RSM Erasmus University

Rotterdam School of Economics
Rotterdam, Block III, Spring 2021

# Evaluating the Search Engine Effectiveness of Cloverleaf

*Analysis of current search engine set-up and recommendations for future keyword optimization*

**593088ah**

**Digital Marketing Strategy**

**Evaluator: Xi Chen**

# Table of Contents

# Executive Summary

The report is segmented into 2 parts, the important findings are summarized as follows:

**Evaluation of Effectiveness**

- CTR of 10.56% and CR of 4.72% are well above benchmark and illustrate the good results of the search engine campaigns in 2012
- Positive correlation between ad rank with CTR and CR strengthens the previous argument. The selected keywords coupled with the ads are also leading to generally good ad ranks, because they fulfill Googles Quality Criteria
- ROI analysis shows that despite good rankings, CTR, and CR the overall search engine effort in 2012 was not profitable and led to a ROI of just 46.2%
- Further research indicates that Cloverleafs ROI in 2012 is below benchmark

**Recommendations for Keyword Optimization:**

- To improve CR it is recommended to use keywords with three keywords and the respective retailer name, to achieve above-average CRs
- To improve CTR it is recommended to use keywords with two keywords, and to include the retailer and brand name, to achieve above-average CTRs
- Same keyword strategy also maximizes the ROI and helps to gain similar results that are claimed by eCommerce benchmarks

**General Remarks:**

- Results are based on Cloverleafs search engine marketing data, which ranges from the 16th of January, 2012 to the 3rd of December 2012
- Preprocessing bridged the gap between minimizing bias in the data and maximizing the number of observations

# Assumptions and Preprocessing

The objective of this paragraph is to streamline assumptions that were necessary to process Cloverleafs search engine marketing data. Because the given conclusions in this report are directly linked to the data quality and the associated preprocessing assumptions, I will shortly outline which detailed preprocessing steps were taken.

## Interaction Performance

Click-Through Rates (CTR) and Conversion Rates (CR) are because of their calculation subject to possible bias. If the given impressions or clicks are zero, the denominator of both indicators will results in an error. Additionally, if the rank of a specific ad equals to zero, this observation will also be discarded from the analysis. The reasoning behind the latter it is that only active ads, which were part of the official bidding process receive a rank from one to infinity. To circumvent calculation errors that bias results either up- or downward, we apply the following rules, leading to a data set of 319 observations.

**Click-Through Rates (CTR):**

1. If clicks = 0 and impressions > 0, CTR will be set to 0
2. If clicks = 0, impression = 0, or adrank = 0, the observations will not be included

**Conversion Rates (CR):**

1. If clicks = 0 and conversions > 0, CR will be set to 0
2. If clicks = 0, impression = 0, or adrank = 0, the observations will not be included

Despite the focus on “paid search” we left observations with a bidprice = 0 in the data, because those either had positive clicks, impressions, conversions or revenues.

## Financial Performance

Similar reasoning can be applied to the calculation of Return on Investment (ROI) or Return on Ad Spend (ROAS) metrics. On top of the previous rules, we applied the following guideline, leading to a data set of 158 observations.

1. If clicks = 0 and bidprice = 0, the observations will not be included

# Evaluation of Effectiveness

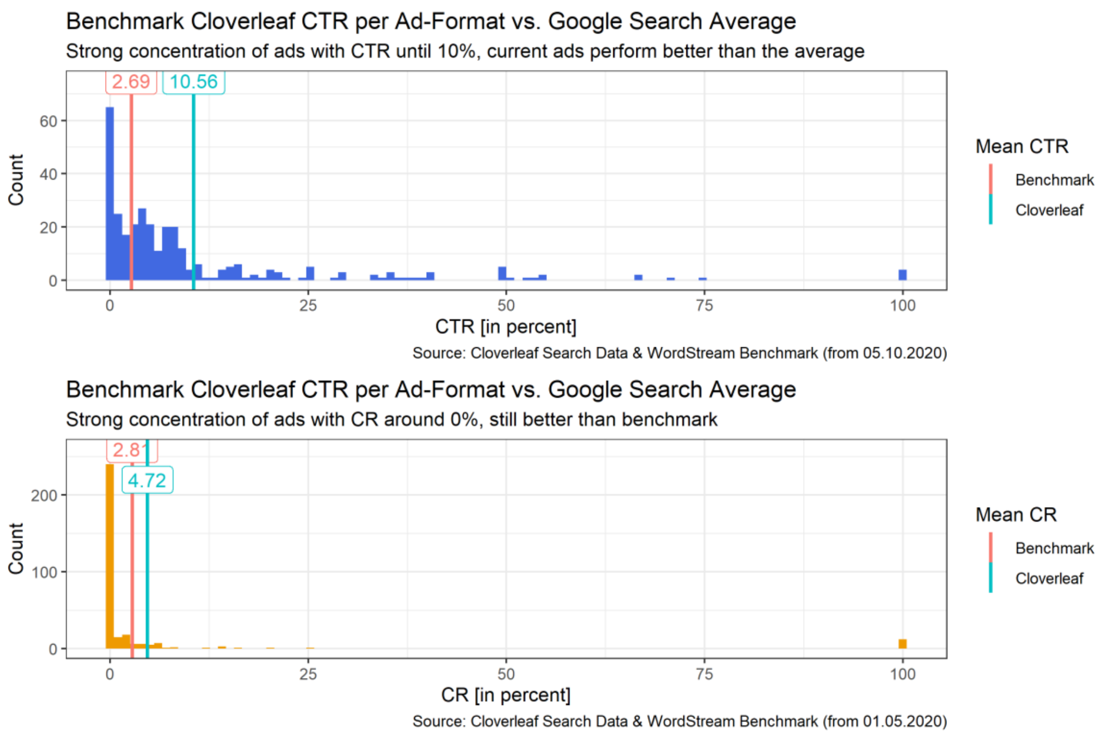
This part analyses the past search engine marketing performance using CTR, CR, ROI and specific industry benchmarks to gauge Cloverleaf's keyword marketing effectiveness. Furthermore, we will also take a look at the correlation between the eventual ad rank of a specific ad to analyze how much CTR and CR contribute to the eventual rank placement in the bidding process.

