TERADATA University Network

DATA CHALLENGE

Teradata Data Challenge 2017

Rise Against Hunger

Authors

Author 1:

- *Isha Singh,*
- *University of Cincinnati,*
- *singhia@mail.uc.edu,*
- *Computer Science,*
- *Level: Master*

Author 2:

- *Shreya Kukreti,*
- *University of Cincinnati,*
- *kukretsa@mail.uc.edu,*
- *Computer Science,*
- *Level: Master*

Author 3:

- *Abhishek Dhanasetty,*
- *University of Cincinnati,*
- *Dhanasak@mail.uc.edu,*
- *Computer Science,*
- *Level: Master*

Faculty Advisor

- *Dr. Karen Davis,*
- *University of Cincinnati,*
- *davikc@ucmail.uc.edu*
- *Computer Science,*
- *Level: Master*

TERADATA University Network

DATA CHALLENGE

1) Abstract

Rise Against Hunger is an NGO that aims to stop hunger across the world. It packages and sends meals all over the world. As a part of our Data Warehousing coursework, we got to work on their data and provide some ideas based on our analysis. We chose to answer the following questions, **Who are their recurring Donors? And Do the giving amounts increases or decreases vary based on donor level? What are the levels ($ amount cutoffs)?** Our analysis gave us an insight into the recurring donors as well as additional information about them. This information can be used to increase their donations and help them move closer towards their goal.

2) Problem and Motivation

Problem: Who are their recurring Donors?
Definition: Since there is no definition mentioned for Recurring Donors, we have addressed the analysis of different types of Recurring Donors based on Frequency mentioned in the Opportunity table (monthly, one time), Giving Amount and Number of Meals Donated
Motivation: Identify Recurring donors, acknowledge them and express gratitude towards their commitment to the organization.

Problem: Do the giving amounts increases or decreases vary based on donor level?
Motivation: If the donations are decreasing for a specific level, the reason behind it can be analyzed. If the donations decrease for all the levels, the reason could be a bad economy during that year. But a drop in the donations for a certain level might indicate a problem that is specific to that level which can be worked upon. These insights can help the NGO to increase their donations.

3) Approaches

Data Cleaning: The Approach

The following steps were followed before starting the data cleaning:
a) Pre-Data Cleaning: The following questions were addressed before cleaning the data

- What are the columns that are required to answer the questions?
- If the columns required are available?
- Are the columns interpretable using data dictionary provided?

Empty columns are eliminated in the end.
b) Schema Preparation: The following schema was created using MySql workbench. We didn't find anything relevant to our analysis in the Contacts table so we are only using Accounts and Opportunities

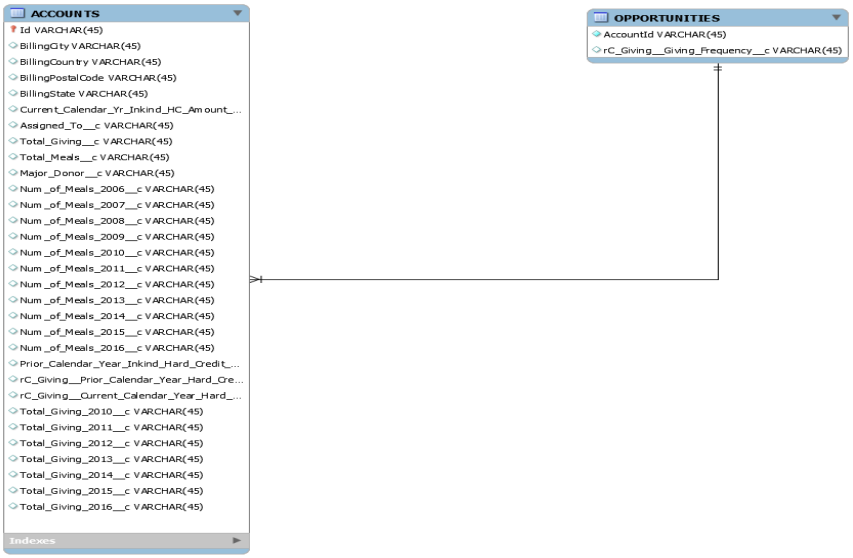


table. Id in the Accounts table acts as a Primary Key for Accounts table. AccountID in Opportunities table is the Foreign Key for this table linked to Id in Accounts table. There is a many-to-one relationship from Opportunities to Accounts.

