Lead Quality Analysis of a Marketing Campaign

Arvind Ramkumar
Email – aramkum4@asu.edu | arvindramkumar1996@gmail.com
Phone – 480(859)-2393

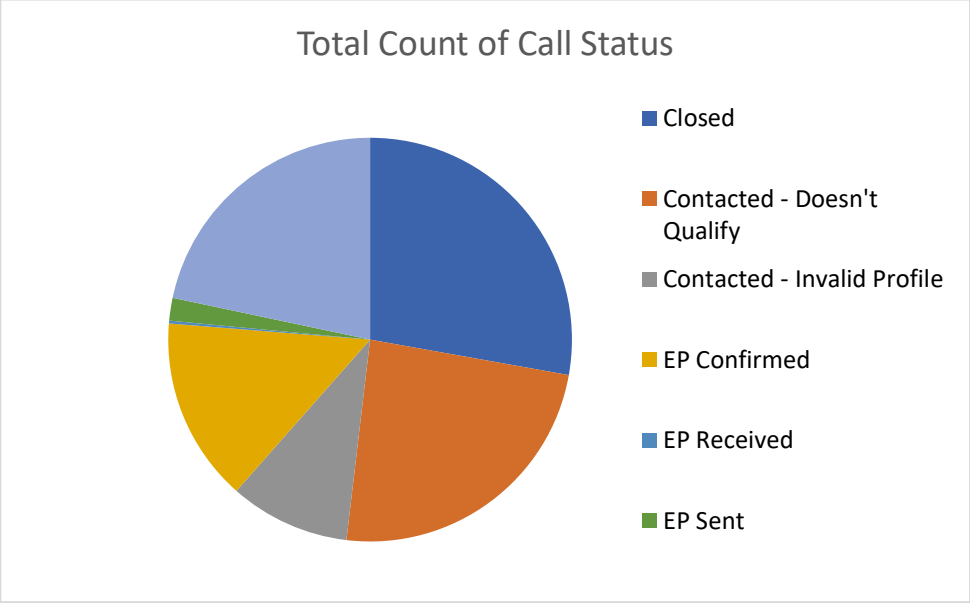
Table of Contents

Variation of Call Status

The Call Status is the most important factor in the entire dataset, and the most important outcome of this data is that the lead (Consumer) should eventually end up buying out product.

So, we first analyze the spread of Call Status. All the blanks are replaced with “Unknown”.

In all the graphs that are produced below is devoid of this category “Unknown”, because that occupies the majority of the graph (Relatively too many unknown), and it makes us difficult to read the graph. The entire interactive dashboard in given in the excel attached.

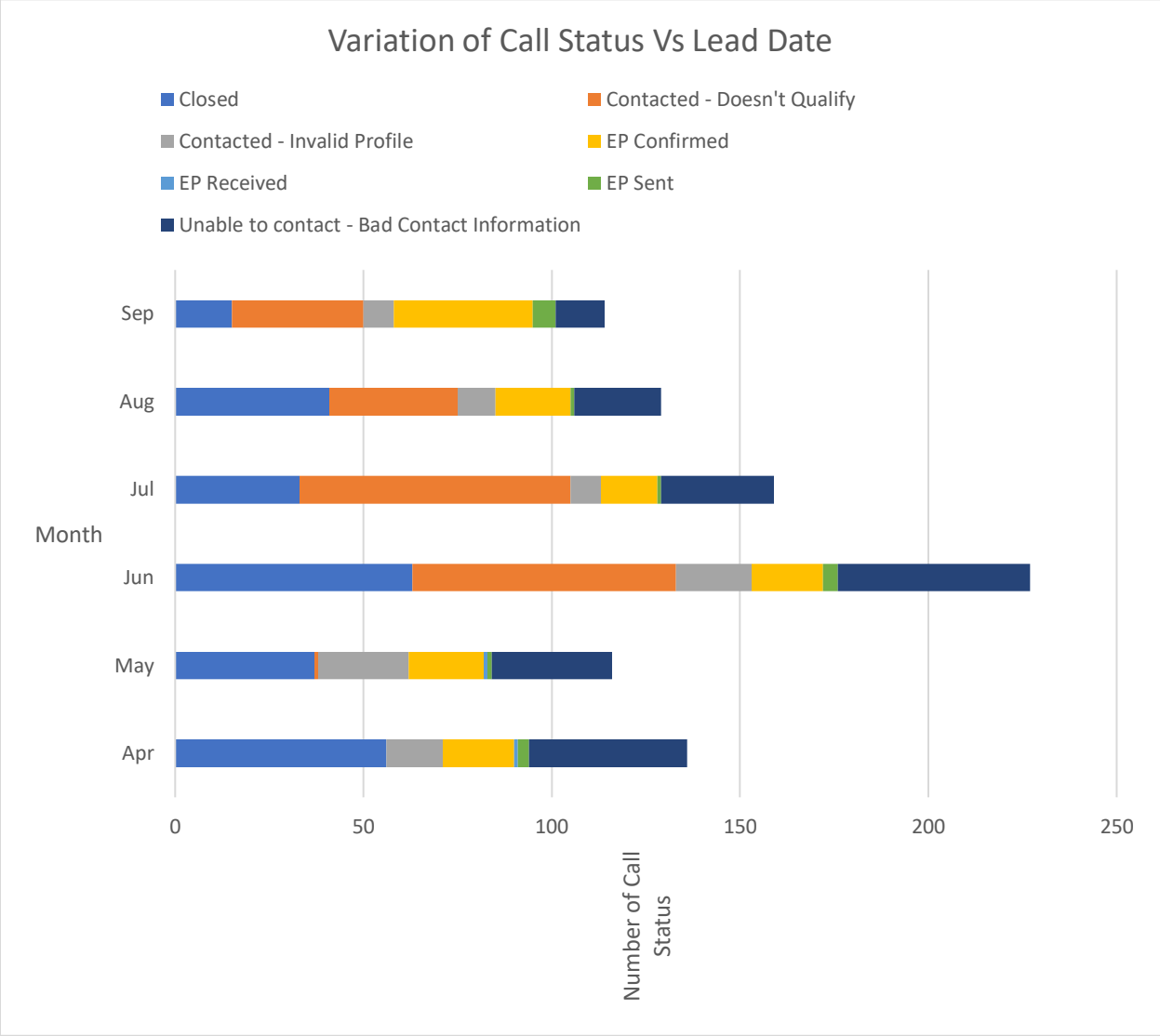


From the above pie chart, we see that there is more than 25% of the leads (Excluding Unknown) is Closed and we need to understand these leads which were closed to understand how to convert all the leads to Closed (Making the consumer to buy the product)

There are few features that seem to be more important factor than any other and they are listed below and are analyzed one by one on how it has affected the Call Status. A detailed interactive Dashboard is also given in the excel attached.

1. Lead Date
2. Publisher Zone Name
3. Widget Name
4. Address Score
5. Phone Score
6. Debt Level
7. Advertiser Campaign Name
8. Partner
9. Landing URL

