

Homework 3: Central Perk Coffee Shop - Selling Cappuccinos With Analytics

Adithya Ramesh, Jie Liu, Brady Engelke, Priya Priya, Sharang Jindal, Zixuan Wang

11/02/2019

Contents

0.1	Situation Overview	1
0.2	What are the client's monthly, weekly, and daily revenue distributions across products? . . .	2
0.3	Should the client focus on maximizing the revenue of existing customers or attract new customers? . . .	9
0.4	Are there distinct groups of existing customers with idiosyncratic behavior?	12
0.5	How can purchasing patterns of business, casual, and new customer segments be leveraged to enhance the customer experience and normalize demand?	15
0.6	Final Recommendations	22

0.1 Situation Overview

The Carlson Consulting Team has been tasked to establish whether the client's current customer base can generate enough revenue to sustain the profitability of the coffee shop by leveraging three years of untapped transactional data. The client's current strategy is to maximize revenue from the existing customer base rather than focusing efforts on getting new customers in the door. This strategy was inspired by management's intuition that strong customer loyalty has allowed the coffee shop to obtain a relatively stable demand year over year. The client is also concerned with extreme demand at certain points in the day/week and would be interested in approaches that would smooth demand into a more uniform frequency. To establish whether there is merit to primarily focus on maximizing the revenue of existing customers, the team visualized the client's daily, weekly, and monthly revenue distributions and assessed the associated contributions of existing and new customers.

It is not in the client's best interest to primarily focus on maximizing existing customer revenue since new customers account for roughly 40% of monthly sales. The team recommends that the client take a more holistic approach and focus efforts on delivering a superior coffee shop experience to new and existing customers. Furthermore, evidence was found that shed light on a more pressing issue for the client. Overall monthly sales have been stagnating at a substantial rate. The team hypothesized that certain aspects of the customer experience can be improved after seeing the market share the client had once obtained, affirmed by the low probability that demand for coffee in New York City is decreasing. The team also discovered that 94% of the client's revenue is generated by a fourth of their product offerings which may be a contributing factor to clientele dissatisfaction. Instead of focusing all efforts on the products that keep customers coming back, the coffee shop is putting forth a lot of effort to provide offerings that are not adding much value to the customer experience.

So, how can the client reverse the trend of falling revenue by refreshing the customer experience for new and existing customers? With this new key question in hand, the team sought to categorize customers into distinguishable groups and illustrate a clear picture of buying patterns unique to each group. By gaining a better understanding of the buying patterns for each type of customer, the client will be better positioned to customize discounts, pricing, and inventory that will delight customers and address the issue of diminishing sales.

0.2 What are the client's monthly, weekly, and daily revenue distributions across products?

Execution and Results: First, the team munged the data and prepared necessary datasets for exploratory analyses.

Load necessary packages and merge yearly datasets

```
library(tidyverse)
library(naniar)
library(lubridate)
library(hms)
library(arules)
library(gridExtra)
library(cowplot)
library(ggpubr)

df_16 <- read.csv('central_park_sales_2016.csv')
df_17 <- read.csv('central_park_sales_2017.csv')
df_18 <- read.csv('central_park_sales_2018.csv')
df <- rbind(df_16, df_17, df_18)

# create transaction_line id
df$tid <- seq.int(nrow(df))
# add unique shorter customer ID. for faster computation down the line
df$cid <- group_indices(df, Customer.ID)
# df$cid[df$cid == 31823] <- NA
df <- subset(df, select = -c(`Customer.ID`))

df$Gross.Sales = as.numeric(gsub("\\$", "", df$Gross.Sales))
df$Net.Sales = as.numeric(gsub("\\$", "", df$Net.Sales))
df$Tax = as.numeric(gsub("\\$", "", df$Tax))
df$Discounts = gsub("\\$", "", df$Discounts)
df$Discounts = gsub("\\)", "", df$Discounts)
df$Discounts = gsub("\\(", "-", df$Discounts)

glimpse(df)

# convert to necessary data types
df$Date <- as.character(df$Date)
df$Time <- as.character(df$Time)
df$Notes <- as.character(df$Notes)
df$Event.Type <- as.character(df$Event.Type)
df$Item <- as.character(df$Item)
df$Category <- as.character(df$Category)
df$Price.Point.Name <- as.character(df$Price.Point.Name)
df$Qty <- as.integer(df$Qty)
df$Gross.Sales <- as.character(df$Gross.Sales)
df$Discounts <- as.character(df$Discounts)
df$Net.Sales <- as.character(df$Net.Sales)
df$Tax <- as.character(df$Tax)
df$Gross.Sales <- as.double(df$Gross.Sales)
df$Discounts <- as.double(df$Discounts)
df$Net.Sales <- as.double(df$Net.Sales)
df$Tax <- as.double(df$Tax)
```

Data cleaning

```
# df %>% filter(Notes != '')
df$Notes[df$Notes == ''] <- NA # 90 out of 221,000 transactions have notes
```

```
# table(df$Event.Type) # mostly payments, 87 refunds
df <- df %>% filter(Event.Type != '') # drop 1 NA
# drop unknown error
df <- df[which(df$Date != 'Unknown Error'), ]
```

```
# table(df$Price.Point.Name)
# df %>% filter(Price.Point.Name == '')
df$Price.Point.Name[df$Price.Point.Name == ''] <- NA # large orders
```

```
# df %>% filter(Price.Point.Name == 'Plain / Choc')
df$Price.Point.Name[df$Price.Point.Name == 'Plain / Choc'] <- NA
# df %>% filter(Price.Point.Name == 'Regular Price')
df$Price.Point.Name[df$Price.Point.Name == 'Regular Price'] <- NA
```

Seperate refunds

```
# table(df$Qty)
refunds <- df %>% filter(Qty < 1)
payments <- df %>% filter(Qty > 0)
```

```
# table(payments$Item)
payments$Item[payments$Item == '\U0001f34bLemonade\U0001f34b'] <- 'Lemonade'
```

Seperate catering orders

```
# table(payments$Category)
# payments %>% filter(Category == 'None')
catering <- payments %>% filter(Category == 'None' & Item == 'Custom Amount')
payments <- payments %>% filter(Category != 'None' & Item != 'Custom Amount')
```

```
miss_var_summary(payments) # missingness is as expected
```

Convert date & time cols and create day of week col

```
payments$Date <- as.Date(payments$Date, format = '%m/%d/%y')
payments$Time <- as.hms(payments$Time)
payments <- payments %>%
  mutate(day_of_week = wday(Date, label = TRUE))
```

Clean up df

```
payments <- payments %>%
  select(Date, Time, day_of_week, Category:Price.Point.Name,
         Notes, cid, Gross.Sales:Tax) %>%
  rename(drink_size = Price.Point.Name,
         cust_id = cid,
         date = Date,
         time = Time,
         category = Category,
         item = Item,
         qty = Qty,
         notes = Notes)
```

Ensure drink_size col and item col integrity

```

# x <- payments %>% filter(is.na(drink_size) == TRUE)
payments$drink_size[payments$category == 'Beans'] <- NA
payments$drink_size[payments$category == 'Cereal'] <- NA
payments$drink_size[payments$category == 'Extras'] <- NA
payments$drink_size[payments$category == 'Food'] <- NA

payments$category <- as.factor(payments$category)
payments$drink_size <- as.factor(payments$drink_size)

payments$item[payments$item == 'Drip LG'] <- 'Drip'
payments$item[payments$item == 'Drip SM'] <- 'Drip'
payments$item[payments$item == 'Latte LG'] <- 'Latte'
payments$item[payments$item == 'Latte SM'] <- 'Latte'
payments$item[payments$item == 'Tea LG'] <- 'Tea'
payments$item[payments$item == 'Tea SM'] <- 'Tea'

```

Discretize time into time of day col

```

payments <- payments %>%
  mutate(time_of_day = ifelse(hour(time) >= 0 & hour(time) < 5, 'late_night',
                              ifelse(hour(time) >= 5 & hour(time) < 11, 'morning',
                              ifelse(hour(time) >= 11 & hour(time) < 14, 'lunch',
                              ifelse(hour(time) >= 14 & hour(time) < 18, 'afternoon',
                              ifelse(hour(time) >= 18 & hour(time) <= 24, 'night', NA))))))

payments$time_of_day <- factor(payments$time_of_day,
                              levels = c('morning', 'lunch', 'afternoon',
                              'night', 'late_night'), ordered = TRUE)

```

```

payments <- payments %>%
  mutate(cust_type = ifelse(cust_id != 31823, 'existing', 'new'))

payments$cust_type <- as.factor(payments$cust_type)

```

Visualize revenue distributions

```

sales <- payments %>%
  group_by(year(date), month(date)) %>%
  summarise(sales = sum(Gross.Sales))

sales$monthid <- seq.int(nrow(sales))

ggplot(sales) + geom_line(aes(x = monthid, y = sales / 1000)) +
  labs(title = 'Revenue Growth Has Stagnated Over The Past 10 Months',
       y = 'Monthly Sales (thousands $)', x = 'Months After Shop Opening',
       subtitle = 'Figure 1') + theme_minimal()

