Coffee Data

Katelyn Donaty

Data Visualization

Exploring and visualizing data by answering the following questions:

- How have sales trended over time?
- Which days of the week tend to be busiest?
- Which products are sold the most/least often?
- Which products drive the most revenue?
- What is the most popular brewed coffee?
- What is the most popular brewed tea?
- What is the most popular bakery item?

Checking for NA values

```
## [1] 0
```

No NA values were found.

Summary stats

```
##  transaction_id   transaction_date   transaction_time   transaction_qty
##  Min.   :     1   Length:149116      Length:149116      Min.   :1.000
##  1st Qu.: 37336   Class :character   Class :character   1st Qu.:1.000
##  Median : 74728   Mode  :character   Mode  :character   Median :1.000
##  Mean   : 74737                                         Mean   :1.438
##  3rd Qu.:112094                                         3rd Qu.:2.000
##  Max.   :149456                                         Max.   :8.000
##     store_id     store_location       product_id      unit_price
##  Min.   :3.000   Length:149116      Min.   : 1.00   Min.   : 0.800
##  1st Qu.:3.000   Class :character   1st Qu.:33.00   1st Qu.: 2.500
##  Median :5.000   Mode  :character   Median :47.00   Median : 3.000
##  Mean   :5.342                      Mean   :47.92   Mean   : 3.382
##  3rd Qu.:8.000                      3rd Qu.:60.00   3rd Qu.: 3.750
##  Max.   :8.000                      Max.   :87.00   Max.   :45.000
##  product_category   product_type       product_detail
##  Length:149116      Length:149116      Length:149116
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

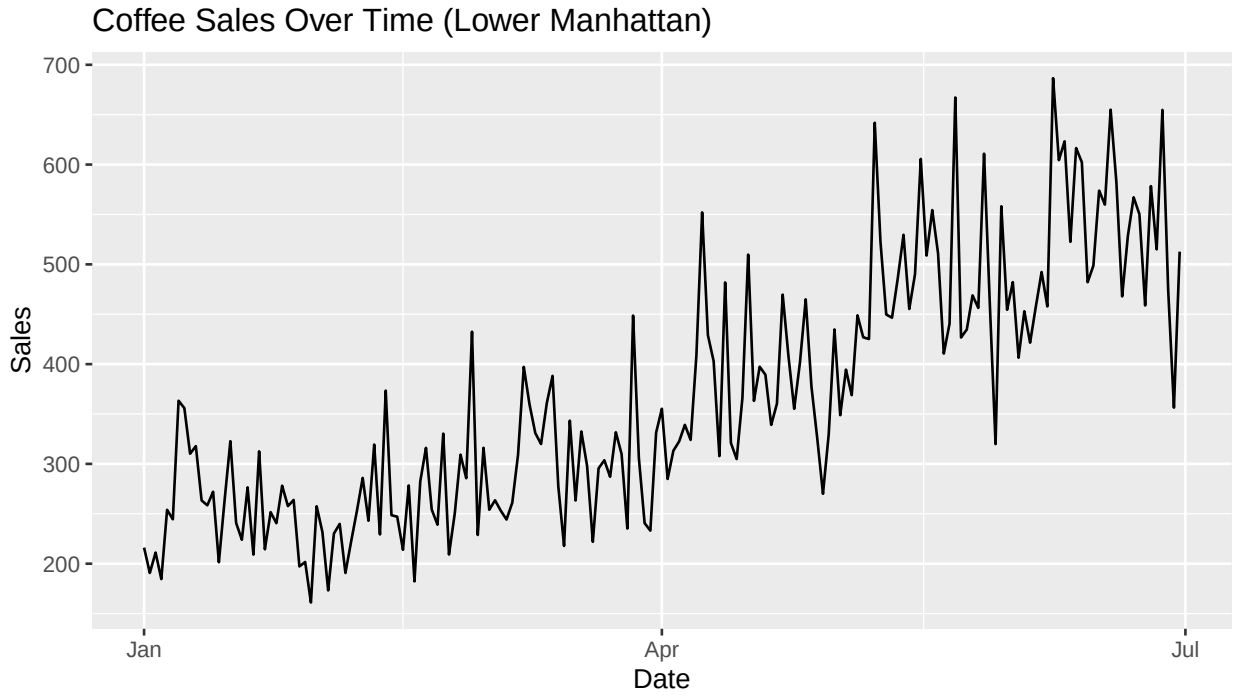
Store locations

```
## [1] "Lower Manhattan" "Hell's Kitchen"  "Astoria"
```

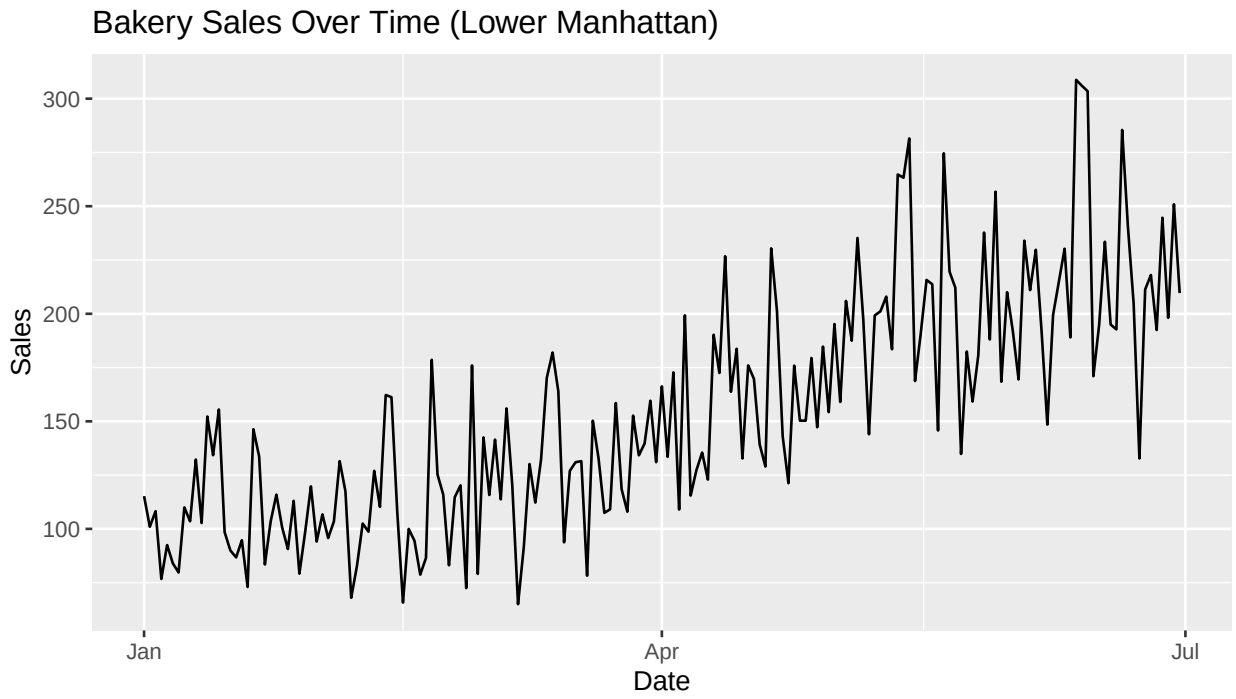
Product categories

```
## [1] "Coffee"             "Tea"                "Drinking Chocolate"
## [4] "Bakery"             "Flavours"           "Loose Tea"
## [7] "Coffee beans"       "Packaged Chocolate" "Branded"
```

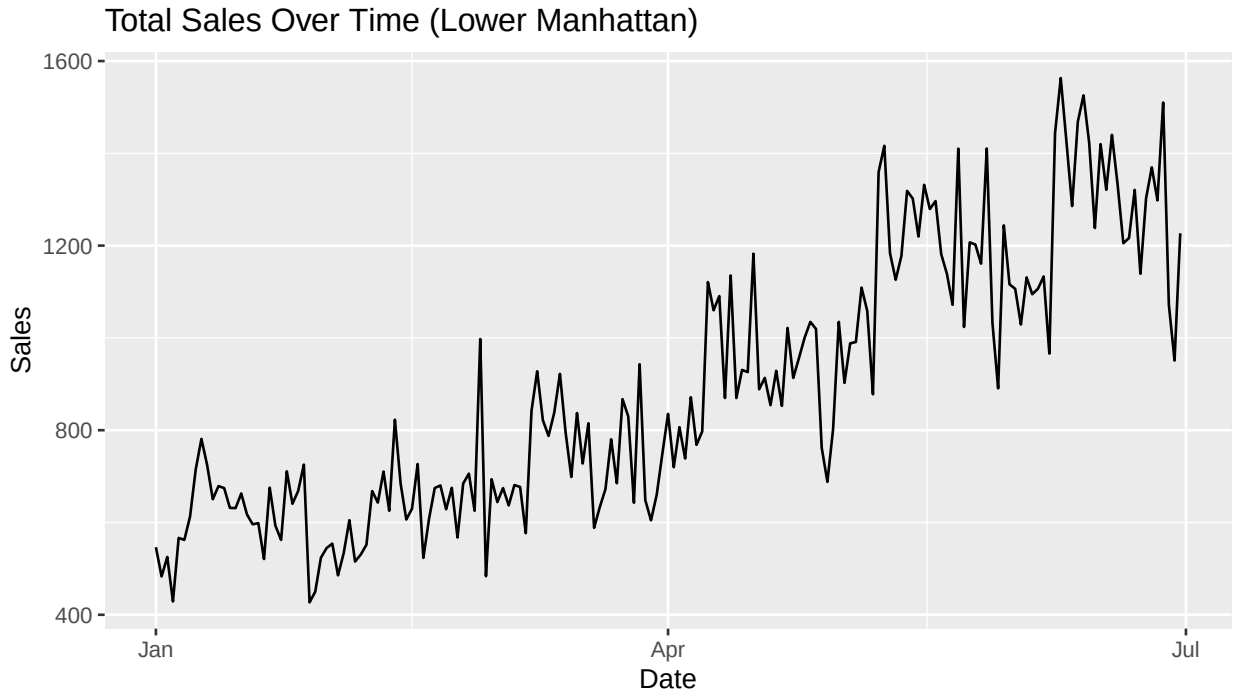
Coffee sales over time (Lower Manhattan store)



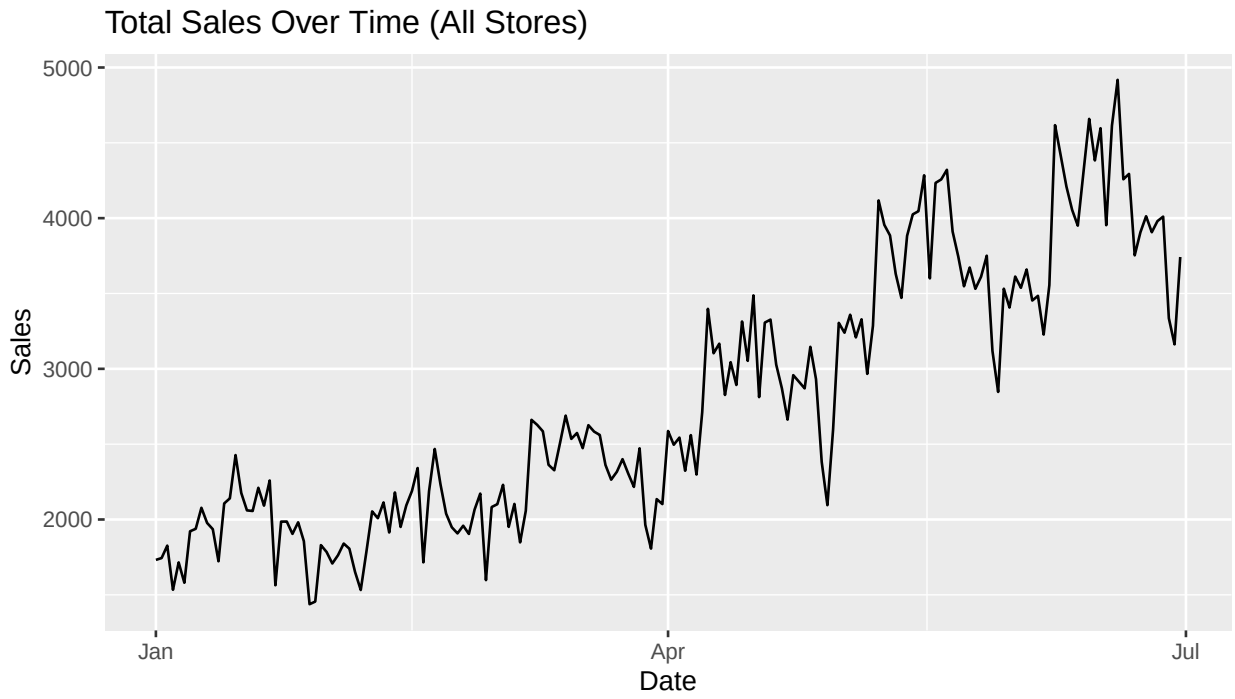
Bakery sales over time (Lower Manhattan store)



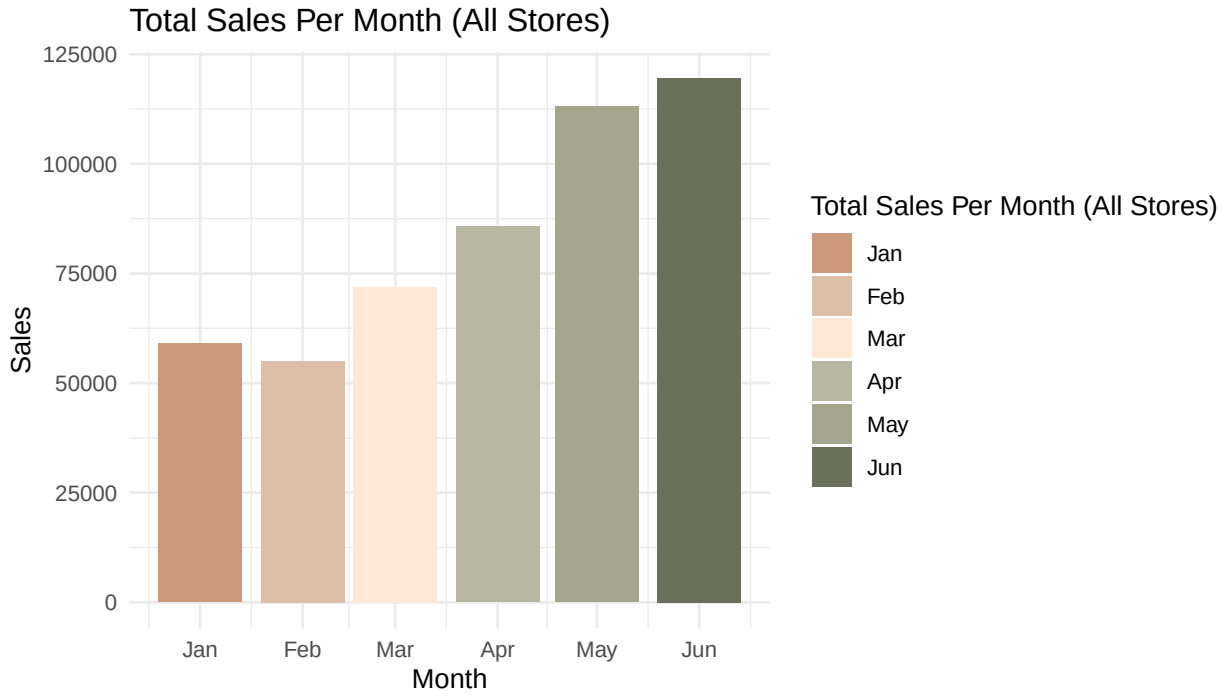