Lead Date

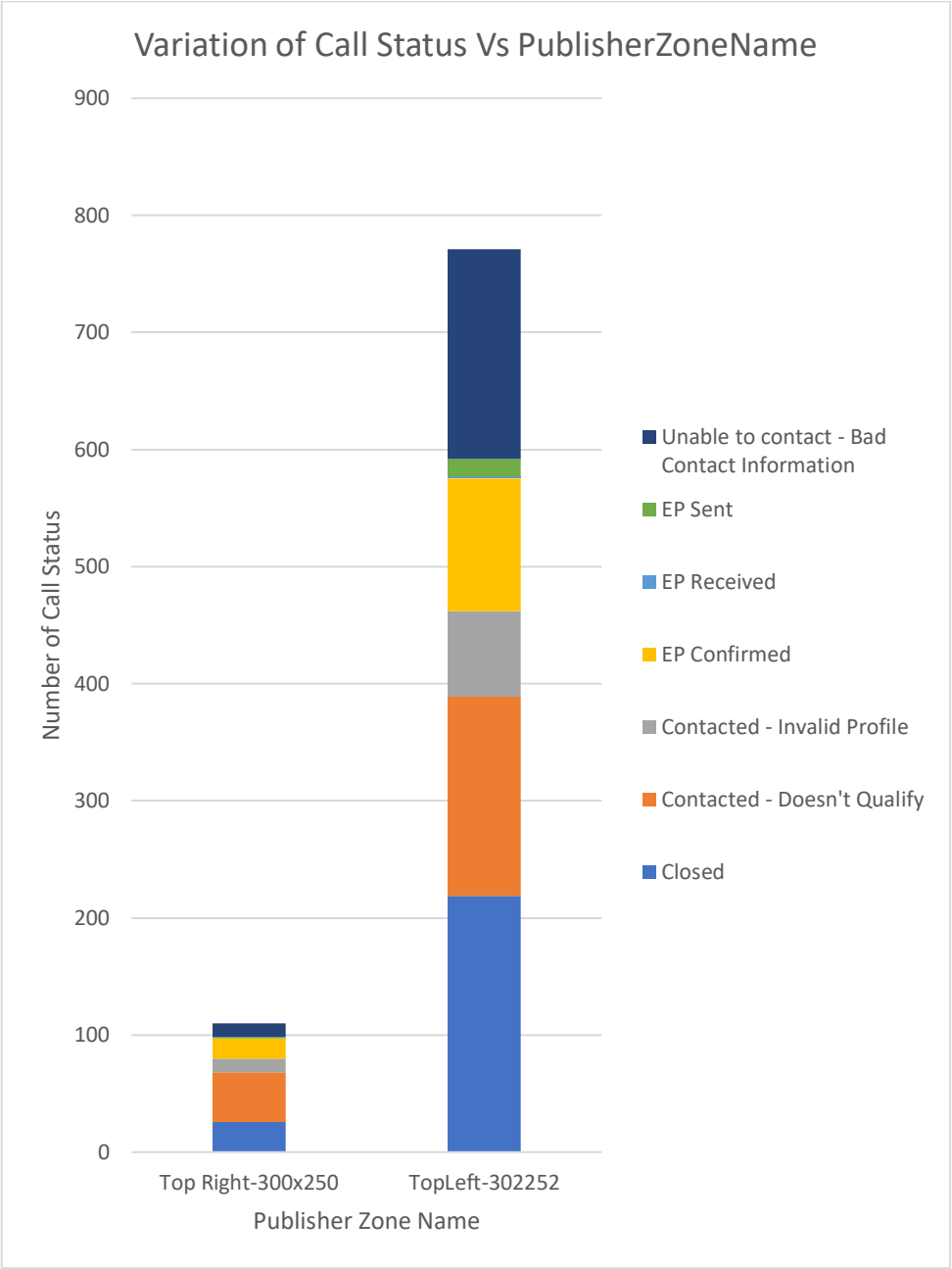


The number of call statuses given during June is double than what we have given in April, but the number of Call Status that has been converted to Count is almost the same for both the months and these are the two months that had a greater number of closed statuses.

The conversion rate to closed is very high in the beginning (April) but with time, the conversion rate has dropped, and it is the least in September. But if we look at the number of unknown, September has least number of unknows, which could be an indication that the leads have improved in quality and decreased in quantity over time. We can understand this better in the section where we calculate the Lead time.

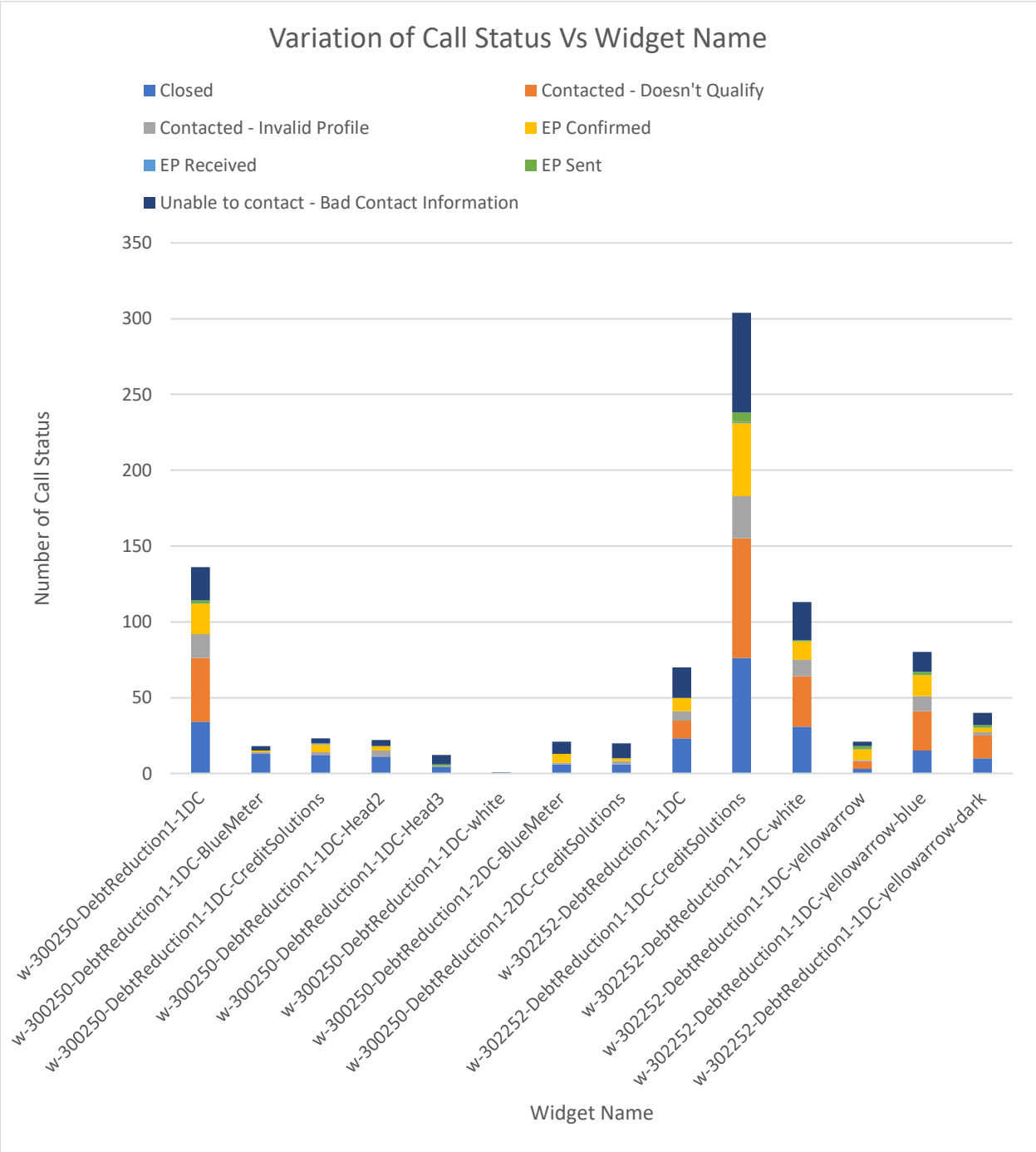
So, the month in which the lead is produced will affect the quality of the leads.

Publisher Zone Name



Here, although we see that, the total number of calls seems to be ridiculously high for the publisher zone name TopLeft – 302252, the conversion rate to closed is almost the same for both (9.59% and 7.96%), but the TopLeft – 302252 performs a little better in terms on Unknown, it has a relatively less Unknown than the other. But this doesn't seem to have a statistical significance, so might not be amongst one of the best measures to obtain lead quality.