```

Revenue Growth Has Stagnated Over The Past 10 Months

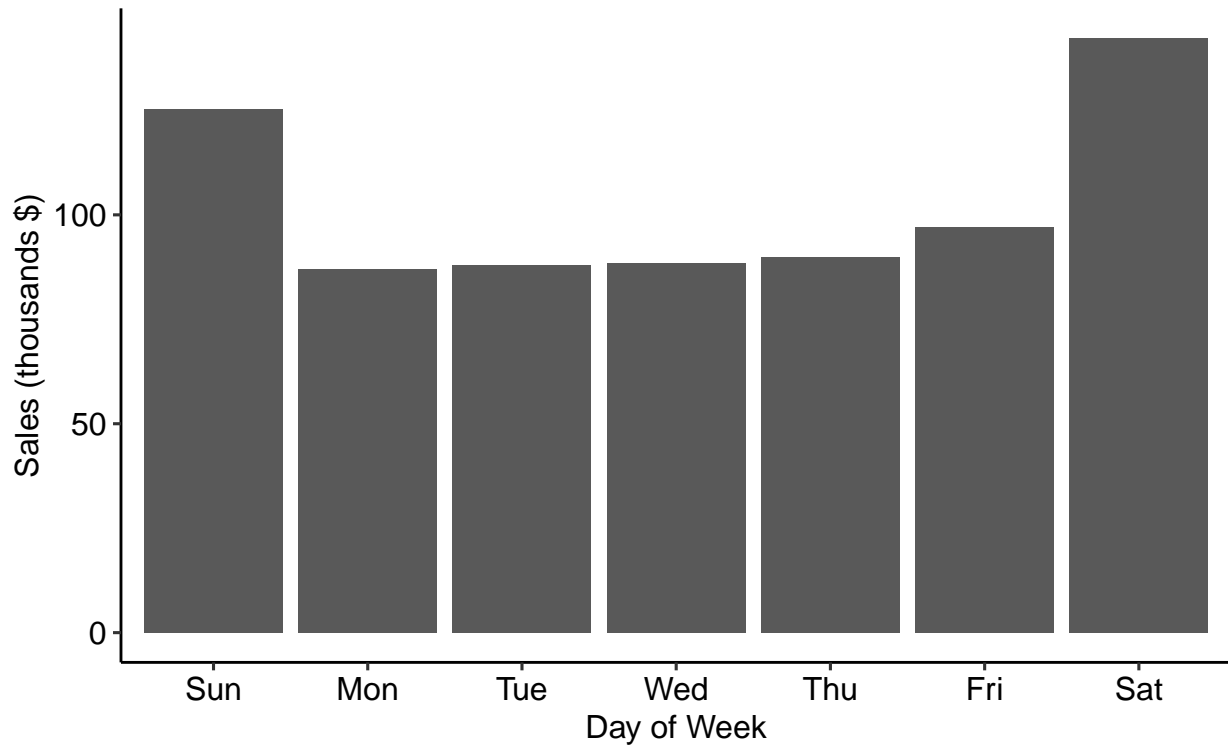
Figure 1



```
ggplot(payments) + geom_col(aes(x = day_of_week, y = Gross.Sales / 1000)) +  
  labs(title = 'Max Revenue is Generated on Sat & Sun', subtitle = 'Figure 2',  
        y = 'Sales (thousands $)', x = 'Day of Week') + theme_pubr()
```

Max Revenue is Generated on Sat & Sun

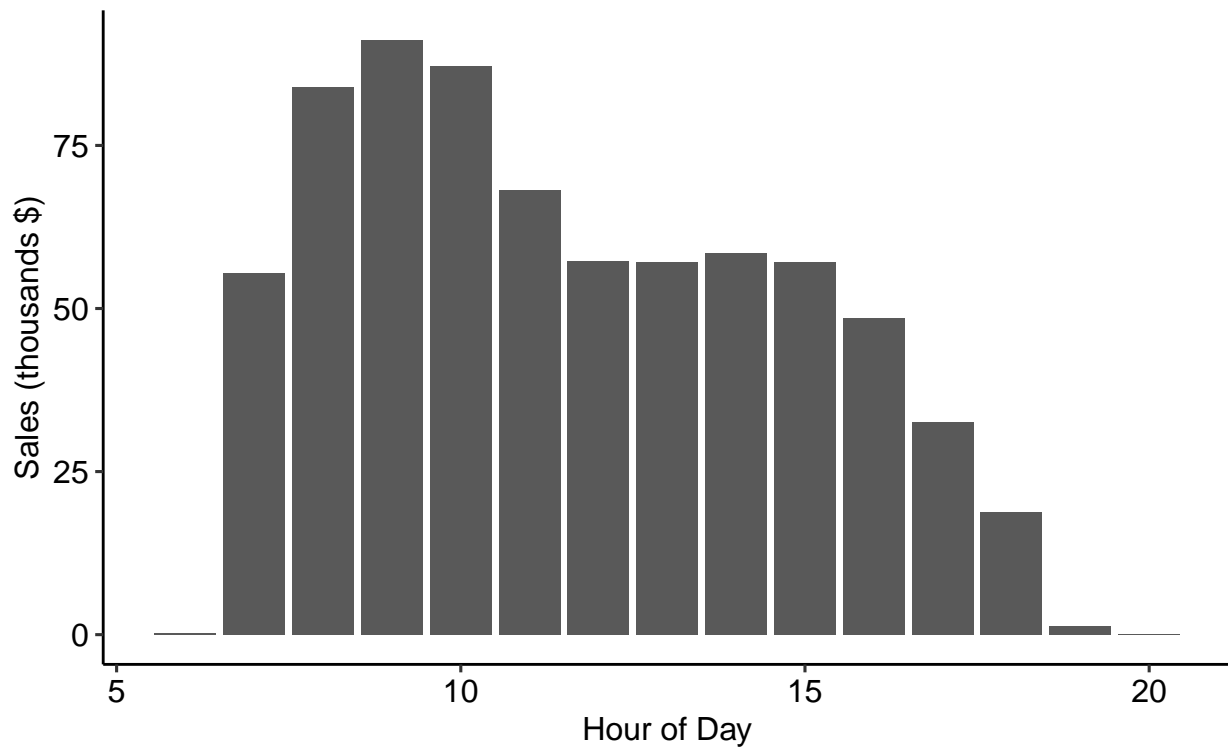
Figure 2



```
payments %>%
  group_by(hour = hour(time)) %>%
  summarize(sales = sum(Gross.Sales)) %>%
  ggplot() + geom_col(aes(x = hour, y = sales / 1000)) +
  labs(title = 'Demand Peaks at 9 am', subtitle = 'Figure 3',
        y = 'Sales (thousands $)', x = 'Hour of Day') + theme_pubr()
```

Demand Peaks at 9 am

Figure 3

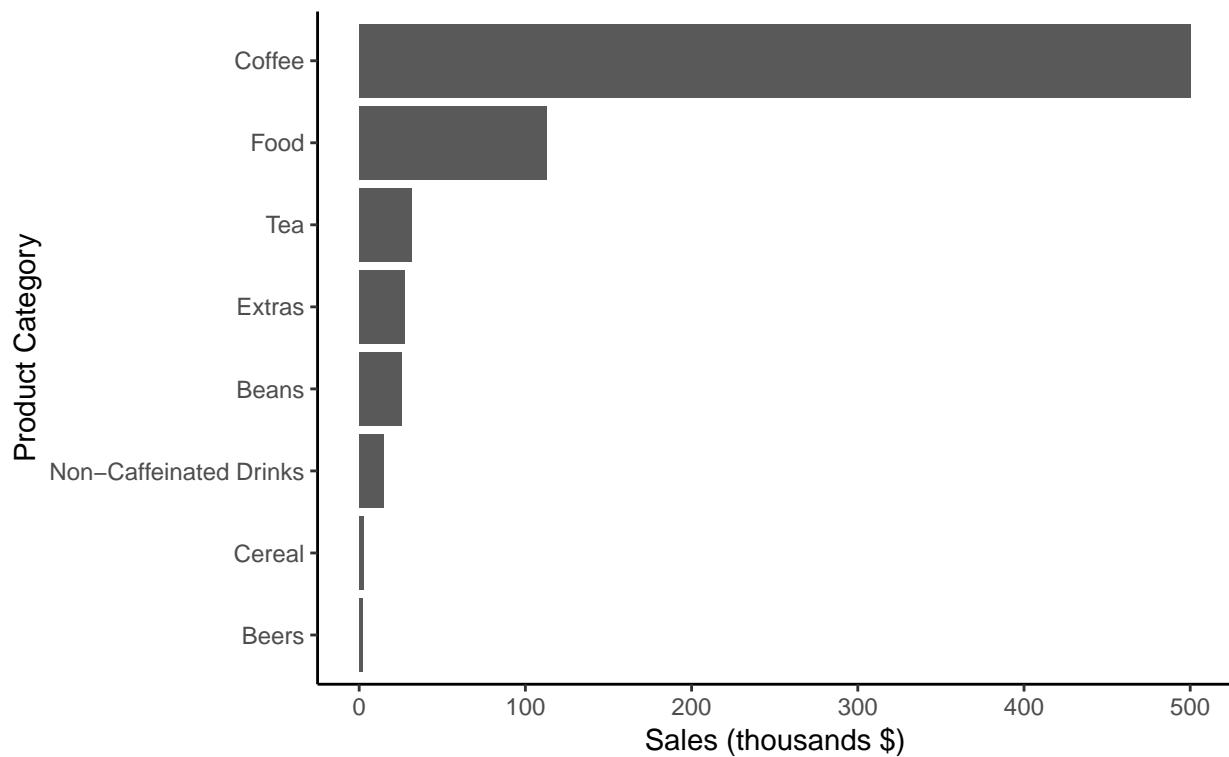


```
y <- payments %>%
  group_by(category) %>%
  summarize(sales = sum(Gross.Sales))

y$category <- factor(y$category, levels = y$category[order(y$sales)])
ggplot(y) + geom_col(aes(x = category, y = sales / 1000)) +
  labs(title = 'Coffee Generates Majority of Revenue',
        subtitle = 'Figure 4', y = 'Sales (thousands $)', x = 'Product Category') +
  coord_flip() + theme_classic()
```

