Findings

<u>TLDR</u>

Are we seeing any lead quality trends over time?

- **Yes, lead quality is declining.** Statistical analysis shows a **significant drop in closed leads** ($p = 0.008$) and an **increase in bad leads** ($p = 0.020$) over time.

What are the key drivers of lead quality?

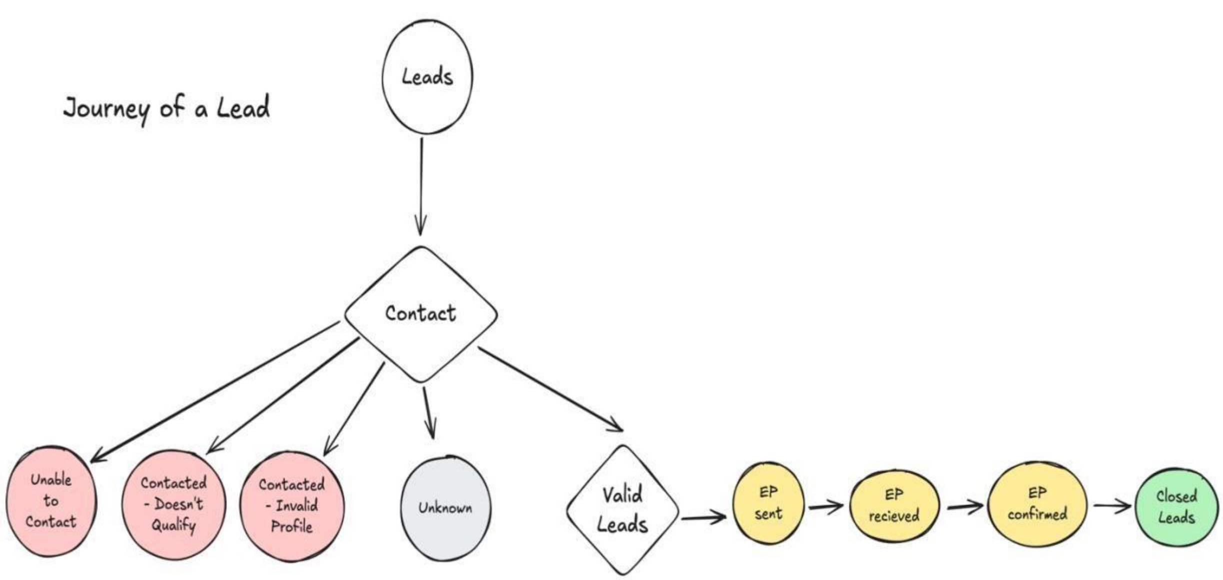
- **Top Factors:**
    - **Ad Placement** – Adknowledge performs best.
    - **Landing Pages** – index8.html is the highest quality.
    - **Keywords** – "Debt" is the best, while "student loan default" generates bad leads.
    - **User Segments** – Users with **$10K-$30K debt** and those in **CA, TX** convert best.
    - **Ad Design** – 1DC-BlueMeter and 1DC-Head2 have the highest close rates.

What can we do to improve lead quality?

- **Shift spend to high-quality sources** (Adknowledge, index8.html).
- **Eliminate low-performing keywords & landing pages.**
- **Target high-value users** (debt levels, geography).
- **Expand top-performing ad formats** (1DC-BlueMeter, 1DC-Head2).

Context

Based on reading the problem statement, this is what I believe the journey of a lead looks like:



As stated in the question, the lead can have the following status:

- Closed Leads are the best. When discussing quality, we need to optimize for increasing the number of closed leads.
- Good Leads: EP Sent / Closed / Received are still in the conversion funnel with a possible path to closure in future and is still considered a desirable state.
- Bad Leads: No Contact / Invalid Profile / Did not Qualify are non-desirable lead traits. The business would want to minimize these to save costs & efforts.
- Neutral Outcomes: These are neither good nor bad outcomes. For this analysis, we would not optimize for this type of outcome. However, in the real world, I would like to minimize the noise.

Q1. Are we seeing any lead quality trends over time (improving, declining)? Are they statistically significant?

To track lead quality and top drivers, I will be tracking the following metrics:

- **Total Lead Count, Closed Lead % and Bad Lead %**
  The total lead count helps us identify any change in lead generation which can affect the effectiveness of the system. We want this number to grow or stay stable. In addition, we want Closed Lead% to go up and Bad Lead% to go down.

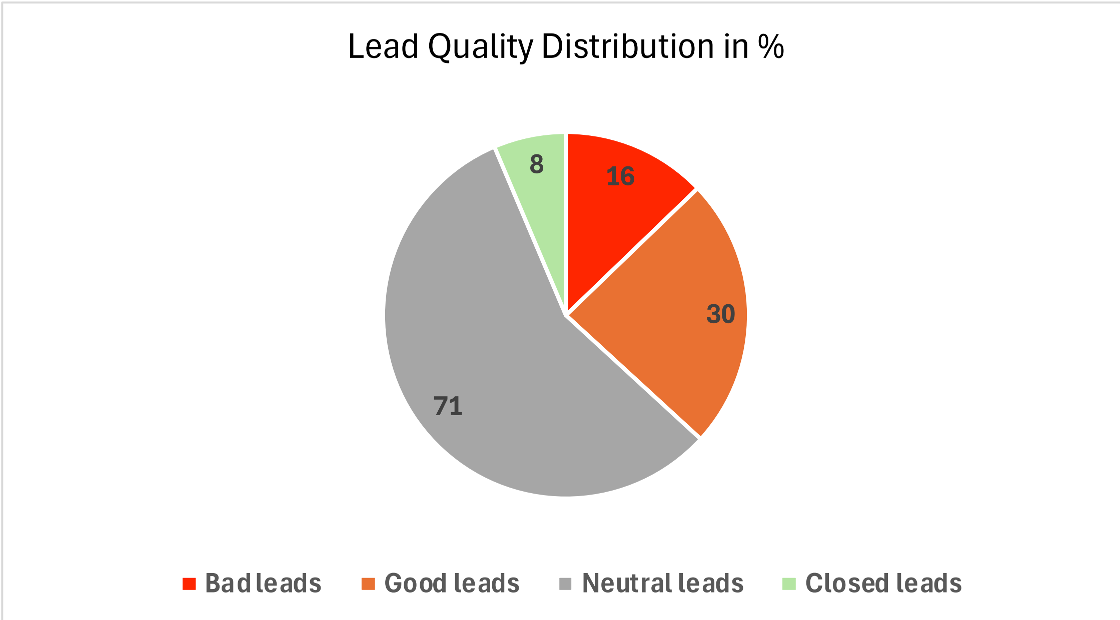
- **Weighted Score** Associates' weights to all outcomes based on relative importance. I am assuming the following scheme:
  - 5: Close Lead
  - 2: Good Leads
  - 0: Neutral Leads
  - -1: Bad Leads

