Restaurant Marketing Campaign A/B Testing

CS555 Final Project (Prof. Alizadeh-Shabdiz, Fall 2022)

Cody Thayer | December 14, 2022

Project Files

> **HTML file** is the most readable and has intractable graphs, but the source code of all the figures is hidden for clarity.
>
> **RMD file** contains all the source code.
>
> **PDF file** is just a pdf version of the RMD file.

```
library(ggplot2)
library(gridExtra)
library(tidyverse)
library(grid)
library(lattice)
library(plotly)
library(kableExtra)
library(GGally)
library(viridis)
library(emmeans)
library(car)


# Colors
colors_3 = c(viridis(25)[4], viridis(25)[10], viridis(25)[15], viridis(25)[19])
colors_4 = c(viridis(25)[1], viridis(25)[4], viridis(25)[10], viridis(25)[15], viridis(25)[19])
```

1. Describe your research scenario and questions(s).

A restaurant chain is changing its menu by adding a few new items. The restaurant wantx to test three different marketing campaigns they are thinking about using to promote their new menu items. The restaurant wants to determine which promotion has the greatest effect on sales of new menu items. The restaurant locations in which the campaigns will be tested are selected at random. Each randomly selected location will begin serving the new menu items, and run one of the three marketing campaigns, also chosen at random. The weekly sales of the new menu items are recorded for the first four weeks of the marketing campaign.

The restaurant chain is hiring a data scientist to evaluate the effectiveness of three different marketing campaigns. The restaurant wants to determine which of the three marketing strategies is most effective at increasing sales of new menu items, so they can confidently implement said marketing campaign at all their restaurant locations.

Research Question: Which marketing campaign(s) results in the greatest increase in sales of new menu items for the restaurant chain?

Many auxillary questions will be answered in our journey to answer the above question, but the above question is the core, guiding inquiry of our investigation.

2. Describe the data set.

Data Source: https://www.kaggle.com/datasets/chebotinaa/fast-food-marketing-campaign-ab-test?resource=download (https://www.kaggle.com/datasets/chebotinaa/fast-food-marketing-campaign-ab-test?resource=download)

```
# Input Data from csv file.
data = read.csv(file = "WA_Marketing-Campaign.csv", header = TRUE); data

# Factor the Data to make it easier to work with.
data$Promotion = factor(data$Promotion); data$Promotion
levels(data$Promotion)
data$week = factor(data$week)
levels(data$week)
tibble(data)

# Create Subsets for each Marketing Campaign to make our life easier.
promo_1 = filter(data, Promotion == 1); promo_1
promo_2 = filter(data, Promotion == 2); promo_2
promo_3 = filter(data, Promotion == 3); promo_3

# Data was luckily very clean, not a lot of data cleaning was necessary.
```

```
print(data)
```

	Description	Values
MarketID	Unique identifier for market.	Range(1:10)
MarketSize	Size of market area by sales	Small, Medium, Large.
LocationID	Unique identifier for store location	
AgeOfStore	Age of store in years	
Promotion	One of three promotions that were tested	Range(1:3)
week	One of four weeks when the promotions were run	Range(1:4)
SalesInThousands	Sales amount for a specific *LocationID*, *Promotion*, and *week*	

SalesInThousands

Lets first look at the **SalesInThousands** variable for each of the three marketing campaigns. The **SalesInThousands** variable is the sales amount in thousands of new menu items for a specific restaurant.

Exploratory Question: Are the sales of new menu items similar for each marketing campaign group?

```
# Summary Statistics
# Mean
mean_1 = mean(promo_1$SalesInThousands); mean_1
mean_2 = mean(promo_2$SalesInThousands); mean_2
mean_3 = mean(promo_3$SalesInThousands); mean_3
mean_tot = mean(data$SalesInThousands); mean_tot

# Standard Deviation
std_1 = sd(promo_1$SalesInThousands); std_1
std_2 = sd(promo_2$SalesInThousands); std_2
std_3 = sd(promo_3$SalesInThousands); std_3
std_tot = sd(data$SalesInThousands); std_tot

# Number of Samples
n = nrow(data); n
k = length(levels(data$Promotion))
n_1 = nrow(promo_1); n_1
n_2 = nrow(promo_2); n_2
n_3 = nrow(promo_3); n_3
```

Marketing Campaign	Sales (In Thousands)	Standard Deviation	Population
Promo 1	58.10	16.55	172
Promo 2	47.33	15.11	188

Marketing Campaign	Sales (In Thousands)	Standard Deviation	Population
Promo 3	55.36	16.77	188
Total	53.47	16.75	548

Table 1.0: Summary statistics for sales of new menu items for all restaurants participating in the new marketing campaigns.

The three different marketing campaigns subsets (Promotion 1, Promotion 2, and Promotion 3) have very symmetric data, all having similar means, standard deviations and population sizes seen in **Table 1.0**. Lets create a boxplot to visualize this symmetry.

```
# Sales Boxplot
data_2 = data
data_2$Promotion = as.character(data_2$Promotion)

sales_promo_box = ggplot(data_2, aes(x=Promotion, y=SalesInThousands, fill=Promotion)) +
  ylab("Sales (in thousands)") +
  xlab("Marketing Campaign") +
  geom_boxplot() + coord_flip() +
  geom_boxplot(color="black") +
  scale_fill_viridis_d() +
  geom_point(shape=16, position=position_jitter(0.1), colour="black", alpha=.50) +
    theme(panel.grid.major = element_line(color="gray90"), panel.grid.minor = element_line(color="gray94"),
        panel.background = element_blank(), axis.line = element_line(colour = "gray90", size =0.5)) +
  theme(panel.border = element_rect(fill=NA, colour = "gray90", size = 0.5, linetype = "solid")) +
  theme(axis.ticks = element_line(color="gray90", size=0.5)) +
  theme(axis.ticks.length = unit(0.18, "cm")) +
  theme(legend.position="none")

sales_promo_box
```
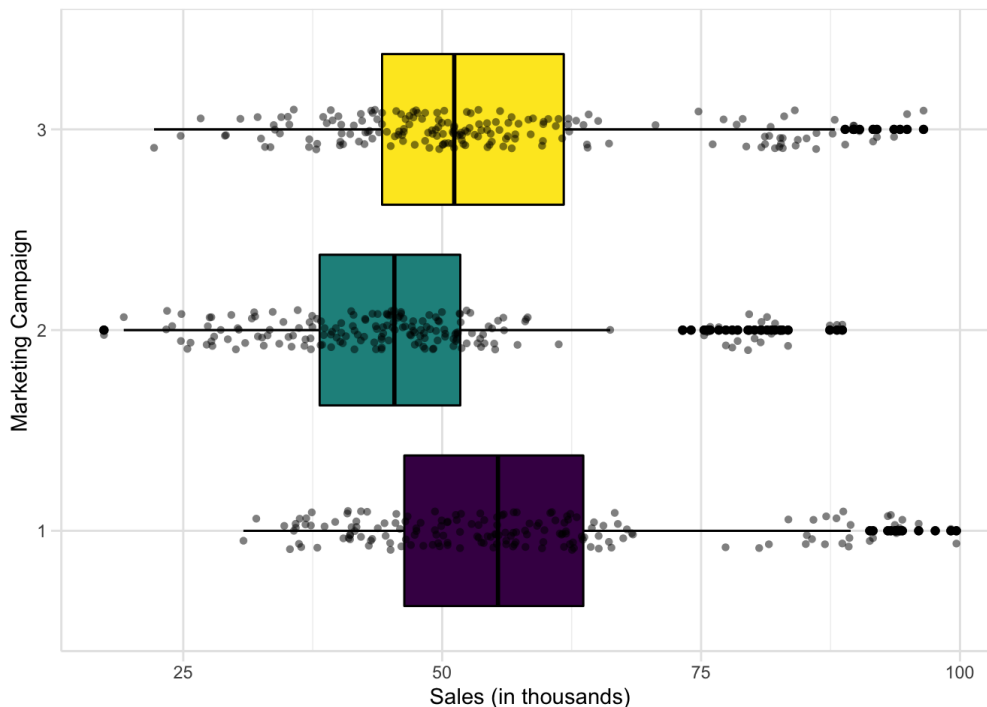

Figure 1.0: Box plot showing the difference in restaurant sales data for each of the three marketing campaigns. Promotion 1 is in purple, Promotion 2 is in green, and Promotion 3 is in yellow.

