Lead Quality Analysis & Optimization Strategy

Prepared for: Aarki

Prepared by: Yi-Chieh (Bella) Kuo

Date: Febuary 10th, 2025

1. Executive Summary – Lead Quality Analysis & Optimization Strategy

In this analysis, I evaluate lead quality trends, identify key drivers affecting lead quality, and assess the feasibility of improving lead quality by 20% to support a Cost Per Lead (CPL) increase from $30 to $33. The results indicate that lead quality has remained stable over time, with no statistically significant upward or downward trend. A key concern is the high proportion of unclassified leads, which limits the reliability of analytical insights and decision-making.

To determine the factors influencing lead quality, I conducted statistical and machine learning-based analyses, identifying **Widget Name, Debt Level, and Partner** as the most significant variables. High-performing widgets, such as *w-300250-DebtReduction1-1DC-BlueMeter* and *w-300250-DebtReduction1-1DC-CreditSolutions*, consistently generate a higher proportion of high-quality leads, while low-performing widgets contribute to inefficiencies. Debt level analysis reveals that leads within the **$7,500 - $20,000** range demonstrate the highest High Quality Rates, making them an optimal target for acquisition efforts. In contrast, leads with debt levels exceeding **$100,000** have the lowest quality and should be deprioritized. Partner analysis highlights **Adknowledge** as the most effective source for high-quality leads, though **Google and Yahoo** remain critical due to their scale, necessitating a balanced investment approach.

Assessing the feasibility of increasing lead quality from **8.0% to 9.6%**, historical data confirms that achieving this target is possible with **targeted optimizations**. A financial analysis comparing the current CPL ($30) and proposed CPL ($33) scenarios indicates an ROI increase from **35.17% to 45.45%**, supporting the viability of this strategic adjustment. To achieve the desired quality uplift, I recommend increasing investment in **high-performing widgets**, reallocating resources toward the **$7,500 - $20,000 debt segment**, and optimizing **partner allocations** to maximize efficiency. Additionally, refining lead classification processes will enhance data accuracy and improve decision-making.

By implementing these strategies, I expect a **sustained improvement in lead quality, higher conversion rates, and increased cost efficiency**. Continuous performance monitoring and iterative optimizations will be essential to maintaining long-term marketing effectiveness and ensuring that CPL adjustments align with sustained lead quality improvements.

2. Lead Quality Trends and Statistical Significance

2.1 Analysis Approach and Rationale

To understand lead quality trends over time, we analyzed the monthly distribution of lead quality categories (High, Medium, Low, and Unknown) along with the Identified Lead Ratio (percentage of classified leads). The purpose of this analysis is to determine if there are notable shifts in lead quality that could indicate improvement or deterioration over time. The Identified Lead Ratio was included to assess the extent to which leads were successfully classified, providing insight into data completeness and consistency.

To classify leads into different quality categories, the following criteria were used:

Lead Quality Category	Criteria
High Quality	Lead status is **'Closed'** (customer conversion achieved)
Medium Quality	Lead status includes **'EP Sent,' 'EP Received,' or 'EP Confirmed'** (progressing in funnel but not yet converted)
Low Quality	Lead status includes **'Unable to Contact,' 'Invalid Profile,' or 'Doesn't Qualify'** (lead is not viable)
Unknown	All other lead statuses that do not fit into the above categories

