

# Investigating Customer's Grocery Shopping Habits During the COVID-19 Pandemic

Capstone Project  
BANA 785  
Professor: Victor Perotti



## **Presented by:**

Dana DiVincenzo, Devren Purdie, Pyone Myat Maw,  
Surbhi Sharma, Grace Matabishi

# Table of contents

1	Project Introduction <i>Data and project goals</i>	Market Basket Analysis <i>Department focused purchasing patterns</i>	5
2	Our Hypothesis <i>Cooking vs. Restaurant department trends</i>	Time Series Predictive Modeling <i>ARIMA</i>	6
3	Exploratory Data Analysis <i>General finds and hypothesis focused</i>	Future Scopes	7
4	Decision Tree Predictive Modeling <i>Feature Importance</i>	Conclusions	8

# Project Introduction

---

## Data

- Wegmans data
- Additional data
  - Covid data Monroe County

## Goal

- Investigate how the COVID-19 pandemic impacted Wegmans' customers shopping habits

# Our Hypothesis

- Customers would change their buying habits.
- With restaurants closed, customers will cook more.
- In tandem, *'restaurant' department* sales would decrease.
- As restaurants restrictions loosened, shopping habits for 'cooking' and 'restaurant' categories would return to normal

Department Categories

Cooking	Restaurant
Asian	Digital Fulfillment
Bakeshop	In-Store Desserts
Cold Cuts & Cheese	Instore Breakfast
Cultured Dairy	Of Submarine Shop
Dairy	Pizza
Dairy Pre-Pack	Restaurants
Eggs	Salads, Sandwiches & Soups
Fresh Seafood	Sushi
Frozen Food	Ultimate Coffee
Frozen Meat	---
Frozen Seafood	---
Grocery	---
Homestyle American	---
Ice Cream	---
In-Store Bread & Rolls	---
Italian	---
Meat	---
Olde World Cheese	---
Produce	---

# Exploratory Data Analysis

Cooking Stats By Income

Hh Income ₪	Avg. Hh Children	Hh Children	Avg. Hoh Age	Avg. Sales	% of Total Count of Wegmans_default_...	% of Total Sales along Table (Down)
15K-20K	0.64	55,310	53	2.7896	1.84%	1.78%
20K-30K	0.64	143,200	55	2.7373	4.72%	4.48%
30K-40K	0.92	262,891	53	2.7831	6.09%	5.88%
40K-50K	0.79	328,213	54	2.8209	8.82%	8.64%
50K-75K	0.99	1,258,694	53	2.8240	26.92%	26.40%
75K-100K	1.11	897,963	55	2.9178	17.17%	17.40%
100K-125K	0.98	397,089	57	2.9467	8.56%	8.76%
125K+	1.06	1,093,408	52	2.9979	21.83%	22.73%
<10K	0.62	84,665	53	2.6839	2.88%	2.68%
NA	0.00	0	49	2.9965	1.18%	1.23%

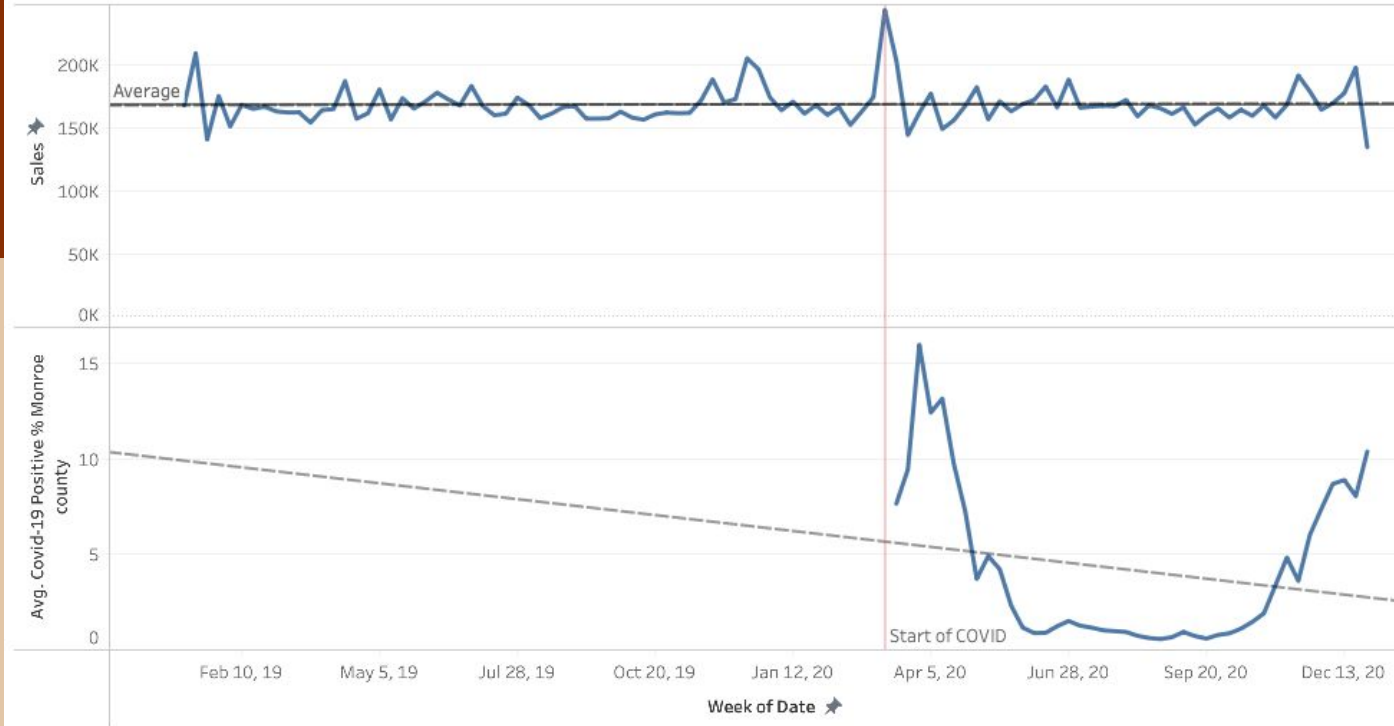
There is not much difference in who buys foods from cooking departments vs those who buys from the restuarant departments. Children, average age and % of sales are all the same. Though, there is an increase in average sales per customer when they purchase from restaurant departments. This makes a lot of since because all restaurant items are higher in price to account for labor and packaging costs.

