The State University of New York - Plattsburgh
School of Business and Economics

Capstone - MSA 590
Spring 2020

Taproot Foundation: Data Challenge

Project Report

Submitted By:
Philip Middleton
Siddant Mehta

Submitted to:
Dr. Rohit Rampal

Submission date:
May 14, 2020

Table of Contents

Teradata/Taproot Foundation Data Challenge

Executive Summary

Taproot foundation is a non-profit company that provides technical support pro bono for other nonprofits. The Taproot Foundation and Teradata teamed up to provide a data challenge for students to test their skill on spreadsheets from a real corporation. Along with the spreadsheets is about 15 questions that can be answered using this data. Many of the questions have broad topics that allow for some analytical creativity.

Our methodology is to create visuals that can answer these questions provided to us. We believe that the visuals are much more interpretable for all involved in the company. Many of the major decision makers in companies do not have an analytical background, so our goal is to provide an analytical understanding of business processes in an easy to understand visual. To add to the question, we also created more analytical questions that can be applied to a deeper understanding of completed projects, sentiment about the project description, and which service or social media network provides the most traffic for Taproot.

In our analytics, much of what is asked of us is to look at frequencies of different aspects in Taproots Business. When Taproot looks into their email data they want to look into the success of an email. What does success mean when looking into emails? We determined different variables and created visuals to understand how they can interpret success in emails. For example, we can see the breakdown of which month is correlated with the highest project recommended emails. Other types of analysis examined types of projects, project completion rate, and project success with more than one volunteer.

The goal of our analysis is to apply practical implications to Taproot. Taproot provides technical support, but in a larger non-profit company not everyone is technical. In fact, many of the projects that they do are in public relation, business planning and marketing strategy (check for other projects in Appendices). The visuales that are created provide a practical understanding of areas of their business that are doing exceptionally well and other areas that could need work. This can help Taproot become more well rounded as a company.

Introduction

The Taproot Foundation is a non profit organization that works with other non profit to help build a stronger IT infrastructure. The types of projects that they engage in involve design, marketing, IT, strategic management, and human resources. Each of their projects helps professionals as pro bono work to help the other nonprofit organizations. Taproot provides technical aid to nonprofit organizations who volunteers don’t have the technical skill necessary to stay up with the technical times. Taproot and Teradata partnered to challenge data students to answer questions using analytical skills. Taproot was kind enough to give Teradata multiple datasets that the students will need to interpret in order to answer the questions needed for the challenge. This type of challenge is a great experience for students to get because datasets in the working world are not all as nice and easy to use as the dataset that students get in classes. This data challenge provides Taproot with different types of analysis from different students and helps the students learn what datasets in a workplace will be like. It is a win win for all parties involved in the data challenge.

Data

Taproot provided multiple different datasets to Teradata for this challenge. Each folder provided carries its own datasets that relate to different aspects within the Taproot business. The five folders include email notifications, project data, project recommendation data, session data, and user activity data. Since the students are not familiar with the business or their terms in the datasets, Taproot provided a data dictionary as well. This helps students work through the datasets and understand each variable. Many of the questions that are asked involve the type of project, the satisfaction rating, the completion time, and the project categories. These variables can help anyone in the foundation understand the success of the foundations or in which areas the foundation needs to work on. These significant variables are key metrics when determining success of the foundation and each of their projects.

Analysis

Analysing open rates, click thru, bounce by email type and user type (volunteer and nonprofit)

Taking a look at the data-sets provided by the Taproot Foundation, we decided to focus our analysis for the question stated above by utilizing the Email Data Set. In order to analyze open rates, click thru, bounce by email type and user type; we utilized the variables Kind, Notifiable Type, Open, Bounce, Delivered, Scheduled For and Created At. Variable "Kind" and "Notifiable Type" provides us with the observations for email type and user type (volunteer and nonprofit) stated below;

- **Nonprofit:** *inactive_nonprofit, inactive_since_nonprofit, nonprofit_webinar, pl_nonprofit_active, pl_nonprofit_inactive,*
- **Volunteer:** *inactive_volenteer, recruiter_volunteer_notification, volunteer webinar, two_hundred_million_inactive_volunteer*

Taking a closer look at the variable "Notifiable Type", *User* Is the only observation available when using the variable "Kind" to filter for Volunteer and Nonprofit.

Please see the formulated calculations below for open rates, click thru and bounce. For Open Rate, we utilized the variable "Open" and "Delivered". In order to calculate Bounce Rate, we utilized the same methodology in calculating the variable "Bounce" by "Delivered", "Processed", and " Scheduled For" separately.

Open Rate:

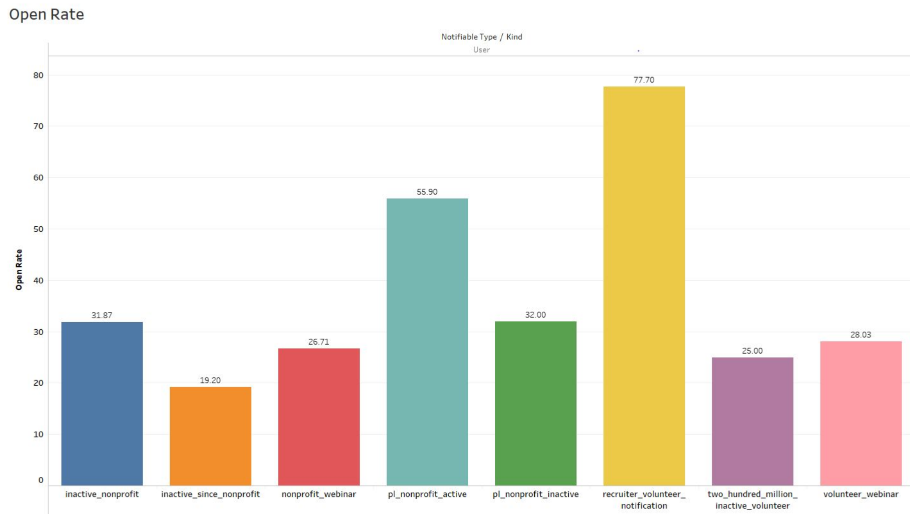
- **Forumal: COUNT([Open])/COUNT([Delivered])*100**

  "Open rate and **click through** would be calculated the same way based off the information given**"**