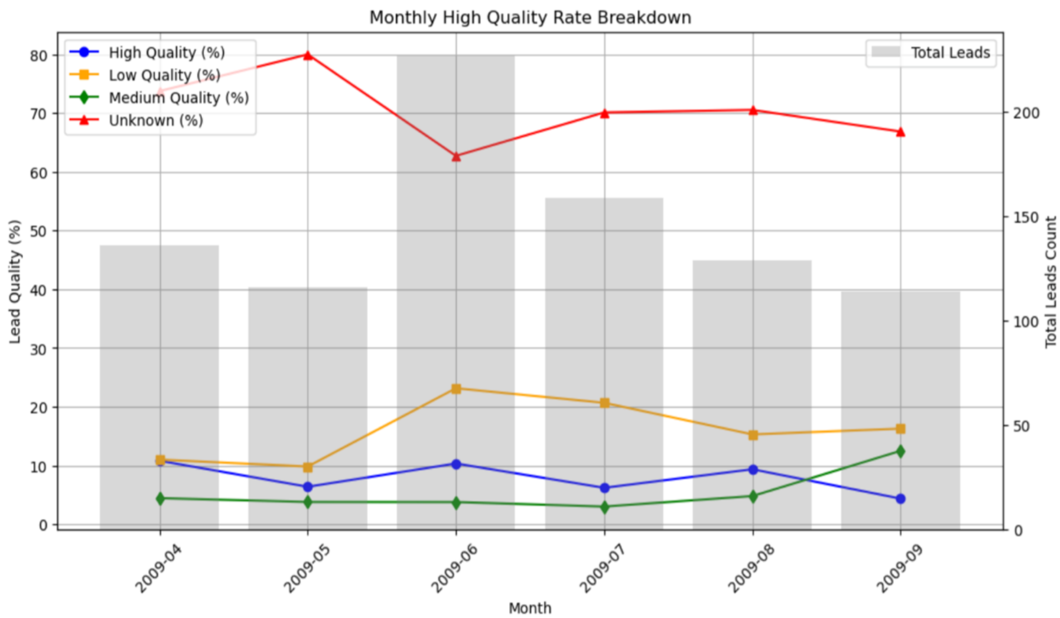
This classification ensures that the analysis effectively captures lead quality trends and provides actionable insights for optimization.

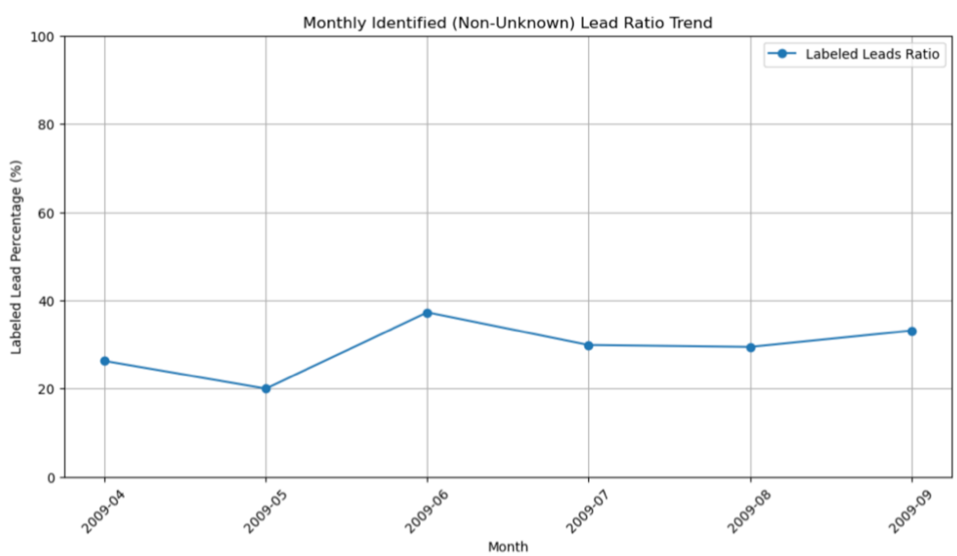
2.2 Lead Quality Trends & Key Observations

The trend analysis highlights several key insights. The **Identified Lead Ratio** remains consistently low, indicating a significant proportion of unclassified leads, which can hinder analytical precision and decision-making effectiveness. Over time, **High, Medium, and Low Quality Leads** do not exhibit a clear upward or downward trend, suggesting relative stability rather than consistent improvement or deterioration in lead quality. Additionally, **Low Quality**

Leadsconsistently outnumber **High Quality Leads**, pointing to potential inefficiencies in the lead sourcing and qualification processes that may require further optimization.

Below is a visualization of the monthly breakdown of lead quality trends:





2.3 Statistical Analysis Validation

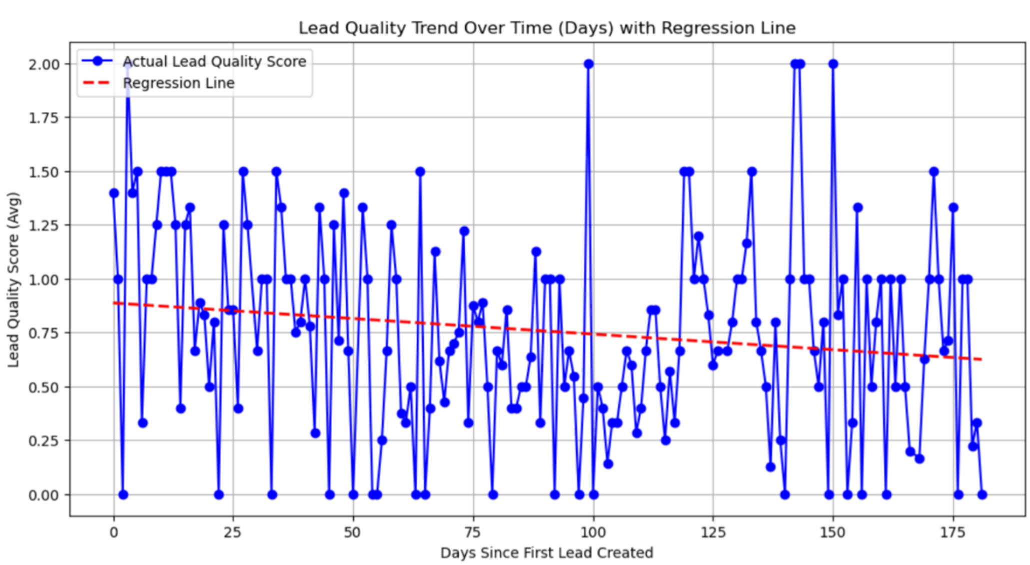
To further examine whether lead quality exhibits a statistically significant trend over time, a linear regression analysis was conducted. **This analysis specifically focuses on leads classified**

as High, Medium, or Low Quality, excluding Unknown leads to ensure a clearer evaluation of identifiable lead quality trends.

