MKT6337. 002

Marketing Predictive Analytics with SAS

**GROUP 07**

**PROJECT REPORT ON**

**Appeal and Donation Analysis**

**SUBMITTED BY GROUP 12**

**APARAJIT**

**JASMINE SETHI**

**KSHITIJ YADAV**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| 1. | PROJECT MOTIVATION | 3 |
| 2. | PROJECT DESCRIPTION | 3 |
| 3. | DATA DESCRIPTION | 4 |
| 4. | DATA EXPLORATION AND STATISTICS | 6 |
| 5. | DEMOGRAPHIC FACTORS ANALYSIS | 12 |
| 6. | APPEAL ANALYSIS | 21 |
| 7. | BEHAVIORIAL FACTOR ANALYSIS | 23 |

**Project Motivation**

NGOs (non-profits) play a significant role in the American society. They are in a unique position that allows them to understand the needs of the society and reciprocate in ways, that often have more impact than the corporate governance of other organizations.

With the advancement of data science, nonprofits are encouraged to apply analytics in their respective processes, with an expectation of higher means of process improvement and effectiveness.

On this note, this project seeks to understand the factors influencing the behavior of the donors and thereafter predict the pattern (propensity to donate and amount donated) in the period 2004 to 2006. The analysis is based on the investigation of 3 sets of factors – zip code level demographics (from the 2000 Census), appeals sent to donors and past donation behavior (i.e., donation behavior in the period 2000-2004).

**Project Description**

The Project has been divided in three parts which are as follows:

Demographical factors analysis

This involves knowing the demographic factors which correlate with past behavior and predicting the future behavior. Also, knowing the customer’s lifetime value and recommendation of demographics for acquiring new donors.

Appeals analysis

The main aim was to know about the response to appeals and to find ways to optimize the targeting of appeals.  Also, to understanding the limitation on using the current data to optimize the targeting of appeals and describing a field of study to overcome these limitations.

Behavioral factors analysis

This involved knowing about prediction of future donations depending on the past amounts. To measure donors recency and frequency and determine does it have a significant impact on future donations behavior.

**Data Description**

The data has been gathered from a downloaded from one of the leading US nonprofit organization. They agreed to release detailed translation and history of the donors it acquired in 2000. The transaction and contact history extend through November 30, 2006. The data is comprehensive and is spread across two files.

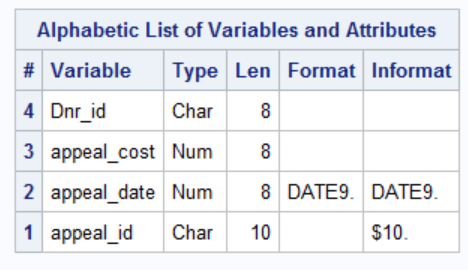
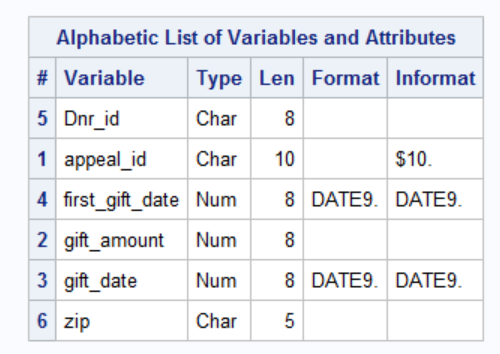
First file consists of Appeals data which has 1,730,598 observations and comprises of all the appeal details sent to the donors. The organization contacts donors by sending appeals to solicit donations (by mail, phone, email etc.). Each month an appeal is sent to a group of donors. There can be multiple appeals sent out in some months – each appeal being part of a different fund-raising campaign. All appeals that are part of the same campaign in each month have the same appeal id. The variables in the dataset are:

Donor\_id: unique donor id.

Appeal\_id: unique appeal id.

Appeal\_date: date that appeal was sent.

Appeal\_cost: cost of that appeal.

The other file is donations dataset. It consists of 192840 observations. This dataset provides all donations made by each donor. The variables in the dataset are.

Donor\_id: unique donor id.

Gift\_date: date of donation.

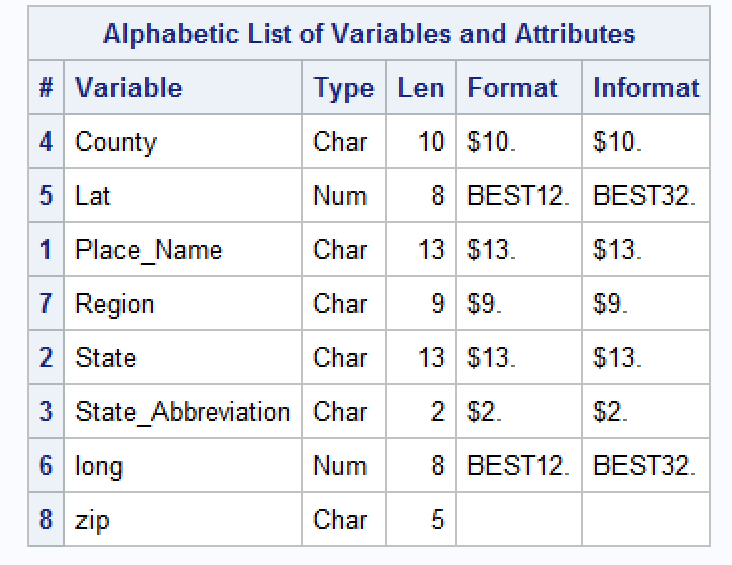
Gift\_amount: donation amount.

Appeal\_id: appeal that donor responded to.

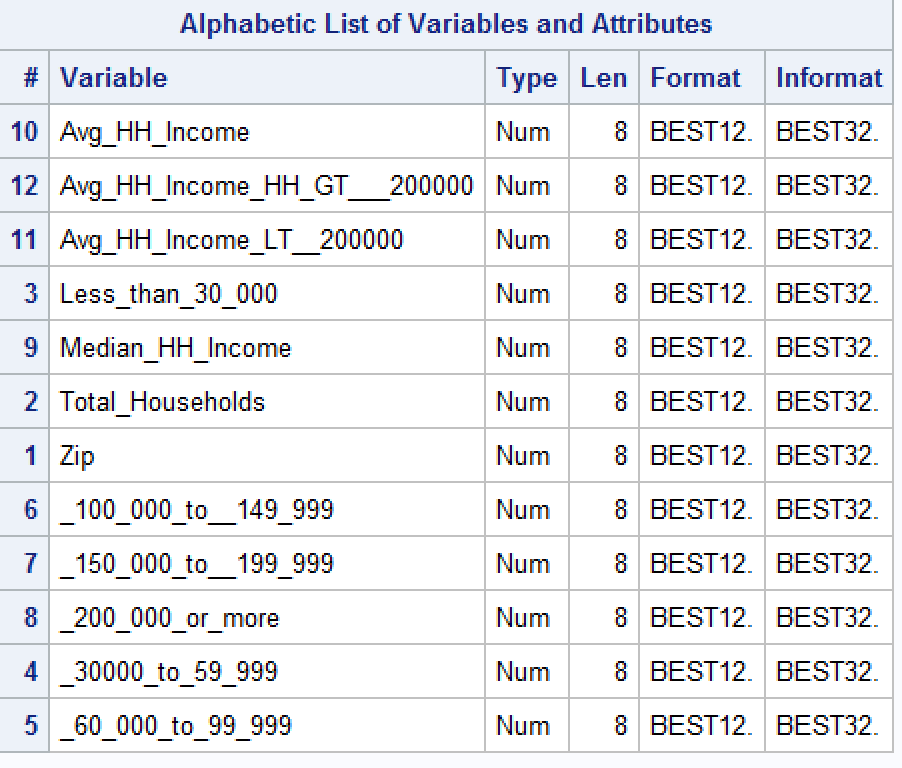
Zipcode: zip code of donor.

First\_gift\_date: The date of the first donation by this donor

For exploratory analysis, we have used data with the city, county and state information for each zip code in the US. The dataset has been stored in a sheet named US Postal Codes. Below are the variables present in Us\_Postal\_codes:

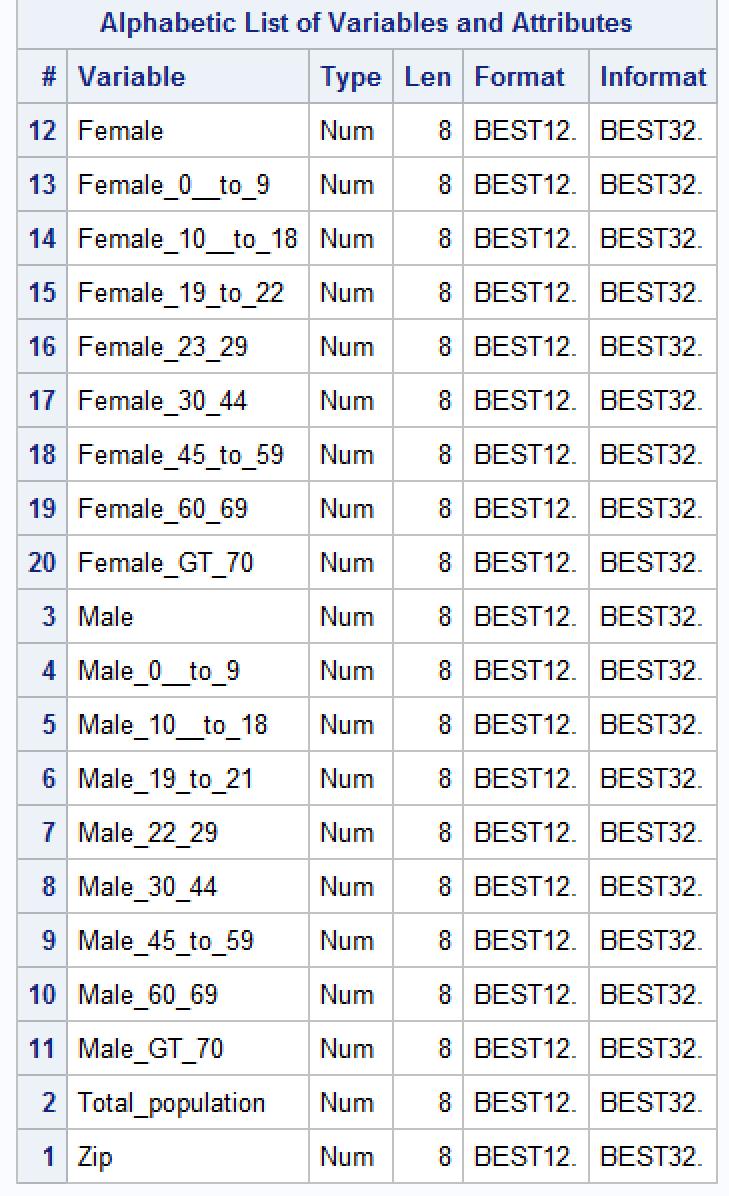
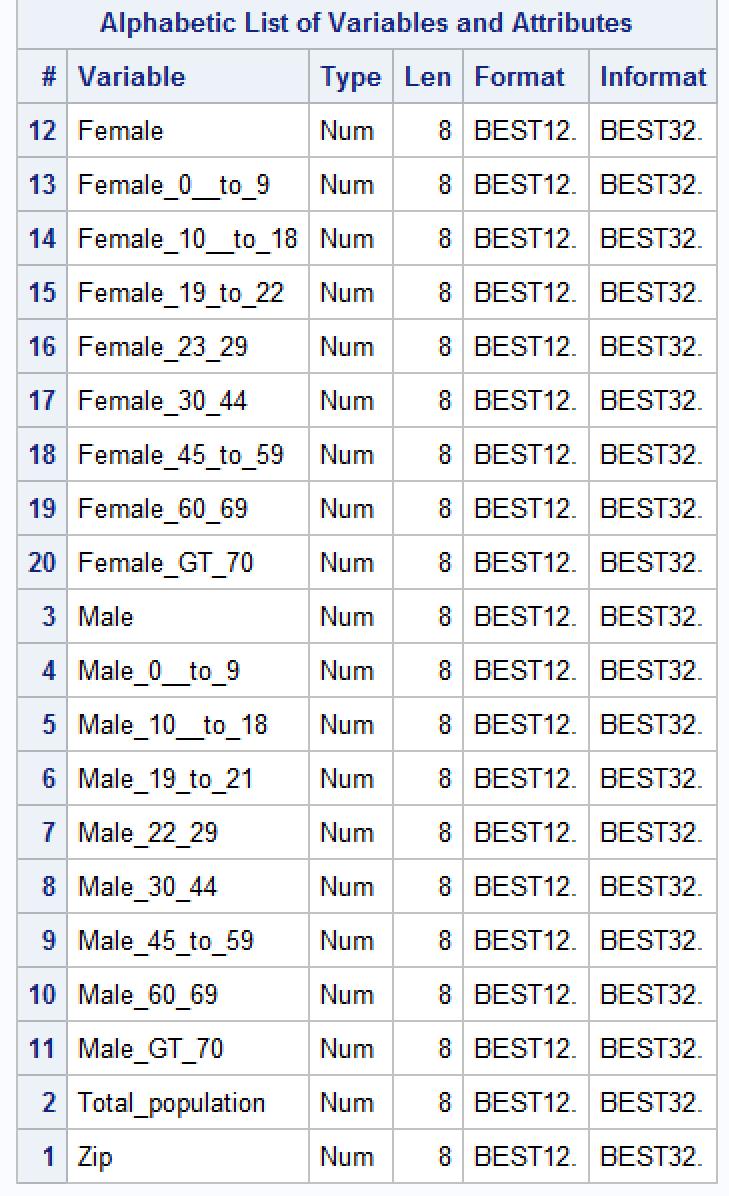


Further data was merged with Household Income Distribution and Household by Age and Family to analyze donations in comparison to income, age and family size. It is stored in Household\_Income\_Distribution and Household\_by\_Age\_and\_Family respectively.