Data Cleaning: What didn't work??

- Google GeoCode Function
- GoogleScript Function:

Data Cleaning: The Process

This was accomplished using MS Excel and the following steps were followed:

- BillingCountry, BillingCity, BillingState were cleaned based on BillingPostalCode as data provided was noisy.
- We considered billing address to identify the locations of the donors.
- We downloaded the postal code data set from the following link: http://federalgovernmentzipcodes.us/
- We used the 'Merge Tables Wizard' add-in for excel to merge two excel datasets based on the Postal code column.
- The result was accurate BillingCountry, BillingCity, BillingState.

We only considered the records having US as BillingCountry because it was more than 50% of the dataset which is a considerable amount of data for analysis. Whatever was not found was assumed to be a non-US Postal code and was deleted. The cleaned Accounts table contains 175490 rows.

Also, the Opportunities table was merged with the Accounts table afterwards and the Account IDs present in both the tables are taken for analysis from the Opportunities table.

Cleaned Datasets

(click on the above link to see the cleaned datasets)

Code Snippets

a) Code for Data Cleaning

The following commands were used for cleaning the columns initially.

```
//find the columns which have one unique value
unique_accounts <- apply(accounts, 2, function(x)length(unique(x)))
//deleting columns with one unique value
deleted_accounts = subset(accounts, select = -c(AccountSource,.....,X2015_Benefit_Level__c))
//selecting the columns that are required
accounts_final = subset(deleted_accounts, select = c(Id,…, Total_Giving_2016__c))
```

b) Code for Data Analysis

The following commands were used for addressing the question **"Who are their recurring Donors?"**

```
uni_count7=0
for(i in 1:dim(accounts)[1]){
  count=0
  if (accounts[i,26]>0){
    count=count+1 }
  if(accounts[i,27]>0){
    count=count+1 }
  if(accounts[i,28]>0){
    count=count+1 }
  if(accounts[i,29]>0){
    count=count+1 }
  if(accounts[i,30]>0){
    count=count+1 }
  if(accounts[i,31]>0){
    count=count+1 }
  if(accounts[i,32]>0){
          count=count+1 }

        if(count==7)
          uni_count7=uni_count7+1
      }
```

The following commands were used for addressing the question **"Do the giving amounts increases or decreases vary based on donor level"**

```
//Level-wise sum calculation per year
l1_2016<-sqldf("select sum(Num_of_Meals_2016__c) from accountsv3 where Num_of_Meals_2016__c >= 1 and Num_of_Meals_2016__c <= 1000")
.
.
.
l6_2016<-sqldf("select sum(Num_of_Meals_2016__c) from accountsv3 where Num_of_Meals_2016__c >= 1000001")
//Creating a matrix containing number of meals donated for different years, one matrix per level
meals_level6_yrvscount = matrix(data = NA, nrow = 11, ncol = 2)
meals_level6_yrvscount[, 1]<- 2006:2016
l6_c2<-c(l6_2006[1,1], l6_2007[1,1], l6_2008[1,1], l6_2009[1,1], l6_2010[1,1], l6_2011[1,1], l6_2012[1,1], l6_2013[1,1], l6_2014[1,1], l6_2015[1,1], l6_2016[1,1])
meals_level6_yrvscount[, 2]<- l6_c2
meals_level6_yrvscount
meals_level1_yrvscount[is.na(meals_level1_yrvscount)] <- 0
```

4) Tools & Analytics

- MS Excel
  - Preliminary Data Analysis
  - Data Cleaning
  - Data Analysis
- R
  - Data Cleaning
  - Data Analysis
- Tableau
  - Data visualization
- MySQL Workbench
  - Schema Preparation

5) Results

Who are the Recurring Donors?
(click on the above link to see the list of recurring donors)