Promotion 1 has the highest mean sales of new menu items, followed by Promotion 3 and then Promotion 2. All have similar variance, which is important when comparing means. However, there seems to be more Promotion 2 outliers. Lets see if that is true.

```
# Identify Outliers using IQR Methology

IQR = summary(promo_1$SalesInThousands)[5] - summary(promo_1$SalesInThousands)[2]; IQR
outliers = filter(promo_1, SalesInThousands > summary(promo_1$SalesInThousands)[5] + (1.5 * IQR) | SalesInThousan
ds < summary(promo_1$SalesInThousands)[3] - (1.5 * IQR)); outliers$SalesInThousands

IQR = summary(promo_2$SalesInThousands)[5] - summary(promo_2$SalesInThousands)[2]; IQR
outliers_2 = filter(promo_2, SalesInThousands > summary(promo_2$SalesInThousands)[5] + (1.5 * IQR) | SalesInThous
ands < summary(promo_2$SalesInThousands)[3] - (1.5 * IQR)); outliers_2$SalesInThousands

IQR = summary(promo_3$SalesInThousands)[5] - summary(promo_3$SalesInThousands)[2]; IQR
outliers_3 = filter(promo_3, SalesInThousands > summary(promo_3$SalesInThousands)[5] + (1.5 * IQR) | SalesInThous
ands < summary(promo_3$SalesInThousands)[3] - (1.5 * IQR)); outliers_3$SalesInThousands

nrow(outliers)
nrow(outliers_2)
nrow(outliers_3)
```

Using the IQR methodology to identity outliers, we observed 12 sales outliers each for both Promotion 1 and 2, and 30 sales outliers for Promotion 3. This is an imbalance we should keep in mind and circle back to before we make any final conclusions about our data.

To answer our exploratory question, **the sales of new menu items were in fact similar for each marketing campaign group but not the same.** Promotion 2 looks to deviate from Promotion 1 and Promotion 2, and we will test to see if that deviation is significant.

AgeOfStore

Lets next look at the *AgeOfStore* variable for each of the three marketing campaigns. The *AgeOfStore* variable is age of each restaurant in the study.

Exploratory Question: Are age of restaurant distributed different in each marketing campaign group?

```
# Scatter Plot Age vs. Sales
x = data$SalesInThousands
y = data$AgeOfStore

age_sales_scatter = plot_ly(data, y = ~SalesInThousands, x = ~AgeOfStore, color = ~Promotion, colors = colors_3,
        type = 'scatter', mode = 'markers',
        marker = list(symbol = 'circle', sizemode = 'diameter',
                      line = list(width = 1, color = 'black'))) %>%
    layout(xaxis = list(title = 'Age of Restaurant',
                    range = c(min(x) - mean(x)/nrow(x), max(x) + mean(x)/nrow(x)),
                    zerolinewidth = 2,
                    tickwidth = 1.5,
                    ticklen = 5,
                    tickprefix="",
                    gridwidth = 1),
        yaxis = list(title = 'Sales (in thousands)',
                    range = c(min(y) - mean(y)/nrow(y), max(y)) + mean(y)/nrow(y),
                    zerolinewidth = 2,
                    tickwidth = 1.5,
                    ticklen = 5,
                    gridwith = 1))

age_sales_scatter
```
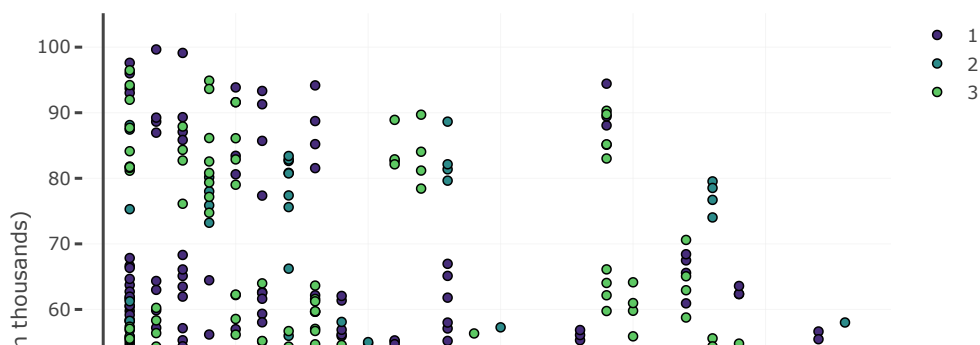
Figure 2.0: *Scatter plot showing the relationship between **AgeOfStore** (x) and **SalesInThousands** (y) for all restaurants conducting marketing campaigns for new menu items. Double-click on the legend to view individual marketing campaigns.*

Looking at the scatter plot of age and sales for restaurants we observe no clear linear relationship between the two variables. We can confirm this with a calculation of the correlation coefficient, r.

```
# Correlation Coefficient
cor(x=data$AgeOfStore, y=data$SalesInThousands)
```

Our the calculated correlation coefficient, r, of store age and store sales is -0.029, which is almost zero, meaning there is little to no linear relationship between the two variables.

```
# Age Summary Statistic
tapply(data$AgeOfStore, data$Promotion, summary)

summary(promo_1$AgeOfStore)
sd(promo_1$AgeOfStore)
summary(promo_2$AgeOfStore)
sd(promo_2$AgeOfStore)
summary(promo_3$AgeOfStore)
sd(promo_3$AgeOfStore)
```

	Promo 1	Promo 2	Promo 3
Mean	8.28	7.98	9.23
Median	6	7	8
Standard Deviation	6.64	6.60	6.65
(Min, Max)	(1, 27)	(1, 28)	(1, 24)
(Q1, Q3)	(3, 12)	(2, 10)	(5, 12)

Table 2.0: *Summary statistics for store age for all restaurants participating in the marketing campaign study.*

```
# Age Boxplot
data_2 = data
data_2$Promotion = as.character(data_2$Promotion)

age_promo_box = ggplot(data_2, aes(x=Promotion, y=AgeOfStore, fill=Promotion)) +
  ylab("Age of Restaurant") +
  xlab("Marketing Campaign") +
  geom_boxplot() + coord_flip() +
  geom_boxplot(color="black") +
  scale_fill_viridis_d() +
  geom_point(shape=16, position=position_jitter(0.1), colour="black", alpha=.50) +
    theme(panel.grid.major = element_line(color="gray90"), panel.grid.minor = element_line(color="gray94"),
        panel.background = element_blank(), axis.line = element_line(colour = "gray90", size =0.5)) +
  theme(panel.border = element_rect(fill=NA, colour = "gray90", size = 0.5, linetype = "solid")) +
  theme(axis.ticks = element_line(color="gray90", size=0.5)) +
  theme(axis.ticks.length = unit(0.18, "cm")) +
  theme(legend.position="none")

age_promo_box
```
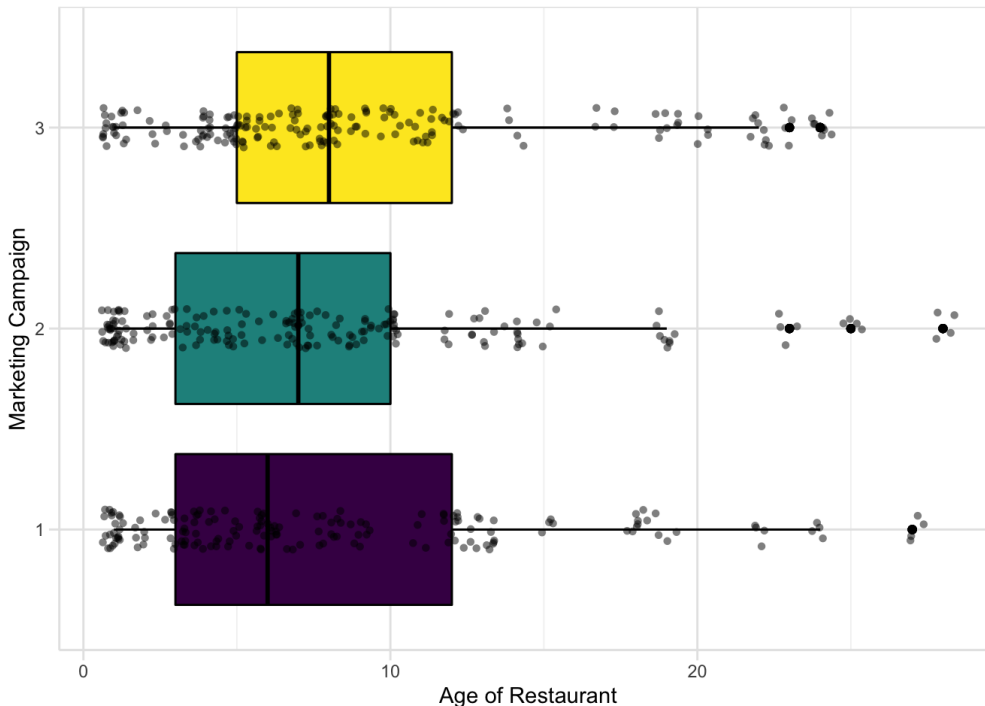