Household\_Income\_Distribution Household\_by\_Age\_and\_Family

finally, it was analyzed on basis of population Age and Gender The dataset is stored in Population\_by\_Age\_and\_Gender.



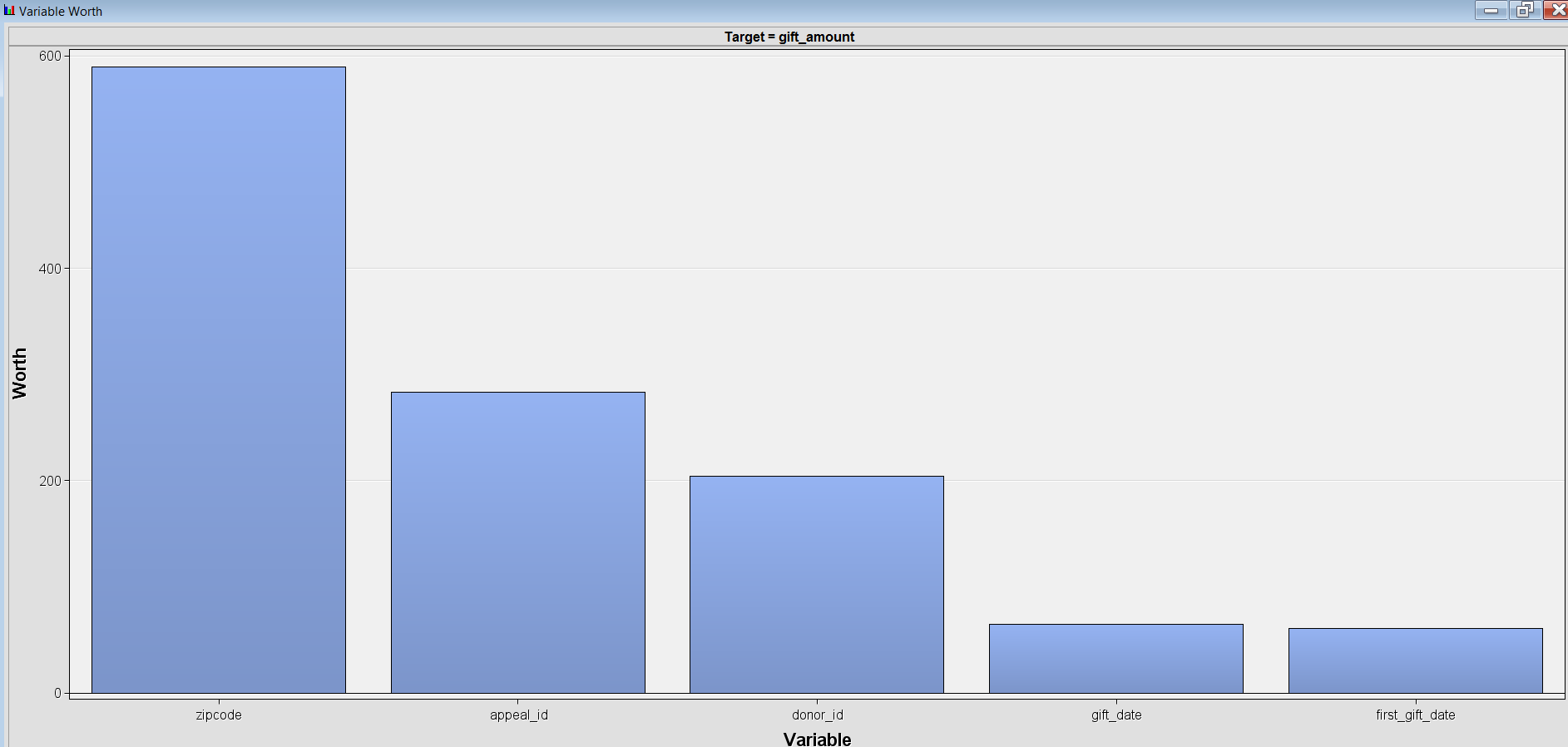
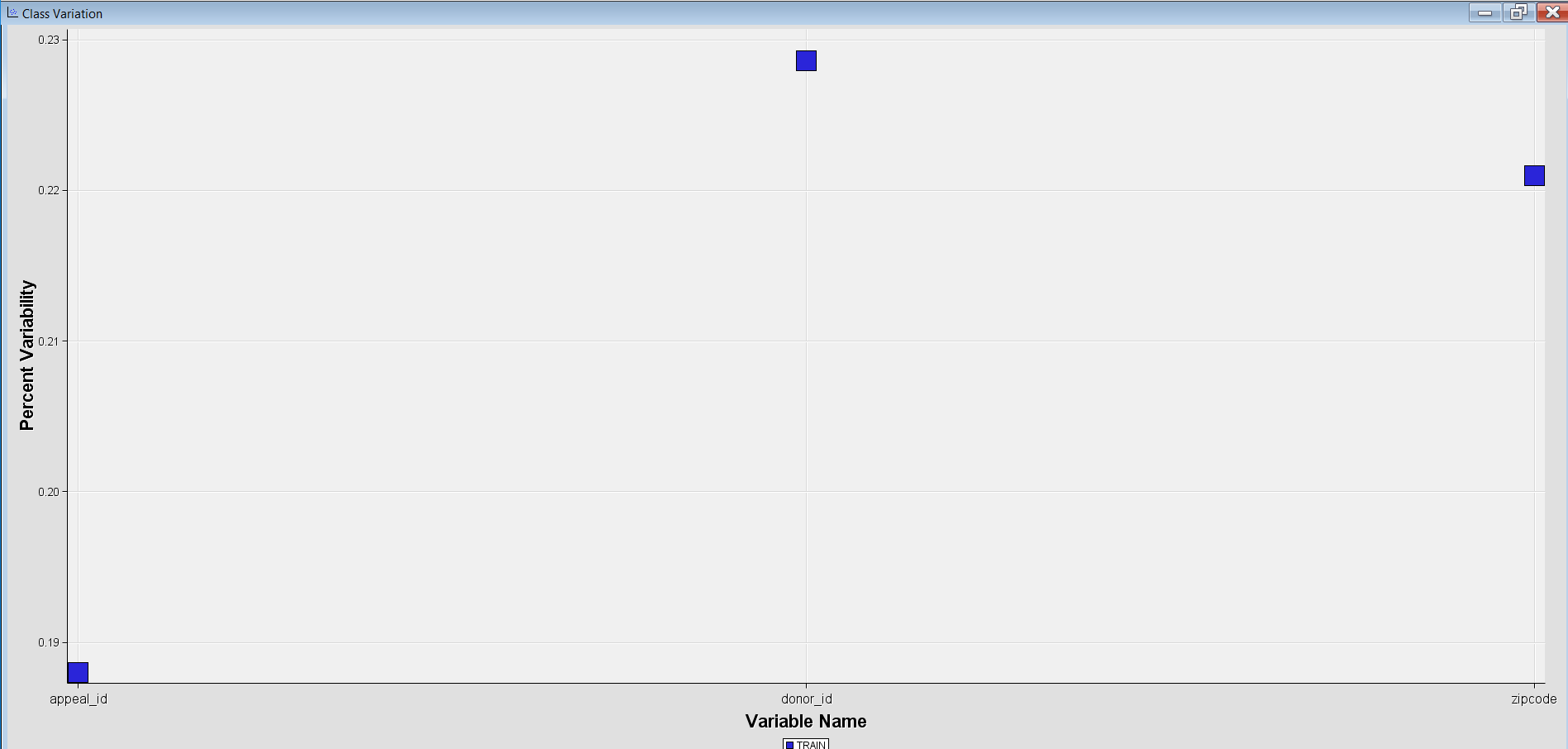
**Data Exploration and Descriptive Statistics**

Appeal and Donation Statistics

Data exploration is the first step in any data analysis and it typically involves summarizing the main characteristics of a dataset.  So, we did an initial exploration of the data to view its most relevant features.

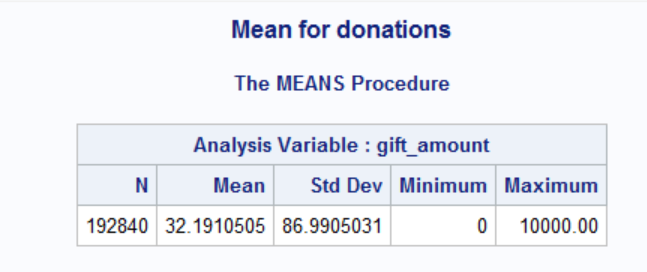
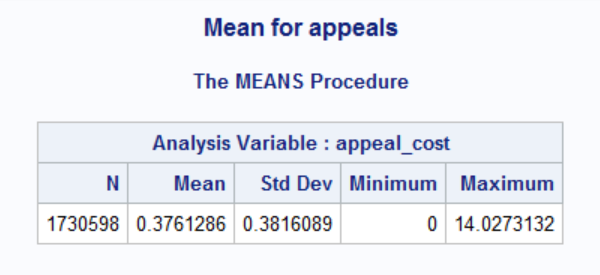
The first thing we did was knowing the variable worth as it was important to analyze which predictors are statistically significant. As can be seen below Zip code and appeal id turned out to be most significant factors.

Also, it is important to know class variability and by how much the data varies. As seen, appeals are less variable. Donor id and zip codes vary allot. It can be inferred that almost same appeals were sent to different donors across different zip.

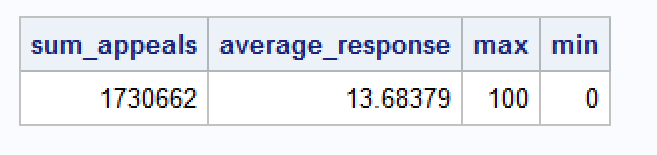
 

Variable Worth Class Variability

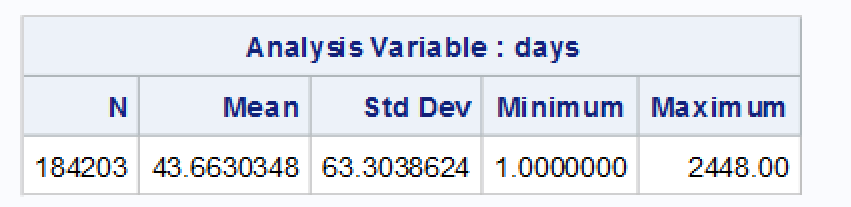
Then we tried finding out the average mean for the appeals cost and the average mean of gift amount. We found out that the average donations made were around $ 32 and maximum was $1000.