Restaurant Stats By Income

Hh Income ₪	Avg. Hh Children	Hh Children	Avg. Hoh Age	Avg. Sales	% of Total Count of Wegmans_default_...	% of Total Sales along Table (Down)
15K-20K	0.54	2,143	53	4.8919	2.06%	2.13%
20K-30K	0.60	6,170	54	4.5670	5.36%	5.18%
30K-40K	0.88	9,616	55	4.7218	5.65%	5.64%
40K-50K	0.70	11,892	54	4.6685	8.77%	8.66%
50K-75K	0.92	46,077	54	4.6031	26.01%	25.32%
75K-100K	1.06	33,785	57	4.7160	16.48%	16.44%
100K-125K	1.00	16,171	57	5.0776	8.33%	8.95%
125K+	1.13	48,431	52	4.7974	22.29%	22.61%
<10K	0.50	3,278	55	4.4973	3.38%	3.21%
NA	0.00	0	47	5.2501	1.67%	1.85%

- general findings and hypothesis focused discoveries

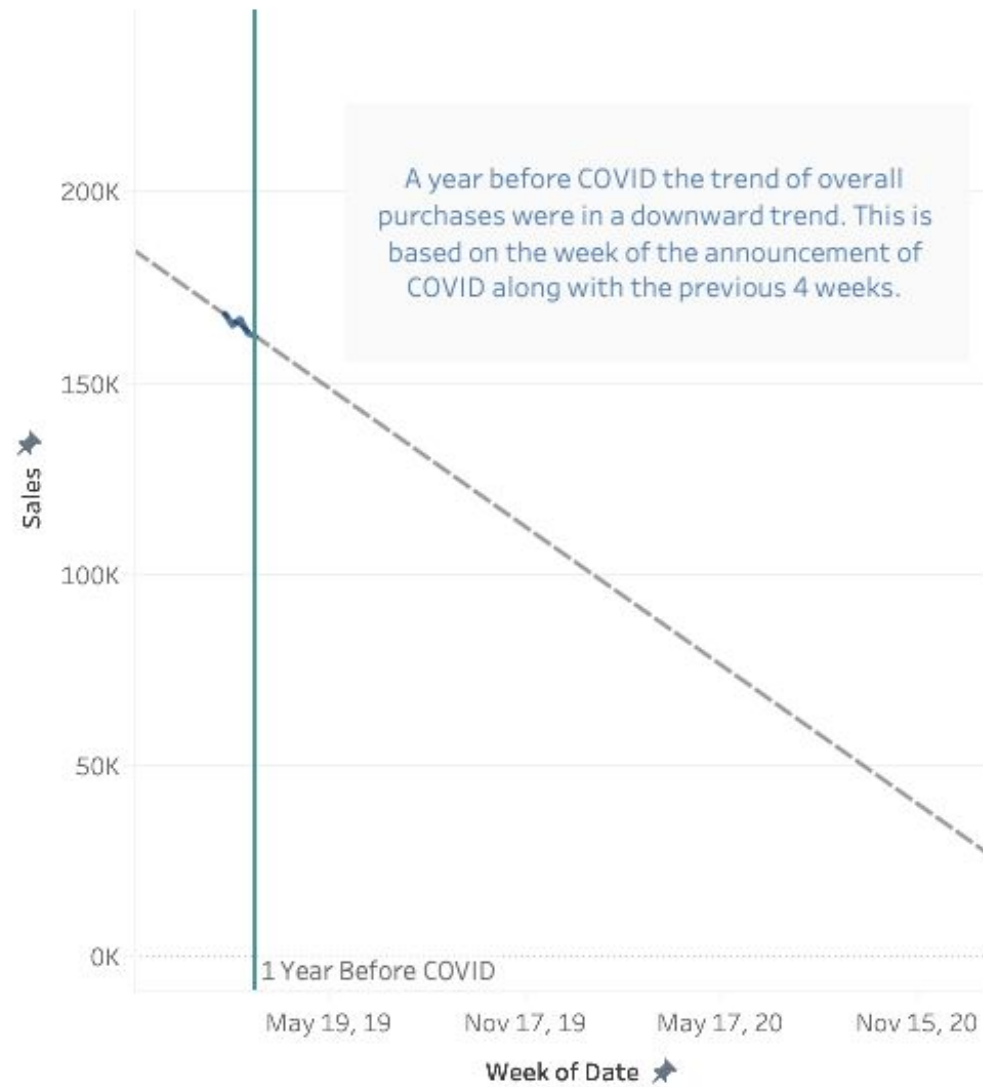
Sales In Relations to Covid 19 % of Cases