Figure 3.0: *Box plot showing the difference in restaurant age data for each of the three marketing campaigns. Promotion 1 is in purple, Promotion 2 is in green, and Promotion 3 is in yellow.*

Both **Table 2.0** and **Figure 3.0** show that the age distributions among the three marketing campaign is very similar. Promotion 1, 2 and 3 have a mean ages of 8.28, 7.98, and 9.23 respectively and all have standard deviations ~ 6.0. All have some clustering around age 1, and have some outliers above Q3. All three marketing campaigns age distributions are not identical but very similar for real world data standards. To answer our exploratory question, **the age of restaurant is not distributed significantly different in each marketing campaign.**

MarketSize

Lets next look at the *MarketSize* variable for each of the three marketing campaigns. The *MarketSize* variable is a categorical variable that describes the market size of each restaurant. There are three market sizes: small, medium and large. The thresholds and criteria for categorization are unknown.

Exploratory Question: Are market size proportions similar in each marketing campaign group?

```
# Stack Bar Plot (Total Sales)
market_bar = plot_ly(data, y=~Promotion, x=~SalesInThousands, color=~MarketSize, colors = colors_4, type = "bar")
%>%
  layout(barmode="stack") %>%
  layout(yaxis = list(title="Marketing Campaign", categoryorder = "array", categoryarray = data$Promotion,
                      linecolor = "white")) %>%
  layout(yaxis = list(minor = list(showgrid = FALSE))) %>%
  layout(xaxis = list(title="Total Sales (in thousands)"), legend = list(title=list(text='Market Size')))

market_bar
```
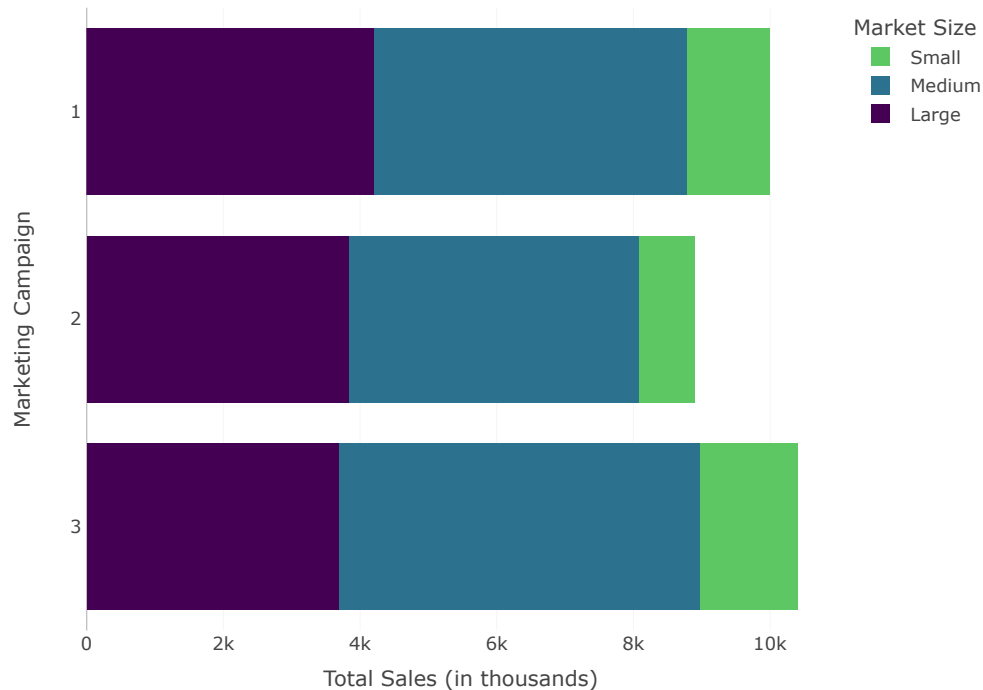

Figure 4.0: *Stacked bar plot showing the difference in total sales between marketing campaigns stratified by market size.*

For all three campaigns the greatest proportion of total sales can be attributed to medium market restaurants, large market restaurants are the next biggest contributor to total sale of new menu items, and small market restaurants contributioned the least amount. The overall trend is similar for all three marketing campaigns.

```
# Stack Bar Plot (Count)
market_data =  data %>%
group_by(Promotion, MarketSize) %>%
summarize(Count=n(), sales_mean=mean(SalesInThousands), sales_total=sum(SalesInThousands)); market_data
```

```
#visualizing data
market_count_bar = plot_ly(market_data, y=~Promotion, x=~Count, color=~MarketSize, colors = colors_4, type = "bar") %>%
  layout(barmode="stack") %>%
  layout(yaxis = list(title="Marketing Campaign", categoryorder = "array", categoryarray = data$Promotion,
                      linecolor = "white")) %>%
  layout(yaxis = list(minor = list(showgrid = FALSE))) %>%
  layout(xaxis = list(title="Restaurant Count"), legend = list(title=list(text='Market Size')))

market_count_bar
```
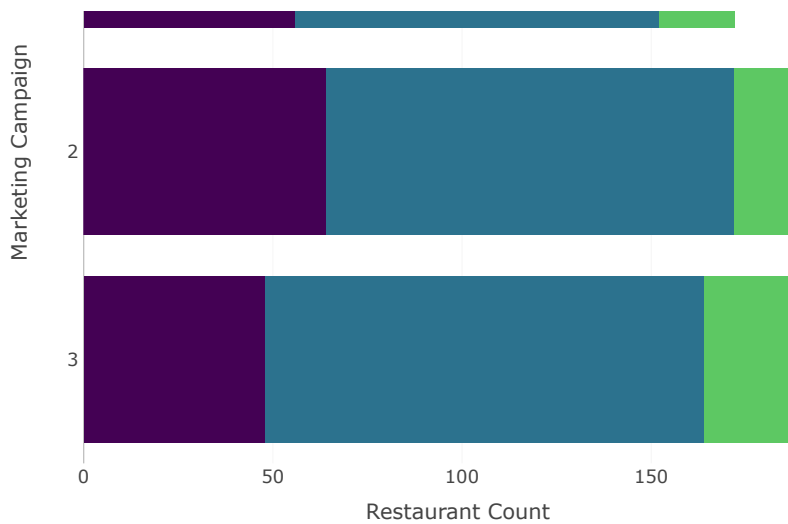
Figure 5.0: Stacked bar plot showing the differences in amount of restaurant in each of the three market sizes (small, medium, large) for each of the three marketing campaigns.

Each marketing campaign has the most amount of medium market restaurants followed by large market restaurants, and the least small market restaurants. Promotion 1, Promotion 2, and Promotion 3 had 56, 64, and 48 large market restaurants respectively, 96, 108, and 116 medium market restaurants respectively, and 20, 16, and 24 small market restaurants respectively. The overall trend is similar for all three marketing campaigns.

