MarginEdge Pulse: Predicting Restaurant Sales

- Bo Davis: bo@marginedge.com
- Julie Davis: juliedavis80@gmail.com
- Yabo Gao: yabo.nancy.gao@gmail.com
- Chelsea Weiss: chelsea.l.weiss@gmail.com

About MarginEdge

- MarginEdge is a software company that provides services to help restaurants manage their operations
- We collect purchase invoices (via mobile app) which lets us see everything a restaurant purchases
- We integrate with labor tools to track staffing costs
- We integrate with Point of Sales systems to track everything the restaurant sells
- We provide management tools for inventory, recipes, ordering, etc. to help the restaurant manage operations
- Servicing over 2,000 restaurants and growing rapidly!

Problem Statement

- Restaurants struggle with variable costs and variable revenue
- Labor and Food generally represent 60% of costs, but both must be scheduled/purchased in advance
- Revenue is highly variable
- Determining how many people to schedule and food to buy requires knowing sales – and there is no good way to predict future sales
- Industry tends to use moving averages which are simply not accurate

Data Collection: Original Dataset

- MarginEdge's daily restaurant sales by category since 2011

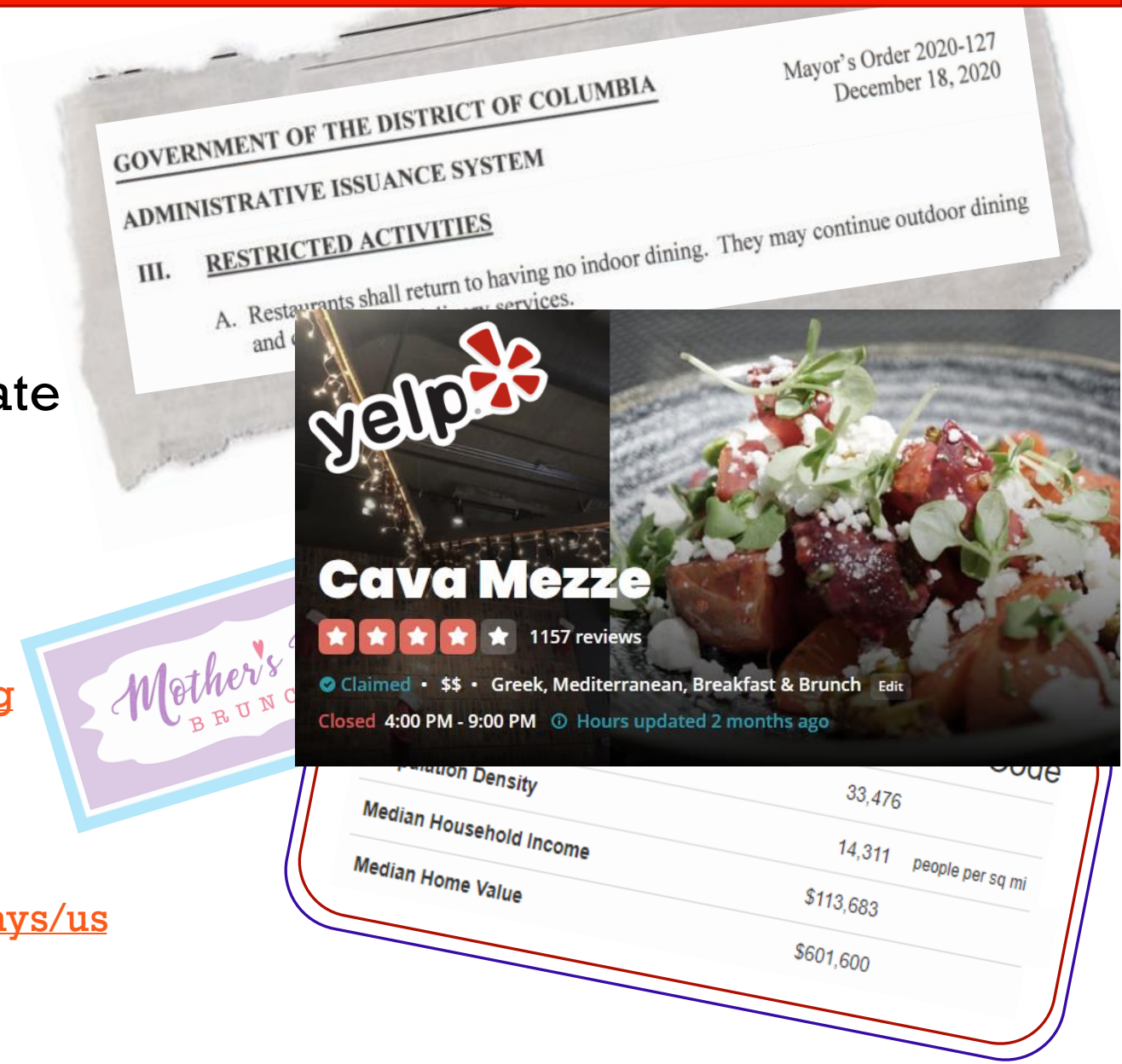
Features		
• Restaurant	• Date	• Zip Code
• Full Service Flag	• POS System	• Total Sales