Bounce Rate: Only 688 observation

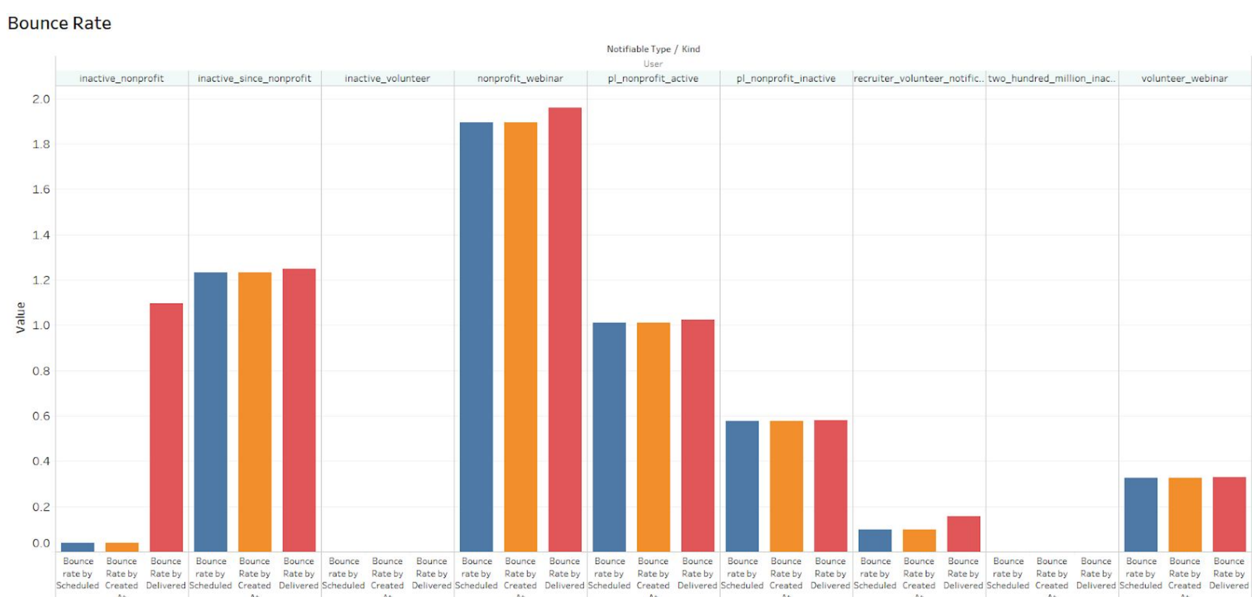
- **Forumal: COUNT([Bounce])/COUNT([Scheduled For])*100**
- **Forumal: COUNT([Bounce])/COUNT([Created At])*100**
- **Forumal: COUNT([Bounce])/COUNT([Delivered])*100**

The chart below represents our results for open rate. As you can see, "recruter_volunteer_notification" has the highest open rate of 77.70%. "inactive_since _nonprofit" has the lowest open rate of 19.20%. We can conclude that inactive users have the lowest open rates followed by users with webinar(volunteer and nonprofit).



inactive_nonprofit	31.87
inactive_since_nonprofit	19.20
nonprofit_webinar	26.71
pl_nonprofit_active	55.90
pl_nonprofit_inactive	32.00
recruiter_volunteer_notification	77.70
two_hundred_million_inactive_volunteer	25.00
volunteer_webinar	28.03

The chart below shows us the distribution for bounce rate for each user email type by "Delivered", "Created At", and " Scheduled For" Although "Created At" and "Scheduled For" have the same results, we do see a differentiation in bounce rate by "Delivered". Non-profit webinar has the highest bounce for the emails that are delivered. We can see that bounce rate is either low or nonexistent for inactive users.



Bounce Rate

Notifiable Type	Kind	Bounce rate by Scheduled	Bounce Rate by Created At	Bounce Rate by Delivered
User	inactive_nonprofit	0.040	0.040	1.095
	inactive_since_nonprofit	1.234	1.234	1.250
	inactive_volunteer	0.000	0.000	
	nonprofit_webinar	1.896	1.896	1.960
	pl_nonprofit_active	1.010	1.010	1.026
	pl_nonprofit_inactive	0.576	0.576	0.580
	recruiter_volunteer_notification	0.098	0.098	0.155
	two_hundred_million_inactive_volunteer	0.000	0.000	0.000
	volunteer_webinar	0.327	0.327	0.329

In conclusion, we can see that inactive users are still in the system. Recruiter Volunteer Notifications seem to be the most predominant usage in this interface. The concluding chart below represents Open Rate and Bounce Rate for Emails Delivered(each user email type).

Open Rate and Bounce Rate for Emails Delivered

Notifiable Type	Kind	Bounce Rate by Delivered	Open Rate
User	inactive_nonprofit	1.09	31.87
	inactive_since_nonprofit	1.25	19.20
	inactive_volunteer		
	nonprofit_webinar	1.96	26.71
	pl_nonprofit_active	1.03	55.90
	pl_nonprofit_inactive	0.58	32.00
	recruiter_volunteer_notification	0.16	77.70
	two_hundred_million_inactive_volunteer	0.00	25.00
	volunteer_webinar	0.33	28.03

<u>Analysing comprehensive breakdown of project recommendations</u>

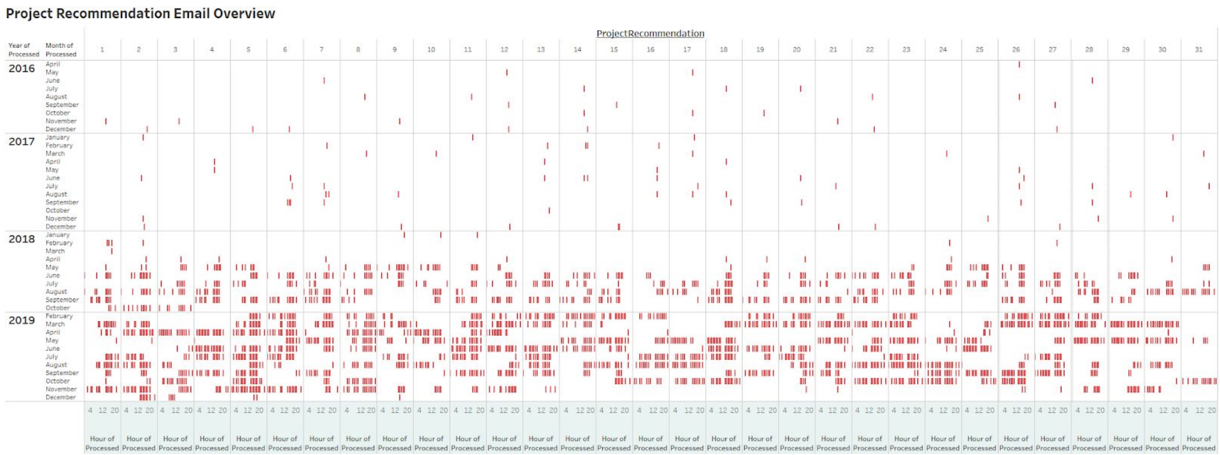
To analyze a comprehensive breakdown of project recommendations, we used the Email Dataset with the following variables;

- **Processed:** Contain information regarding year, month, day, and time. We only want to look at emails that are processed for more of an accurate result.
- **Notification Type:** This variable is used to filter for "project recommendation" and taking "Sum of Records" (Note: there are 410,664 Project Recommendation observations after filtering Notification Type)

Tableau: Created a Visualization that shows project recommendation by year, month, day, and time(hours).

- **Project Recommendation Email Overview**
- **Project Recommendation Email by** Day Sent
- **Project Recommendation Email by** Month
- **Project Recommendation Email by** Day of the Week
- **Project Recommendation Email by** Quarter for 2018 and 2019

The chart below provides us with an overview of project recommendation email by year, month, day and time. This chart is a great illustration of activity that has occured for project recommendation email over a periodic time. We see that project recommendation emails were not significant in the year of 2016 and 2017. We see a significant incline of project recommendation emails in May of 2018 all the way to the end of November 2019.