Coffee Generates Majority of Revenue

Figure 4



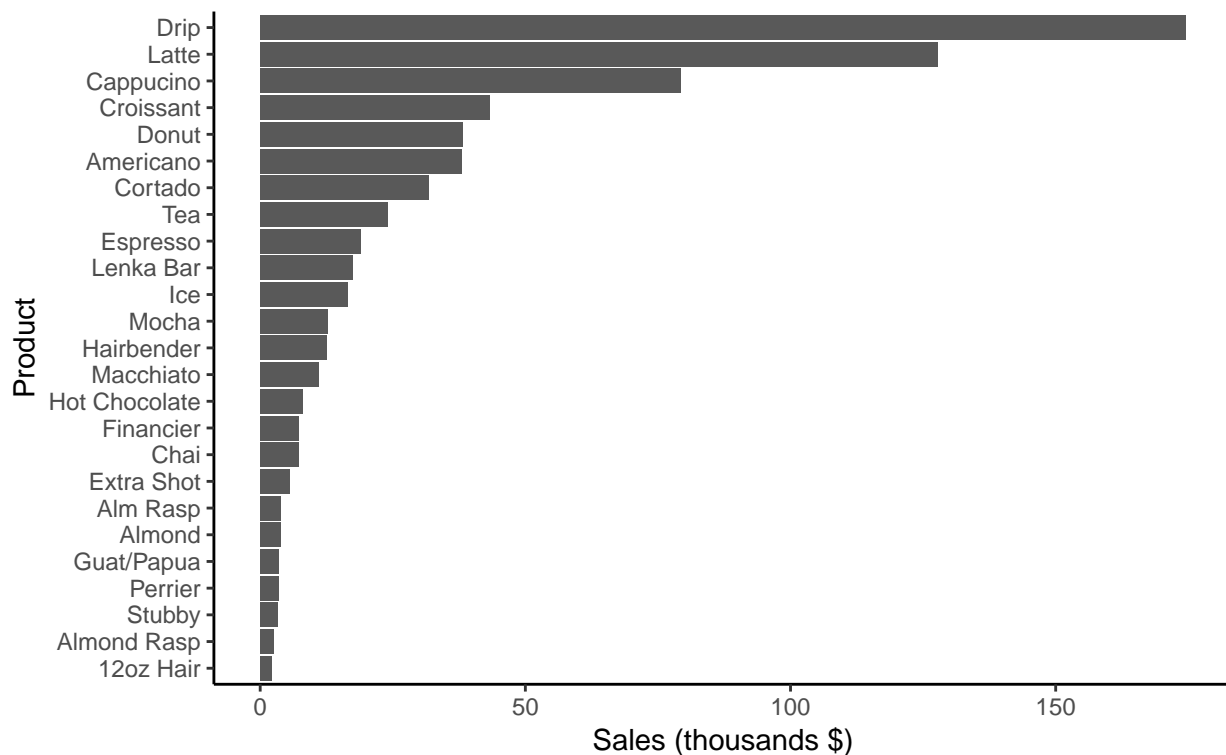
```
revenue_generating_products <- payments %>%
  group_by(item) %>%
  summarize(sales = sum(Gross.Sales)) %>%
  filter(sales > 5000)

z2 <- payments %>%
  group_by(item) %>%
  summarize(sales = sum(Gross.Sales)) %>%
  filter(sales > 2000)

z2$item <- factor(z2$item, levels = z2$item[order(z2$sales)])
ggplot(z2) + geom_col(aes(x = item, y = sales / 1000)) + coord_flip() +
  labs(title = 'Most Products Do Not Generate Substantial Revenue',
        subtitle = 'Figure 5', y = 'Sales (thousands $)', x = 'Product') + theme_classic()
```


Most Products Do Not Generate Substantial Revenue

Figure 5



```
rgp_sum <- sum(revenue_generating_products$sales)
payments_sum <- sum(payments$Gross.Sales)
z_perc <- rgp_sum / payments_sum
```

Interpretation & Conclusions: Figure 1 illustrated a concerning trend of diminishing sales over the past 10 months. This is something the coffee shop must seek to address. Figures 2 & 3 confirmed that there are indeed spikes and lulls in demand throughout the day and week, confirming this aspect of the client's intuition. Demand increases on the weekends and is relatively steady throughout the week with a slight increase on Friday. During a typical day, demand peaks at 9 am, linearly drops until noon, plateaus until 3 pm, and then drops again until 6 pm.

Figure 4 clarified that coffee is by far the most popular item with food, tea, extras, and beans being important revenue contributors as well. Revenue from beans seems to be disappointingly low for a popular coffee shop. Figure 5 and the code snip directly above demonstrated that there are 18 products out of 66 that account for 94% of revenue. This raised concern amongst the team that the client is wasting their efforts on the breadth of product offerings when it'd be prudent to acutely focus on the revenue-generating products. This would help the client avoid stocking unnecessary inventory and deliver higher quality on best selling products.

0.3 Should the client focus on maximizing the revenue of existing customers or attract new customers?

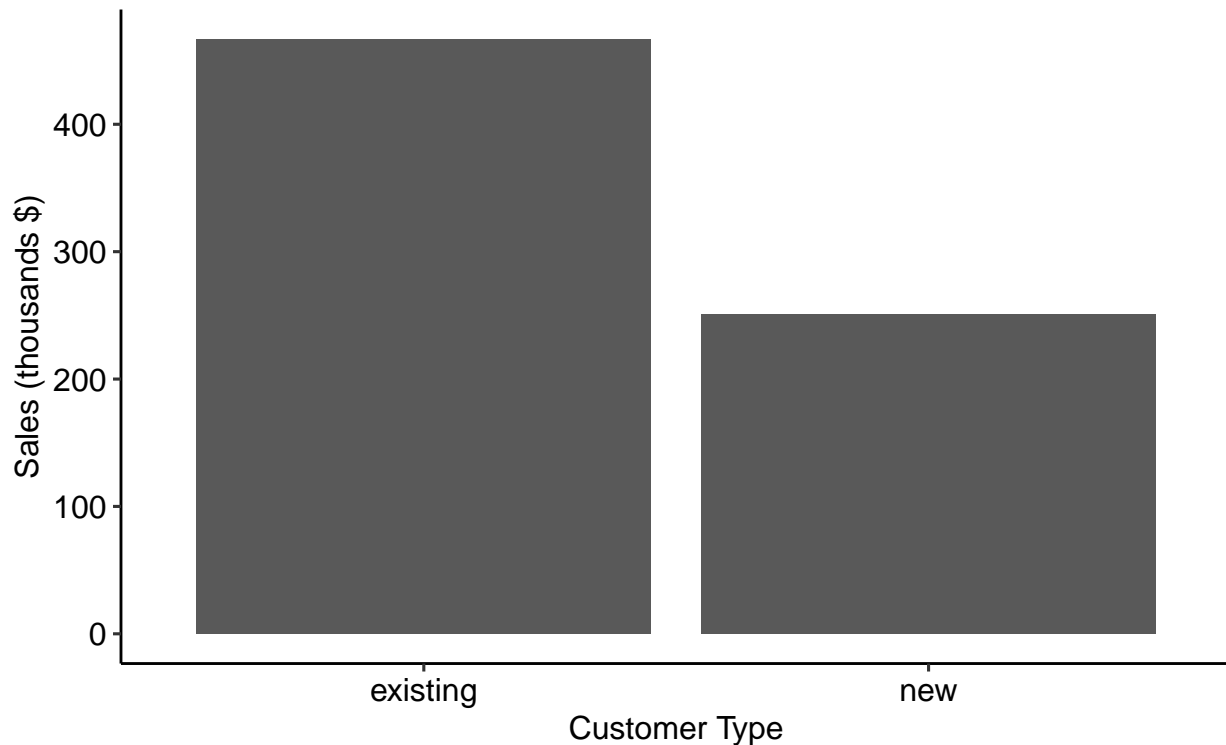
Assumption: An existing customer is one with an id.