Below is the regression output detailing the statistical results:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      LeadQuality_Score   R-squared:                       0.014
Model:                            OLS   Adj. R-squared:                  0.013
Method:                 Least Squares   F-statistic:                     12.42
Date:                Fri, 21 Feb 2025   Prob (F-statistic):           0.000446
Time:                        16:45:26   Log-Likelihood:                -1120.6
No. Observations:                 881   AIC:                             2245.
Df Residuals:                     879   BIC:                             2255.
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               0.9075      0.060     15.222      0.000       0.791       1.025
LeadCreated_Num    -0.0021      0.001     -3.525      0.000      -0.003      -0.001
==============================================================================
Omnibus:                     9197.995   Durbin-Watson:                   1.885
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              117.377
Skew:                           0.535   Prob(JB):                     3.25e-26
Kurtosis:                       1.567   Cond. No.                         204.
==============================================================================
```

In this analysis, **LeadQuality_Score** is defined as **2 for High Quality, 1 for Medium Quality, and 0 for Low Quality**, allowing us to numerically represent lead quality levels. **LeadCreated_Num** represents the number of days since the earliest lead in the dataset, serving as a proxy for time progression. This setup enables us to determine whether lead quality has changed systematically over time.



The regression results show a coefficient of **-0.0021** for **LeadCreated_Num**, suggesting a very slight decline in lead quality as time progresses. While the **p-value of 0.000 (<0.05)** indicates statistical significance, the **R^2 value of 0.014** suggests that time accounts for only a small portion of the variation in lead quality. This means that while the decline is statistically significant, the extremely low R^2 value and small absolute coefficient suggest that the relationship is **not strong enough to be considered meaningful**.

2.4 Key Findings and Next Steps

In summary, the analysis indicates that **lead quality has remained relatively stable over time, with no substantial upward or downward trend**. While the regression analysis detected a statistically significant decline, the small coefficient magnitude and low R^2 value suggest that time alone does not meaningfully impact lead quality. Additionally, the **high proportion of unclassified leads** and the **consistent predominance of low-quality leads** highlight inefficiencies in the lead qualification process. These findings underscore the need for **improving lead classification mechanisms and refining sourcing strategies** to enhance overall lead quality.

Building on these insights, the next section will **delve into the key drivers of lead quality**, analyzing the critical factors that influence lead classification and performance. This deeper exploration will help identify actionable strategies to improve lead qualification and optimize lead generation efforts.

3. Statistical & Machine Learning-Based Variable Analysis

3.1 Chi-Square Test for Feature Significance

To determine which categorical variables have a statistically significant impact on lead quality, we conducted a chi-square test on five selected features: Widget Name, Debt Level, Partner, PublisherZone, and AdvertiserCampaign. These features were chosen based on their relevance to lead generation and potential influence on lead quality. The chi-square test assesses whether observed differences in lead quality distribution across these features are statistically significant.

The results indicate that **Widget Name, Debt Level, and Partner have strong statistical significance (p-values < 0.05)**, confirming their direct impact on lead quality, whereas PublisherZone and AdvertiserCampaign do not show statistical relevance. This insight guides further analysis in understanding the role of these key factors in driving high-quality leads.

Feature	Chi-Square Score	p-value	Significance
Widget Name	60.78	0.0001	Highly Significant
Debt Level	40.52	0.00005	Highly Significant
Partner	17.38	0.026	Significant
PublisherZone	1.28	0.526	Not Significant
AdvertiserCampaign	0.78	0.678	Not Significant

3.2 Feature Importance via Machine Learning

To complement the chi-square test, we employed Random Forest and XGBoost as our chosen machine learning models to determine feature importance. Random Forest was selected due to its ability to handle categorical features effectively and provide intuitive feature importance rankings through ensemble decision trees. XGBoost, on the other hand, was chosen for its robustness in handling complex relationships within the data and its strong performance in predictive modeling.