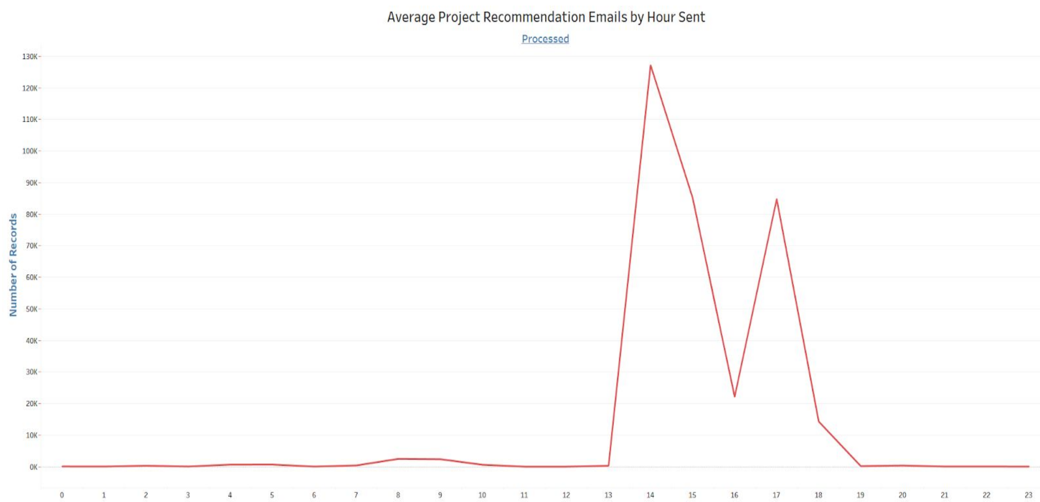
The chart below represents project recommendation emails by year and quarter.

Project Recommendation by Year and Quarter

2016			2017				2018				2019			
Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
5	10	16	14	15	25	13	10	9,791	16,670	575	58,925	107,758	99,403	49,063

The line graphs below represent average project recommendation emails by hour sent and day sent.The graphs take into consideration the average activity for all the years and months; as well as days for hour sent. As you can see, project recommendation emails are predominantly sent between the hours of 1pm and 7pm. There is a significant increase in project recommendation emails at 2pm and then 5pm. There is generally more activity at the beginning of the month compared to the end.

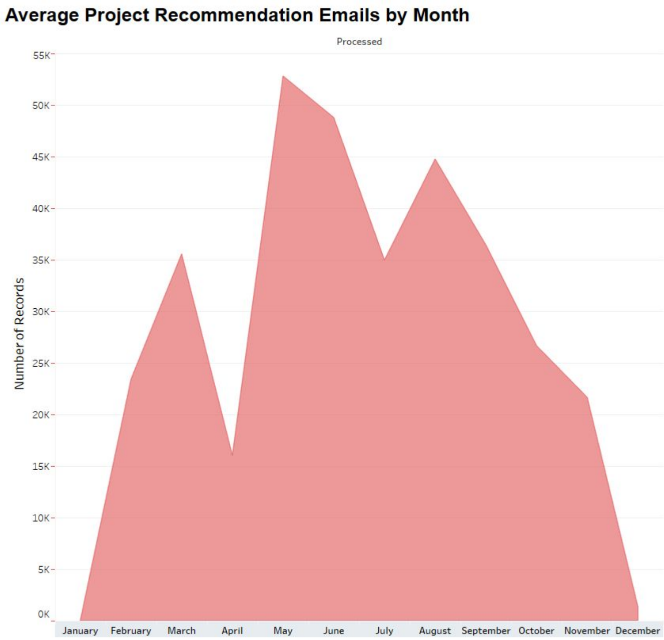
Day of Processed	
1	8,622
2	11,826
3	7,465
4	10,005
5	21,126
6	15,335
7	11,811
8	19,393
9	7,716
10	6,052
11	16,360
12	11,784
13	13,196
14	11,917
15	9,148
16	9,980
17	6,942
18	13,837
19	10,512
20	12,044
21	10,149
22	13,449
23	14,063
24	11,380
25	7,292
26	9,591
27	12,003
28	8,996
29	7,607
30	7,310
31	5,382



Hour of Processed	
0	65
1	41
2	309
3	56
4	653
5	675
6	35
7	396
8	2,478
9	2,366
10	607
11	7
12	8
13	292
14	127,202
15	85,176
16	22,159
17	84,792
18	14,307
19	184
20	365
21	47
22	47
23	26

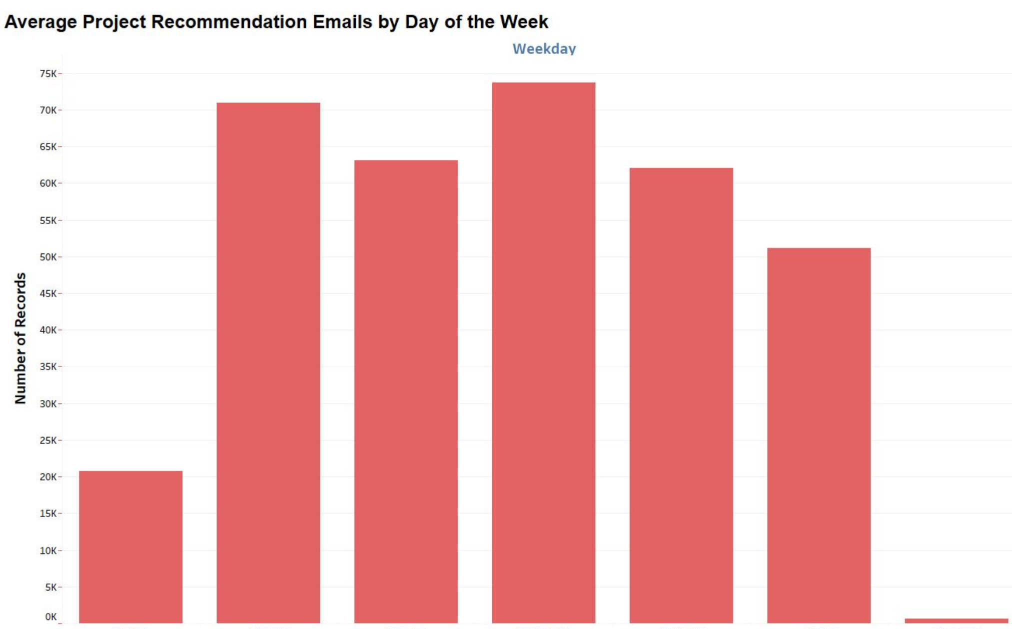
The tables to the right of the line graphs provide us with the average *number* for project recommendation emails by hour sent and day sent.

The graphs below represent the average project recommendation emails by month and week. Month of May has the highest project recommendation emails. There seem to be a staggered normal distribution for project recommendation emails from the month of January to December. Taking a look at the days of the week, Wednesday has the highest flow in average project recommendation emails. The weekends have the lowest average of project recommendation emails with Saturday being the bare minimum.



Month of Processed	
January	7
February	23,394
March	35,548
April	16,017
May	52,785
June	48,767
July	34,946
August	44,739
September	36,423
October	26,660
November	21,649
December	1,358

The tables below provide us with the average *number* for project recommendation emails by week and month.