  Weighted Score = (5 Closed Lead + 2 * Good Lead + (-1) * Bad Lead) / Total Leads*

Note: The weights are logically selected, but we can use other factors like cost associated and returns on closed lead to select better weights.

Additionally, I will analyze Week over Week (WoW) data. Daily data analysis is volatile and Month over Month data is sparse due to the small data sample (6 months).

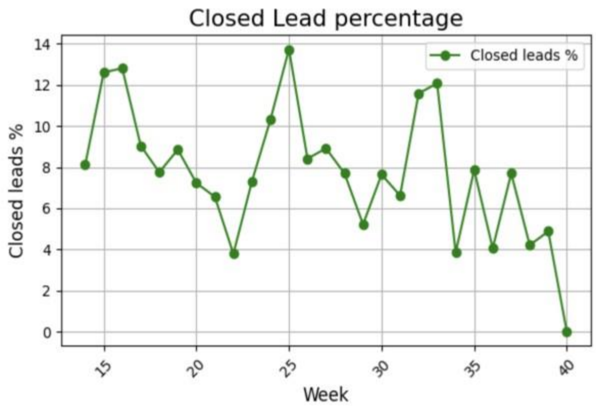
Let's look at the Lead Outcome Distribution to get an aggregated view

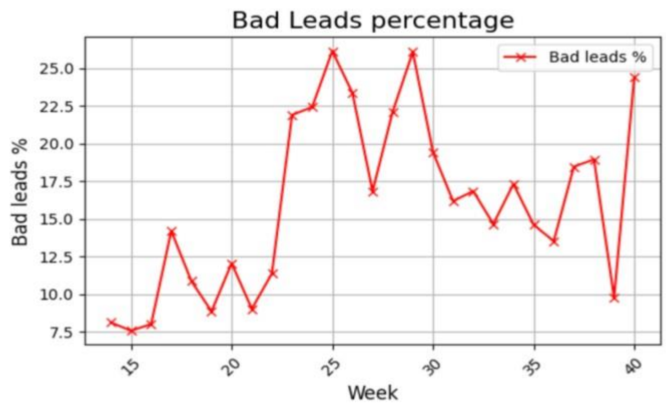


Approx 8% leads close and 16% go bad

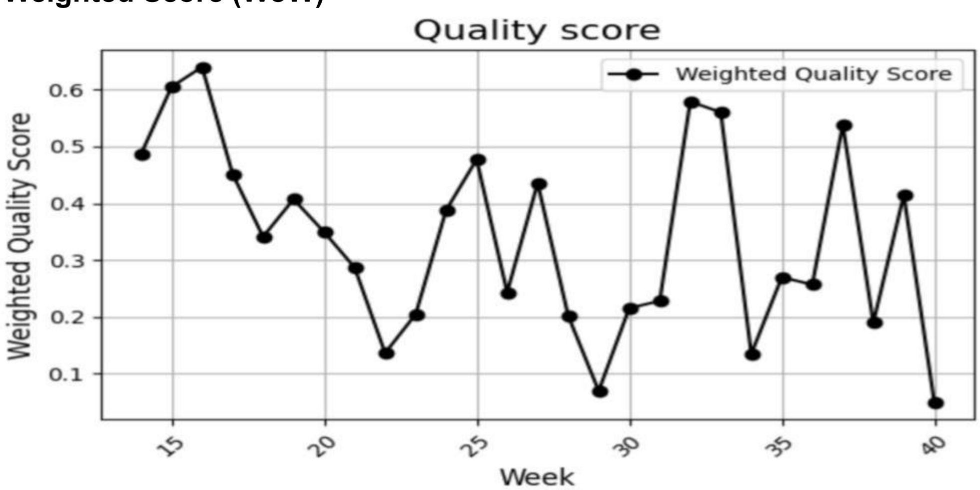
Total Leads (WoW)

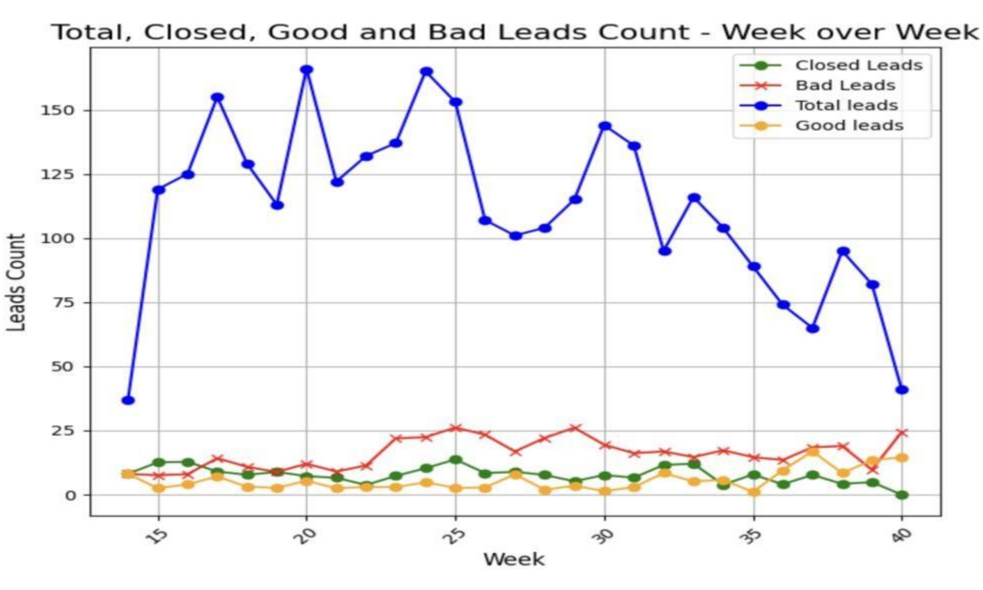
Closed Leads% (WoW) and **Bad Leads% (WoW)**





