Statinators Project: Coffee Sales

Kara Kochansky, Addy McGarry, Abbie Qiu, Jenna Rishi

Today's agenda

- Topic + Motivation
- Introducing the Data
- Graphs 1-3
- Graphs 4-6
- Conclusion, discussion, critique

Research Question: How do season and time of day affect the type of coffee people buy?

Motivation

Coffee is something we all have in common as something we like! Interesting to investigate when we have differences in academic interests

Hypothesis

Colder months → more hot chocolate and lattes

Mornings → more espresso drinks like Americanos

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle
  - “Coffee Sales Dataset” by Saad Ali Yaseen

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle
  - “Coffee Sales Dataset” by Saad Ali Yaseen
- Collected data from a coffee vending machine

Int



- saienkov on Kaggle
- Ali Yaseen
- machine

Int

-
-


PREMIUM
Gourmet
Coffee
HOT DRINKS
MENU


MENU
Premium Blend
Decaf Coffee
French Vanilla
Café Mocha
Hot Chocolate
White Chocolate Caramel
English Toffee
Caffè Vanile
French Vanilla Mocha

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle
  - “Coffee Sales Dataset” by Saad Ali Yaseen
- Collected data from a coffee vending machine
- 3547 rows and 11 columns

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle
  - “Coffee Sales Dataset” by Saad Ali Yaseen
- Collected data from a coffee vending machine
- 3547 rows and 11 columns
  - Rows: Individual coffee transactions
  - Columns: time, day, month, payment method, cost, coffee type, and sorting columns

Introducing the Data:

- “Coffee Sales” published by Yaroslav Isaienkov on Kaggle
  - “Coffee Sales Dataset” by Saad Ali Yaseen
- Collected data from a coffee vending machine
- 3547 rows and 11 columns
  - Rows: Individual coffee transactions
  - Columns: time, day, month, payment method, cost, coffee type, and sorting columns
- We created a variable called “Seasons”

Introducing the Data:

coffee_sales	[3547 rows x 11 columns] <spec_tbl_df>	
hour_of_day	10 12 12 13 13 15 16 18 19 19 19 10 10 11 14 16 17 10 11 12 1...	dbl [3547]
cash_type	"card" "card" "card" "card" "card" "card" "card" "card" "c...	str [3547]
money	38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7 33.8 33.8 28....	dbl [3547]
coffee_name	"Latte" "Hot Chocolate" "Hot Chocolate" "Americano" "Lat...	str [3547]
Time_of_Day	"Morning" "Afternoon" "Afternoon" "Afternoon" "Afternoon...	str [3547]
Weekday	"Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "S...	str [3547]
Month_name	"Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "...	str [3547]
Weekdaysort	5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 1 1 1 1 2 2...	dbl [3547]
Monthsort	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ...	dbl [3547]
Date	"2024-03-01" "2024-03-01" "2024-03-01" "2024-03-01" "202...	Date
Time	"10:15:50" "12:19:22" "12:20:18" "13:46:33" "13:48:14" "15:39:4...	hms

av Isaienkov on Kaggle

ad Ali Yaseen

ng machine

actions

payment method, cost, coffee type, and sorting

ns”

Introducing the Data:

coffee_sales	[3547 rows x 11 columns] <spec_tbl_df>	
hour_of_day	10 12 12 13 13 15 16 18 19 19 19 10 10 11 14 16 17 10 11 12 1...	dbl [3547]
cash_type	"card" "card" "card" "card" "card" "card" "card" "card" "c...	str [3547]
money	38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7 33.8 33.8 28....	dbl [3547]
coffee_name	"Latte" "Hot Chocolate" "Hot Chocolate" "Americano" "Lat...	str [3547]
Time_of_Day	"Morning" "Afternoon" "Afternoon" "Afternoon" "Afternoon...	str [3547]
Weekday	"Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "S...	str [3547]
Month_name	"Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "...	str [3547]
Weekdaysort	5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 1 1 1 1 2 2...	dbl [3547]
Monthsort	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ...	dbl [3547]
Date	"2024-03-01" "2024-03-01" "2024-03-01" "2024-03-01" "202...	Date
Time	"10:15:50" "12:19:22" "12:20:18" "13:46:33" "13:48:14" "15:39:4...	hms