Weekday	
Sunday	20,787
Monday	70,949
Tuesday	63,106
Wednesday	73,713
Thursday	62,005
Friday	51,150
Saturday	583

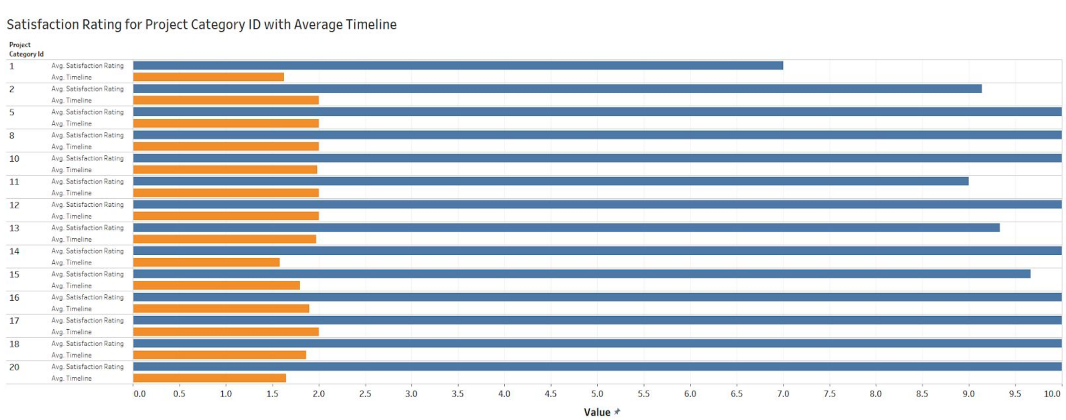
In order to increase the flow of business, some departmental changes must have been made in May of 2018 as we see an incline from 10 project recommendation emails in the month of April to 74 project recommendation emails in the month of May. The increase of project recommendation emails is steady throughout the remainder of the year all the way into the end of 2019. March of 2019 has the highest occurrence of project recommendation emails (at 280). No project recommendation emails were sent in December 2018. Consequently, only 15 project recommendation emails were sent in December of 2019 compared to November 2019 which had 150 project recommendation emails. An assumption can be made saying, due to holidays in December, there is a steady decrease of project recommendation emails. The 5th and 8th of every month (Year 2016 to 2019) has the highest sum of project recommendation emails at 40,519 combined.

<u>Percentage of nonprofit and volunteers users drop off prior to completing registration</u>

In order to calculate the percentage of nonprofit and volunteer users that drop off prior to completing registration, we used the Conversions Dataset. We utilized the pivot table below to run our analysis. The pivot table consists of three variables(Channel, ID, and User ID).Taking a look at Row Label and Count of id, it provides us with channeled components along with a count of users that can either be voluntary or nonprofit. Using the "GETPIVOTDATA" function, we can calculate the grand total of “Count of id” along with its users(volunteer and nonprofit). We then subtract the results of 57800 by 125 and divide by 57800 to get the total percentage of drop off at 99.78%.

<u>The project types that produce the best outcomes in survey’s</u>

In order to analyze project types that produce the best outcomes in survey’s, we utilized the Project Export Dataset to run our analysis. The variable “Satisfaction Rating” provides us with a range of satisfaction from 1 through 10. Project type can be categorized by the variable “ Project Category Id” which ranges from 1 thorough 20. We also included an averaged timeline for each project.The graph and chart below represent the satisfaction rating for project category ID with their average timeline.



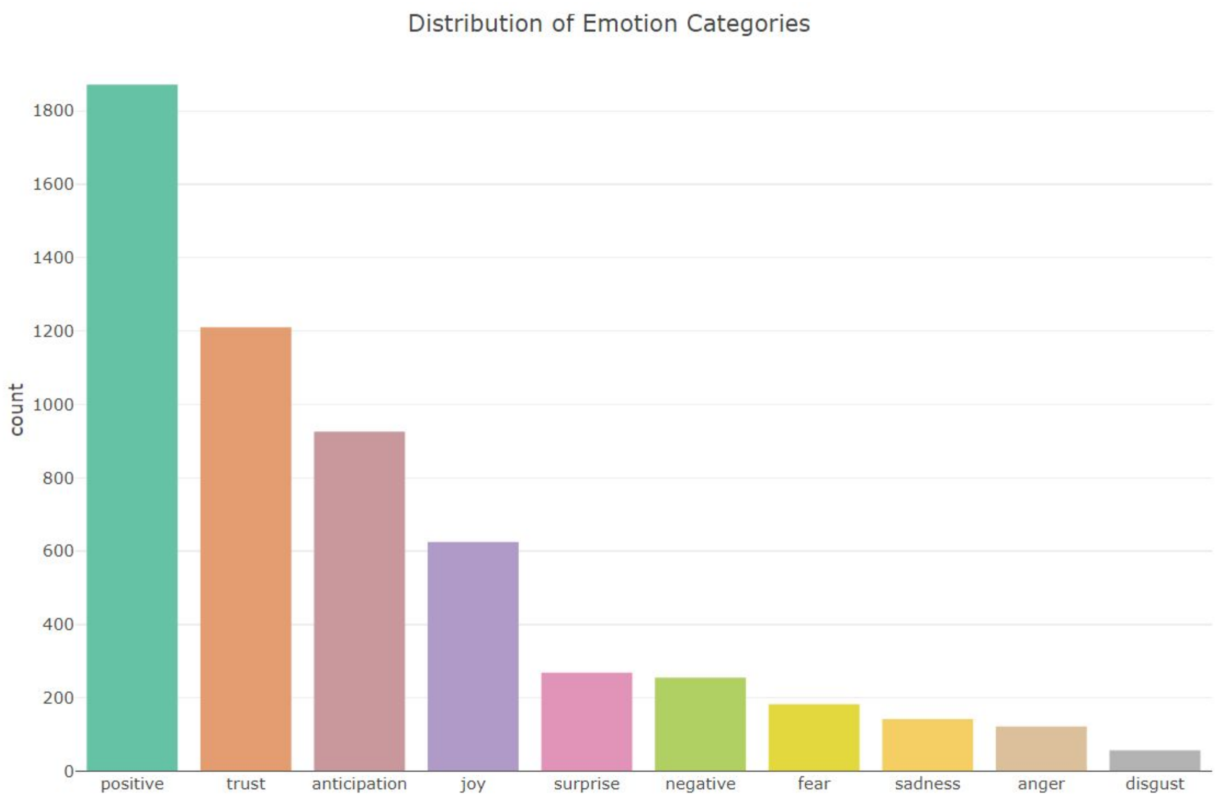
Project Category Id	Avg. Satisfaction Rating	Avg. Timeline
1	7.000	1.625
2	9.143	2.000
5	10.000	2.000
8	10.000	2.000
10	10.000	1.985
11	9.000	2.000
12	10.000	2.000
13	9.333	1.974
14	10.000	1.581
15	9.667	1.800
16	10.000	1.900
17	10.000	2.000
18	10.000	1.867
20	10.000	1.647

As you can see, there doesn't seem to be a direct correlation between the average timeline and satisfaction rating. Project 1 has the lowest satisfaction rating whereas project 5,8,10,12,14,16,17,18,19,20 have equaliventy the highest. We can conclude by saying that all the projects have above average satisfaction ratings.