Weighted Score (WoW)





Statistical Analysis

I am using **linear regression** to quantify the relationship between the week and lead quality. This will enable me to perform a statistical analysis on the three parameters I have defined for measuring lead quality.

- Closed leads percentage
- Bad leads percentage
- Adjusted Weighted score

Variables used

- Independent variable - Weeks
- Dependent variable - Closed leads percentage/Bad leads percentage/Adjusted Weighted score

Hypothesis

- Null hypothesis - There is no linear relationship between time and lead quality parameters
- Alternate hypothesis - There is a linear relationship and the quality parameters are changing over time

Closed Lead percentage

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.247
Model:                            OLS   Adj. R-squared:                  0.217
Method:                 Least Squares   F-statistic:                     8.213
Date:                Wed, 29 Jan 2025   Prob (F-statistic):            0.00831
Time:                        22:37:13   Log-Likelihood:                -65.280
No. Observations:                  27   AIC:                             134.6
Df Residuals:                      25   BIC:                             137.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         10.3282      1.057      9.775      0.000       8.152      12.504
x1            -0.1998      0.070     -2.866      0.008      -0.343      -0.056
==============================================================================
Omnibus:                        0.810   Durbin-Watson:                   1.480
Prob(Omnibus):                  0.667   Jarque-Bera (JB):                0.622
Skew:                           0.353   Prob(JB):                        0.733
Kurtosis:                       2.770   Cond. No.                         29.6
==============================================================================
```

Based on the above results, we can say that given **p-value of 0.008**, I can conclude with 95% confidence that the relationship between week and closed lead % is statistically significant, suggesting a decline in the closed lead % over time

Bad Lead percentage

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.198
Model:                            OLS   Adj. R-squared:                  0.166
Method:                 Least Squares   F-statistic:                     6.172
Date:                Wed, 29 Jan 2025   Prob (F-statistic):             0.0200
Time:                        22:37:16   Log-Likelihood:                -82.453
No. Observations:                  27   AIC:                             168.9
Df Residuals:                      25   BIC:                             171.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.7821      1.996      5.903      0.000       7.672      15.893
x1             0.3272      0.132      2.484      0.020       0.056       0.598
==============================================================================
Omnibus:                        1.465   Durbin-Watson:                   0.970
Prob(Omnibus):                  0.481   Jarque-Bera (JB):                1.349
Skew:                           0.477   Prob(JB):                        0.509
Kurtosis:                       2.462   Cond. No.                         29.6
==============================================================================
```

Based on the above results, given **the p-value of 0.020,** we can conclude with 95% confidence that there is a statistically significant relationship between week and average bad lead percentage suggesting an increase in avg bad % over time.

