Take Home Test Analysis

Kalon Goonetilleke

Setup

I called the following libraries

```
library("gridExtra")
library("readxl")
library("tidyverse")
library("dplyr")
library("ggplot2")
library("lubridate")
library("rtsplot")
library("xts")
```

I read in the Excel File and assigned it to df. We have 174,226 rows.

```
df <- read_excel("/Users/kalong/Desktop/CODE3/dataset.xlsx")
df
tidy_names(df)
```

I converted creative_name to creative_name_low in Excel.
I proceeded to select creative name lowercase and assign it to creativedf

```
creativedf <- select(df, creative_name_low)
#verifying my method
creativedf
```

Question 1

How many versions of creative run?
There were 10 versions of creative run. They are:

1. *Black*
2. *Black**new*
3. *Grey*
4. *Grey**new*
5. *Orange*
6. *Orange**new*
7. *Teal*
8. *Teal**new*
9. *White*
10. *White**new*

```
> creativedf %>%
+   group_by(creative_name_low) %>%
+   summarise( n = n())
# A tibble: 10 x 2
   creative_name_low     n
   <chr>             <int>
 1 black              3594
 2 black**new          407
 3 grey                  6
 4 grey**new             2
 5 orange            74252
 6 orange**new        8812
 7 teal               3464
 8 teal**new           441
 9 white             74454
10 white**new         8794
```

```
count(unique(creativedf))
A tibble: 1 x 1
      n
  <int>
     10
```

Question 2

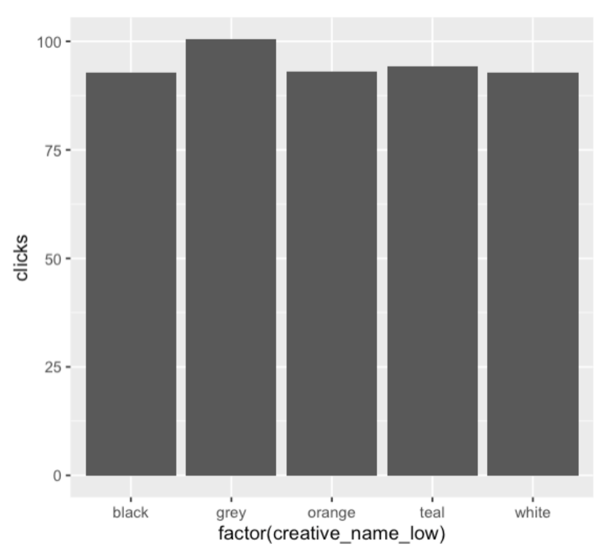
Did the newer version of creative have any effect? How much?

I separated it into two dataframes that only have new creatives and another dataframe that has standard creatives (excludes new creative string).

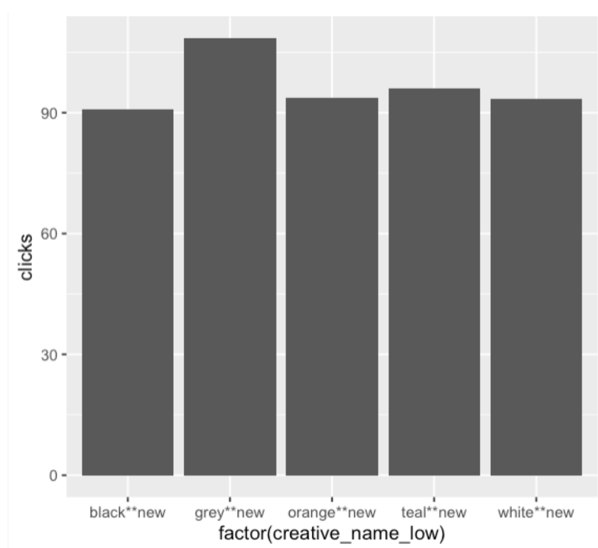
```
#Create a dataframe that only has new creatives
newcre <- ndf %>% filter(grepl('new', creative_name_low))
newcre
#Extract data I want
newcreatives <- select(newcre, creative_name_low, impressions, clicks, conversions)

#Create a dataframe that has 'standard' creatives
stdcre <- ndf %>% filter(!grepl('new', creative_name_low))
#Extract Data I want
stdcreatives <- select(stdcre, creative_name_low, impressions, clicks, conversions)
```

Comparing Clicks

I assigned ggp1 to standard creatives and then ggp2 to new creatives. I plotted the averages of the clicks per each creative on a bar graph as shown below.