Sales Subcategories		
• Food	• Beer	• Retail
• Liquor	• Nonalcoholic Beverage	• Other
• Wine		

- Potential obstacles? Varying effects from…
  - Covid-19
  - Seasonality
  - Assumption that restaurants sold to their demand
  - Holidays
  - Economy
  - Price Range
  - Cuisine Type

Data Collection: Additional Data Sources

- Restaurant reviews and price ranges
  - Source: https://www.yelp.com
- Covid-19 restaurant restrictions by state
  - Source: https://www.huschblackwell.com
- Demographics by zip code
  - Source: https://www.unitedstateszipcodes.org
- U.S. Holidays
  - Source: https://www.timeanddate.com/holidays/us



Data Preparation

Final Dataset for Training and Testing

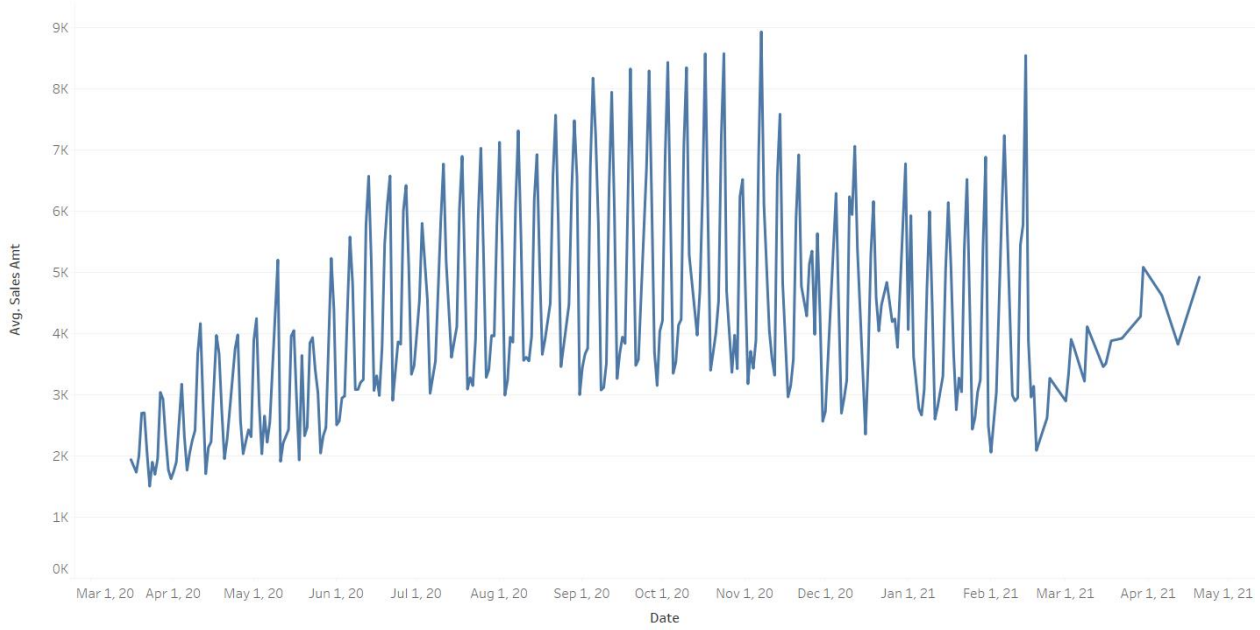
	Observations (n)	
Description of Change to Dataset	**Change**	**Cumulative**
MarginEdge original dataset	385,177	385,177
Excluded: Restaurants where difference between total sales and subcategories is greater than $10,000	(17,389)	367,788
Combined: POS records by day and restaurant	(264)	367,524
Added: Missing days from 2016+ for restaurants with $0 sales *(presumed to be closed)*	30,482	398,006
Excluded: Transactions prior to 2016	(8,414)	389,592
Excluded: Transactions where MarginEdge did not have historical sales data for 56 weeks + prior to sale date	(133,621)	255,971
Excluded: Two restaurants with poor data quality	(2,593)	253,378
Ending: Observations included for modeling		**253,378**