Execution and Results:

```
ggplot(payments) + geom_col(aes(x = cust_type, y = Gross.Sales / 1000)) +
  labs(title = 'New Customers Provide a Large Portion of Revenue',
        subtitle = 'Figure 6', y = 'Sales (thousands $)', x = 'Customer Type') +
  theme_pubr()
```

New Customers Provide a Large Portion of Revenue

Figure 6



```
e_sales <- payments %>%
  filter(cust_type == 'existing') %>%
  group_by(year(date), month(date)) %>%
  summarise(sales = sum(Gross.Sales))

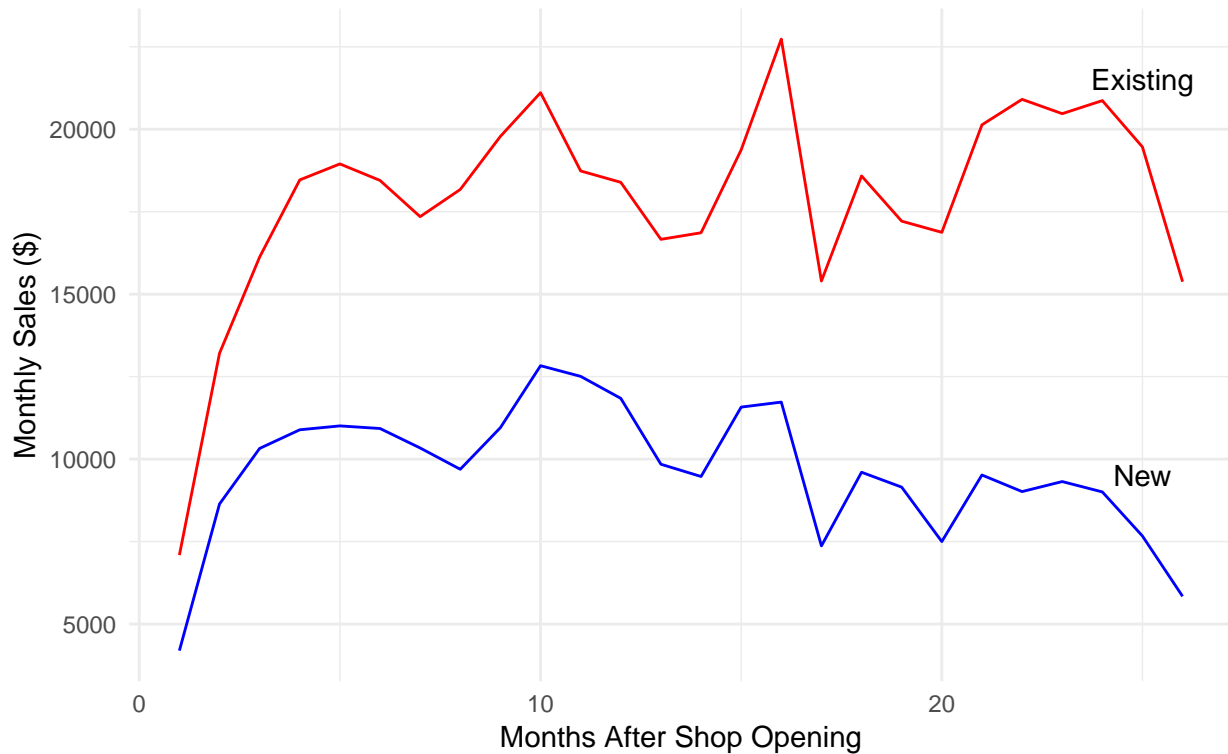
n_sales <- payments %>%
  filter(cust_type == 'new') %>%
  group_by(year(date), month(date)) %>%
  summarise(sales = sum(Gross.Sales))

e_sales$monthid <- seq.int(nrow(e_sales))
n_sales$monthid <- seq.int(nrow(n_sales))

ggplot() +
  geom_line(data = e_sales, aes(x = monthid, y = sales), color = 'red') +
  geom_line(data = n_sales, aes(x = monthid, y = sales), color = 'blue') +
  labs(title = 'Revenue Has Been Decreasing For Both Customer Types',
       subtitle = 'Figure 7', y = 'Monthly Sales ($)',
       x = 'Months After Shop Opening') +
  geom_text(aes(x = 25, y = 21500, label = "Existing")) +
  geom_text(aes(x = 25, y = 9500, label = "New")) +
  theme(legend.title = element_blank()) + theme_minimal()
```

Revenue Has Been Decreasing For Both Customer Types

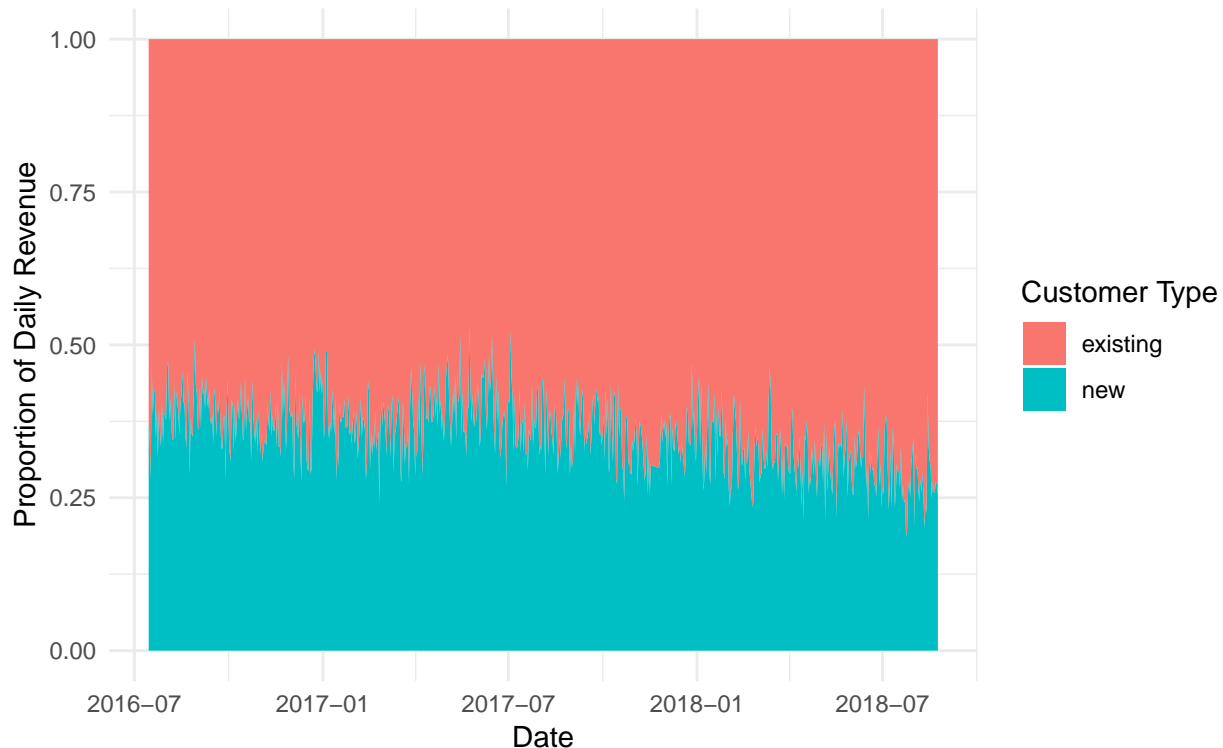
Figure 7



```
daily_sales <- payments %>%  
  group_by(date, cust_type) %>%  
  summarise(sales = sum(Gross.Sales))  
  
ggplot(daily_sales) + geom_area(aes(x = date, y = sales, fill = cust_type),  
                                position = 'fill') +  
  labs(title = 'New Customers Generate 40% of Daily Revenue',  
        subtitle = 'Figure 8', y = 'Proportion of Daily Revenue',  
        x = 'Date', fill = 'Customer Type') + theme_minimal()
```

New Customers Generate 40% of Daily Revenue

Figure 8



Interpretation & Conclusions: No question existing customers do generate most of the client's revenue. However, new customers do account for roughly 40% of sales per month. Thus, neither customer segment should be deprioritized.

Revenue has been in stagnation from existing and new customers for months. Although, a reduction in new customer revenue is expected since new customers naturally convert to existing customers over time. What is concerning is that there was no influx of existing customer revenue as new customer revenue decreased. It is unclear as to whether this is a customer retention or new customer appeal issue. The team hypothesized that customer needs, in general, are not being fully met by the coffee shop and that there are certain aspects of the customer experience that can be customized which will rehabilitate the sales trend for both customer segments.

0.4 Are there distinct groups of existing customers with idiosyncratic behavior?

How many unique customers are there?

```
payments %>%
  filter(cust_id != 31823) %>%
  select(cust_id) %>%
  unique()
```

Interpretation & Conclusions: There are 143,403 out of 221,355 transactions with a customer id with 31,811 unique customers.

```
df_i_cols <- payments %>% filter(cust_id != 31823) %>%
  select(cust_id, qty, Gross.Sales, Discounts, day_of_week, drink_size) %>%
  mutate(weekday = ifelse(day_of_week == 'Sun' | day_of_week == 'Sat', 0, 1),
         weekend = ifelse(day_of_week == 'Sun' | day_of_week == 'Sat', 1, 0),
         drink_size_num = ifelse(drink_size == 'SM', 1,
```

```
ifelse(drink_size == 'Regular', 2, 3)))
```

Feature engineer a purchasing pattern ratio: weekday trips / total trips

```
customers <- df_i_cols %>%  
  filter(is.na(cust_id) == FALSE) %>%  
  group_by(cust_id) %>%  
  summarize(total_trips = n(),  
            quantity = sum(qty),  
            sales = sum(Gross.Sales),  
            discounts = sum(Discounts),  
            wd_sum = sum(weekday),  
            we_sum = sum(weekend)) %>%  
  mutate(week_ratio = wd_sum / (wd_sum + we_sum))
```