Data Analysis and Visualization

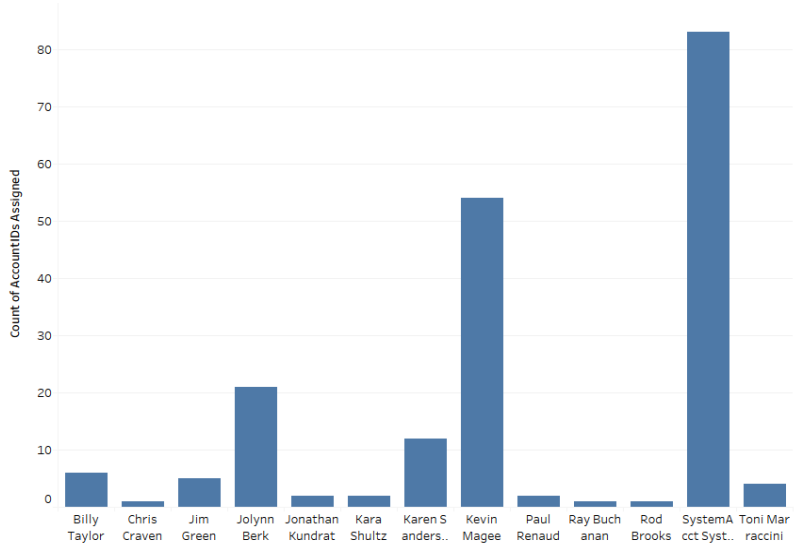
Location-wise Recurring Donors based on Frequency (monthly, one time) in Opportunity Table

- We merged the Opportunity and Accounts table and the rows which were found in both the tables were used to do the location-wise recurring donors analysis.

Insight:

- Most of the recurring donors are located on the East Coast of the USA.
- Advertising on the west coast can be improved to get more donations from that region.

Analysis of Recurring Donor(Monthly) Accounts Assignees



- This analysis is done based on the Assigned_to column in the Accounts table.

Insight:

- This graph shows SystemAcct System has the maximum number of recurring donors followed by Kevin Magee.
- These people can be appreciated with quarterly bonus so that they are motivated.
- They can be asked to take sessions about how to convert a normal donor to a recurring donor so that other employees can perform better too.

Location-wise Major Donor Analysis

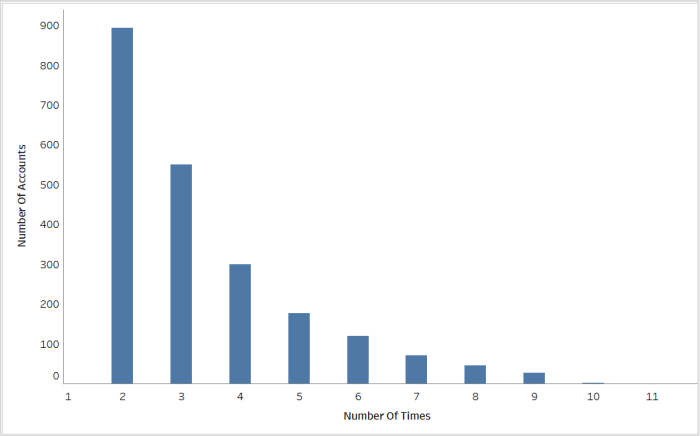
- This analysis was done based on the Accounts table (Major_donor__c column).

Insight:

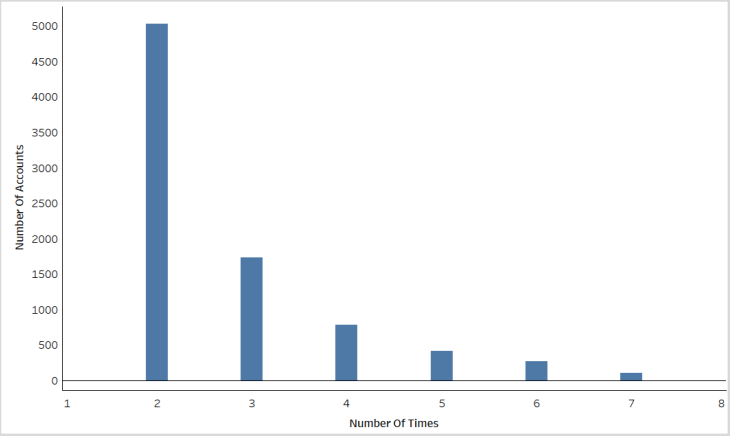
- Most of the major donors are located on the East Coast of the USA.