Lets consolidate these trends in **Figure 4.0** and **Figure 5.0** into one scatter plot, to best visualize the different amount of restaurants in each market size for each of the three marketing campaigns, and how those markets contributed to the marketing groups sales.

```
market_count_sales = plot_ly(market_data, y = market_data$Count, x = market_data$sales_mean, color = ~Promotion,
colors = colors_3, type = 'scatter', mode = 'markers', sizes = c(10, 100),
        marker = list(symbol = 'circle', sizemode = 'diameter',
                      line = list(width = 2, color = 'black')),
        text = ~paste('Marketing Campaign:', Promotion, '<br>Market:', MarketSize, '<br>Restaurant Count:', Coun
t, '<br>Total Sales:', sales_total)) %>%
layout(xaxis = list(title = 'Mean Sales (in thousands)',
                    range = c(0, 100),
                    zerolinewidth = 2,
                    tickwidth = 1.5,
                    ticklen = 5,
                    tickprefix="$",
                    gridwidth = 2),
       yaxis = list(title = 'Restaurant Count',
                    range = c(0, 150),
                    zerolinewidth = 2,
                    tickwidth = 1.5,
                    ticklen = 5,
                    gridwith = 2))

market_count_sales
```
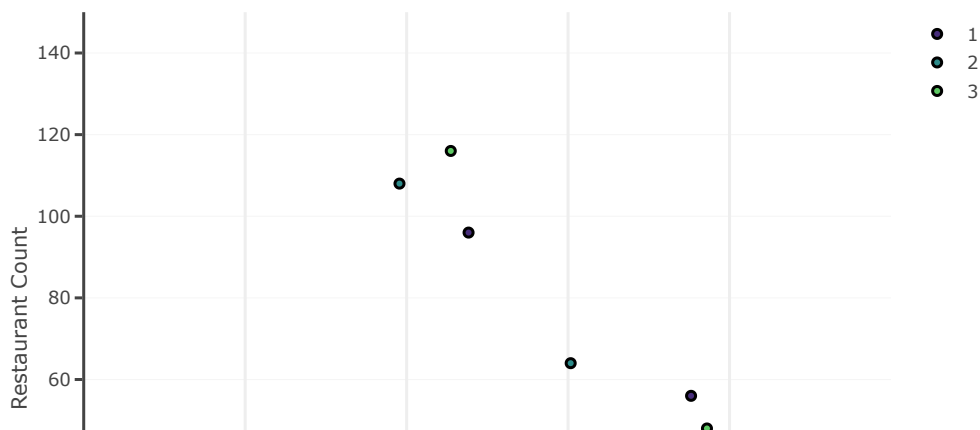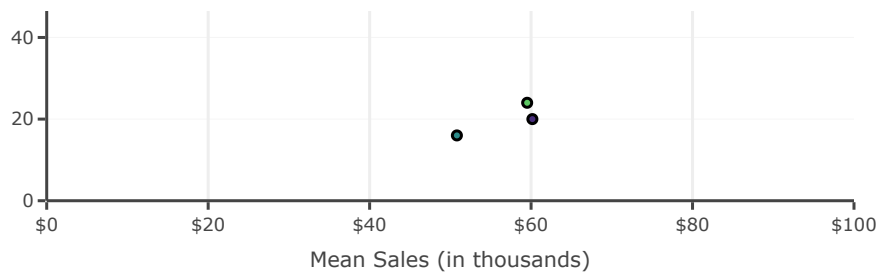
Figure 6.0: Scatter plot showing the relationship between Mean Sales (x) and Restaurant Count (y) for all restaurants conducting marketing campaigns for new menu items. Additionally, if you hover over each point the Total Sales of each sub group is listed.

The big take away from **Figure 6.0** is that **Promotion 2 had the most large market restaurants, but the lowest total large market sales (**the total sales can be seen if you hover over the data point in **Figure 6.0**). In fact,all of Promotion 2's restaurants, no matter the market size, under-performed, but Promotion 2's large market restaurants were the most significant under-performing market for the Promotion 2 cohort. **Figure 6.0** identifies the reason Promotion 2 has the lowest mean sales, but also shows that large market restaurants under-performed more than Promotion 2's small and medium market restaurants.

Answering our exploratory question, the market size proportions were similar for each marketing campaign group, and each market size contributed proportionality based on their market size to the total sales of the marketing campaign group. This trend is visualized in **Figure 6.0**, which shows three clusters. a small market cluster, a medium market cluster, and a large market cluster. The medium market cluster had the lowest mean sales, while the large market had the highest mean sales.

Week

Lastly, lets look at the *Week* variable for each of the three marketing campaigns. The *Week* variable ranges from week 1 through week4, time stamping the four weeks of marketing campaign study. Sales amounts for each restaurant were recorded for each of the four weeks.

Exploratory Question: Are some campaigns trending differently? Low sales in week 1, and higher in week 4?

```
# stacked bar plot, percent
week_bar = plot_ly(data, y=~Promotion, x=~SalesInThousands, color=~week, colors = colors_4, type = "bar") %>%
  layout(barmode="stack") %>%
  layout(yaxis = list(title="Marketing Campaign", categoryorder = "array", categoryarray = data$Promotion,
                      linecolor = "white")) %>%
  layout(yaxis = list(minor = list(showgrid = FALSE))) %>%
  layout(xaxis = list(title="Total Sales (in thousands)"), legend = list(title=list(text='Week')))

week_bar
```
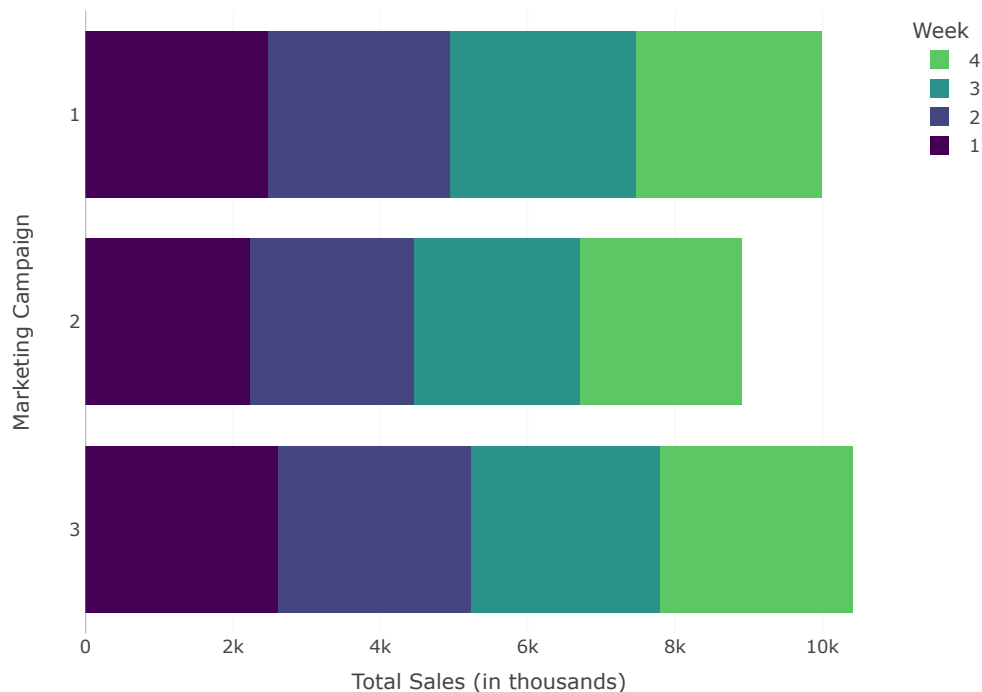
There are no clear differences in marketing campaign performance week to week for any of the three promotions. There are no visible trends that shows an increase or decrease in marketing campaign performance week by week for any promotion. Each promotions weekly performance from week 1 through week 4 is consistent, each week accounting for about a quarter of the monthly sales.

Overall Conclusion

The data for all three underlying populations, Promotion 1 restaurants, Promotion 2 restaurants, and Promotion 3 restaurants, is extremely similar when we look at *SalesInThousands*, and its relation with all the relevant variables: *AgeOfStore*, *MarketSize*, and *Week*. Symmetry in our data is what were are most concerned about when comparing underlying populations, and we observed no significant irregularities. We can confidently proceed to comparing the *SalesInThousands* means for each marketing campaign.

3. Describe the statistical methods you plan to use.

Analysis of Variance (ANOVA)

The goal of a one-way analysis of variance (ANOVA) is to compare means across groups. There are many procedures to compare means across two groups, but ANOVA is a procedure that extends this methodology to a setting where there are more than two groups. In our scenario, we have three groups, the three different marketing campaigns, and we are interesting in comparing mean sales of new menu items between the three campaigns.