## CTR, CR, and Benchmark Performance (Q1.1)

Considering the preprocessing steps outlined before, the aggregated results on data set level are 10.56% for CTR and 4.72% for CR. This means, that on average, 10.56% of users who see an ad, that was used in the one-year span of the search engine marketing campaign, also click on it. Moreover, of those users who clicked the ad, 4.72% also convert in accordance to the conversion goals.

Benchmark figures from WordStream confirm, that the usual industry average CTR and CR for eCommerce websites and shops are significantly lower (WordStream, 2020b). The nature of different brands, businesses, and products in the eCommerce industry, makes it difficult to build benchmark figures (WordStream, 2020a). Despite the differences, no official benchmarks figures above a CTR of 2.81% or a CR of 3.93% could be identified (StoreGrowers.com, 2021). All previous sources considered Google Ads and the Google search engine exclusively for their figures, which is crucial considering the origin of the given search engine data. These common findings indicate that Cloverleafs current strategy is solid in terms of CTR and CR, as illustrated in Graph 1.

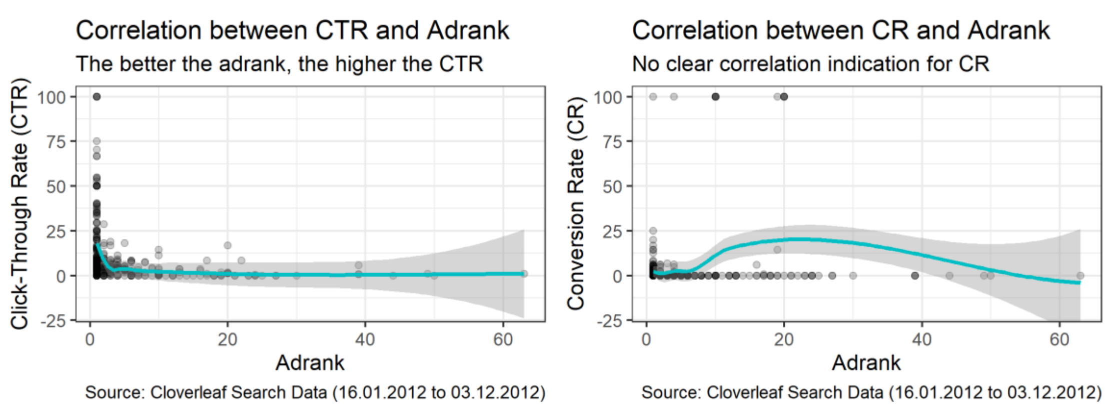
Taking a closer look at the distribution of the CTR metrics, one can observe that there is a strong concentration around 0%, with relatively many occurrences of ads with up to 10% in CTR. A less diverse distribution can be observed for the CR, which shows an expressively more ads around 0% CR. Nevertheless, both metrics show occurrences until the 100% mark, as shown in Graph 1.




*Graph 1: CTR and CR Benchmarking of Cloverleaf in the eCommerce Space*

## Correlation Analysis (Q1.2)

We analyze the correlation between the ad rank with the respective performance metrics, to understand to what extent CTR and CR drive the eventual ad rank, that is set by the keyword auction. Graph 2 shows, that just looking at the data, an inference about the relationship is only possible for CTR, but not for CR:



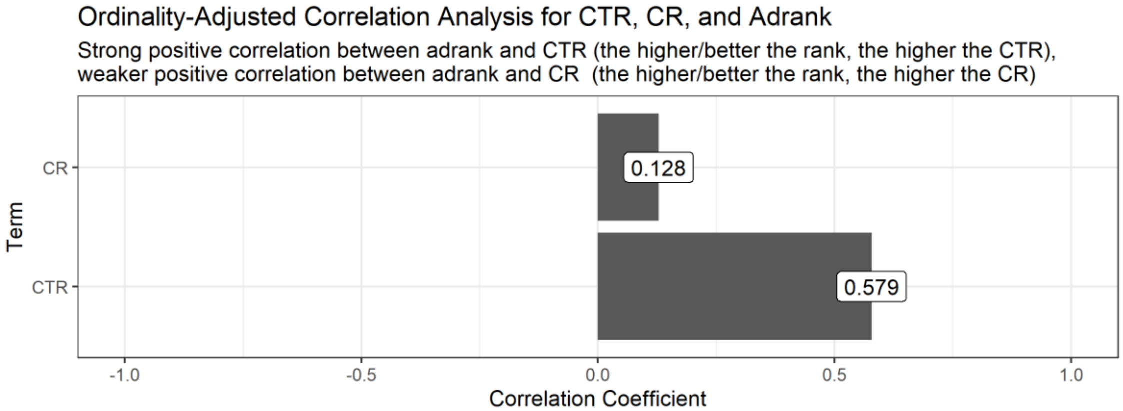

*Graph 2: Correlation Analysis between Ad Rank, CTR, and CR*

The latter shows a possibly misleading indication, meaning that ads that were ranked between the 8$^{th}$ and 20$^{th}$ place perform better in terms of CR, than ads with a higher / better rank (1$^{st}$ to 7$^{th}$). A closer look into Graph 3 and the Spearman correlation creates a more unified picture, by accounting for the ordinal character of the ad rank variable.

Because of the adjustment for the rank variable, we see that both CTR and CR are positively correlated with the resulting ad rank. In other words, the higher CTR or CR of the respective ad, the better / higher is also the rank of the ad, and vice versa.

