DETAILED ANALYSIS REPORT

Web Analytics at Quality Alloys, Inc.

The visit data indicates that a "reasonable number" of people are visiting the website. However, it's difficult to judge whether the numbers of visitors are large or small, without having the equivalent web data for comparable companies.

Looking more closely at the data analysis (See the Results-1) on the Weekly Visits worksheet, and financials draw the following conclusions:

Results-1

```
# IMPORTING DATA INTO R (Mainly to examine various aspects related to website visits)

visit <- read.csv("clipboard", sep = "\t", header = TRUE)

# CORRELATION EXAMINING THE RELATIONSHIP BETWEEN TOTAL VISITS AND UNIQUE VISITS
cor(visit$visits, visit$unique_visit)

## [1] 0.9994881

#                          DESCRIPTIVE                          STATISTICS
summary(visit)

##                  week         visits       unique_visit       page_views
##   Apr 12 - Apr 18: 1    Min.    : 383    Min.    : 366.0    Min.    : 793
##   Apr 19 - Apr 25: 1    1st Qu.: 596    1st Qu.: 540.0    1st Qu.:1602
##   Apr 26 - May 2 : 1    Median : 842    Median : 790.0    Median :1910
##   Apr 5 - Apr 11 : 1    Mean    :1052    Mean    : 989.2    Mean    :2173
##   Aug 10 - Aug 16: 1    3rd Qu.:1244    3rd Qu.:1175.0    3rd Qu.:2410
##   Aug 16 - Aug 22: 1    Max.    :3726    Max.    :3617.0    Max.    :5291
##            (Other)                                                :60
##    pages_visit          avg_time          bounce_rate          new_visit
##   Min.    :1.420    Min.    : 28.00    Min.    :0.5400    Min.    :0.7900
##   1st Qu.:2.025    1st Qu.: 59.75    1st Qu.:0.6100    1st Qu.:0.8400
##   Median :2.235    Median : 75.50    Median :0.6600    Median :0.8700
##   Mean    :2.258    Mean    : 74.94    Mean    :0.6718    Mean    :0.8688
##   3rd Qu.:2.575    3rd Qu.: 92.50    3rd Qu.:0.7300    3rd Qu.:0.8900
##   Max.    :3.180    Max.    :120.00    Max.    :0.8600    Max.    :0.9500
##
##             promotion      Revenue            Profit          lbs.sold
##  Initial         :14    Min.    :133967    Min.    : 32825    Min.    : 3826
##  Post Promotion :14    1st Qu.:372374    1st Qu.:111886    1st Qu.:12789
```

```
##  Pre-promotion   :21    Median :484857    Median :137928    Median :17216
##  Promotion       :17    Mean   :495440    Mean   :150898    Mean   :17342
##                         3rd Qu.:613587    3rd Qu.:187468    3rd Qu.:21128
##                         Max.   :951216    Max.   :275218    Max.   :31969
##
##                                                                 inquiries
##                      Min.                                   :    1.000
##                                1st                  Qu.:    5.000
##                                Median                  :    6.000
##                      Mean                                   :    6.394
##                                3rd                  Qu.:    8.000
##                           Max.                                  :16.000
##
```

Unique visits are highly correlated with visits (r = .99). About an average of 86% of visits are new visits—that is, they are generated by individuals who have not been to the QA website previously, during that week. (The monthly percentage of new visits is about the same—though this information is not presented in the case.) In other words, it appears that visitors check out the QA website and don't come back, at least not immediately.

Visitors don't spend a lot of time on the QA website. The average time spent in the website is one minute, 15 seconds. The average bounce rate is 67%. That is, on average 67% of the visitors leave the QA website from the same page they entered. The average pages viewed per visit is 2.26. It would seem therefore that visitors either find what they're looking for quickly (perhaps contact information) or don't find enough to interest them for any length of time. We don't know which it is (and neither does QA).

The promotion apparently had a large impact on the number of visitors to the site. However, the number of visitors is not a good predictor of the number of sales. As the visitors spike, seemingly as a result of the promotion, sales actually decrease. This is clear in the charts (Figures 1 and 2 below) generated in the solution.

```
# SUMMARIZE VISIT BASED ON PROMOTION
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'

##    The    following    objects    are    masked    from    'package:stats':
##
##     filter, lag

##    The    following    objects    are    masked    from    'package:base':
##
##     intersect, setdiff, setequal, union

group          <-          group_by          (visit,          promotion)
sum   <-   summarize(group,   visit   =   mean(visits,   na.rm   =   TRUE))

#                                USING                                GGPLOT2
library(ggplot2)

#      BAR      CHART      FOR      VISIT      BASED      ON      PROMOTIONS
ggplot(data=sum,        aes(x=promotion,        y=        visit))        +
  geom_bar(stat="identity",                              fill="steelblue")+
  geom_text(aes(label=     round(visit,     2)),     vjust=-0.3,     size=4)+
  theme_minimal()
```