It is important to remember that when we conduct an ANOVA test, we are comparing across several groups while considering all of the data together. We have three marketing campaigns, if we did not use an ANOVA test, we would likely resort to two-sample t-testing different combinations of the three marketing campaigns. However, conclusions are often different when all means are considered instead of just a pair of means. ANOVA allows us to take into account all three means before we make a conclusion. The difference in the test statistic, F, from the test statistic of a two-sample t-test is fundamental to this idea of taking into account more than two means. The equation for the test statistic for an ANOVA test is shown below:

$$ F = \frac{\text{Between Group Variance}}{\text{Within Group Variance}} $$

The deviation between an individual point and the respective group mean is representative of the **within group variance**. In our scenario, this is the difference between the sales of an individual restaurant who is using Promotion 1 and the mean sales of all restaurant who is using Promotion 1. The deviation between the group mean and the overall mean is representative of the **between group variance.** In our scenario, this is the difference between the Promotion 1 mean sales and the means sales of all three marketing campaigns (and the same for the other two campaigns). These deviations are squared and the sum over all points is taken to obtain the sum of squares within and the sum of squares between, respectively.

That is a lot of words. The important take away from the above description of F is that the **ANOVA methodology takes into account the overall mean when analyzing variance of a data set.** And because the overall mean is a function of all groups in the data set, the ANOVA test is different than simply performing multiple two-sample t-test, which only uses the means of two groups.

A good rule of thumb when conducting ANOVA tests is **if the between group variance is large and the within group variance is small, then the null hypothesis is often rejected** in favor of the alternative hypothesis that the underlying means across group are not all equal. In our sales data, we observe the opposite. As seen in **Figure 1.0**, the between group variability is small and the within group variability is large. Following our rule of thumb above, we should expect to fail to reject our null hypothesis that the mean sales of the three marketing campaigns are equal. After we conduct our ANOVA test below, we will see if our hypothesis is correct.

Tukey's Pairwise T-Tests

Simply put, the pairwise t-test for an ANOVA test allows us determine between which groups the differences (we deemed significant in our ANOVA test) lie. The equation for the test statistic for a Tukey's pairwise t-test in the context of an ANOVA test is shown below:

$$t = \frac{\bar{x}_{promo-1} - \bar{x}_{promo-2}}{\sqrt{s^{2}_{p}(\frac{1}{n_{promo-1}} + \frac{1}{n_{promo-2}})}}$$

The key take away from this equation, is that the t-statistic is a function of the variance from the ANOVA, s^{2}_{p}, as opposed to using the pooled estimate of the variance from only groups involved in the pairwise t-test. This is a what differentiates a Tukey's pairwise t-tests for ANOVA from a regular two-sample t-test. If we only used the data from groups we would have ended up with a slightly different estimate of the variance, and thus a slightly different t-statistic. Because we **assume that all groups have the same variance when conducting an ANOVA test,** we do not want to use different estimates of the variance in each pairwise comparison.

Given that the estimate from ANOVA of the variability is based on more information (more observations are included), it will be more accurate than that obtained from using the data from just two of the groups. Our estimate of variability in an ANOVA test, s^{2}_{p}, is therefore calculated from a larger sample size (all groups, not just the groups involved in the pairwise t-test). This estimate of variability is more accurate, and is one of the primary advantages of ANOVA pairwise t-tests.

Something to keep in mind when running multiple pairwise t-tests is that the more pairwise tests we run the more errors we are likely to observe due to chance alone. The more you do just about anything the more likely you are to make a mistake because everytime you do something you give yourself the opportunity to make a mistake. If you do nothing, you can't make a mistake. In our scenario, an error is a **Type I Error, a false positive conclusion** (rejecting the null hypothesis when in reality the null hypothesis is in fact true).

In order to reduce the amount of Type I Errors we have to make it harder for us to reject the null hypothesis during our pairwise t-tests. One way of doing this is to increase our confidence level (decreasing α). Therefore, as we increase the number of pairwise t-tests we run, we also increase our confidence level for those pairwise t-tests in order to balance out our increasing chance for errors. This allows us to **maintain a specified family-wise Type I error rate**. Family-wise Type I error rate meaning the error rate when looking at all pairwise t-tests, not just the Type I error rate of individual pairwise t-tests.

Procedures for controlling the family wise type I error rate at a pre-specified level are called multiple comparison procedures. Luckily for us, we are only conducting three pairwise tests, but in our ANOVA testing we will still use the **Tukey multiple comparison procedure**, a complex procedure that requires statistical software (we will be using R). This will reduce our Type 1 error rate.

Analysis of Covariance (ANCOVA)

Analysis of Covariance (ANCOVA) is really just a variation of a linear regression model, and an extension of the ANOVA methodology. We are still interesting in the different group means like we are when doing ANOVA, but when doing ANCOVA, we are adjusting those group means by another variable.

AgeOfStore is this new variable we will be introducing, and we want to create an ANCOVA model to see how the age of the restaurants in a certain promotional group impacts sales. The question we are asking when conducting our ANCOVA test is: **does the mean restaurant sales still differ by marketing campaign when the age of restaurants is taken into account?**

Both ANCOVA and ANOVA testing involve mean analysis; however, the means of the underlying populations are different. When we conduct an ANOVA test we are comparing non-adjusted group means, where as in and ANCOVA test we are comparing adjusted group means. When we conduct our ANOVA testing we will not take into account *AgeOfStore* when analyzing each marketing campaigns mean sales. However, when we conduct our ANCOVA test, we do take into account *AgeOfStore*, and adjust our means accordingly before analyzing marketing campaign mean sales variance. In other words, when we are conducting an ANCOVA test, we follow the same procedural steps of an ANOVA test just with adjusted underlying population means.

4. Report your results.

ANOVA Global F-Test

Test Question: Do new menu item sales means vary by marketing campaign?

Since we have three different promotions, we will start will start by doing an analysis of variance (ANOVA), global f-test. If there were only two promotional campaigns, we could simply do a pairwise t-test to see if the two groups differ from each other. Reasons are discussed above.

STEP 1: Set up the hypotheses and select the alpha level.

$H_{0}: \mu_{promo-1} = \mu_{promo-2} = \mu_{promo-3}\text{ (All promotions result in the same amount of sales)}$

$H_{a}: \mu_{promo-1} \ne \mu_{promo-2} \text{ or } \mu_{promo-2} \ne \mu_{promo-3} \text{ or } \mu_{promo-1} \ne \mu_{promo-3}\text{ (Not all promotions result in the same amount of sales)}$

$\alpha = 0.05$

STEP 2: Select the appropriate test-statistic, f.

$f = \frac{\text{Mean Squares Between}}{\text{Mean Squares Within}}$

STEP 3: State the decision rule.

Use R to find $F_{k-1, n-k, \alpha}$ associated with a right hand tail probability of $\alpha = 0.05$.

```
qf(0.95, df1=k-1, df2=n-k)
```

Reject $H_{0}: \mu_{promo-1} = \mu_{promo-2} = \mu_{promo-3}$ if either are True:

1. $f \ge F_{2, 546, 0.05} = 3.01$
2. $p < 0.05$

STEP 4: Compute the test statistic, f, and associated p-value, p using aov().

```
# One-Way ANOVA
test_1 = aov(formula = SalesInThousands ~ Promotion, data = data); test_1
summary_test_1 = summary(test_1); summary_test_1
```

$f = 21.95$

$p = 6.7*10^{-10} \sim 0$

STEP 5: State your conclusion.

Because $f = 21.95 > 3.01 = F_{2,546, 0.05}$, and $p = 6.7*10^{-10} < 0.05 = \alpha$, we reject the $H_{0}: \mu_{promo-1} = \mu_{promo-2} = \mu_{promo-3}$

We have sufficient evidence that at the $\alpha = 0.05$ level that there is a difference in mean sales between marketing campaigns.

To answer our test question, since the global f-test was significant, new menu items sales means do in fact vary by marketing campaign. We now want to look into the pairwise comparisons of the sales means for each promotion.

Tukey's Pairwise t-Tests

Test Question: Between which marketing campaigns does our difference lie?

Our global f-test has determined that there is a difference in mean sales between marketing campaigns, so we will conduct multiple Tukey Pairwise t-tests to determine which marketing campaign's mean sales differ.

The same Tukey's pairwise t-testing procedure shown for the comparison of Promotion 1 and Promotion 2 mean sales will be used for comparing Promotion 2 and Promotion 3 mean sales, and Promotion 1 and Promotion 3 mean sales.

Promotion 1 - Promotion 2

STEP 1: Set up the hypotheses and select the alpha level.