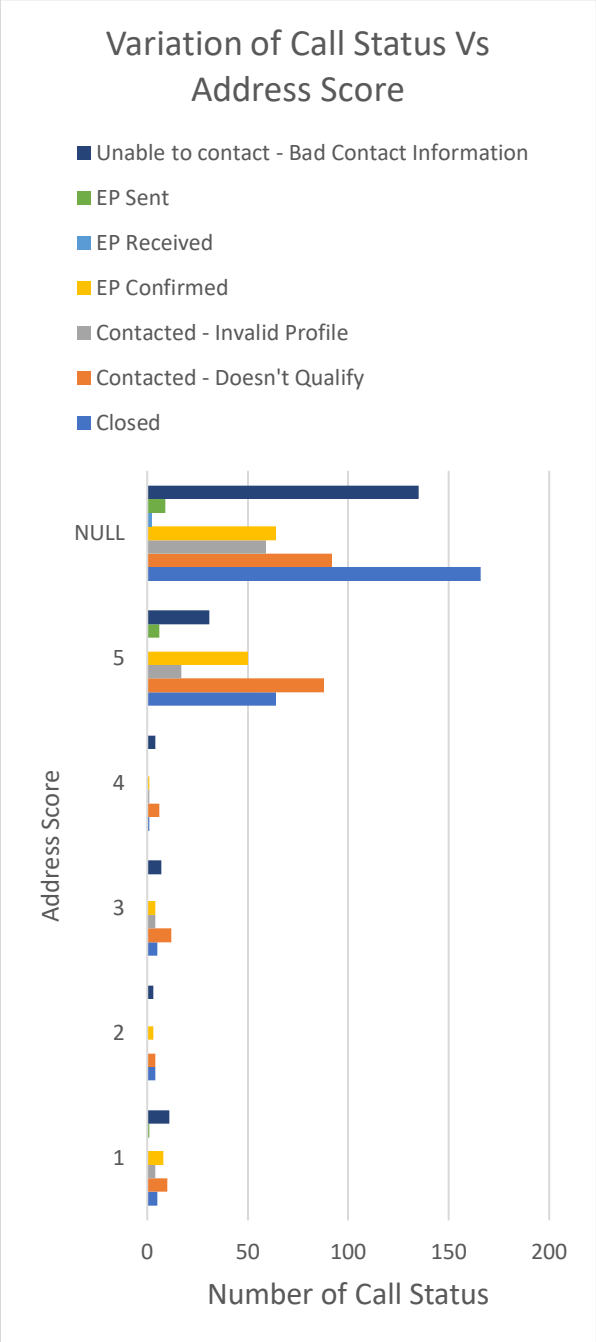
Widget Name

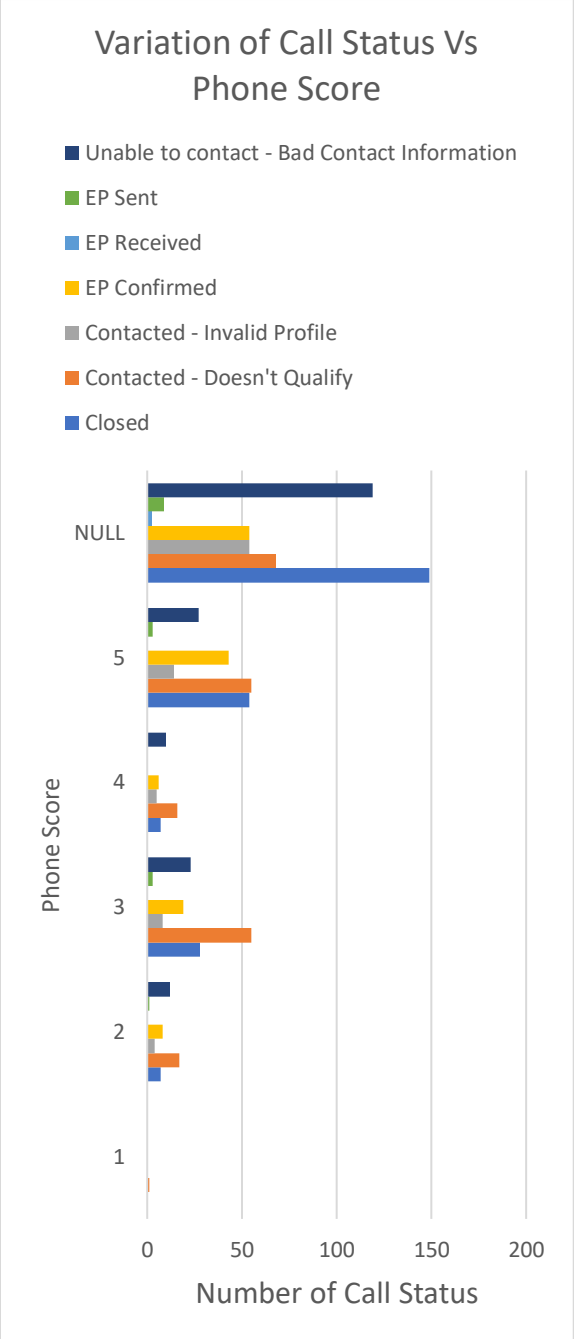


The problem here is the number of call status for w-300250-DebtReduction1-2DC-CreditSolutions is ridiculously high, so correspondingly the closed is also high. We need to better understand based on the other call status in each widget as well.

Because of the vast differences, and this is what the customer sees, this will play a major role in determining the quality of the lead.

Address Score and Phone Score

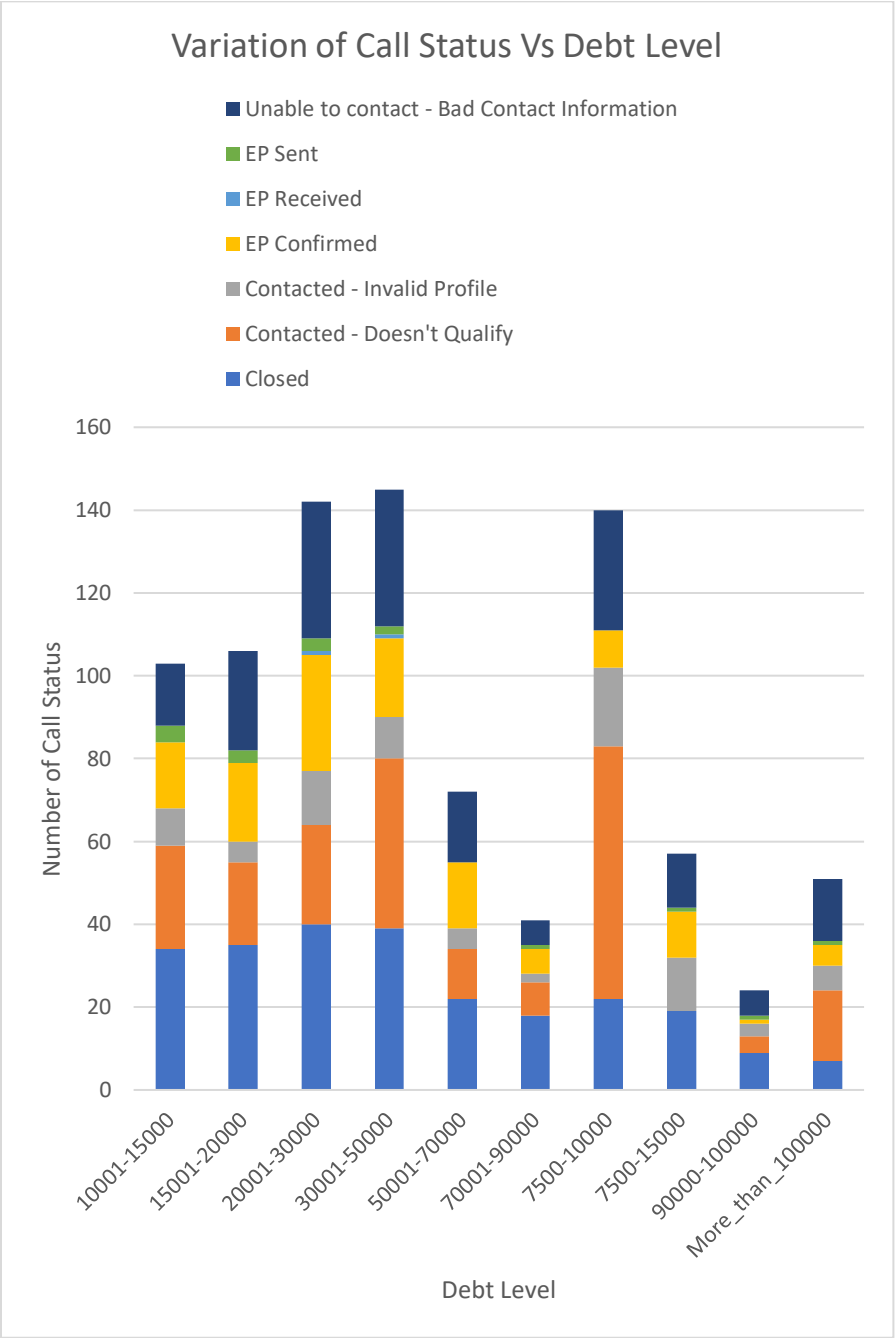




The Address Score and the Phone Score is one of the most viable parameters to determine the quality lead, in spite of the lack of information for so many leads. All the prospect customers are likely to give us the correct details rather than consumers who have no motivation to purchase. It is evident from the above table that the conversion rate to closed is high if the Score is 5 (Perfect Match)

So, these scores will play a vital role in determining the Quality of the lead.