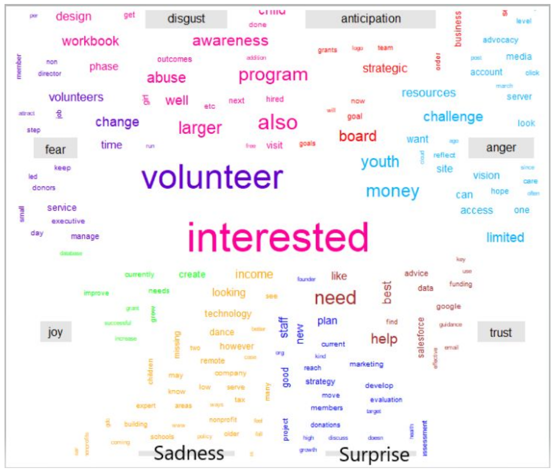
Provide an overview of sentiments associated with completed projects

In order to answer the question above, we utilize the Session Export Data sets. We first filtered the data sets for projects that were completed and then ran our analysis using R Studio. We have a variable in our data set called " description" where it provides us an understanding of completed projects. The objective of this analysis will be to focus on driving sentiments from the " description" column to gain a better understanding of the emotional tone behind the series of words presented. Focus will be to look at completed projects from the years 2016 to 2019.

We start off our analysis by going through a vigorous data cleaning process. Once the data has been cleaned, we utilize sentiment NRC to obtain the direction of emotions associated with completed projects. The chart below represents the distribution of the emotions by category. As you can see, the first four predominant sentiments are " positive", "trust", "anticipation", and "joy". Due to our results, We can state that there is happiness and excitement towards projects that have been completed.



The word cloud below represents the most frequent words associated with the distribution of categories.



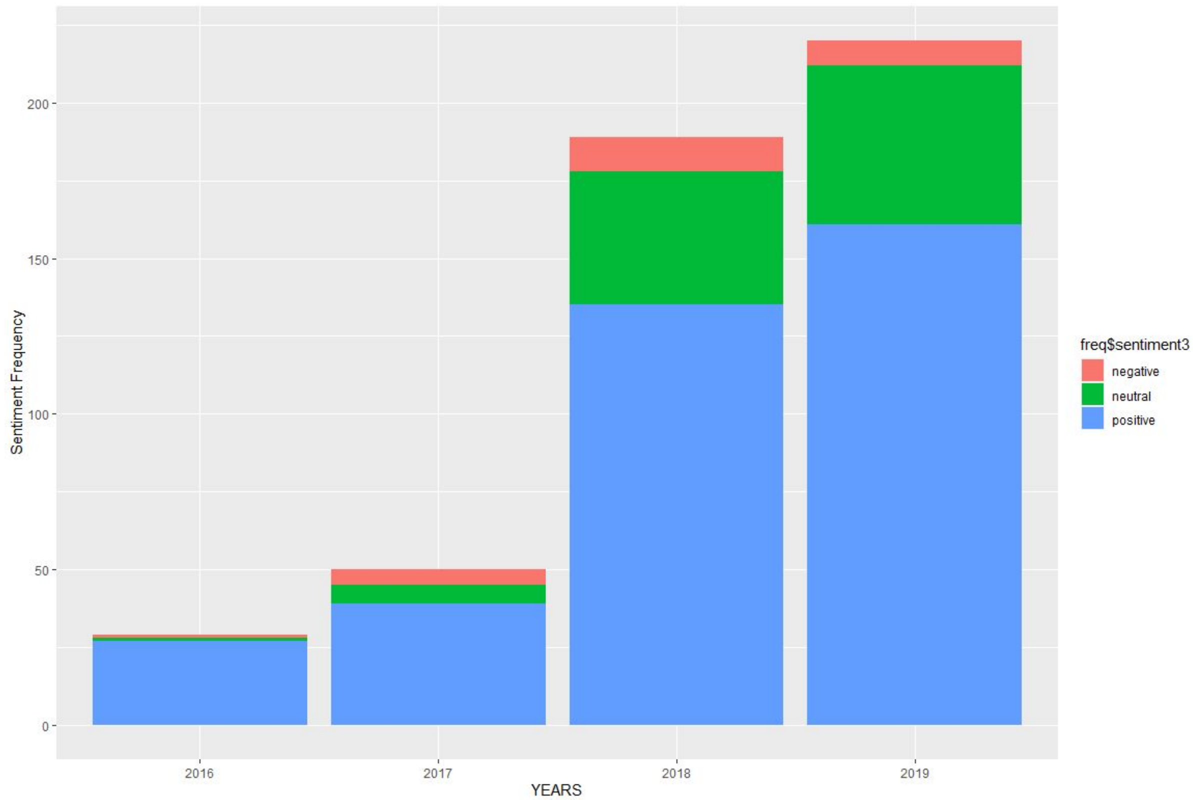
The charts below portray an overview of the sentiments for completed projects by year. As you can see, the first chart represents the mean sentiments by year. The year 2017 has the highest mean sentiment compared to the year 2018, where it has the lowest. Our results for all four years show positive sentiment which dictates satisfaction amongst the candidates descriptions. Second and third charts provide us with the frequency of positive, negative and neutral emotions for all four years. Our results indicate that 2019 has the highest completed projects whereas 2016 had the lowest. One factor to note, sentiments generally ranges between 1 and -1. Although our analysis is out of range, it still provides us an idea of the general sentiment.

Chart 1 & 2

	date	meanSentiment
1	2016	2.72413793103448
2	2017	2.78
3	2018	1.55555555555556
4	2019	1.64090909090909

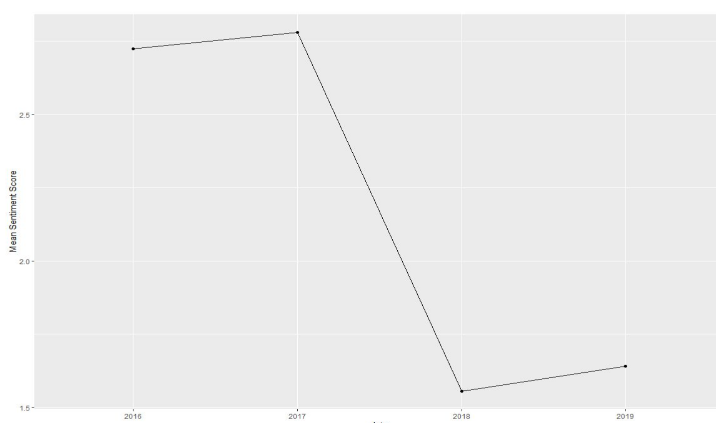
	date	negative	neutral	positive
1	2016	1	1	27
2	2017	5	6	39
3	2018	11	43	135
4	2019	8	51	161

Chart 3



The fourth chart provided us a visualization of mean sentiment score by year. Although 2016 and 2017 had the lowest frequency of completed projects, we can see that it also provides us the highest mean sentiment score. Since 2018 and 2019 had the highest number of observations, we can state that the mean sentiment score for the last two years provides us with the most accurate result.

Chart 4



Our analysis was a great representation of how candidates portray their emotions for the projects completed. The year 2019 portrays the most accurate results since the number of observations exceed the prior years. This analysis can be conducted for future years as the number of completed projects increase. Overall, we can conclude by saying that the completed projects portray a positive outlook for the organization as a whole.