Analysis of Meals and Giving Amount donated based on Number of Accounts vs the Frequency of donation

Meals Donated (2006-2016)



Giving Amount (2010-2016)

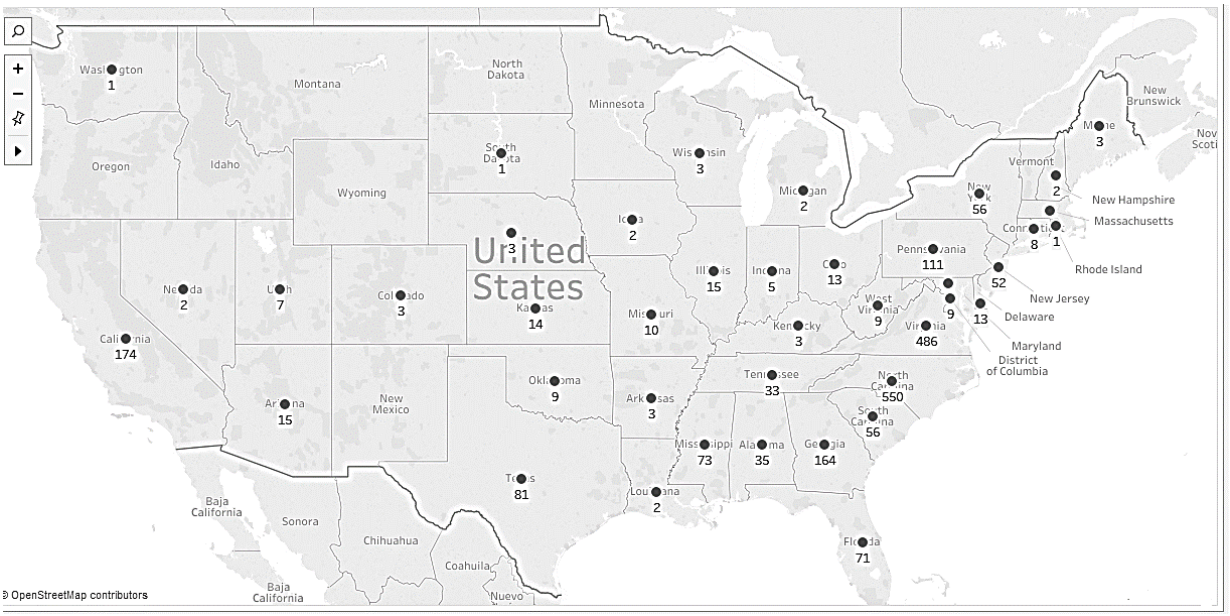


- This analysis is done to analyze who are the loyal donors and how many times they have donated.

Insight:

- There are a few donors who have been there with the organization for almost all the years.

Location-wise Recurring Donors Analysis (Meals Donated)