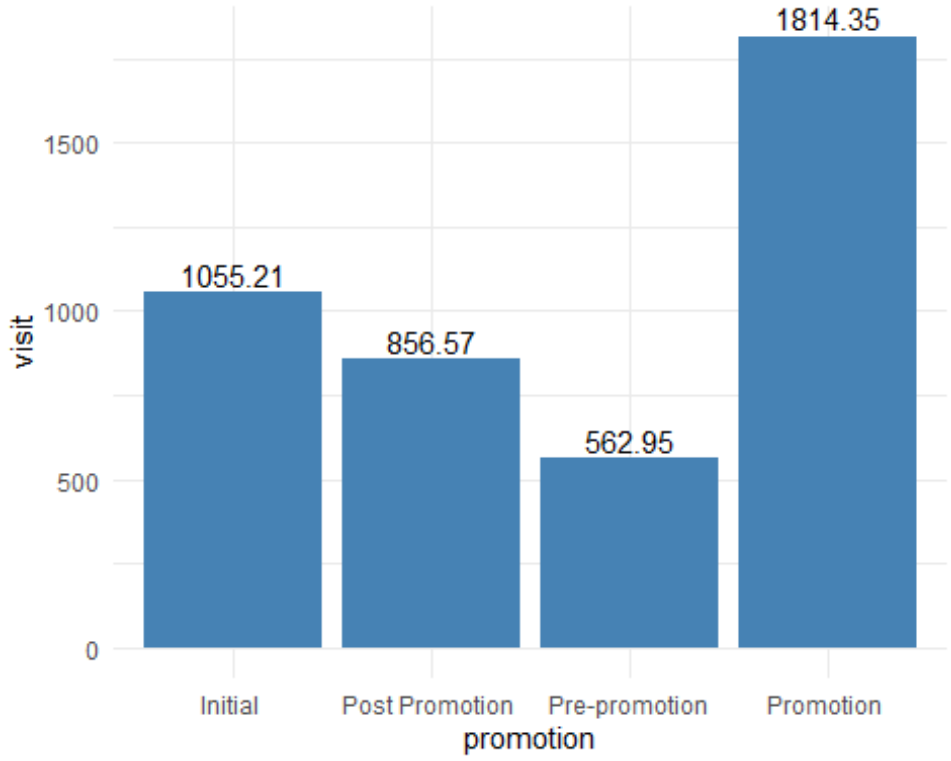

Figure 1: Bar chart (visit vs. promotion)

```
#        SUMMARIZE        REVENUE        BASED        ON        PROMOTION
group          <-          group_by          (visit,          promotion)
sum_1<-   summarize(group,   revenue   =   mean(Revenue,   na.rm   =   TRUE))
```

```
#      BAR      CHART      FOR      REVENUE      BASED      ON      PROMOTIONS
ggplot(data=sum_1,       aes(x=promotion,       y=       revenue))       +
  geom_bar(stat="identity",                              fill="steelblue")+
  geom_text(aes(label=    round(revenue,    2)),    vjust=-0.3,    size=4)+
  theme_minimal()
```

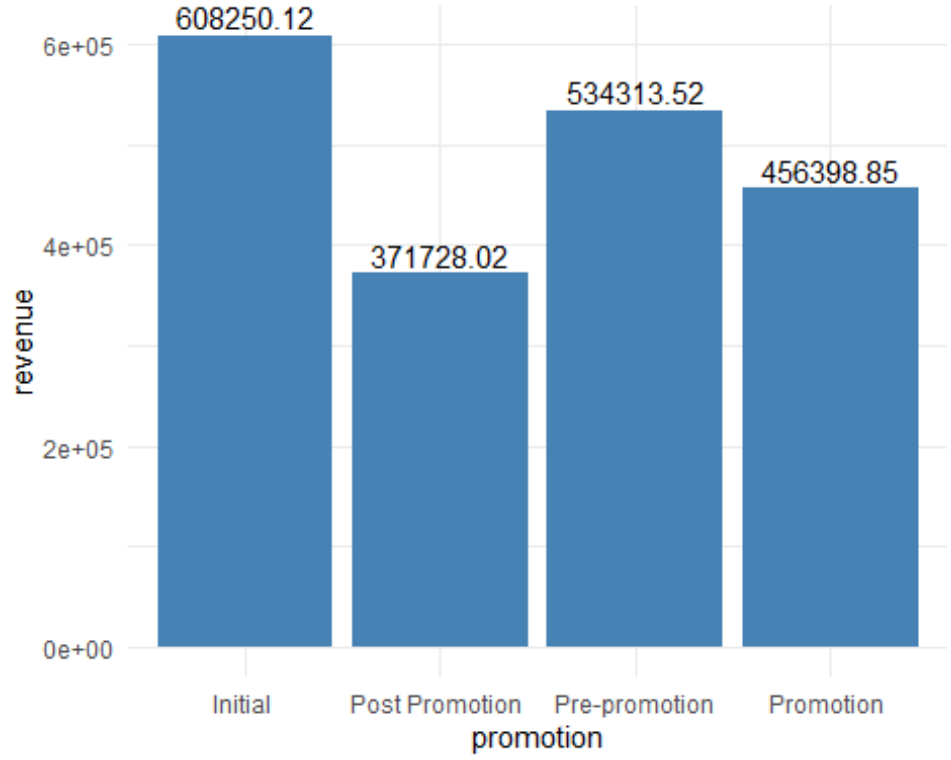

Figure 2: Bar chart (revenue vs. promotion)