Create normalization function

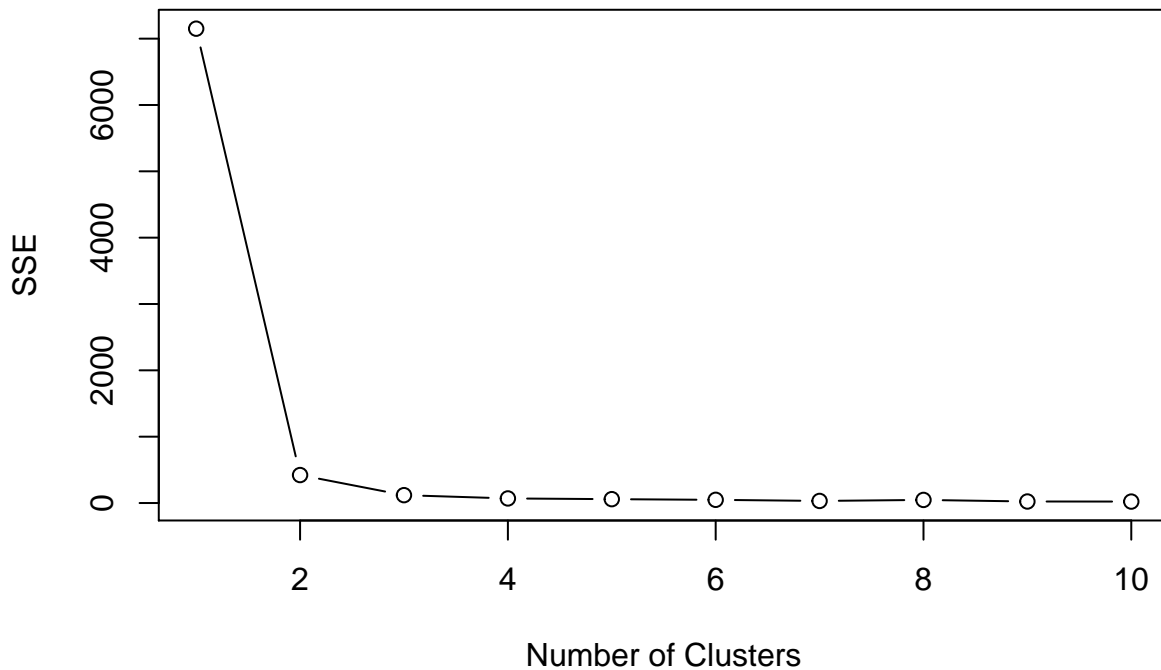
```
normalize <- function(x){  
  return ((x - min(x))/(max(x) - min(x)))  
}
```

Normalize features

```
customers2 <- customers %>%  
  select(total_trips: week_ratio) %>%  
  transmute(tot_trips = normalize(total_trips),  
            quantity = normalize(quantity),  
            sales = normalize(sales),  
            discounts = normalize(discounts),  
            week_ratio = normalize(week_ratio))
```

Identify # of clusters

```
SSE_curve <- c()  
for (n in 1:10) {  
  kcluster <- kmeans(customers2, n)  
  sse <- sum(kcluster$withinss)  
  SSE_curve[n] <- sse  
}  
plot(1:10, SSE_curve, type="b", xlab="Number of Clusters", ylab="SSE")
```



Interpretation & Conclusions: Clearly, there are two distinct groups of customers with unique behavior. Below is a summary table of the discrepancies between the average behavior of group 1 and group 2.

```
k = 2
kcluster <- kmeans(customers2, k, nstart = 20)

clustered_customers <- customers %>%
  mutate(Cluster = kcluster$cluster)

clustered_means <- clustered_customers %>%
  select(Cluster, total_trips:discounts, week_ratio) %>%
  group_by(Cluster) %>%
  summarise_all("mean") %>%
  mutate(cluster_size = kcluster$size)
knitr::kable(clustered_means)
```

Cluster	total_trips	quantity	sales	discounts	week_ratio	cluster_size
1	3.448049	3.795382	11.41151	-0.0069013	0.0465318	14119
2	5.353832	5.756613	17.27545	-0.0267409	0.9722199	17692

Interpretation & Conclusions: One group consists of “business” professionals who are grabbing coffee during the week whereas the other group is the “casual”, networking coffee drinkers who only come on the weekends. The business professionals are more valuable on average, utilize more discounts, and are larger in group size than the casual, weekend drinkers.

Merge cluster assignments to overall payments dataset. Assigning “New” customers to cluster 3

```
customer_trans <- payments %>%
  merge(y = clustered_customers, by = 'cust_id', all.x = TRUE) %>%
  select(cust_id, cust_type, Cluster, date:day_of_week, time_of_day,
         category:Tax, total_trips:discounts, week_ratio) %>%
  rename(sales_sum = sales, discounts_sum = discounts)
```

```
customer_trans$Cluster[is.na(customer_trans$Cluster) == TRUE] <- 3

# be sure to assign this mutation correctly as cluster assignment varies
customer_trans <- customer_trans %>%
  mutate(cluster = ifelse(Cluster == 1, 'Casual',
                          ifelse(Cluster == 2, 'Business', 'New')))
customer_trans$cluster <- as.factor(customer_trans$cluster)

Create dataframes for drinks and non-drinks
drinks <- customer_trans %>%
  filter(is.na(drink_size) == FALSE)

primary_drink_options <- c('Drip', 'Americano', 'Cappuccino',
                          'Espresso', 'Latte', 'Tea', 'Cortado')

non_drinks <- customer_trans %>%
  filter(is.na(drink_size) != FALSE)

primary_food_options <- c('Croissant', 'Donut', 'Lenka Bar')

glimpse(customer_trans)
```

0.5 How can purchasing patterns of business, casual, and new customer segments be leveraged to enhance the customer experience and normalize demand?

Execution and Results:

```
clust1_sales <- customer_trans %>%
  filter(cluster == 'Business') %>%
  group_by(year = year(date), month = month(date)) %>%
  summarise(sales = sum(Gross.Sales))

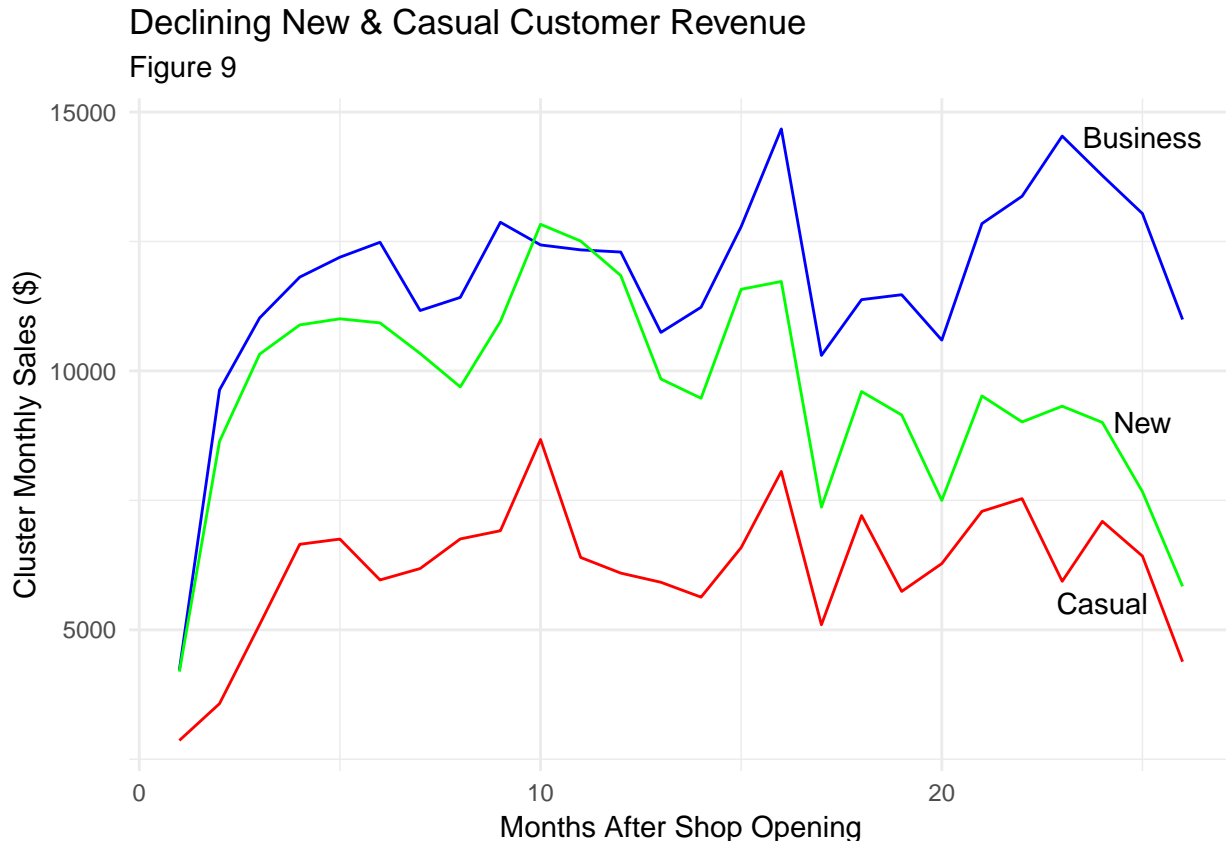
clust2_sales <- customer_trans %>%
  filter(cluster == 'Casual') %>%
  group_by(year = year(date), month = month(date)) %>%
  summarise(sales = sum(Gross.Sales))

clust3_sales <- customer_trans %>%
  filter(cluster == 'New') %>%
  group_by(year = year(date), month = month(date)) %>%
  summarise(sales = sum(Gross.Sales))

clust1_sales$monthid <- seq.int(nrow(clust1_sales))
clust2_sales$monthid <- seq.int(nrow(clust2_sales))
clust3_sales$monthid <- seq.int(nrow(clust3_sales))

ggplot() + geom_line(data = clust1_sales, aes(x = monthid, y = sales),
                     color = 'blue') + # Business
  geom_line(data = clust2_sales, aes(x = monthid, y = sales),
            color = 'red') + # Casual
  geom_line(data = clust3_sales, aes(x = monthid, y = sales),
            color = 'green') + # New
```

```
labs(title = 'Declining New & Casual Customer Revenue',
      subtitle = 'Figure 9', y = 'Cluster Monthly Sales ($)',
      x = 'Months After Shop Opening') +
geom_text(aes(x = 25, y = 14500, label = "Business")) +
geom_text(aes(x = 25, y = 9000, label = "New")) +
geom_text(aes(x = 24, y = 5500, label = "Casual")) +
theme(legend.title = element_blank()) + theme_minimal()
```