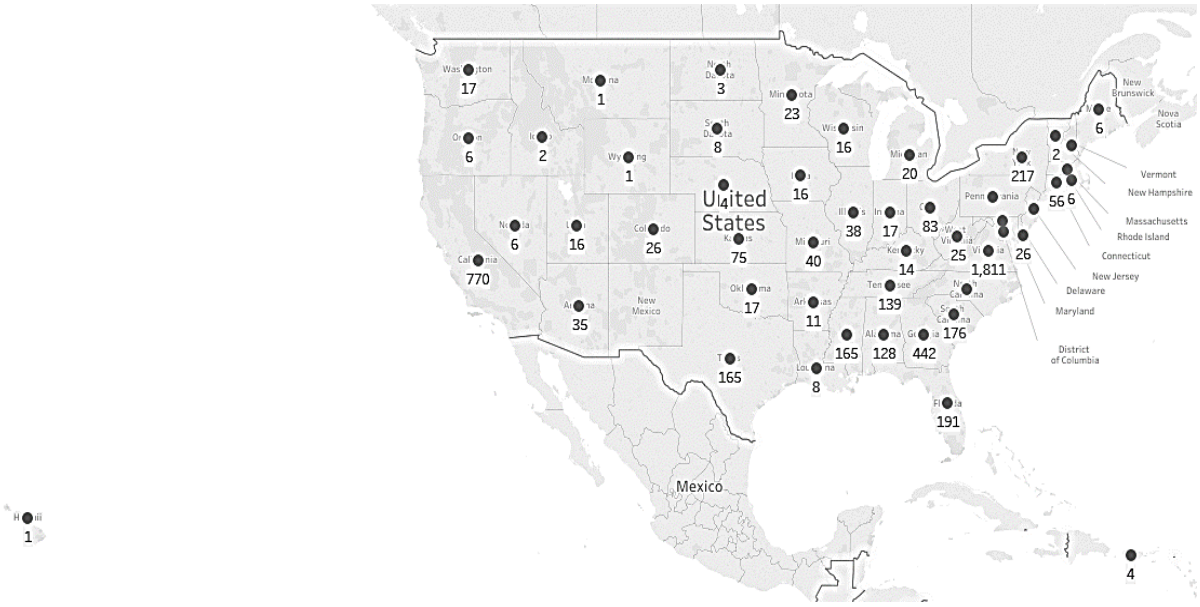


- This recurring donor analysis is conducted based on how many times they have donated meals from year 2006 to 2016.
- If an account has donated for two or more times, it is considered as a **recurring donor account**.

Insights:

- Most of the recurring donor accounts are in east coast.
- On the west coast, California has a huge number of recurring donors.

Location-wise Recurring Donors Analysis (Giving Amount)



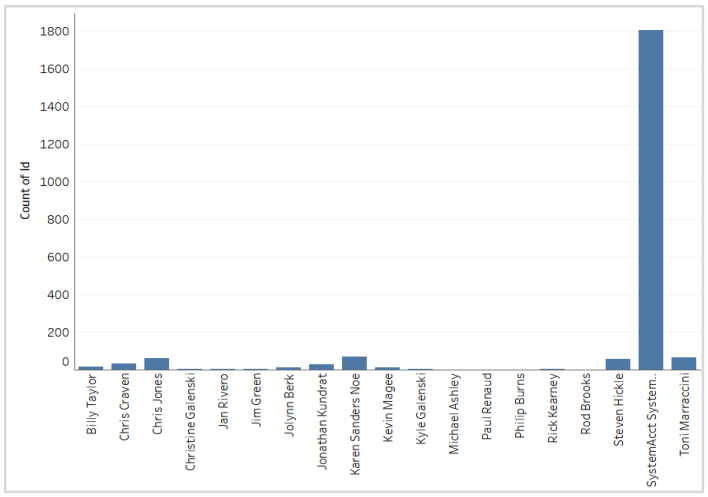
- This recurring donor analysis is conducted based on how many times they have donated from year 2010 to 2016.
- If an account has donated for two or more times, it is considered as **recurring donor account**.

Insights:

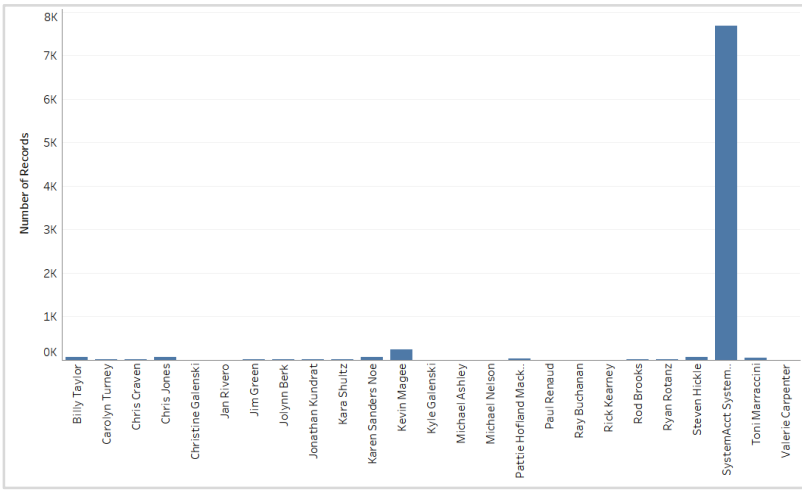
- As it can be seen from the image above, most of the donors are in east coast
- Virginia has the maximum number of recurring donors.