Further, when we look at these two variables together, creating a scatter plot and calculating the correlation coefficient. The scatter plot, as is implied by these charts (**See Figure 3**), indicates no pattern, and the correlation coefficient is close to zero (-.059). This indicated that there exists no pattern or relationship between number of visits and revenue.

```
#EXAMINATION OF THE RELATIONSHIP (visit vs. Revenue) USING SCATTER PLOT AN
D                                                                CORRELATION
ggplot(visit, aes(x= visits, y= Revenue)) + geom_point()
```

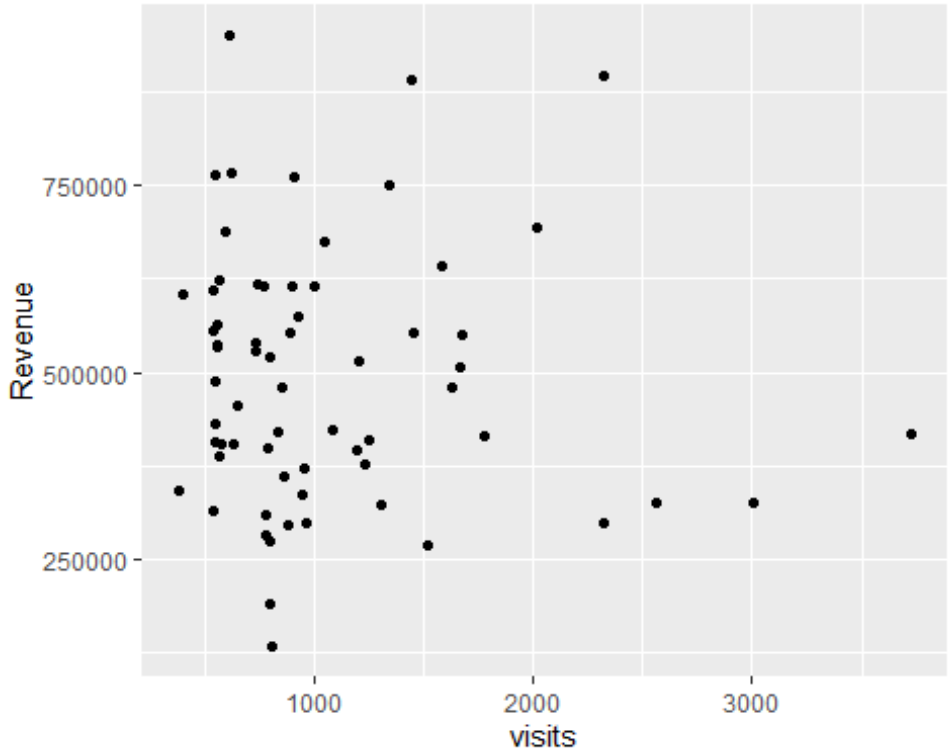

Figure 3: Scatter plot (visit vs. Revenue)

```
cor(visit$visits, visit$Revenue)

## [1] -0.05939183
```

However, when we create a scatter plot of revenue versus pounds and calculating this correlation coefficient serves to set a baseline, and it reported a strong, positive linear relationship between two variables. In addition, the scatter plot showing the relationship between revenue, and pounds of material sold also carry a high correlation/association.

```
#EXAMINATION OF THE RELATIONSHIP (Revenue, lbs.sold, Profit)
par(mfrow=c(1,2))
plot(visit$Revenue,                                    visit$lbs.sold)
plot(visit$Revenue,visit$Profit)
```

