HHUSA FACEBOOK REGISTRATION CAMPAIGN ANALYSIS

Inquiry and Research into User and Financial Results

Abstract

HHUSA has been running a Facebook Registration campaign since December 18th. The ads are targeted at military personnel and veterans and are running in geographical areas where there are large military bases. Initial reports from Hubspot and Facebook showed nearly 45,000 people have gone to their website from social media. HHUSA goal is to get transitioning military members and veterans to register for their free services to better help them transition to the civilian workforce. To support HHUSA, UNCC has been engaged to perform an empirical study into the results of the Facebook Registration campaign.

Rachit Choksi, Shravya Maroli, Ankita Sharma,
Ali Algarmi, and Josh Peters

1.0 INTRODUCTION

Hire Heroes USA (HHUSA) empowers U.S. military members, veterans, and spouses to succeed in the civilian workforce. Through their free, signature workshops and online programs, clients are individually partnered with a highly trained Veteran Transition Specialist who works with the client to: create a tailored civilian resume that effectively highlights skills and achievements, translate military experience into civilian terminology, learn effective job searching, networking and interviewing techniques, and get connected with companies that actively hire veterans. As a result, HHUSA confirms hired on average more than 60 clients per week. Though this success has allowed growth in both scope and impact, there are still roughly 500,000 unemployed veterans that exist at any given time. To better reach this population, HHUSA would like to analyze existing client, social media, and volunteer data to determine if there are opportunities for further improvements to our systems.

HHUSA has been running a Facebook Registration campaign since December 18th. The ads are targeted at military personnel and veterans and are running in geographical areas where there are large military bases. Initial reports from Hubspot and Facebook showed nearly 45,000 people have gone to their website from social media. HHUSA goal is to get transitioning military members and veterans to register for their free services to better help them transition to the civilian workforce.

2.0 OBJECTIVE

To support HHUSA, UNCC has been engaged to perform an empirical study into the results of the Facebook Registration campaign. Our objective will be to research and articulate the findings pertaining to the following questions:

i. Can we trace any of the registrants since December 18 specifically to ads targeted to users;
ii. Can we show a historical cause-effect benefit relating to marketing/branding awareness (social media, web, etc) and the results it had on registration for HHUSA services;
iii. Can we trace specific trends (patriotic holidays such as Memorial Day, Independence Day, and Veterans Day) to certain days;
iv. Can we show a historical cause-effect relationship between social media messaging/online presence and actual annual donations; and
v. Has there been a measurable effect on fundraising over the years from the brand awareness that is created by having a website, social media, etc.

The aforementioned research questions will be the focus of our analysis. Upon conclusion, the UNCC research team will have opine on the results of existing strategies and determine if it has predictive probability for success based on our observation into the current data.

3.0 DATA WRANGLING

HHUSA has provided the UNCC research team with the data sets needed to perform our analysis. The data is comprised of HHUSA's varies media channels. The following deliverables were provided to the research team:

- Facebook data files
- Hubspot SM data files
- LinkedIn data files

- Twitter data files
- Google analytics files
- Donor perfect files

To get the data in the appropriate format, the research team performed the following procedures:

i. Facebook page and post data had different number of columns for each file - starting from 350, it varied up to 3,000.
ii. Using excel, the research team kept only the common columns and combined it. GA data had different datasets in the same file, so the team split all of them and created different files. Most of the string columns in all files had irregular characters at the end, so we used substr operator in SAS enterprise guide to remove it.
iii. Dates were imported as string, so to maintain it as date format we had to do some manipulations in excel as well as date functions in SAS enterprise guide to maintain the datatype. Numbers were imported as string because of irregularities in characters. So, we removed it using a substr operator.
iv. All twitter files had the same number of columns, so we just combined it using excel.

The software used by our team for analytics were SAS 9.4, EG, and E-Miner.

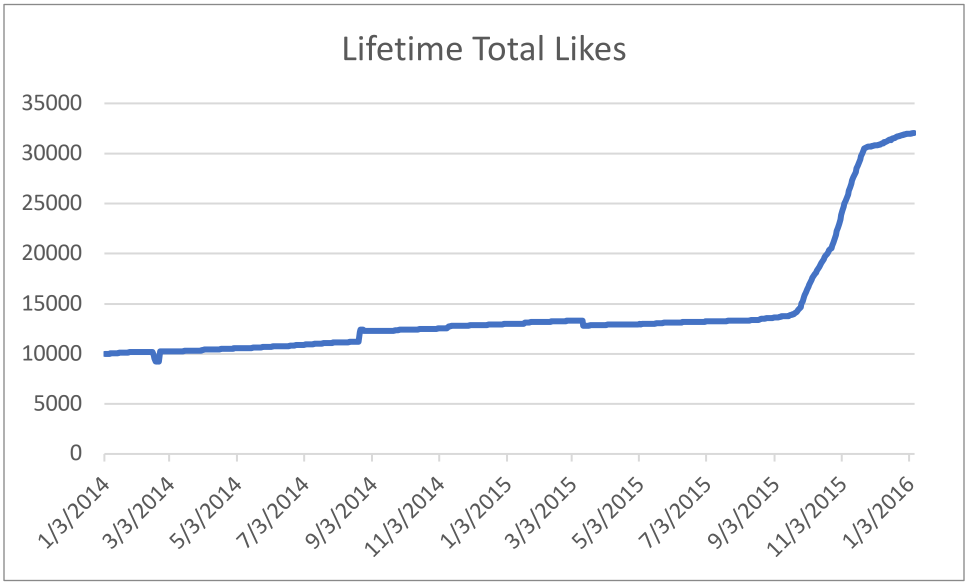
4.0 DATA EXPLORATION AND DESCRIPTIVE ANALYSIS

To help transition military members and veterans to the civilian workforce, Hire Heroes USA (HHUSA) marketing strategy has integrated several social media platforms to support their initiative. As with any strategy, HHUSA needs to be able to generate funds to ensure adequate resources. Currently, HHUSA has an active presence online for which they use to promote and curate content and events to their members and partners. Using their multiple media channels, HHUSA can get their message to the public and by extension potentially boost contributions to their cause. Since January 3, 2014, HHUSA has obtained a strong following on Facebook as well as on social media accounts. Facebook is by far the largest of the platforms with a total page likes of 32,044 as of January 7, 2016. Given HHUSA focus on running campaigns through Facebook, we will determine if these efforts have yielded a boost in brand awareness and annual donations.

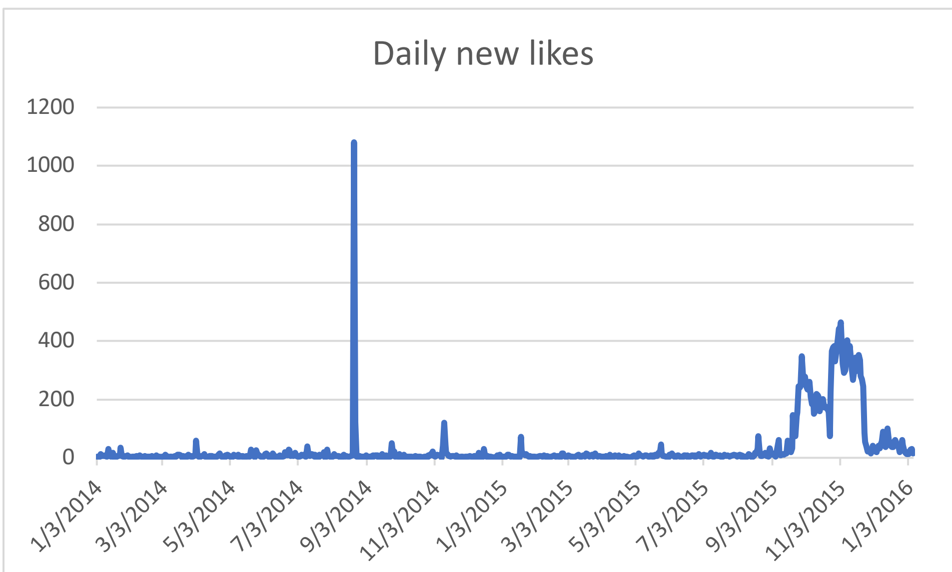
The number of people HHUSA Page reached broken down by how many times people saw any content about the Page had the highest number of daily on a daily total and page post frequency distribution shows a relatively consistent trend across the stratified averages.

The histograms illustrate that on average users are most like going to see HHUSA's content pushed through the Page once.