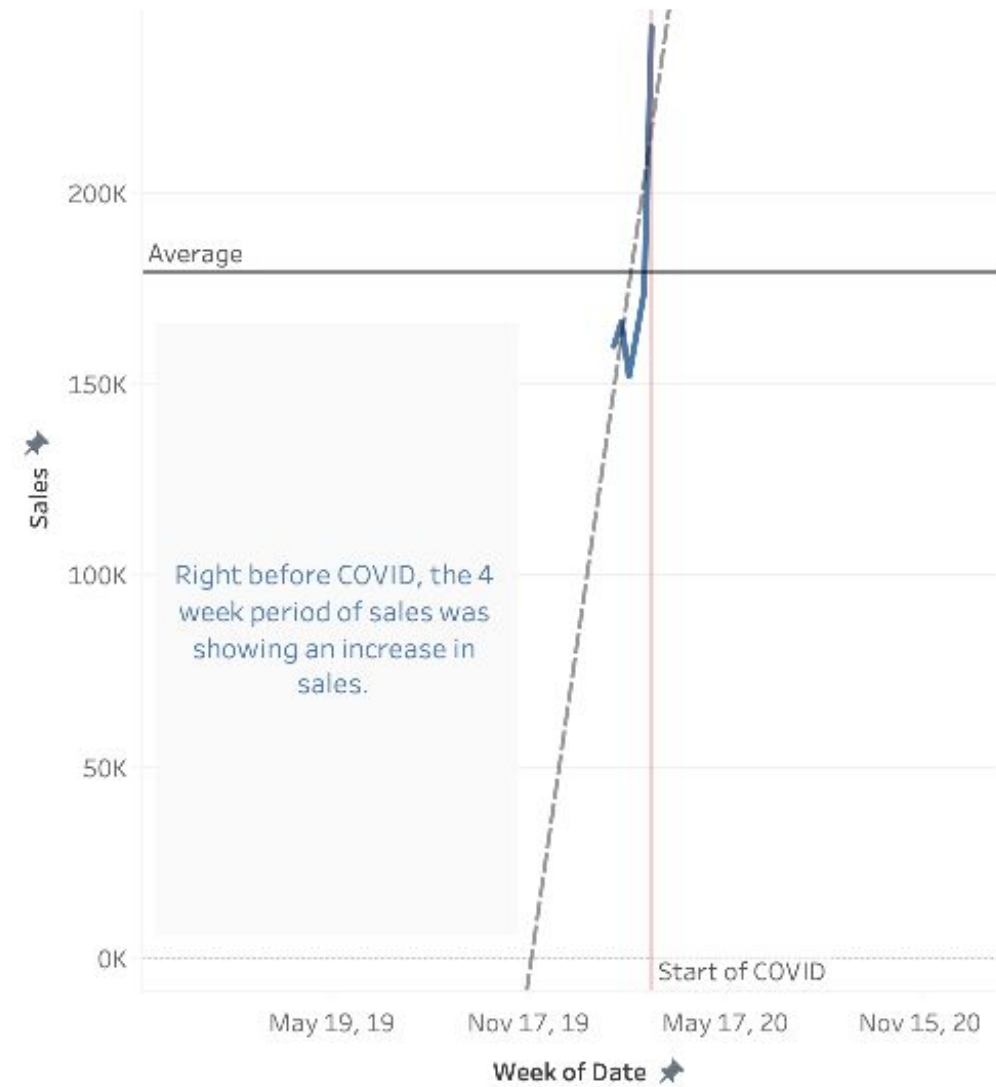
Instacart Sales



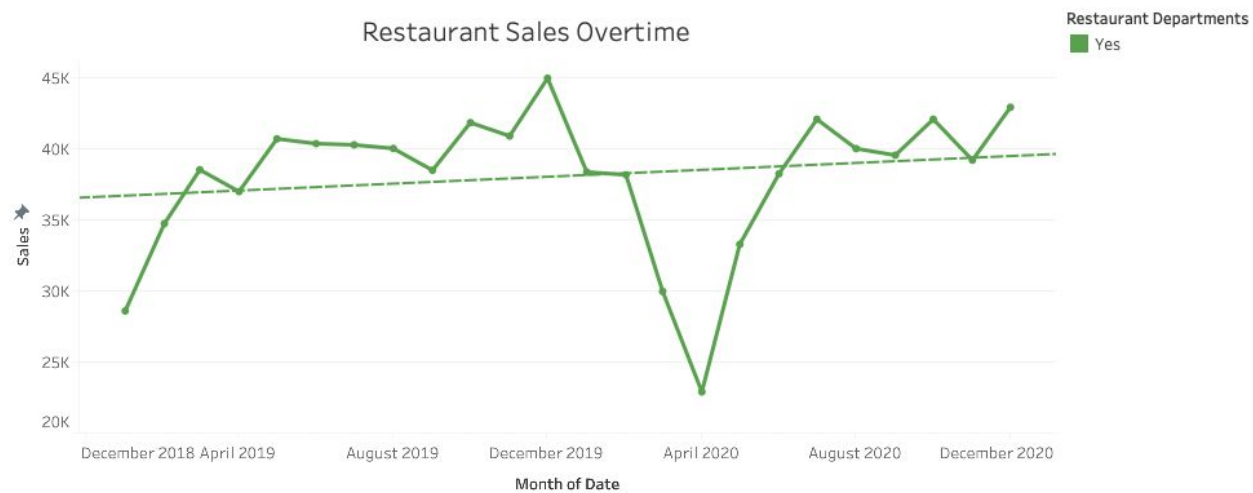
### Sales 2019 4 Weeks BC



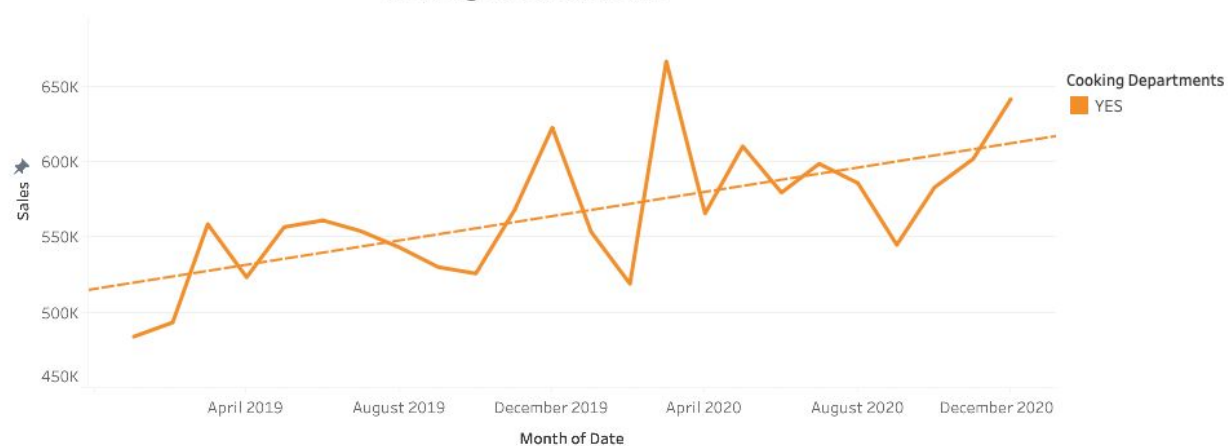
### Sales 2020 4 Weeks BC



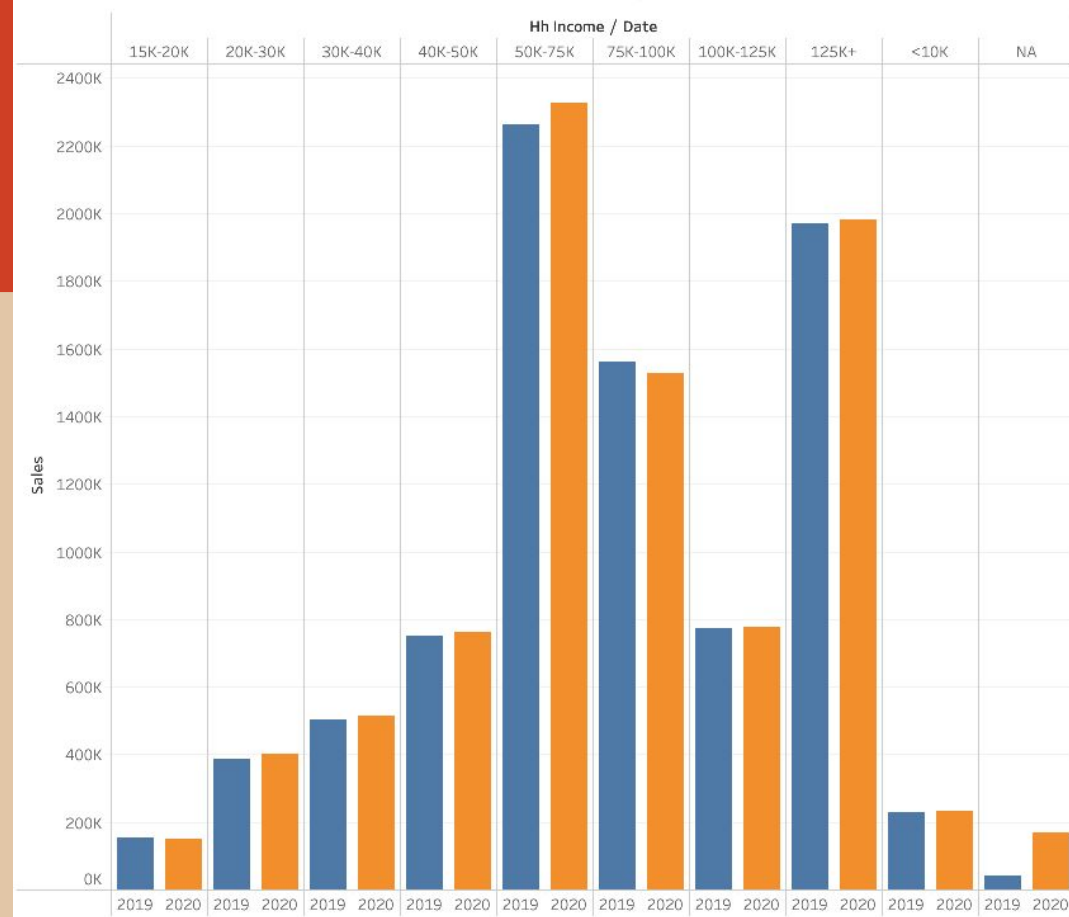
### Restaurant Sales Overtime



### Cooking Sales Overtime



### Sales by Income






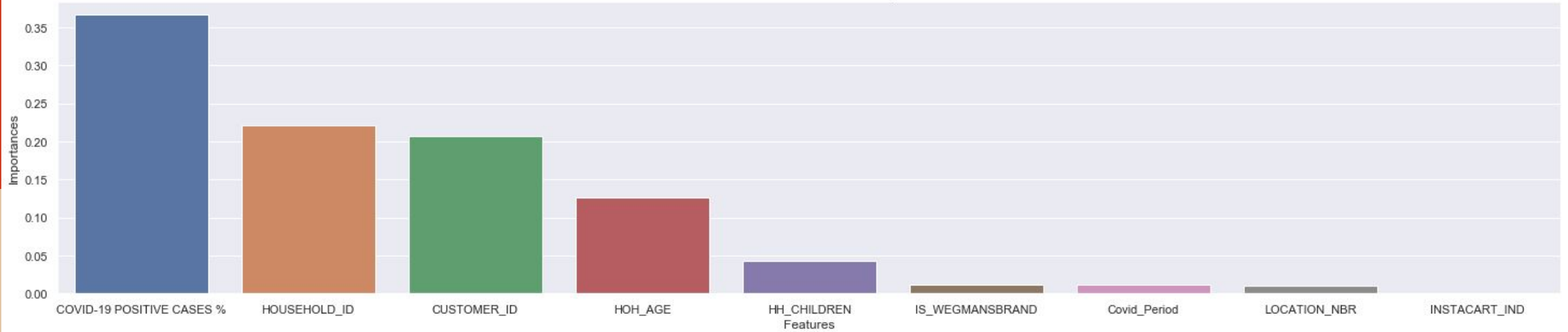
# 4

## Decision Tree Predictive Modeling

- Selected 9 features
- Feature importance is clear
- Relationships can be viewed easily
- Predicted continuous values of Wegmans Cooking Sales
- Generated daily forecasted cooking sales (2019-2020)

 Wegmans Sales Prediction.csv Date Key	#Wegmans Sales Prediction.csv Actual	#Wegmans Sales Prediction.csv Predicted
1/6/2019	5,742.23	5,868.40
1/7/2019	3,121.68	3,360.67
1/8/2019	2,516.96	2,639.72
1/9/2019	3,331.79	3,366.83
1/10/2019	2,093.75	2,097.22
1/11/2019	3,762.85	3,783.55
1/12/2019	5,287.08	5,386.66
1/13/2019	5,726.65	5,821.66
1/14/2019	3,158.37	3,211.95
1/15/2019	2,805.03	2,717.87
1/16/2019	3,740.51	3,821.75
1/17/2019	5,293.94	5,333.61
1/18/2019	7,529.21	7,547.18
1/19/2019	4,659.48	4,707.00
1/20/2019	990.57	1,031.45