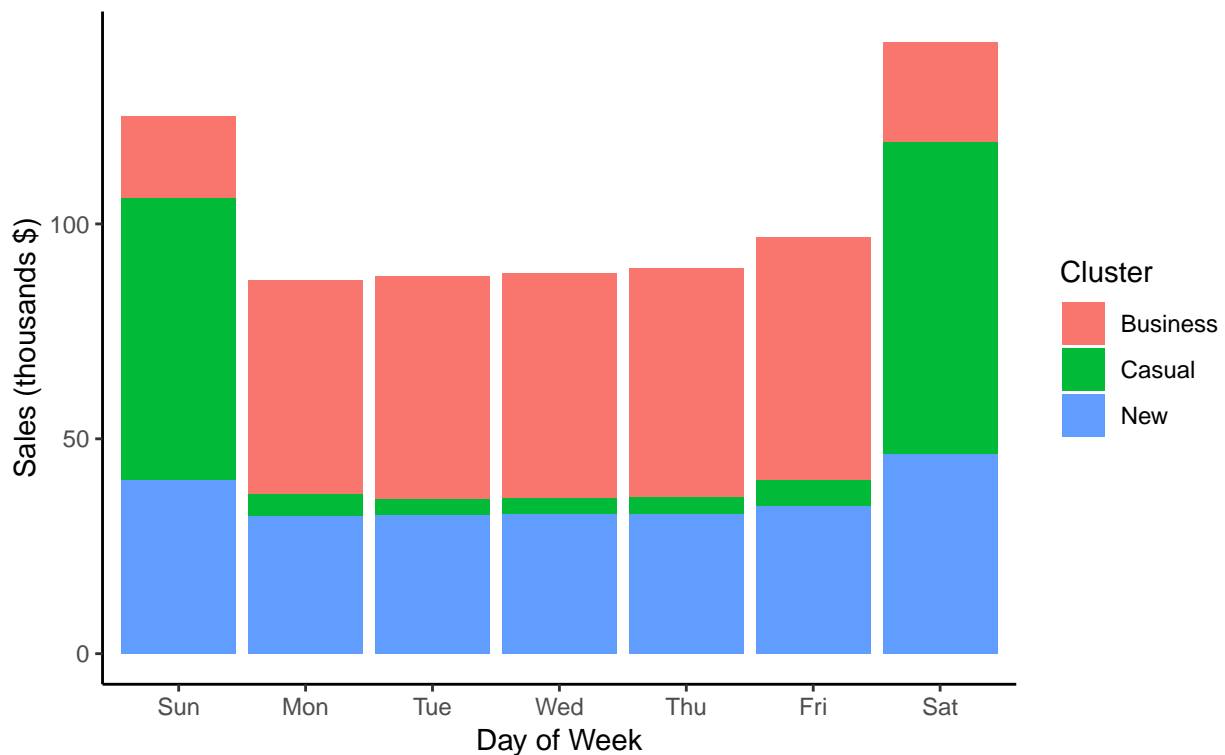


Interpretation & Conclusions: The “business” customers are bringing in nearly double the revenue as the casual coffee drinkers. The coffee shop seems to be meeting the needs of the “business” customers better than the “casual” or “new” customers. “New” customers have stopped coming at a linear rate for over 15 months. Similarly, but less extreme, is the negative linear trend for “casual” customer revenue. These negative trends are the problem of interest the final recommendations aim to address.

```
customer_trans %>%
  group_by(day_of_week, cluster) %>%
  summarize(sales = sum(Gross.Sales)) %>%
  ggplot() + geom_col(aes(x = day_of_week, y = sales / 1000, fill = as.factor(cluster))) +
  labs(title = 'Purchasing Pattern of Each Cluster',
        subtitle = 'Figure 10', y = 'Sales (thousands $)',
        x = 'Day of Week', fill = 'Cluster') + theme_classic()
```


Purchasing Pattern of Each Cluster

Figure 10

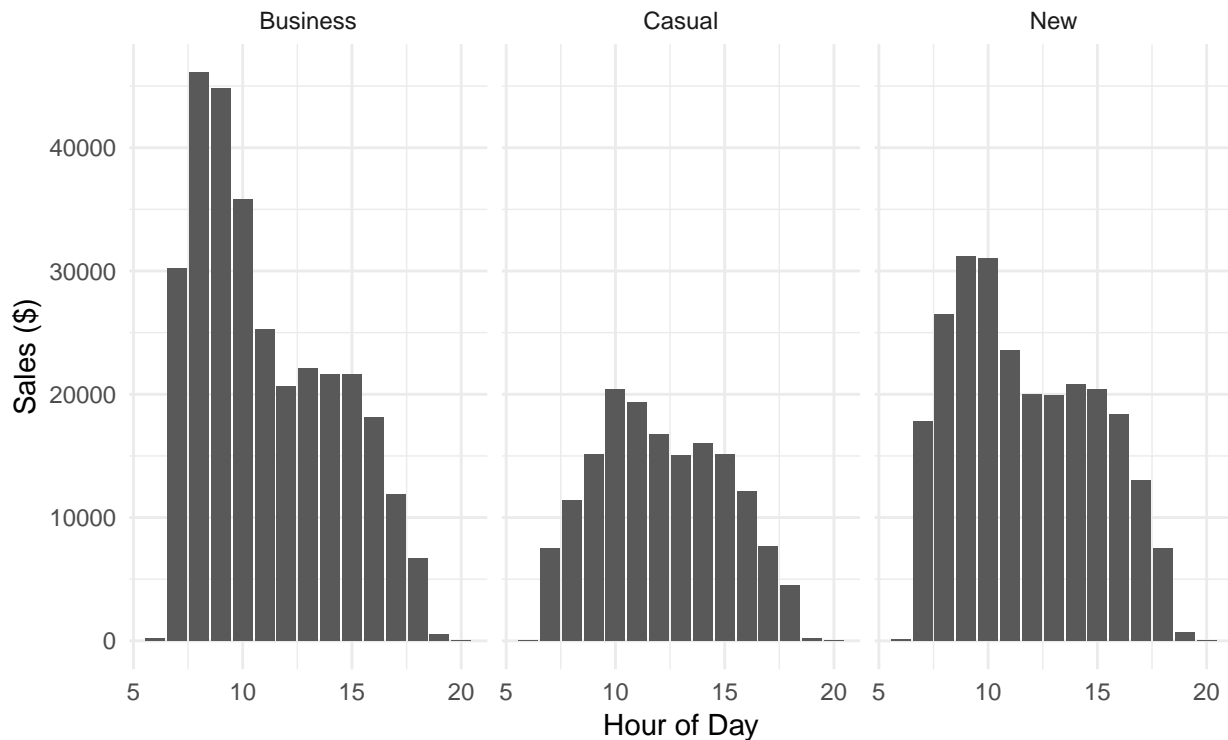


Interpretation & Conclusions: An effective promotional strategy to pull some of the demand from Sat & Sun into the middle of the week, particularly from “casual” customers, would smooth the weekly demand distribution as well as provide an incentive for “casual” customers to come in more frequently. This is an inference that could be leveraged to customize the experience for “casual” customers which would address the mediocre demand obtained from this cluster over the past 15 months.

```
ggplot(customer_trans) + geom_col(aes(x = hour(time), y = Gross.Sales)) + facet_grid(. ~ cluster) +  
  labs(x = 'Hour of Day', y = 'Sales ($)', subtitle = 'Figure 11',  
    title = 'Daily Revenue Distribution by Cluster') + theme_minimal()
```

Daily Revenue Distribution by Cluster

Figure 11



Interpretation & Conclusions: Understanding that “business” customers typically come during the week from figure 10, the best time to offer a promotional program to attract “casual” customers to come on weekdays would be from noon to 6 pm since most “business” customers enjoy their coffee in the morning. “Casual” customers have a relatively uniform daily demand. Thus, it is not unreasonable to assert that the client could incentivize the “casual” customers to adjust the day at which they come since manipulation of their preferred time of day to come is not required.