Ggp1 (standard creatives)



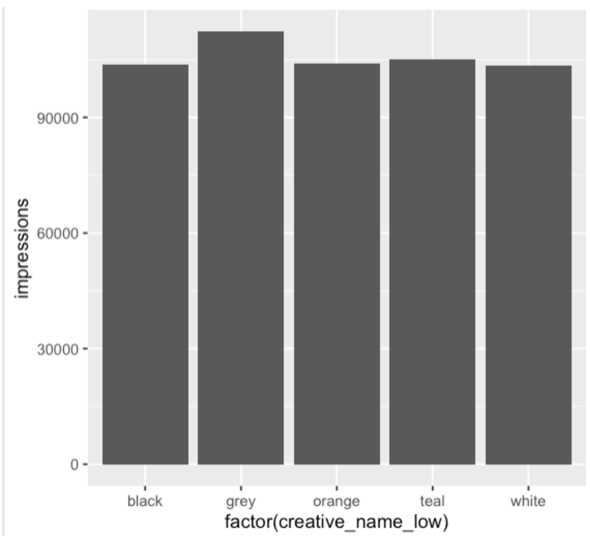
Ggp2 (new creatives)

```
> summary(newcreatives$clicks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.00   50.00   94.00   93.63  138.00  184.00
> summary(stdcreatives$clicks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.00   49.00   93.00   92.99  137.00  184.00
```

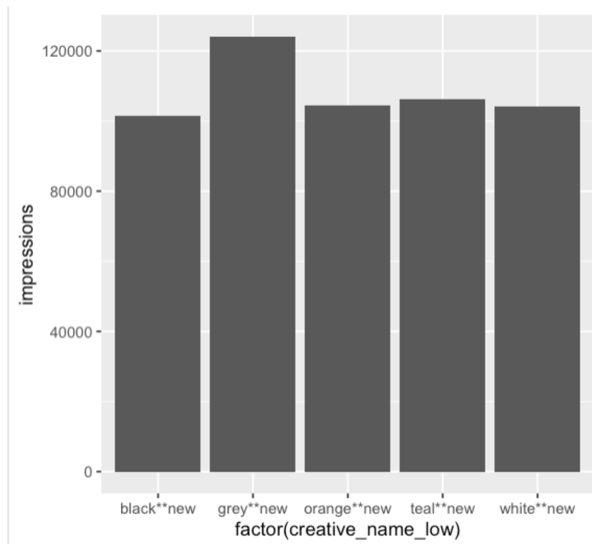
From the bar chart and summary of both creatives, we can tell that new creatives generated approximately similar numbers of clicks. New creatives have a marginally higher mean. *I would state that new creatives generate marginally more clicks than the standard creative. This has a low level of significance.* ***We can see that grey new had a larger significant increase in generating clicks relative to grey standard.***

Comparing Impressions

I assigned ggp3 to standard creatives and then ggp4 to new creatives. I plotted the average number of impressions on a bar graph as shown below.



Ggp3 (standard creatives)

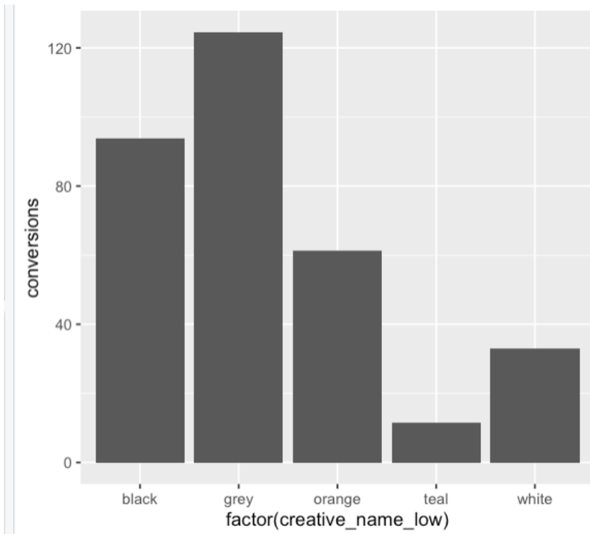


Ggp4 (new creatives)