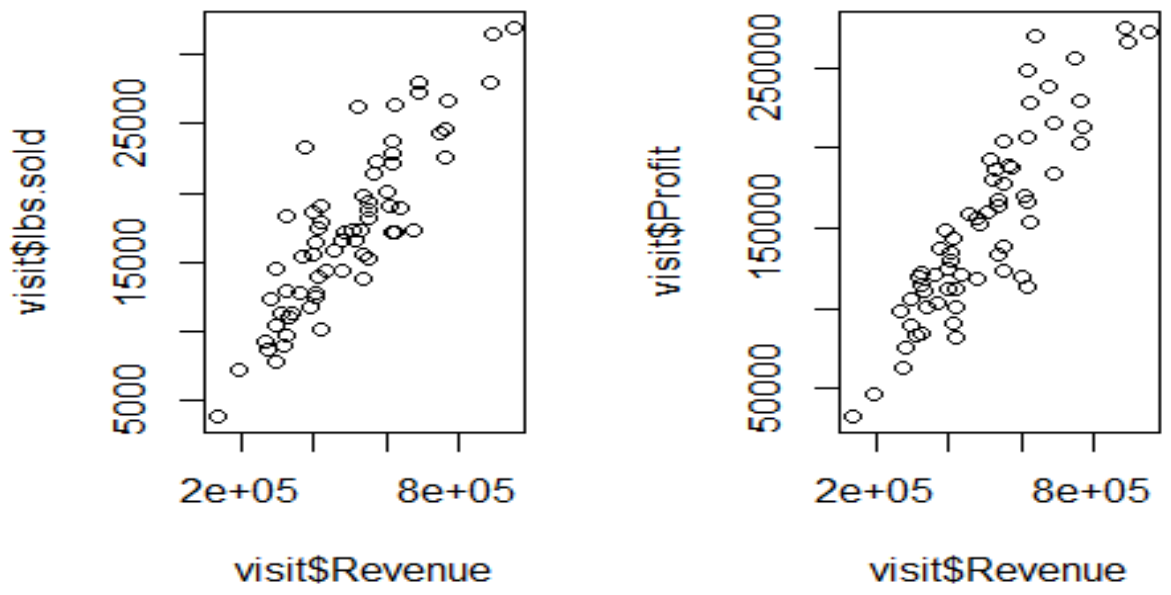

Figure 4: Scatter plot (visit, revenue, profit)

```
#                                                                    CORRELATION
cor(visit$Revenue, visit$lbs.sold)

## [1] 0.8689297

cor(visit$Revenue, visit$Profit)

## [1] 0.8872251
```

Management at QA was not particularly cognizant of the level of website traffic, and was not explicitly aware that visits were unrelated to sales. This being said, they were not particularly surprised at the lack of relationship between visits and profit/pounds of material sold.

One likely explanation for these results is that QA customers typically make a purchase when they have a contract in hand. Therefore, they won't respond immediately to promotional material. It can reasonably be argued though, that the value of the website is that it helps provide name recognition for the company, so that when they are ready to make a purchase they will think of QA. Beyond this, it provides a sense of legitimacy (what reputable

organization doesn't have a website?), as well as handy access to contact information. There are of course other possible explanations—for example, poor website design and functionality or that a website in this line of business is not terribly relevant.

If we need further insights, we can also analyze the data by looking at other variables by period, though the exercises don't explicitly ask they do. Figure 5 gives the number of inquiries received by QA by period. The numbers generally are small; in any case, there's no jump in inquiries related to the spike in visits.

```
# SUMMARIZE INQUIRIES BASED ON PROMOTION
group <- group_by (visit, promotion)
sum_2<- summarize(group, inquiries = mean(inquiries, na.rm = TRUE))

# BAR CHART FOR INQUIRIES BASED ON PROMOTIONS
ggplot(data=sum_2, aes(x=promotion, y= inquiries)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label= round(inquiries, 2)), vjust=-0.3, size=4)+
  theme_minimal()
```