- Observations: 253,378
- Features: 166 *(from 13)*
- Unique Restaurants: 353
- Locations:

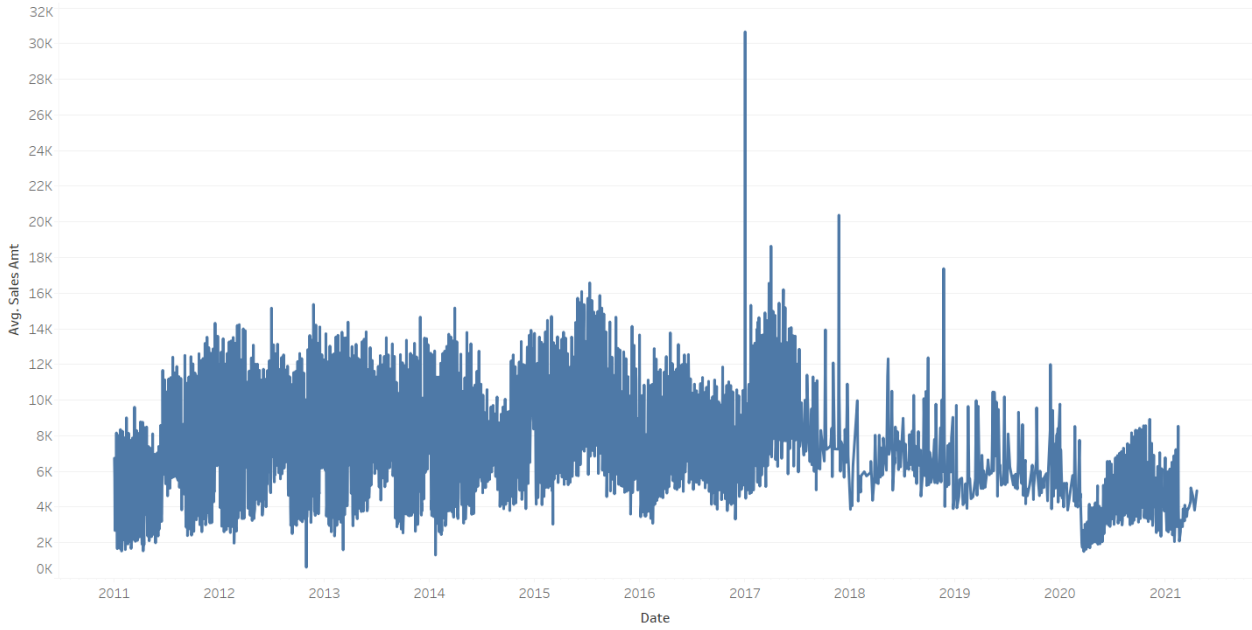
  VA, MD, DC
- Date Range:

  01/01/16 – 04/21/21

Plot of All Restaurants' average sales since pandemic over time

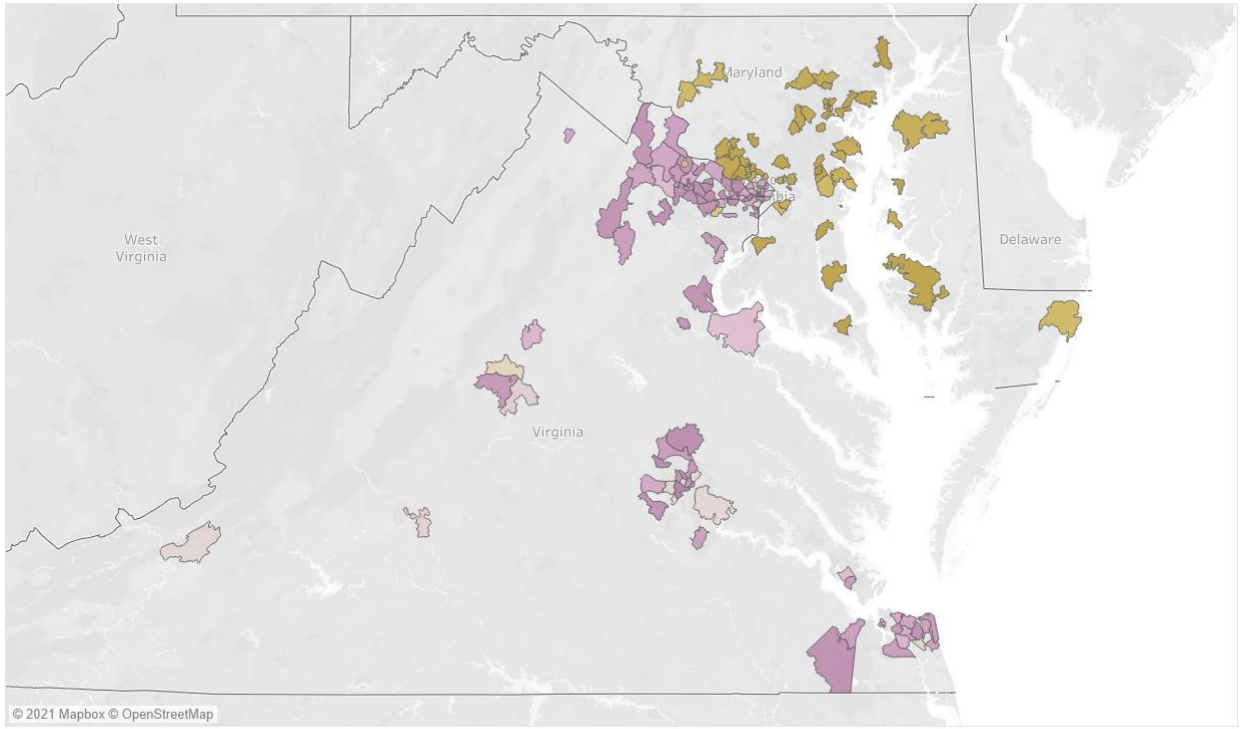


Cumulative Sales Over Time

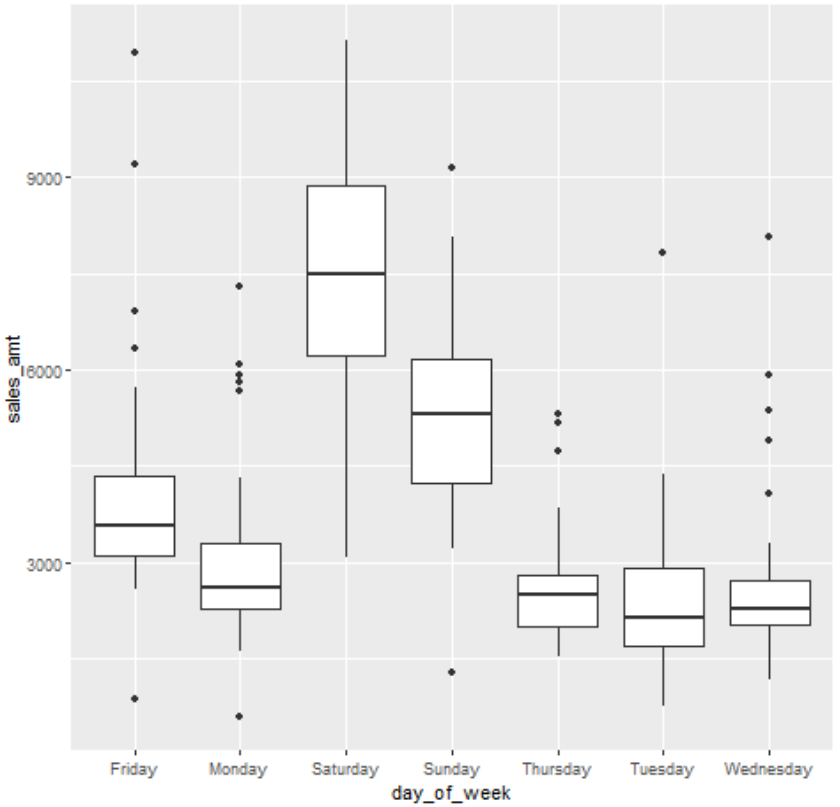


Exploratory Data Analysis: Change Over Time

Plot of Seating Capacity by Zip Code



Boxplot of Wasabi Tysons Sales by Day of Week



Exploratory Data Analysis: Categories

Metrics

- R^2
- Root Mean Squared Error
- Absolute Daily Variance as a Percentage of Actual Sales
- Absolute Daily Variance as a Percentage of Actual Sales with a 5% Tolerance in Error

Methods

- Simple Linear Regression (LR)
- Ordinary Least Squares Regression (OLSR)
- Weighted Simple Linear Regression (Weighted LR)
- Weighted Ordinary Least Squares Regression (Weighted OLSR)
- Long Short-Term Memory (LSTM)
- Stepwise Regression (SR)
- Multiple Linear Regression (MLR)
- Decision Tree Regression (DT)
- Autoregressive Integrated Moving Average (ARIMA)