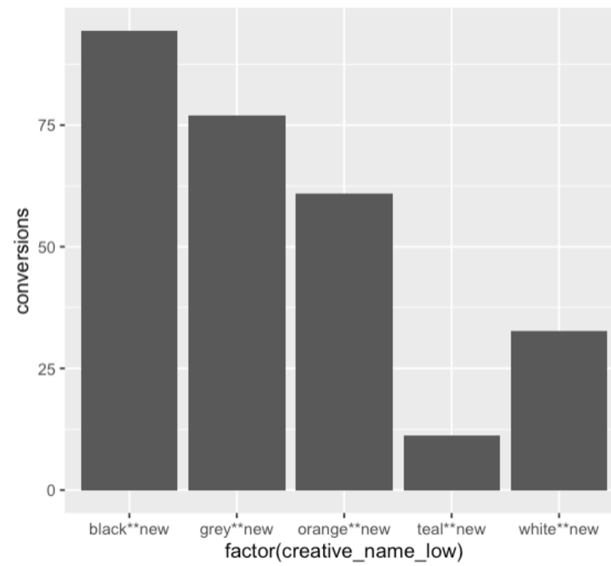
```
> summary(newcreatives$impressions)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    782   54663  103142  104364  151820  296951
> summary(stdcreatives$impressions)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
    789.6   53949.5  102128.8  103694.0  150871.4  329474.0
>
```

I can see here that the mean of impressions of **new creatives** is slightly higher than that of mean impressions of standard creatives. However, we can see that the maximum value of standard impressions is higher than that of max value of new creatives. There is an outlier in standard creatives that skews the data a bit. *I would not confirm that new creatives is better than standard creatives or vice versa. However, we can confirm that grey new tends to be exceptional at generating impressions. There is a more significant increase in impressions relative to what standard grey was generating prior.*

Comparing Conversions
I assigned ggp5 to standard creatives and then ggp6 to new creatives. I plotted these on a bar graph as shown below.



Ggp5 (standard creatives)



Ggp6 (new creatives)

```
> summary(newcreatives$conversions)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   17.00   30.00   47.07   56.00  313.00
> summary(stdcreatives$conversions)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   17.00   29.00   47.36   57.00  314.00
```

I can verify that the mean conversions of standard creatives is marginally higher than the mean of conversions of new creatives. We can see that black new is performing better than black standard, relative to other creatives overall. Further, we see that grey new is performing worse than grey standard and generating fewer clicks than before. Overall, we can see the overall version of creative didn't have a significant effect on conversions. We also see that the max of conversions in standard creatives is 1 conversion lower than that of new creative conversions. There is not a significant effect in the overall change of conversions from standard creative to new creatives, but within the creatives themselves, there is a much larger effect.

Question 3

Show the **CPA** and **CVR** for each campaign type that was run.

I defined **CPA** to be cost per acquisition.
Where **CPA** is total amount spent / total attributed conversions. The corresponding **CPA** values can be found in the ‘total’ dataframe.

```
q3df <- select(df, campaign_name, creative_name_low, month,
               spend, impressions, clicks, conversions)
head(q3df)

spend <- q3df %>%
  group_by(campaign_name) %>%
  summarise( sumspend = sum(spend))

sumofconv<- q3df %>%
  group_by(campaign_name) %>%
  summarise(sumconversion = sum(conversions))

total <- merge(sumofconv, spend)
head(total)

total$CPA <- (total$sumspend / total$sumconversion)
```

```
> tail(total)
                  campaign_name sumconversion  sumspend      CPA
103 sc-VISA-sleep--conversions         47313  233873.3 4.943109
104   sc-VISA-sleep-conversions       354787 1750559.6 4.934114
105     sc-VISA-You--conversions       47826  238506.9 4.986971
```

Exported to spread sheet.