Both models identify Widget Name, Debt Level, and Partner as the top factors influencing lead quality.

Feature	Importance Score (RF)	Importance Score (XGBoost)
Widget Name	42.2%	23.0%
Debt Level	37.1%	21.3%
Partner	14.8%	19.8%
AdvertiserCampaign	4.3%	18.8%
PublisherZone	1.4%	17.1%

3.3 Feature Significant Conclusion

The results from both the chi-square test and machine learning analysis provide strong evidence that **Widget Name, Debt Level, and Partner are the key determinants of lead quality**. Given their high statistical significance and predictive importance, the following sections will conduct an in-depth analysis of these three factors to extract actionable insights and optimize lead generation strategies.

4. Factor-Specific Analysis

4.1 Widget Name Analysis

The analysis of widget performance reveals substantial variance in lead quality. Some widgets, such as **w-300250-DebtReduction1-1DC-BlueMeter (72.2%)** and **w-300250-DebtReduction1-1DC-CreditSolutions (52.2%)**, consistently produce a high proportion of high-quality leads, making them strong candidates for increased investment. These widgets demonstrate a balanced distribution of lead quality categories, with a notably low proportion of low-quality leads.

Conversely, several widgets, including **w-302252-DebtReduction1-1DC-yellowarrow (18.7%)** and **w-302252-DebtReduction1-1DC-yellowarrow-blue (18.7%)**, generate a disproportionately high number of low-quality leads, reducing their overall efficiency. These widgets exhibit a low High Quality Rate and a high Low Quality Rate, indicating inefficiencies in their ability to convert leads into valuable opportunities.

Overall, increasing investment in **high-performing widgets** while discontinuing or adjusting **low-performing widgets** is essential for improving lead quality and maximizing return on investment.

LeadQuality / WidgetName	High Quality	Low Quality	Medium Quality	TotalCount	HighQualityRate	HighQualityCount
w-300250-DebtReduction1-1DC-white	1.000000	0.000000	0.000000	1	1.000000	1
w-300250-DebtReduction1-1DC-BlueMeter	0.722222	0.222222	0.055556	18	0.722222	13
w-300250-DebtReduction1-1DC-CreditSolutions	0.521739	0.217391	0.260870	23	0.521739	12
w-300250-DebtReduction1-1DC-Head2	0.500000	0.363636	0.136364	22	0.500000	11
w-302252-DebtReduction1-1DC-yellowarrow-dark	0.400000	0.400000	0.200000	25	0.400000	10
w-302252-DebtReduction1-1DC	0.396552	0.448276	0.155172	58	0.396552	23
w-302252-DebtReduction1-1DC-white	0.387500	0.450000	0.162500	80	0.387500	31
w-300250-DebtReduction1-1DC	0.361702	0.404255	0.234043	94	0.361702	34
w-302252-DebtReduction1-1DC-CreditSolutions	0.337778	0.417778	0.244444	225	0.337778	76
w-300250-DebtReduction1-1DC-Head3	0.333333	0.583333	0.083333	12	0.333333	4
w-300250-DebtReduction1-2DC-CreditSolutions	0.300000	0.600000	0.100000	20	0.300000	6
w-300250-DebtReduction1-2DC-BlueMeter	0.285714	0.428571	0.285714	21	0.285714	6
w-302252-DebtReduction1-1DC-yellowarrow-blue	0.277778	0.425926	0.296296	54	0.277778	15
w-302252-DebtReduction1-1DC-yellowarrow	0.187500	0.250000	0.562500	16	0.187500	3

4.2 Debt Level Analysis

Debt levels significantly impact lead quality. Leads with a debt level between **$7,500 - $20,000** demonstrate **moderate to strong High Quality Rates (30.1% - 42.1%)**, indicating that this segment is a reliable target for optimized marketing efforts. In particular, the **$10,001 - $20,000** range achieves a **High Quality Rate of 42.1%**, making it one of the most promising categories for lead conversion.

Conversely, leads with debt levels exceeding **$100,000** exhibit the **lowest High Quality Rate (20.6%)** while simultaneously having the highest proportion of **low-quality leads (61.7%)**, suggesting a significant inefficiency in targeting this segment. This indicates that prospects in this category may be less engaged or less likely to convert.

Although the **$70,001 - $100,000** category has the **highest High Quality Rate (50.9%)**, the total number of leads in this range is relatively low, making it less statistically reliable for decision-making. The small sample size suggests that this result may not be as consistent as other segments with higher lead volumes.

Overall, targeting leads within the **$7,500 - $20,000** range presents the best opportunity for high-quality lead acquisition, whereas efforts should be minimized for leads with debt levels **above $100,000** to improve efficiency and return on investment.

LeadQuality	High Quality	Low Quality	Medium Quality	TotalCount	HighQualityRate	HighQualityCount
DebtLevel						