coffee_df	[3547 rows x 12 columns] <tbl_df>	
hour_of_day	10 12 12 13 13 15 16 18 19 19 19 10 10 11 14 16 17 10 11 12 1...	dbl [3547]
payment_met	"card" "card" "card" "card" "card" "card" "card" "card" "c...	str [3547]
money	38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7 33.8 33.8 28....	dbl [3547]
coffee_name	"Latte" "Hot Chocolate" "Hot Chocolate" "Americano" "Lat...	str [3547]
time_of_day	"Morning" "Afternoon" "Afternoon" "Afternoon" "Afternoon...	str [3547]
weekday	"Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "S...	str [3547]
month_name	"Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "...	str [3547]
weekday_sort	5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 1 1 1 1 2 2...	dbl [3547]
month_sort	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ...	dbl [3547]
date	"2024-03-01" "2024-03-01" "2024-03-01" "2024-03-01" "202...	Date
time	"10:15:50" "12:19:22" "12:20:18" "13:46:33" "13:48:14" "15:39:4...	hms
season	"Spring" "Spring" "Spring" "Spring" "Spring" "Spring" "...	fct(4) [3547]

Introducing the Data:

coffee_sales	[3547 rows x 11 columns] <spec_tbl_df>	
hour_of_day	10 12 12 13 13 15 16 18 19 19 19 10 10 11 14 16 17 10 11 12 1...	dbl [3547]
cash_type	"card" "card" "card" "card" "card" "card" "card" "card" "c...	str [3547]
money	38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7 33.8 33.8 28....	dbl [3547]
coffee_name	"Latte" "Hot Chocolate" "Hot Chocolate" "Americano" "Lat...	str [3547]
Time_of_Day	"Morning" "Afternoon" "Afternoon" "Afternoon" "Afternoon...	str [3547]
Weekday	"Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "S...	str [3547]
Month_name	"Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "...	str [3547]
Weekdaysort	5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 1 1 1 1 2 2...	dbl [3547]
Monthsort	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ...	dbl [3547]
Date	"2024-03-01" "2024-03-01" "2024-03-01" "2024-03-01" "202...	Date
Time	"10:15:50" "12:19:22" "12:20:18" "13:46:33" "13:48:14" "15:39:4...	hms

coffee_df	[3547 rows x 12 columns] <tbl_df>	
hour_of_day	10 12 12 13 13 15 16 18 19 19 19 10 10 11 14 16 17 10 11 12 1...	dbl [3547]
payment_met	"card" "card" "card" "card" "card" "card" "card" "card" "c...	str [3547]
money	38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7 33.8 33.8 28....	dbl [3547]
coffee_name	"Latte" "Hot Chocolate" "Hot Chocolate" "Americano" "Lat...	str [3547]
time_of_day	"Morning" "Afternoon" "Afternoon" "Afternoon" "Afternoon...	str [3547]
weekday	"Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "Fri" "S...	str [3547]
month_name	"Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "Mar" "...	str [3547]
weekday_sort	5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 1 1 1 1 2 2...	dbl [3547]
month_sort	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ...	dbl [3547]
date	"2024-03-01" "2024-03-01" "2024-03-01" "2024-03-01" "202...	Date
time	"10:15:50" "12:19:22" "12:20:18" "13:46:33" "13:48:14" "15:39:4...	hms
season	"Spring" "Spring" "Spring" "Spring" "Spring" "Spring" "...	fct(4) [3547]