campaign_name	sumconversion	sumspend	CPA
sc_BBY_._branding	12131	236462.5654	19.49242
sc_BBY_Austin_branding	13133	255321.4171	19.44121
sc_BBY_in_branding	11133	220965.7777	19.84782
sc_BBY_Nashville_branding	12022	234568.3044	19.51159
sc_BBY_NOLA_branding	12659	236683.4575	18.69685
sc_BBY_Office_branding	13686	269291.6248	19.67643
sc_BBY_pivot_branding	12428	248095.0321	19.96259

I defined **CVR** to be the conversion rate where the users saw ad and took action as a result
CVR is (number of conversions/number of impressions) * 100

```
q3df
q3df$CVR <- ((q3df$conversions / q3df$impressions) * 100)
head(q3df)
```

Output/result:

```
  campaign_name                     creative_name_low month spend impressions clicks conversions     CVR
  <chr>                             <chr>             <dbl> <dbl>       <dbl>  <dbl>       <dbl>   <dbl>
1 sc-BBY-Nashville-branding         white                11 660.       178191    132          13 0.00730
2 sc-VISA-Portland-conversions      orange                5 890.       210176    176         197 0.0937
3 sc-GAP-Portland-fanacquisition    white                 8  40.9        8039      9           1 0.0124
```

Exported to Excel Sheet.

Question 4

Which campaign type was more efficient at driving conversions? By how much?
To find this, I plotted a graph of conversions relative to campaign type. In order to sort by campaign type, I had to filter by branding, conversions, fanacquistions.

```
q4df<- select(df, campaign_name, conversions)
head(q4df)
tail(q4df)

campbrand <- q4df %>% filter(grepl('branding', campaign_name))
campconver <- q4df %>% filter(grepl('conversions', campaign_name))
campfana <- q4df %>% filter(grepl('fanacquisition', campaign_name))

campbrandsum<- campbrand %>%
  summarise(sumbrandconversion = sum(conversions))
campbrandsum

campconversum<- campconver %>%
  summarise(sumconverconver = sum(conversions))
campconversum

campfanasum<- campfana %>%
  summarise(sumcampfanasum = sum(conversions))
campfanasum
```

I obtained the sum of **conversions** for *brand*, *conversion*, and *fan acquisitions.*

Type of Campaign	Sum of Conversions
Conversion	5,483,600
Brand	1,381,526
FanAcquisitions	1,381,300

From this data, we can see that the campaign ‘conversion’ generated the most amount of conversions, which was 5,483,600. Next up was brand, and then fan acquisitions which had similar amounts of conversions each. The campaign type ‘CONVERSION’ was the most efficient at driving conversions.
When looking at summary statistics:
The mean was 94.37 for conversions
The mean was 23.78 for fan acquisitions
The mean was 23.8 for branding.

On average, there were 70 more conversions carried out for the conversion type campaign relative to brand and fan acquisitions. We can safely state that the conversion campaign looks successful via data analysis.

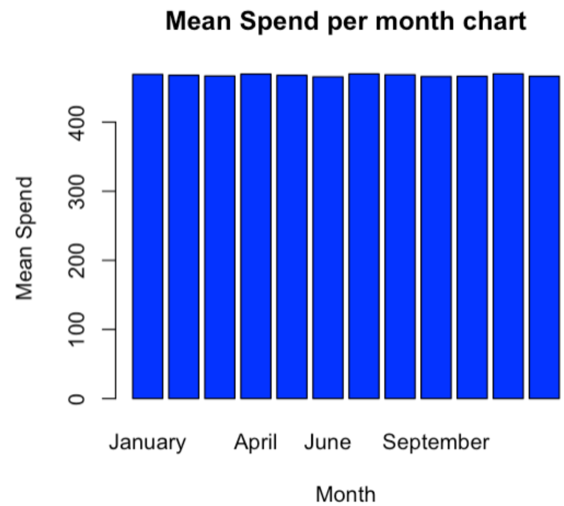
Question 5

Are there any time period trends? If so, within which metrics and what dimension are they sensitive to?
Assign months to different time periods
What metrics/dimensions are time periods sensitive to?

I created a dataframe for Question 5, also known as q5df. I included all the necessary metrics.