These are the projects that take the longest and shortest to complete

One integral part of business is knowing how long each project will take. Knowing this is a critical part in planning and forecasting for Taproots business. Each type of project may take longer or shorter than expected but when talking with clients it is helpful to have a ballpark estimate of how long the project is going to take.

Our method of examining which is the longest and shortest projects to complete involved creating a new spreadsheet, filtering the data then interpreting the results. We took the project categories data that was provided and copied the data over to a new sheet. In this new sheet we put a filter on the *created_at* and *updated_at variables.* With these two filters we can determine which project takes the longest and shortest to complete.

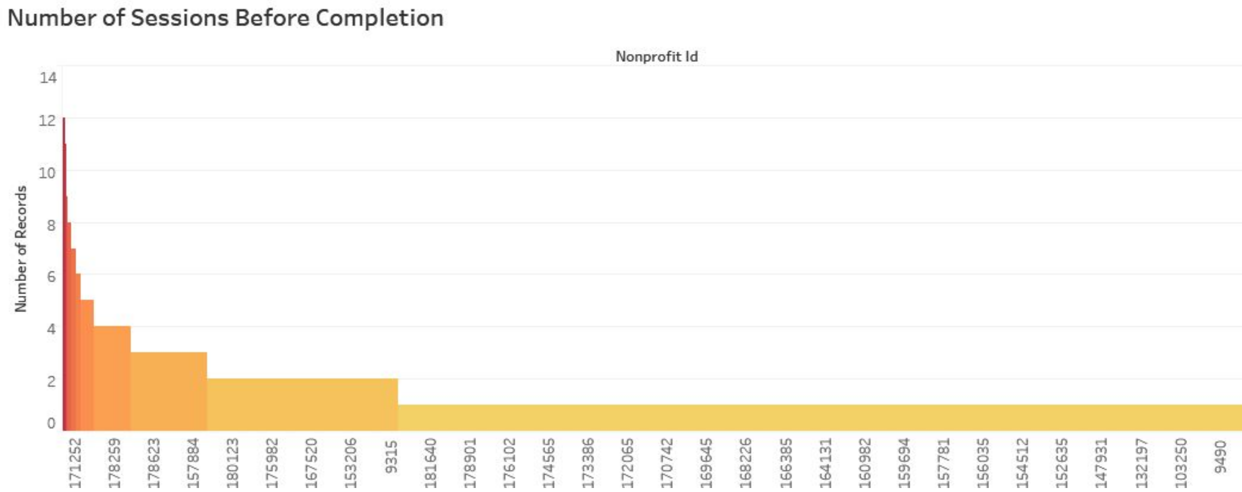
The results show that the IT infrastructure projects take the longest to complete. In comparison, Accounting and Financing projects take the least amount of time to complete. Another interesting insight that we caught onto with this data is that their *created_at* variable all have similar time stamps on each project. Each project was started on 04-04-2016 and then the *updated_at* variables changed. With this information we can conclude that Taproot uses a small sample size of projects to give an estimate of how long or short a project will take

Amount of sessions it takes a non-profit to complete a project

Taproot would be interested in knowing how many sessions each project would take on average. This information can be helpful to know when scheduling projects and also integration strategies when helping the other non profit companies. We looked into the session export dataset that Taproot provided. In this dataset we can see each project, when they were created at, completed at, the description of the project and other important variables that would help us find the average amount of sessions a project would take to complete.

With all of this information our strategy is to identify and filter the variables that were most important when looking into this question. Taproot is looking into non-profit projects, so we took the non-profit ID number and the count of each non-profit ID which related to the amount of sessions each non-profit company had before completion.

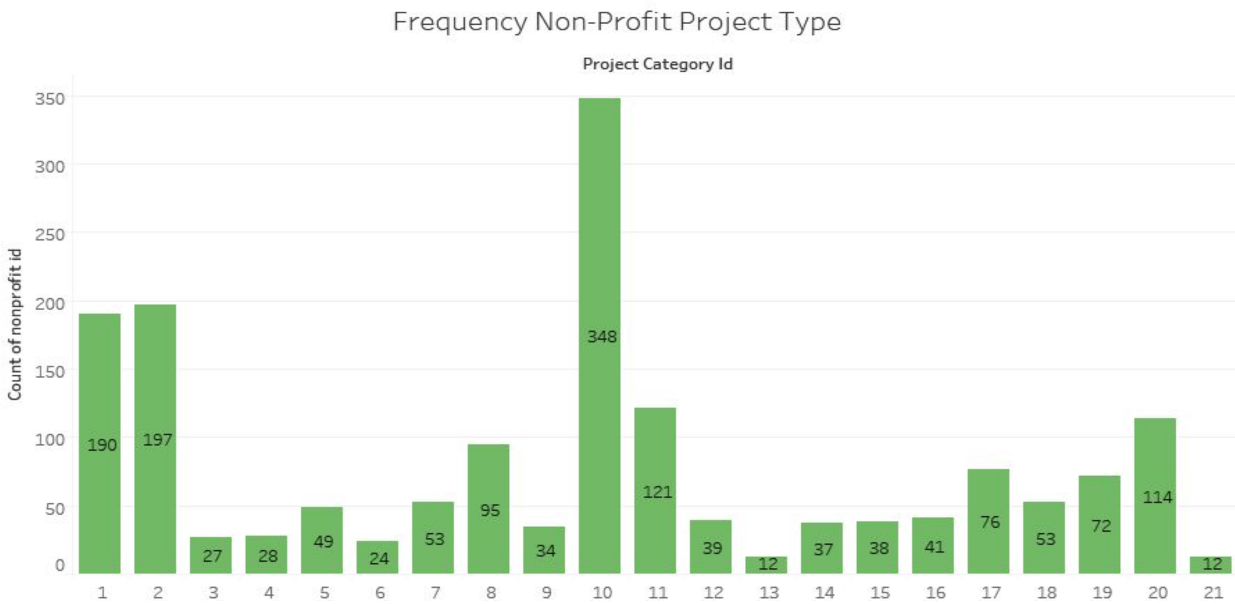
At this point we took the average of all the project sessions that all of the non-profit companies had. The average amount of sessions it takes to complete a non-profit project is 3 sessions. The graph below shows that the range goes from 12 sessions to 1 session. We can successfully say that it takes an average of 3 sessions to complete a project for a non-profit organization.



Amount of nonprofits that are conducting the same projects

To find out the frequency of nonprofits that are conducting the same project can help Taproot focus their efforts towards the most frequent task. To find this information we looked at the sessions export datafile. This dataset contains a lot of information about project types, the person working on that project, the non-profit ID, and other contributing factors to an understanding of the different projects.

To best find the results, we needed to identify the best variables to answer the questions. We put the data into a pivot table, this allows us to filter the dataset by non-profit ID. Each non-profit ID is correlated to the project that the non-profit company is working on. The projects are labeled numerically. To find out what each project is we looked at the project category dataset to see which each number and their corresponding project type. We found that there are 22 different types of project (see in appendix). We plotted this information as a barchart to visually see the highest frequency of projects.