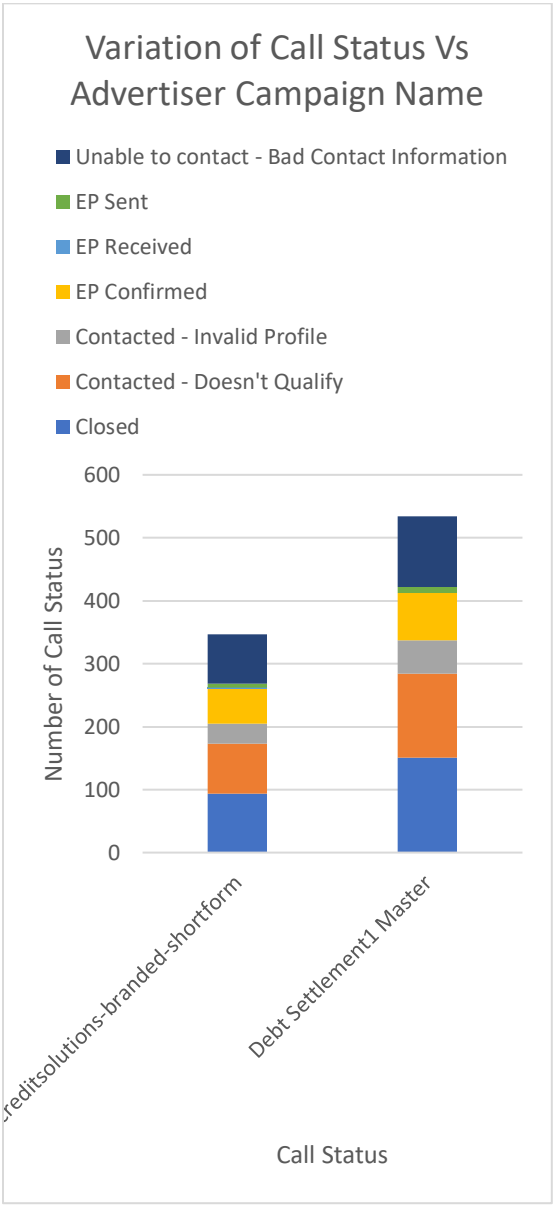
Debt Level

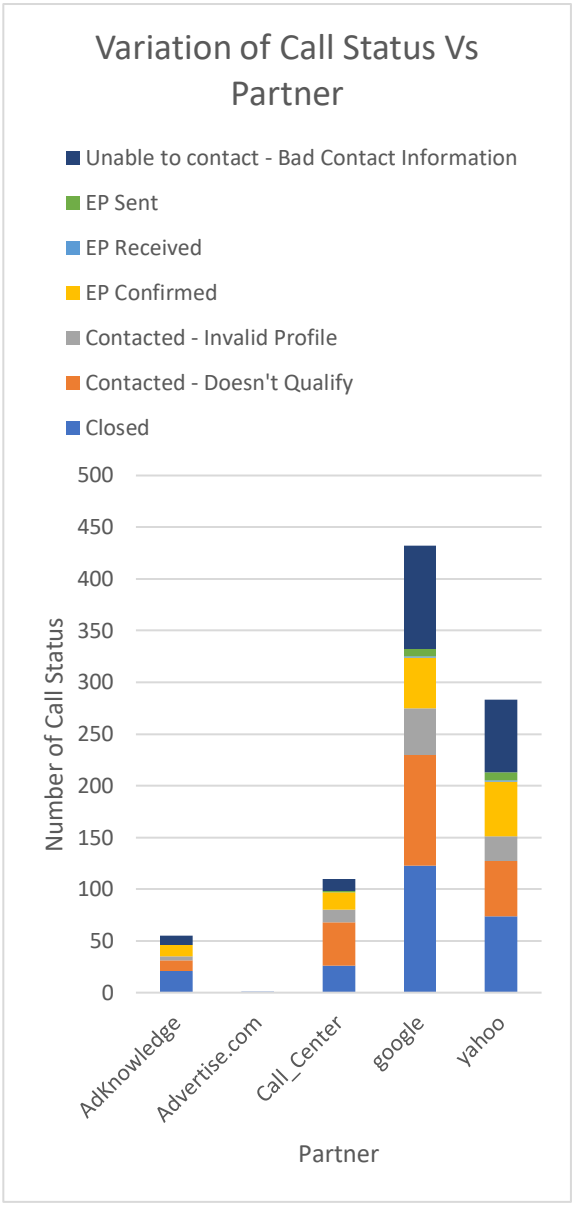


The entire dataset is related to debt and with the given understanding of the data, it is related to solve the debt crisis. So, the debt level of the visitor is one of the most important factors to consider in understanding the prospects of the visitor turning into a consumer.

So, to calculate the lead quality, this is one of the most important criteria to consider