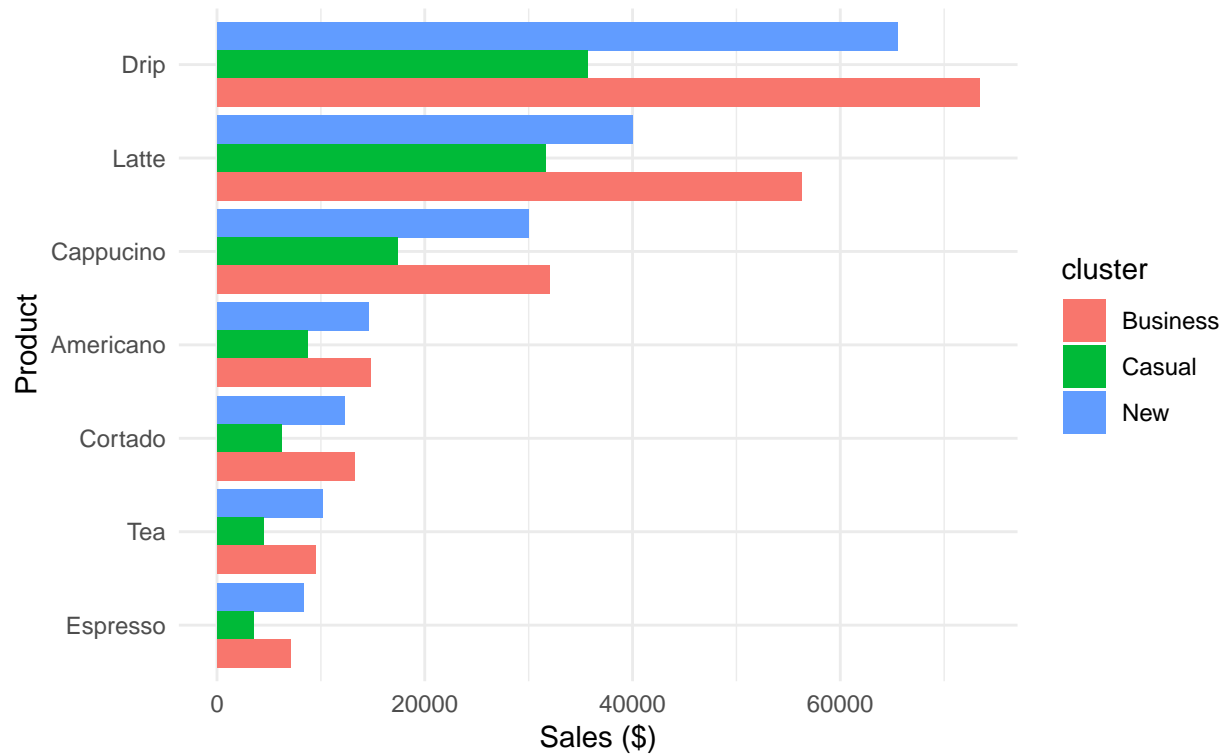
```
d <- drinks %>%
  filter(item %in% primary_drink_options) %>%
  group_by(item, cluster) %>%
  summarise(sales = sum(Gross.Sales))

d$item <- factor(d$item, levels = c('Espresso', 'Tea', 'Cortado', 'Americano',
                                   'Cappuccino', 'Latte', 'Drip'), ordered = TRUE)

ggplot(d) + geom_col(aes(x = item, y = sales, fill = cluster), position = 'dodge') +
  coord_flip() + labs(x = 'Product', y = 'Sales ($)', subtitle = 'Figure 12',
                     title = 'Coffee Revenue Distribution by Cluster') + theme_minimal()
```

Coffee Revenue Distribution by Cluster

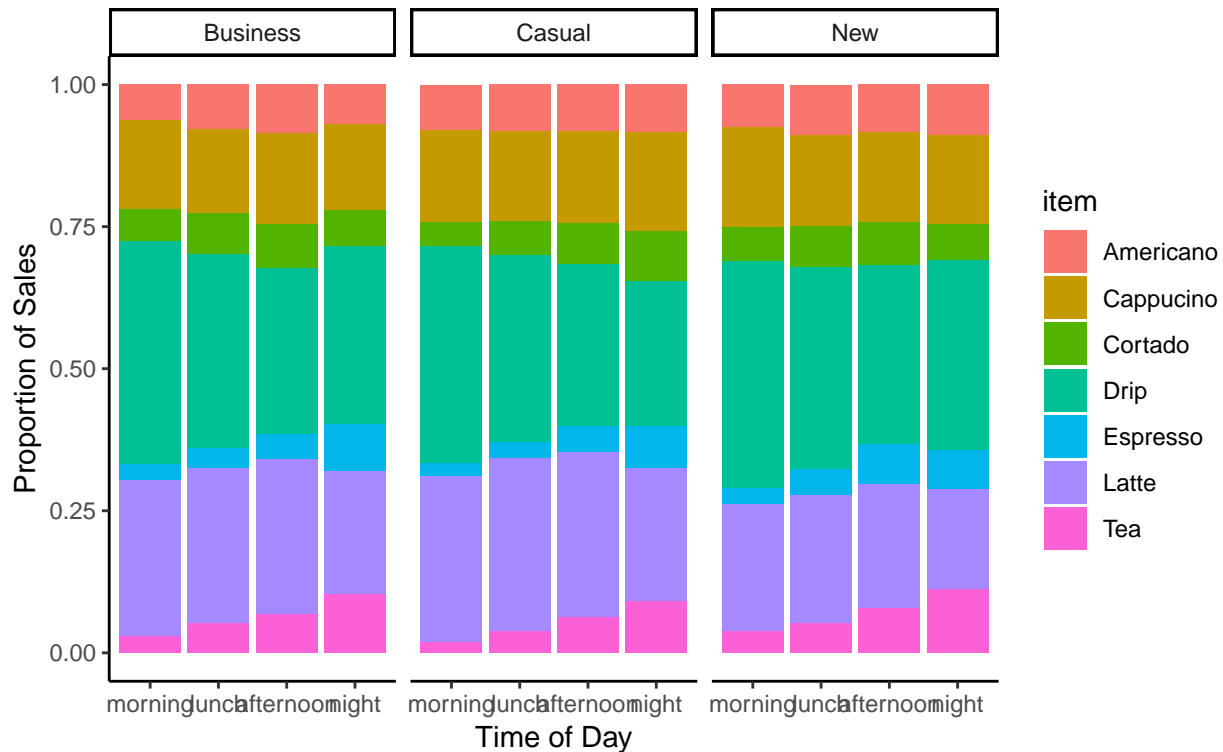
Figure 12



```
customer_trans %>%
  group_by(time_of_day, cluster, item) %>%
  summarise(sales = sum(Gross.Sales)) %>%
  filter(item %in% primary_drink_options) %>%
  ggplot() + geom_col(aes(x = time_of_day, y = sales, fill = item), position = 'fill') +
  facet_grid(. ~ cluster) +
  labs(x = 'Time of Day', y = 'Proportion of Sales', subtitle = 'Figure 13',
       title = 'Coffee Sales Throughout Day by Cluster') + theme_classic()
```

Coffee Sales Throughout Day by Cluster

Figure 13



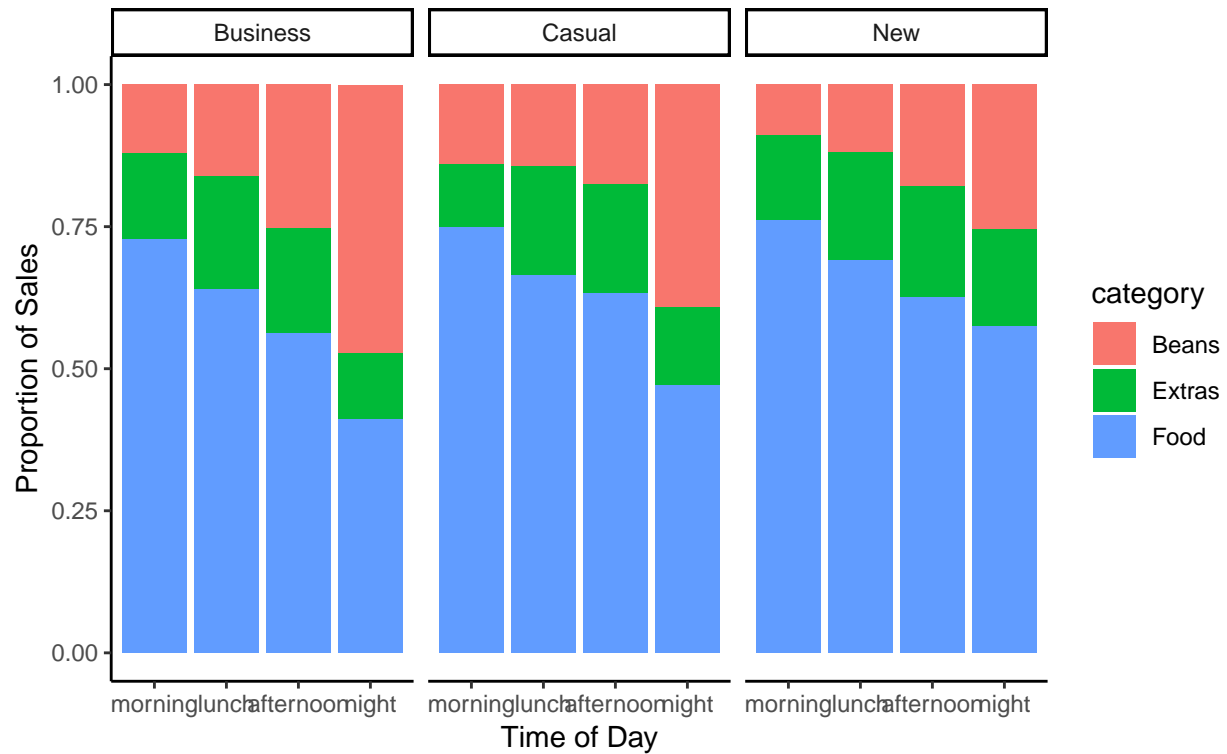
Interpretation & Conclusions: To attract “casual” customers into the coffee shop during the week, the client is going to have to develop a creative marketing program. The team proposes that the client host a happy hour from 2 - 5 pm, Monday - Thursday. The happy hour can be promoted daily to all customers but a majority of marketing efforts should be targeted at the “casual” customer cluster.