Total sales over time (Lower Manhattan store)



Total sales over time (all stores)

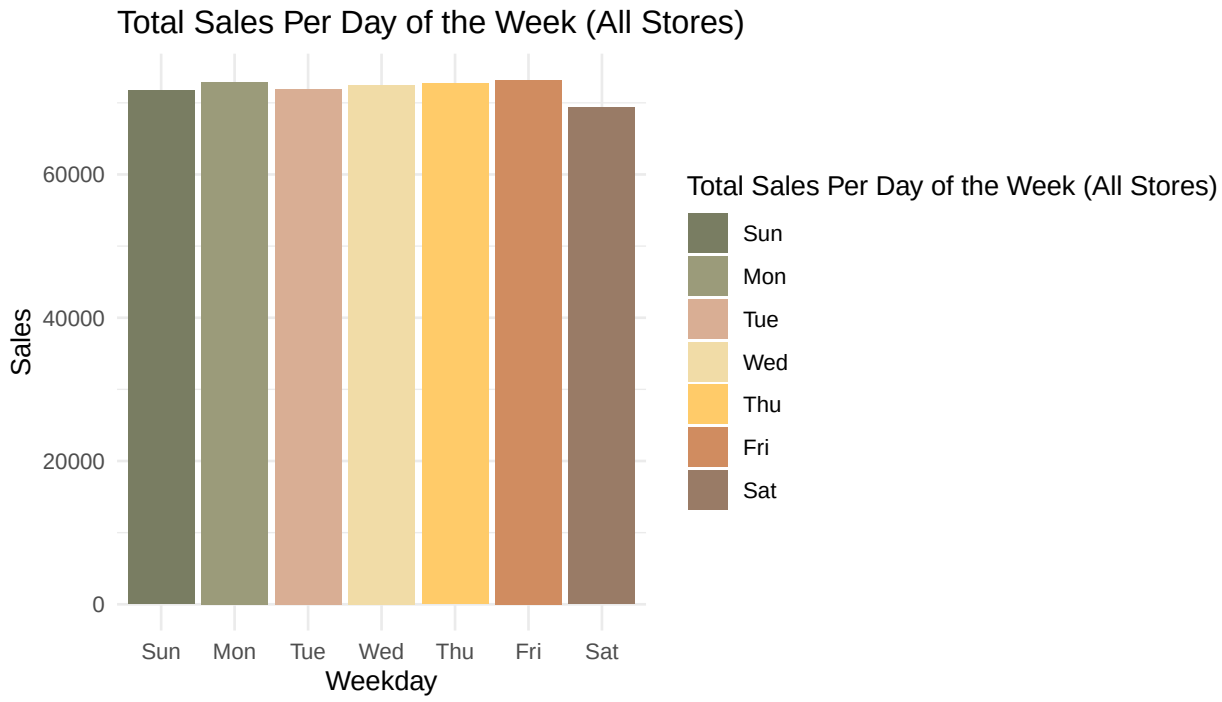


Total sales per month (all stores)



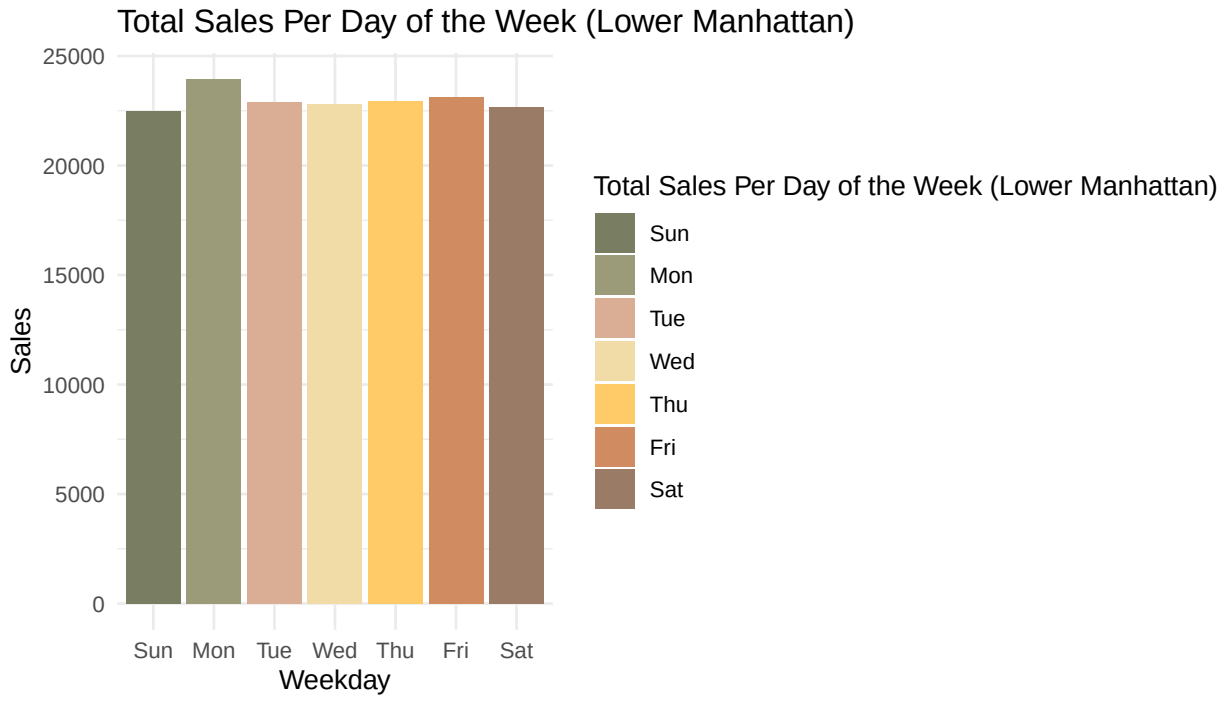
Total sales per day of the week (all stores)

Mondays and Fridays are the busiest overall



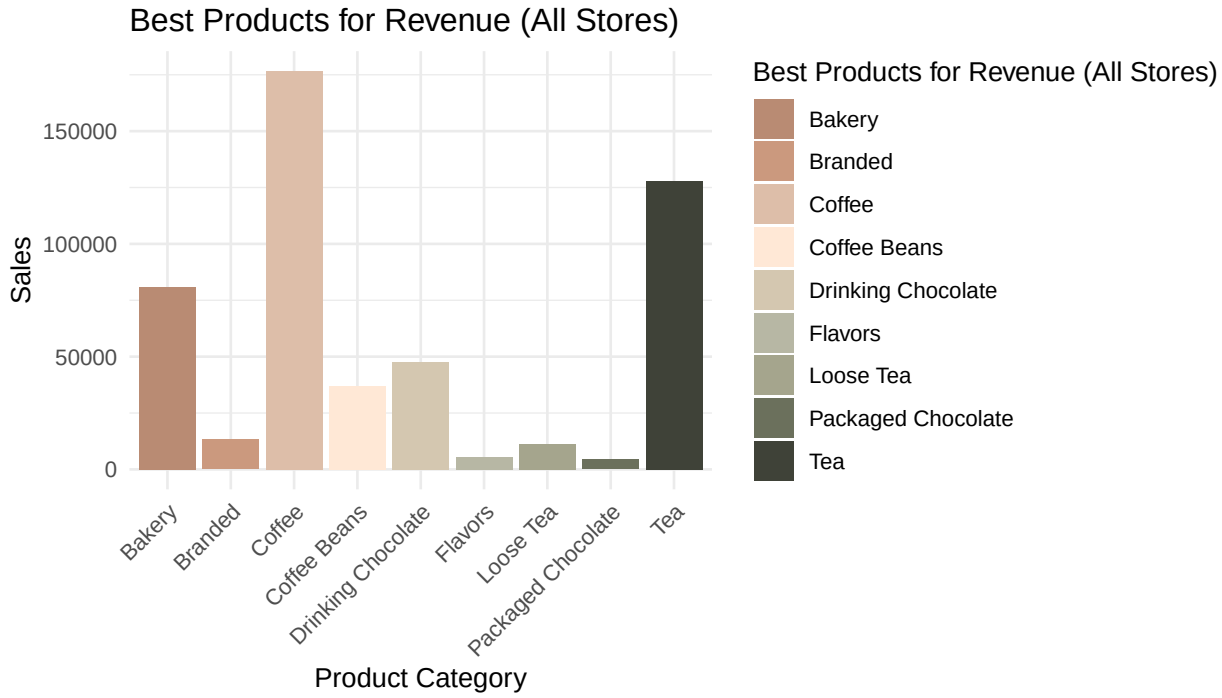
Total sales per day of the week (Lower Manhattan)

Mondays and Fridays are the busiest



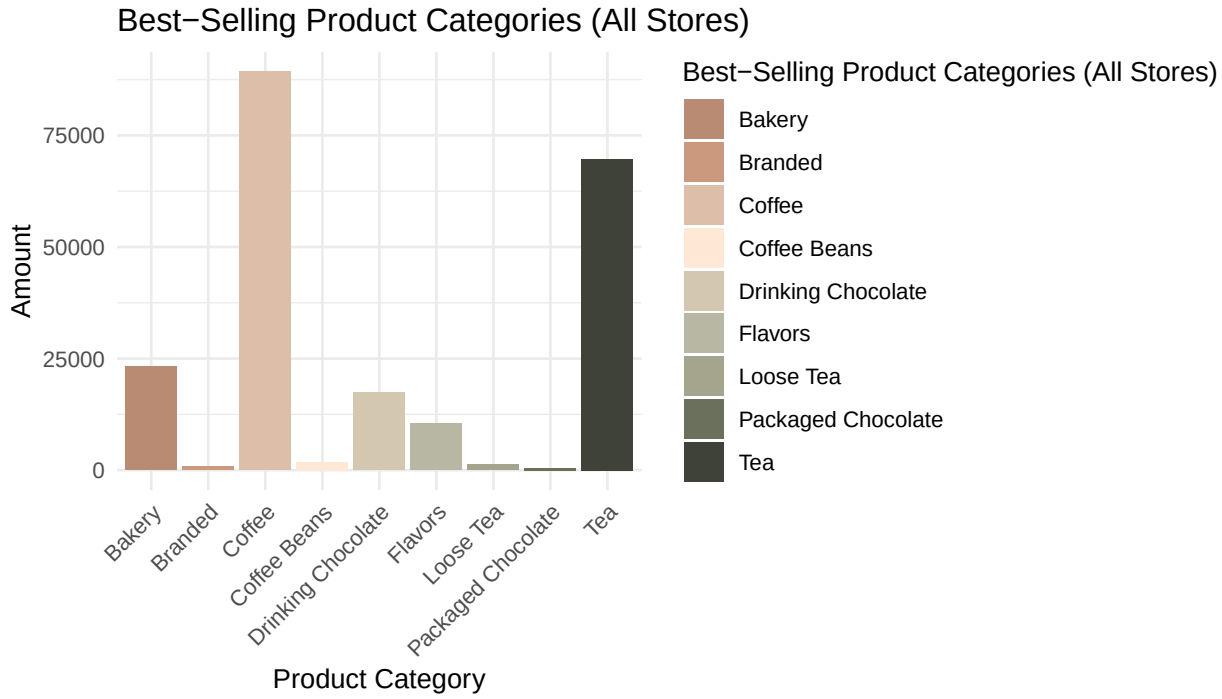
Best product categories for revenue (all stores)

Coffee and Tea (not Loose Tea) bring in the most profit, followed by Bakery items

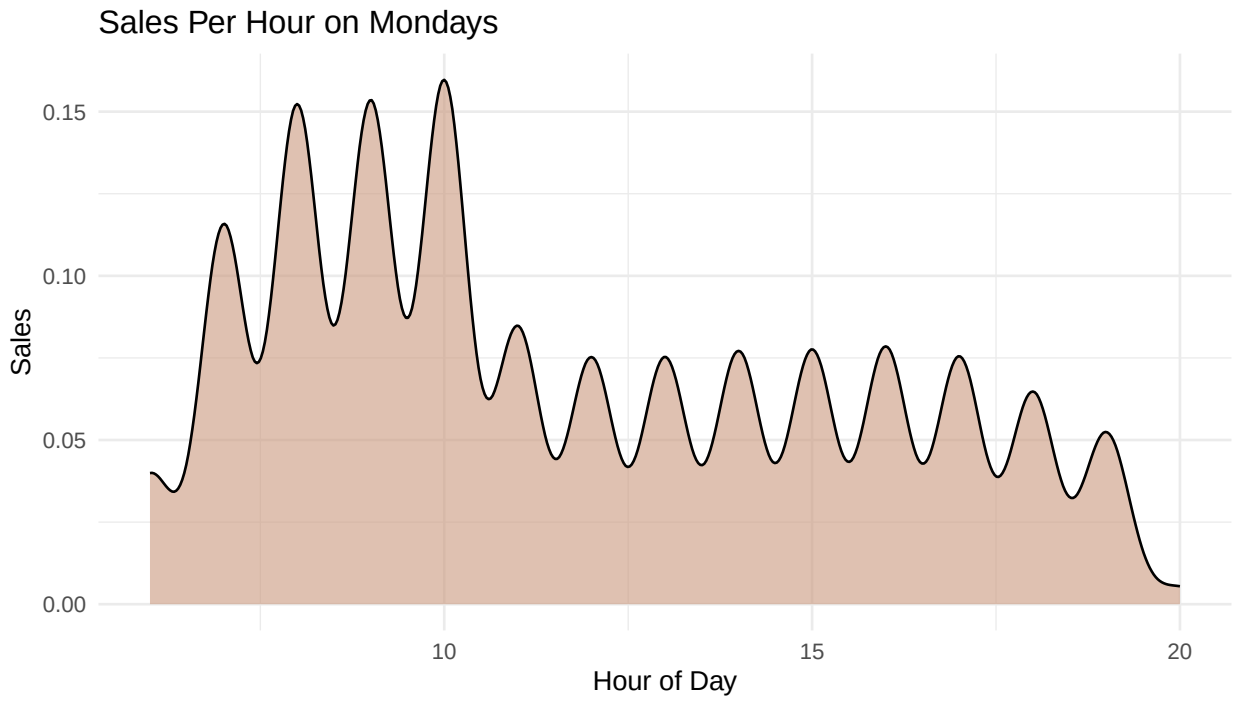


Best-selling products (all stores)

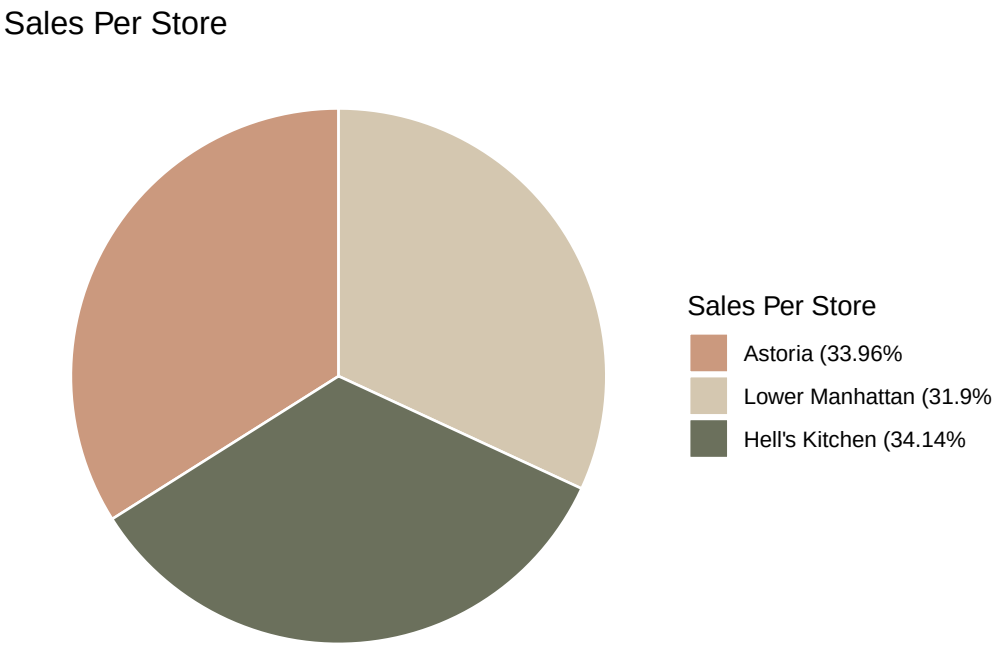