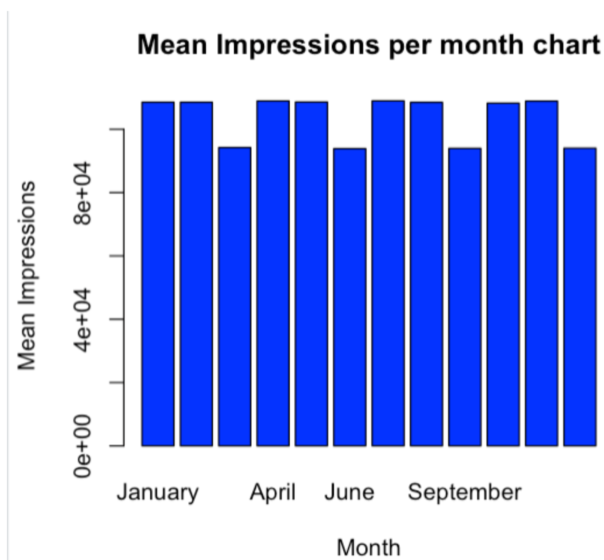
```
q5df <- select(q3df, month, creative_name_low, spend, impressions, clicks, conversions, CVR)
```

Spend



Upon analyzing mean spend, we see that month to month spend is relatively the same with a mean spend of about $467. We do see some months will marginally (and relatively) lower spend, like March ($467) June ($466), September, October, and December (also $466 respectively).

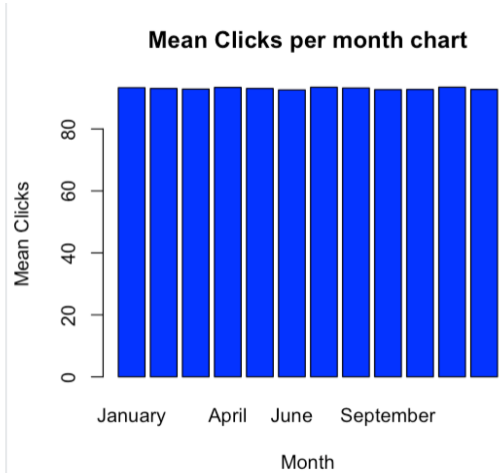
Impressions



Upon analyzing average impressions, we can see a trend of 2 months with high impressions followed by a month of low impressions. For example January and February will have exceptional average impressions in lockstep. However, March will result in a lower number of impressions. This continues until December.

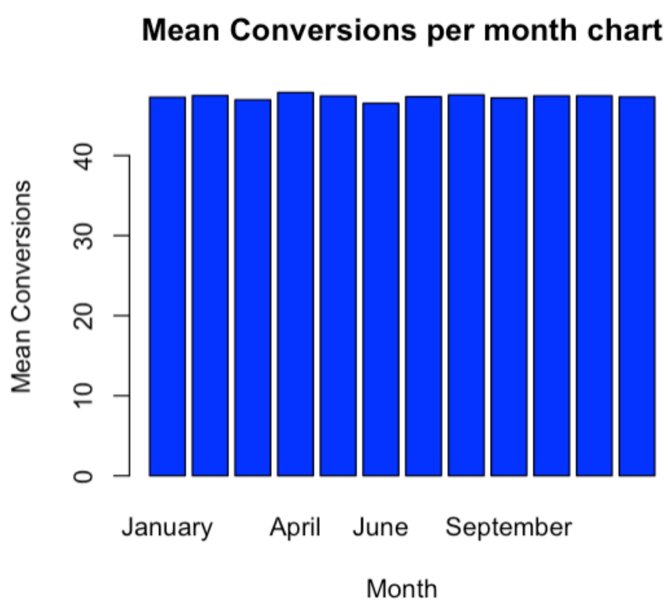
Further, I can see the average number of impressions is 103,782.7. Every 3rd month however, we can see that the average impressions dips to under 95,000 impressions which is a significant drop. Overall these are good impressions and refers to a lot of potential for clicks and conversions.

Clicks



When analyzing mean clicks per month, I was able to find that the mean clicks stay relatively the same hovering at about 92 or 93. The average was 93 clicks per month for this analysis. I would say that clicks remain average and don't change month-to-month.

