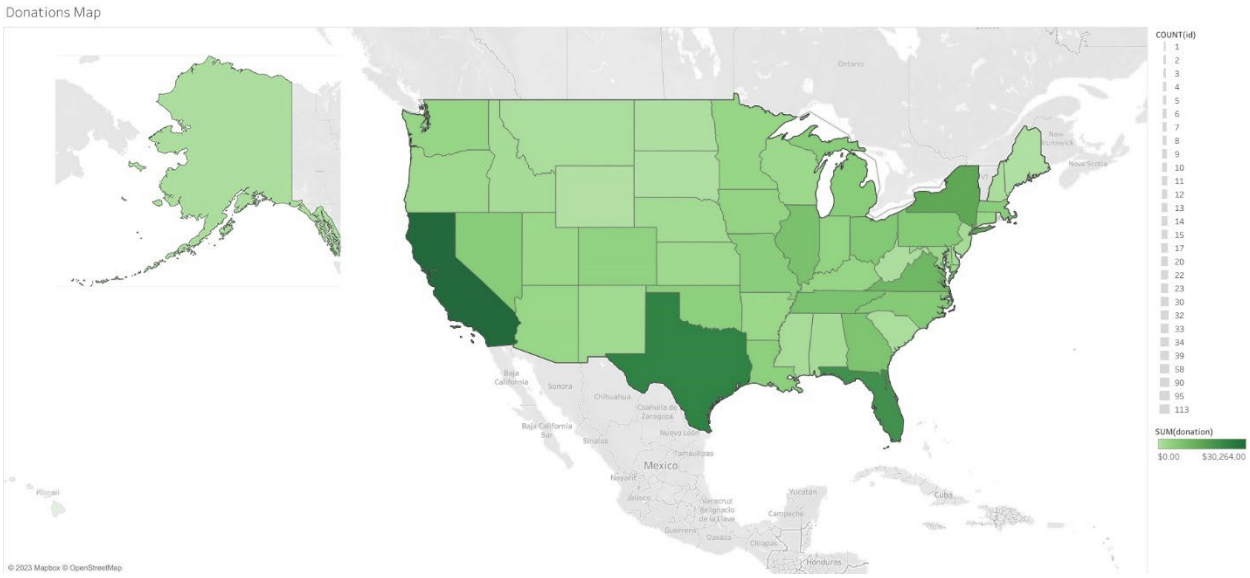
```
SELECT Donation_Data.donation, Donation_Data.gender, Donation_Data.state,
Donation_Data.job_field, Donor_Data2.donation_frequency,
Donor_Data2.university
FROM Donation_Data
JOIN Donor_Data2
ON Donation_Data.id = Donor_Data2.id
ORDER by donation DESC
LIMIT 10;
```

donatn	gender	state	job_field	frequency	university
500	Male	Michigan	Support	Yearly	Walasik
500	Male	New York	Product Management	Monthly	Leithgoe
499	Female	Virginia	Legal	Yearly	Sparhawk
499	Female	Delaware	Sales	Yearly	Antoszewski
498	Male	Wisconsin	Sales	Monthly	Trotton
497	Male	New York	Research and Development	Weekly	Rockcliffe
494	Male	California	Support	Weekly	Cominetti
494	Female	California	Product Management	Monthly	Coates

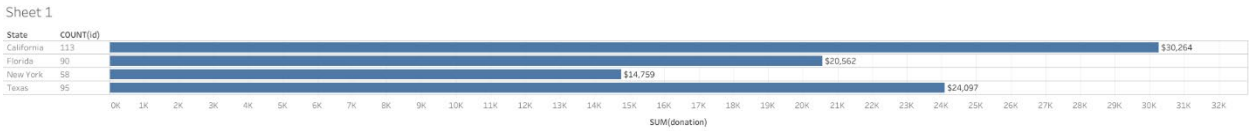
494	Male	California	Human Resources	Monthly	Baumber
493	Male	Maryland	Product Management	Monthly	Armatidge