Adjusted Weighted Score

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.137
Model:                            OLS   Adj. R-squared:                  0.103
Method:                 Least Squares   F-statistic:                     3.983
Date:                Wed, 29 Jan 2025   Prob (F-statistic):             0.0570
Time:                        22:37:32   Log-Likelihood:                 12.362
No. Observations:                  27   AIC:                            -20.72
Df Residuals:                      25   BIC:                            -18.13
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.4412      0.060      7.406      0.000       0.318       0.564
x1            -0.0078      0.004     -1.996      0.057      -0.016       0.000
==============================================================================
Omnibus:                        1.689   Durbin-Watson:                   1.632
Prob(Omnibus):                  0.430   Jarque-Bera (JB):                1.320
Skew:                           0.345   Prob(JB):                        0.517
Kurtosis:                       2.166   Cond. No.                         29.6
==============================================================================
```

Based on the above results, we can say that given **p-value of 0.057** I can conclude with 90% confidence that the relationship between week and quality score is statistically significant, suggesting a decline in the quality score over time

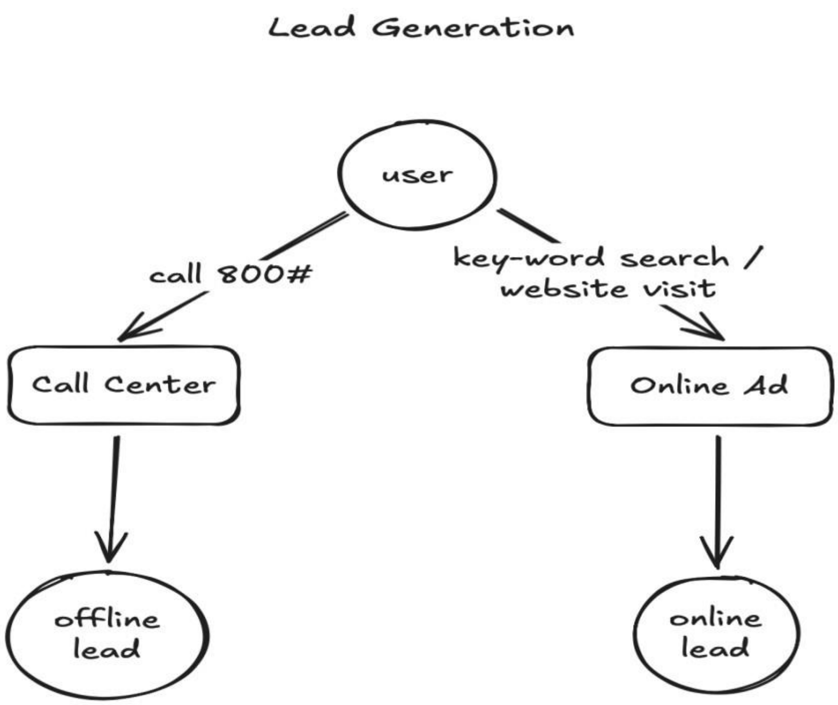
<u>Some insights</u>

- In Week 25, both closed leads and bad leads increased, likely due to a rise in total leads.
- Week 33 was a strong month, marked by a notable increase in closed leads.
- Although bad lead quality decreased last month, closed leads remained stagnant, while good leads increased, indicating a bottleneck in the closing process
- Statistically, there has been a significant decline in closed leads and a notable increase in bad leads over time

Q2. 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

Leads can be generated by two methods : From Call-Center by Calling 800# or By Filling an online form.



Lets deep dive into the following parameters to understand lead quality - •
Publisher Campaign Name (Online vs Offline)

- **Partner**
- **Landing Page Url**
- **Keywords**

Aggregate quality metric for Call-Center vs Online Ad

Publisher Campaign Name	Total leads %	Closed leads %	Bad leads %	Weighted Quality Score
Debt Reduction Call Center	9	9.59	24.35	0.369