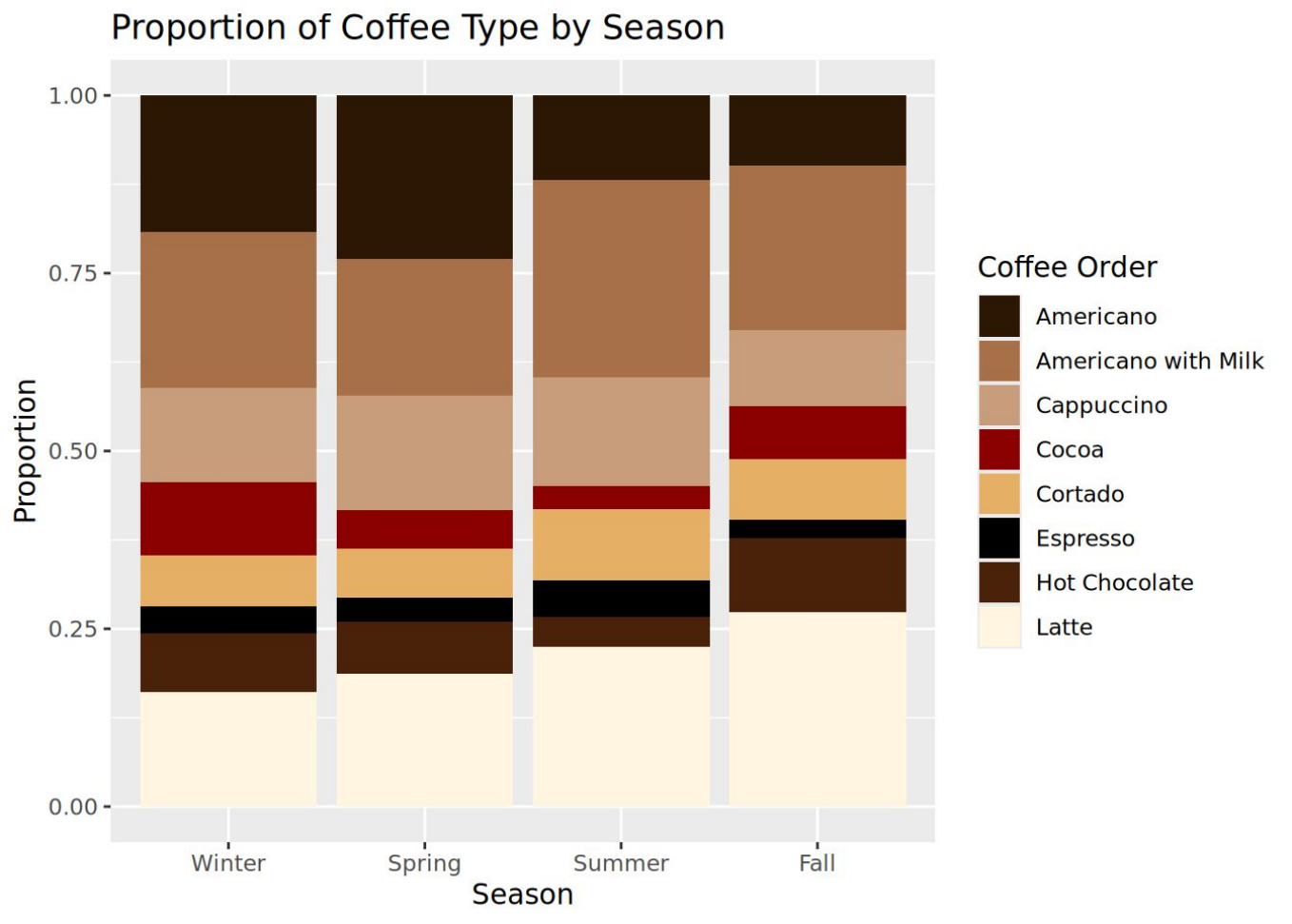
coffee_name



season



time_of_day



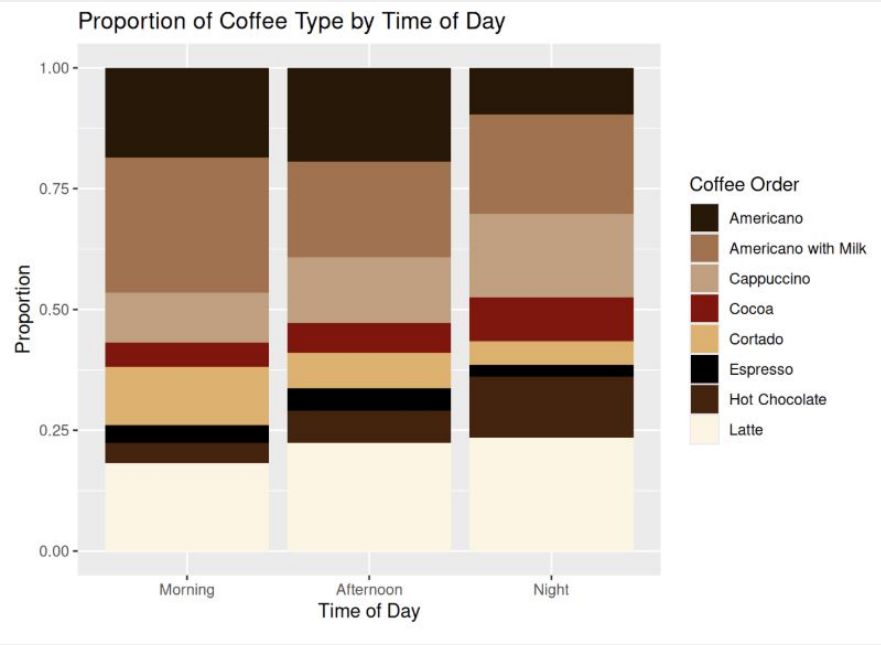
Exploratory Data Analysis: First Graph

→ Lattes: peak in fall (a little over 25%), lowest in winter (around 15%)