Decision Trees Feature Importances



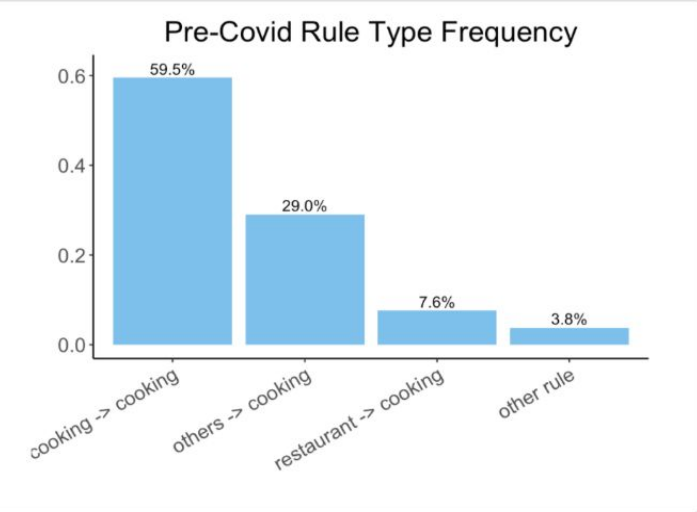
## Top 30 Sales By CustomerID

Customer Id	
3206.0	39,881
1305.0	39,660
3116.0	32,863
1546.0	31,316
3682.0	30,362
3309.0	30,148
220.0	29,092
3629.0	28,459
3014.0	28,216
318.0	28,062
3038.0	27,380
2072.0	26,692
6965.0	26,381
133.0	26,309
682.0	26,188
2463.0	25,581
4133.0	25,268
3053.0	25,187
2266.0	24,720
2098.0	24,717
55.0	24,644
658.0	24,620
2334.0	24,237
1130.0	24,048
1556.0	24,032
2352.0	23,999
401.0	23,833
4190.0	23,767
3595.0	23,586
2588.0	23,300

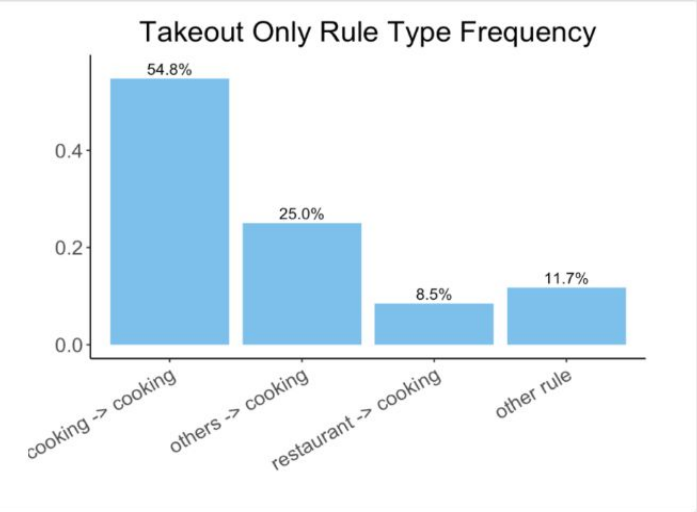
## Bottom 30 Sales By CustomerID

Customer Id	
2363.0	4.650
1586.0	4.420
4351.0	4.020
6095.0	4.000
1660.0	3.990
2561.0	3.950
2937.0	3.950
4101.0	3.930
6434.0	3.230
5806.0	3.220
996.0	3.090
6139.0	2.940
3892.0	2.880
757.0	2.870
2865.0	2.290
1577.0	2.280
0	2.150
6264.0	1.940
2202.0	1.480
1091.0	1.440
4718.0	1.430
5214.0	1.430
7026.0	1.420
2452.0	1.040
4485.0	0.710
4852.0	0.000
4909.0	0.000
5505.0	0.000
5758.0	0.000
6578.0	0.000