Appeal and Donor response times were calculated. The average response time for appeals has been 13.683 as few appeals did not get any response and some got late response. On an average people donated money in 43 days and maximum being 2448.



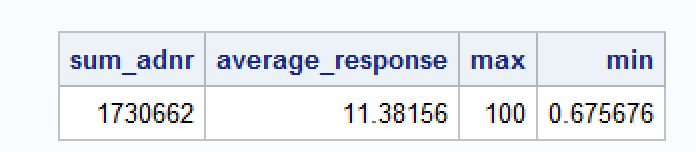
Appeal Response Rate



Appeal Response in Days

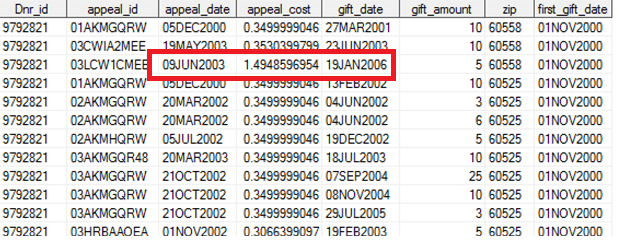
Also, there were around **1535157** that were not responded at all.

The average response rate of donors turned out to be 11.38 as many donated within a day and some never donated.

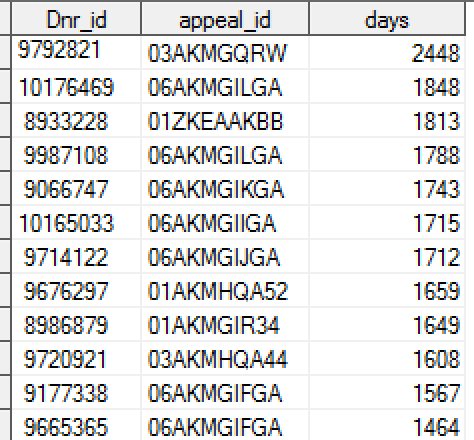


Donation Response Rate

One donor id made donation after 2448 days. The appeal was sent in 2003 and the amount was donated in 2006. Below is the snapshot for one id.



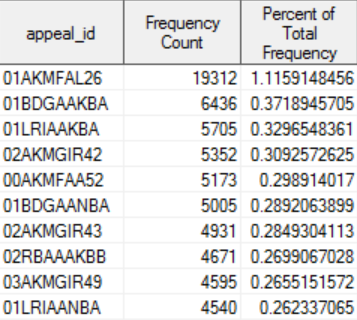
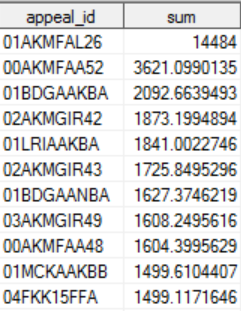
Donors with maximum difference between the appeal date and donated date are:



Delay in Donation

It was important to know which appeals were most frequent and it was interesting to know one particular appeal was sent 19312 times and total appeal cost associated to it was $14484. The other interesting thing we noticed was the appeal\_id ‘000AKMFAA52’ had appeal cost around $3621 which was more than various other much more frequent appeals.

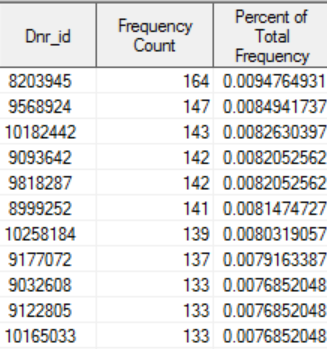
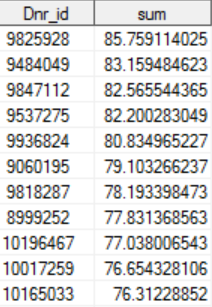
Below are the snapshots for the most frequent appeals and the appeal id with maximum appeal cost.

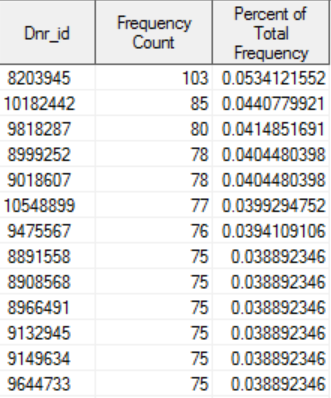
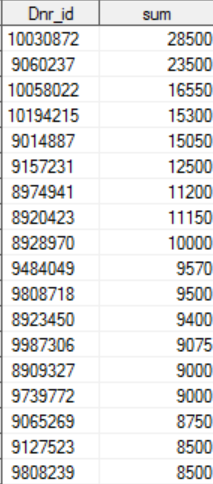
Appeal Frequency Appeal Cost

Also, all the donors were not sent all appeals. Below is the screenshot of the frequency of appeals sent to each donor and the sum of appeal cost for each donor. Though maximum number appeals were sent to donor id 8203945 but the total appeal cost was $73.395 and he donated only 103 times out of 160 times. The total amount donated by this id was only $3325.

While donor id 9825928 was sent fewer appeals but appeals with higher appeal cost. The appeals cost sent to him summed up to $85.759 and the donation amount received was $4800. The maximum amount of donations was made by donor id 100030872 around $28500 while the total appeals sent to this id was way too less.

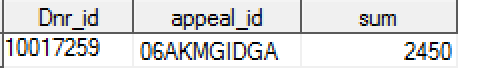
 

Appeals Frequency Appeal cost sum for donors

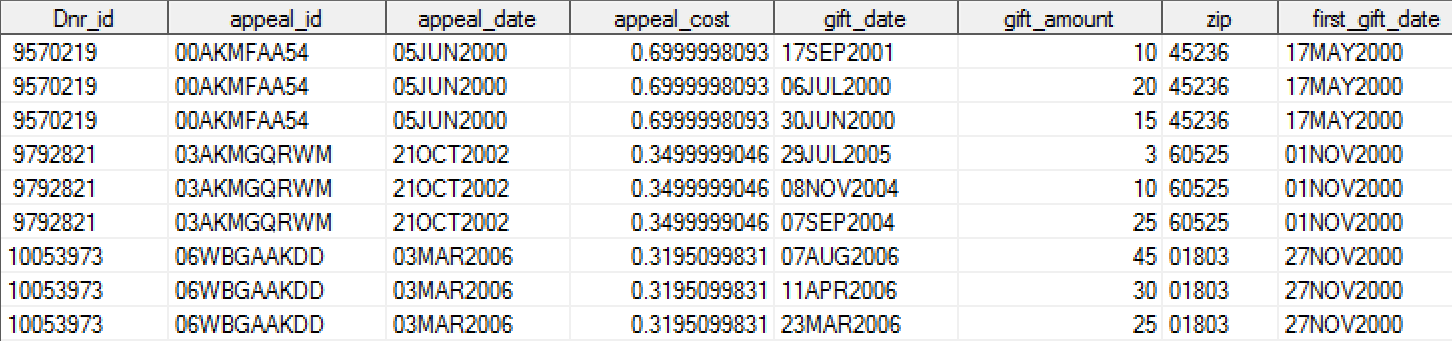
Donors donation frequency Donated sum by each donor

Also, donor id 10017259 paid more than 20 times for a particular appeal but total sum donated was around $2450.

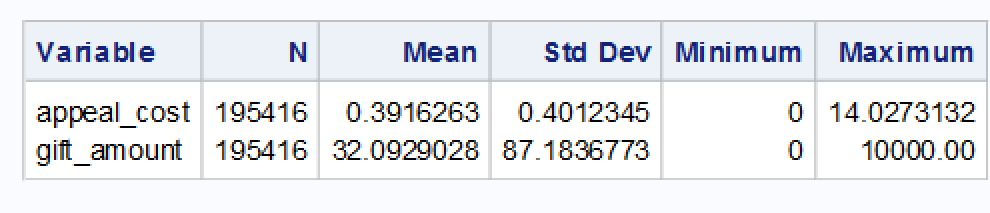


Amount Donated by 10017259

Also, three ids donated thrice for same appeal id.



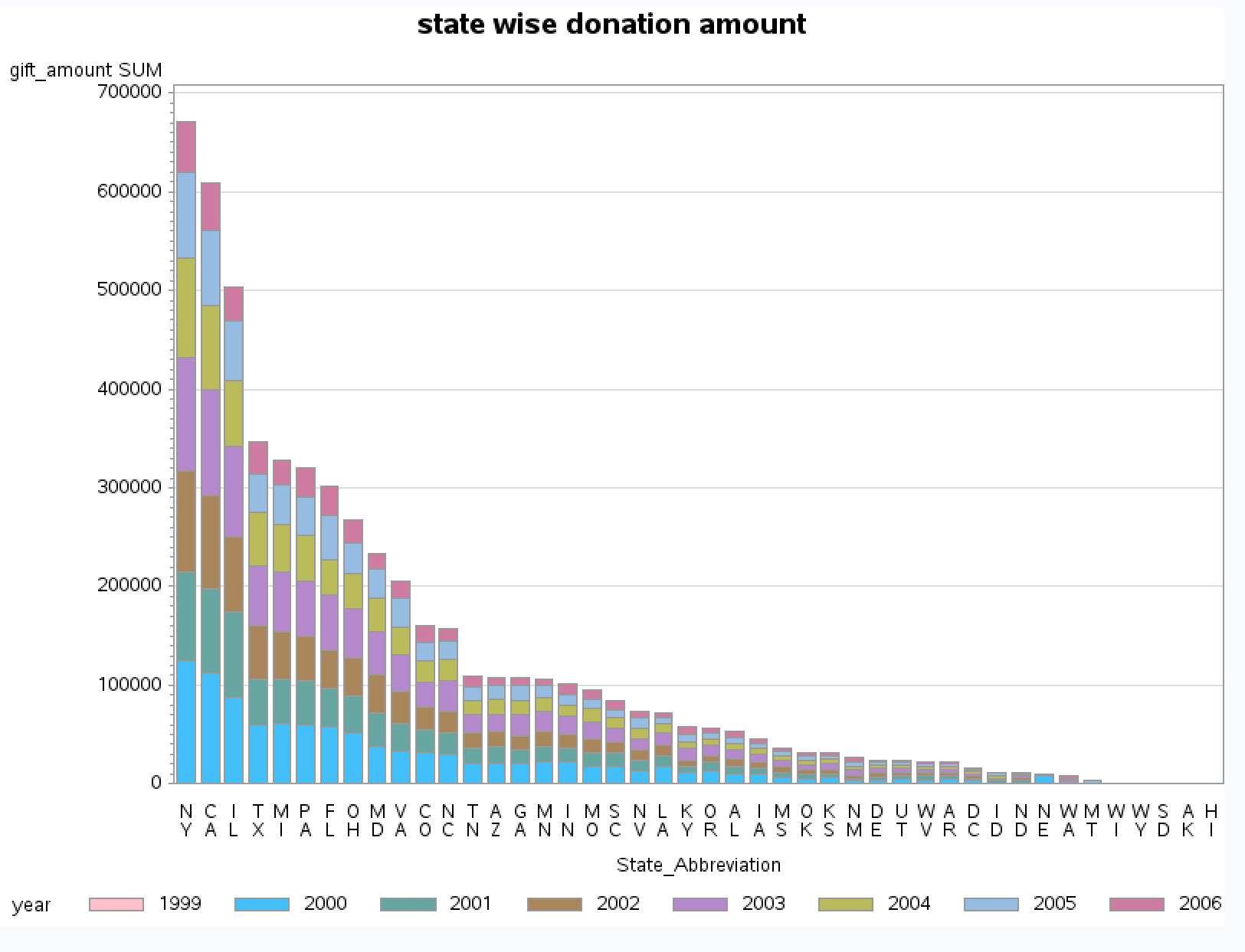
The descriptive stats of appeal cost and the gift amount are as follows:



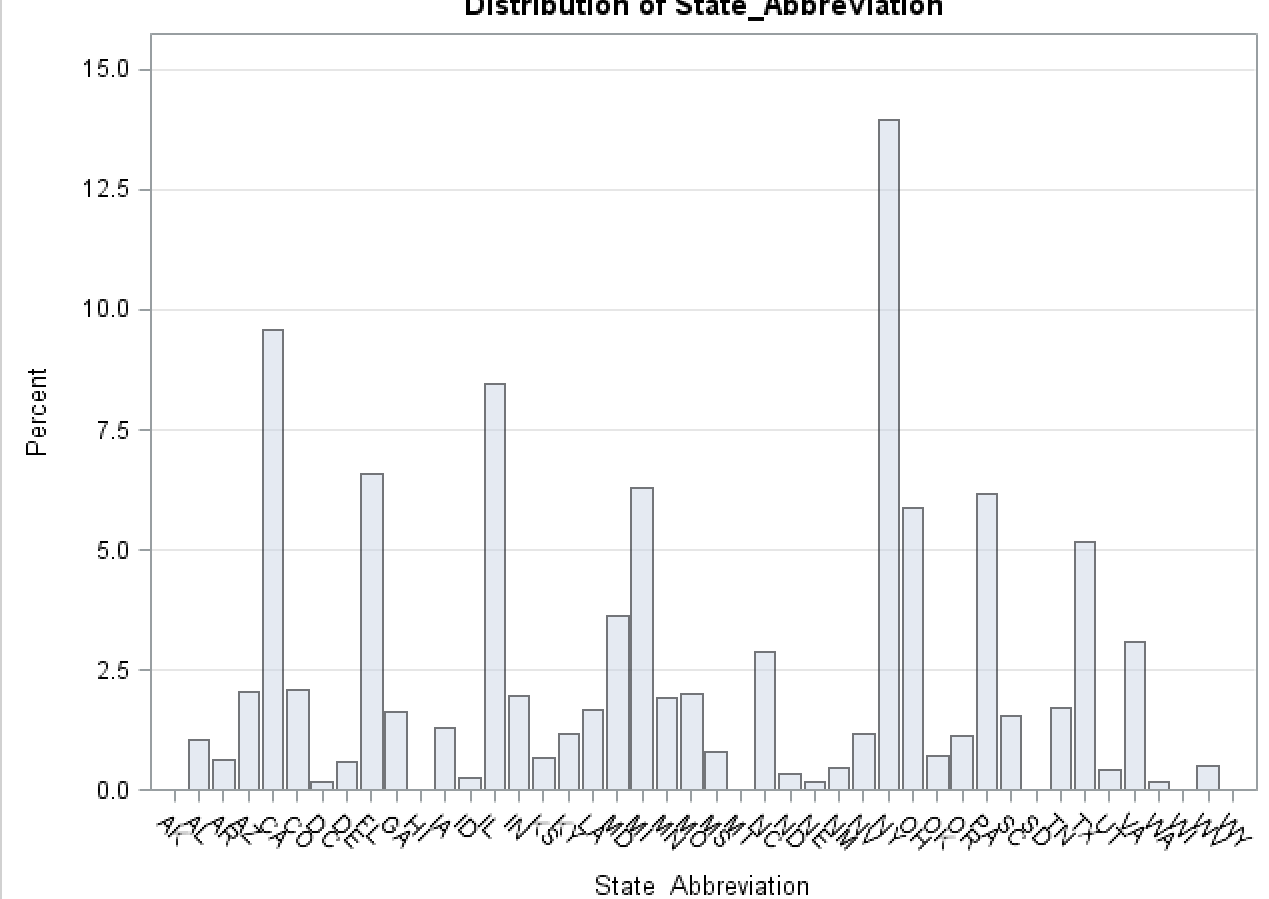
Appeal Cost and gift amount

Demographics Statistics

To find out which states donated maximum amount, the donations and appeals sheet was merged with the Us\_Postal Code sheet. It was noted that New York had maximum amount of donations as well as frequency of donations in total in all years. Followed by California and Illinois. Graph below shows donations made by each state.

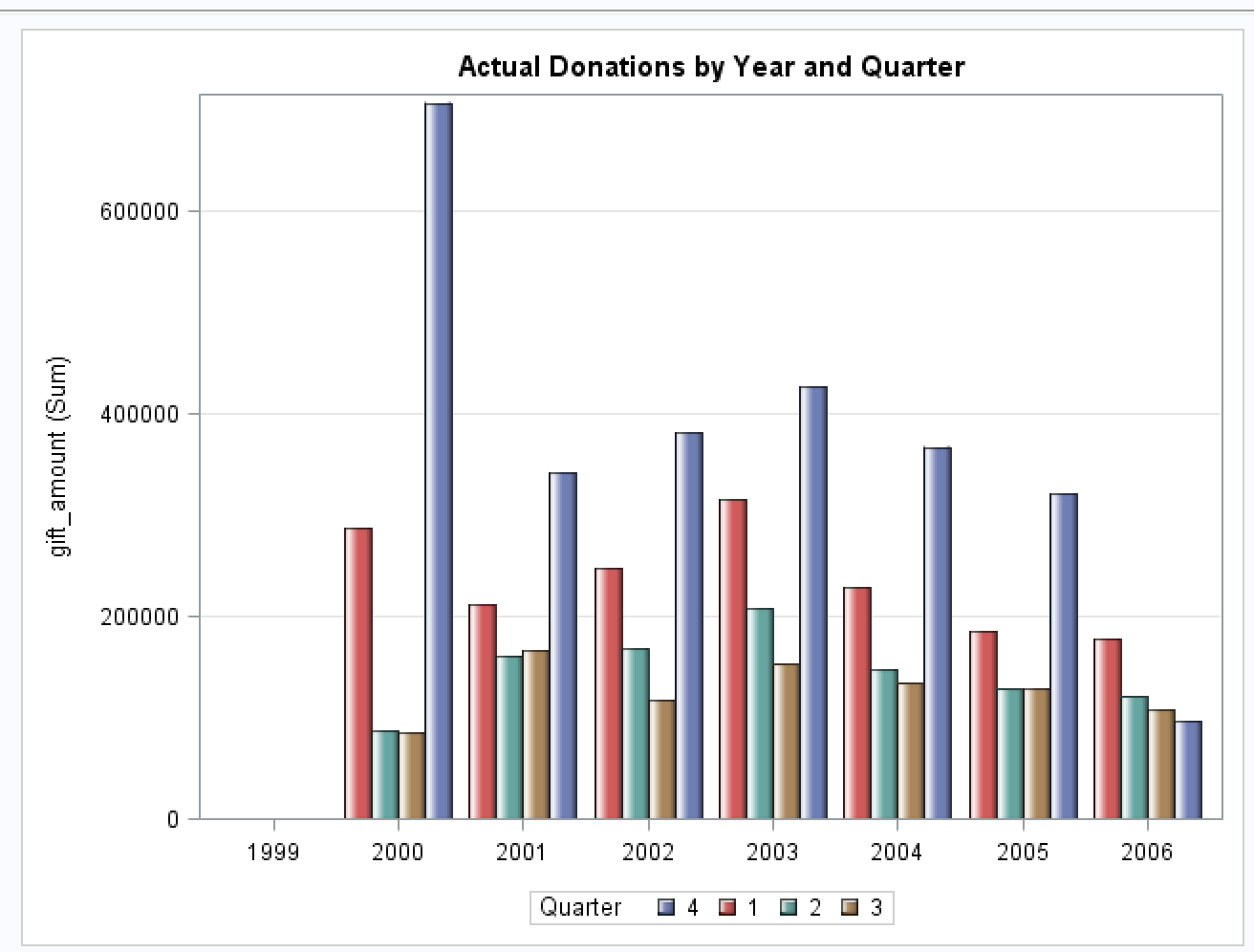
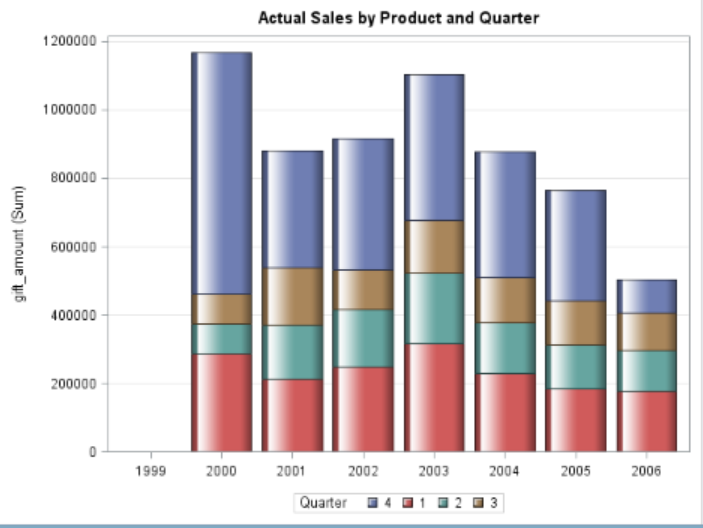


Donations by each state in each year



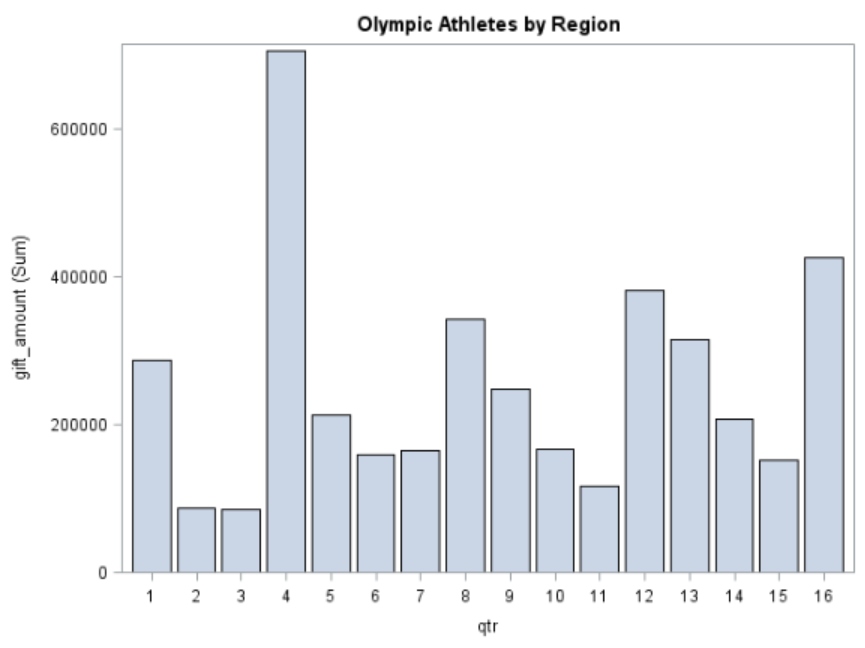
Frequency of donations by each state

Maximum donations were made in the last quarter of each year. Also, the maximum donations happened in the starting year which is 2000. We can infer that donations were made at the end of each year and people donated maximum initially and later their donation trend kept on decreasing.

Quarterly donations in each year Yearly donations

To be noted the maximum donations happened in the fourth quarter and after that in 16 quarter i.e. dec 2003.



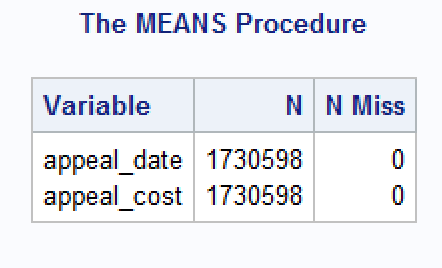
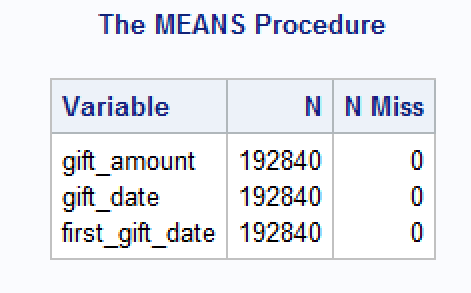
**Demographic Factors Analysis**

**Objective** The objective of this exercise is to see if demographical factors (from US 2000 census data) have any impact on the donation behavior. Mainly our focus was on noticing demographical trends and their correlation with customer lifetime value and the ability to accurately predict future behavior basis the existing data. Basis this we also aimed to identity certain regions which could be targeted for customer acquisition.