Debt Reduction Inc	91	7.96	15.94	0.339

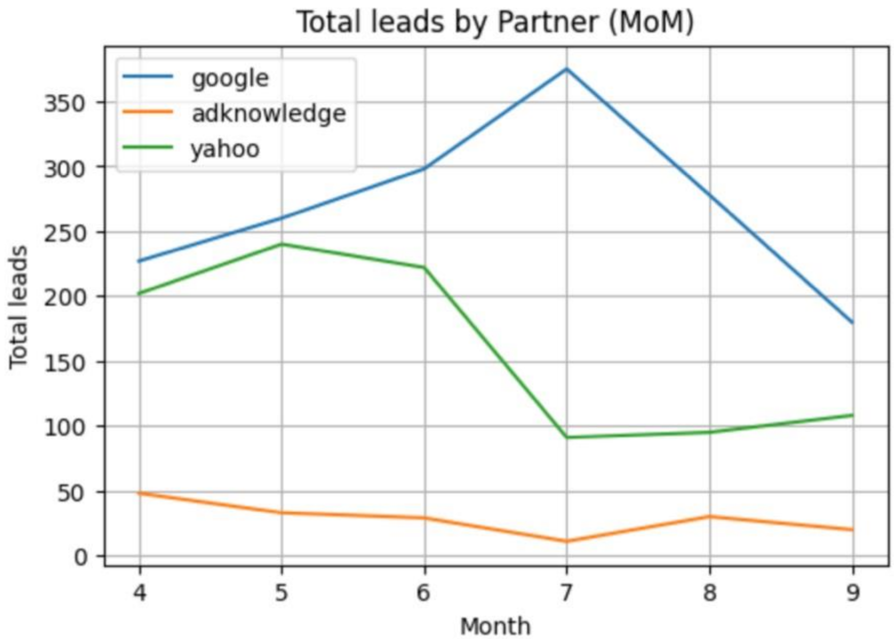
Insights :

The total lead count for Online Ads (DebtReductionInc) is 91% of all leads. Only 9% leads are generated via call-centers.

The call-center leads have a slightly higher rate of closing and weighted-score.
They surprisingly have a higher bad lead rate, which could indicate a gap in

- call center staff entering incorrect info or generating bad leads for those who do not qualify.
- people sharing bad information

Partner



Quality Matrix

CallStatusEnum	Total leads %	Closed leads %	Bad leads %	Weighted Quality Score
Partner adknowledge	6	12.2	13.4	0.608187
Google	54	7.60	15.5	0.294808
Yahoo	32	7.72	15.3	0.362213

Insights:

Google dominates as the leading partner, contributing 54% of the total leads, followed by Yahoo. Both Google and Yahoo exhibit similar lead quality performance

Landing page URL

Landing page URL by partners

Landing Page URL	Partner	Total Leads %
Google	http://www.debtreductioninc.com/index8.html http://www.debtreductioninc.com/indexl 2. html http://www.debtreductioninc.com/indexl 1 .html	47% 4% 3%
Yahoo	http://www.debtreductioninc.com/index8.html	32%
Call center	https://callcenter.inadcoads.com/callcenter/si... http://www.inadcoad.com/callcenter/signius/cal...	6% 6%
adknowledge	http://www.debtreductioninc.com/index8.html	6%

Quality Matrix

LandingPageURL	Total leads %	Closed leads %	Bad leads %	Weighted Quality Score
http://www.debtreductioninc.com/index8.html	84	8.30	14.9	0.36
https://callcenter.inadcoads.com/callcenter/signius/callcenterdri.html	6	11.70	30.31	0.40
http://www.debtreductioninc.com/index12.html	4	1.61	22.6	-0.11
http://www.debtreductioninc.com/indexll.html	3	6.97	18.6	0.18
http://www.inadcoad.com/callcenter/signius/callcenterdri.html	3	4.81	10.8	0.27

Insights:

- **Google** utilizes three landing pages, with landing page index 8 standing out not only for the highest lead count but also for generating good lead quality.
- **However**, two other URLs (index 2) appear to be driving poor-quality leads, possibly due to being redirect URLs.
- **The** call center at "https://callcenter.inadcoads" is yielding high-quality leads, whereas the one at "**http://www.inadcoad.com/**" shows lower quality.

Keywords

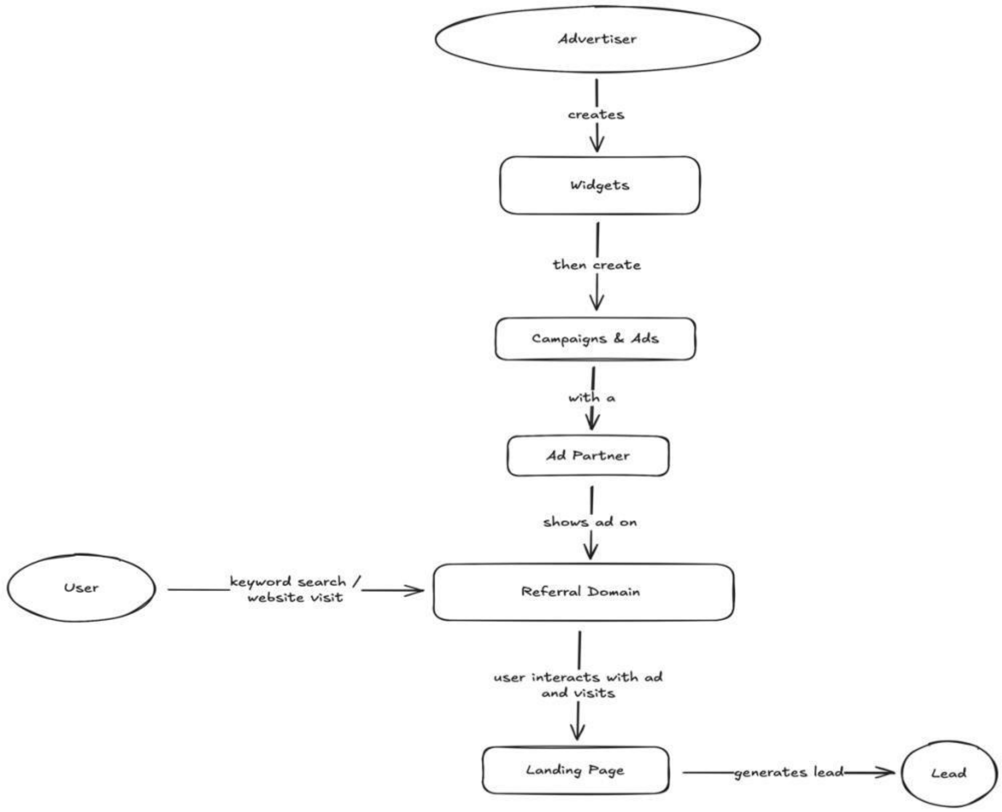
Top 10 Searched Keywords and their quality

Keyword	Total leads count	Closed leads %	Bad leads %	Weighted Quality Score
debt	157	8.92	12.74	0.471338
credit card payments	104	6.73	19.23	0.201923