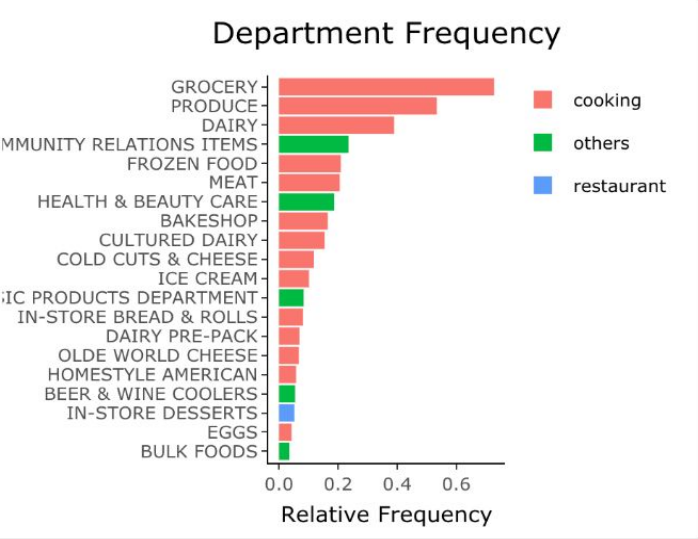
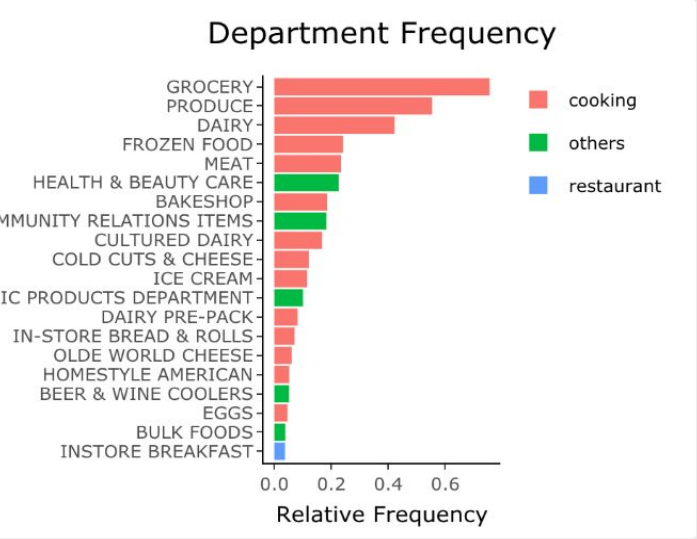
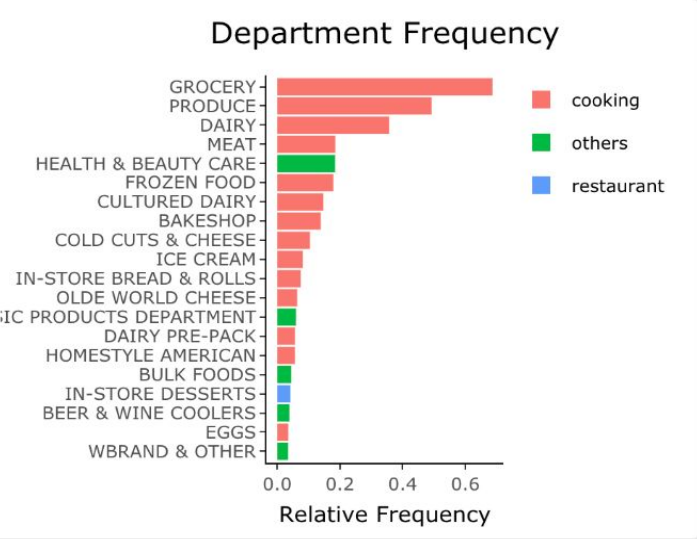
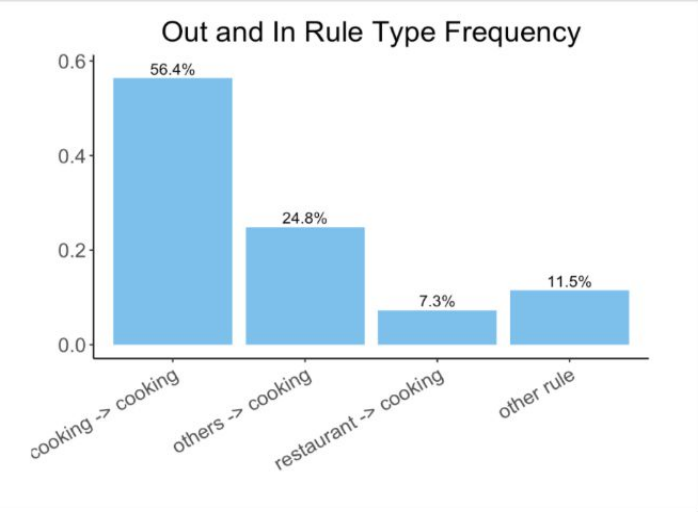
Pre-Covid (1/6/2019 to 2/28/2020)



Takeout Only (3/1/2020/ to 6/3/2020)



Indoor and Outdoor Dining (6/4/2020 to 1/2/2021)




# Time Series Predictive Models - Cont.

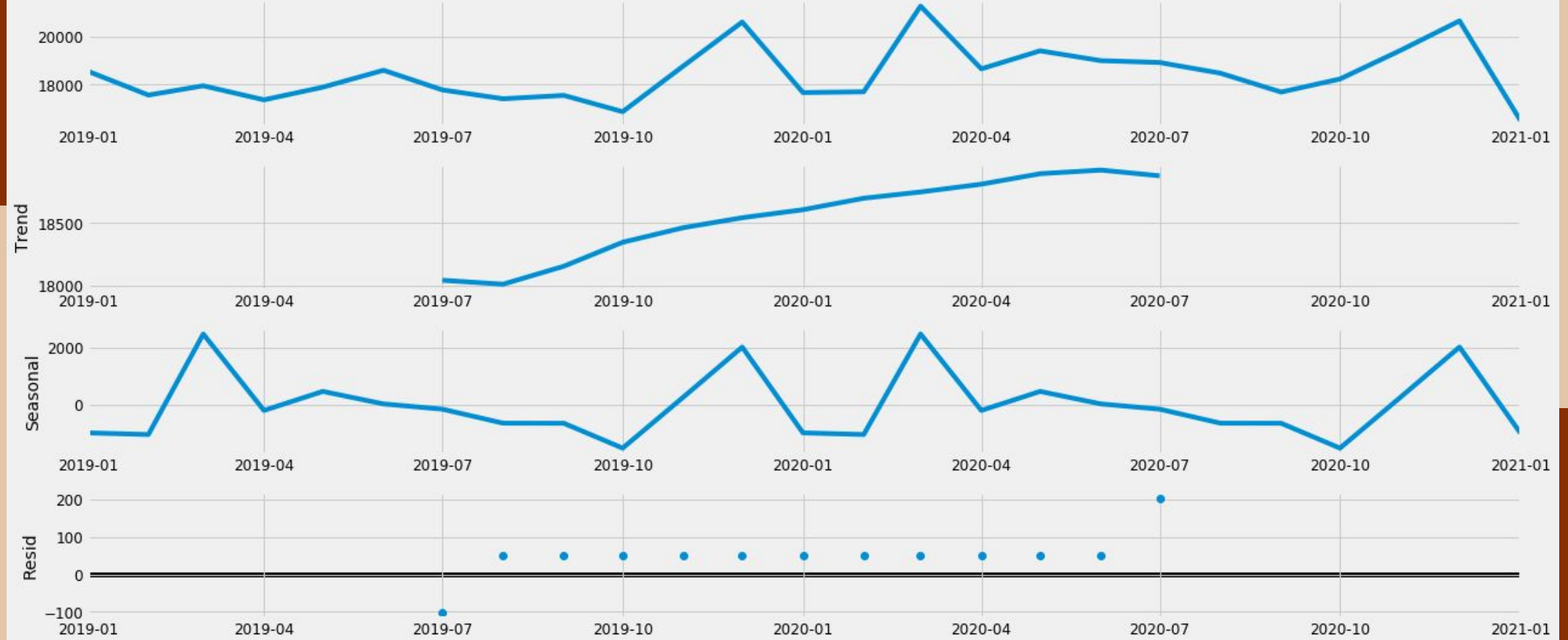
## Autoregressive integrated moving average (ARIMA)