Advertiser Campaign Name and Partner

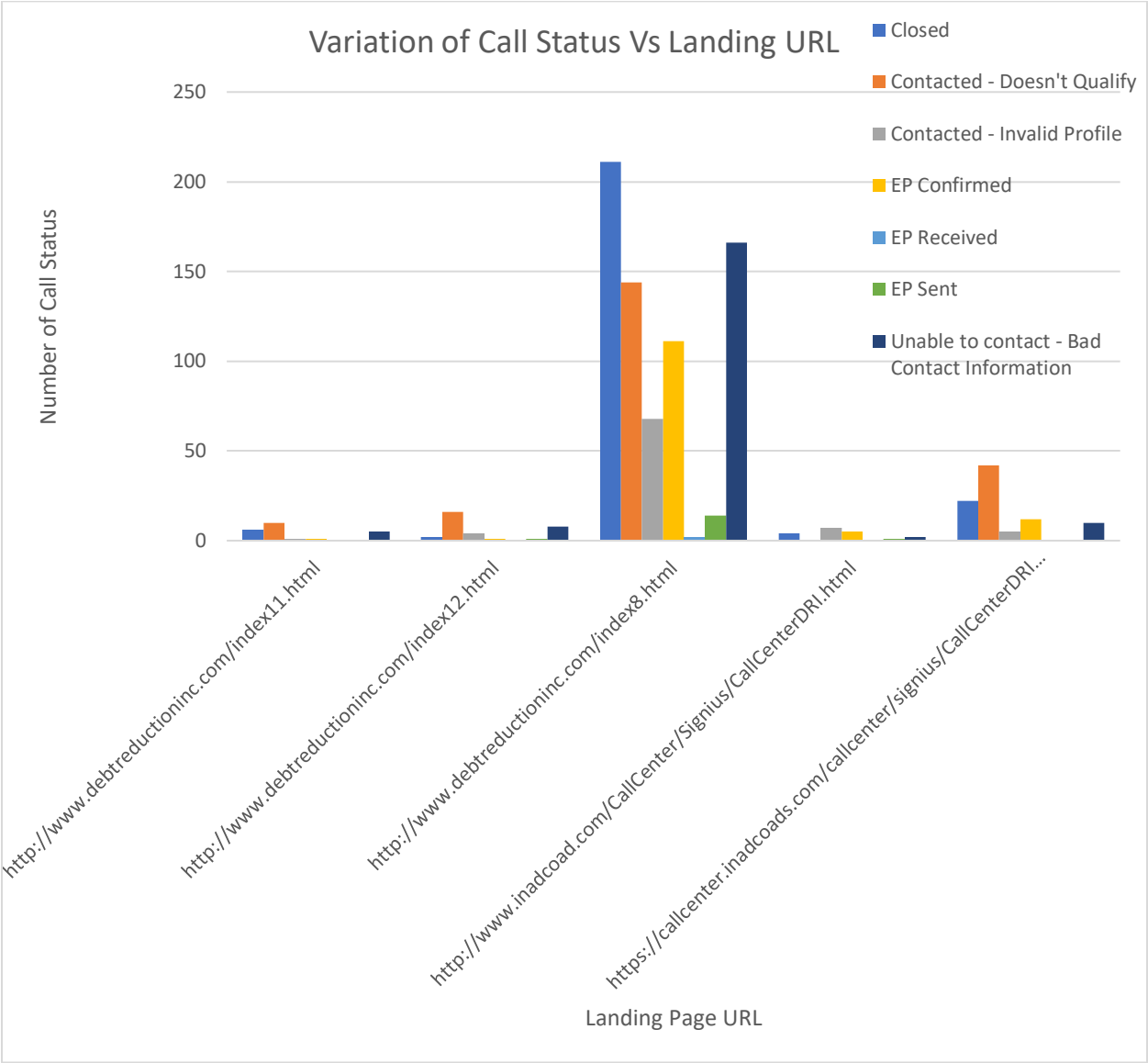




The call status variation over the Advertiser Campaign doesn't seem to affect the conversion rate to Closed. So, it is highly unlikely that the Advertiser Campaign will affect the Quality of the lead.

But, from the graph Call Status Vs Partner has a wide variation based on the conversion rate. So it is highly likely that the Partner has a very serious impact on the quality of the lead.

Landing Page URL



The landing page URL, the page in which the visitor lands will have higher impact in the prospects of the visitor converting to Closed. So, we need to understand and incorporate the Landing Page URL to estimate the Quality of the lead.

Estimate Quality of Lead

- The Lead Quality is a factor that allows us to identify the how likely are the leads are to become a paying customer. The higher the lead quality, the more likely the lead become the customer.

- There is no specific calculation to estimate the quality of the lead, but universally accepted method to calculate the Quality of the lead is the BANT Approach.

- B – Budget
- A – Authority
- N – Need
- T – Timing

- But with the given data, it is not possible to estimate these parameters individually, so we use a basic but effective way to obtain the Quality of the lead.

- **Step 1:** Decide the parameters that will affect the lead quality and these parameters will be used to calculate the lead quality. These parameters are obtained from the intuition gathered from above (From the Call Status Variations Graph).

- **Step 2:** For each parameter chosen, we rank based on the Call Status conversion rate.

    - Closed, EP Sent, EP Received, EP Confirmed – Higher percentage is desired.
    - Unknown, contacted – Doesn't Qualify, contacted – Invalid Profile, Unable to contact - Bad Contact Information –Low percentage is better.

    - For Example, consider the month parameter (April – September), We need to have a high Closed Status. So, the month with high percentage of closed will have Rank 1, and this is the case for EP Sent, EP Received and EP Confirmed as well.
    - On the contrary, the Unknown, contacted – Doesn't Qualify, contacted – Invalid Profile, Unable to contact - Bad Contact Information, lower percentage is preferred. So, the month with the least value is given Rank 1.

- **Step 3:** After giving Rank individually for each call status, across every category in each parameter, we now take the average over all the call status over every category in each parameter.

    - In the example, for every month, the 8 different ranks obtained from 8 different call status, we take average, and the minimum most average rank has top Preference in terms of lead quality. Here, September has the least average rank and they are top in the order and July has high average rank, so they are last in the order. We do this for every parameter.

Example Explained to determine the Desirable order of the categories in the parameter

Count of CallStatus	CallStatus							
Months	Closed	Contacted - Doesn't Qualify	Contacted - Invalid Profile	EP Confirmed	EP Received	EP Sent	Unable to contact - Bad Contact Information	Unknown
Apr	10.81%	0.00%	2.90%	3.67%	0.19%	0.58%	8.11%	73.75%
May	6.38%	0.17%	4.14%	3.45%	0.17%	0.17%	5.52%	80.00%
Jun	10.34%	11.49%	3.28%	3.12%	0.00%	0.66%	8.37%	62.73%
Jul	6.20%	13.53%	1.50%	2.82%	0.00%	0.19%	5.64%	70.11%
Aug	9.36%	7.76%	2.28%	4.57%	0.00%	0.23%	5.25%	70.55%
Sep	4.36%	10.17%	2.33%	10.76%	0.00%	1.74%	3.78%	66.86%

The first value (10.81%) against the Month April and the Call Status Closed is obtained from the (Number of Calls in Closed in April / Number of Calls in April). On a similar basis, all the other values are calculated.

Now, we know that

- Closed, EP Sent, EP Received, EP Confirmed – Higher percentage is desired.
- Unknown, contacted – Doesn't Qualify, contacted – Invalid Profile, Unable to contact - Bad Contact Information –Low percentage is better.

So, the desirable order with respect to the "Closed" status is April (Rank 1 – 10.81%) – June (Rank 2 – 10.34%) – Aug (Rank 3 – 9.36%) – May (Rank 4 – 6.38%) – July (Rank 5 – 6.20%) – September (Rank 6 – 4.36%) (Check the table below). This is similar for EP Sent, EP Received, EP Confirmed as well.

But considering just these, wouldn't make sense. So, we give ranking for the other category as well for which least percentage will have least rank. Consider for "Unknown", the desirable order is June (Rank 1 – 62.73%) – September (Rank 2 – 66.86%) – July (Rank 3 – 70.11%) – August (Rank 4 – 70.55%) – April (Rank 5 – 73.75%) – May (Rank 6 – 80%). This is similar to other cases contacted – Doesn't Qualify, contacted – Invalid Profile, Unable to contact - Bad Contact Information as well.

This method is performed on all the parameters that we are interested in study.

Count of CallStatus										
Months	Closed	Contacted - Doesn't Qualify	Contacted - Invalid Profile	EP Confirmed	EP Received	EP Sent	Unable to contact - Bad Contact Information	Unknown	Average	Order
Apr	1	1	4	3	1	3	5	5	2.875	2
May	4	2	6	4	2	6	3	6	4.125	5
Jun	2	5	5	5	3	2	6	1	3.625	4
Jul	5	6	1	6	3	5	4	3	4.125	5
Aug	3	3	2	2	3	4	2	4	2.875	2
Sep	6	4	3	1	3	1	1	2	2.625	1

After giving the individual ranks, we need to summarize them together and we have ensured that all the desirable months are given least rank. So, we average the ranks across the months and from the average, the month with the least rank is the highly desirable prospect of converting it into a closed and as well minimizing all the nuisance status.

So, this highly desirable Month will have the high lead quality, and here the order of desirability is – September, April, August, June, May, July.

This is done for all the parameters and the desirable order for every parameter is given below in order.

Desirable Order

Month	Widget Name	Address Score	Phone Score	Debt Level	Partner	Landing Page URL
September	w-300250-DebtReduction1-1DC	2	5	70001-90000	Advertise.com	https://callcenter.inadcoads.com/callcenter/signius/CallCenterDRI.html
April	w-300250-DebtReduction1-1DC-Head2	5	3	10001-15000	AdKnowledge	http://www.debtreductioninc.com/index8.html
Aug	w-300250-DebtReduction1-2DC-CreditSolutions	NULL	NULL	15001-20000	yahoo	http://www.inadcoad.com/CallCenter/Signius/CallCenterDRI.html
June	w-302252-DebtReduction1-1DC-CreditSolutions	1	1	20001-30000	Call_Center	http://www.debtreductioninc.com/index11.html
July	w-300250-DebtReduction1-2DC-BlueMeter	3	2	50001-70000	google	http://www.debtreductioninc.com/index12.html
May	w-302252-DebtReduction1-1DC	4	4	90000-100000		
	w-302252-DebtReduction1-1DC-yellowarrow-blue			30001-50000		
	w-300250-DebtReduction1-1DC-CreditSolutions			7500-15000		
	w-302252-DebtReduction1-1DC-yellowarrow			7500-10000		
	w-302252-DebtReduction1-1DC-yellowarrow-dark			More_than_100000		
	w-300250-DebtReduction1-1DC-Head3					
	w-302252-DebtReduction1-1DC-white					
	w-300250-DebtReduction1-1DC-BlueMeter					
	w-300250-DebtReduction1-1DC-white					

- **Step 4:** Now, after we have decided the desirable order, we calculate the lead quality. Highly desirable category is given higher rank. Here for example, September – Rank 6, April – Rank 5, August – Rank 4, June – Rank 3, July – Rank 2, May – Rank 1. We do this for every parameter, and we develop a rank column accordingly.