student loan default	51	0.00	23.53	-0.196078
loan default help	44	0.00	43.18	-0.431818
credit services	21	9.52	23.81	0.238095
debt cures	21	4.76	33.33	-0.095238

Insights

- **Debt** is the best performing keyword. While others like "student loan default" bring in mostly bad leads that do not close. We can optimize campaign setup by excluding these keywords

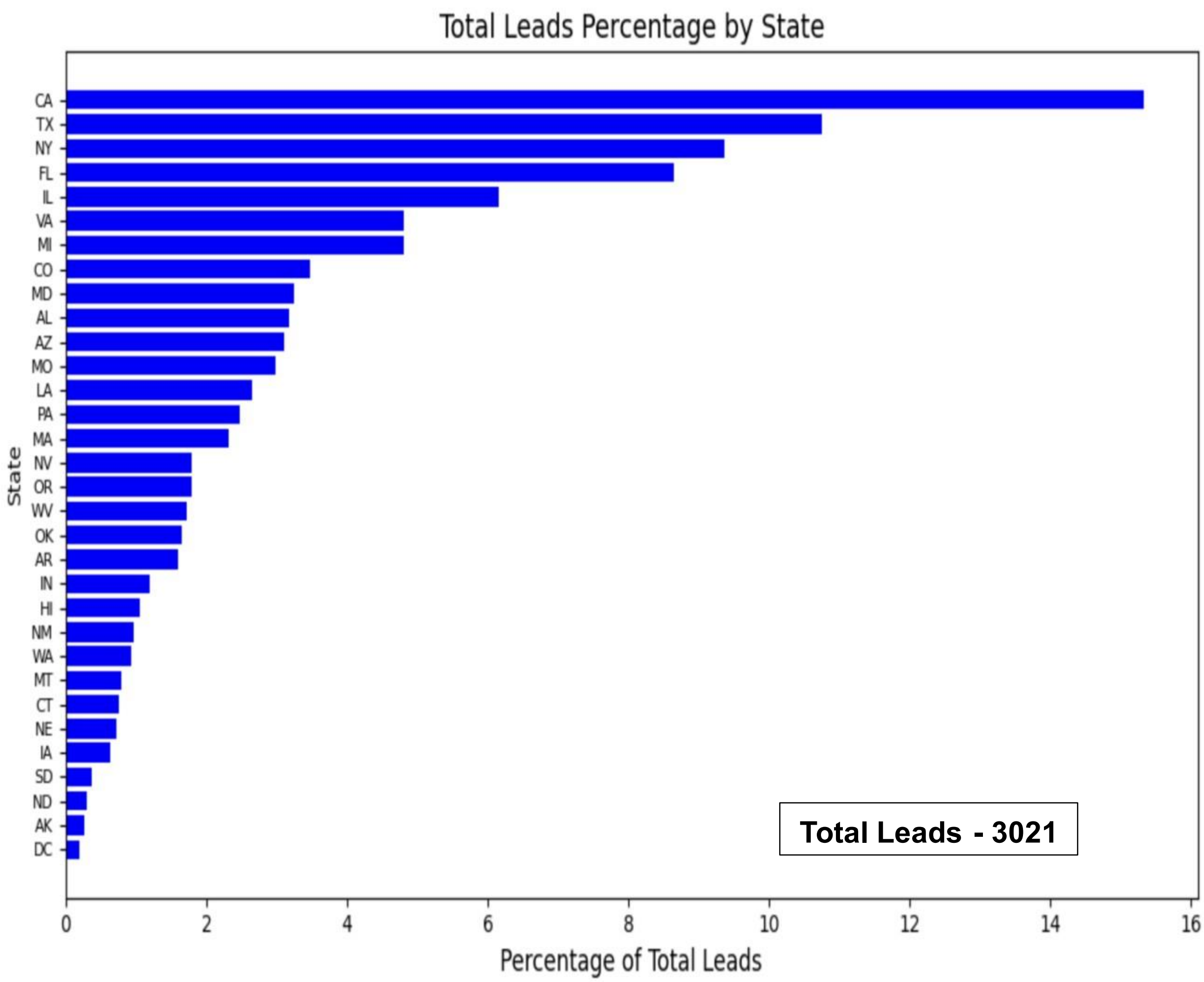
Kind of person



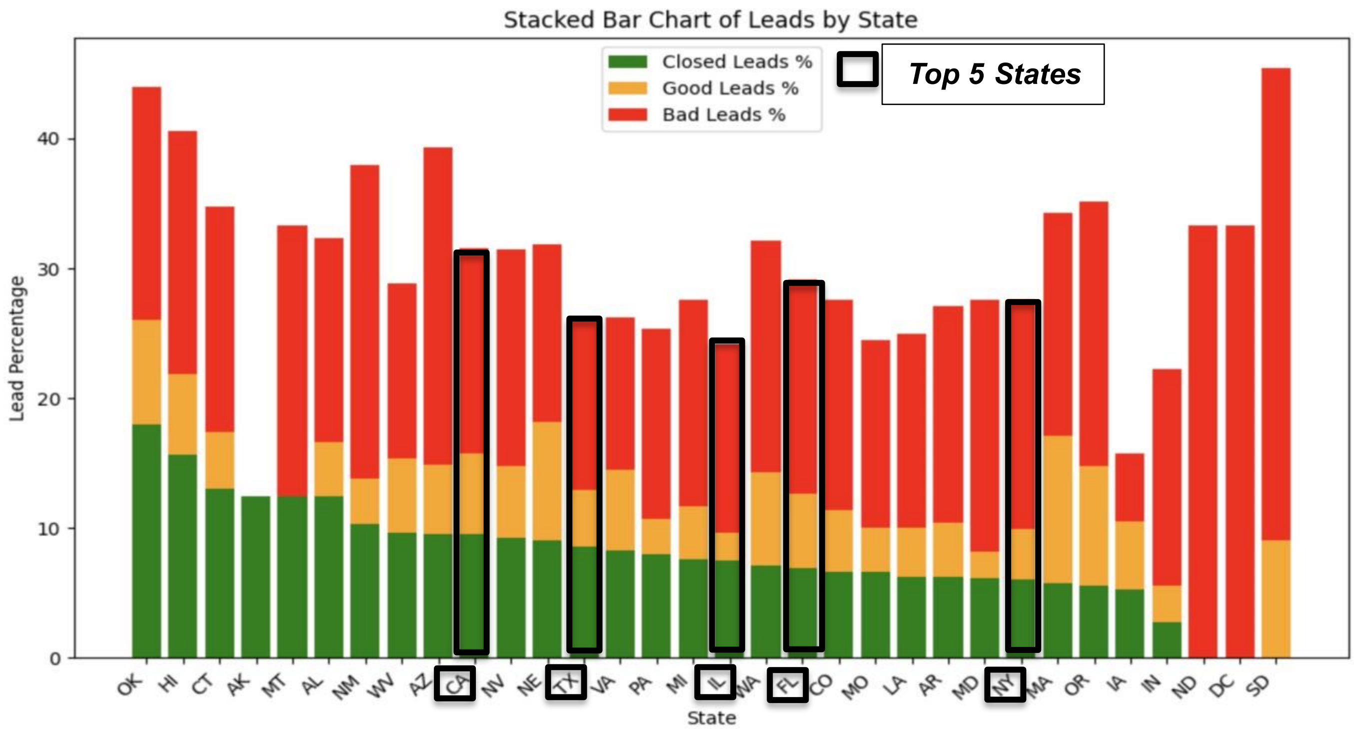
Lets deep dive into the following to understand customer and their affect on lead quality

- **State**
- **Debt Level**
- **Phone Score and Address Score**

Total Lead Count by state



Lead Quality by state



Insights

- CA, TX, NY, FL, and IL have the highest lead counts, which is in account with their population.
- CA and TX also show higher closed lead % than the average closed lead %, while NY and IL exhibit higher bad lead %.

Debt Level

Matrix

Debt Level	Total Leads %	Closed leads %	Good leads %	Bad leads %	Neutral leads %	Weighted Quality Score
7500-10000	15	5	2	25	68	0.04
7500-15000 10001-15000	9	7	4	10	79	0.34
	10	12	7	17	65	0.55
15001-20000 20001-30000	14	9	5	12	74	0.42

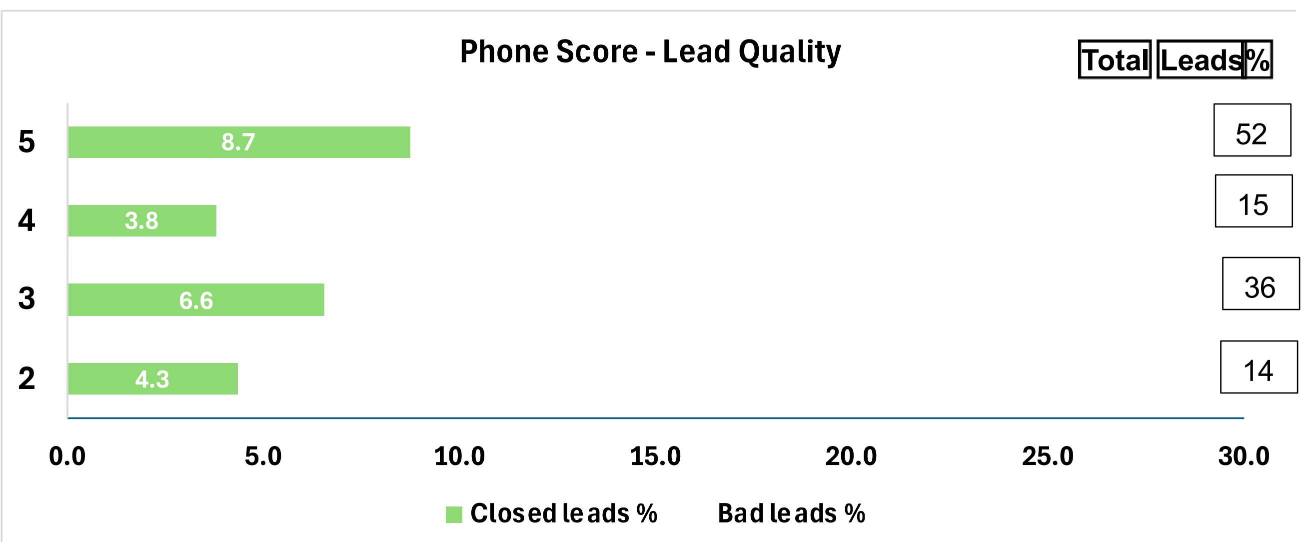
	15	9	7	15	69	0.43
30001-50000	16	8	4	17	71	0.31
50001-70000	8	9	7	14	71	0.44
70001-90000	4	14	5	12	69	0.67
90000-100000	3	10	2	14	73	0.40
More_than_100000	6	4	3	20	73	0.05

Total Leads - 3021

Insights

- Customers with debt levels between $10,000 - $30,000 have a higher closed lead %.
- While customers with higher debt levels have lower total leads %, they still demonstrate a higher closed lead %.

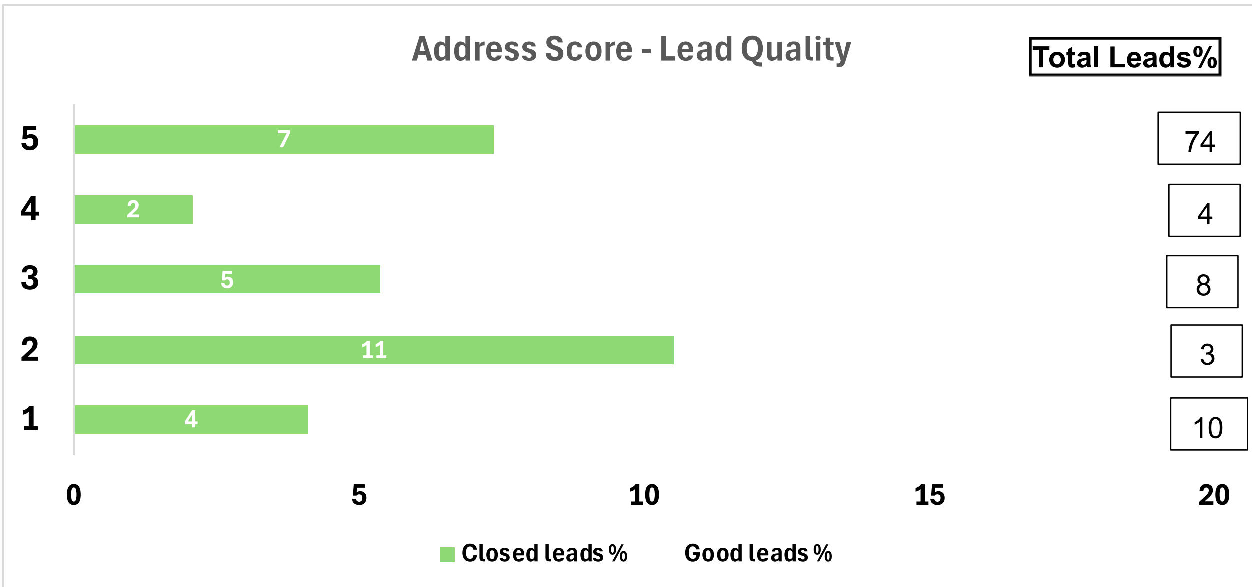
Phone Score



Insufficient data for score 1

Total Leads - 1393

Address Score



Total Leads - 1171

<u>Insights:</u>

- Both phone score and address score show that their higher values are associated with better lead quality

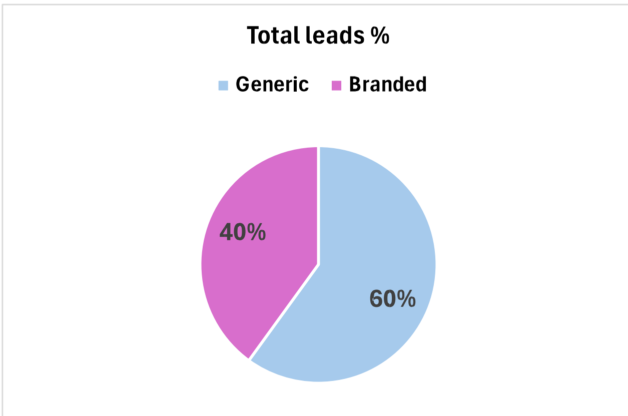
What Type of Ad

Lets deep dive into the following parameters to understand lead quality

- Logo Type
- Widget Name (1DC (1 page) vs 2DC (2 page) and Design Type

Logo Type

Generic vs With Brand Logo



Campaign Type	Closed leads %	Bad leads %
Generic	8.320	16.419
Branded	7.794	15.755

<u>Insights:</u>

The split between total leads generated from generic ads and those with brand logo is 60 / 40.

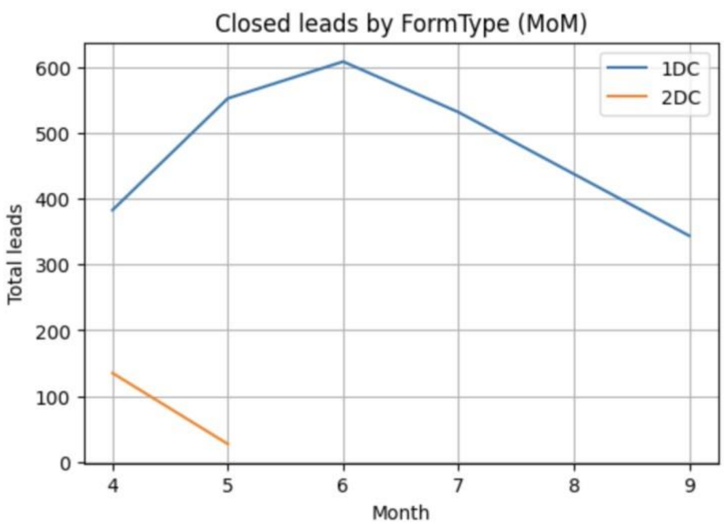
- Surprisingly the closed lead percentage for Generic ads is higher.