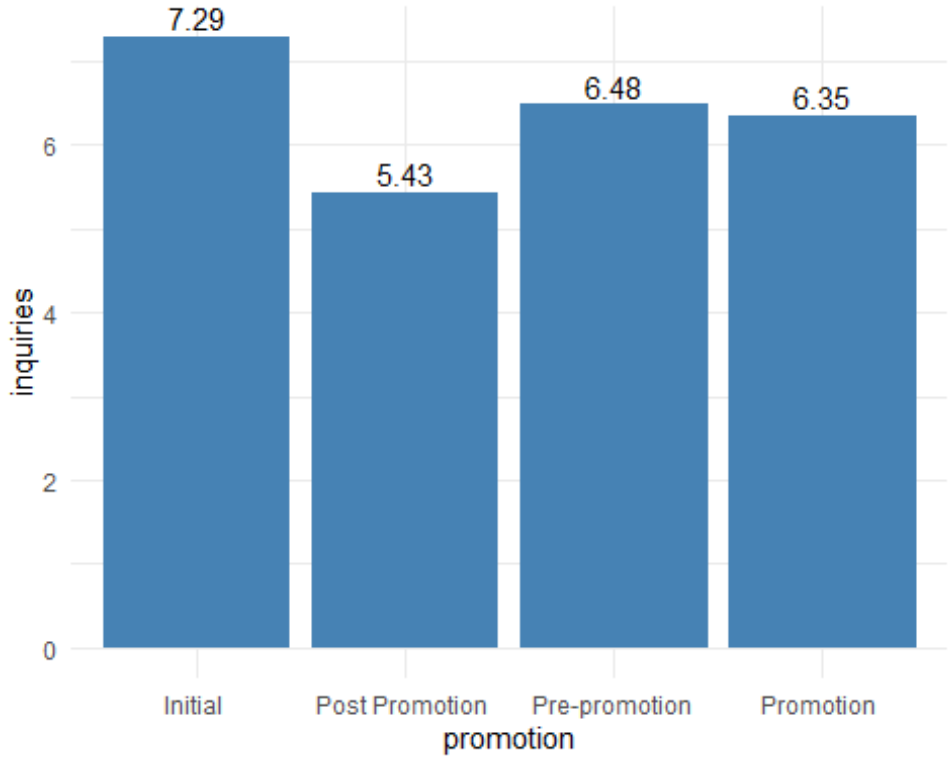

Figure 5: Bar chart (inquires vs. promotion)

```r
#          SUMMARIZE          OTHER          IMPORTANT          VARIABLES

sum_3<-  summarize(group,  page_views  =  mean(page_views,  na.rm  =  TRUE))
sum_4<-  summarize(group,  pages_visit  =  mean(pages_visit,  na.rm  =  TRUE))
sum_5<-   summarize(group,   avg_time   =   mean(avg_time,   na.rm   =   TRUE))
sum_6<-  summarize(group,  bounce_rate  =  mean(bounce_rate,  na.rm  =  TRUE))
sum_7<-   summarize(group,   new_visit   =   mean(new_visit,   na.rm   =   TRUE))


#              BAR              CHART              FOR              VARIABLES

ggplot(data=sum_3,       aes(x=promotion,       y=       page_views))       +
  geom_bar(stat="identity",                                fill="steelblue")+
  geom_text(aes(label=    round(page_views,    2)),    vjust=-0.3,    size=4)+
  theme_minimal()
```

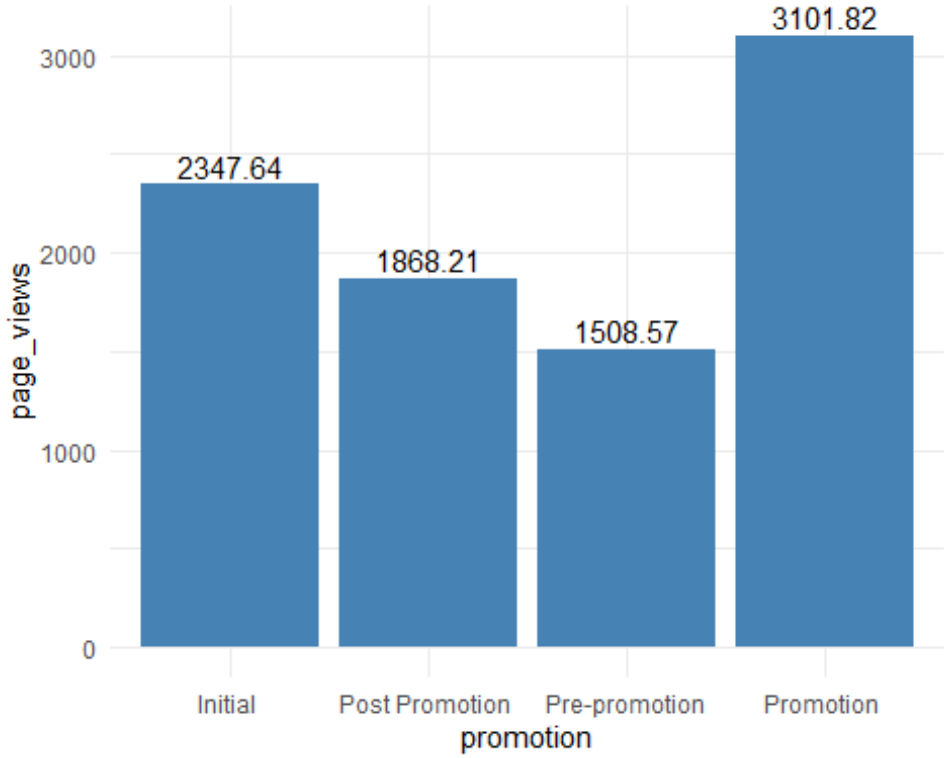

Figure 6: Bar chart (page views vs. promotion)

```r
ggplot(data=sum_4,       aes(x=promotion,       y=       pages_visit))       +
  geom_bar(stat="identity",                                  fill="steelblue")+
  geom_text(aes(label=    round(pages_visit,    2)),    vjust=-0.3,    size=4)+
  theme_minimal()
```