$H_{0}:\mu_{promo-1} = \mu_{promo-2}$

$H_{a}: \mu_{promo-1} \ne \mu_{promo-2}$

$\alpha = 0.05$

STEP 2: Select the appropriate test-statistic, t.

$t = \frac{\bar{x}_{promo-1} - \bar{x}_{promo-2}}{\sqrt{s^{2}_{p}(\frac{1}{n_{promo-1}} + \frac{1}{n_{promo-2}})}}$

STEP 3: State the decision rule.

```
qt(0.975, df = n-k)
```

Reject $H_{0}:\mu_{promo-1} = \mu_{promo-2}$ if either are True:

1. $t \ge T_{n-k,\frac{\alpha}{2}} = 1.96\text{ or } t \le -T_{n-k,\frac{\alpha}{2}}= -1.96$
2. $p < \frac{\alpha}{2} = 0.025 \text{ (Right hand tail probability)}$
3. 0 is not within the bounds of our confidence interval.

STEP 4: Compute the t-statistic, t, confidence interval, and associated p-value, p.

```
# Manual Calculation of t_value
t_value = (mean_1 - mean_2) / sqrt((std_1**2/n_1 + std_2**2/n_2)); t_value
# Welch t-test
t.test(promo_1$SalesInThousands, promo_2$SalesInThousands)

# Tukey Comparison of Means
TukeyHSD(test_1)
# Non-Adjusted p-values (will not use just for clarity)
pairwise.t.test(data$SalesInThousands, data$Promotion, p.adj="none")
```

$t = 6.43$

$\text{Confidence Interval = (6.77, 14.8)}$

$p = 4.29*10^{-10} \sim 0$

STEP 5: State your conclusion.

Because $p = 4.29*10^{-10} < 0.025 = \frac{\alpha}{2}$, and $t = 6.43 \ge 1.96 = T_{545,0.025}$, we reject the $H_{0}:\mu_{promo-1} = \mu_{promo-2}$

Using the Tukey Pairwise T-Test, we have sufficient evidence that at the $\alpha = 0.05$ level that the mean sales of restaurants using Promotion 1, $\mu_{promo-1} = 58.10$, and restaurants using Promotion 2, $\mu_{promo-2} = 47.33$, are different from one another. The observed difference in mean sales between restaurants conducting Promotion 1 and restaurants conducting Promotion 2 was 10.77 thousand dollars.

We are 95% confident that the true difference between the sales means is between 6.77 and 14.8 thousand dollars. Another interpretation of our null hypothesis states that $H_{0}:\mu_{promo-1} - \mu_{promo-2} = 0$, testing to see if the difference between the two means is zero. Since a difference of 0 is not in our confidence interval, we can again reject the $H_{0}:\mu_{promo-1} = \mu_{promo-2}$.

Promotion 2 - Promotion 3

```
t_value = (mean_2 - mean_3) / sqrt((std_2**2/n_2 + std_3**2/n_3)); t_value
t.test(promo_3$SalesInThousands, promo_2$SalesInThousands)
```

Because $p = 1.57*10^{-6} < 0.025 = \frac{\alpha}{2}$, and $t = 4.88 \ge 1.96 = T_{545,0.025}$, we reject the $H_{0}:\mu_{promo-2} = \mu_{promo-3}$

Using the Tukey Pairwise T-Test, we have sufficient evidence that at the $\alpha = 0.05$ level that the mean sales of restaurants using Promotion 2, $\mu_{promo-2} = 47.33$, and restaurants using Promotion 3, $\mu_{promo-3} = 55.36$, are different from one another. The observed difference in mean sales between restaurants conducting Promotion 2 and restaurants conducting Promotion 3 was -8.03 thousand dollars.

We are 95% confident that the true difference between the sales means is between -4.12 and -11.95 thousand dollars. Another interpretation of our null hypothesis states that $H_{0}:\mu_{promo-2} - \mu_{promo-3} = 0$, testing to see if the difference between the two means is zero. Since a difference of 0 is not in our confidence interval, we can again reject the $H_{0}:\mu_{promo-2} = \mu_{promo-3}$.

Promotion 1 - Promotion 3

```
t_value = (mean_1 - mean_3) / sqrt((std_1**2/n_1 + std_3**2/n_3)); t_value
t.test(promo_1$SalesInThousands, promo_3$SalesInThousands)
```

Because $p = 0.24 > 0.025 = \frac{\alpha}{2}$, and $t = 1.55 < 1.96 = T_{545,0.025}$, we fail to reject the $H_{0}:\mu_{promo-1} = \mu_{promo-3}$

Using the Tukey Pairwise T-Test, we have sufficient evidence that at the $\alpha = 0.05$ level that the mean sales of restaurants using Promotion 1, $\mu_{promo-1} = 58.10$, and restaurants using Promotion 3, $\mu_{promo-3} = 55.36$, are different from one another. The observed difference in mean sales between restaurants conducting Promotion 1 and restaurants conducting Promotion 3 was -2.73 thousand dollars.

We are 95% confident that the true difference between the sales means is between -1.26 and 6.74 thousand dollars. Another interpretation of our null hypothesis states that $H_{0}:\mu_{promo-1} - \mu_{promo-3} = 0$, testing to see if the difference between the two means is zero. Since a difference of 0 is in our confidence interval, we also fail to reject the $H_{0}:\mu_{promo-1} = \mu_{promo-3}$.

ANCOVA

Test Question: Does the mean restaurant sales still differ by marketing campaign when the age of restaurants is taken into account?

Through our investigation of our data in Part 1, we have no reason to suspect a correlation between **SalesInThousands**, and **AgeOfStore**, **MarketSize**, or **Week**. However, for completeness, lets look at the **AgeOfStore** variable, and conduct a quick ANOVA test using R to confirm that it has no effect on the variance in mean sales between marketing campaigns.

```
# Interested in mean sales, Adjusting for Age
# ORDER MATTERS when you use aov()
ancova_model = aov(SalesInThousands ~ Promotion + AgeOfStore, data = data)
Anova(ancova_model, type="III")
```

Table 5.0: *ANOVA Table for our ANCOVA Model that takes into account both marketing campaign (**Promotion**) and **AgeOfStore**.*

We see in **Table 5.0** that **AgeOfStore** is not significantly associated with the mean sales of a restaurant with a large $p = 0.3326$, while the marketing campaign remains to be significantly associated with the mean sales with a small $p = 5.49*10^{-10}$. Additionally, the f-statistic for the marketing campaign variable has not changed, remaining high at around 22.

Therefore, we do not have sufficient evidence at the $\alpha = 0.05$ level that the difference in amount of sales between marketing campaigns is associated with the **AgeOfStore**. This means that the mean sales of each marketing campaign will not change significantly when we take into account **AgeOfStore**. Therefore, there is no need to look at the pairwise comparisons of the sales numbers for each promotion after adjusting for **AgeOfStore**.

5. State your conclusions and discuss any limitations.

Conclusions

We concluded that **globally that their was a significant different between marketing campaigns mean sales**. In the context of our question, this indicates that we need to figure out between which marketing campaigns the difference lies, so we can identify the specific marketing campaign that is generating the most sales.

Table 6.0: *Summary of Tukey's Pairwise T-Test results.*

Comparison	t-statistic	p_{tukey} value	Conclusion
Promo 1 - Promo 2	6.43	0	Significant difference between means.
Promo 2 - Promo 3	4.88	0.0000055	Significant difference between means.
Promo 1 - Promo 3	1.55	0.24	Insignificant difference between means.

When looking at the differences in mean sales between marketing campaigns, both **Promotion 1 and Promotion 3 were significantly different that Promotion 2, but not significantly different from one another.** Both Promotion 1 and Promotion 3 have higher means than Promotion 2, meaning these are the two campaigns that resulted in more new menu items sales than than the Promotion 2 marketing campaign. Both Promotion 1 and Promotion 3 mean sales figures were different from one another, but not significantly different at the $\alpha = 0.05$ level. Therefore, either the Promotion 1 or Promotion 3 marketing campaign should be preferred over the Promotion 2 marketing campaign in order to increases the total sales of new menu items. To answer our research question, **Promotion 1 and Promotion 3 marketing campaigns are most likely to result in the greatest increase in sales of new menu items for the restaurant chain.**

```
TukeyHSD(test_1, conf.level = 0.99)
```

Even if we conduct a Tukey Pairwise T-Test at the $\alpha = 0.01$ level, the results are the same, so the restaurant should feel confident moving forwards with either Promotion 1 or Promotion 3 marketing campaigns, and should consider looking into other marketing campaign factors, costs of the campaigns, ease of implementation, staying power, etc. to determine which of the two campaigns to implement across all restaurants.