→ Americanos: peak in spring (a little under 25%), lowest in fall (around 10%)

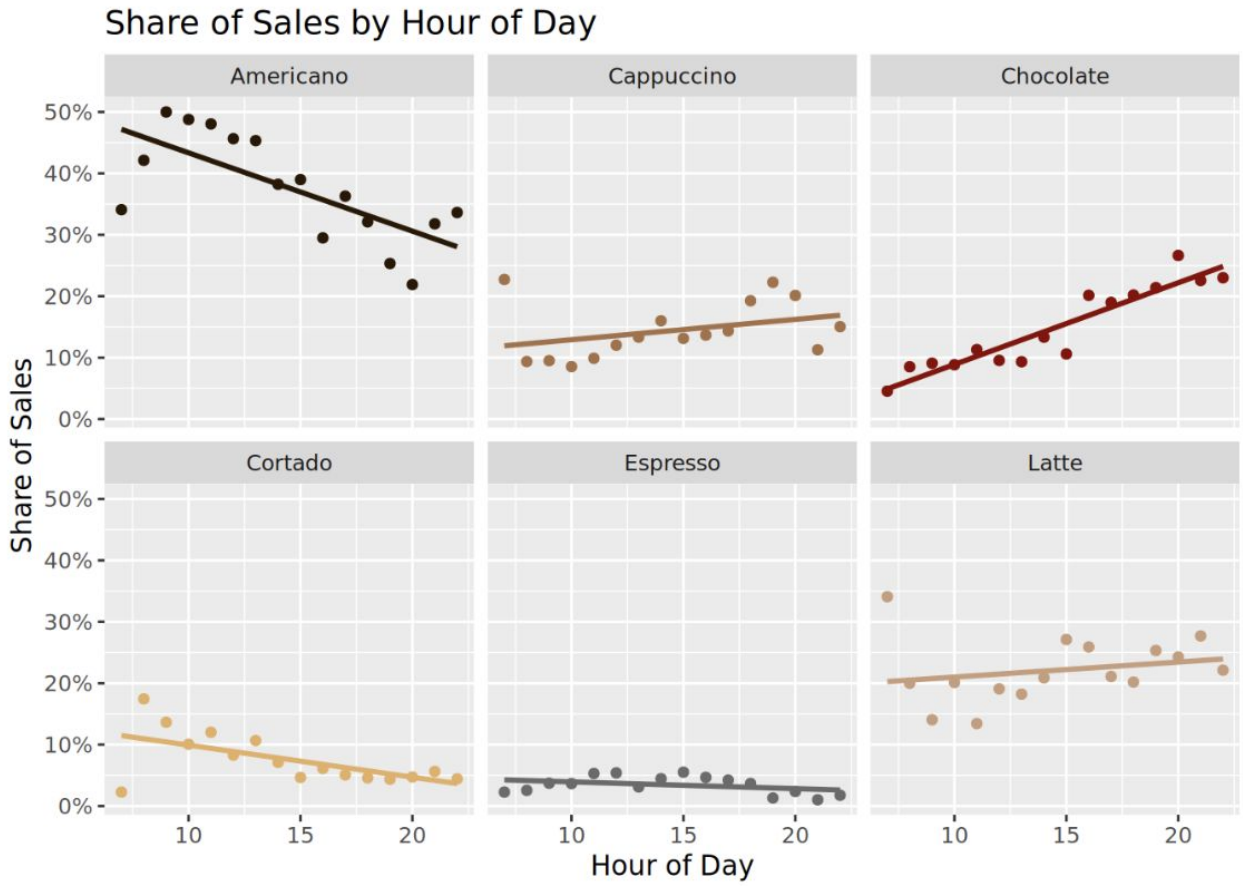
→ Chocolate drinks (Cocoa & Hot Chocolate): (slightly under 20%) in winter vs (slightly under 10%) in summer