Coffee and Tea (not Loose Tea) sells the most, followed by Bakery items



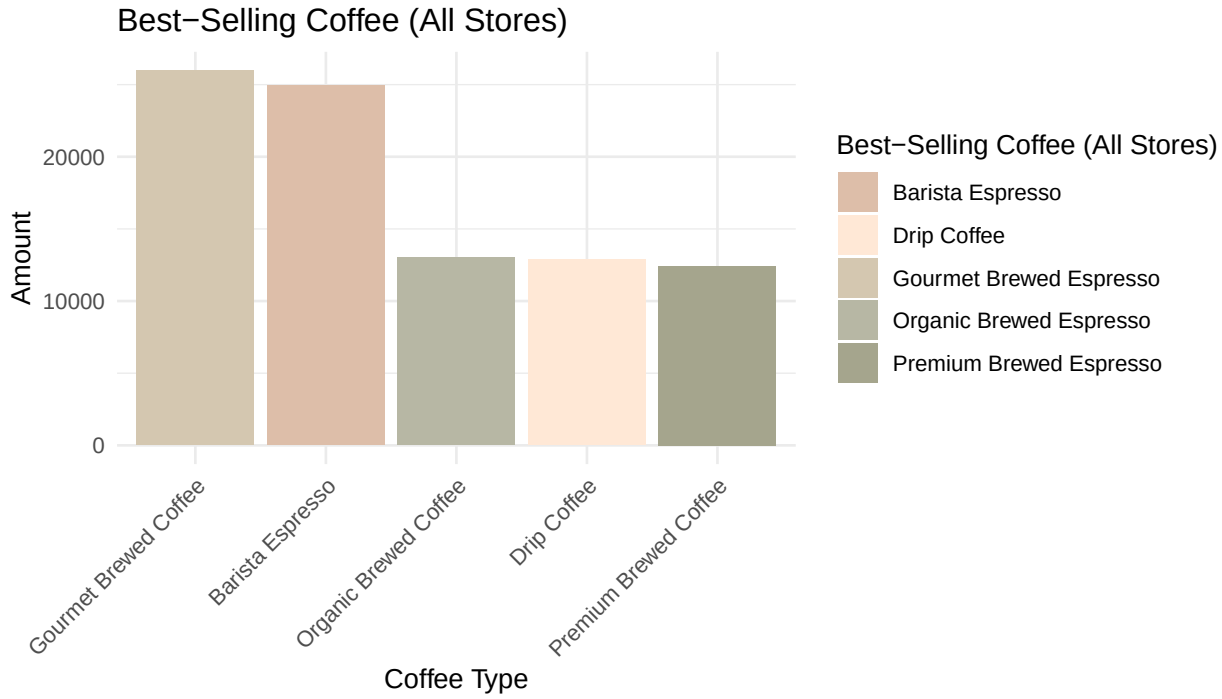
Busiest hours (Monday, all stores)



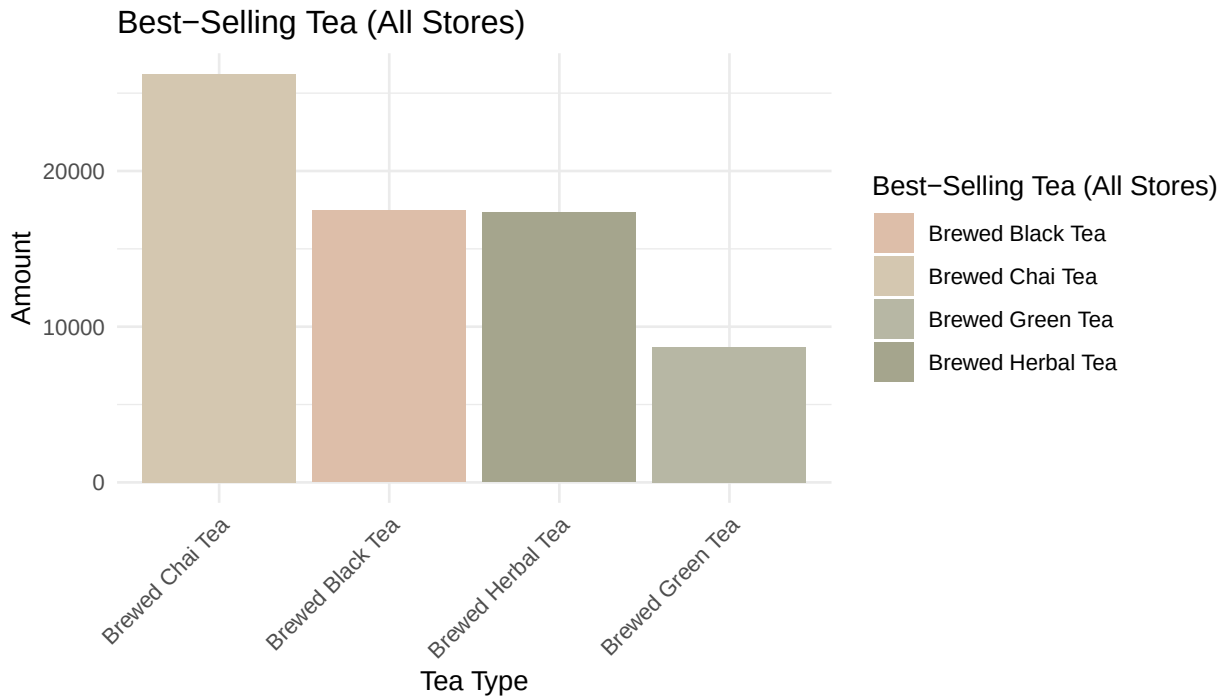
Sales per store



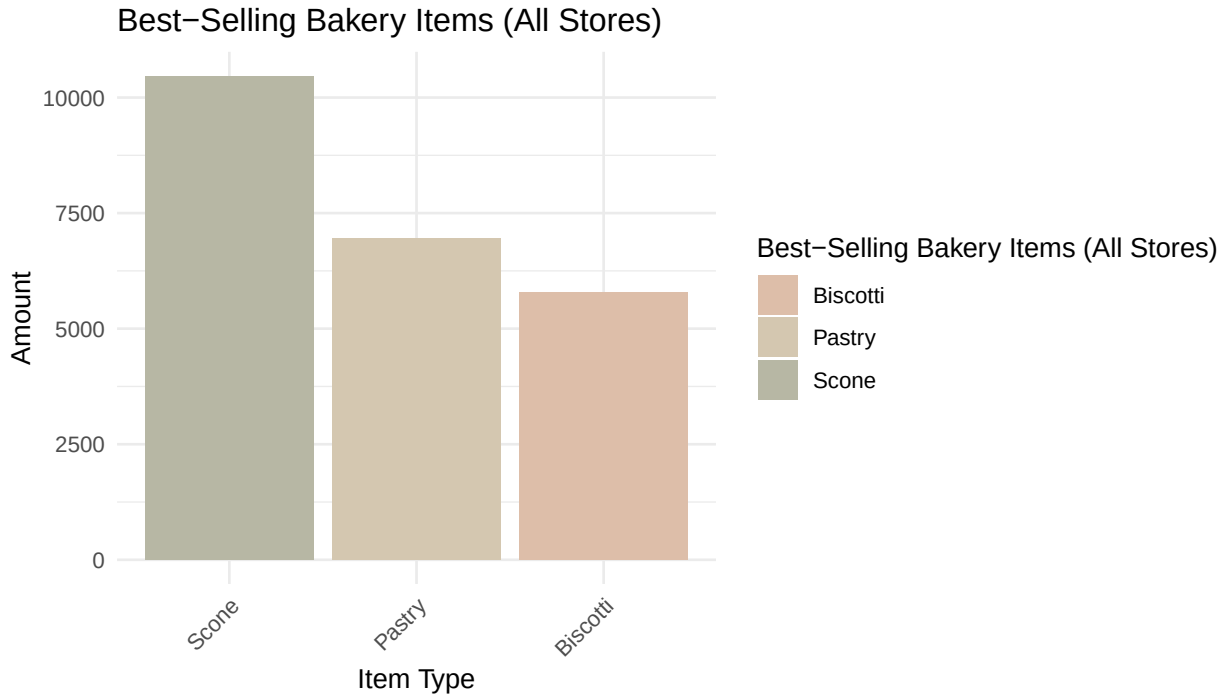
Most popular coffee across all stores



Most popular tea



Most popular bakery



Predictions Using Linear Regression, Random Forest, and Support Vector Machines

Predicting future data by answering the following questions:

- What is the expected coffee sales trend for July to December?
- What is the expected total revenue for the end of December?

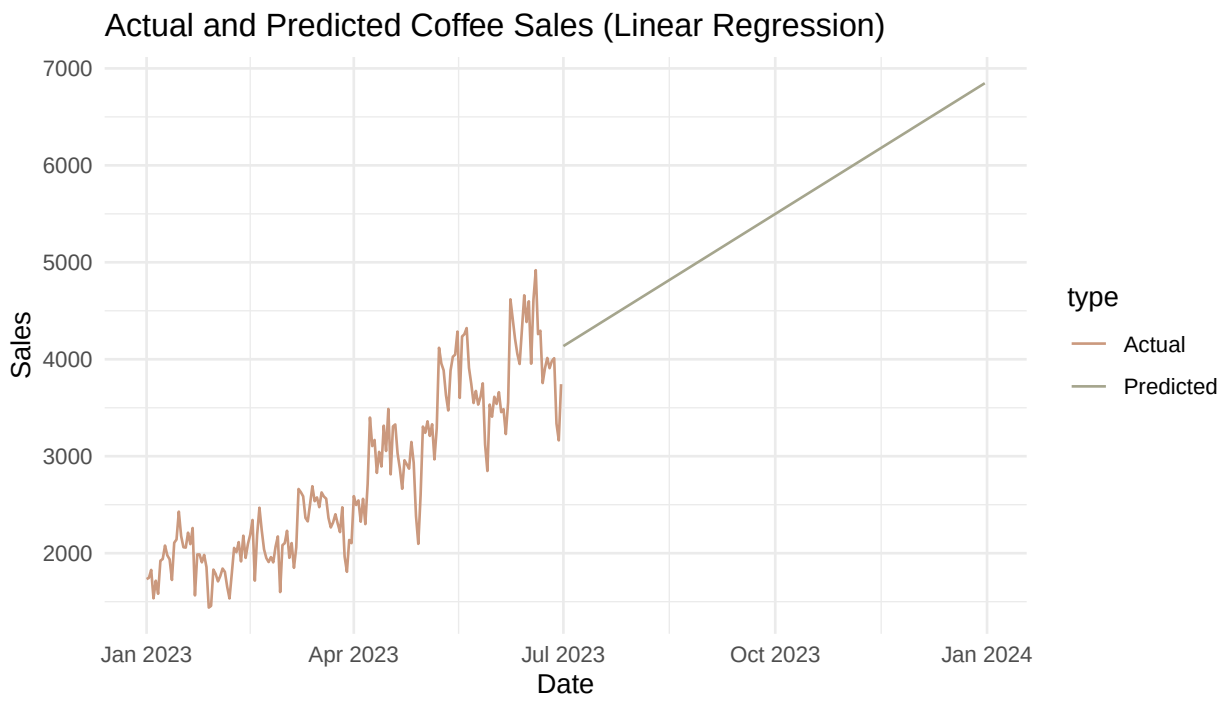
Coffee Sales for July-December (Linear Regression)

Linear Regression Model Summary

```