Models: OLS and Linear Regression

Date	Actual	OLSR	LR	Weighted LR	Weighted OSLR
04/15/21	$ 2,449	$ 4,903	$ 5,391	$ 5,470	$ 7,086
04/16/21	$ 4,851	$ 4,830	$ 5,216	$ 5,475	$ 7,104
04/17/21	$ 9,986	$ 4,854	$ 5,274	$ 5,481	$ 7,122
04/18/21	$ 7,492	$ 4,879	$ 5,332	$ 5,486	$ 7,140
04/19/21	$ 3,211	$ 4,842	$ 5,245	$ 5,492	$ 7,158
04/20/21	$ 2,566	$ 4,866	$ 5,303	$ 5,497	$ 7,176
04/21/21	$ 3,109	$ 4,891	$ 5,362	$ 5,503	$ 7,194
Total	$ 33,663	$ 34,065	$ 37,123	$ 38,404	$ 49,982

Metric	OLSR	LR	Weighted LR	Weighted OSLR
R^2	**0.070**	0.070	0.063	0.063
Root Mean Squared Error	**7,346**	7,642	7,770	9,014
Ave Abs. Daily Variance %	**47.3%**	51.1%	52.8%	67.6%
Daily Var % w/ Tolerance	**43.0%**	46.1%	47.8%	62.6%

Models: LSTM

Date	Actual	Bidirectional LSTM
04/15/21	$ 2,449	$ 3,132
04/16/21	$ 4,851	$ 3,900
04/17/21	$ 9,986	$ 3,521
04/18/21	$ 7,492	$ 5,307
04/19/21	$ 3,211	$ 2,063
04/20/21	$ 2,566	$ 2,903
04/21/21	$ 3,109	$ 3,025
Total	$ 33,663	$ 23,852