Visualizations

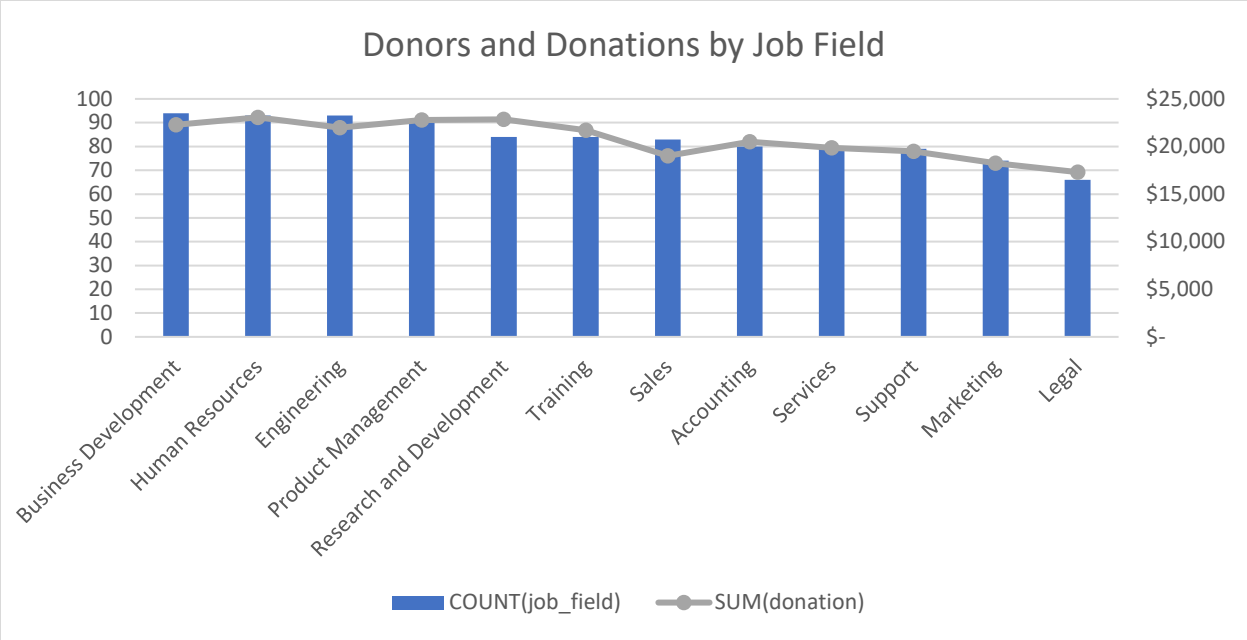




On this map you can see the states with the highest total donations highlighted in the darkest green shade. California, Texas, and Florida stand out as those generating the highest amount of donations. New Yorkers contribute at high levels as well.



According to [an article published] on The Ascent, Florida is attracting the nations highest earners because there is no state income tax and housing prices are affordable. The nation's high-income earners also live in California, Texas, and many in New York as well.



job_field	COUNT
Business Development	94
Human Resources	93
Engineering	93
Product Management	90
R and D	84
Training	84
Sales	83
Accounting	80
Services	80
Support	79
Marketing	74
Legal	66

job_field	donations
Human Resources	$ 23,060
R and D	$ 22,862
Product Management	$ 22,798
Business Development	$ 22,266
Engineering	$ 21,968
Training	$ 21,721
Accounting	$ 20,504
Services	$ 19,858
Support	$ 19,475
Sales	$ 19,009
Marketing	$ 18,255
Legal	$ 17,309

Findings and Recommendations

From my analysis of the two datasets provided I found

- There are $1,000 donors.
- Collectively they have contributed $249,085.
- The highest donation amount received is $500.
- The lowest donation amount received is $5.
- The average donation is $249.09.
- There are more female donors than male but collectively male donors have contributed more.
    - 492 donors are male and contribute a total of $127,628.
    - 508 donors are female and contribute a total of $121,457.
- States with the most donors are California, Texas, and Florida.
    - California has 113 donors contributing a total of $30,264.
    - Texas has 95 donors contributing a total of $24,097.
    - Florida has 90 donors contributing a total of $20,562.

California is where the country's highest income earners live, which might help to explain why many of our donors live there and contribute the most in donations. However, because the cost of living is high in California, many high-income earners are moving to more affordable locations like Texas and Florida where there is no state income tax, and the housing is more affordable.

The top states mentioned above may continue to be great places to find new donors but we should keep in mind that our most generous donors may migrate to more affordable states in the coming years. States that are experiencing a positive net change in high earning households like Arizona, North Carolina, and South Carolina may be good places to concentrate efforts.

The Job Fields with the most donors are Business Development, Human Resources, Engineering, Product management, Research and Development, and Training. These fields are also where the most donations come from although in a different order. The Job Fields with the least donors and donations are Sales, Accounting, Services, Support, Marketing, and Legal.

I gather from this that the top donors work in fields where they would have the greatest interest in hiring, training well educated, highly qualified new applicants. The people who participate in a company's success internally or their companies' own talent pools are the concern of those who work in those top six job fields. Those that work in sales, services, support, and marketing are more interested in customers than people who contribute internally. Those who work in accounting and legal help to maintain their businesses' success by supporting the legal and financial structure of the company.

Highlighting that Education for All is helping to support bright minds for the next generation of workers would be a strong selling point that would be especially relevant to those who work in job fields where we currently receive the most donors and donations.

Conclusion

Education for All has collected $249, 085 from 1,000 generous donors. The datasets provided give some insight into our success so far and where we might increase our efforts next year.

So far, we have reached interested parties in 48 states and the District of Columbia (We don't yet have donors in Vermont or Rhode Island). In the states where the highest income earners live, we have reached out to the most donors and collected the most money from these states. We should continue our efforts in these places as we reach out to potential donors in every state.

Our most generous donors have attended university but out of the entire pool of donors there are some who have not attended university and still contribute to our cause.

Our donors also work in a variety of different job fields. It seems that those who contribute the most tend to be those who work in recruitment and retention of highly qualified employees as well as those who work in fields were developing teams of highly educated people is most advantageous for successful projects.

The information we have on donors so far is somewhat helpful. We can start an email marketing campaign to let our donors know how we are using their donations, what our plans are, and how they can help by increasing the frequency and amount of their donations and by sharing news about Education for All with their friends and colleagues. We can also ask for input from them through polls and surveys. Because we know what colleges they went to we can reach out to alumni associations and local organizations with similar goals.

Going forward, we should consider collecting more useful information to better understand what incentivized our donors to contribute and what they would like to see because of their support. Information like the date the donor signed up, what event or campaign did they respond to, and whether a gift was offered in return for donations or special recognition.