- Since there are multiple widgets designs it is not fair to use the above to decide which is better. We should look at equivalent designs where one has branding and other doesn't to make this decision.

1DC (1 page) vs 2DC (2 page)

Form Type	Total leads %	Closed leads %	Bad leads %	Weighted Quality Score
IDC	95	8.14	16.33	0.34
2DC	5	7.40	12.9	0.33

<u>Insights:</u>

- The 1DC form makes up for 95% of the total ad generated lead. This can be attributed to potential drop-off of customer between the two form pages.
- The closed lead % is higher for 1DC vs 2DC, but so is the Bad Lead% as an easy UI might invite more people who are just clicking for fun/curiosity or are non-qualifying.
- 1DC format is better in this case as it is generating a much higher number of leads, we can see 2 DC was discontinued after two months from chart below.



Design Type

Design Name	Total leads %	Closed leads %	Bad leads %	Weighted Quality Score
1DC-CreditSolutions	37	7.8	15.7	0.34
1DC	21	9.2	19.0	0.37
1DC-white	14	7.3	15.8	0.27
1DC-yellowarrow-blue	8	6.5	21.1	0.25
1DC-yellowarrow-dark	4	7.4	18.5	0.26
1DC-BlueMeter	3	14.1	4.3	0.68
1DC-Head2	3	12.4	9.0	0.60
2DC-BlueMeter	3	6.9	10.3	0.38
1DC-Head3	2	5.3	9.3	0.20

Insights

- We can see from comparison of different designs that the 1DC-CreditSolutions and 1DC (generic) are biggest in-term of generating leads and have overall decent lead quality.
- The two widgets 1DC-BlueMeter and 1DC-Head2 have high closing rate and overall score. They also have a low bad rate percentage. It is worth investigating those two designs to see what they are doing right. We also need to investigate why their share in overall lead % is small.

Q3 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?

1. Optimizing Ad Placement and Partners

- **Call Center vs. Online Ads:** Call center leads have a slightly higher closing rate but also a higher bad lead rate. Improving call center data quality (e.g., better screening, verification processes) could enhance lead quality.
- **Partner Selection:** Ads from **Adknowledge** have the highest weighted quality score (0.608), significantly outperforming Google (0.294) and Yahoo (0.362). Increasing the budget for Adknowledge ads could improve lead quality.

2. Refining Landing Pages

- **Best Performing Landing Page:** index8.html performs best across platforms. Prioritizing traffic to this page while discontinuing or modifying lower-quality pages (e.g., index12.html, index11.html) can improve conversions.