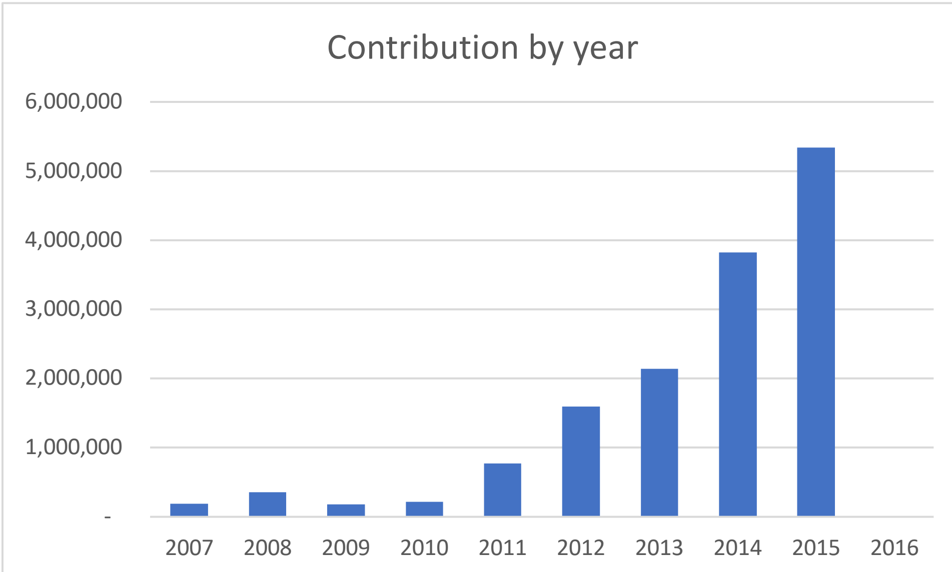
HHUSA's Page from 2014 to 2016 for new and lifetime users experienced constant growth. Lifetime growth had stayed gradual until September 2015 where page likes surged and did not level off until November 2015.



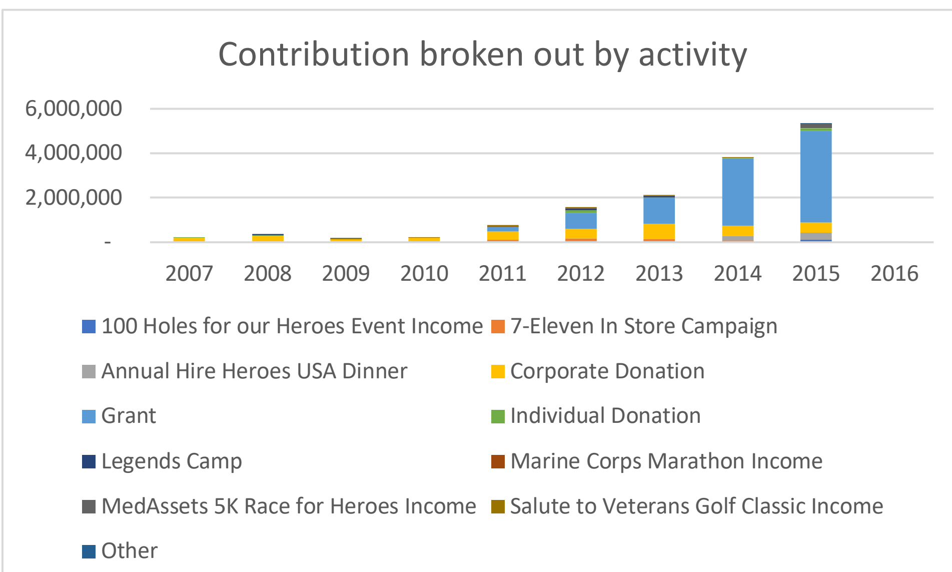
Additionally, lifetime likes had a small shift in September 2014 which when observed at the daily likes level we can see a significant spike in the numbers of likes for that period.



The next set of data that was explored was historical financial records. HHUSA has experience significant growth in total contribution over the past several years.

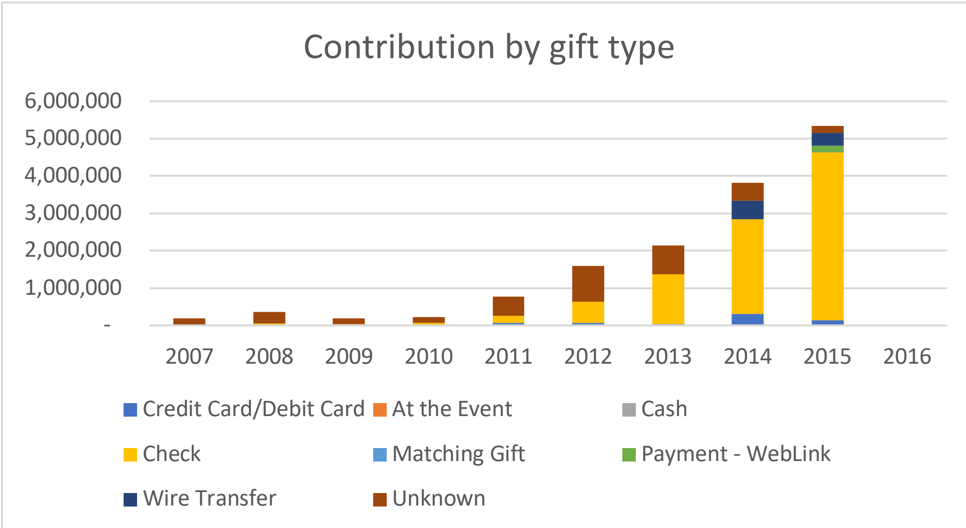


Examining total revenue by year, we see growth follow an exponential trend starting from 2011. Delineating further, HHUSA's revenue by year is broken out by contribution activity: Holes for our Heroes Event Income, Eleven in Store Campaign, Annual Hire Heroes USA Dinner, Corporate Donation, Grant, Individual Donation, Legends Camp, Marine Corps Marathon Income, MedAssets 5K Race for Heroes Income, and Salute to Veterans Golf Classic Income.



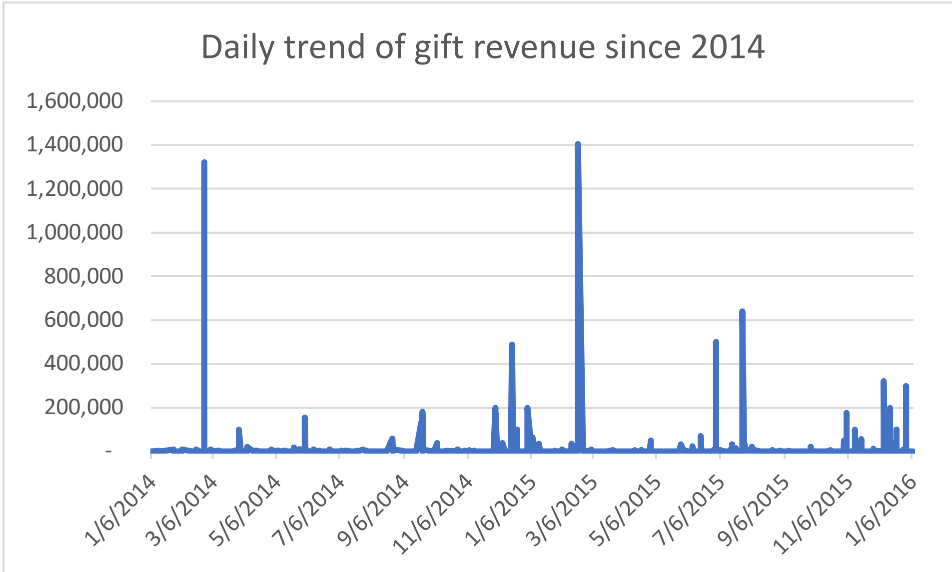
We see the largest spread in growth coming from grants. These contributions have significantly contributed to HHUSA bottom line. In February 2014 and December 2014, HHUSA received a $1.3 million and $1.2 million contribution respectively. The following year HHUSA received grant contributions for $1.5 million and $1.1 million in February and July respectively.

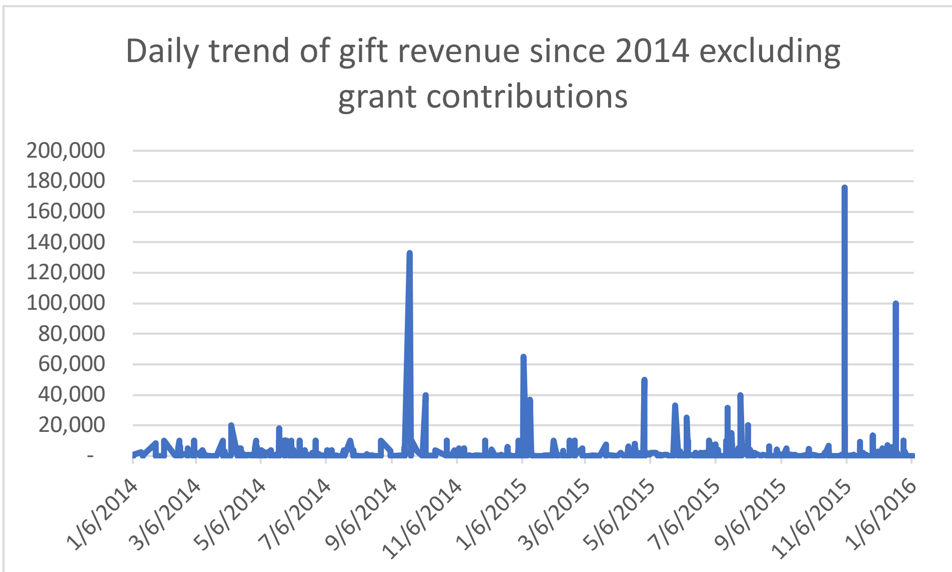
Next, we looked at the breakout of contribution by gift type.