##
## Call:
## lm(formula = total_sales ~ transaction_date, data = predicted_coffee_sales_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1106.23  -244.57   -11.25   251.29   961.15
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.854e+05  1.074e+04  -26.57   <2e-16 ***
## transaction_date  1.482e+01  5.523e-01   26.83   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 388.2 on 179 degrees of freedom
## Multiple R-squared:  0.8008, Adjusted R-squared:  0.7997
## F-statistic: 719.7 on 1 and 179 DF,  p-value: < 2.2e-16
```

Actual and Predicted Coffee Sales (Linear Regression)



Predicted total revenue (Linear Regression)

Dec 31, 2023: $1,514,940
Jan 1, 2023: $1,732.80

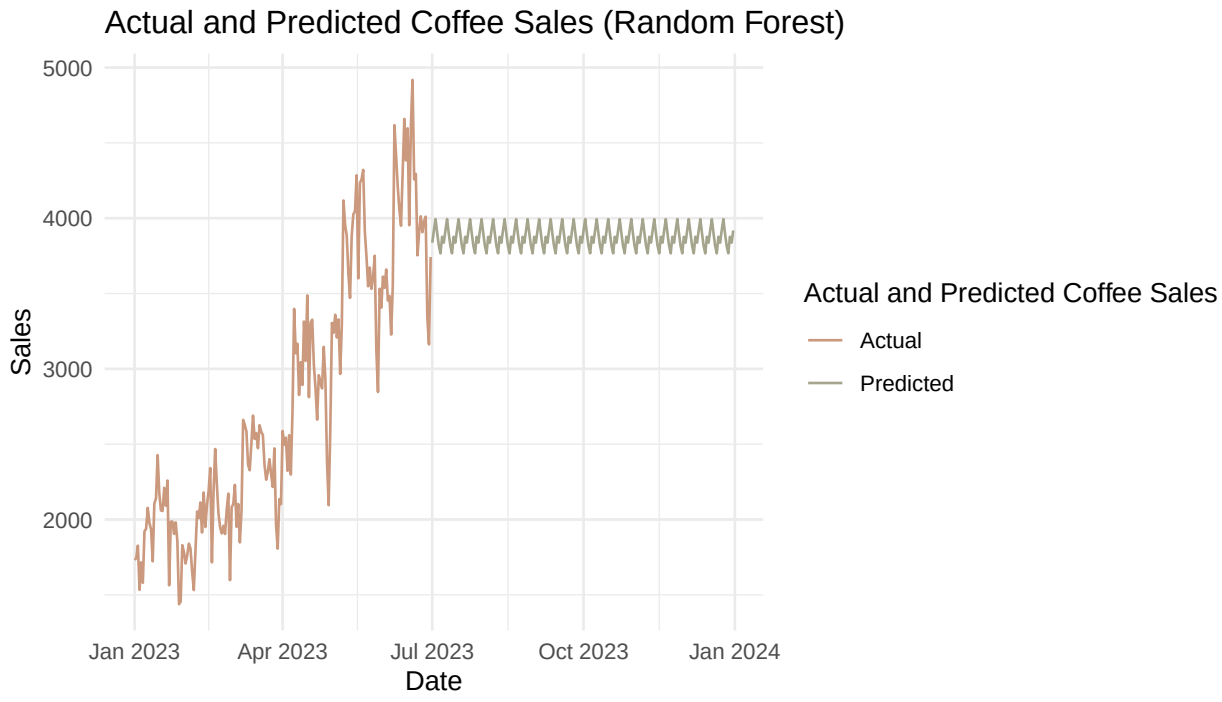
Coffee Sales for July-December (Random Forest)