In case of the CTR, this makes perfect sense. Google uses a myriad of factors to set the eventual PPC that needs to be paid per click. Beside the CTR, also the quality of the ad, quality of the landing page, keyword relevance, historical performance etc. matters. This helps users to identify products and services they need, but also rewards marketers, because their offerings are demanding. In consequence, this leads to an increase in search engine usage. A positive correlation of 0.579 means that the CTR has a considerable effect on ad rank. This is not surprising because CTR is among the listed features the one most user-oriented metric and therefore supports the analysis of effective and growing search engine usage for users. The higher the average CTR of ads and content the tailor-made is the content to a user's needs. The overall intensity also means that combined with the (on average) higher-than-benchmark CTR that Cloverleafs past strategy yielded good results.

The CR is also positively correlated, which makes sense because CR is a good indicator for ad, landing page quality, and general search relevance. These factors in turn are part of Googles Ad Score and used to determine the eventual ad rank. But, CR measures actual conversions, which rely heavily on the aforementioned characteristics and are influenced directly by the landing and web page, less by the ad. Therefore, a lower correlation intensity is logical, especially because not every click leads to a complete conversion.




*Graph 3: Spearman Correlation Results between Ad Rank, CTR, and CR*

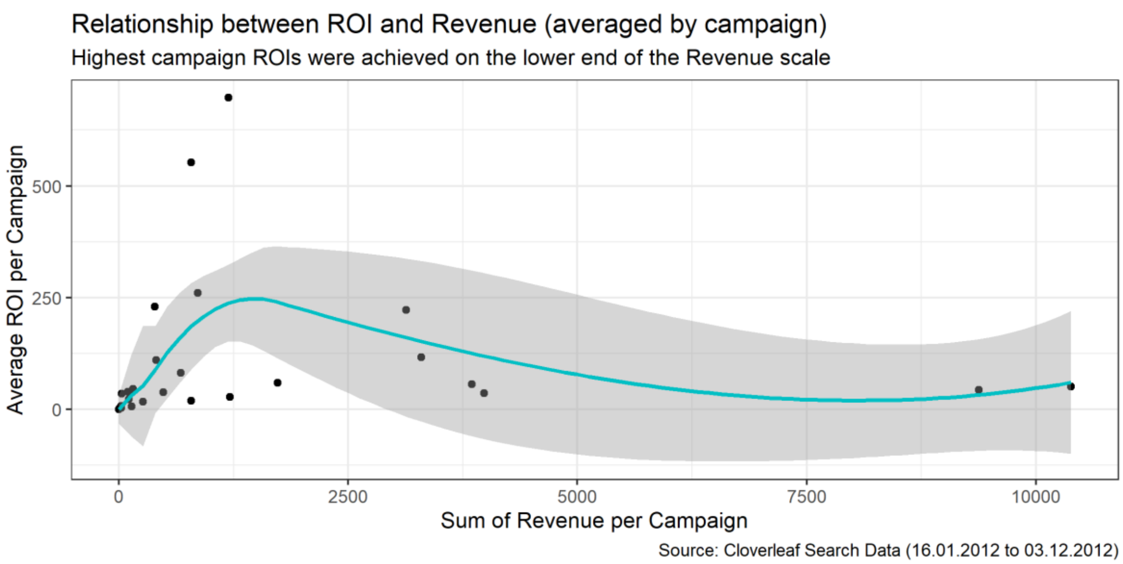
## Measuring Financial Effectiveness with ROI (Q1.3)

If considering correlation, CTR, CR, and benchmarks we arrived at the conclusion that the search engine marketing strategy led to good results. Nevertheless, we did not account for the financial side of the equation. For this we compute the ROI to assess the financial gain Cloverleaf got from the ads and campaigns. Because we assume a profit margin of 3.5%, we calculate ROI using:

- Costs = Bidprice * Clicks
- Profit = Revenue * Profit Margin
- ROI = Profit / Costs

Based on a given profit margin, we assume that costs are already included in the profit margin figure. This is also the usual standard in marketing and finance alike. In a traditional profit and loss statement (P&L) it is assumed that marketing costs are part of the “Selling, General, and Administrative Expenses” (SG&A), and therefore included in the profit margin. Doing otherwise would result in partly negative ROI, which is by definition not possible because the measure is normalized from zero to infinity.

For the underlying ROI calculation we computed ROI on the most detailed level: the ad level (= observation level). Then, we aggregated all results on campaign level and aggregated in the last step on data set level. This gradual calculation ensures that we keep the maximum number of observations, which is especially important for later keyword optimization suggestions. The average ROI is with 46.2% below 100%, meaning that from the invested marketing budget, only 46.2% were retrieved back in form of returns. In other words, the search engine keyword marketing campaign consumed more budget than it earned, despite the good results described in previous parts. ROI benchmarks for eCommerce shops in industries, such as apparel, jewelry, mass merchants, office supplied and sporting goods range from 297% to 607%, making the current ROI results far below the usual benchmark (StoreGrowers.com, 2021). We also observe that high ROI is not correlated with high revenue, as shown in Graph 4. After reaching a maximum ROI, the average sum of revenue per campaign decreases.