70001-100000	0.509434	0.320755	0.169811	53	0.509434	27
10001-20000	0.420732	0.323171	0.256098	164	0.420732	69
30001-50000	0.375000	0.413462	0.211538	104	0.375000	39
50001-70000	0.366667	0.366667	0.266667	60	0.366667	22
20001-30000	0.338983	0.389831	0.271186	118	0.338983	40
7500-15000	0.301471	0.544118	0.154412	136	0.301471	41
More_than_100000	0.205882	0.617647	0.176471	34	0.205882	7

4.3 Partner Analysis

Partner performance plays a crucial role in lead quality. **Adknowledge** demonstrates the highest High Quality Rate (46.7%) among scalable partners, making it a strong candidate for increased investment. Its relatively balanced distribution between high, medium, and low-quality leads suggests it is a consistent performer in generating valuable leads.

Google and Yahoo generate the highest total lead volumes, with **325 and 230 leads**, respectively. Although their High Quality Rates (37.8% for Google and 32.1% for Yahoo) are lower than that of Adknowledge, this is expected given the larger scale of leads they generate. Instead of shifting all investment to Adknowledge, maintaining a strong presence on Google and Yahoo remains essential to sustaining lead volume. The lower High Quality Rate is a natural consequence of the scale, and these platforms still contribute significantly to overall lead acquisition. However, if a downward trend in lead quality is observed over time, further analysis should be conducted to determine whether strategic adjustments, such as refined targeting or budget reallocation, are necessary.

Overall, **Adknowledge presents the best balance between quality and volume, making it the strongest investment candidate. Meanwhile, Google and Yahoo should remain key acquisition channels due to their scale, with continuous monitoring to ensure lead quality remains stable over time.**

LeadQuality Partner	High Quality	Low Quality	Medium Quality	TotalCount	HighQualityRate	HighQualityCount
advertise.com	1.000000	0.000000	0.000000	1	1.000000	1
adknowledge	0.466667	0.288889	0.244444	45	0.466667	21
call_center	0.382353	0.352941	0.264706	68	0.382353	26
google	0.378462	0.446154	0.175385	325	0.378462	123
yahoo	0.321739	0.408696	0.269565	230	0.321739	74

5. Lead Quality Improvement & ROI Impact

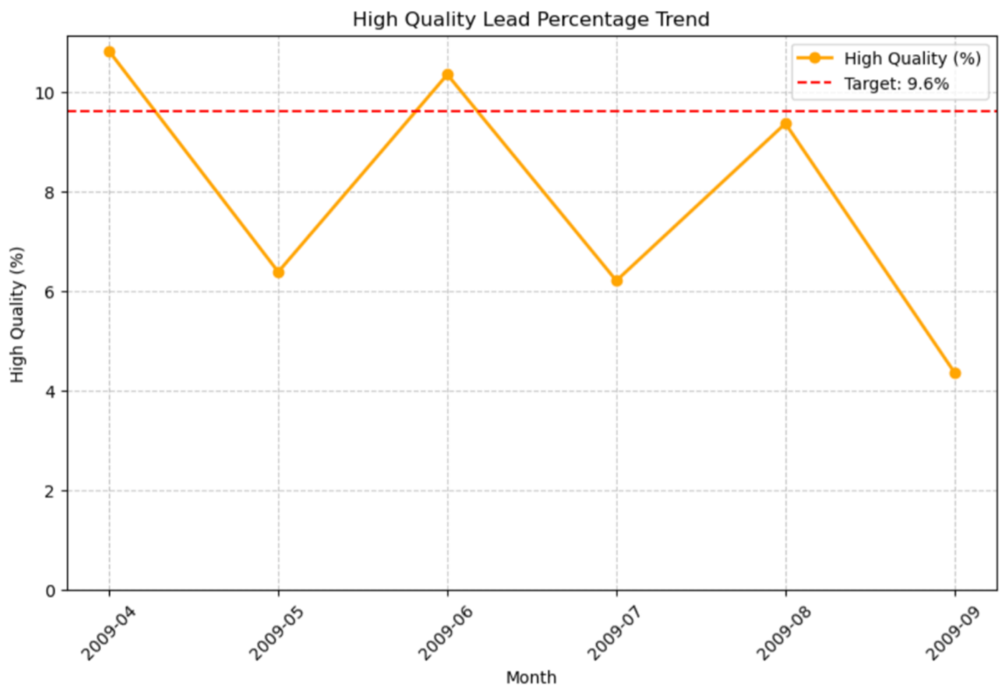
The advertiser has proposed increasing the **Cost Per Lead (CPL) by 20%** (from **$30 to $33**) if the **Lead Quality improves by 20%** (from **8.0% to 9.6%**). To evaluate the feasibility of this proposal, we analyze historical lead quality trends, assess the potential for achieving the 9.6% target, and calculate the corresponding **Return on Investment (ROI)** to determine whether this trade-off is financially viable.

5.1 Feasibility of Achieving 9.6% Lead Quality

Based on historical data, the highest observed Lead Quality rate reached **10.81%**, indicating that achieving the **9.6% target** is feasible. However, this requires targeted optimizations, particularly in reducing the proportion of "Unknown" leads and improving classification accuracy.

Based on the available data, the total number of leads, after filtering out "Unknown" Lead Quality cases, is **881**. Among these, the estimated number of High Quality Leads is **71.45**, resulting in a **current High Quality Rate of 8.11%**. This serves as the baseline for evaluating the feasibility of achieving the proposed 9.6% target.

	Month	High Quality (%)	Low Quality (%)	Medium Quality (%)	Unknown (%)	Total Leads
0	2009-04	10.810811	11.003861	4.440154	73.745174	518
1	2009-05	6.379310	9.827586	3.793103	80.000000	580
2	2009-06	10.344828	23.152709	3.776683	62.725780	609
3	2009-07	6.203008	20.676692	3.007519	70.112782	532
4	2009-08	9.360731	15.296804	4.794521	70.547945	438
5	2009-09	4.360465	16.279070	12.500000	66.860465	344



5.2 ROI Analysis & Advertiser Proposal Evaluation

To determine the financial impact of this proposal, we calculate ROI under both the **current scenario (CPL = $30, Lead Quality = 8.11%)** and the **proposed scenario (CPL = $33, Lead Quality = 9.6%)**.

5.2.1 ROI Calculation Assumptions & Data Sources

- **Total Cost ($)** = CPL × Total Leads (Advertiser Proposal) → Assumption: Total Leads remain constant (881)
- **Total Revenue ($)** = High Quality Leads × Revenue per Converted Lead → Assumption: Revenue per Converted Lead is fixed at $500
- **ROI (%)** = (Total Revenue - Total Cost) / Total Cost × 100 → Assumption: ROI is a static calculation, ignoring potential long-term effects

	CPL ($)	High Quality (%)	Total Leads	High Quality Leads	Total Cost ($)	Total Revenue ($)	ROI (%)
0	30	8.11	881	71.4491	26430	35724.55	35.166667
1	33	9.60	881	84.5760	29073	42288.00	45.454545

5.2.2 Conclusion