As mentioned in the instructions the data was split in two parts – one dataset containing records prior to 2004 and one containing data post 2004 (Based on gift date). This part of the analysis mainly deals with data prior to donation dates of 1st Jan 2004. Before we moved to analyzing the data, some preprocessing was required which is mentioned in the below section

**Data pre- processing**

**Missing values:** The datasets required minimal pre-processing. Both datasets were checked for any missing values but the data was found to be reasonably robust with no missing values.

Appeals Donations

**Importing the US Census data:** Five additional datasets were provided which contained information about the US 2000 annual census. These files contained Zip code level information about the population, income groups, family and non-family households, age, Urban and rural households etc. The mentioned information was at the zip code level.

Proc Import was used to import all the five datasets into SAS.

**Data Aggregation:**  In order to perform further analysis both appeals and donations datasets were aggregated basis the unique donor\_id and appeal\_id present in the data. Proc SQL was used to execute this step.

As a next step we merged the donations data with appeals using appeal\_id and donor\_id .Merge command within a data step was used for this. This created an aggregate master data set which could be used in further analysis.

This dataset was sorted by zip codes using Proc Sort and merged with the master dataset.

***Modification to the US\_postal\_codes dataset* –** *From the US census website(*[www.census.gov](http://www.census.gov)*) the US census regions were downloaded for all US states and this metric was added to the dataset.*



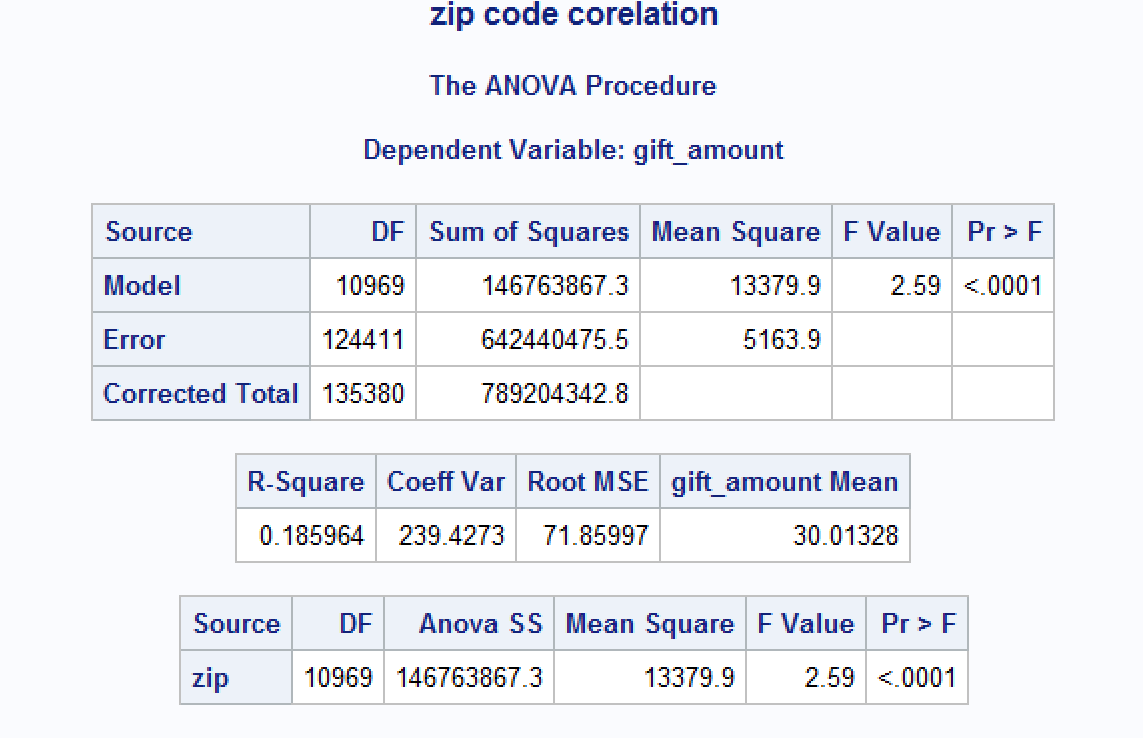
**Data Analysis:**

**ANOVA-** To check if the mean for gift\_amount varied with change in Zip an anova was run which had the following hypothesis-

Null Hypothesis: The means for Zip and gift\_amount are equal

Alternate hypothesis: The means for Zip and gift\_amount are not equal

After running proc ANOVA the output received was as follows –

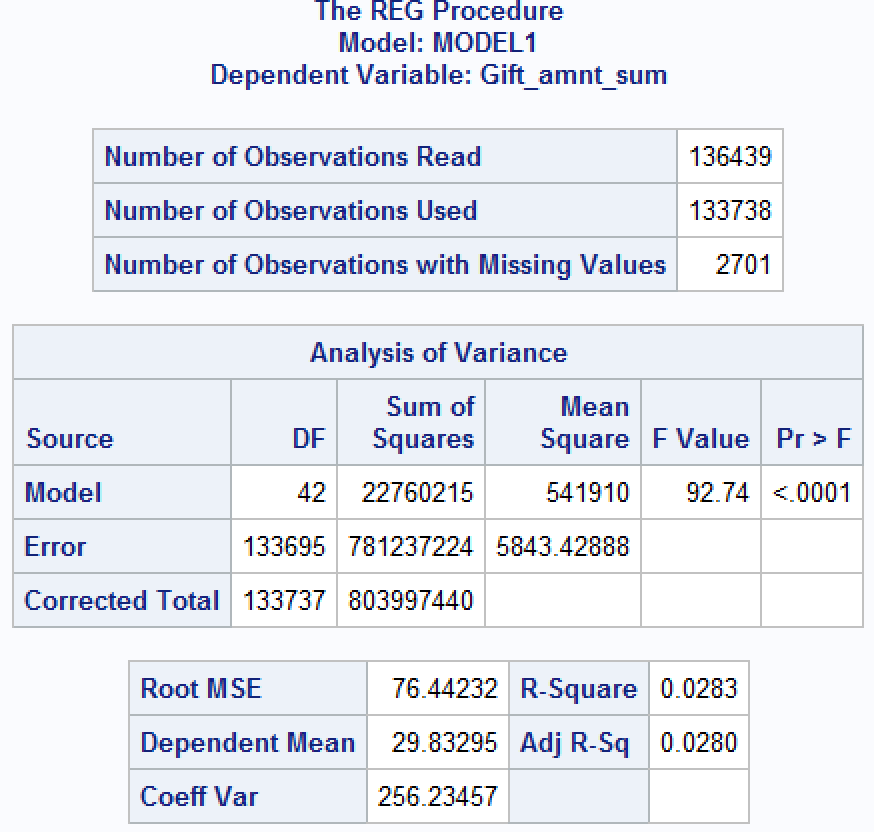
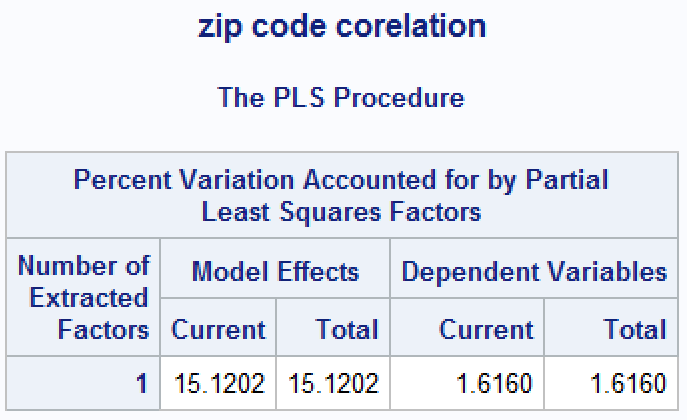


Zip Code Correlation

This helped us conclude that there is variation in donation amounts with change in zip codes.

Next we moved to running multiple regression models to further establish the relationship

**REGRESSION ANALYSIS:** as a first step, a linear regression using all numerical and categorical variables was run without any modifications to the data. Proc Reg was used with multiple combinations like forward, backward , stepwise and MAXR. Even after excluding factors with Variance Inflation Factor greater than 20 a robust model that could explain the variance could not be found. We even tried running a partial least squares model using Proc PLS but the results were still less than satisfactory.

**Linear Model**  **PLS**

Since no results could be obtained through these models, as a next step we tried to modify the data and some new variables were created and some existing variables were dropped –

*working\_male=Male\_19\_to\_21 + Male\_22\_29 + Male\_30\_44 + Male\_45\_to\_59;*

*retd\_male=Male\_60\_69 + Male\_GT\_70;*

*working\_female=Female\_19\_to\_22 + Female\_23\_29 + Female\_30\_44 + Female\_45\_to\_59;*

*retd\_female=Female\_60\_69 + Female\_GT\_70;*

*avg\_per\_household=Total\_population/Total\_Households;*

We classified males and females aged 18 to 60 as working and above 60 as retired. Also average number of people per household was calculated by dividing the total population by the number of households. The donation amount was also binned into the following categories-

*if Gift\_amount le* ***10*** *then cat1="lt\_10";*

*if Gift\_amount >* ***10*** *and Gift\_amount<=****50*** *then cat1="10\_50";*

*if Gift\_amount >* ***50*** *and Gift\_amount<=****100*** *then cat1="50\_100";*

*if Gift\_amount ge* ***100*** *then cat1="mt\_100";*

Cat1 had a value of Lt\_10 if the donation amount upto $10, 10\_50 if it was between $10 to $50 , 50\_100 if it was between $50 and $100 and mt\_100 if it was more than 100.

Also as a last step, basis the census region modification in the dataset 4 dummy variables were created for each of the regions South, East, Midwest and Northeast.

*if region='South' then South=****1****;else South=****0****;*

*if region='West' then West=****1****;else West=****0****;*

*if region='Northeast' then Northeast=****1****;else Northeast=****0****;*

*if region='Midwest' then Midwest=****1****;else Midwest=****0****;*

Using feature selection , some variables that we deemed surplus were dropped from the dataset-

*drop Avg\_HH\_Income Family\_HH\_Other\_witht\_own\_CHILD\_ Family\_households\_HH\_15\_to\_34*

*Family\_households\_HH\_35\_to\_54 Family\_households\_HH\_55\_to\_74 Family\_households\_HH\_75\_\_or\_olde*

*Family\_households\_\_\_Other\_with\_o Lat Less\_than\_30\_000 Nonfamily\_HH\_1\_person\_HH*

*Nonfamily\_HH\_MT\_2\_persons\_HH Nonfamily\_households\_HH\_15\_to\_34*

*Nonfamily\_households\_HH\_35\_to\_54 Nonfamily\_households\_HH\_55\_to\_74*

*Nonfamily\_households\_HH\_75\_\_or\_o long Total\_population \_100\_000\_to\_\_149\_999*

*\_150\_000\_to\_\_199\_999 \_200\_000\_or\_more \_30000\_to\_59\_999 \_60\_000\_to\_99\_999*

*Female\_0\_\_to\_9 Female\_10\_\_to\_18 Female\_19\_to\_22 Female\_23\_29 Female\_30\_44*

*Female\_45\_to\_59 Female\_60\_69 Female\_GT\_70 Male\_0\_\_to\_9 Male\_10\_\_to\_18 Male\_19\_to\_21*

*Male\_22\_29 Male\_30\_44 Male\_45\_to\_59 Male\_60\_69 Male\_GT\_70 Female Male*

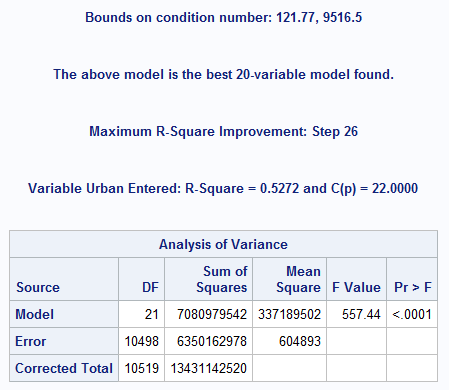
*state\_abbreviation total\_households;*

All the numerical variables were then standardized with mean equal to zero and a standard deviation of one.

**Regression Analysis:** It was not no surprise that no correlation was found in the regression analysis because the appeals and donations datasets were at donor level and the demographic factors data was aggregated.

Therefore as a next step, the donations and appeals datasets were merged and aggregated at zip code and then merged with the demographics data. A new variable was created turnaround time which was the time taken by a person to respond to an appeal. Average values across zip codes were for the turnaround time and these values were added to the dataset.