*Graph 4: Relationship between ROI and Ad Revenue*

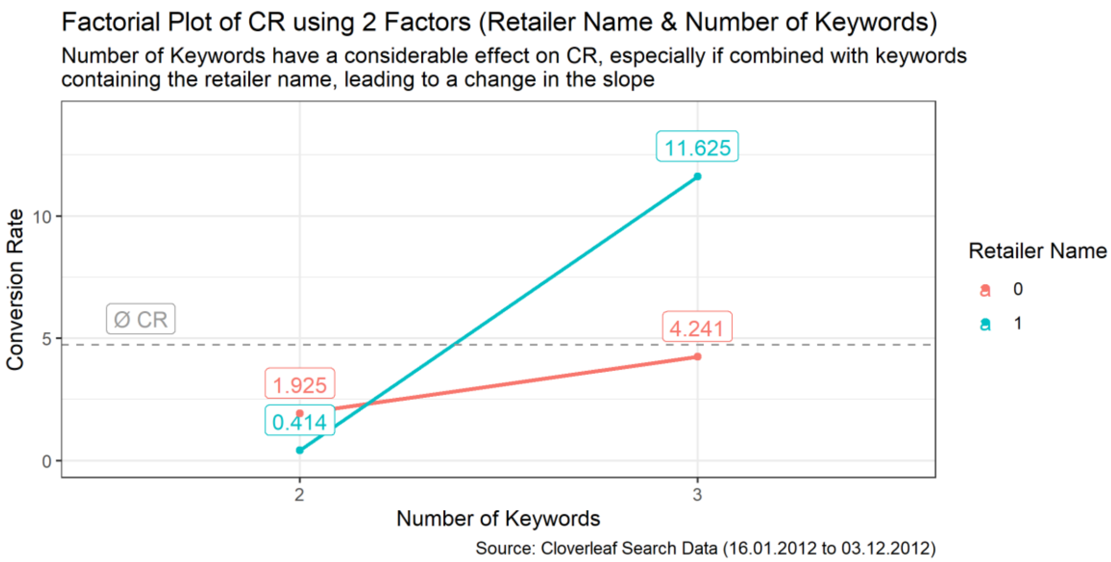
# Recommendations for Keyword Optimization

For suitable strategy changes and keyword optimization, we must study the effect of factors we can directly influence in our search engine set-up. Specifically, we must understand how CTR, CR and ROI are driven by differences in number of keywords, and the inclusion of retailer or brand names. This report focuses on factor interaction and does not consider statistical significance or reliability, those issues are outlined in the Appendix.

## Effect of Keyword Length and Retailer Names on CR (Q2.1)

Once we split the effect of retailer names and keyword length on CR, we can observe interesting effects in this setting. Firstly, the longer the keyword length, the higher the achieved CR on average. Secondly, the inclusion of retailer names leads to significantly higher CR than their omission. Thirdly, the interaction between retailer names and keyword length is considerable, because the slope changes also leading to a CR higher than the average CR, calculated across all ads.

Based on these interactions, the first strategy assumption would be to exclusively setup keywords including three keywords and also to include the retailer name. This is the only way to increase the average CR across all campaigns, because all other factor combinations lead to below-average CR performance, as seen in Graph 5.



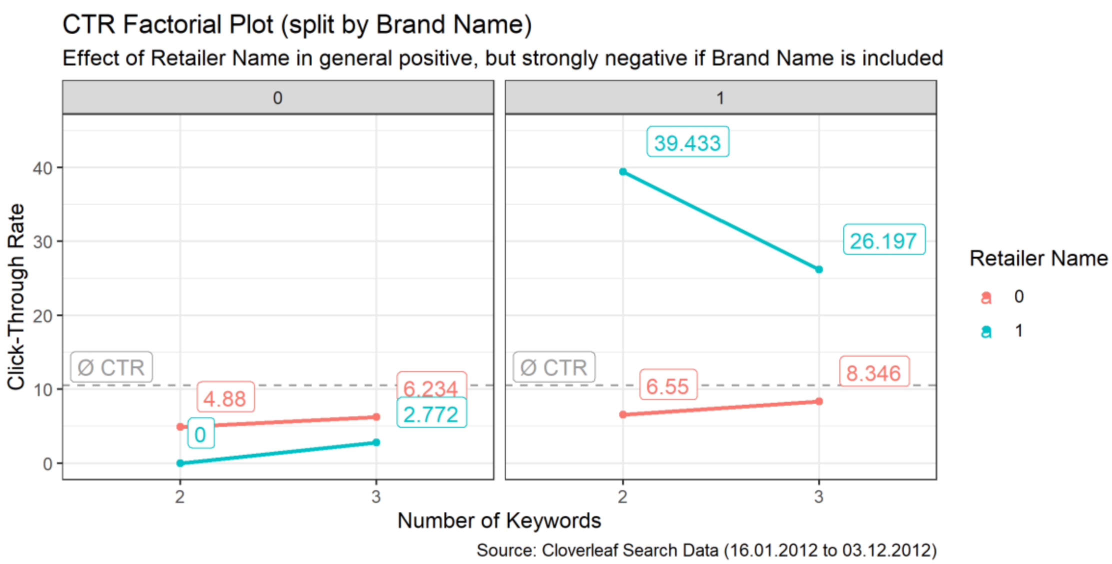

*Graph 5: Effect of Retailer Name and Number of Keywords on CR*

## Effect of Length, Retailer, and Brand Names on CTR (Q2.2)

The results from the factorial table, were illustrated in an interaction plot to make the results more comprehensible. The concrete recommendations to increase CTR are threefold: Firstly, to not use any keywords that do not include the brand name, because they always lead to a below-average CTR. Secondly, to include the retailer name, because this is the main contributor to above-average CTR. Thirdly, only using two keywords as this leads to higher CTR on average than listings with three keywords. For specific numbers and interactions, refer to Graph 6 and Table 1.

| Average CTR | | | | | |
|---|---|---|---|---|---|
| With Retailer Name (=1) | | | Without Retailer Name (=0) | | |
| Keyword Length | With Brand (=1) | Without Brand (=0) | Keyword Length | With Brand (=1) | Without Brand (=0) |
| 2 | 39.433% [1)] | 0% [1)] | 2 | 6.550% | 4.880% |
| 3 | 26.197% [1)] | 2.772% [1)] | 3 | 8.346% | 6.243% |