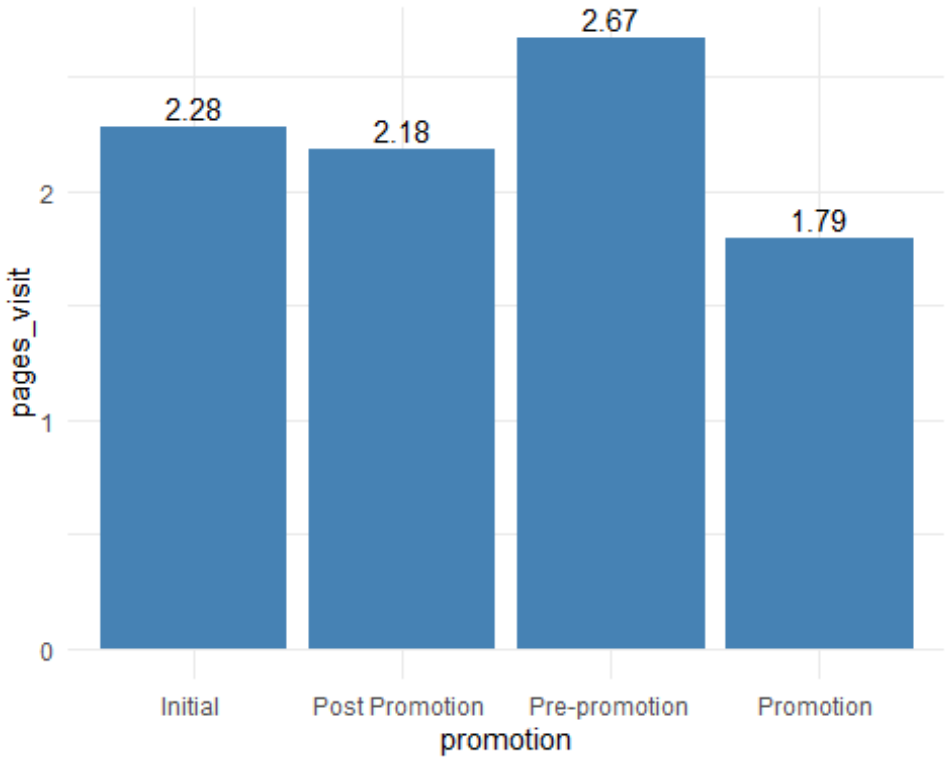

Figure 7: Bar chart (page visit vs. promotion)

```
ggplot(data=sum_5,       aes(x=promotion,       y=       avg_time))       +
  geom_bar(stat="identity",                                fill="steelblue")+
  geom_text(aes(label=    round(avg_time,    2)),    vjust=-0.3,    size=4)+
  theme_minimal()
```

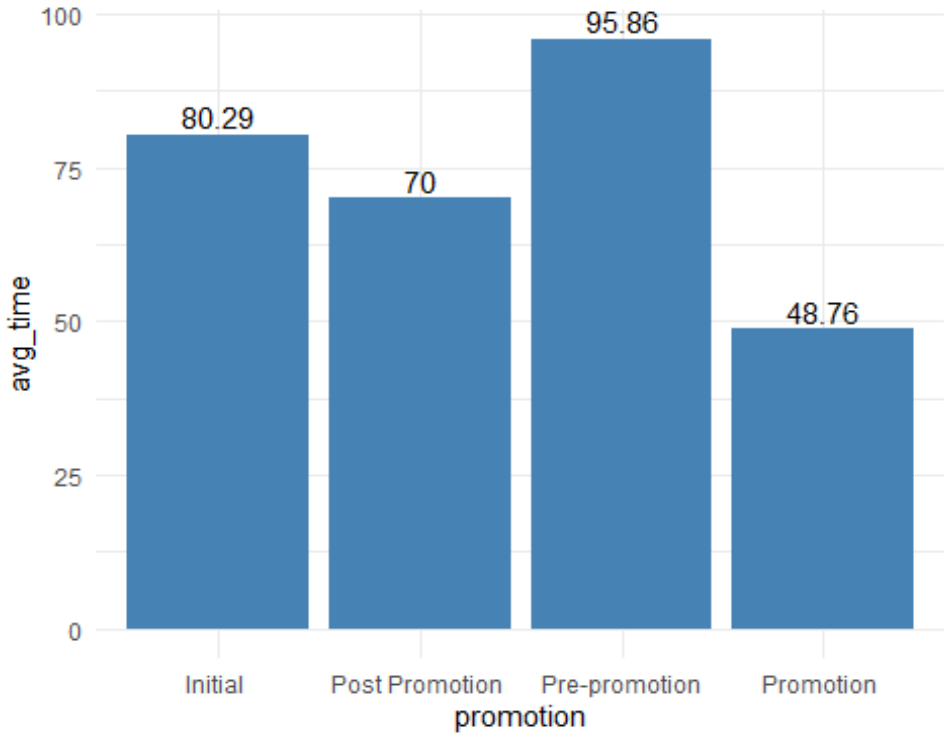

Figure 8: Bar chart (Average time (in seconds), vs. promotion)

```
ggplot(data=sum_6,      aes(x=promotion,      y=      bounce_rate))      +
  geom_bar(stat="identity",                              fill="steelblue")+
  geom_text(aes(label=   round(bounce_rate,   2)),   vjust=-0.3,   size=4)+
  theme_minimal()
```