Upon assessing figures 12 & 13, the team advises the happy hour menu to revolve around the following list of coffee drinks: latte, cappuccino, americano, tea, espresso, and cortado. Lattes, cappuccinos, and americanos can be considered as the staple drinks of the happy hour since they are very popular amongst all customers. Drip is very popular as well but figure 13 illustrates a diminishing customer interest in drip coffee as the day progresses. Tea, espresso, and cortados should be included as specialty drinks for the happy hour since they are relatively popular amongst “casual” and “new” customers when compared to the popularity amongst “business” customers. For example, “business” customers purchase by far the most drip coffee and lattes but the same cannot be said about cortados, tea, or espressos. Figure 13 shows that these specialty drinks become more popular as the day progresses as well.

```
customer_trans %>%
  group_by(time_of_day, cluster, category) %>%
  summarize(sales = sum(Gross.Sales)) %>%
  filter(category %in% c('Beans', 'Food', 'Extras')) %>%
  ggplot() + geom_col(aes(x = time_of_day, y = sales, fill = category),
    position = 'fill') +
  facet_grid(. ~ cluster) +
  labs(x = 'Time of Day', y = 'Proportion of Sales', subtitle = 'Figure 14',
    title = 'Non-coffee Sales Throughout Day by Cluster') + theme_classic()
```

Non-coffee Sales Throughout Day by Cluster

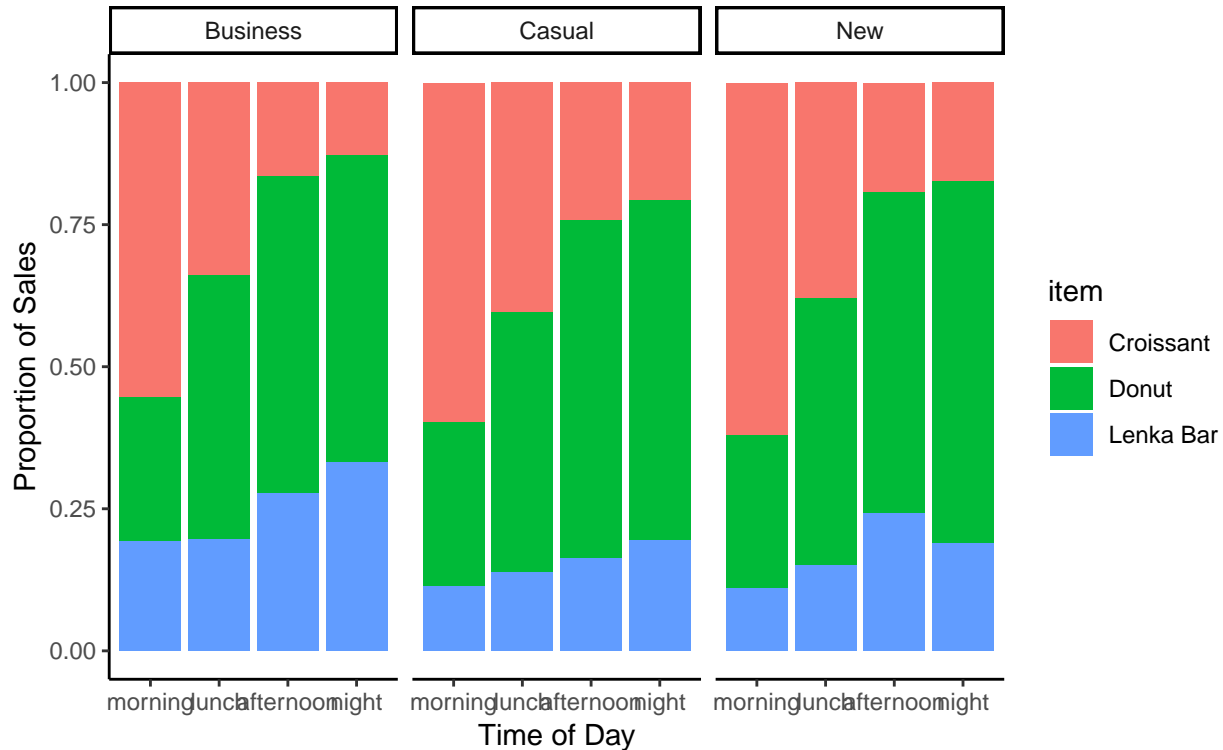
Figure 14



```
non_drinks %>%
  group_by(time_of_day, cluster, item) %>%
  summarise(sales = sum(Gross.Sales)) %>%
  filter(item %in% primary_food_options) %>%
  ggplot() + geom_col(aes(x = time_of_day, y = sales, fill = item),
    position = 'fill') +
  facet_grid(.~ cluster) +
  labs(x = 'Time of Day', y = 'Proportion of Sales',
    subtitle = 'Figure 15', title = 'Food Sales Throughout Day by Cluster') +
  theme_classic()
```

Food Sales Throughout Day by Cluster

Figure 15



Interpretation & Conclusions: In addition to the coffee products offered during happy hour, the client should incorporate donuts as well since customer preferences shift heavily to donuts as the day progresses. Donuts should be considered as a staple item.

The team still believes bean sales could be improved. There seems to be a pattern of customers purchasing beans as it gets later in the day. A potential strategy to address the low bean sales that would also persuade customers to come to the new happy hour would be to offer a buy-one-get-one-free deal on beans if a customer comes to three weekday happy hours. This promotion should be ongoing, incentivizing customers to continue to take advantage of the deal, in turn increasing the trip frequency substantially of these customers. Additionally, this promotional program would appeal to more new customers and incentivize them to create a customer id with the shop so they can be considered for the buy-one-get-one-free deal on beans.

0.6 Final Recommendations

Hopefully, it is clear as to why the team adjusted the primary focus of this analysis from maximizing the revenue of the client's existing customer base to tailoring the customer experience to recapture the falling contribution in sales from "new" and "casual" customers. Figures 1, 6, 7, & 8 served as the evidence that led the team down this path of analysis. Figures 9 and 10 served as inspiration for recommendation 1 & figure 5 motivated the second supplemental recommendation.

1)

The client should launch a new happy hour program from 2 - 5 pm, Monday - Thursday. This happy hour should be seen as a promotional product, specifically targeted to the "casual" and "new" customer segments. Lattes, cappuccinos, and americanos will serve as the staple drinks of the happy hour with tea, espresso, and cortados serving as the specialty drinks. Complimenting the drink menu, donuts should be offered also. This menu was selected based on the progression of customer preferences throughout the day (figures 13 & 15) and the relative popularity of the drinks amongst "casual" and "new" customers compared to "business" customers (figure 12). The staple items offered on the happy hour menu should be discounted at 20% and

specialty items at 10%, forfeiting the gross margin of the staple products during happy hour yet maintaining half of the gross margin from the specialty item sales. The staple items will likely be the main draw for “new” and “existing” customers to come to the happy hour thus these are the products that require the substantive discount. A second incentive that will ensure the happy hour attracts new customers and boosts the disappointing sales of beans is to offer a buy-one-get-one-free deal on beans to customers that come to the happy hour three times. This incentive should be offered on an ongoing basis with the happy hour.

Key Benefits of This Recommendation: 1) The extreme weekend demand of “casual” customers will be spread across the week because of the happy hour and deal on beans. 2) “New” and “casual” customers will likely increase the frequency at which they come to the coffee shop, addressing the diminishing sales generated by these two customer segments. 3) “New” customers will be incentivized to create a customer id to be considered for the deal on beans. Thus, helping to increase the loyalty of these customers. 4) Beans sales are likely to rise as the customer base becomes aware of the quality of Central Perk’s beans.

If the happy hour becomes too successful and customers only purchase the staple products on the menu, the client may have to take some reactionary measures to ensure that the ancillary revenue generated by the happy hour exceeds the costs incurred by the discounts provided via the happy hour.

2)

Central Perk should stop procuring all drink and food options not contained in the revenue_generating_products vector. All products beyond the products contained in this vector provide little to no revenue for the coffee shop. All efforts that go into procuring, promoting, and stocking these products are likely a waste of resources for the client. The investment in these products would be better off allocated to funding the margin reduction anticipated from future happy hour sales.

If the client does choose to implement this tactic, there is a certain level of risk of losing customers that have come to the shop for the low revenue-generating products. As shown in figure 5, this would only be a small fraction of sales and is likely more economical to cater to the broader preferences of customers.