    - Call status is also taken into consideration and the desirability order is given based on the knowledge, like Closed means the quality of the lead is very high and so on.

Score	Call Status
8	Closed
7	EP Confirmed
6	EP Received
5	EP Sent
4	Unable to contact - Bad Contact Information
3	Contacted - Invalid Profile
2	Contacted - Doesn't Qualify
1	Unknown

    - Also, another thing to consider is in the Address Score and the phone score, Null has higher desirability, but on a real case, we do not want it to be null. So we push it to the least desirability.

- **Step 5:** For every lead now, we calculate the total score.

- **Step 6:** We calculate the lead quality, and it is given by the (total score/ max score possible)

- The max score possible is the sum of number of categories in all the parameters, and here we have 60 total categories, so the maximum possible score is 60.

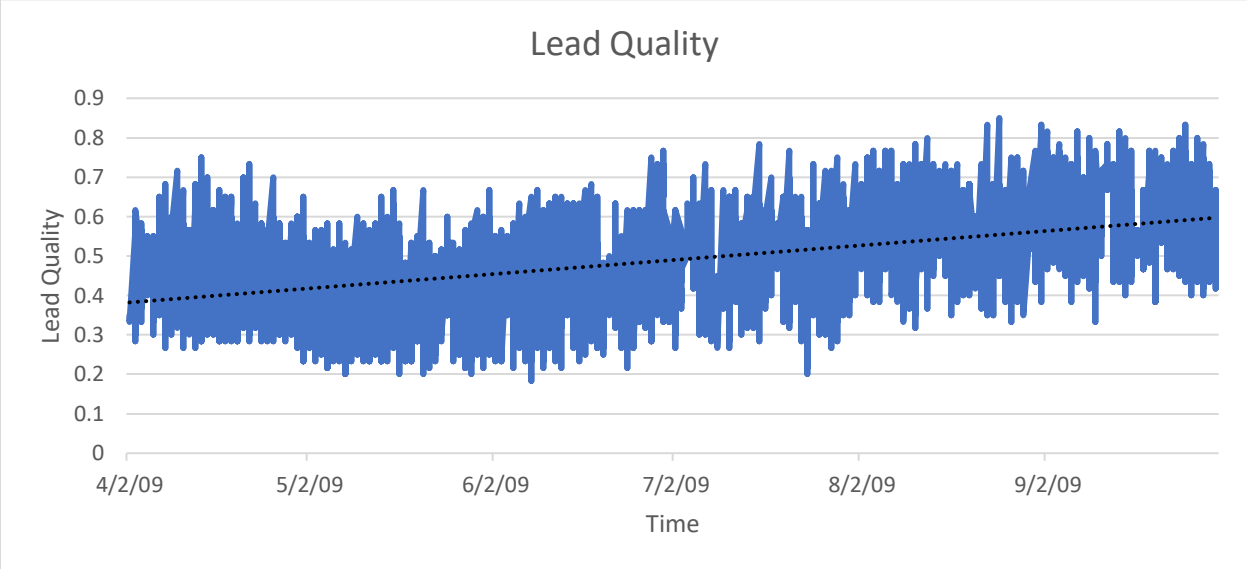
Example to calculate the Quality of some random lead.

Parameter Name	Value	Score
Month	August	4
Widget Name	w-302252-DebtReduction1-1DC-CreditSolutions	11
Address Score	3	3
Phone Score	3	5
Partner	Google	1
Landing Page URL	http://www.debtreductioninc.com/index8.html	4
Call Status	Unable to contact - Bad Contact Information	4
Debt Level	7500-10000	2
Total Score		34

The Quality of this lead is 34/60 = 0.56667 = 56.67%

Question 1 – Are we seeing any lead quality trends over time? Are they statistically significant?

The Time Series Plot is given below and at the first instance, based on the trend line it could be seen that there seems to be an increasing trend over time, i.e., The Quality of the lead seems to be increasing over time.



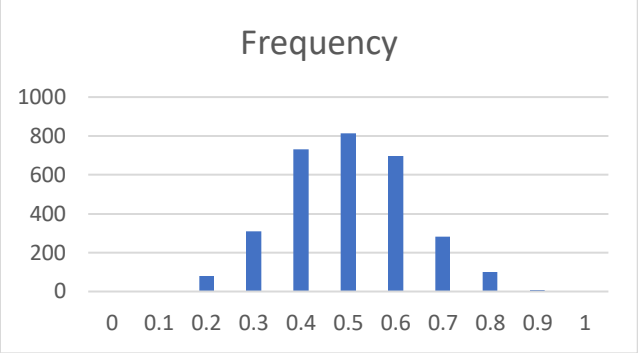
To prove the statistical significance we compare, the Average of consecutive months individually whether the Average lead quality has improved over month

Hypothesis Testing – t -test to compare the means of Two samples

Null Hypothesis: H0: Average Lead Quality of Month 1 = Average Lead Quality of Month 2
Alternate Hypothesis: H1: Average Lead Quality of Month 1 < Average Lead Quality of Month 2

To Reject the Null Hypothesis, p-value < 0.05 and t0 < -t(α, v)

To perform the t test to compare the means of two samples, we first need to prove the assumption of Normality. So, we plot the histogram to see whether the lead quality is distributed normally.



So, we can proceed with the T test as the normality assumption is satisfied.

April Vs May

μ1	April
μ2	May
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	1.48746E-19
But , t0 ≦ - t(α,v)	"9.14 ≦ -1.64"
So, result Contradiction. So we check May Vs April	

May Vs April

μ1	May
μ2	April
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	1.48746E-19
Also , t0 < - t(α,v)	"-9.14 < -1.64"
Average Lead Quality of May is Higher than April	

May Vs June

μ1	May
μ2	June
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	8.40067E-19
Also , t0 < - t(α,v)	"-8.92 < -1.64"
Average Lead Quality of June is Higher than May	

June Vs July

μ1	June
μ2	July
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	2.98524E-10
Also , t0 < - t(α,v)	"-6.24 < -1.64"
Average Lead Quality of July is Higher than June	

July Vs August

μ1	July
μ2	August
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	1.02921E-33
Also , t0 < - t(α,v)	"-12.53 < -1.64"
Average Lead Quality of August is Higher than July	

August Vs September

μ1	May
μ2	June
Null Hypothesis	*H0 : μ1 = μ2*
Alternate Hypothesis	*H1 : μ1 < μ2*
P Value is Much Lesser than 0.5	1.37097E-07
Also , t0 < - t(α,v)	"-5.19 < -1.64"
Average Lead Quality of September is Higher than August	

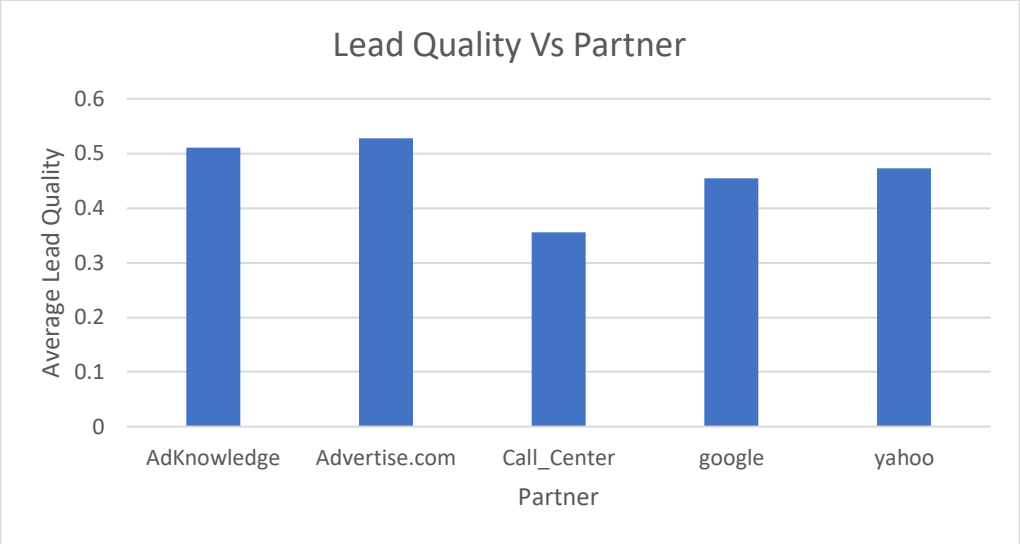
So, it is quite evident that during the first month, April – May, there was a drop in the trend for the lead quality and after that there is a statistically significant increasing trend is observed.