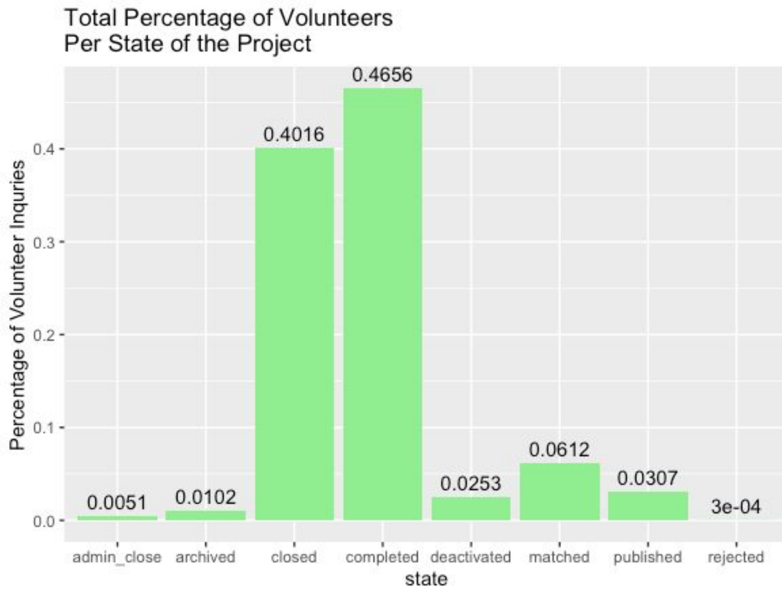


As we can see from this bar chart, project 10 (see in appendix) is the most frequently used project by non-profit companies. Project category ID number 10 (see in appendix) is Market Strategy project, followed by business planning (project ID 2) (see in appendix), and Accounting & Financing (project ID 1) (see in appendix). We are able to show that Taproot needs to focus their energy mainly on these project lines because they are the most frequently used projects by non-profit companies.

Project Rates with Multiple Volunteers

In theory, projects with multiple volunteers generally have the highest project completion rate. Our goal is to test this theory. With larger projects the more volunteers they have can create an easier work flow for the client and for Taproot volunteers. This is a great way to increase efficiency with your business if you have extra volunteers that are ready to work.

The approach we took to answer this questions started with finding the datasets. We used the *project_export* datasheet and the *project_categories* data to run analysis in R coding language. We identified *id, description, state, and project_inquiries_count* as significant variables in the *project_export* data. We filtered this information by *project_inquiries_count* and then grouped them by their project *state.* Our next step was to look at the total amount of volume for this project and their state to get an accurate count for rate for projects with multiple volunteers.

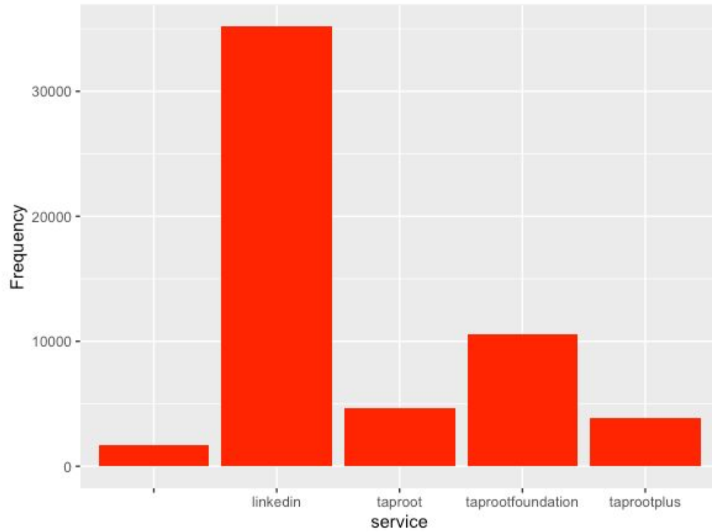


From this chart we can see that there is a much higher rate of completed projects with multiple volunteers. We see that out of all the projects that have more than 1 volunteer inquiry is just under 50% of those projects ended up completed. Which stresses the importance of having one or more volunteers per project. Even a smaller percentage had projects published.

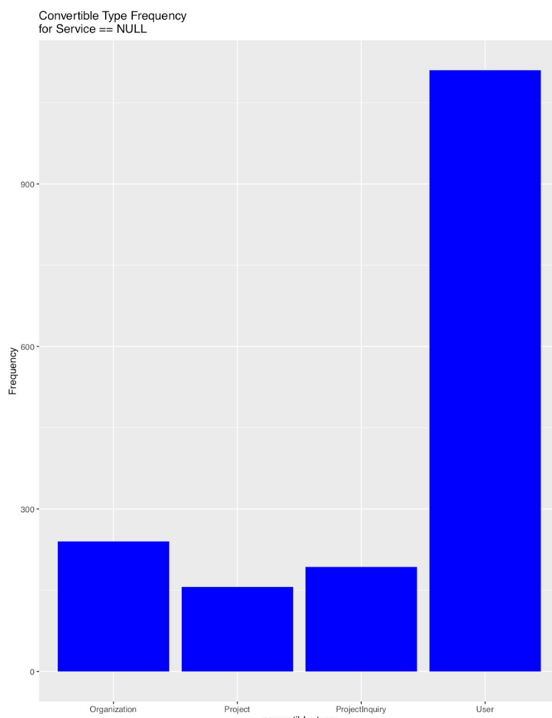
<u>Top Services for Projects</u>

Using the conversions dataset we wanted to know what is the top service that Taproot and client use for projects. These are services that projects are mentioned on. This can be helpful for taproot to know where their projects are being talked about the most. This can be helpful for their marketing strategy to see which platform to focus more of their marketing to. As a non-profit company word-of-mouth marketing is very important. From this we can also see the conversion to a completed project based on the conviersions dataset we were provided.

Our first plunge into this dataset was to remove NAs. This step significantly reduced our dataset. After removing NA's, we see that the only projects left are predominantly posts from facebook or twitter. This gave us insight that projects shared through social media are more neat and contain more information on project details. Projects on social media tend to have more fill in information on this spreadsheet.



The bar chart proves that social media dominates project details in the data. Linkedin dominants as the main service for projects, as taproot and other taproot extremities are the other popular outlets on the taproot platform, which makes sense. However, we see that a service with no name has the 5th highest frequency, this could be attributed to that fact that a project is missing information on what service it is being completed through. Further investigation shows that services that haven't been identified or entered in properly have a corresponding NA for the user_id.



Conclusion

To conclude our analysis with Teradata and Taproot we can successfully determine that Taproot is on an overall upward trend. Many of the projects that they start are completed, especially projects that have more than one volunteer on the job. We can confirm Taproots growth from 2016 to 2019 with the sentiment barchart. In this barchart we are able to see the upward trend and the drastic increase in completed projects in 2019. One area of improvement that we found that Taproot could work on is data completion. In our *Top Service of Projects* analysis section, we can see that projects with no name have a very high frequency. This can cause problems for their corporate organization or they can lose an important project because the data is not completed. Taproot will continue to grow because they provided a helpful service for nonprofits that do not have the technical infrastructure to handle this day and age.

In this data challenge our team learned a lot about corporate datasets. These datasets are messy and do not provide adequate direction about the data. Many time variable names are different or they are the same name for different meanings. We ran into this issue with the "ID" variable. In each dataset there was an ID variable but did not always mean the same type of ID, it could be the employee ID, the project ID, or some other ID for all we know. How we moved past this was by combining variables that we knew were the same, for example, nonprofit ID was a common variable that we used to provide analysis. As data analysts we will run into difficult datasets and this capstone project gave us experience in understanding messy data, cleaning data, and then performing analysis to answer larger business questions. Team Chocolate Milk is more prepared for our future as data analysts and we have found it very beneficial to see the valuable lessons we have learned in real world situations.