*Table 1: Overview of Retailer, Brand Name and Keywords Interaction on CTR*




*Graph 6: Effect of Retailer, Brand Name and Keywords on CTR*

## Effect of Length, Retailer, and Brand Names on ROI (Q2.3)

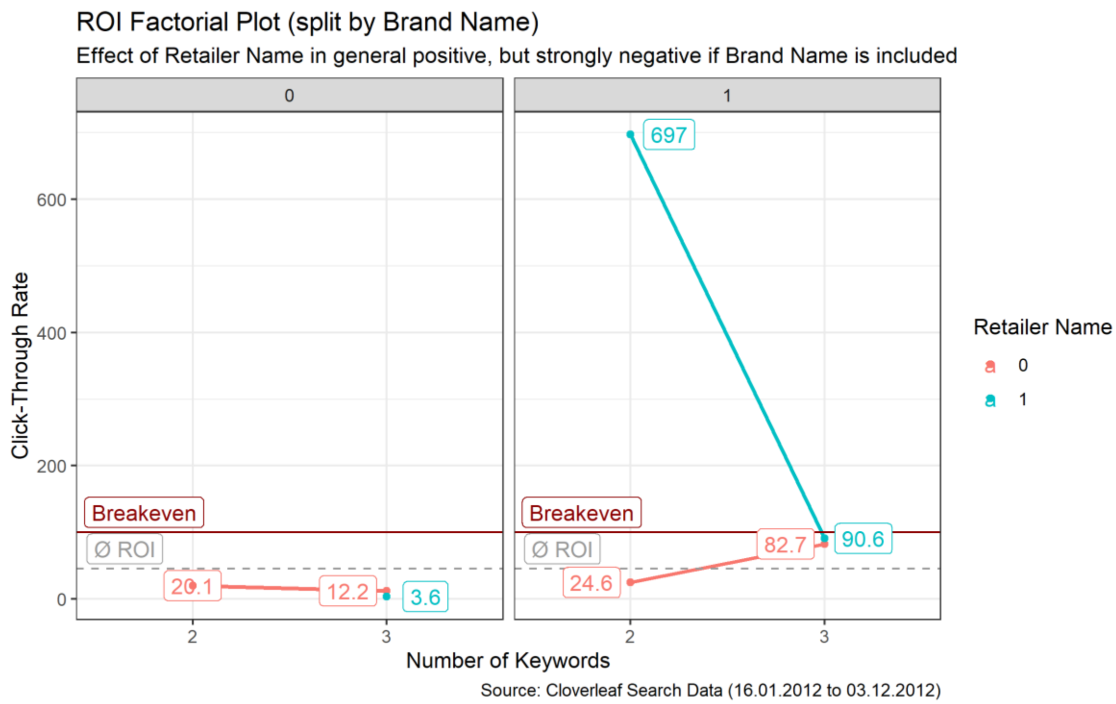
The concrete recommendations to increase ROI are threefold and similar to the recommendations for CTR increase: Firstly, to not use any keywords that do not include the brand name, because they always lead to a below-average ROI and even below-breakeven returns. Secondly, to include the retailer name. This time it is not the main contributor to the difference and only works well if combined with the last recommendation. Thirdly, only using two keywords as this leads to higher ROI on average

[1] See Appendix C for more relevant information, regarding the reliability of these figures.

than listings with three keywords. This interaction between two keywords, retailer and brand name inclusion is the only combination that leads to ROI that is on par with the industry benchmark. It is also the only combination that leads to above-breakeven ROI. This strategy is exactly in line with the previous optimization for CTR. For specific numbers and interactions, refer to Graph 7 and Table 2.

| **Average ROI** | | | | | |
|---|---|---|---|---|---|
| With Retailer Name (=1) | | | Without Retailer Name (=0) | | |
| Keyword Length | With Brand (=1) | Without Brand (=0) | Keyword Length | With Brand (=1) | Without Brand (=0) |
| 2 | 697% 2) | - | 2 | 24.6% | 20.1% |
| 3 | 90.6% 2) | 3.6% 2) | 3 | 82.7% | 12.2% |

*Table 2: Overview of Retailer, Brand Name and Keywords Interaction on ROI*




*Graph 7: Effect of Retailer, Brand Name and Keywords on ROI*

[2] See Appendix C for more relevant information, regarding the reliability of these figures.

# Sources

StoreGrowers.com. (2021, January 18). *30+ Ecommerce Metrics Benchmarks (2021) - Store Growers*. https://www.storegrowers.com/ecommerce-metrics-benchmarks/#cvr

WordStream. (2020a, May 1). *Google Shopping Ads Benchmarks: Average CPC, CTR, Monthly Budget, & More* | *WordStream*. https://www.wordstream.com/blog/ws/2019/04/01/shopping-ads-benchmarks

WordStream. (2020b, October 5). *Google Ads Benchmarks for Your Industry* | *WordStream*. https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks

# Appendix

## Appendix A: Data Structure

The given analyses change between three different levels of aggregation. Generally every calculation is computed on the lower level. Summaries, such as averages, are build on different aggregation levels, depending on the desired output

1. Ad-Level: This is the level with the most detail and includes ads and keywords using different formats. If not specified otherwise, this level is used for the given calculations, especially because impressions, clicks, conversions, and revenue can change from ad to ad, even if it is part of the same campaign
2. Campaign Level: First aggregation level and only applied for preparing the ROI results. Despite not being necessary, it gives a better overview on how the average ROI can be compared with the average revenue (see Graph 4).
3. Data Set Level: For all averages that refer to the overall CTR, CR, or ROI, this aggregation level was used.

## Appendix B: Data Reliability

According to statistical theory, reliable predictions for future scenarios should only be included once a certain number of observations are included per group or cohort. In both cases, for the optimization of CTR and ROI, specific groups do not have enough observations to make a robust judgement. The applied threshold for marking recommendations as possibly erroneous is at < 30 observations. In the tables below we summarize the number of observations and mark the previous strategy recommendation in bold, to highlight if the eventual recommendation was affected by the analysis.