Random Forest Model Summary

```
##                 Length Class  Mode
## call               4    -none- call
## type               1    -none- character
## predicted        181    -none- numeric
## mse              500    -none- numeric
## rsq              500    -none- numeric
## oob.times        181    -none- numeric
## importance         3    -none- numeric
## importanceSD       0    -none- NULL
## localImportance    0    -none- NULL
## proximity          0    -none- NULL
```

```
## ntree             1    -none- numeric
## mtry              1    -none- numeric
## forest           11    -none- list
## coefs             0    -none- NULL
## y               181    -none- numeric
## test              0    -none- NULL
## inbag             0    -none- NULL
## terms             3    terms  call
```

Actual and Predicted Coffee Sales (Random Forest)



Predicted total revenue (Random Forest)

Dec 31, 2023: $1,216,976
Jan 1, 2023: $1,732.80

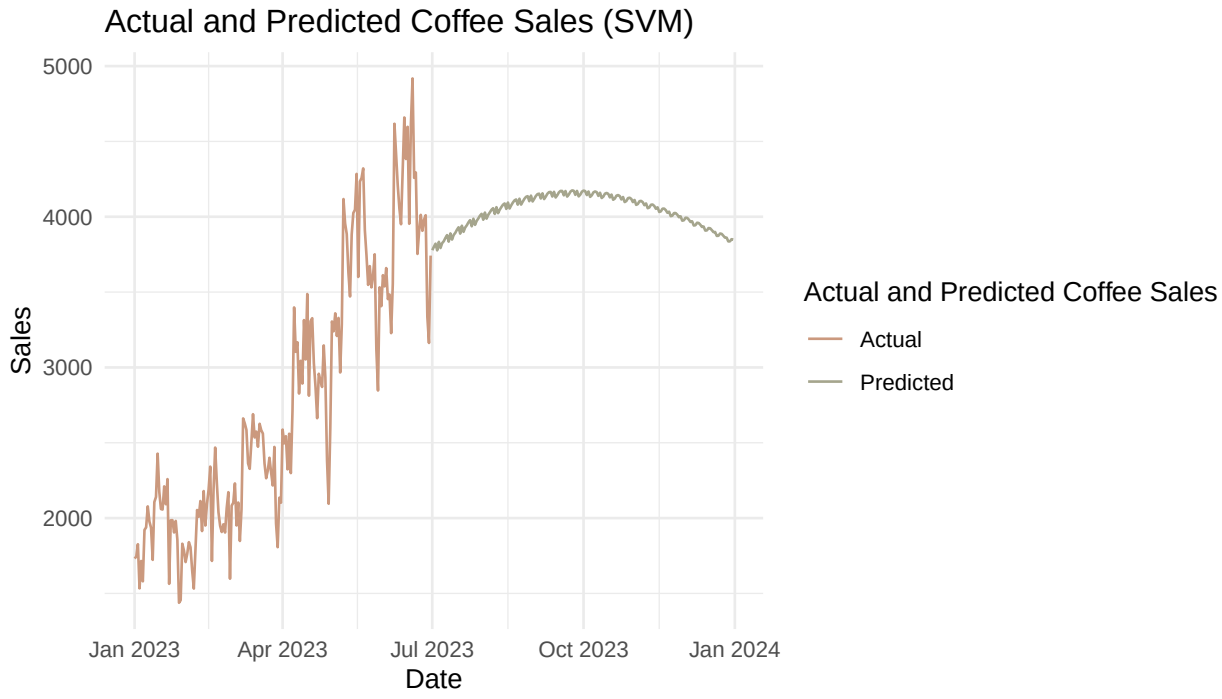
Coffee Sales for July-December (SVM)

SVM Model Summary

```
##
## Call:
## svm(formula = total_sales ~ date_numeric + day_of_week + month, data = svm_data,
##     type = "eps-regression")
##
##
## Parameters:
##    SVM-Type:  eps-regression
```

```
##  SVM-Kernel:  radial
##        cost:  1
##       gamma:  0.05263158
##     epsilon:  0.1
##
##
## Number of Support Vectors:  141
```

Actual and Predicted Coffee Sales (SVM)



Predicted total revenue (SVM)

Dec 31, 2023: $1,247,064
Jan 1, 2023: $1,732.80