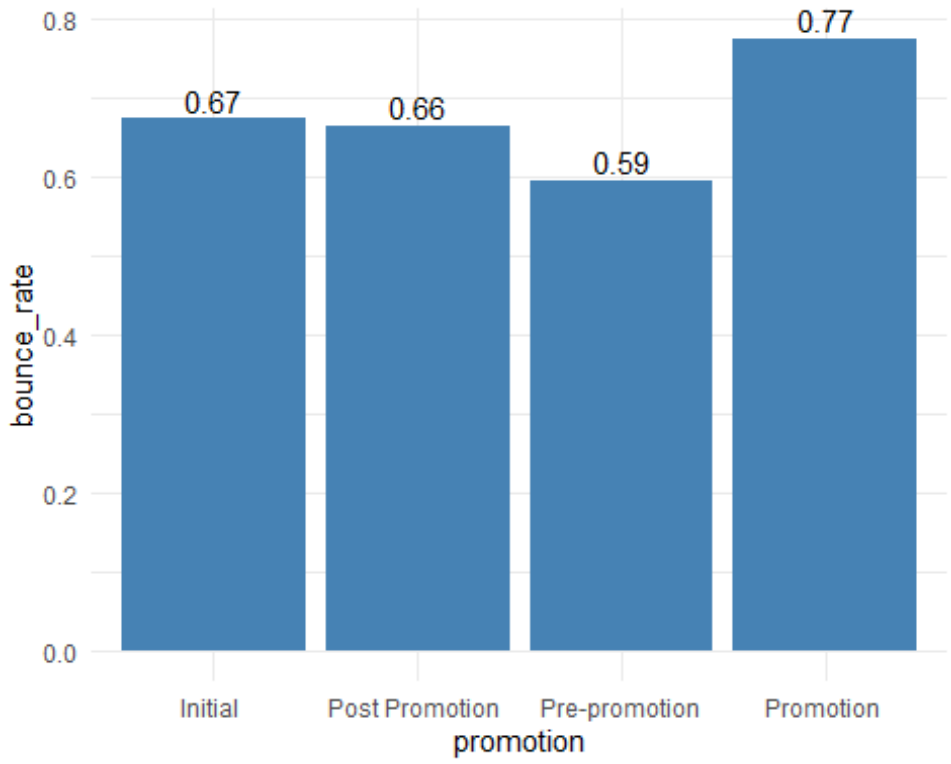

Figure 9: Bar chart (Bounce rate (in percentage), vs. promotion)

```
ggplot(data=sum_7,      aes(x=promotion,      y=      new_visit))      +
  geom_bar(stat="identity",                              fill="steelblue")+
  geom_text(aes(label=   round(new_visit,   2)),   vjust=-0.3,   size=4)+
  theme_minimal()
```

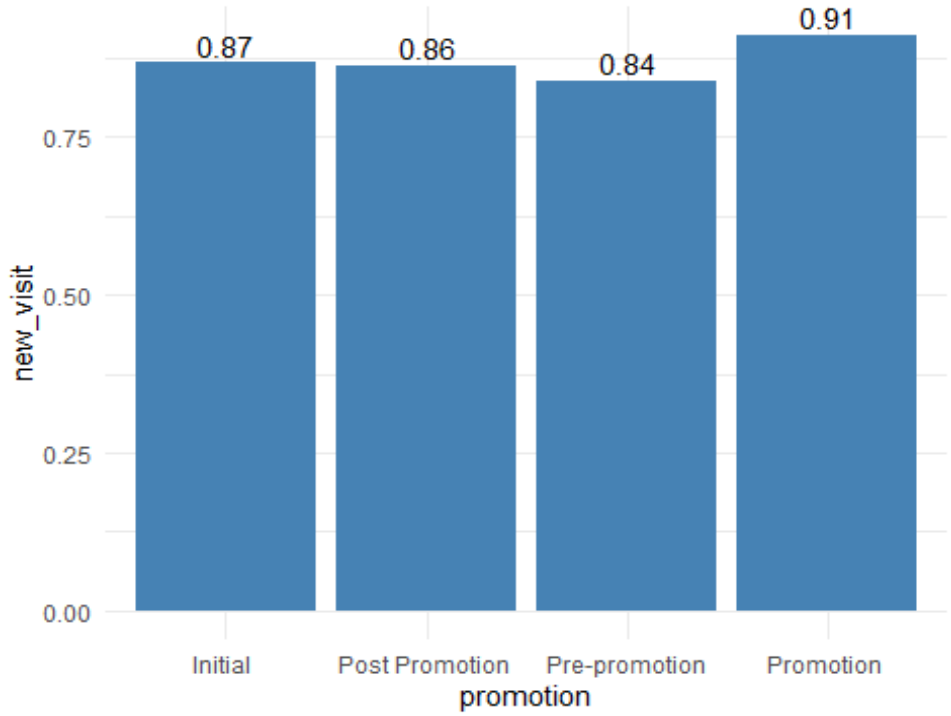

Figure 10: Bar chart (new visit (in percentages) vs. promotion)

Figures 7–9 uniformly indicate that the promotion visitors were *less* engaged with the website than visitors during other periods. That is, visitors during the promotion period viewed fewer pages, had a higher bounce rate, and spent less time on the website. **Figure 10** indicates a slight rise in the percentage of new visitors during this period.