This analysis is not meant to undermine the described results in Q2.2 and Q2.3, but just to give the view, that often marketing decisions need to balance data reliability and costs for keyword and campaign diversification that make it possible to include enough data points per eventual group.

| Number of Observations per Group: Average CTR | | | | | |
|---|---|---|---|---|---|
| With Retailer Name (=1) | | | Without Retailer Name (=0) | | |
| Keyword Length | With Brand (=1) | Without Brand (=0) | Keyword Length | With Brand (=1) | Without Brand (=0) |
| 2 | **4** | 4 | 2 | 91 | 47 |
| 3 | 23 | 13 | 3 | 58 | 34 |

*Table 3: Number of Observations per Group, CTR Calculaton*

| Number of Observations per Group: Average ROI | | | | | |
|---|---|---|---|---|---|
| With Retailer Name (=1) | | | Without Retailer Name (=0) | | |
| Keyword Length | With Brand (=1) | Without Brand (=0) | Keyword Length | With Brand (=1) | Without Brand (=0) |
| 2 | **1** | 0 | 2 | 67 | 22 |
| 3 | 10 | 4 | 3 | 22 | 15 |

*Table 4: Number of Observations per Group, ROI Calculaton*

## Appendix C: Software

The underlying data and graphs were implemented using R and the Integrated Development Environment (IDE) RStudio. All imported functionalities and packages can be seen in the attached script in Appendix D.