- use of lagged moving averages to smooth time series data
- implicitly assume that the future will resemble the past
- predicted the seasonal trends
- generate forecasting sales for 10 years (2019 to 2029)

6

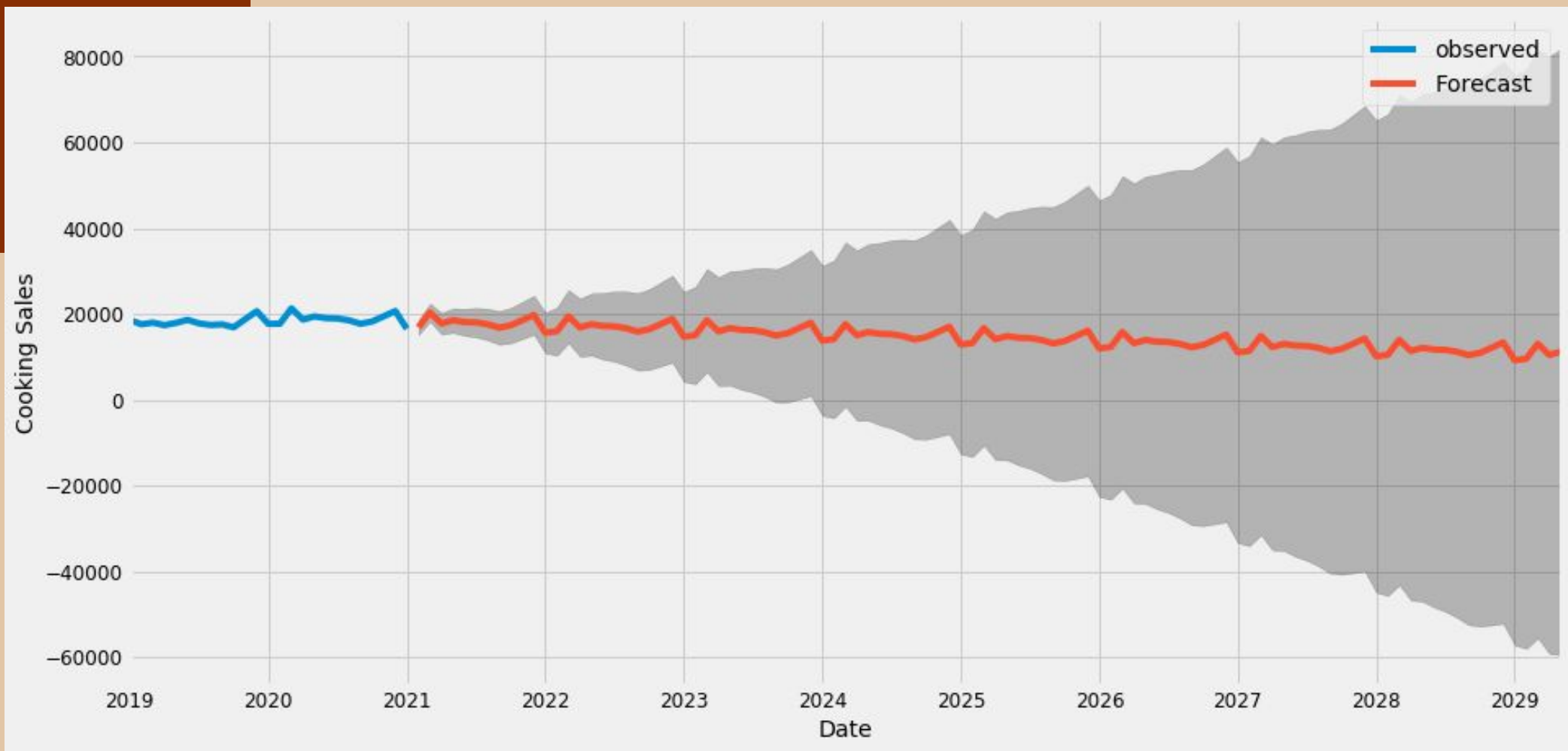
 Wegmans Cooking Sales ARIMA Model Truth.csv	# Wegmans Cooking ...	# Wegmans Co...
Date Key	Forecaste	Truth
2/1/2020	19,553.59	17,702.90
3/1/2020	17,806.58	21,276.34
4/1/2020	19,888.61	18,659.82
5/1/2020	19,707.01	19,411.37
6/1/2020	20,056.57	19,001.93
7/1/2020	18,463.81	18,924.29
8/1/2020	18,365.03	18,477.92
9/1/2020	18,639.67	17,691.48
10/1/2020	17,245.94	18,235.63
11/1/2020	19,832.91	19,434.90
12/1/2020	21,447.51	20,660.38
1/1/2021	17,870.03	16,515.61