Contribution by gift type within the different cohorts indicate most of contributions are received by checks. This is consistent with expectations given that most contributions received by grants. Another interesting observation was the relatively significant amount from unknown gift types. This number made up most of the contributions from 2007 to 2011 and about half of the total contributions in 2012. This trend away from unknown sources of gift type is a positive trend and most like can be attribute to better accounting practices and software upgrades.

Looking for structural changes, it appears the HHUSA operations started picking up in early 2012 and again significant surge in contributions in early 2015. This is of course not taking into consideration the large grants that was received previously mentioned.



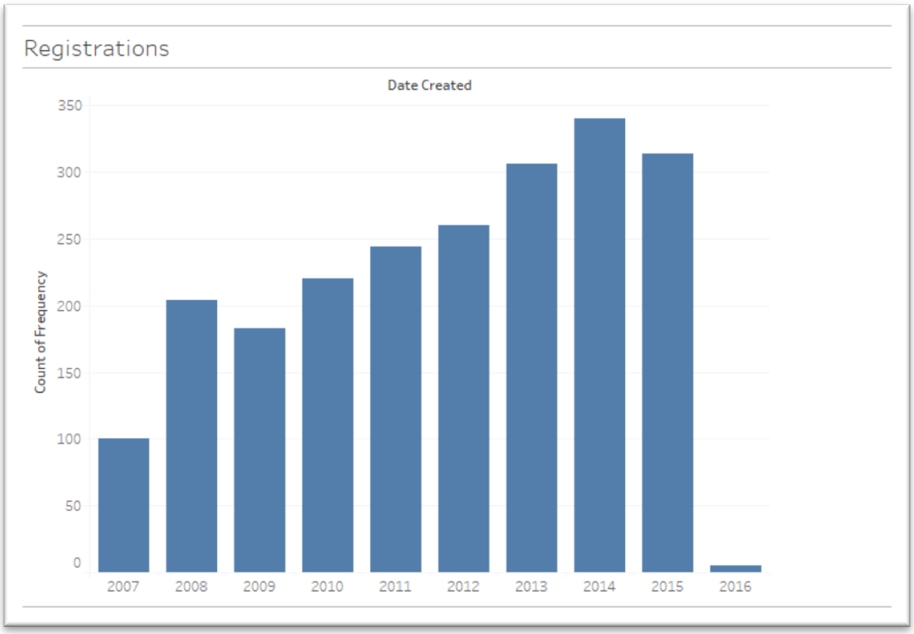


5.0 RESEARCH INQUIRY AND METHODOLOGY

5.1 Cause-effect relationship of online marketing/branding awareness

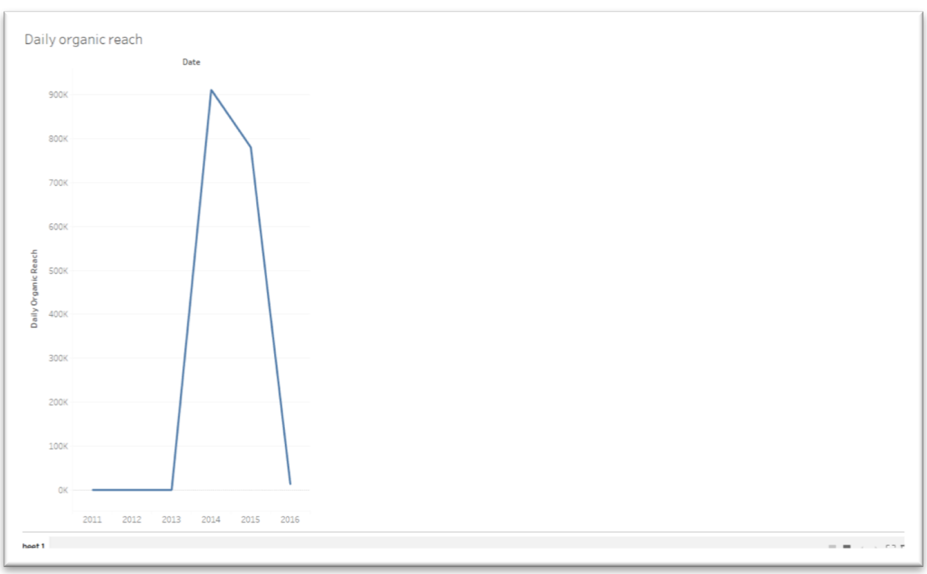
To comment on the historical cause-effect benefit based on registrations for the free services provided by Hire Heroes USA, we need to know the registrations the company received, which was not present in marketing data. However, the client data of Hire Heroes has the accounts created daily. This can help us in knowing a rough understanding of how many registrations might have been possible every day based on the record type.

To calculate this, we first import account data to SAS E-guide and use aggregate function and group-by using Query builder to identify the number of registrations that happened each day. If we plot it in tableau we get the below result.



No registration data is available for the year 2016. Hence, we have not considered year 2016 in the analysis.

Now considering Facebook page data (social media data). Organic Reach in the FB Page data is the number of unique users (Fans or non-Fans) who saw your Page post in News Feed, Ticker or on your Page



As you can see in the above graph, years 2014 and 2015 only have real values. 2011-2013 has 0 values. Taking these years into consideration wouldn't provide us with valid results.

Considering all the above analysis, we have decided to find historic cause-effect benefit only for the years 2014 and 2015.

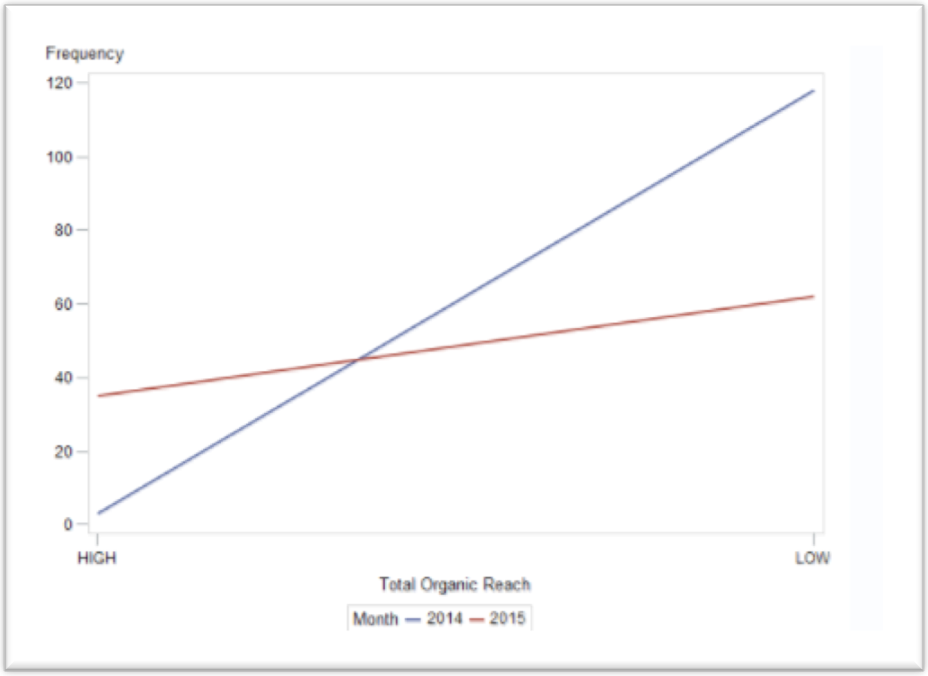
Preparation of data to do experimentation to understand the cause-effect benefit.

- The average of organic reach is 2500 compared for all days in 2014 and 2015. The organic reach is considered low for a year if the average of organic reach for the year is less than 2500 and as high If it more.
- The average of frequency of account created for all days in 2014 and 2015 is 9. The accounts created for the year is low if it is less than this 9 and high if it is more than 9.

Based on this understanding, the below table was created where response is the accounts created.