Metric	Bidirectional LSTM
R-Squared	0.344
Root Mean Squared Error	6,539
Ave Abs. Daily Variance %	35.2%
Daily Var % w/ Tolerance	30.4%

Models: Stepwise Regression and Multiple Linear Regression

		Stepwise Regression (Forward Selection)				Multiple Linear Regression			
Training Data:	Restaurants	*353*	*1*	*353*	*1*	*353*	*1*	*353*	*1*
	Time frame	*2016+*	*2016+*	*Post Covid*	*Post Covid*	*2016+*	*2016+*	*Post Covid*	*Post Covid*
Date	Actual	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
04/15/21	$ 2,449	$ 4,386	$ 3,468	$ 4,410	$ 3,307	$ 4,666	$ 4,413	$ 4,824	$ 4,343
04/16/21	$ 4,851	$ 6,268	$ 5,355	$ 5,688	$ 5,525	$ 7,504	$ 6,494	$ 7,612	$ 6,564
04/17/21	$ 9,986	$ 7,249	$ 7,630	$ 6,228	$ 7,910	$ 9,594	$ 9,123	$ 9,544	$ 9,215
04/18/21	$ 7,492	$ 5,161	$ 5,152	$ 5,007	$ 5,023	$ 6,283	$ 6,353	$ 6,412	$ 6,306
04/19/21	$ 3,211	$ 3,421	$ 2,649	$ 3,836	$ 2,627	$ 3,954	$ 3,639	$ 4,218	$ 3,737
04/20/21	$ 2,566	$ 3,819	$ 2,610	$ 4,079	$ 2,241	$ 3,345	$ 3,635	$ 3,611	$ 3,301
04/21/21	$ 3,109	$ 4,218	$ 2,773	$ 4,420	$ 2,466	$ 3,588	$ 3,794	$ 3,817	$ 3,542
Total	$ 33,663	$ 34,524	$ 29,636	$ 33,669	$ 29,099	$ 38,935	$ 37,451	$ 40,038	$ 37,008

Metric	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
R-Squared	0.698	0.868	0.699	0.849	0.786	**0.883**	0.785	0.875
Root Mean Squared Error	7,500	**7,160**	7,353	7,163	8,117	7,910	8,200	7,892
Ave Abs. Daily Variance %	32.7%	**21.3%**	37.1%	22.7%	25.2%	23.1%	28.0%	21.6%
Daily Var % w/ Tolerance	27.7%	**16.5%**	32.1%	17.7%	20.5%	18.1%	23.1%	16.6%

Models: Stepwise Regression Formula Example

42 Variables Selected

Predicted Sales = -2257.32 + 1164.57*[Open Status] + 1482.73*[Seating Capacity Allowed]

Impact from Date

+ 62.72[quarter] + -308.45*[is Monday] + 688.47*[is Friday] + 2086.19*[is Saturday] + 1119.18*[is Sunday] + -10114.44*[is Christmas] + -2598.98*[is Thanksgiving] + 4221*[is Columbus Day] + -6902.09*[is Easter] + -1117.04*[is Father's Day] + 3346.3 *[is Good Friday] + -1916.94*[is Halloween] + 5057.46*[is Labor Day] + 5887.42*[is Martin Luther King Day] + 4967.12*[is Memorial Day] + 4370.91*[is President's Day] + 1617.52*[is Valentine's] + 2074.08*[is Veteran's Day]