# SALES







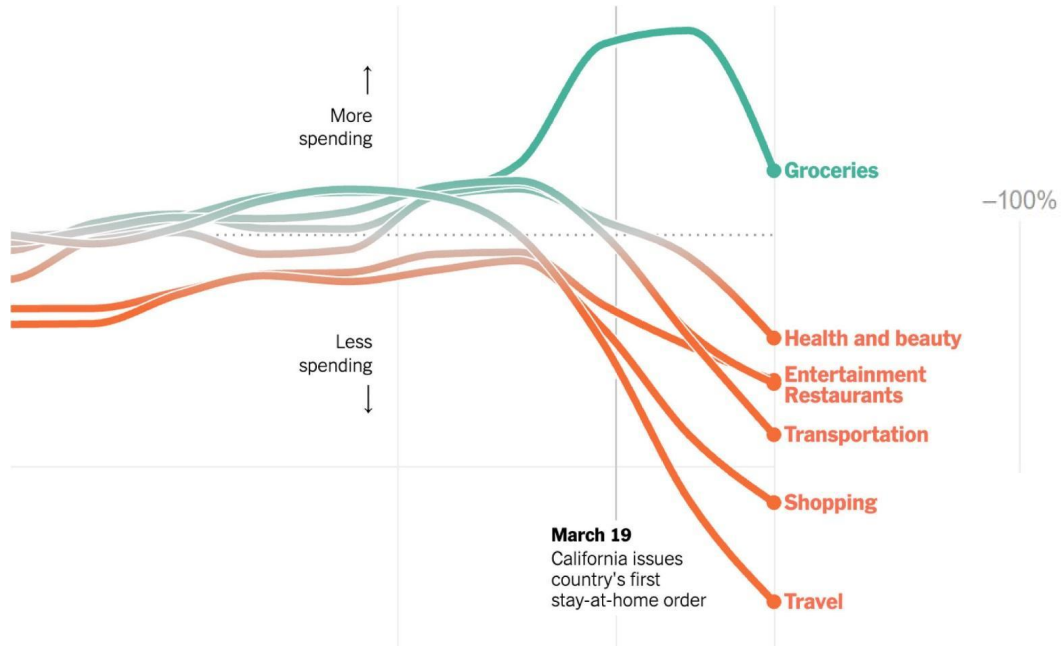




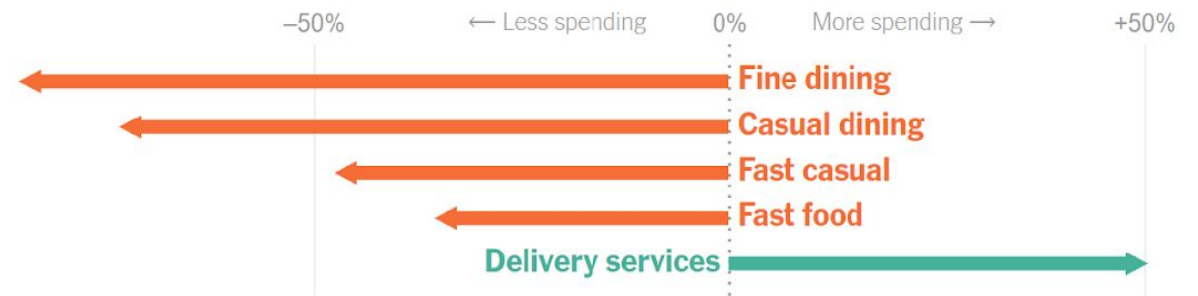
# Future Scopes .....

- Customer lifetime value analysis to determine the values and churn of customers
- Cluster analysis to achieve more effective customer marketing
- Set rules/threshold in decision tree - E.g., If a customer spends more than \$100.
- Sentiment analysis to hear out what people said about wegmans brands on social media

# Conclusions

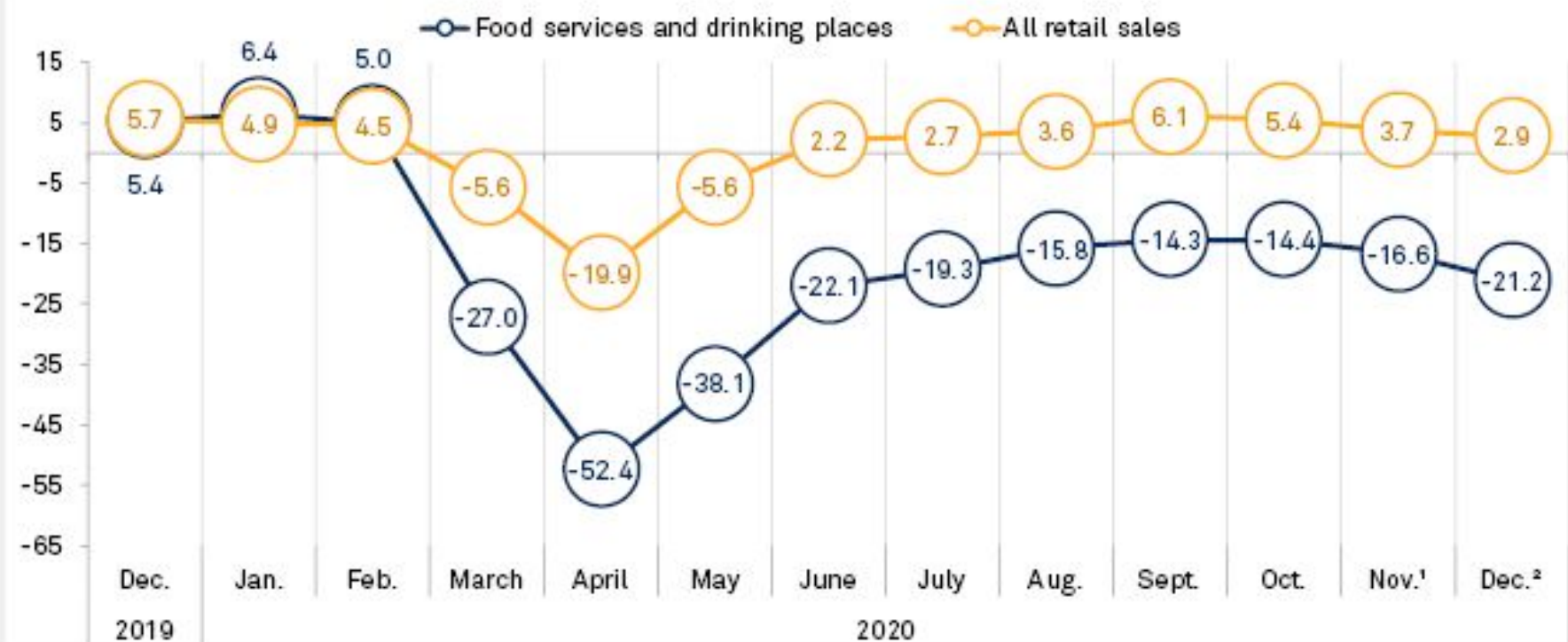


## Restaurant sales have plummeted.



Change in spending from 2019 for the week ending April 1.

## YOY change in sales for food services, drinking places (%)



Data compiled Jan. 15, 2021.

Based on U.S. government data released Jan. 15, 2021.

Data includes seasonally adjusted estimates of monthly sales.

<sup>1</sup> Represents a preliminary estimate.

<sup>2</sup> Represents an advance estimate.

Source: U.S. Census Bureau

Questions ?