	Total Organic Reach	Month	Response	Frequency
1	LOW	2014	0	204
2	LOW	2014	1	118
3	LOW	2015	0	153
4	LOW	2015	1	62
5	HIGH	2014	0	15
6	HIGH	2014	1	3
7	HIGH	2015	0	64
8	HIGH	2015	1	35

Using SAS E-guide, plotted the below Multiple lines plot. The graph shows that for both the years 2014 and 2015, the frequency of response rate = 1 is increases, when Organic reach decreases.



However, while running logistic regression, the results were not significant.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	834.561	834.275
SC	839.044	852.207
-2 Log L	832.561	826.275

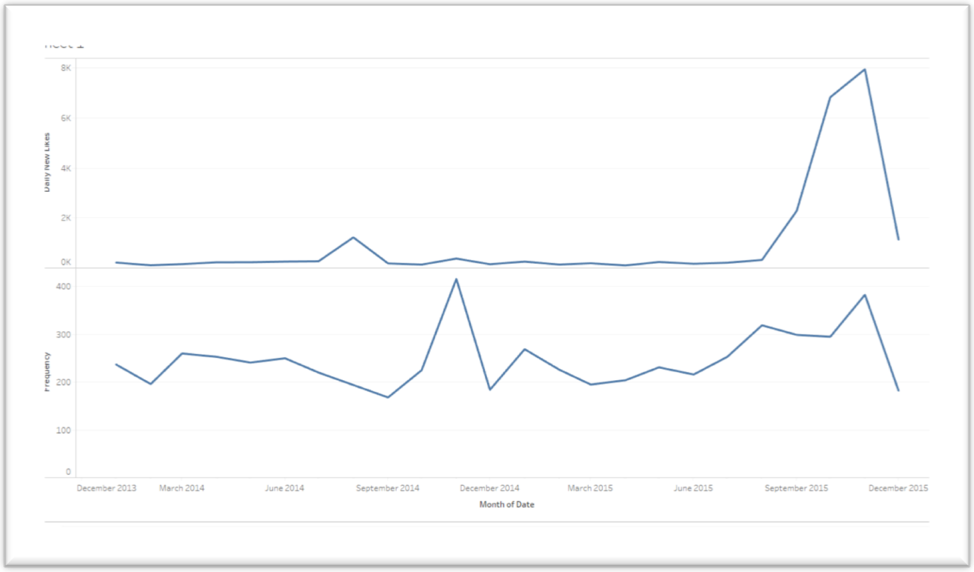
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6.2856	3	0.0985
Score	5.9777	3	0.1127
Wald	5.8037	3	0.1216

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Month	1	0.8798	0.3483
Total Organic Reach	1	1.2098	0.2714
Month*Total Organic	1	3.8614	0.0494

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-0.9159	0.1732	27.9503	<.0001
Month	2014		1	-0.1625	0.1732	0.8798	0.3483
Total Organic Reach	HIGH		1	-0.1906	0.1732	1.2098	0.2714
Month*Total Organic	2014	HIGH	1	-0.3404	0.1732	3.8614	0.0494

One of the recommendations can be that, for each of the statistics (around 300), which we have from Facebook data like total reach, organic reach, video views etc. the above experimentation procedure can be run to understand the effect on response for each of the case.

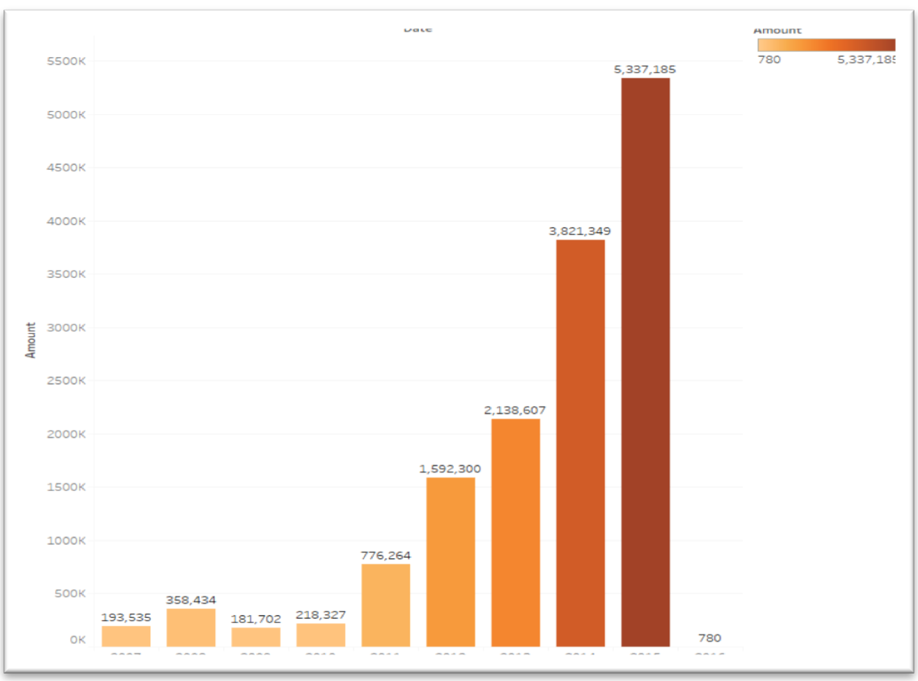
However, most of the individual statistics was plotted against the response metric (account created) in tableau and the interesting plot was found for daily new likes.



As seen above, as the daily new likes increases, the frequency of accounts created is increased in most cases. However, November (2014 and 2015) have more likes as well as accounts created. Recommendation is that this month can be analyzed to know more about this underlying effect.

5.2 Measurable effect on fundraising from online brand awareness

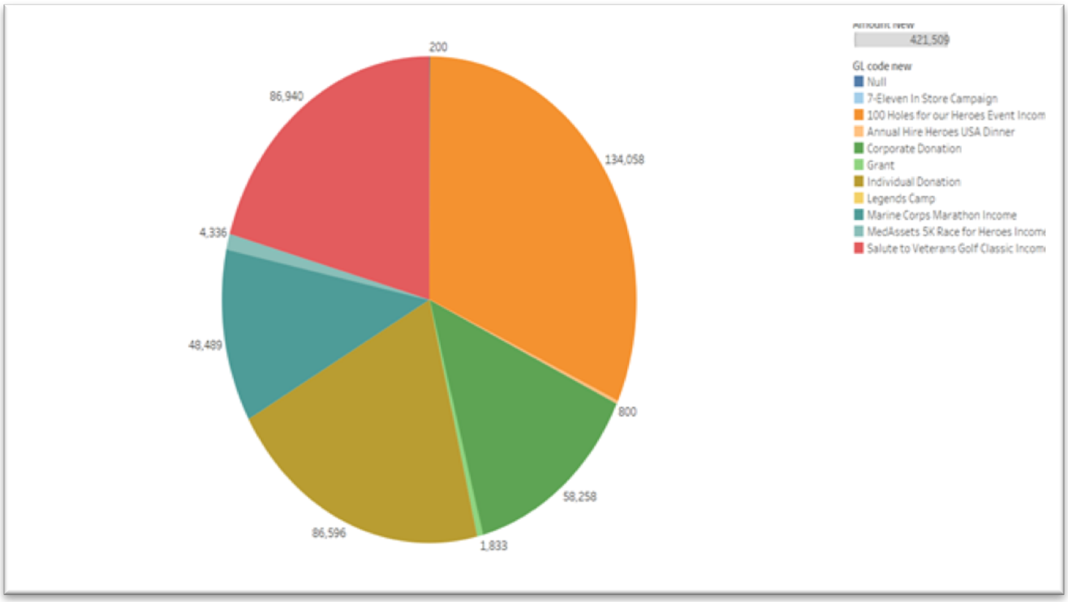
To calculate a measurable effect on fundraising over the years from the brand awareness created by social media, we have showed relationship between the total donation received on yearly basis.



We can see that donation amount is gradually increasing from 2007 to 2010. But from 2011, there is a drastic increase in donation amount and from there it is also increasing gradually for every year except 2015.

We are not able to find out the actual reason for why there is not much donation amount received from 2007 to 2010 because we have data available for social media only from 2011 to 2016.

The pie chart displays various type of donation received and the amount.



We can analyze that "100 Holes for our Heroes Event Income" has received maximum donation of $134058 and "Salute to veterans Golf Classic Income" has received donation of $86,940.