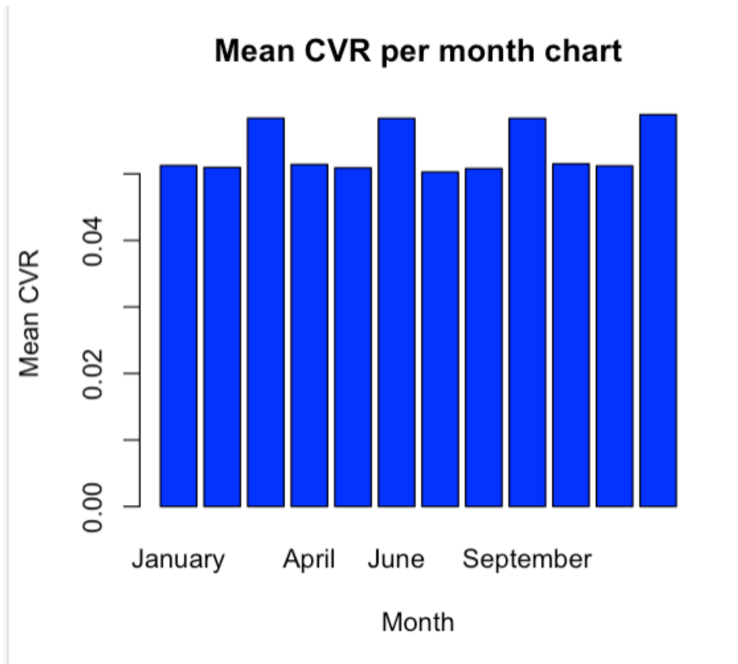
Conversions



We can see here that mean conversions stay at about 47 across all months. We tend to see some months that have marginally lower conversions, like March and June. This could be attributed to the correlation that was seen earlier when analyzing impressions. I noticed that the average conversions dropped to 46.5 in June and 47 in March along with September and December (relatively lower than other average conversions).

The reduced impression rate in March, June, September, and December translates to lower conversions that we can see in March and June in the conversion chart above. We can't really attribute this to spend because it is a marginal change of about $1 and has little significance. However, this is still something to keep in mind. We are spending similar amounts yet the impressions are dropping significantly and impacting the conversion rate. There is some sort of interference here.

CVR (conversion rate)



We can see that in March, June, September, and December, the average conversion rate spikes up to above 0.0510%. We can see here that some seasonality exists every 2 months. I would use ARIMA time series modelling to get more information on this but I didn't have enough time. We can deduce that conversion rates increase to a consistent and significant level, every 3 months.

Question 6

How many creative colors were run in February?

```
q6df <- select(q5df, month_name, creative_name_low)


febcreative <- q6df %>%
  filter(month_name == "February")

febcreative %>%
  group_by(creative_name_low) %>%
  summarise()
```

I actually did this an easier way with filer(month_name = “february” = TRUE) and it worked but I had an issue with my compiler so I just did it like this.

We obtained the creative colors run in February as:

1. Black
2. Black**new
3. Orange
4. Orange**new
5. Teal
6. Teal**new
7. White
8. White**new

```
1 black
2 black**new
3 orange
4 orange**new
5 teal
6 teal**new
7 white
8 white**new
```

Question 7

If this client asked you what CPA you'd project for next month what would be the most important questions you want to ask prior to providing a response?
Cost per acquisition is the average cost of marketing spent to acquire 1 new customer.

1. I would first ask them what their spend budget would be for next month.
2. I would also ask them if they will be willing to spend their budget mainly on conversion type campaigns as seen previously. Since this is a key driver in conversions, we could potentially negotiate a similar split of campaign type as seen in the data.
3. I would also ask what month the next one would be. Since CPA isn’t a static value, we see a lot of seasonality. We can categorize them into slow months (like January, February, April May, and so on). It makes sense to spend less money if the next month is a slow month and not a fast month (March, June).

4. In order to project CPA, we should also try to determine if the next month is a peak month.
5. Is a competitor entering the market next month/quarter?
6. What is consumer demand like for the product we are selling? Stagnant? Growing? Expanding rapidly? We can adjust guidance by using this.
7. What is the general news and consensus like about this product? We can create different baselines for good market conditions to an extreme poor market condition to model different market sentiment.