Impact from Historical Sales and Seating Capacity Allowed

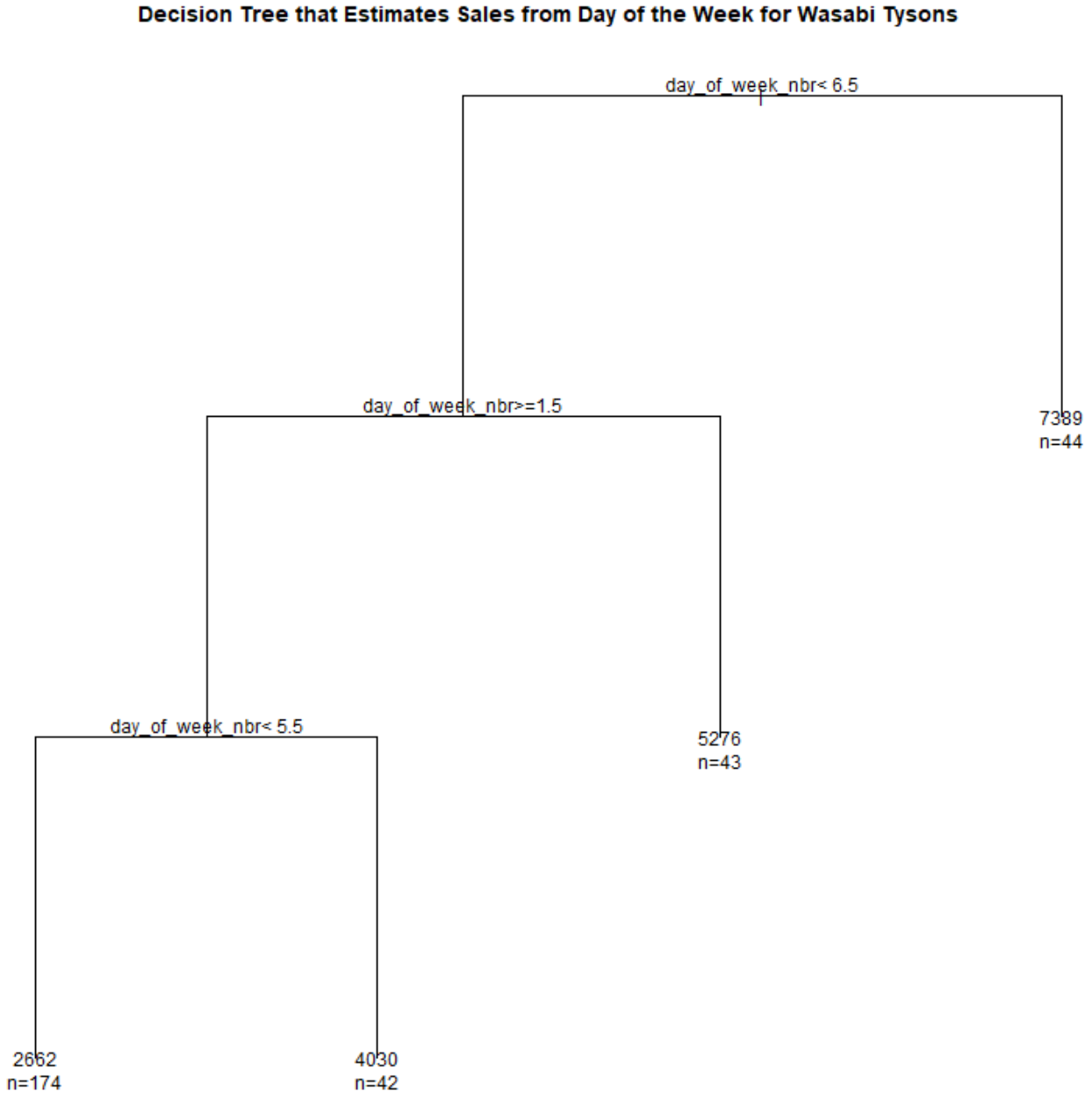
+ 0.34*[Ave. Sales in Past 7 Days] + 0.39*[Ave. Sales on Same Past 2 Weekdays] + 0.47*[Ave. Prior Year Sales a Week Before and After]+ -6551.74*[Ave. Seating Capacity Allowed a Week Before and After in Prior Year] + -0.16*[Ave.Prior Year Sales 2 Weeks Before] + -0.38*[Ave.Prior Year Sales 4 Weeks Before] + 5825.29*[Ave. Seating Capacity Allowed 4 Weeks Before in Prior Year] + -0.01*[Ave. Prior Year Sales 3 Similar Weekdays] + 0.29[Ave. Prior Year Sales on Same Weekday]