Whom the Recurring Donors Accounts are Assigned To

Meals Donated



Giving Amount

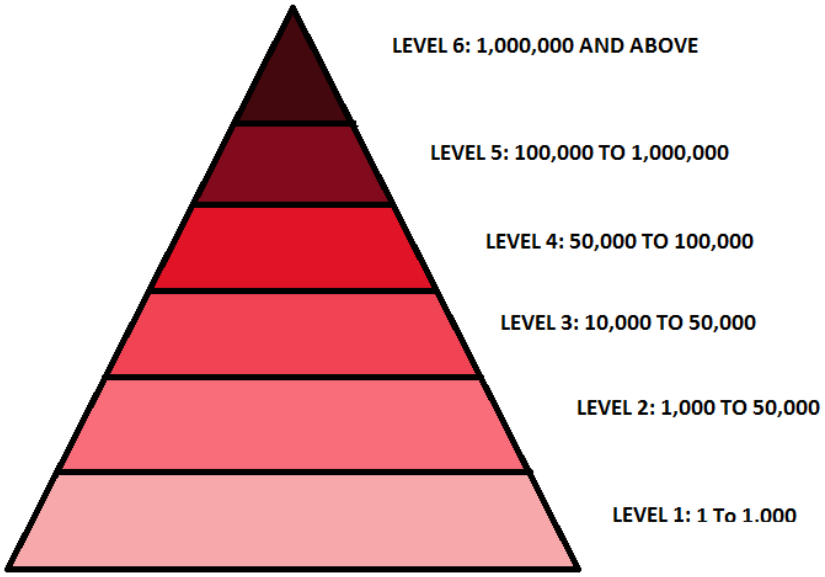


- This analysis is done based on the Assigned_to column in the Accounts table.

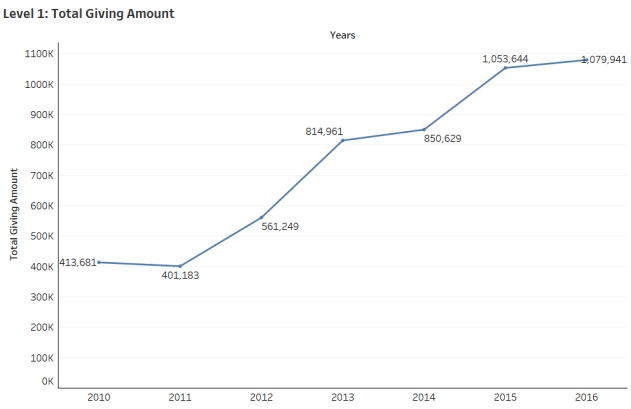
Insights:

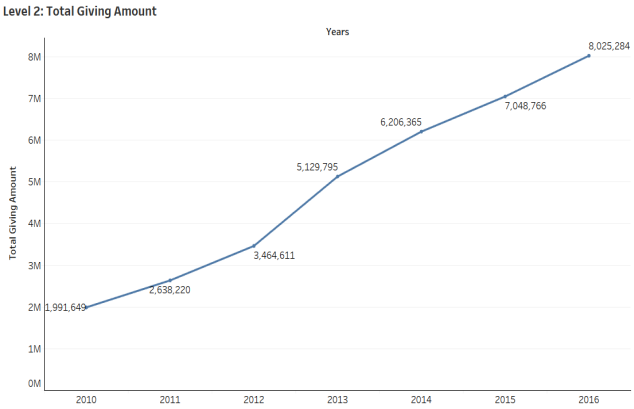
- The graphs show that SystemAcct System has the maximum number of recurring donor accounts in both the cases.