6.0 EVALUATION OF OUR RESEARCH

After our initial inquiries we explored the variables within Facebook Insights and DonorPerfect to determine which variables would be critical for our analysis. The following variables were selected:

Variable	Short Description	Long Description
VAR0003	Daily New Likes	Daily: The number of new people who have liked your Page (Unique Users)
VAR0011	Daily Organic Reach	Daily: The number of people who visited your Page, or saw your Page or one of its posts in news feed or ticker. These can be people who have liked your Page and people who haven't. (Unique Users)
VAR0026	Daily Logged-in Page Views	Daily: Page Views from users logged into Facebook (Total Count)
VAR0045	Daily Paid impressions of your posts	Daily: The number of impressions of your Page posts in an Ad or Sponsored Story. (Total Count)
VAR0048	Daily Total Consumers	Daily: The number of people who clicked on any of your content. Stories that are created without clicking on Page content (ex, liking the Page from timeline) are not included. (Unique Users)
VAR0072	Daily count of fans online	Daily: The number of people who liked your Page and who were online on the specified day. (Unique Users)
VAR0115	Daily People Talking About This	Daily: The number of people sharing stories about your page. These stories include liking your Page, posting to your Page's timeline, liking, commenting on or sharing one of your Page posts, answering a question you posted, responding to one of your events,

		mentioning your Page, tagging your Page in a photo or checking in at your location. (Unique Users)

After identifying the variables, we then ran descriptive statistics for each of the variables selected. The variables are as of January 3, 2014 through January 7, 2016.

Variable	Description	Mean	Std Dev	Minimum	Maximum
VAR0003	Daily New Likes	33.47	86.64	0	1081
VAR0011	Daily Organic Reach	2318.61	9027.85	17	170296
VAR0026	Daily Logged-in Page Views	22.16	27.50	0	513
VAR0045	Daily Paid impressions of your posts	1961.88	10903.54	0	90720
VAR0048	Daily Total Consumers	153.66	561.12	0	4362
VAR0072	Daily count of fans online	11790.17	5207.94	0	29308
VAR0115	Daily People Talking About This	78.68	143.25	0	1709

N=735 for each variable

```
/*Descriptive statistics*/
proc means data=modfbdata2;
var VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0063 VAR0072 VAR0115
run;
```

We noticed many of the variables had a minimum of zero, which is not uncommon given the data set is on a daily time interval. Looking a daily total reach and daily organic reach both variables have a minimum and maximum daily content seen of 17 and 170,296 respectively. This suggest that most of the individuals that interact with HHUSA's page are organic with a slight standard deviation of 3,084 non-organic viewers. This is also the case with daily impressions which has a higher average than reaches for both total and organic users. HHUSA's average daily customers and consumptions is 154 and 256 respectively.

6.1 Trends in user activity and cause-effect relationship between online presence and actual annual donations

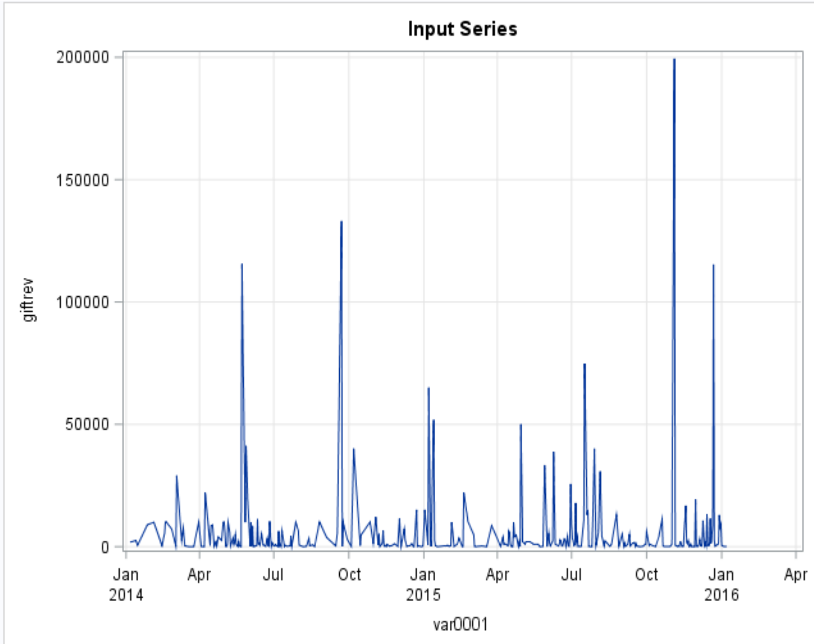
A OLS model was estimated from the time-series data to determine a historical cause-effect relationship between social media messaging/online presence and actual annual donations. Each observation i includes a scalar response y_i. In a linear regression the response variable is a linear function of the regressors:

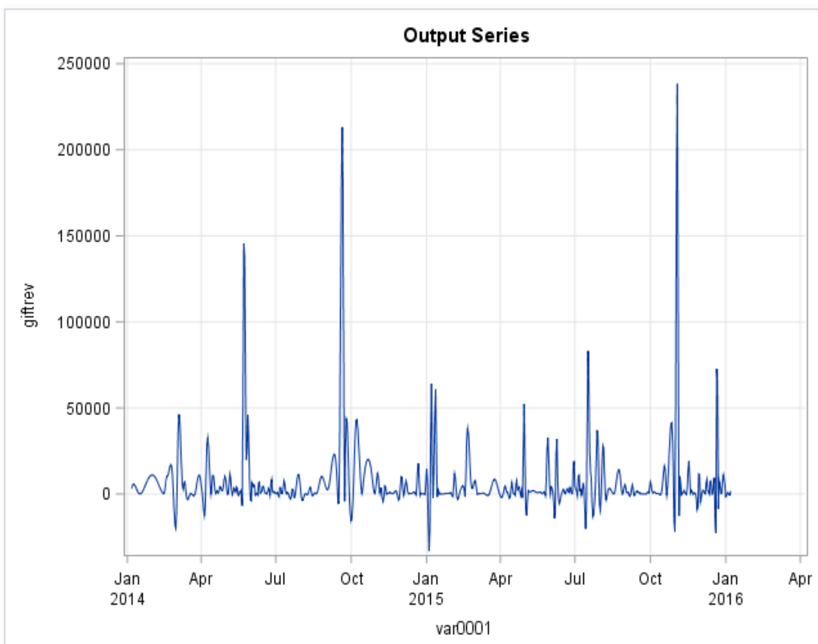
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \ldots, n$$

Where β is parameter estimate; x_i is the regressors of n data points; ε_i's are unobserved scalar random variables which consist of the residuals between the actually observed responses y_i. Amount data was provided on an irregular time interval, therefore we used a spline interpolation from the irregular days to daily to get data from Facebook Insights and general ledger in the same time interval. Furthermore, due to grants being significant single gifts this contribution category was excluded. Spline interpolation results are consistent with actuals.

```
/*Spline interpolation*/
proc expand data=teststat6 out=teststat7 to=day plots=(input output);
id var0001;
convert giftrev / observed=(beginning,average);
```

```
run;
```





Our first iteration of the model was not full rank. That is, least-squares solutions for the parameters are not unique and statistics were misleading (i.e. bias in the estimation). We further developed the model to ensure situations where the coefficient may change erratically in response to small changes in the data (i.e. multicollinearity), and thus the model was full rank. We then used a White test to check for heteroscedasticity and used Newey-West correction for heteroscedasticity and autocorrelation to have a more robust model.

$$\text{giftrev}_i = \beta_0 + \beta_1\text{var0003}_i + \beta_2\text{var0011}_i + \beta_3\text{var0026}_i + \beta_4\text{var0045}_i + \beta_5\text{var0048}_i + \beta_6\text{var0072}_i + \beta_7\text{var0115}_i + \varepsilon_i$$

```
/*OLS model with heteroscedasticity and multicollinearity test*/
proc reg data=modmasterfile01;
model giftrev=VAR0003 VAR0011 VAR0045 VAR0048 VAR0063 VAR0072 VAR0115 / vif
white;
output out=modmasterfile3 p=yhat;
run;

/*OLS model with Newely-West corrections*/
proc model data=modmasterfile01;
parms a0 b1 b2 b3 b4 b5 b6 b7;
giftrev=a0+b1*var0003+b2*var0011+b3*var0026+b4*var0045+b5*var0048+b6*var0072+
b7*var0115;
fit giftrev / gmm kernel=(bart,1,0);
run;
```