# Appendix D: R Script

```
# libraries
library(xlsx) # excel import
library(dplyr) # data preprocessing and data wrangling
library(corrr) # simple correlation analysis
library(tsibble) # dealing with time-series tibbles
library(ggplot2) # general plotting
library(patchwork) # side-by-side plotting
library(AlgDesign) # first stage interaction plot
library(tidyverse) # advanced data wrangling

# load the data
sdat <- read.xlsx("cloverleaf_search_data.xlsx",
                  sheetIndex = 1,
                  as.data.frame = T) %>%
  mutate(ID = seq(1:nrow(.)))

# transform to homogeneous date formats
character_dates <- sdat %>%
  filter(grepl("/", datestring)) %>%
  mutate(date = as.Date(datestring, format = "%m/%d/%Y"))

numeric_dates <- sdat %>%
  filter(!grepl("/", datestring)) %>%
  mutate(date = as.Date(as.numeric(datestring), origin = "1899-12-30"))

# combine data sources
sdat <- character_dates %>%
  full_join(numeric_dates) %>%
  arrange(ID) %>%
  select(-datestring, -ID) %>%
  relocate(date, .before = "advertID") %>%
  mutate(id = seq(1:nrow(sdat)))


# ADVANCED PRE-PROCESSING
# transform to long data format to facilitate comprehensive analysis
sdat_longer <- sdat %>%
  select_if(is.numeric) %>%
  select(-advertID) %>%
  pivot_longer(cols = c(impressions, clicks, bidprice, conversions, numberofwords,
                        retailer, brandname, adQuality, landQuality, revenue, adrank),
    names_to = "variables",
    values_to = "value")

# check distribution for all variables
sdat_longer %>%
  ggplot(aes(value)) +
  stat_density() +
  labs(title = "Distribution of all Variables",
       x = "Values",
       y = "Density",
       caption = "Source: Cloverleaf Search Data & WordStream Benchmark (from 01.05.2020)") +
  facet_wrap(~variables, scales = "free") +
  theme_bw()

ggsave("01_app_variables.png", width = 8, height = 6)
```

```r
# check distribution for quality and rank variables
sdat_longer %>%
  filter(variables %in% c("adQuality", "landQuality", "adrank", "bidprice")) %>%
  ggplot(aes(value)) +
  geom_histogram(bins = 60) +
  labs(title = "Close Look into Quality, Rank, and Price Variables",
       subtitle = "Quality and rank variables seem to include zeroes, indicating possible data 
errors",
       x = "Values",
       y = "Count",
       caption = "Source: Cloverleaf Search Data & WordStream Benchmark (from 01.05.2020)") +
    facet_wrap(~variables, nrow = 2, scales = "free") +
  scale_x_continuous(n.breaks = 6) +
  theme_bw()

ggsave("02_app_variables.png", width = 8, height = 6)

# check distribution for clicks and impression
sdat_longer %>%
  filter(variables %in% c("clicks", "impressions")) %>%
  ggplot(aes(value)) +
  geom_histogram(bins = 300) +
  labs(title = "Close Look into Clicks and Impressions",
       subtitle = "High concentration of zero clicks and impressions that may impair further 
calculations",
       x = "Values",
       y = "Count",
       caption = "Source: Cloverleaf Search Data & WordStream Benchmark (from 01.05.2020)") +

  facet_wrap(~variables, nrow = 2, scales = "free") +
  scale_x_continuous(n.breaks = 6) +
  theme_bw()

ggsave("03_app_variables.png", width = 8, height = 6)

# TASK 1.1: CTR AND CR ANALYSIS
# overall CTR and CR
sdat <- sdat %>%
  mutate(CTR = ifelse(is.nan(clicks/impressions), # correct for zero cases
                      0,
                      (clicks/impressions)) * 100,
         CR = ifelse(is.nan(conversions/clicks), # correct for zero cases
                     0,
                     (conversions/clicks)) * 100)

# delete entries for which the identified criterias hold true
aj <- sdat %>%
  filter(clicks == 0 & impressions == 0 | adrank == 0)

# delete data that may lead to bias
sdat <- sdat %>%
  anti_join(aj)

metrics_mean <- sdat %>%
  summarise(mean_CTR = mean(CTR),
            mean_CR = mean(CR))
```

```r
p_CTR <- sdat %>%
  ggplot() +
  geom_histogram(aes(CTR), fill = "royalblue", bins = 100) +
  geom_vline(aes(xintercept = metrics_mean[,"mean_CTR"], color = "Cloverleaf"), size = 1) +
  annotate(geom = "label", x = metrics_mean[[1]], y = 75.0, label = round(metrics_mean[[1]],2),
colour = "#00BFC4") +

  geom_vline(aes(xintercept = 2.69, color = "Benchmark"), size = 1) +
  annotate(geom = "label", x = 2.69, y = 75.0, label = 2.69, colour = "#F8766D") +

  labs(x = "CTR [in percent]",
       y = "Count",
       colour = "Mean CTR",
       title = "Benchmark Cloverleaf CTR per Ad-Format vs. Google Search Average",
       subtitle = "Strong concentration of ads with CTR until 10%, current ads perform better
than the average",
       caption = "Source: Cloverleaf Search Data & WordStream Benchmark (from 05.10.2020)") +
  theme_bw()


p_CR <- sdat %>%
  ggplot() +
  geom_histogram(aes(CR), fill = "orange2", bins = 100) +
  geom_vline(aes(xintercept = 2.81, color = "Benchmark"),size = 1) +
  annotate(geom = "label", x = 2.81, y = 260, label = 2.81, colour = "#F8766D") +
  geom_vline(aes(xintercept = metrics_mean[,"mean_CR"], color = "Cloverleaf"), size = 1) +
  annotate(geom = "label", x = metrics_mean[[2]], y = 220.0, label = round(metrics_mean[[2]],2),
colour = "#00BFC4") +

  labs(x = "CR [in percent]",
       y = "Count",
       colour = "Mean CR",
       title = "Benchmark Cloverleaf CTR per Ad-Format vs. Google Search Average",
       subtitle = "Strong concentration of ads with CR around 0%, still better than benchmark",
       caption = "Source: Cloverleaf Search Data & WordStream Benchmark (from 01.05.2020)") +
  theme_bw()

p_CTR / p_CR

ggsave("04_hist.png", width = 9, height = 5)


# TASK 1.2: CORRELATION COEFFICIENTS
# check data by description and smoothing function
p_corr_CTR <-
  sdat %>%
  ggplot(aes(adrank, CTR)) +
  geom_point(alpha = .2) +
  geom_smooth(method = "loess", colour = "#00BFC4") +

  labs(title = "Correlation between CTR and Adrank",
       subtitle = "The better the adrank, the higher the CTR",
       x = "Adrank",
       y = "Click-Through Rate (CTR)",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  coord_cartesian(ylim = c(-20,100)) +
  theme_bw()

p_corr_CR <-
  sdat %>%
  ggplot(aes(adrank, CR)) +
  geom_point(alpha = .2) +
  geom_smooth(method = "loess", colour = "#00BFC4") +

  labs(title = "Correlation between CR and Adrank",
       subtitle = "No clear correlation indication for CR",
       x = "Adrank",
       y = "Conversion Rate (CR)",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  coord_cartesian(ylim = c(-20,100)) +
  theme_bw()
```

```
p_corr_CTR + p_corr_CR

ggsave("05_visualcorrelation.png", width = 8, height = 3)


# compute actual spearman correlation coefficients
sdat %>%
  select(adrank, CTR, CR) %>%
  correlate(method = "spearman") %>%
  rearrange() %>%
  shave() %>%
  focus(adrank) %>%
  mutate(term = reorder(term, adrank)) %>%
  mutate(adrank = adrank * (-1)) %>%
  ggplot(aes(adrank, term)) +
  geom_col() +
  geom_label(aes(label = round(adrank, 3))) +
  labs(title = "Ordinality-Adjusted Correlation Analysis for CTR, CR, and Adrank",
       subtitle = "Strong positive correlation between adrank and CTR (the higher/better the
rank, the higher the CTR), \nweaker positive correlation between adrank and CR  (the
higher/better the rank, the higher the CR)",
       x = "Correlation Coefficient",
       y = "Term") +
  coord_cartesian(xlim = c(-1,1)) +
  theme_bw()