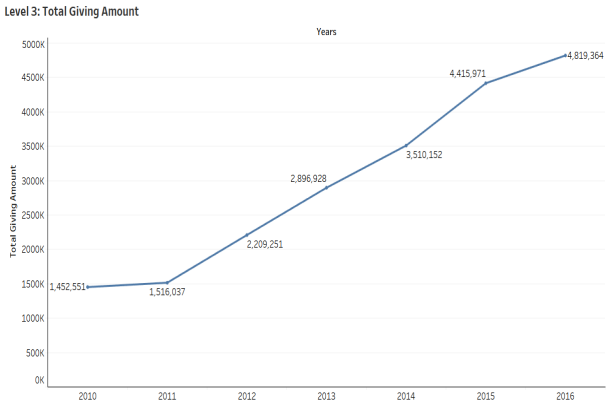
What are the levels ($ amount cutoffs)?
Since the donor levels are not specified, we have considered the following donor levels for our analysis by looking the max and min values of the various columns:

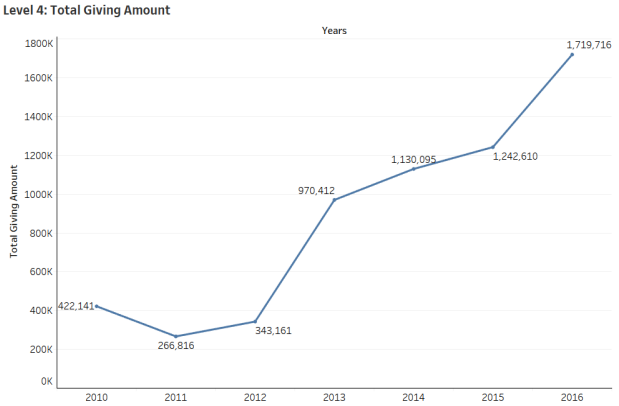


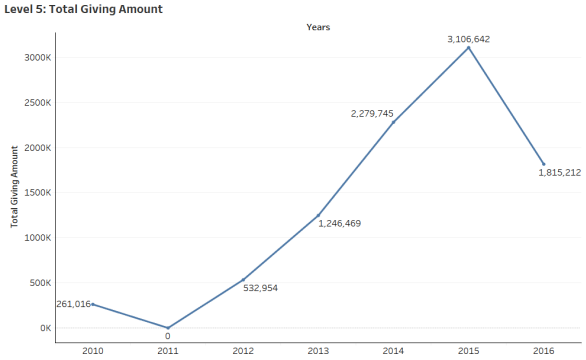
Do the giving Amounts increases or decreases vary based on donor level? The following plots show how the total **Giving Amount** is changing (increasing or decreasing) based on each Donor Level.