In regards to the 30 sales outliers found in the Promotion 2 subset discussed in our exploration of the data, these will have no impact on our final conclusions. These outliers all fall above the mean, and therefore drag the mean up, giving Promotion 2 an artificially large mean. If these outliers were removed, along with the outliers for the other marketing campaigns, the difference between Promotion 2 sales mean and the sales means of Promotion 1 and Promotion 3 would be even more significant. Therefore, we can disregard the effect of these outliers when making our conclusion.

We concluded that **globally that their was a significant different between marketing campaigns** (Promo-1, Promo-2, Promo-3) mean sales. In the context of our question, this indicates that we need to figure out where the difference lies, so we can identify the specific marketing campaign that is generating the most sales.

When looking at the differences in mean sales between marketing campaigns, both **Promotion 1 and Promotion 3 were significantly different that Promotion 2, but not significantly different from one another.** Both Promotion 1 and Promotion 3 have higher means than Promotion 2, meaning these are the two campaigns that resulted in more new menu items sales than than the Promotion 2 marketing campaign. Both Promotion 1 and Promotion 3 mean sales figures were different from one another, but not significantly different at the $\alpha = 0.05$ level. Therefore, **either the Promotion 1 or Promotion 3 campaign should be preferred over the Promotion 2 campaign in order to increases the total sales of new menu items.**

```
TukeyHSD(test_1, conf.level = 0.99)
```

Even if we conduct a Tukey Pairwise T-Test at the $\alpha = 0.01$ level, the results are the same, so the restaurant should feel confident moving forwards with either Promotion 1 or Promotion 3 marketing campaigns, and should consider looking into account other marketing campaign factors, costs of the campaigns, ease of implementation, staying power, etc.

Our **ANCOVA testing** adjusting for *AgeOfStore* showed no correlation between *AgeOfStore* and *SalesInThousands* means for each marketing campaign. Meaning, even when *AgeOfStore* was taken into account, we did not change our conclusion from our original ANOVA test, and maintain that at the $\alpha = 0.05$ level, mean sales difference between for Promotion 1, Promotion 2, and Promotion 3 are associated with marketing campaign.

Lastly, we have a three topics we wanted to circle back upon after we tested our data. In regards to the 30 sales outliers found in the Promotion 2 subset discussed in our exploration of the data, these will have no impact on our final conclusions. These outliers all fall above the mean, and therefore drag the mean up, giving Promotion 2 an artificially large mean. If these outliers were removed, along with the outliers for the other marketing campaigns, the difference between Promotion 2 sales mean and the sales means of Promotion 1 and Promotion 3 would be even more significant. Therefore, we can disregard the effect of these outliers when making our conclusion.

After concluding mean sales for Promotion 2 to be significantly different from Promotion 2 and 3, we should give more credence to our discussion regarding **Figure 6.0**. And I would recommend that the restaurant chain investigate Promotion 2's large market restaurants to understand why they may have under-performed. This may give clarity to some of the causes of the association we found between marketing campaign and mean sales.

Finally, our data did not follow the ANOVA rule of thumb that **if the between group variance is small and the within group variance is large then it is likely that the underlying means are all equal.** Despite this, we did find the underlying means were not all equal, and there was significant difference between underlying population means. This shows the important of testing, doing the actual math (thank you R), even when the data is visually communicating a seemingly clear conclusion.

Assumptions

In order to make inferences using ANOVA methodology, we have to make sure three conditions are met:

1. **Independent, random samples from each group in the population.**

Our data source stated that each restaurant in the marketing study was chosen at random from all of the companies restaurants. Furthermore, the marketing campaign assigned to each restaurant was also chosen at random. Each restaurant was only tested once, for one marketing campaign over one month. Assumption 1 is therefore met.

2. **Outcome of interest in group/population is normally distributed with unknown mean**
3. **Outcome of interest in each group/population has a similar unknown standard deviation**