Models: Decision Trees

Training Data:		*Past 45 Days*	*Post Covid*	
Date	Actual	Day of the Week	Day of the Week	Mo. & Day of the Wk
04/15/21	$ 2,449	$ 2,662	$ 2,662	$ 2,469
04/16/21	$ 4,851	$ 4,030	$ 4,030	$ 4,030
04/17/21	$ 9,986	$ 7,389	$ 7,389	$ 8,840
04/18/21	$ 7,492	$ 5,276	$ 5,276	$ 5,276
04/19/21	$ 3,211	$ 2,662	$ 2,662	$ 2,469
04/20/21	$ 2,566	$ 2,662	$ 2,662	$ 2,469
04/21/21	$ 3,109	$ 2,662	$ 2,662	$ 2,469
Total	$ 33,663	$ 27,342	$ 27,342	$ 28,022

Metric	Day of the Week	Day of the Week	Mo. & Day of the Wk
R-Squared	**0.983**	**0.983**	0.956
Root Mean Squared Error	**6,958**	**6,958**	7,157
Ave Abs. Daily Variance %	20.6%	20.6%	**16.9%**
Daily Var % w/ Tolerance	15.7%	15.7%	**12.3%**

Sketch of Decision Tree



Models: Arima

Training Data:		2016 +	Past 6 Weeks	Past 6 Weeks	Past 5 Weeks excl. Spring Break
Date	Actual	Holt-Winters	Holt-Winters	ARIMA (Std)	ARIMA (Std)
04/15/21	$ 2,449	$ 2,997	$ 3,628	$ 3,448	$ 2,881
04/16/21	$ 4,851	$ 4,996	$ 6,721	$ 6,038	$ 5,389
04/17/21	$ 9,986	$ 9,569	$ 10,732	$ 8,887	$ 9,851
04/18/21	$ 7,492	$ 7,664	$ 8,064	$ 7,158	$ 7,561
04/19/21	$ 3,211	$ 3,643	$ 3,617	$ 3,350	$ 3,111
04/20/21	$ 2,566	$ 2,574	$ 2,957	$ 2,982	$ 2,615
04/21/21	$ 3,109	$ 3,108	$ 3,581	$ 3,744	$ 2,902
Total	$ 33,663	$ 34,552	$ 39,300	$ 35,607	$ 34,310

Metric	Holt-Winters	Holt-Winters	ARIMA (Std)	ARIMA (Std)
R-Squared	**0.992**	0.966	0.959	0.990
Root Mean Squared Error	7,799	8,325	**7,788**	7,823
Ave Abs. Daily Variance %	5.1%	16.7%	14.3%	**4.5%**
Daily Var % w/ Tolerance	**2.1%**	27.7%	20.5%	32.1%

Result: Best Model

Training Data:		2016 +
Date	Actual	Holt-Winters ARIMA
04/15/21	$ 2,449	$ 2,997
04/16/21	$ 4,851	$ 4,996
04/17/21	$ 9,986	$ 9,569
04/18/21	$ 7,492	$ 7,664
04/19/21	$ 3,211	$ 3,643
04/20/21	$ 2,566	$ 2,574
04/21/21	$ 3,109	$ 3,108
Total	$ 33,663	$ 34,552

Metric	
R-Squared	0.992
Root Mean Squared Error	7,799
Ave Abs. Daily Variance %	5.1%
Daily Var % w/ Tolerance	2.1%

Future Work

- Simultaneously predict sales for all restaurants
- Change parameters automatically based on individual's need (over predict or under predict)
- Automatically monitor predictions across the entire client base
- Adjust to outliers and new datapoints automatically
- Remove outdated data points automatically

Product Enhancements

- The community of interest for this product is extremely large
- 600,000 restaurants in the United States with $850 Billion in Annual Sales
- Serve as the "bedrock" for future predictive software
- Help restaurants prevent future losses and generate continuous profits