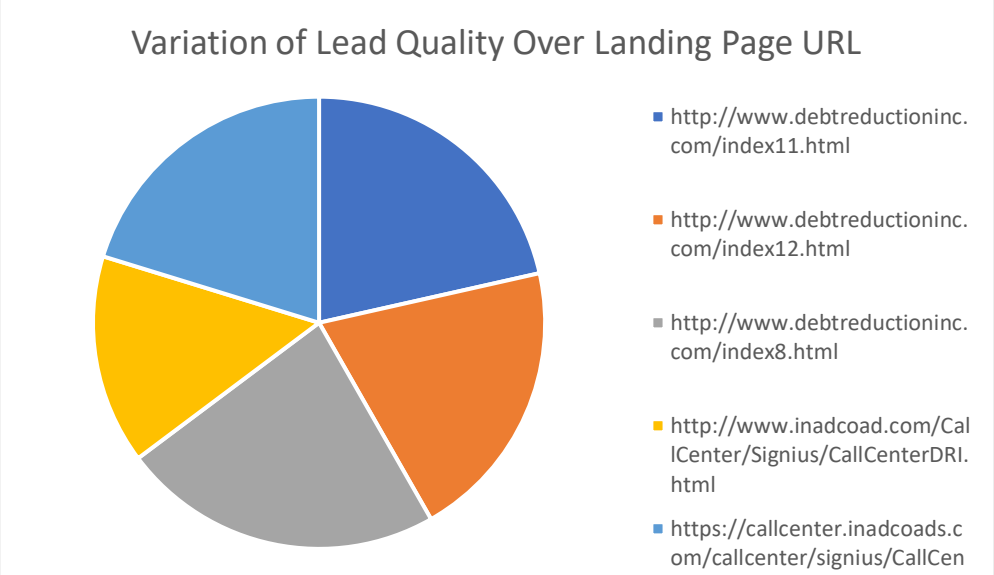
Question 2 – What can we learn about the drivers of "lead quality" from this dataset? What segments - where the ad was shown, what kind of person filled out the ad, what kind of ad did they see - have differing lead quality rates?

Where was the Ad shown?

The Ad shown can be understood by Partner and Landing URL. With the below given Bar chart and the Pie chart, it is not that significant that the partner and Landing URL has an impact over the Lead Quality.

But with respect to the partner, Call Center seems to have a relatively less Lead Quality than the others





What Kind of Person filled out?

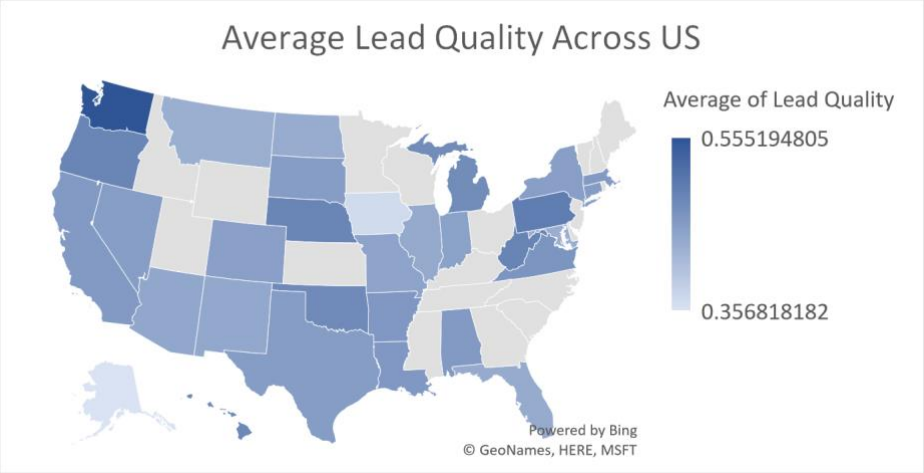
Type of Person

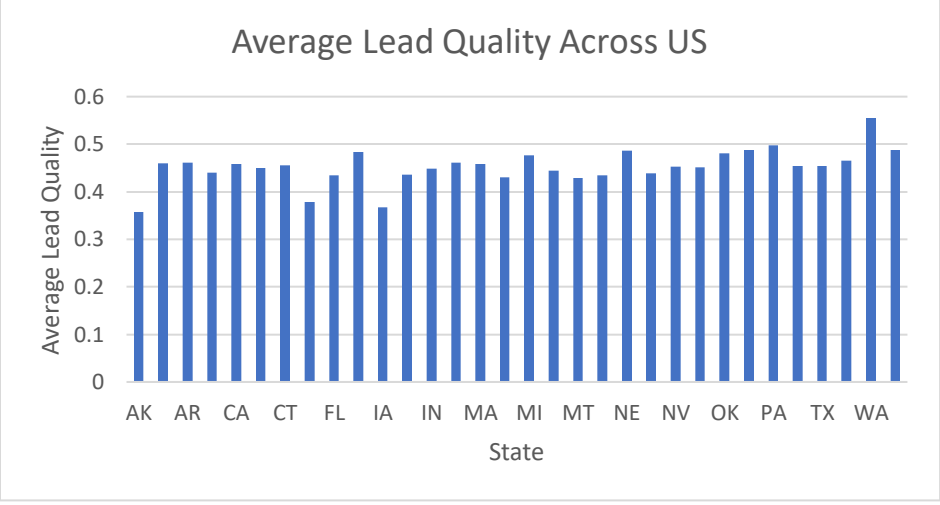
The type of person can be understood by the location and his debt level

Based on Location

From the bar chart below, there isn't much change in lead quality over region, except for AK, DC and IA. But the map given below gives us more insights. We are able to understand the value of lead quality relative to all states.

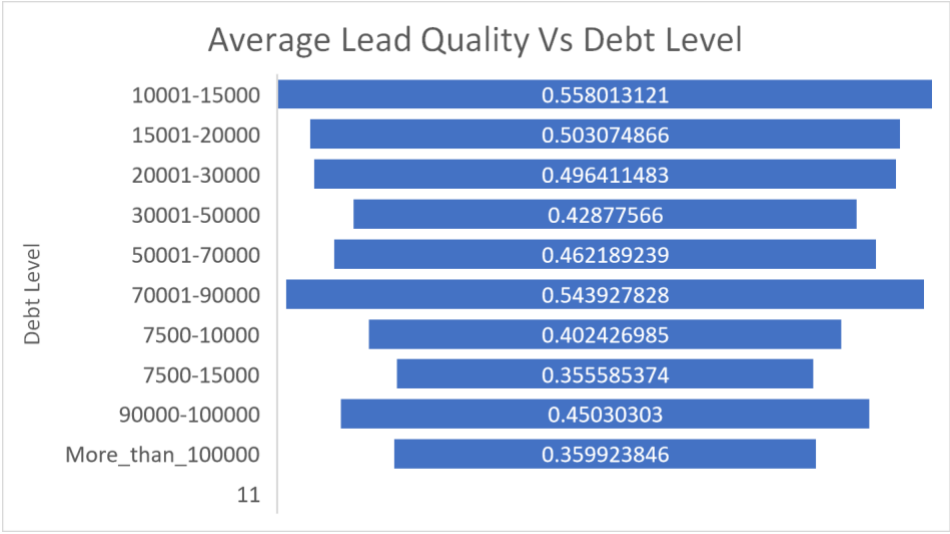
It is observed that few of the states in the central US and Eastern US have a relatively low lead quality than in the West Coast or the southern part. So, we might concentrate more along the West Coast and Southern part than the other part.