Exploratory Data Analysis: Second Graph



Proportion of coffee type by time of day

- Morning, afternoon, and night
- Changing proportions of Americanos, hot chocolate, and cappuccinos over day
- Consistent latte sales throughout the day



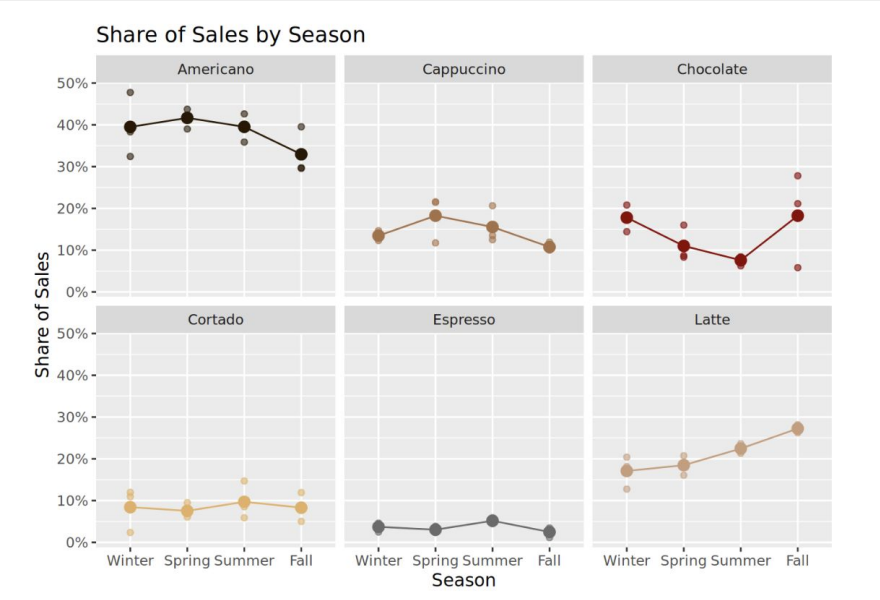
Exploratory Data Analysis: Third Graph

→ Americanos: Clearest decline, falling from 47% of sales in the morning to only 28% at night

→ Cortados: Significant decline of about 0.05% per hour

→ Chocolate drinks (Cocoa & Hot Chocolate): Strongest positive trend, gaining 1.3 percentage points per hour (5% to 25%)