Since **ROI increases from 35.17% to 45.45%**, the proposal is **financially viable** and supports the feasibility of targeting a 9.6% Lead Quality rate through strategic optimizations. To successfully achieve the goal, combining the **factor analysis from the previous section**, we will implement **targeted strategies** in the next section to optimize **widget selection, debt level targeting, and partner allocation**, ensuring a sustainable improvement in lead quality and maximizing return on investment.

6. Strategy and Recommendations

Based on the factor analysis, several strategic recommendations can be made to optimize lead quality and improve return on investment. To successfully achieve the **9.6% Lead Quality rate**, the following strategies should be considered:

6.1 Optimizing High-Performing Widgets

Investment should be **increased** in high-performing widgets that consistently generate high-quality leads (e.g., w-300250-DebtReduction1-1DC-BlueMeter and w-300250-DebtReduction1-

1DC-CreditSolutions). These widgets have demonstrated **High Quality Rates above 40%**, making them reliable for sustaining lead improvements.

Conversely, **low-performing widgets** with High Quality Rates below 30% should be **discontinued or re-optimized** to minimize inefficiencies. This targeted allocation of resources will enhance **overall lead quality and improve cost-effectiveness**, ensuring that marketing investments are directed towards the most productive channels.

High-Performing Widgets (Investment Recommended)	High Quality Rate (%)
w-300250-DebtReduction1-1DC-BlueMeter	72.2%
w-300250-DebtReduction1-1DC-CreditSolutions	52.2%
w-300250-DebtReduction1-1DC-Head2	50.0%
w-302252-DebtReduction1-1DC-yellowarrow-dark	40.0%

Low-Performing Widgets (Recommended for Removal due to Low High Quality Rate and Small Volume)	High Quality Rate (%)	Total Lead Count
w-300250-DebtReduction1-1DC-Head3	33.3%	12
w-300250-DebtReduction1-2DC-CreditSolutions	30.0%	20
w-300250-DebtReduction1-2DC-BlueMeter	28.6%	21
w-302252-DebtReduction1-1DC-yellowarrow-blue	27.8%	54
w-302252-DebtReduction1-1DC-yellowarrow	18.7%	16

6.2 Targeting Optimal Debt Levels

For **debt level targeting**, the optimal range for lead acquisition lies between **$7,500 - $20,000**, as these segments show the highest High Quality Rates, particularly **$10,001 - $20,000 (42.1%)**. Leads with debt levels exceeding **$100,000** exhibit the lowest High Quality Rate (20.6%) and the highest proportion of low-quality leads (61.7%), making them inefficient for conversion. While **$70,001 - $100,000** has a high High Quality Rate (50.9%), its total lead count is too low to be a reliable reference. As a result, marketing efforts should focus on **mid-range debt levels** while reducing investment in leads with excessively high debt.

Debt Level Category	High Quality Rate (%)	Recommendation
$10,001 - $20,000	42.1%	Increase investment
$7,500 - $10,000	30.1%	Maintain investment
$70,001 - $100,000	50.9%	Sample size too low to prioritize
More than $100,000	20.6%	Reduce investment

6.3 Refining Partner Allocation

Regarding **partner performance**, **Adknowledge** demonstrates the highest High Quality Rate (46.7%) among scalable partners, making it a strong candidate for increased investment. However, **Google and Yahoo**, despite their lower High Quality Rates (37.8% and 32.1%, respectively), generate the largest lead volumes. Instead of shifting all investment to Adknowledge, maintaining a strong presence on Google and Yahoo is crucial to sustaining lead volume, as their lower High Quality Rate is a natural consequence of their scale. If a decline in lead quality is observed over time on these platforms, further analysis should be conducted to determine whether strategic adjustments, such as refined targeting or budget reallocation, are necessary.

Partner	High Quality Rate (%)	Total Lead Volume	Recommendation
Adknowledge	46.7%	45	Increase investment
Google	37.8%	325	Maintain, monitor lead quality
Yahoo	32.1%	230	Maintain, monitor lead quality

6.4 Expected Impact

By implementing these strategies, we expect to achieve **higher lead quality and improved conversion rates** through the optimization of widget selection. Allocating resources towards high-performing widgets while discontinuing low-quality ones will enhance overall lead efficiency. Additionally, these adjustments will lead to **increased cost efficiency** by reducing expenditures on underperforming sources, ensuring that marketing investments yield better returns. Furthermore, refining partner selection and targeting optimal debt levels will contribute to a **stronger return on investment (ROI)** by focusing efforts on segments with the highest conversion potential. To sustain these improvements, continuous performance monitoring and **data-driven adjustments** will be essential in refining lead acquisition strategies and maintaining long-term marketing effectiveness.