The Rsquare value obtained foe the regression model was 0.52. With this we can conclude that demographic factors are correlated with the aggregated donation behavior.



**Model estimate**

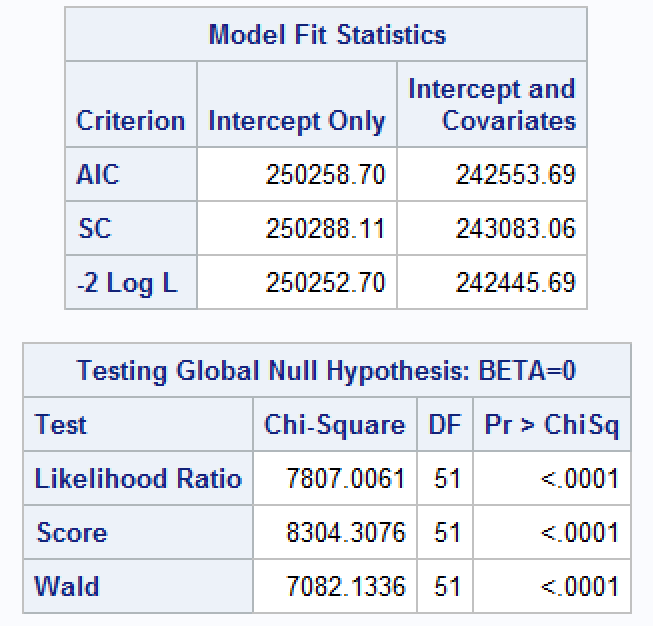


**Parameter Estimates**

**Key Insights:**

* From the model, we can see that the demographic factors are correlated with donation behavior
* Average household income does not have much of an impact on the propensity to donate, the wealthy do not necessarily donate more
* Married families with kids have a greater tendency to donate than married couples without kids
* Median household income is a significant variable with positive coefficient, so cities with higher median incomes would donate more
* Retired males tend to donate more than retired females
* Weather a particular area is urban or rural or semi urban does not influence donation behavior
* Working females donate more than working males

**Multinomial LOGIT:** After the above mentioned modifications, we now had a dependent variable (Cat1) which was categorical in nature with multiple levels. So a multinomial logit model could be run. The results obtained from the model are shown below –

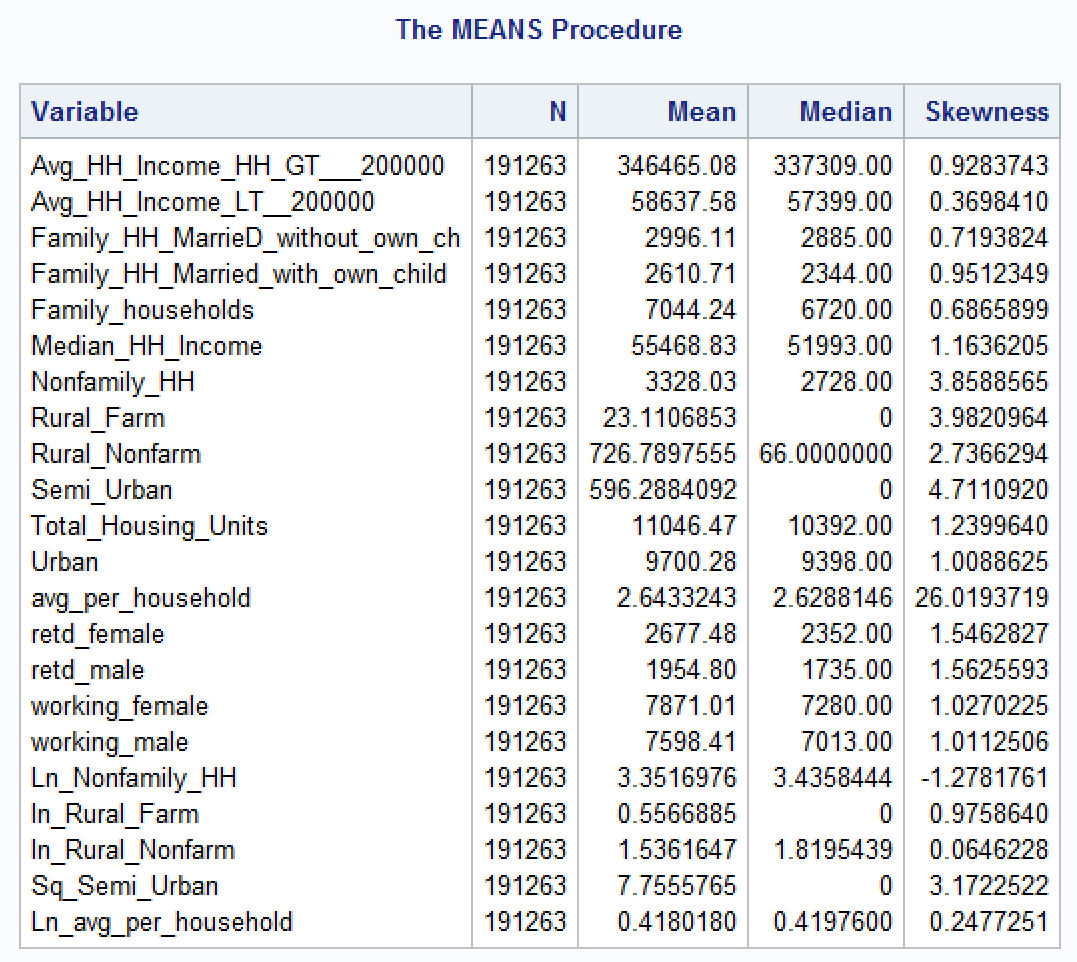


The MCfadden’s R square value calculated for this model was **0.0311** meaning these predictors are not very good interpreters of the donation amount. **Therefore , it is safe to conclude that demographic factors do not show a strong correlation with past donation behavior.**

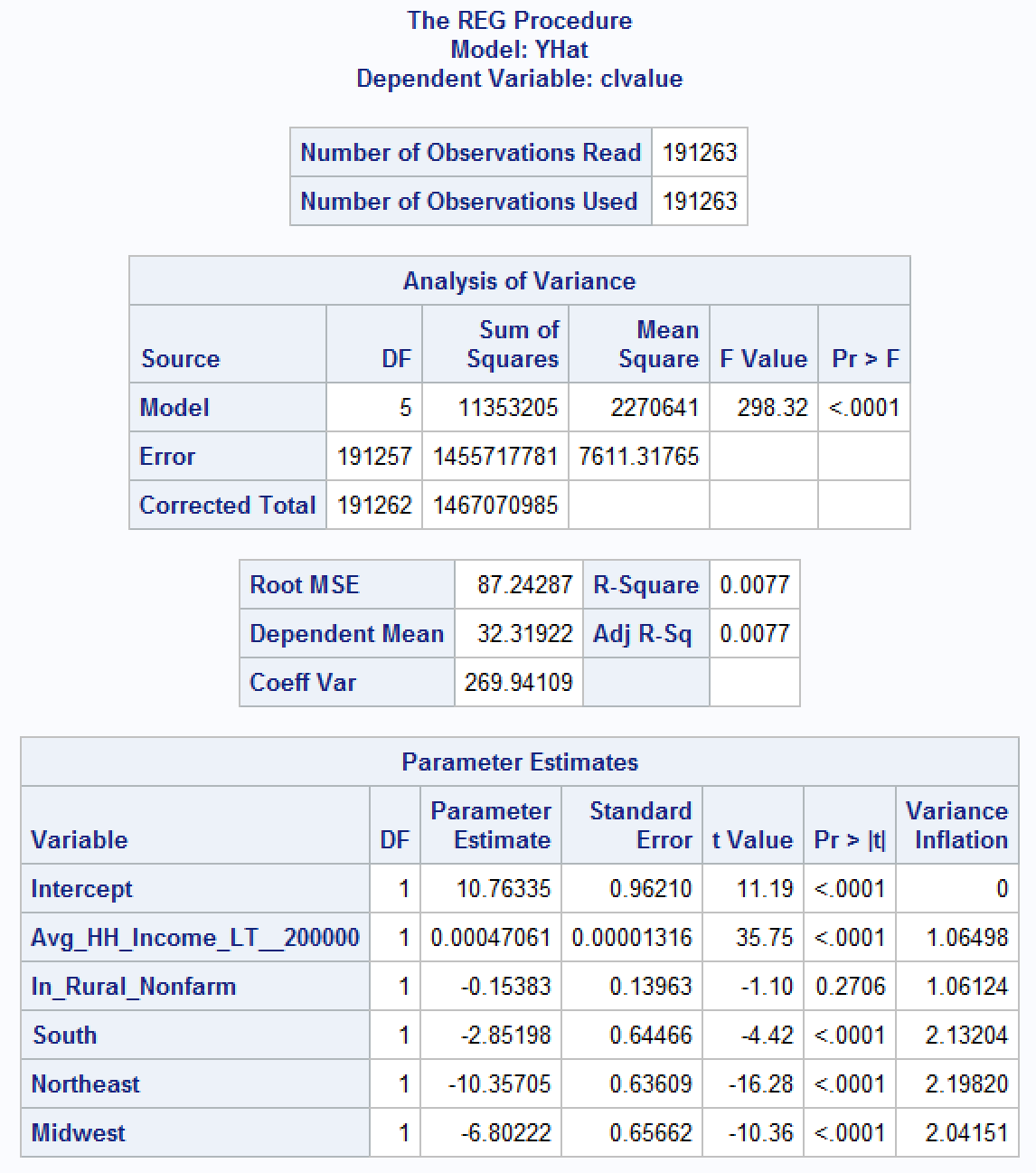
This would not be a good predictor of future donation behavior since no significant correlation was found with present donation behavior.

**Customer Lifetime Value Analysis:** Customer lifetime value was calculated by adding the appeal costs(negative values) to the donation amounts. In this section we try to find the correlation of CLV with demographic factors.

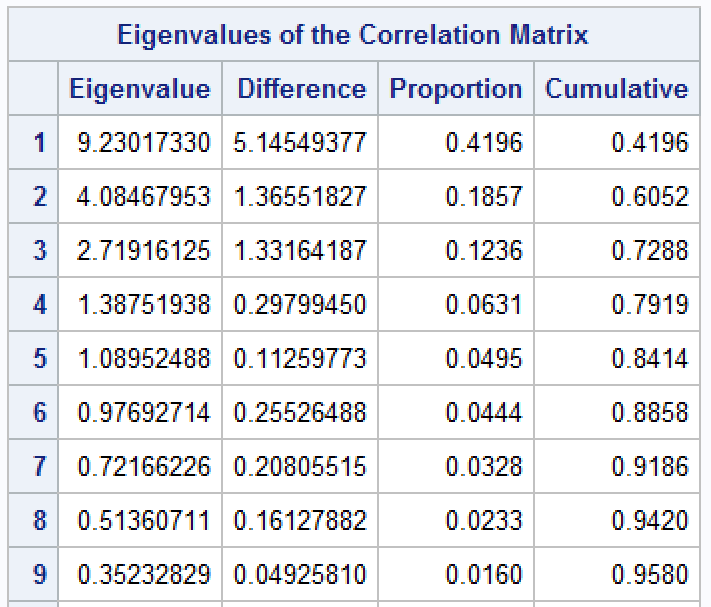
Some of the variables had significant positive skewness so they were normalized taking either log or square root values.



Post this data was again standardized using Proc Standard and then multiple regression models were run. Then selectively insignificant variables were eliminated and multiple regression models were run.