Lets graphs some histograms of our outcome of interest, *SalesInThousands,* for each group, *Promotion*, in order to check for normality and variance similarity.

```r
# Distributions
## Promotion 1
breaks = seq(from = 20, to = 100, by=2)


promo_1_hist = ggplot(promo_1, aes(x=SalesInThousands)) +
  geom_histogram(aes(y = ..density..), breaks = breaks, color='gray80', fill="gray80") +
  ggtitle('Promotion 1 Sales Density Histrogram') +
  xlab("Sales (in thousands)") +
  ylab("Density") +
  geom_density(color = "black", fill=colors_4[1],  alpha=0.5) +
  geom_vline(xintercept = mean(promo_1$SalesInThousands), linetype="dashed", color='darkblue', size = 0.8) +
  annotate("text",color='darkblue', size = 3, x=mean(promo_1$SalesInThousands)+1.25, y=0.0075,
           label= paste("Mean Sales =", round(mean(promo_1$SalesInThousands), 2)), angle=90) +
  theme(panel.grid.major = element_line(color="gray90"), panel.grid.minor = element_line(color="gray94"),
        panel.background = element_blank(), axis.line = element_line(colour = "gray90", size =0.5)) +
  theme(panel.border = element_rect(fill=NA, colour = "gray90",
                                    size = 0.5, linetype = "solid")) +
  theme(axis.ticks = element_line(color="gray90", size=0.5)) +
  theme(axis.ticks.length = unit(0.18, "cm")) +
  theme(axis.line = element_line(colour = "gray90",
                                 size = 0.5, linetype = "solid")) +
  theme(axis.text.x = element_text(face="plain", color="black",
                                   size=8, angle=0),
        axis.text.y = element_text(face="plain", color="black",
                                   size=8, angle=0)) +
  theme(axis.title = element_text(face="plain", color="black",
                                  size=12, angle=0)) +
  theme(plot.title = element_text(face="plain", color="black",
                                  size=14, angle=0))
# promo_1_hist
```

```r
## Promotion 2
promo_2_hist = ggplot(promo_2, aes(x=SalesInThousands)) +
  geom_histogram(aes(y = ..density..), breaks = breaks, color='gray80', fill="gray80") +
  ggtitle('Promotion 2 Sales Density Histrogram') +
  xlab("Sales (in thousands)") +
  ylab("Density") +
  geom_density(color = "black", fill=colors_4[3],  alpha=0.5) +
  geom_vline(xintercept = mean(promo_2$SalesInThousands), linetype="dashed", color='darkblue', size = 0.8) +
  annotate("text",color='darkblue', size = 3, x=mean(promo_2$SalesInThousands)+1.25, y=0.01,
           label= paste("Mean Sales =", round(mean(promo_2$SalesInThousands), 2)), angle=90) +
  theme(panel.grid.major = element_line(color="gray90"), panel.grid.minor = element_line(color="gray94"),
        panel.background = element_blank(), axis.line = element_line(colour = "gray90", size =0.5)) +
  theme(panel.border = element_rect(fill=NA, colour = "gray90",
                                    size = 0.5, linetype = "solid")) +
  theme(axis.ticks = element_line(color="gray90", size=0.5)) +
  theme(axis.ticks.length = unit(0.18, "cm")) +
  theme(axis.line = element_line(colour = "gray90",
                                 size = 0.5, linetype = "solid")) +
  theme(axis.text.x = element_text(face="plain", color="black",
                                   size=8, angle=0),
        axis.text.y = element_text(face="plain", color="black",
                                   size=8, angle=0)) +
  theme(axis.title = element_text(face="plain", color="black",
                                  size=12, angle=0)) +
  theme(plot.title = element_text(face="plain", color="black",
                                  size=14, angle=0))

# promo_2_hist
```

```
## Promotion 3
promo_3_hist = ggplot(promo_3, aes(x=SalesInThousands)) +
  geom_histogram(aes(y = ..density..), breaks = breaks, color='gray80', fill="gray80") +
  ggtitle('Promotion 3 Sales Density Histrogram') +
  xlab("Sales (in thousands)") +
  ylab("Density") +
  geom_density(color = "black", fill=colors_4[4],  alpha=0.5) +
  geom_vline(xintercept = mean(promo_3$SalesInThousands), linetype="dashed", color='darkblue', size = 0.8) +
  annotate("text",color='darkblue', size = 3, x=mean(promo_3$SalesInThousands)+1.25, y=0.01,
          label= paste("Mean Sales =", round(mean(promo_3$SalesInThousands), 2)), angle=90) +
  theme(panel.grid.major = element_line(color="gray90"), panel.grid.minor = element_line(color="gray94"),
        panel.background = element_blank(), axis.line = element_line(colour = "gray90", size =0.5)) +
  theme(panel.border = element_rect(fill=NA, colour = "gray90",
                                    size = 0.5, linetype = "solid")) +
  theme(axis.ticks = element_line(color="gray90", size=0.5)) +
  theme(axis.ticks.length = unit(0.18, "cm")) +
  theme(axis.line = element_line(colour = "gray90",
                                    size = 0.5, linetype = "solid")) +
  theme(axis.text.x = element_text(face="plain", color="black",
                                    size=8, angle=0),
        axis.text.y = element_text(face="plain", color="black",
                                    size=8, angle=0)) +
  theme(axis.title = element_text(face="plain", color="black",
                                    size=12, angle=0)) +
  theme(plot.title = element_text(face="plain", color="black",
                                    size=14, angle=0))
# promo_3_hist
```

```
grid.arrange(promo_1_hist, promo_2_hist, promo_3_hist)
```
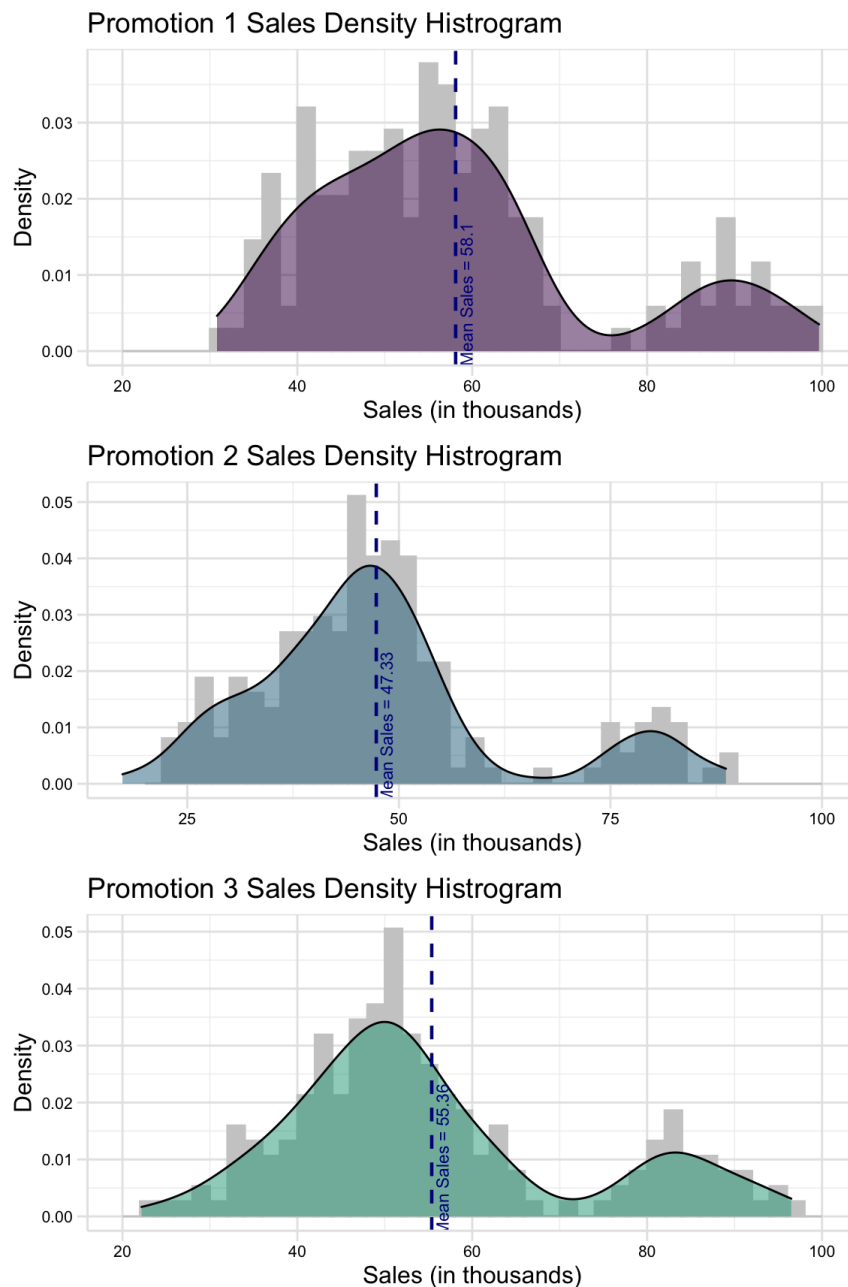
Figure 10.0: *Sales histograms for restaurants using Promotion 1 (purple), Promotion 2 (blue), and Promotion 3 (green).*

All three density histograms are in the same scale, making it easy for us to compare both normality and variance. When we explored the data in Part 1, the data for each sub population (Promotion 1, Promotion 2 and Promotion 3) had very similar distributions (see **Table 2.0**). Our density histograms in **Figure 10.0** confirm that similarity.

All three marketing campaigns distributions are extremely symmetric. Although none of the histograms are perfectly normal, which is expected as perfect normality is rare in practice, they are somewhat normal. All histograms have one major peak around 53 thousand dollars, and a second minor peak at around 80 thousand dollars. All histograms skew right which is expected for most monetary distributions. The deviation from normality is not severe enough to warrant concern for a couple reasons. Firstly, ANOVA testing is very robust especially when we are using a large sample size of 588 restaurants. Secondly, the deviations from normality are similar between the underlying populations.

All three marketing campaigns show very similar variance for *SalesInThousands* at around 16.75. Visually, all three distributions look very similar in width, indicating similar variance among the three marketing campaigns sales. According to the Central Limit Theorem, as our sample size increases, our variance decreases. Because each marketing campaign population has around 170 restaurants each, our symmetric variance can be partly attributed to the rather large, equally proportioned underlying population sizes. If we had one underlying population with a disproportionately small population size, say Promotion 1 had 30 restaurants instead of 172, we would likely see a variance that was not similar to the other underlying populations. Luckily this is not the case. The general **rule of thumb** used in practice to check for variance similarity is that **the largest sample standard deviation should be no more than twice as large as the smallest sample standard deviation.** Our data follows this rule.

In conclusion, both all three assumptions are met.

Limitations

From our exploration of the data, the only variable with association to a restaurant's *SalesInThousands* was the marketing campaigned used, *AgeOfStore* was shown to have no correlation in our ANCOVA test and when we did a simple correlation coefficient calculation. However, There may be **lurking variables** that have a strong influence on a restaurant's *SalesInThousands* that were not measured by the company in our study. Although *SalesInThousands* and Promotion are associated, it is highly likely that the marketing campaign is not the only variable that is responsible for differences in sales figures.

For example, maybe the lower sales observed in was due to that underlying population having more restaurants in poorer areas with people who cannot afford to eat out, or maybe affluent area that don't eat at the restaurant often. To continue, maybe the market demographics of the underlying populations are different. If Promotion 2 was a digital marketing campaign and the Promotion 2's market demographics skewed older than the Promotion 1 and Promotion 2 markets, that demographic difference may have caused the decrease in the effectiveness of the marketing campaign, not the campaign itself. There are many lurking variables, and near impossible to account for all of them.

If time and capital were inconsequential, one way in which the company could vet our conclusions would be to switch the marketing campaigns each restaurant cohort uses and regather sales data. For instance, all restaurants that used Promotion 1 will use Promotion 2, Promotion 2 will use Promotion 3, and Promotion 3 will use Promotion 1. If our conclusion that Promotion 2 marketing campaign was least effective at increasing sale holds, we would be more confidence in the effect of the marketing campaign on sales. Even though this may answer some of our questions about demographic differences discussed above, we can never account for every lurking variable, and we can never be fully certain about causation, only strength of association. We would need to design and conduct many more studies to suggest causation.