So, it seems likely that individuals who received the promotion simply checked the website briefly.

OTHER IMPORTANT ANALYSIS:

In addition to the above-stated analysis, one can use the demographic sheet to get detailed insights about the customers. From the analysis, we see that over half of the QA website visitors over this period come from referring sites; of these, about three-quarters come via Google AdWords (the first two in the list). Some 30% of the visitors come from search engine traffic, with over three-quarters of these coming from Google searches.

Likewise, QA doesn't know if the number of visitors to its website and their visit characteristics are typical for other companies in their space. It would seem reasonable to get some benchmark data from industry organizations, consultants, or some other sources. These places may be good resources as well for ideas as to how better monetize their web presence.

It would seem to make sense for QA to periodically verify that the company appears towards the top of the listings of companies that appear on a Google search of relevant terms.

Overall, from this data analysis one can walk away with an appreciation for the immense range of data resulting from web-based interactions that are available for free and with little effort (in this case via Google Analytics). The data presented here is just a subset of all that's collected—for example, the geographic data can be broken down further (e.g., by country). Ideally, the case should serve as a motivator to understand the importance of quantitative analysis (basic descriptive statistics, data grouping, summary, and visualizations), while at the same time making it clear that available data may not provide the answers all questions, but rather point at the need for further data gathering and analysis.

In addition to the Google Analytics data, the AdWords program provides a wealth of data too—not included here. Further, both programs offer well-designed dashboards and other analytics tools. QA hasn't looked at the AdWords data at all. QA is no doubt not unusual in "having more data than it knows what to do with."

Finally, we can also apply the basic assumptions using the current data, specifically the assumption or normality which we discussed in some of the earlier classes. One data set (pounds of material sold) is pretty normal while the other (daily visits) is not, so you all get a basis for comparison. This sets the stage for you all to better understand distributions in general, as well as really understand how we can assess normal probabilities via z-scores or using skewness (please refer some of the earlier session notes).

OVERVIEW OF THE DATA ANALYSIS

FINDINGS-1

Visits per week indicate a big bump due to the promotion.

Unique visits track visits. Referring back to the Weekly Visits worksheet—we found that the correlation between these two variables are quite high. Also, we calculated the average percentage of new visits as 86.88%. We can conclude that, for better or worse, most visitors are new visitors (at least over the week). The related data (page views, time on site, etc.) indicate that visitors come to the site, look around briefly, leave, and don't come back (at least not during the same week).

Revenue, profit, and pounds sold seem to move together—this makes sense. These, however, do not seem to correspond at all to the number of website visits. In particular, there's no observable bump during the promotion period. Beyond this, there doesn't seem to be any dramatic pattern in these figures (aside from a regular decrease every year around July 4, when QA closes for a week).

FINDINGS-2

There is a huge jump in visits associated with the promotion and a corresponding increase in the standard deviation of visits during this period. This is very clear from the plot of the data. During post-promotion, the mean settles down to a level somewhat higher than that of pre-promotion, and the spread returns to roughly that of the pre-promotion data. The mean in each period is greater than the median, which would seem to indicate that the data are skewed right.

The pattern in the financial data is entirely different. Mean and median values decrease over the four periods. The standard deviations remain roughly constant. In each period the mean is again greater than the median, but the effect seems less pronounced than for the visit data.

FINDINGS-3

The promotion effect on visits/unique visits is quite clear in these plots. The steadily decreasing pattern in the financial data is clear as well. You can also observe the lack of relationship between these data sets.

FINDINGS-4

From the analysis we can also observe a linear relationship between revenue and pounds sold. The scatter plot and correlation coefficient confirm this.

FINDINGS-5

There is no clear linear relationship between the revenue and visits. The scatter plot and correlation coefficient confirm this.

An additional point worth noting—the other "visit" data (page views, bounce rate, etc.) indicate that visitors are spending precious little time on the website, and visitors during the promotion period (as noted above) seem the least engaged.

FINDINGS-6 (please do the necessary analysis/calculations using R)

Histogram and Empirical Rule calculations indicate that the pounds of material sold per week data can be reasonably approximated by a normal distribution.

The histogram indicates that daily visits are not normal. The z-scores confirm (a) more data clustered around center than normal and (b) more data at far right than at far left. The positive skewness further confirms these observations.

FINDINGS-7 (please do the necessary analysis/calculations using R)

It is advised use pie charts here (please see demographic sheet). Some observations are:

Over half (56%) of the visitors come from referring sites. Of the top ten referring sites, some 80% of these come from two Google AdWords related sites.