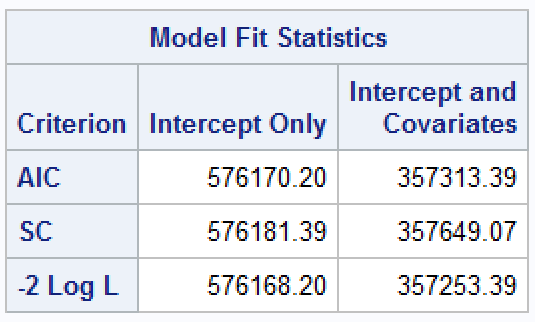
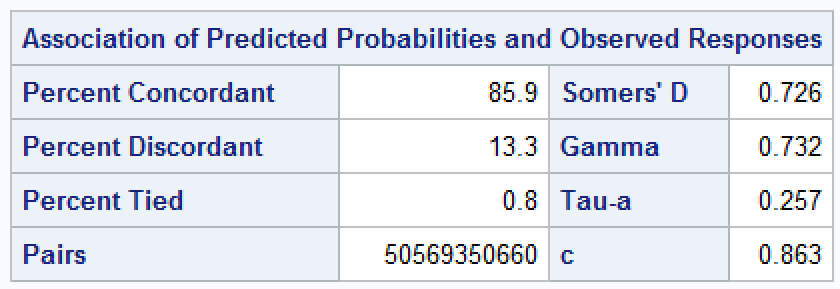


As can be seen from the above table the model is poor with very low R SQUARE value. We also tried running Principle component analysis using Proc Princomp to see if any of the principle components might help improve the model taking first nine eigen vectors.



Running this model also did not yield the desirable results so we can conclude that **customer lifetime value is not influenced by demographic factors.**

**Geographical Area Recommendations:** To run this analysis we modified the data so that a conditional logit could be run on the dataset. The dummy variables created for regions, appeal cost, appeal month and donation month were included in the model.

Model fit statistics of conditional logit

The Mcfadden R square value for this modle was 0.379.

The following insights can be drawn from the model –

1. The coefficient for appeal\_cost is -0.0946 which means basically higher appeal cost does not warrant higher donations
2. **Northeast** and **midwest** have postive coefficents so for a targeted donation campaign we can focus on states falling in these regions-



3)Appeals sent in June and december have the highest positve coefficients

4)Donations for the month of January have a negative coefficient whereas for December the coefficient is positive

In summary we would recommend the states mentioned in the above table for acquiring new donors as these have a positive impact on our regression model.

**Appeal Analysis**

**Objective** The purpose of this analysis was to check if appeals influence donations and are donations dependent on cost of appeals or timing of the appeal.

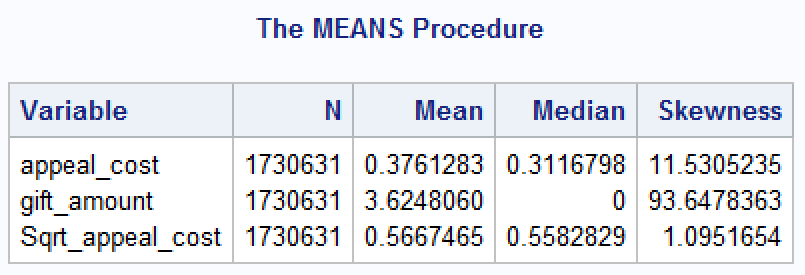
**Data Pre-Processing:** The appeals dataset is much larger than the donations dataset, implying that a lot of appeals went unanswered. We sorted both the datasets on their donor\_ids and appeal\_ids and the merged the datasets keeping all records from the appeals dataset. A new variable donate was created which had the value 1 if the person donated and 0 if the appeal went unanswered. Gift month and appeal month were also extracted.

Next this dataset was again merged with the donations dataset after sorting by donor\_id. This was done to map the donor zip codes against each donor\_id. The irrelevant fields were dropped and basis the zip codes we could now merge this data with the US\_postal\_codes data to map the respective regions for donors.

Dummy variables were created for each of the four regions just like the previous analysis. (South, West, Midwest and Northeast)

Quarters in which donations and appeals were made was also extracted as new columns – Donations\_qtr and Appeals\_qtr to check seasonality in the data.

Appeal\_cost had a positive skew and its square root was taken to remove skewness as confirmed by running Proc Means.

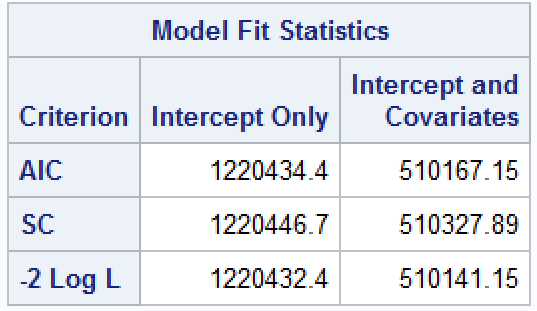
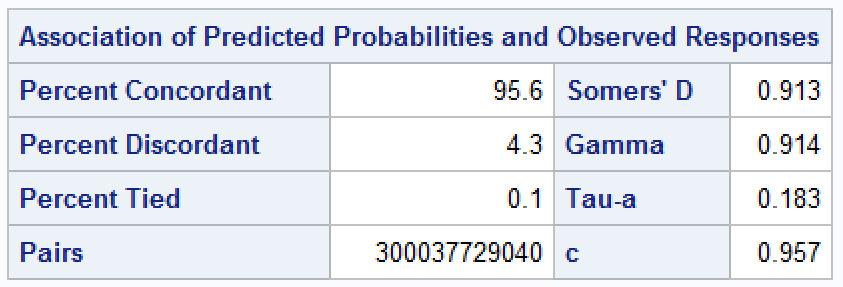


Now the dataset had a binary outcome in donate (0,1) which could be fitted to a logistic or a probit model.

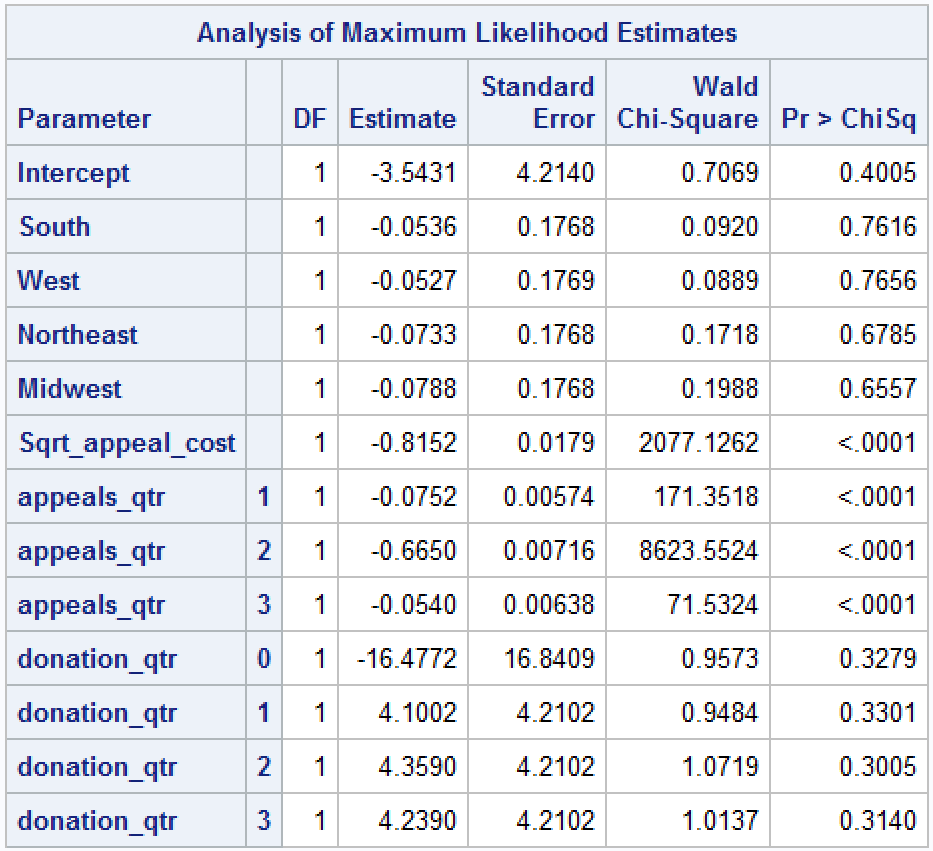
**Logistic Regression:** A logistic regression model was run with donate as the dependent variable. The model equation was as follows:

*donate (event='1')= South West Northeast Midwest Sqrt\_appeal\_cost appeals\_qtr donation\_qtr;*

Basically, we were trying to ascertain whether a person donating is dependent on his region , cost of the appeal , the quarter in which the appeal was sent and the quarter in which the donation was received.

As we can see from the above statistics this is a good model with a MCfadden R square value of 0.58



The key insights that can be drawn from the model are-

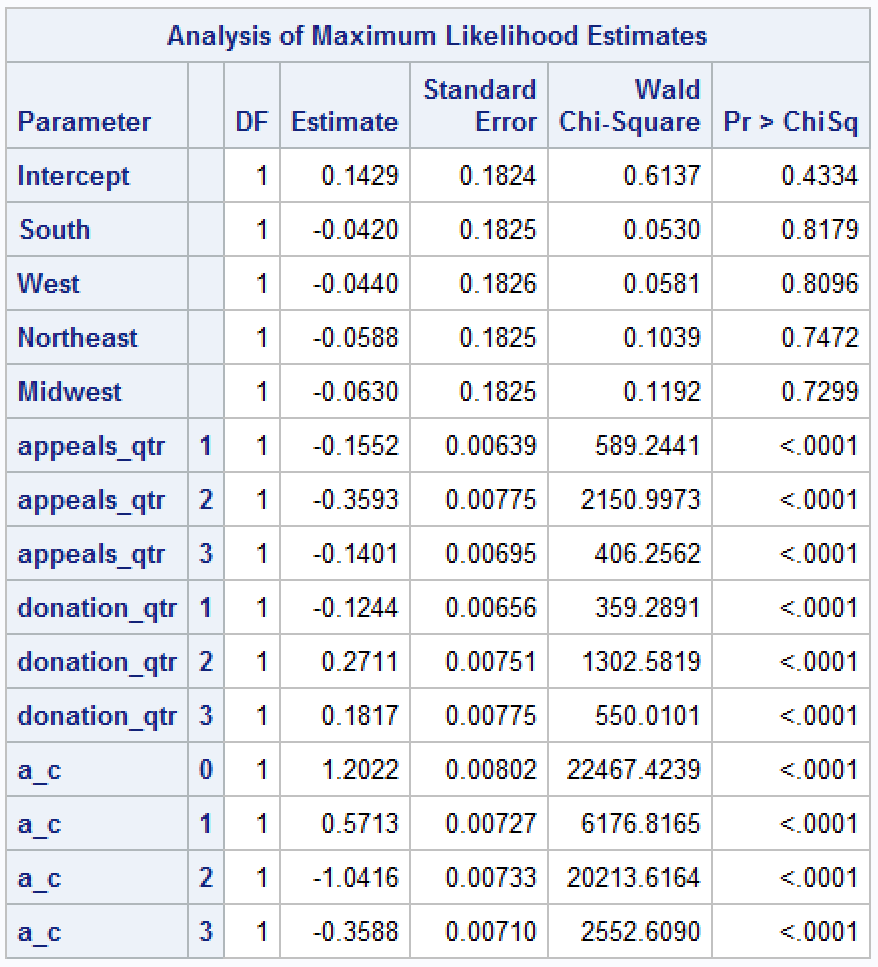
1. The coefficient for sqrt\_appeal\_cost is negative meaning **higher cost of sending an appeal does not imply that a person will donate more**
2. Quarter four is the best and most significant for sending appeals , Qtr 1,2 and 3 though significant in the model have negative co-efficients.

To further ascertain our findings, we ranked the appeal costs into 5 categories – 0 being the lowest and 4 being the highest. Proc Rank was used to achieve this. The basic idea was to bin appeal costs and then check if higher appeal costs meant higher donation amounts.

After processing the data a logistic model was run with donate(0,1) being the dependent variable. The model equation is shown below-

*model donate (event='1')= South West Northeast Midwest Sqrt\_appeal\_cost appeals\_qtr donation\_qtr a\_c;*

Here we tested weather a person’s propensity to donate is a factor of his region the appeal quarter , donation quarter or the rank variable created(a\_c\_). The outcome is shown below



Rank 4 has been set as a reference. From this output we can interpret that rank 0 ($0 to $0.29) and 1 ($0.29 to $0.32) which corrspond to the lower ranges of appeal cost have the maximum impact on donations.

Therefore , it is safe to conclude that higher appeal costs do not warrant higher donations.