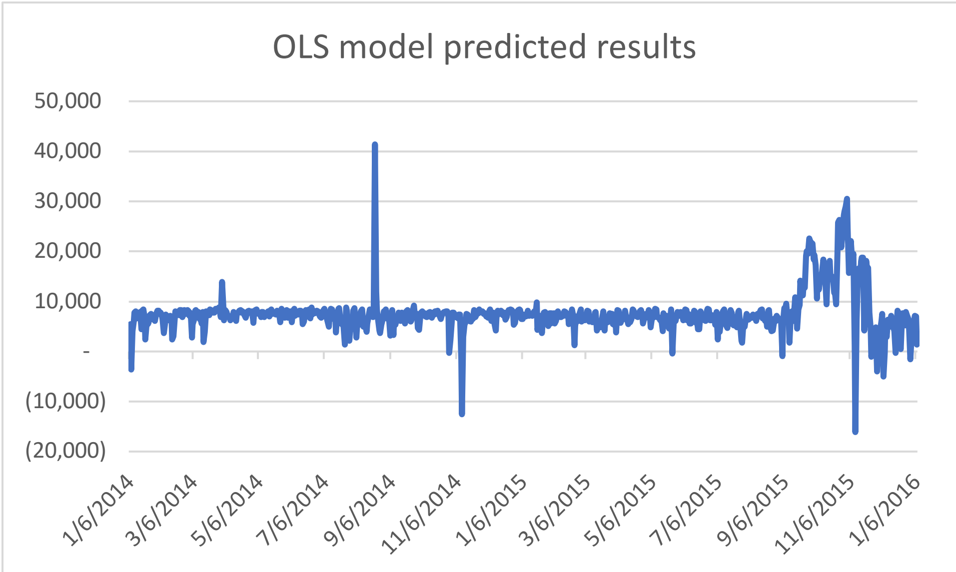
Table: OLS model estimates for the dependent variables

Parameter	Estimate	Approx Std Err	t Value
Var0003	**88.08773****	42.3288	2.08
Var0011	-0.07753	0.0501	-1.55
Var0026	-54.6474	42.9953	-1.27
Var0045	-0.15597	0.1355	-1.15
Var0048	2.816962	3.3376	0.84
Var0072	-0.0663	0.1306	-0.51
Var0115	**-31.4224****	14.2266	-2.21

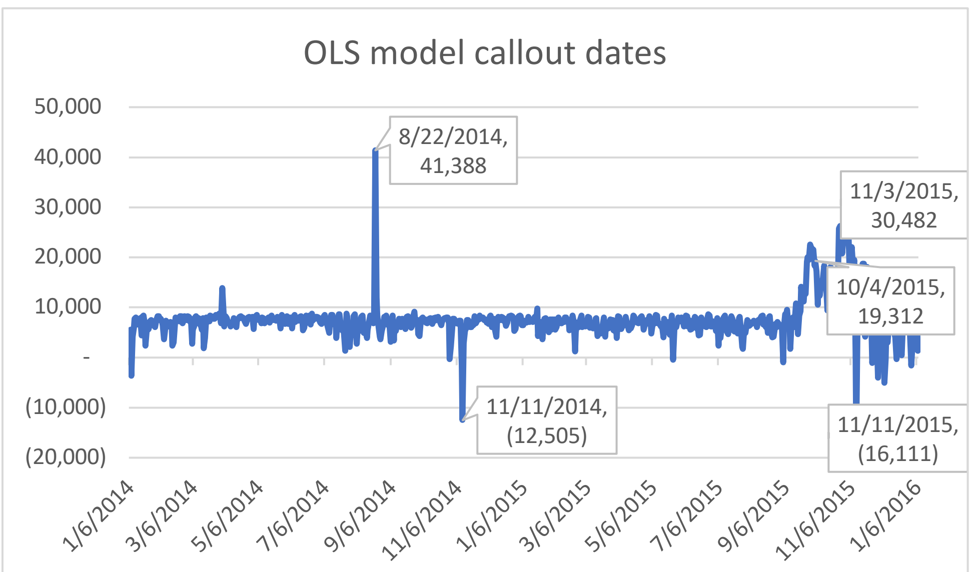
** p < .05

The $R^2 = .0368$, $Adj\ R^2 = .0275$, and F-statistic $= 3.55$ (p < .01). The R^2 value only explains approximately 4% of the variation in contribution revenue. However, our F-statistic is less than the critical value meaning the coefficients as a group are jointly no different from zero. Therefore, we reject the null as a group (results are significant).

Daily new likes coefficient is 88.088 meaning that for every marginal change in daily new likes, on average, contributions are going to increase by $88.1. This suggests that campaigns via posted on Facebook, given the number of users that liked the page, HHUSA should exact a marginal gain of $88.1 *ceteris paribus*. Conversely, the parameter results for the other variables were negative except for daily total consumers. We noted during the development of the model that the coefficients were sensitive to changing variables when correcting for collinearity. Individually these results were not statistically significant. We suggest that the negative coefficients can be understood as marginal loss opportunity given strong user activity being generated by the Facebook page. That is, the marginal loss on not capturing user activity for a specific day.

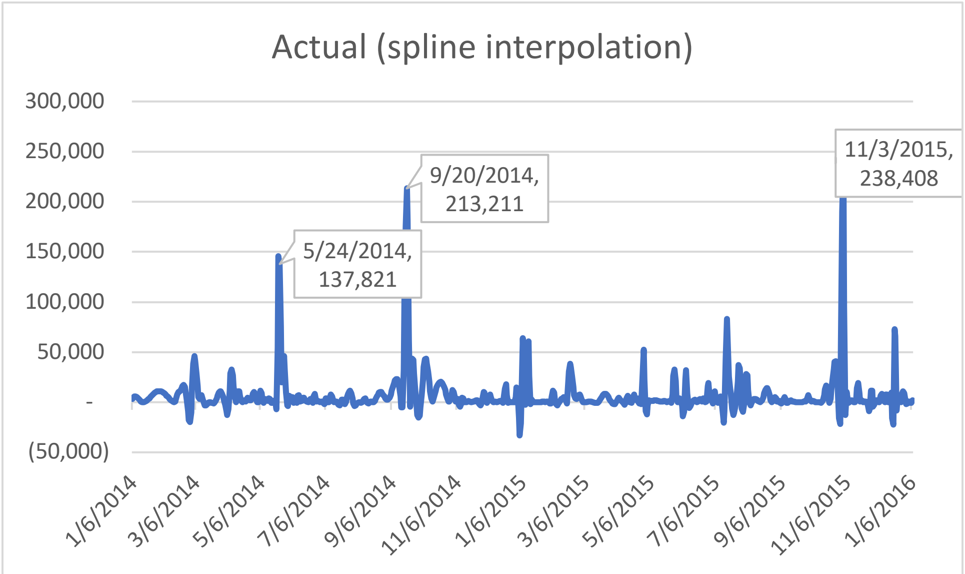


Next, the model developed by the team was used to forecast if we can trace specific trends (Veterans Day and other HHUSA events) using the OLS model.



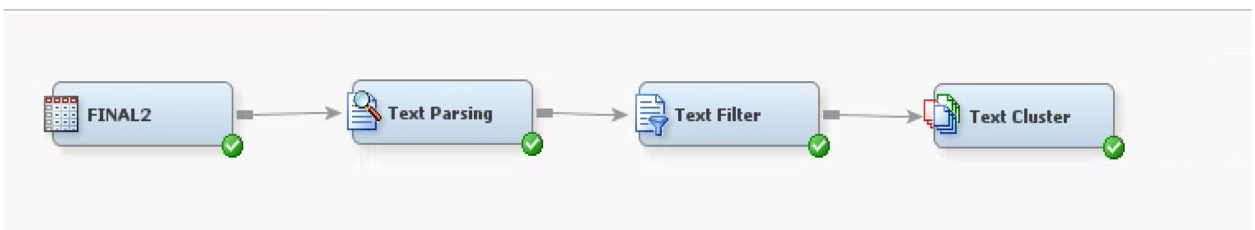
Note: Veterans Day (11/11); Salute to Veterans Golf Classic Income (11/3)

The OLS model can quantify marginal value based on page activity. For instance, 8/22 generated significant activity following the 100 Holes for our Heroes Event and possibly lead to the $60,000 donations received in August of that year and the $133,112 and $182,511 in corporate donation in the following month. No significant events or activity took place on Veterans Day therefore a negative impact to donation revenue (i.e. loss opportunity). The 11/3 Salute to Veterans Golf Classic received significant Facebook activity leading up to and on the day of the event. Therefore, the model predicts a positive impact to donation revenue due to strong page activity and exposure.

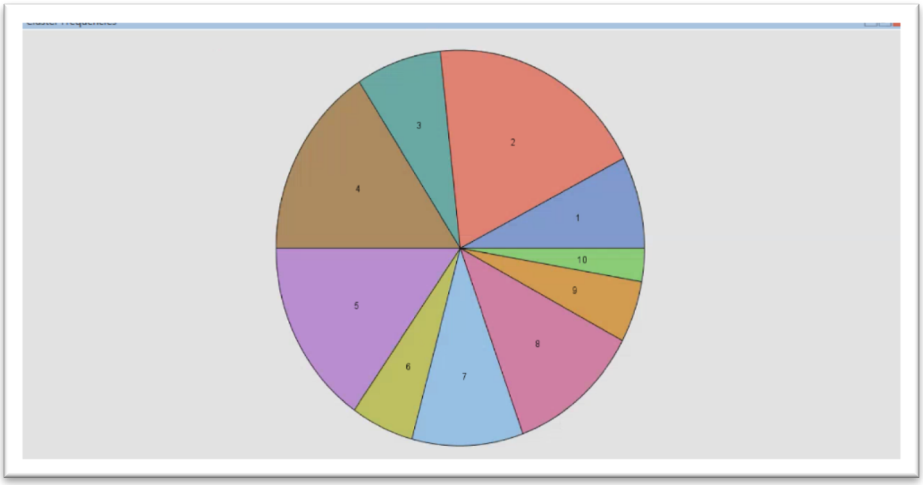


The only other significant event that the model did not predict to have a material impact to donation revenue was MedAssets 5K Race for Heroes and the WWE 10% donation of every dollar spent on WWEShop.com. The events in Plano and Alpharetta raised $165,000 to go towards our programs and services and WWE donation was not readily determinable.