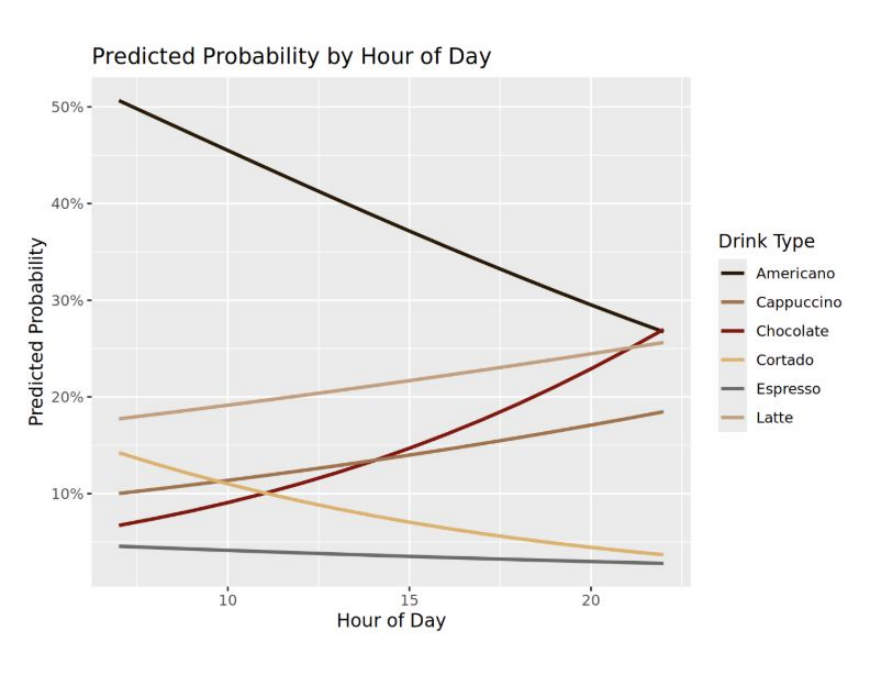
Exploratory Data Analysis: Fourth Graph



Share of sales by season

- Lattes: Clearest seasonal pattern, rising from 17% in winter to 27% in fall
- Chocolate drinks: 10% drop in sales from winter to summer
- Changes in sales for Americanos, cappuccinos, cortados, and espresso changes were not significant

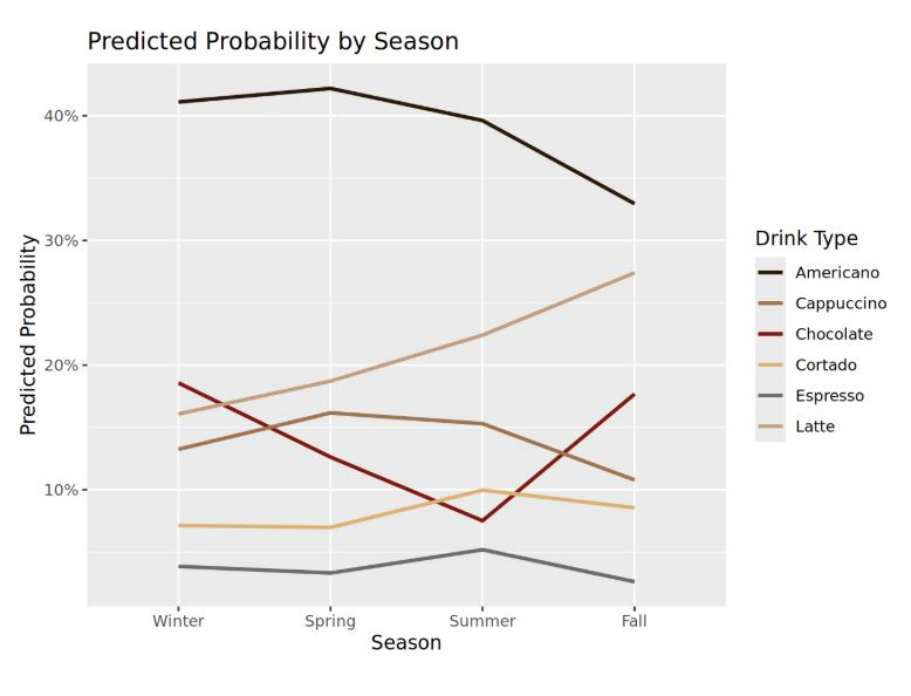
Exploratory Data Analysis: Fifth Graph



Predicted probability of each coffee type by hour of day

- Hours from 0-23
- Predicted purchase of americano, cortado, and espresso decrease
  - Chocolate, latte, chocolate increase
- Americano and chocolate have steepest slopes
- Americano has highest probability at start of day at ~50%

Exploratory Data Analysis: Sixth Graph



Predicted probability of each coffee type by season

- Winter, spring, summer, fall
- Chocolate has a sharp turnaround post summer at around 8%
- Americanos slightly increase then consistently decrease from around 40%
- Espresso stays pretty consistent starting and ending around 4%

Conclusions, discussion & critique

- Correlation present between type of drink and time of day as well as season
  - Specifically with Cappuccinos, Lattes, and Americanos
  - Greatest slopes and changes: Americano, Chocolates, Lattes
- Likely influenced by external factors:
  - Drink availability
  - New flavor alternatives
  - Temperature
- Lack of information on dataset
- Minimal understanding of exact details of drinks and availability

Sources

Dataset:

https://www.kaggle.com/datasets/saadaliyaseen/coffee-sales-dataset/data