Coffee Sales Data Analysis Project

Justin Miller - gk8149

Introduction

The goal of this report is to analyze beverage sales data to identify products driving revenue as well as those that may benefit from promotions to increase sales. Additionally, this report will investigate seasonal sales trends to inform cafe staffing requirements. More specifically, this analysis will answer the two following questions: which beverages are the strongest performers? How does sales volume vary over time?

Data Description

This data set was sourced from Kaggle with transaction dates between March 2024 and March 2025 and contains 3,547 observations and 11 variables. It provides information on transactions of beverages sold and some customer information like payment type, but is primarily composed of the time of transaction (e.g., date, time, etc). The most relevant variables for this analysis are `beverage_name`, `hour_of_day`, `month_abbrev`, and `amount`. It is not clear where the coffee shop is located or what currency is used for payment.

`beverage_name` is a character data type that indicates the kind of coffee beverage purchased. The `hour_of_day` variable is a number data type with values between 6 and 22 (24-hour clock). The `month_abbrev` is a character data type which will be coerced into a factor. These time-related variables will help determine sales trends and staffing needs. The `amount` variable is a number data type where the currency is unknown (not US dollars). Both of these will be used to look at revenue drivers.

Note: a subset of data `sales_no_overlap` was created to avoid inflating March's transaction volume. For the purposes of this report, it was important to identify a typical month for revenue and staffing so the 23 day overlap in March that contained 319 transactions were removed. This subset contains data from March 2024 through February 2025.

Results

### Revenue Drivers

```
# A tibble: 8 x 5
  beverage_name      mean_price num_sold approx_rev approx_share_revenue
  <chr>                   <dbl>    <int>      <dbl>                <dbl>
1 Latte                    35.5      757     26874.               0.239
2 Americano with Milk      30.6      809     24747.               0.220
3 Cappuccino               35.9      486     17438.               0.155
4 Americano                26.0      564     14653.               0.130
5 Hot Chocolate            36.0      276      9933.               0.0885
6 Cocoa                    35.6      239      8520.               0.0759
7 Cortado                  25.7      287      7385.               0.0658
8 Espresso                 20.8      129      2690.               0.024
```

Each beverage had two or more prices listed under amount so the average was taken and used to calculate approximate revenue. The approximate revenue is the `mean_price * num_sold` and the approximate share of revenue aims to estimate the revenue a specific beverage generated. This is not exact because it is using the mean price rather than the sum of prices.

Lattes and americanos with milk are primary revenue drivers, making up about 46% of total revenue. Cappuccinos and americanos are also steady contributors at about 28% of revenue between the two. Though americanos are purchased more frequently than cappuccinos, their mean sales price is lower suggesting the greatest opportunity to increase revenue is by promoting cappuccinos.

Volume Sold and Average Price per Beverage
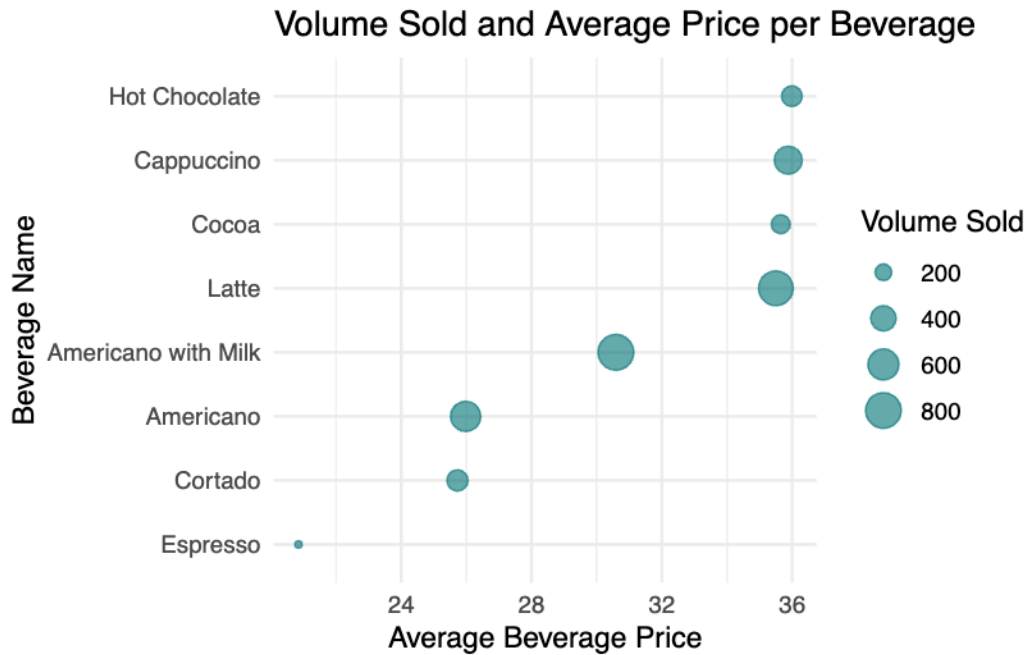
This plot shows the average beverage price with points relative to the volume of drinks sold. There is no clear relationship between average price and volume sold, indicating price may not be a driving factor on type of beverage purchased. Additionally, this helps visualize the promotional potential of hot chocolate and cocoa where the volume sold is lower but average price is high and marginal increases in sales could increase revenue more quickly.