In addition, we also did text mining analysis in order to support our trend results. The ideas of that is to find evidence of most frequent terms that may be related to our finding in the trend. I used SAS Enterprise Miner tool to do this part



we started by importing data that related to the trend started to 2013. We used text parsing, text editing and text clustering. We used IDF as term weight method and set cluster to 10 but we kept other parameters as default values. Here is our results:



Clusters

Cluster ID	Descriptive Terms	Frequency	Percentage	Coordinate 1	Coordinate 2	Coordinate 3	Coordinate 4
1	est bundle ps4 vetnet black iii ops +bid +'trade card' +prize +trade google duty card 3pm ...	54	7%	0.157013	0.257253	0.025917	-0.0093
2	transition uso veteran +event +day +serve army +donation country next workshop +career +center +family emp...	139	19%	0.278085	-0.03584	0.017194	-0.0987
3	+donate +purchase aweamess cole kenneth bodybuilding.com menå wearhouse label +shop site +sale mem...	55	8%	0.391779	-0.02224	-0.29701	-0.0206
4	race medassets marine marathon corps team +photo +runner air weekend +great happy heroes +run hire ...	116	16%	0.236786	0.02493	-0.00582	-0.1244
5	+transition +spouse +workshop +'transition workshop' nov 17-18 8th +coach free +register civilian +service +'s...	109	15%	0.3842	-0.1337	0.21239	0.11808
6	quot +attempt +hole meyer july +golf carl 4th +raise +fund awareness +reason +run +effort +open ...	42	6%	0.279123	0.044724	-0.05843	-0.1704
7	brian stann ceo golf president classic +american salute 'annual salute' annual 7th november veterans quot +...	70	10%	0.343828	0.091495	-0.00409	-0.2028
8	+veteran crisis line month +word tuesday +spread +job support +show linkedin +help +know +tip +hire ...	85	12%	0.266381	-0.03037	-0.08744	-0.0935
9	corporate +'corporate partner' partner hireheroesusa hire services +dedicate usa +position heroes managemen...	35	5%	0.228208	0.0362	-0.09954	-0.1243
10	blog +week post 'week's blog' search job +post +reason services +calendar carl +prepare meyer +job +center	21	3%	0.136462	-0.00973	0.022464	-0.0870

in our results, we found some terms that could be linked to the trend. We focused on the three most frequent clusters, which are 2, 4 and 5. We found some evidence terms that are related to our trend. For example, by looking to second cluster, we have some terms like *events*, *donation, workshops* and others like the names of some events and workshops. Also, clusters 4 and 5 can include similar terms that are also related to the trend we have in forecasting analysis. In contrast, clusters 9 and 10 showed lowest frequent values. By seeing these clusters, they were not related to our findings.

7.0 CAUSALITY TEST

To test the hypotheses that a cause-effect relationship between social media messaging/online presence and actual annual donations exist, we will use a Granger causality test. The Granger causality test is a statistical hypothesis test for determining whether one-time series is useful in forecasting another. It should be noted that the Granger test finds only predictive causality.

A time series of social media messaging/online presence will be tested to assert Granger-cause annual donations to increase. This is through a series of F-tests on lagged values of our dependent variables (and with lagged values of the independent variable also included), that our regressors provide statistically significant information about future values of donations.

The test is undertaken as

$$y_t = \alpha + \sum_{j}^{k} \beta_j y_{t-j} + \sum_{j}^{k} \delta_j x_{t-j} + \epsilon_t$$

```
/*Granger Causality test*/
proc varmax data=modmasterfile01;
model VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115 giftrev / p=6;
causal group1=(VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115)
group2=(giftrev);
causal group1=(giftrev) group2=(VAR0003 VAR0011 VAR0026 VAR0045 VAR0048
VAR0072 VAR0115);
run;
```

The test is an F-test on the δ's being jointly equal to zero. If we reject the null hypothesis then X is said to "Granger-cause" Y. Our results indicate that social media messaging/online presence Ganger-cause an increase in donation revenue.

Granger-Causality Wald Test			
Test	DF	Chi-Square	Pr > ChiSq
1	42	31.09	0.8923
2	42	61.63	0.0257

Test 1: Group 1 Variables:	VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115
Group 2 Variables:	giftrev

Test 2: Group 1 Variables:	giftrev
Group 2 Variables:	VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115

Based on the aforementioned output we can assert that online presence granger-cause donations.