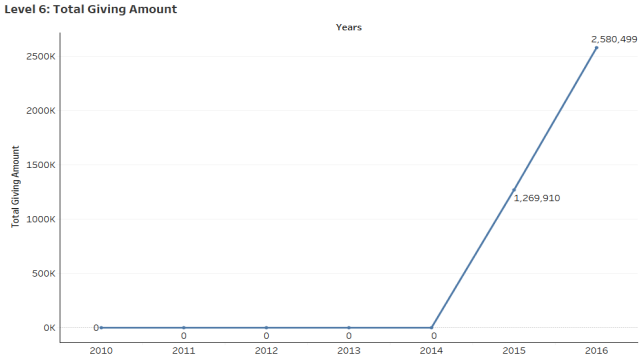








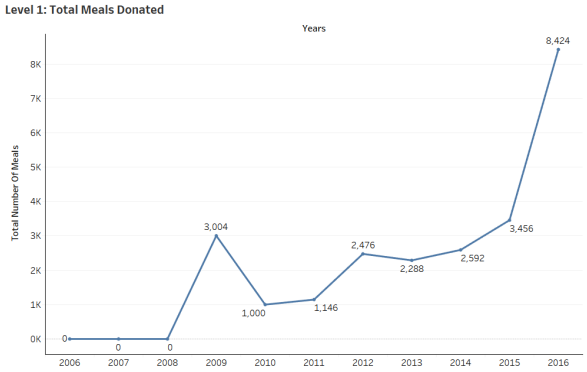


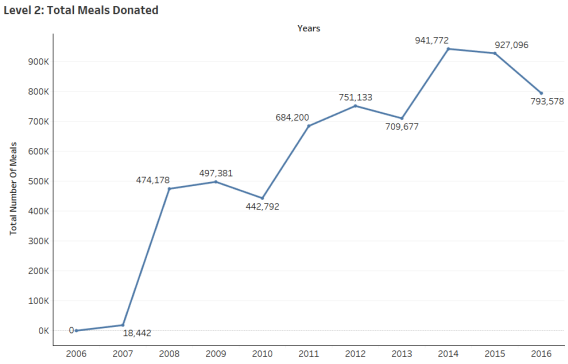


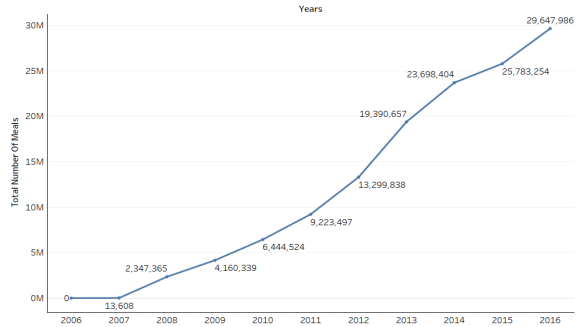
Insights:

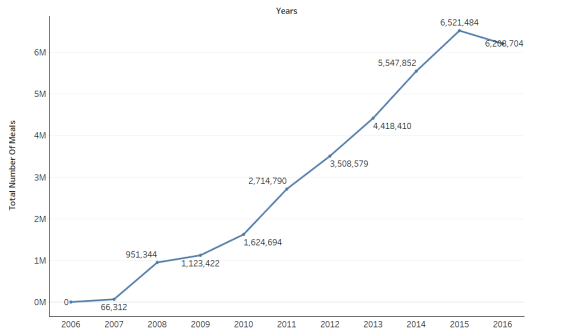
- The overall trend of donation is increasing for most of the donor levels.
- There were no donations in Level 6 till 2014, but after that the donations have increased from year 2015 to 2016.
- The level 5 donations have taken a dip from 2015 to 2016. Since the donation amount is very high in this level, if one or two donors decide not to donate, total donation will decrease by significant amount.
- Rise in the Level6 donations show that the NGO has got some new big donors in year 2015.

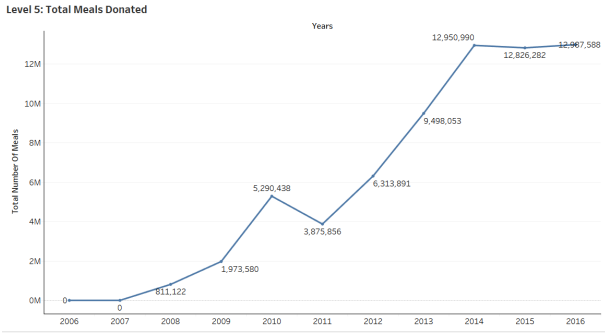
We have also analysed how the total number of donations in terms of **Meals Donated** are changing in different Donor Levels. Here are the plots for each level:

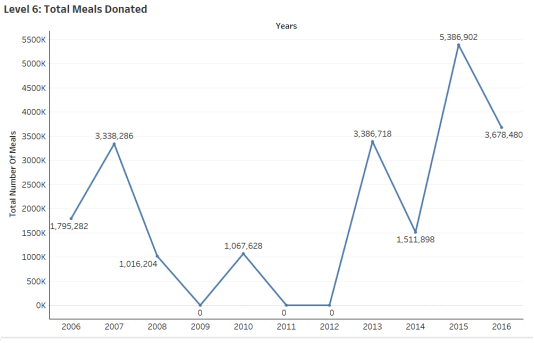












Insights:

- In three out of six levels, the meal donations have decreased from year 2015 to 2016.
- In Level1, there is huge increase in donations from year 2015 to 2016 which is due to a large increase in the number of small donors.
- In Level3 and Level4 donations, the Meals Donated show an increasing trend overall.
- In Level5 donations, the Meals Donated shows almost an increasing trend.
- In Level6, the total donations are increasing and deceasing in alternate years. The reason behind this could be some accounts donating large amounts in alternate years.

6) References

- http://federalgovernmentzipcodes.us/
- *https://productforums.google.com/forum/ - !topic/docs/doQed0PgcsA;context-place=forum/docs*