## Seasonal Revenue Trends (March 2024 – February 2025)
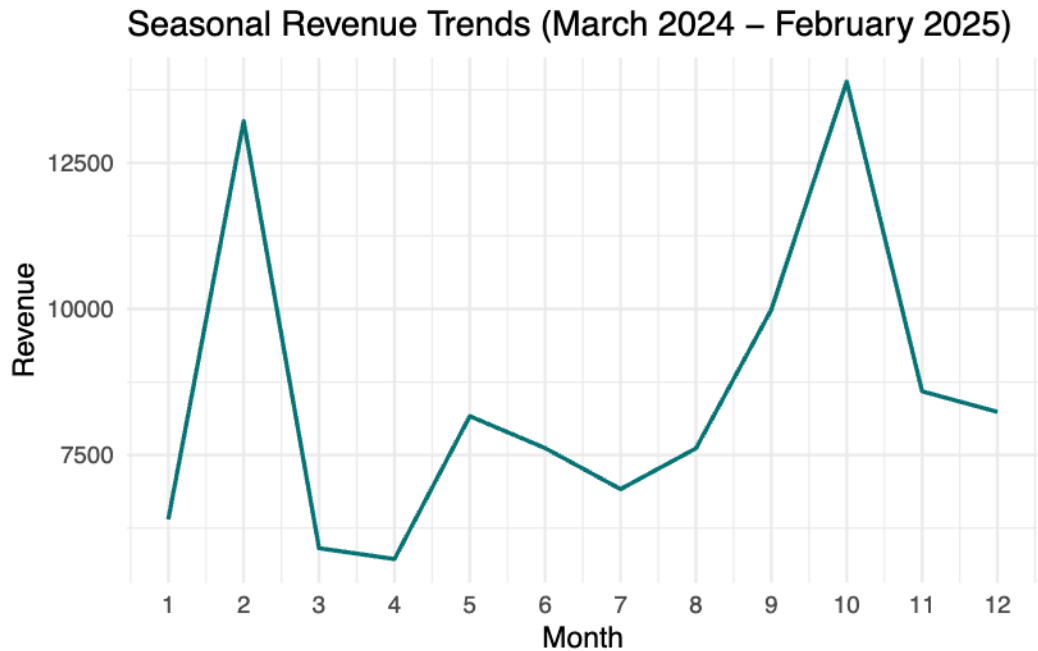
This line graph uses the subset data `sales_no_overlap` where January and February consist of transactions from 2025 and March through December consist of transactions from 2024.

Revenue shows two clear spikes in February and October and two troughs in March and April. Revenue appears to be seasonal but steady mid-year between May and August.

Transaction Times & Staffing

Beverages Sold by Hour of Day – Monthly
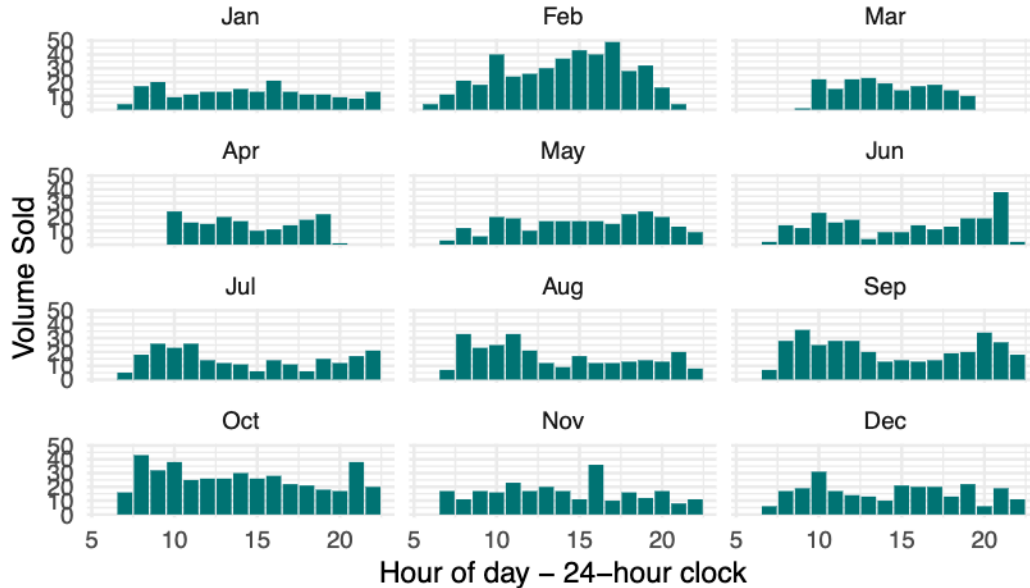
This plot uses the subset of data `sales_no_overlap` and shows the the number of beverages sold in each hour of the day (between 6AM and 10PM, though not all months register transactions every hour). All year around, sales seem to peak in the mid-morning hours, followed by consistent sales into the afternoon and steady or declining sales into the evening. Notably, June, September and October see a second peak during evening hours. March and April see less sales volume in less number of hours than other months, suggesting possible difference in operations. Because location of this coffee shop is unknown, it's difficult to know what may be influencing these fluctuations (e.g., weather, holidays, etc.).

Conclusion

This coffee shop leads in sales and revenue with lattes and americanos with milk. There is strong potential to grow revenue through incentives for hot chocolate, cocoa, and cappuccinos which see steady purchase volume with high average prices. Monthly revenue is seasonal but steady with strong performances in February and October and room to grow in March and April. For staffing, sales are strongest in the morning and into the early afternoon so increased staffing earlier in the day year-round with all-day coverage for February and October is suggested. There is strong potential for promotions in the afternoon to increase already steady sales during that time. In a future analysis, currency type, coffee shop location, and other product purchase information could be helpful in providing more helpful insights.