8.0 CONCLUSION

In summary, our research has indicated that HHUSA's online presence has contribute to its overall mission and financial achievements. We found evidence for registrants since December 18th was most likely influenced by direct ad campaigns based on the results from text mining. We showed a benefit relating to branding awareness via social media was strongly correlated with registration for HHUSA services. We traced specific trends to certain days and events based own our analysis using an OLS estimator. Our research also demonstrated a historical cause-effect relationship between social media messaging/online presence and actual annual donations. And we measured the effects on fundraising over the years from the brand awareness that is created by having a website and social media. To further demonstrate the causal effect on having an online presence, we performed a causality test to support the assertion that the Facebook campaign has a direct impact on donations. Based on our accumulated research and results obtained, we believe the Facebook Registration Campaign has yielded a predictive probability for success.

```
proc sql;
create table fbd07 as
select * from fbd071
join fbd072
on fbd071.var0001=fbd072.var0001;
quit;

proc sql;
create table fbd08 as
select * from fbd081
join fbd082
on fbd081.var0001=fbd082.var0001;
quit;

proc sql;
create table fbd09 as
select * from fbd091
join fbd092
on fbd091.var0001=fbd092.var0001;
quit;

proc sql;
create table fbd10 as
select * from fbd101
join fbd102
on fbd101.var0001=fbd102.var0001
join fbd103
on fbd102.var0001=fbd103.var0001;
quit;

proc sql;
create table modfbd1 as
select * from fbd01
union
select * from fbd06;
quit;

proc sql;
create table modfbd2 as
select * from fbd07
union
select * from fbd08;
quit;

proc sql;
create table modfbd3 as
select * from fbd09
union
select * from fbd10;
quit;

proc sql;
create table fbmasterfile as
select * from modfbd1
union
select * from modfbd2
union
```

```
select * from modfbd3;
quit;

data fbmasterfile;
set fbmasterfile;
array a(*) _numeric_;
do i=1 to dim(a);
if a(i) = . then a(i) = 0;
end;
drop i;

proc sql;
create table teststat as
select
var0001,var0002,var0003,var0030,var0048,var0051,var0060,var0063,var0115
from fbmasterfile
where var0001>='01jan2014'd;
quit;

proc sgplot data=teststat;
series x=var0001 y=var0003;
series x=var0001 y=var0030;
series x=var0001 y=var0048;
series x=var0001 y=var0051;
series x=var0001 y=var0060;
series x=var0001 y=var0063;
series x=var0001 y=var0115;
run;

proc sql;
create table teststat2 as
select
var0001,char0010,char0011,char0012,char0013,char0014,char0015,char0016,char00
17,char0058,char0059,char0060,char0061
from fbd07
union
select
var0001,char0010,char0011,char0012,char0013,char0014,char0015,char0016,char00
17,char0058,char0059,char0060,char0061
from fbd08
union
select
var0001,char0010,char0011,char0012,char0013,char0014,char0015,char0016,char00
17,char0058,char0059,char0060,char0061
from fbd09
union
select
var0001,char0010,char0011,char0012,char0013,char0014,char0015,char0016,char00
17,char0058,char0059,char0060,char0061
from fbd10
where var0001>='01jan2014'd;
quit;

proc sgplot data=teststat2;
series x=var0001 y=char0010;
series x=var0001 y=char0011;
series x=var0001 y=char0012;
```

```
series x=var0001 y=char0013;
series x=var0001 y=char0014;
series x=var0001 y=char0015;
series x=var0001 y=char0016;
series x=var0001 y=char0017;
series x=var0001 y=char0058;
series x=var0001 y=char0059;
series x=var0001 y=char0060;
series x=var0001 y=char0061;
run;

proc sql;
create table teststat3 as
select date as var0001, amount
from donordata
where var0001>='01jan2014'd;
quit;

proc sgplot data=teststat3;
series x=var0001 y=amount;
run;

proc sql;
create table individualdonation as
select date as var0001, amount, gl_code
from donordata
where var0001>='01jan2014'd and gl_code='Individual Donation';
quit;

proc sgplot data=individualdonation;
series x=var0001 y=amount;
run;

proc sql;
create table modfbdata as
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
```

```
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
from fbd06
union
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
```

```
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
from fbd07
union
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
```

```
from fbd08
union
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
from fbd09
union
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
```

```
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
from fbd10;
quit;

proc sql;
create table modfbdata2 as
select VAR0001,VAR0003,
VAR0008,
VAR0011,
VAR0030,
VAR0029,
VAR0017,
VAR0020,
VAR0023,
VAR0026,
VAR0028,
VAR0033,
VAR0036,
VAR0039,
VAR0045,
VAR0048,
VAR0051,
VAR0060,
```

```
VAR0063,
VAR0072,
VAR0115,
CHAR0010,
CHAR0011,
CHAR0012,
CHAR0013,
CHAR0014,
CHAR0015,
CHAR0016,
CHAR0017,
CHAR0034,
CHAR0035,
CHAR0036,
CHAR0037,
CHAR0038,
CHAR0039,
CHAR0040,
CHAR0041,
CHAR0058,
CHAR0059,
CHAR0060,
CHAR0061,
CHAR0086,
CHAR0087,
CHAR0088,
CHAR0089
from modfbdata
where var0001>='03jan2014'd;
quit;

proc means data=modfbdata2;
var VAR0003 VAR0008     VAR0011     VAR0030     VAR0029     VAR0017
      VAR0020     VAR0023     VAR0026     VAR0028     VAR0033     VAR0036
      VAR0039     VAR0045     VAR0048     VAR0051     VAR0060     VAR0063
      VAR0072     VAR0115     CHAR0010    CHAR0011    CHAR0012    CHAR0013
      CHAR0014    CHAR0015    CHAR0016    CHAR0017    CHAR0034    CHAR0035
      CHAR0036    CHAR0037    CHAR0038    CHAR0039    CHAR0040    CHAR0041
      CHAR0058    CHAR0059    CHAR0060    CHAR0061    CHAR0086    CHAR0087
      CHAR0088    CHAR0089;
run;

proc sql;
create table modfbdata3 as
select * from modfbdata2 m
join teststat3 t on t.var0001=m.var0001;
quit;

proc reg data=modfbdata3;
model amount=VAR0003 VAR0008 VAR0011 VAR0030 VAR0029 VAR0017 VAR0020 VAR0023
VAR0026 VAR0028 VAR0033 VAR0036 VAR0039 VAR0045 VAR0048 VAR0051 VAR0060
VAR0063 VAR0072 VAR0115 CHAR0010 CHAR0011 CHAR0012    CHAR0013 CHAR0014
CHAR0015 CHAR0016 CHAR0017 CHAR0034 CHAR0035 CHAR0036 CHAR0037 CHAR0038
CHAR0039 CHAR0040 CHAR0041 CHAR0058 CHAR0059 CHAR0060 CHAR0061 CHAR0086
CHAR0087 CHAR0088 CHAR0089;
output out=modfbdata4 p=yhat;
run;
```

```
proc reg data=modfbdata3;
model amount=VAR0003 VAR0029 VAR0026 VAR0028 VAR0048 VAR0051 VAR0060 VAR0063
VAR0072 VAR0115 CHAR0013 CHAR0017 CHAR0034 CHAR0035 CHAR0036 CHAR0037
CHAR0041;
output out=modfbdata5 p=yhat;
run;

proc sql;
create table teststat5 as
select sum(amount) as dailytotal, var0001
from teststat3
group by var0001;
quit;

proc expand data=teststat5 out=teststat4 from=day to=day;
id var0001;
convert dailytotal;
run;

proc sql;
create table modfbdata3 as
select * from modfbdata2 m
join teststat4 t on t.var0001=m.var0001;
quit;

proc reg data=modfbdata3;
model dailytotal=VAR0003 VAR0008 VAR0011 VAR0029 VAR0026 VAR0028 VAR0039
VAR0045 VAR0048 VAR0051 VAR0060 VAR0063 VAR0072 VAR0115 CHAR0010 CHAR0011
CHAR0012    CHAR0013 CHAR0014 CHAR0015 CHAR0016 CHAR0017 CHAR0034 CHAR0035
CHAR0036    CHAR0037 CHAR0038 CHAR0039 CHAR0040 CHAR0041/ vif;
output out=modfbdata4 p=yhat;
run;

proc reg data=modfbdata3;
model dailytotal=VAR0003 VAR0011 VAR0028 VAR0045 VAR0048 VAR0072 VAR0115 /
vif;
output out=modfbdata4 p=yhat;
run;

proc reg data=modfbdata3;
model dailytotal=CHAR0034 CHAR0035 CHAR0036     CHAR0037 CHAR0038 CHAR0039
CHAR0040    CHAR0041 /vif;
output out=modfbdata4 p=yhat;
run;

proc sql;
create table masterfile2 as
select m.var0001, var0002 from  fbmasterfile fb
join modfbdata3 m on fb.var0001=m.var0001;
quit;

proc sql;
create table masterfile as
select * from modfbdata3 m
join masterfile2 m2 on m2.var0001=m.var0001;
quit;
```

```
data masterfile;
set masterfile;
array a(*) _numeric_;
do i=1 to dim(a);
if a(i) = . then a(i) = 0;
end;
drop i;

data modmasterfile; set masterfile;
lngiftrev=log(dailytotal);
lnvar0002=log(var0002);
lnvar0011=log(VAR0011);
lnvar0045=log(VAR0045);
lnvar0048=log(VAR0048);
lnvar0072=log(VAR0072);
lnvar0115=log(var0115);
expvar0011=var0011**2;
expvar0048=var0048**2;
expvar0072=var0072**2;
run;

data modmasterfile;
set modmasterfile;
array a(*) _numeric_;
do i=1 to dim(a);
if a(i) = . then a(i) = 0;
end;
drop i;

proc reg data=modmasterfile;
model lngiftrev=lnvar0002 var0003 var0011 var0072 / vif;
output out=modmasterfile2 p=yhat;
run;

proc model data=modmasterfile;
parms a0 b1 b2 b3 b4;
lngiftrev=a0+b1*lnvar0002+b2*var0003+b3*var0011+b4*var0072;
fit lngiftrev / gmm kernel=(bart,15,0);
run;

proc sgplot data=modmasterfile;
series x=var0001 y=dailytotal;
run;

proc sql;
create table teststat6 as
select sum(amount) as giftrev, var0001
from nongrantgift
group by var0001;
quit;

proc expand data=teststat6 out=teststat7 to=day plots=(input output);
id var0001;
convert giftrev / observed=(beginning,average);
run;
```

```
proc sql;
create table modfbdata4 as
select * from modfbdata2 m
join teststat7 t on t.var0001=m.var0001;
quit;

proc sql;
create table masterfile3 as
select m.var0001, var0002 from  fbmasterfile fb
join modfbdata4 m on fb.var0001=m.var0001;
quit;

proc sql;
create table masterfile01 as
select * from modfbdata4 m
join masterfile3 m3 on m3.var0001=m.var0001;
quit;

data modmasterfile01; set masterfile01;
lngiftrev=log(giftrev);
lnvar0002=log(var0002);
lnvar0003=log(var0003);
lnvar0011=log(VAR0011);
lnvar0045=log(VAR0045);
lnvar0048=log(VAR0048);
lnvar0063=log(var0063);
lnvar0072=log(VAR0072);
lnvar0115=log(var0115);
run;

data modmasterfile01;
set modmasterfile01;
array a(*) _numeric_;
do i=1 to dim(a);
if a(i) = . then a(i) = 0;
end;
drop i;

proc reg data=modmasterfile01;
model giftrev=VAR0003 VAR0011 VAR0045 VAR0048 VAR0063 VAR0072 VAR0115 / vif;
output out=modmasterfile3 p=yhat;
run;

proc model data=modmasterfile01;
parms a0 b1 b2 b3 b4 b5 b6 b7;
giftrev=a0+b1*var0003+b2*var0011+b3*var0026+b4*var0045+b5*var0048+b6*var0072+
b7*var0115;
fit giftrev / gmm kernel=(bart,1,0);
run;

proc varmax data=modmasterfile01;
model VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115 giftrev / p=6;
causal group1=(VAR0003 VAR0011 VAR0026 VAR0045 VAR0048 VAR0072 VAR0115)
group2=(giftrev);
causal group1=(giftrev) group2=(VAR0003 VAR0011 VAR0026
```