**Behavioral factors analysis**

**Objectives –**

* Is past donation amount of a donor a good predictor of the future donation amount by that donor?
* What patterns do we see in Recency and Frequency of past donations and Frequency of Future donation? Discuss them.
* Do Recency and Frequency have significant impact on future donation behaviors?

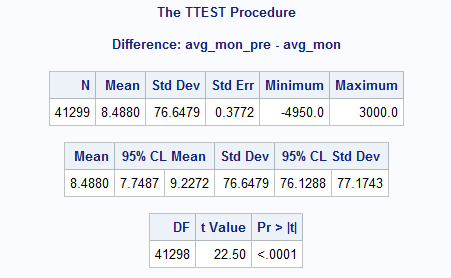
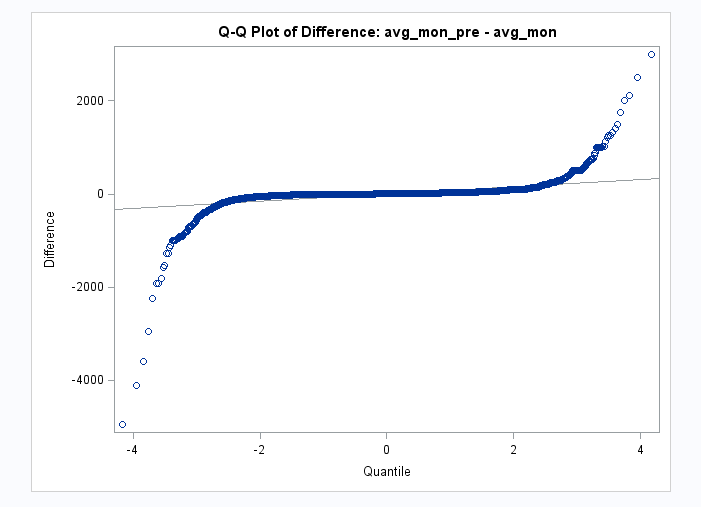
**Description:** RFM (recency, frequency, monetary) analysis is a marketing technique used to determine quantitatively which customers are the best ones by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary). RFM analysis is based on the marketing axiom that "80% of your business comes from 20% of your customers."

Using RFM analysis, customers are assigned a ranking number of 1,2,3,4, or 5 (with 5 being highest) for each RFM parameter. The three scores together are referred to as an RFM "cell" . The database is sorted to determine which customers were "the best customers" in the past, with a cell ranking of "555" being ideal.

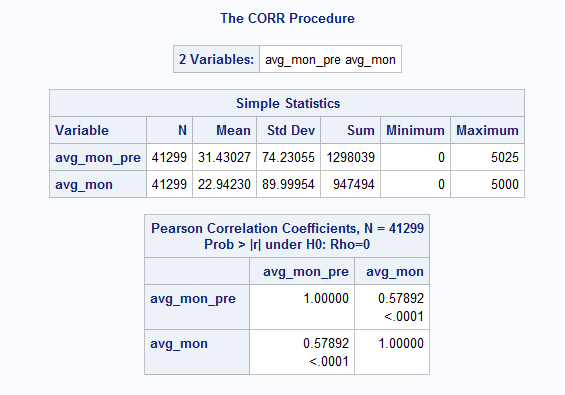
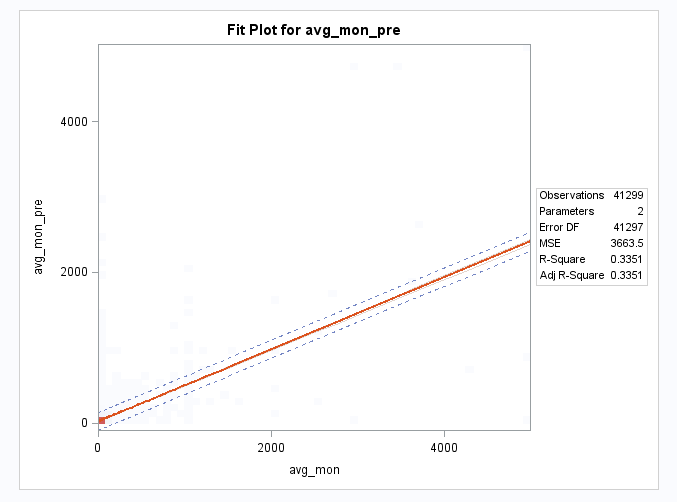
Using this approach, we used to process the data and found out recency, frequency and monetary value for Donations and Appeal as a function of Quarter. **We decided to divide the time frame into Quarter as using Quarter gives us the necessary amount of data points to conduct our analysis. Using Year would have yielded very few data points and using months would have yielded too many.** For monetary value, we used average monetary value so as to standardize it and see its impact. As per our 1st objective, we wanted to determine whether past donation amount of a donor is a good predictor of the future donation amount by that donor. For this we simply used **Paired TTest.** The results of the TTest were as follow –

***H0: μd = μ0*** *- The population mean of the differences (μd) equals the hypothesized mean of the differences (μ0).*

***H1: μd ≠ μ0*** *- The population mean of the differences (μd) does not equal the hypothesized mean of the differences (μ0).*

As seen above, we can see that P value is < 0.001 which mean we reject Null hypothesis, means that the means are not equal. But analyzing from Q-Q plot pf the differences, we can see that majority of data lies near 0. Hence, we further analyzed the data using Correlation and Regression. We got the following chart –

As seen in the chart, we can see that the correlation is approximately ≈ .6, which is a decent correlation factor, but the regression line seen in the chart is a linear, hence **we can conclude that past monetary amount of a donor is somewhat a good factor to predict his future donation.**

Coming to our second objective, we attempted various approach to see the effect of past recency and frequency on future donation frequency. We ran models such as Correlation Matrix, Regression, Frequency tables etc. and saw that there was a pattern emerging between the same.

**Impact of Recency on Future Donation:** To see the impact of Recency on Future frequency, we divide our past recency into two parts.

* Part 1 – In this section we included only those Donors (Made a column and updated it with value 1 if donated) who donated in the latest year of Pre-data i.e. year 2003.
* Part 2 – In this section we included only those Donors (Made a column and updated it with value 0 if not donated in the latest year) who haven’t donated anything in the latest year of Pre-data that is there is no donation from their side in the year 2003.

Now, after dividing the data into two parts, we saw that those donors who were active in donating in the last year of Pre-data (2000- 2003), donated in future as well. It can be clearly seen below –

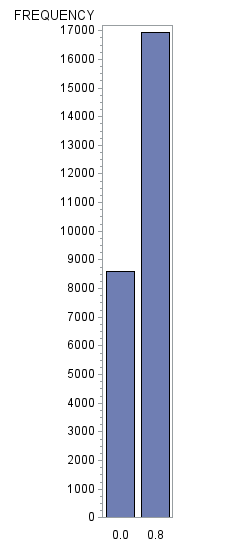
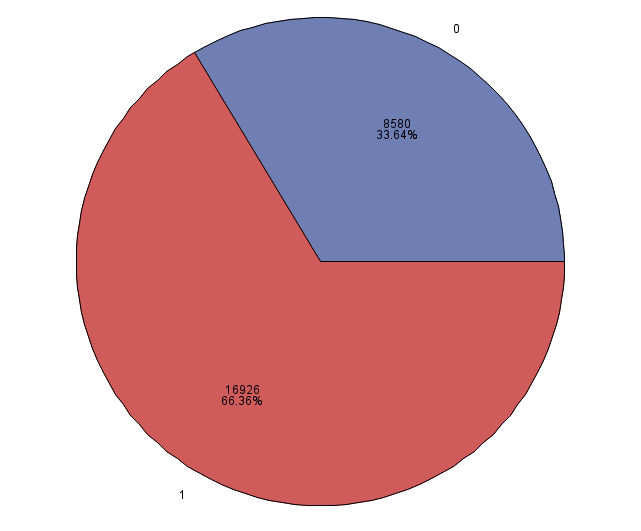


Chart for Donors who donated in the past one year

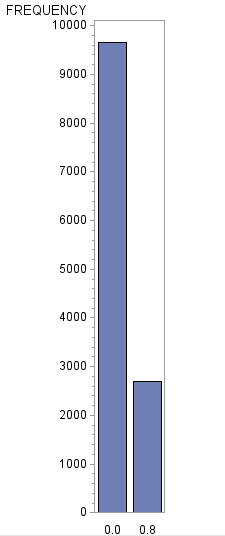
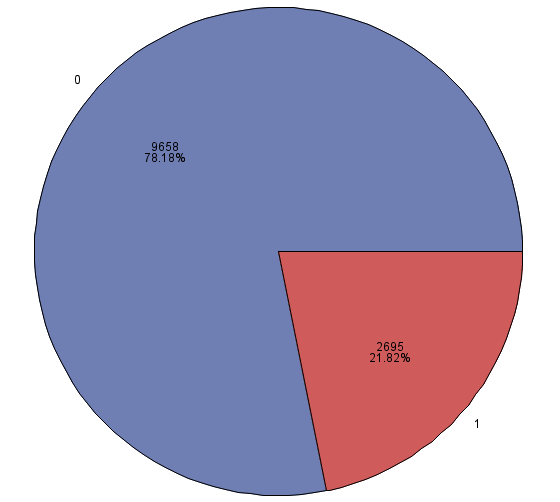


Chart for Donors who didn’t donate in the past one year

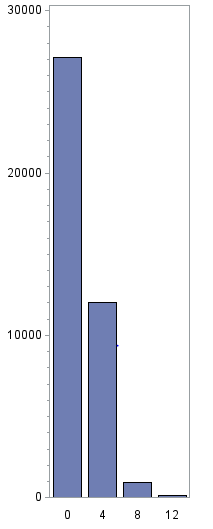
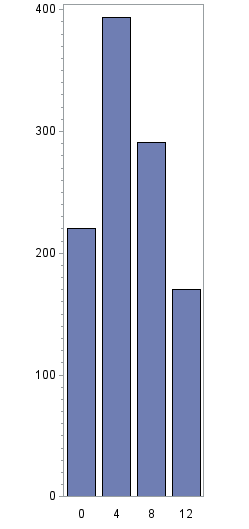
Here the ‘1’ are those donors who donated in the latest year (year 2003) and ‘0’ are those who didn’t. We can clearly see that there is a higher probability of donating for those donors who donated in the latest year than those who don’t.

**Impact of Frequency on Future number of Donation –**

Similarly, to see the impact of Frequency on future number of donations we divided our data set into 2 parts –

* Part 1 – Donors having donated more than 8 in the Pre-Data (year 2000 – 2003)
* Part 2 - Donors having donated less than or equal to 8 in the Pre Data (Year 2000 – 2003 )

After processing the data, we clearly saw a correlation in numbers. We can see in the chart below –

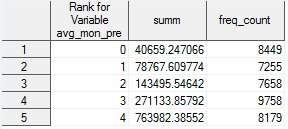


Frequency greater than 8 Frequency Less than equal to 8

As seen in the Bar chart above, it is very evident if donors actively donate throughout the year (more than once every quarter, hence the cutoff ‘8’) tend to donate in the future as well (Year 2004 – 2006). Whereas if donors are less frequent or do not donate at-least once per quarter, they tend to donate less or doesn’t donate only in future.

The possible reasons for the explanation of this behavior is that, that **Recency**, or the number of days that have gone by since a customer completed an action is the **most powerful predictor** of the customer repeating this action. **As each day that goes by after the customer completed the action, the customer gets less and less likely to repeat it**. So, conducting RFM Analysis helps us to understand which all donors based on their previous behavior are going to make donations and those who are not. Since we segmented and separated those customers who were active in the last year in pre-data (year 2003) i.e. they were more active recently, hence are more likely to remain active in the future and those who weren’t have defected and stopped donating. This can be clearly seen in our analysis.

Also, we did an analysis of Pareto’s Principal, that states, “80% Value comes from 20% Customer”.



We can clearly see from the analysis above; 73% Value comes from 22% customer. Hence, we can see that Pareto’s Law somewhat holds true over here.

Hence, we can conclude that Recency and Frequency does have significant impact on frequency of donation in the future. The more recent a customer has donated, the more likely he going to continue donating in future as well, which is the main principal of RFM Analysis. Hence, our data follows this principal as well, and we can use RFM Analysis to get Managerial insights as to which all customer to target, how to target and when to target.