- **Call Center Landing Page Discrepancy:** One call center URL (callcenter.inadcoads.com) produces high-quality leads, while another (inadcoad.com) performs worse. Investigating and standardizing best practices across call center pages may help.

3. Keyword Optimization

- **High-Performing Keywords:** "Debt" has the highest lead quality (8.92% closed, 12.74% bad, weighted score 0.471). Investing more in this keyword could increase quality.
- **Poor-Performing Keywords:** Keywords like "student loan default" (0% closed, 23.53% bad) and "loan default help" (0% closed, 43.18% bad) should be excluded or reduced in bidding.

4. Targeting Higher-Quality Users

- **Debt Level:** Users with debt levels between **$10,000 - $30,000** show the highest closed lead rates. Adjusting targeting to focus on these segments (rather than very high or very low debt levels) could improve lead quality.

- **Geographic Focus:** CA and TX have above-average closed lead rates. Increasing ad spend in these states while reducing investment in states with high bad lead rates (like NY, IL) could help.

5. Ad Design and Format Adjustments

- **Branding:** Surprisingly, **generic ads** (8.32% closed) outperform branded ads (7.79%). Testing similar designs without branding for better performance could be an option.
- **Best Performing Ad Designs: 1DC-BlueMeter (14.1% closed, 4.3% bad, 0.68 weighted score) and 1DC-Head2 (12.4% closed, 9.0% bad, 0.60 weighted score)** have the best performance. Investigating what makes these successful and replicating their features in other designs could improve overall lead quality.
- **Form Type:** The **1DC (one-page) form** has a higher closed lead rate (8.14%) than 2DC (7.40%), despite more bad leads. It should remain the primary form, but small adjustments to filter out low-quality leads may help.

Action Plan to Increase Lead Quality by 20%

1. **Shift ad spend to high-performing partners** (Adknowledge) and **optimize Google/Yahoo ad targeting**.
2. **Redirect traffic to high-performing landing pages** (index8.html, callcenter.inadcoads.com).
3. **Exclude low-quality keywords** and increase bids on "debt" and other high-performing terms.
4. **Focus ads on users with $10,000 - $30,000 in debt** and in high-performing states (CA, TX).
5. **Expand successful ad designs** (1DC-BlueMeter, 1DC-Head2) and test generic branding.
6. **Improve call center processes** to filter out unqualified leads before submission.

By implementing these optimizations, the lead quality should improve, making the proposed **CPL increase to $33** worthwhile.