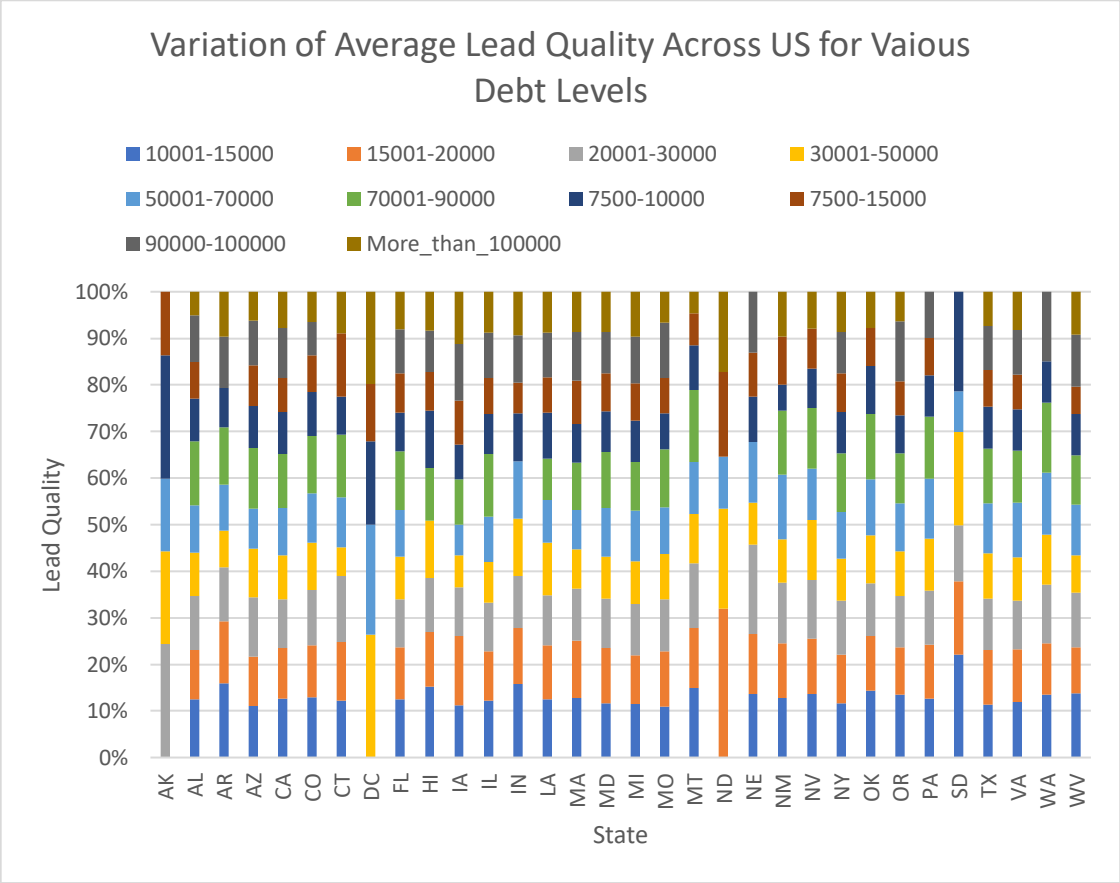


Based on Debt Level

People with debt level between 10001 – 15000 and 70001 – 90000 seems to give us a better lead quality than others. So, we might want to target the consumers who have this amount of debt level.



Based on Debt Level and Region



What Kind of Ad did they see?

It could be understood by the Widget Name

There seems to be a significant difference in the Lead Quality based on the type of widget the consumer sees. There are a few widgets like w-302252-DebtReduction1-1DC-yellowarrow, w-300250-DebtReduction1-1DC-CreditSolutions and a few others and the type of widgets that we show to the consumers should be one among these few that has high Quality Leads.

Question 3 - If the advertiser says they will increase our CPL by 20% (i.e., $30 to $33) if we increase our lead quality by 20% (i.e., from 8.0% to 9.6%), do we see any opportunities to do that here? What kinds of things could we do?

It is possible to increase the Quality of our lead by focusing on the parameters that affect our lead quality to a better extent. So, to determine the factors that affect the Lead Quality, we generate the Correlation matrix between Lead Quality and all the parameters that we are interested in studying.

The Correlation Matrix will help us in understanding the most important factor that will affect the Lead Quality

	Lead Quality	*Month*	*Widget Name*	*Address Score*	*Phone Score*	*Debt Level*	*Partner*	*Landing Page URL*
Lead Quality	1							
Month	0.29703948	1						
Widget Name	0.581984599	-0.045001542	1					
Address Score	0.518234994	0.134433309	0.009959356	1				
Phone Score	0.499364385	0.010588058	0.026833472	0.763327706	1			
Debt Level	0.457195338	0.039725635	0.017764749	0.087507369	0.042349341	1		
Partner	0.113280818	0.093693212	-0.015627753	-0.147148821	-0.156572965	0.00718192	1	
Landing Page URL	0.094155173	0.138310153	-0.128810817	-0.069683369	-0.052283822	0.04464118	0.21423419	1

The Most important factors that affect the Lead Quality that could be altered to improve the Quality of the lead are Widget Name, Address Score, Phone Score and Debt Level.

It is hard to quantify the measure of improvement the lead quality, but we are quite sure on what measure are to be taken to improve the quality of the lead.

From the above table we know that Widget Name, Address Score, Phone Score, Debt Level has a greater impact on the quality of the lead and from all the previous analysis, we definitely know which of these categories relatively performs better than others and we need to focus on in future to concentrate more on these categories that will help us in obtaining the increment of 20% in the lead quality.

Widget Name – We concentrate More on the widgets including “w-300250-DebtReduction1-1DC”, “w-300250-DebtReduction1-1DC-Head2”, “w-300250-DebtReduction1-2DC-CreditSolutions”, “w-302252-DebtReduction1-1DC-CreditSolutions”, “w-300250-DebtReduction1-2DC-BlueMeter”. So, these patterns of widget should be given more to the viewers, because these widget types have a higher potential to convert to Call Status to Closed and so thereby the Lead Quality is also not compromised.

Address Score and Phone Score – The Leads the has a high Address Score and Phone Score should be followed more than other leads because, the viewers who do really want to become the customer tend to be more inclined in giving the correct details. So, if we focus on leads that have high scores (4 and 5), we can possibly increase the quality of the lead.

Debt Level – This is the most important parameter in understanding the improvement of lead quality. People with debts come to service, and the group of people with specific debt level including 10001 – 15000 and 70001 – 90000 seems to have a higher potential of becoming a customer. So, instead of having leads from all debt levels, we can target group with specific debt level, so the Quality of each lead will be improved and also the total average quality could be improved.

Recommendations and Future Works

1. It is advisable to narrow down our leads based on the debt level and from the region they are from. This will help us to increase the CPL.
2. Also, the follow could be done not for everyone but for the viewers with High Address and Phone Score
3. Instead of having a wide variety of Widget Name, Partner Name, Landing URL, we can sue specific categories, that is possibly giving us high Lead Quality and high Conversion rate to closed (The most desirable)
4. The mathematical equivalent of lead quality could be refined more using other parameters including region, Ad Group, Marketing Campaign, Publisher Zone Name and Advertiser Campaign Name.
5. We might want to study the ratio at which the lead quality increased over time and understand the factors that has affected the increase and try to incorporate those ideas to improvise the Lead Quality.
6. If it is possible to obtain the revenue earned from each lead, then it is possible to see the profitability for the company. Sometime profit is deceptively success. We might have a higher number of customers, but the profit could be low or sometimes vice versa. So, if available the revenue from each closed lead, it would be better to quantify the Lead Quality.
7. Lead Quality could be defined in two ways
    a. Closing Potential – How easy it will be to convert that prospect into a customer.
    b. Revenue Potential – How much money the prospect could generate over a life time or a fixed period of time.
8. As mentioned, we have worked only with the closing potential so far. So, if we are able to have the data about the Revenue potential and together it is always a better measure to quantify the quality leads.