ggsave("06_spearmancorrelation.png", width = 8, height = 3)


# TASK 1.3: COMPUTE ROI

# costs part of the profit margin
sdat_roi <- sdat %>%
  mutate(costs = bidprice * clicks,
         profit = (revenue * 0.035),
         roi = (profit/costs) * 100) %>%
  filter(clicks > 0 & bidprice > 0)

metrics_roi <- sdat_roi %>%
  summarise(mean(roi))

# how many observations are deleted for ROI calculation
nrow(sdat) - nrow(sdat_roi)

sdat_roi_summary <- sdat_roi %>%
  group_by(advertID) %>%
  summarise(mean_roi = mean(roi),
            sum_revenue = sum(revenue),
            sum_profit = sum(profit)) %>%
  mutate(scale = sum_revenue/sum(sum_revenue),
         wmean_roi = scale * mean_roi) %>%
  arrange(desc(mean_roi))

# mean roi accross all campaigns
sdat_roi_summary %>%
  summarise(mean_roi_all_campaigns = mean(mean_roi))

# relationship between roi and revenue
sdat_roi_summary %>%
  ggplot(aes(sum_revenue, mean_roi)) +
  geom_point(alpha = .3) +
  geom_smooth(method = "loess", colour = "#00BFC4") +
  labs(title = "Relationship between ROI and Revenue (averaged by campaign)",
       subtitle = "Highest campaign ROIs were achieved on the lower end of the Revenue scale",
       x = "Sum of Revenue per Campaign",
       y = "Average ROI per Campaign",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  theme_bw()

ggsave("07_ROI+RevenueRelationship.png", width = 8, height = 4)
```

```
# TASK 2.1: FACTORIAL PLOT FOR CR

# DPLYR BASED INTERACTION PLOTS (USED) ------

# observations per group (Appendix C)

sdat %>%
  select(CR, numberofwords, retailer, id) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer)) %>%
  group_by(retailer, numberofwords) %>%
  count()

# create summary statistics
sdat %>%
  select(CR, numberofwords, retailer) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer)) %>%
  group_by(retailer, numberofwords) %>%
  summarise(mean_CR = round(mean(CR),3)) %>%

  # bring to ggplot
  ggplot(aes(numberofwords, mean_CR, colour = retailer, label = mean_CR)) +
  geom_line(aes(group = retailer), size = 1) +

  geom_hline(aes(yintercept = metrics_mean[[2]]), linetype = "dashed", colour = "gray60") +
  annotate(geom = "label", x = 0.6, y = 5.8, label = "يCR", colour = "gray60") +

  geom_label(vjust = -0.5) +
  geom_point(pch = 19) +
  labs(title = "Factorial Plot of CR using 2 Factors (Retailer Name & Number of Keywords)",
       subtitle = "Number of Keywords have a considerable effect on CR, especially if combined 
with keywords \ncontaining the retailer name, leading to a change in the slope",
       x = "Number of Keywords",
       y = "Conversion Rate",
       colour = "Retailer Name",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  coord_cartesian(xlim = c(1,2),
                  ylim = c(0,14)) +
  theme_bw()

ggsave("08_FactorialPlotCR.png", width = 8, height = 4)


# QUESTION 2.2: CTR BASED ON CASES (RETAILER NAME; BRAND; KEYWORD LENGTH)

# observations per group (Appendix C)
sdat %>%
  select(numberofwords, retailer, brandname, CTR) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer),
         brandname = factor(brandname)) %>%
  group_by(retailer, brandname, numberofwords) %>%
  count()


# create summary statistics
sdat %>%
  select(numberofwords, retailer, brandname, CTR) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer),
         brandname = factor(brandname)) %>%
  group_by(retailer, brandname, numberofwords) %>%
  summarise(mean_CTR = round(mean(CTR),3)) %>%
```

```r
# to ggplot
  ggplot(aes(numberofwords, mean_CTR, colour = retailer, label = mean_CTR)) +
    geom_line(aes(group = retailer), size = 1) +

    geom_hline(aes(yintercept = metrics_mean[[1]]), linetype = "dashed", colour = "gray60") +
    annotate(geom = "label", x = 0.65, y = 13.0, label = "يCTR", colour = "gray60") +

    geom_label(vjust = -0.5, hjust = -0.3)+
    geom_point(pch = 19) +
    labs(title = "CTR Factorial Plot (split by Brand Name)",
         subtitle = "Effect of Retailer Name in general positive, but strongly negative if Brand
Name is included",
         x = "Number of Keywords",
         y = "Click-Through Rate",
         colour = "Retailer Name",
         caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  facet_wrap(~brandname) +
    coord_cartesian(xlim = c(1,2),
                    ylim = c(0,45)) +
    theme_bw()

ggsave("09_FactorialPlotCTR.png", width = 8, height = 4)


# QUESTION 2.3: ROI BASED ON CASES (RETAILER NAME; BRAND; KEYWORD LENGTH)

# observations per group (Appendix C)
sdat_roi %>%
  select(numberofwords, retailer, brandname, roi) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer),
         brandname = factor(brandname)) %>%
  group_by(retailer, brandname, numberofwords) %>%
  count()

# summary statistics
sdat_roi %>%
  select(numberofwords, retailer, brandname, roi) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer),
         brandname = factor(brandname)) %>%
  group_by(retailer, brandname, numberofwords) %>%
  summarise(mean_roi = round(mean(roi),1)) %>%

  # to ggplot
  ggplot(aes(numberofwords, mean_roi, colour = retailer, label = mean_roi)) +
  geom_line(aes(group = retailer), size = 1) +

  geom_hline(aes(yintercept = metrics_roi[[1]]), linetype = "dashed", colour = "gray60") +
  annotate(geom = "label", x = 0.65, y = 75.0, label = "يROI", colour = "gray60") +

  geom_hline(yintercept = 100, colour = "red4") +
  annotate(geom = "label", x = 0.75, y = 130, label = "Breakeven", colour = "red4") +

  geom_label(position = position_dodge(width = 0.8, preserve = c("total"))) +
  geom_point(pch = 19) +


  labs(title = "ROI Factorial Plot (split by Brand Name)",
       subtitle = "Effect of Retailer Name in general positive, but strongly negative if Brand
Name is included",
       x = "Number of Keywords",
       y = "Click-Through Rate",
       colour = "Retailer Name",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  facet_wrap(~brandname) +
  theme_bw()

ggsave("10_FactorialPlotROI.png", width = 8, height = 5)
```

```r
## APPENDIX ------

# CHECKING IF SAME STRATEGY WOULD BE PROPOSED FOR CR WITH SAME SET-UP

sdat %>%
  select(numberofwords, retailer, brandname, CR) %>%
  filter(numberofwords == 2 | numberofwords == 3) %>%
  mutate(numberofwords = factor(numberofwords),
         retailer = factor(retailer),
         brandname = factor(brandname)) %>%
  group_by(retailer, brandname, numberofwords) %>%
  summarise(mean_CR = round(mean(CR),1)) %>%

  # to ggplot
  ggplot(aes(numberofwords, mean_CR, colour = retailer, label = mean_CR)) +
  geom_line(aes(group = retailer), size = 1) +

  geom_hline(aes(yintercept = metrics_mean[[2]]), linetype = "dashed", colour = "gray60") +
  annotate(geom = "label", x = 0.65, y = 6.0, label = "يCR", colour = "gray60") +

  geom_label(vjust = -0.5, hjust = -0.3)+
  geom_point(pch = 19) +
  labs(title = "CTR Factorial Plot (split by Brand Name)",
       subtitle = "Effect of Retailer Name in general positive, but strongly negative if Brand
Name is included",
       x = "Number of Keywords",
       y = "Click-Through Rate",
       colour = "Retailer Name",
       caption = "Source: Cloverleaf Search Data (16.01.2012 to 03.12.2012)") +
  facet_wrap(~brandname) +
  coord_cartesian(xlim = c(1,2),
                  ylim = c(0,20)) +
  theme_bw()


ggsave("11_FactorialPlotCR2.png", width = 8, height = 4)
```