Appendix

Project Category Chart

id	name
1	Accounting & Finance
2	Business Planning
3	Evaluation
4	Project Management
5	Program Design
6	Research
7	HR Management
8	Board Development
9	Staff Development
10	Marketing Strategy
11	Brand Development
12	Messaging
13	Multimedia
14	Design
15	Copy writing/editing
16	Public Relations
17	IT Infrastructure
18	CRM
19	Website design
20	Website development
21	Mobile Development
22	Grant writing/development

Weekly Scrum

Product backlog for this project (by question):

1. ~~Please analyze Open rates, **click thru**, bounce by email type and user type (volunteer and nonprofit)~~
    - We will be using **EmailDataSet**
    - Variables to be used: Columns (14, open), (12, delivered), Filter through (2, kind) for volunteer and nonprofit.
    - (12, delivered), (10, processed), (3, scheduled_for) are identical
    - (3, scheduled_for) has less data missing but some dates are structured differently.
    - (16, bounce_reason) filter for inputs that have a reason.
    - (14, open) and filter for dates (make the open binary).

~~2. Please analyze Comprehensive breakdown of Project Recommendation email by day sent, day of month, time of year etc.~~

- We will be using **EmailDataSet**
- Variables to be used: Columns (7, notifiable_type) and filter for "ProjectRecommendations"
- (12, delivered), (10, processed), (3, scheduled_for) are identical
- (3, scheduled_for) has less data missing but some dates are structured differently.
- We took column (3, scheduled_for) and created columns "S-DAY", "S-MONTH" AND "S-YEAR"

3. Need to know Activation/reactivation trends for failed and closed projects ** **Did NOT Answer**
    - We will be using **project_export with potential integration with project injuries and project recommendation**
    - Variables to be used: column (F, completed) and filter for "closed"

4. ~~What % of nonprofit and volunteers users drop off prior to completing registration.~~
    - User activity Dataset

5. ~~Which project types produce the best outcomes in survey data?~~
    - Project Category and Project Recommendation

- Commons looked at (Satisfaction rating) from 1-10
- Project type (colomon)
- Project catorigary ID and you match it up with ID in Project Category
- For project names, we match it up with Project Type.

6. Which project types perform better on the better? Match rate, days to match, # of inquiries received etc. ** **Did NOT Answer**
    - Project categories, Project Inquires
7. Which nonprofit issues areas perform better on the platform? Match rate, days to match, # of inquiries ** **Did NOT Answer**
    - Conversions, All web analytics (Maybe)
8. ~~Which project types take the longest to complete? Shortest?~~
    - Use Project Catoragary Data set
    - Updated at and subtract it by Time created
    - Project categories(created at, updated at => long to short)
    - Shortest is not there!! Sort from shortest to longest
9. Which volunteer skills have the greatest success on the platform? Finding projects? Matching? Completing? Returning? ** Did NOT Answer
    - All web analytics, project categories

10. Are there any trends in user activity and location? ** **Did NOT Answer**
    - Conversions, All Webs, time series per office
11. ~~Any analysis of projects with multiple volunteers?~~
    - Project export
    - Project categories

a. Project and inquiry status rates for projects with multiple volunteers
b. Average days to complete for projects with multiple vols
c. Project title and description, type analysis for projects with multiple volunteers

12. ~~How many sessions does a nonprofit create before starting a project?~~
    - ~~Project inquiry, project categorization, project export.~~
13. ~~How many nonprofits are conducting sessions on the same or related subject?~~
    - ~~Project categorization, project inquiries, session exports~~

Note: Find and replace project category ID with ID from project category dataset.

Sprint

Week 8 04/01/20

~~Question 2: Please analyze Comprehensive breakdown of Project Recommendation email by day sent, day of month, time of year etc.~~

- We will be using **EmailDataSet**
- Variables to be used: Columns (7, notifiable_type) and filter for "ProjectRecommendations"
- (12, delivered), (10, processed), (3, scheduled_for) are identical
- (3, scheduled_for) has less data missing but some dates are structured differently.
- We took column (3, scheduled_for) and created columns "S-DAY", "S-MONTH" AND "S-YEAR"

~~Question 8: Which project types take the longest to complete? Shortest?~~

- Use Project Catoragary Data set
- Updated at and subtract it by Time created
    - Project categories(created at, updated at => long to short)
    - Shortest is not there!! Sort from shortest to longest

Week 9 04/08/20

~~Question 1: Please analyze Open rates,~~ **~~click thru~~**, ~~bounce by email type and user type (volunteer and nonprofit)~~

- We will be using **EmailDataSet**
- Variables to be used: Columns (14, open), (12, delivered), Filter through (2, kind) for volunteer and nonprofit.
- (12, delivered), (10, processed), (3, scheduled_for) are identical
- (16, bounce_reason) filter for inputs that have a reason
- (13, bounce) and filter for dates
- (14,open) and filter for dates (make the open binary)
- Comlom perform delivery - if the email makes into the recipient

Week 10____04/15/20

10. Are there any trends in user activity and location?
    - Conversions, All Webs, time series per office
    - We need to take more time on this question. This data is very messy.
11. Any analysis of projects with multiple volunteers?
    - Project export
    - Project categories
        - Look at absolute numbers or averages

i. Project and inquiry status rates for projects with multiple volunteers
ii. Average days to complete for projects with multiple vols
iii. Project title and description, type analysis for projects with multiple volunteers

~~Question 5: Which project types produce the best outcomes in survey data?~~
- Project Category and Project Recommendation
- Commons looked at (Satisfaction rating) from 1-10
- Project type (colomon)
- Project catorigary ID and you match it up with ID in Project Category
- For project names, we matched it up with Project Type.

Week 11____04/22/20

11. Any analysis of projects with multiple volunteers?
    - Project export
    - Project categories
        - Look at absolute numbers or averages

iv. Project and inquiry status rates for projects with multiple volunteers
v. Average days to complete for projects with multiple vols
vi. Project title and description, type analysis for projects with multiple volunteers

~~Question 12: How many sessions does a nonprofit create before starting a project?~~
- ~~Project inquiry, project categorization, project export.~~

~~Question 13: How many nonprofits are conducting sessions on the same or related subject?~~

- Project categorization, project inquiries, session exports

~~Question 4: What % of nonprofit and volunteers users dropoff prior to completing registration User activity Dataset.~~

Week 12____04/29/20

6. ~~Which project types produce the best outcomes in survey data?~~
    - Project Category and Project Recommendation

Created question: ~~Provide an overview of sentiments associated with completed projects ?~~

- Filtered through the project_export dataset for completed projects and ran our analysis using RStudio.

Created question: ~~Amount of sessions it takes a non-profit to complete a project~~

Created question: ~~Amount of nonprofits that are conducting the same projects~~

The beginning stage of preparing the final report

Week 13____05/06/20

We evaluated our final analysis and completed the project report .

Week 14____05/13/20

Gathering our documentations in preparation for final submission.