Business, Economics and Financial Data

By: Derek Sweet, Emanuele Zangrando and Ivan Kulazhenkov

WHAT IS SCAREFEST?

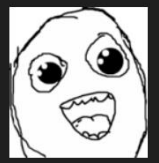
Halloween entertainment with scary actors offering 4 haunted attractions over 17 nights on weekends in September and October

Haunted Hayride

A narrated story suitable for all ages of the forest's recent outbreak by a guide who will be by your side until the very end

Scare Factor:

Family Friendly

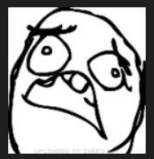


Haunted Maze

Good luck finding your way out as terror lurks around every turn and dead end in the Terror Zone Maze

Scare Factor:

Mid Level

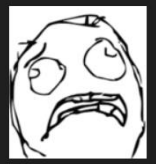


Forest Walk

Do you have what it takes to walk the wooded trail alone amidst the specters of death that surely wait for you?

Scare Factor:

Mid Level

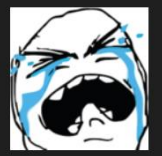


Haunted House

Try your luck and see if you can survive the terrors that wait just beyond the fence at the Castle of the Dead

Scare Factor:

Most Intense

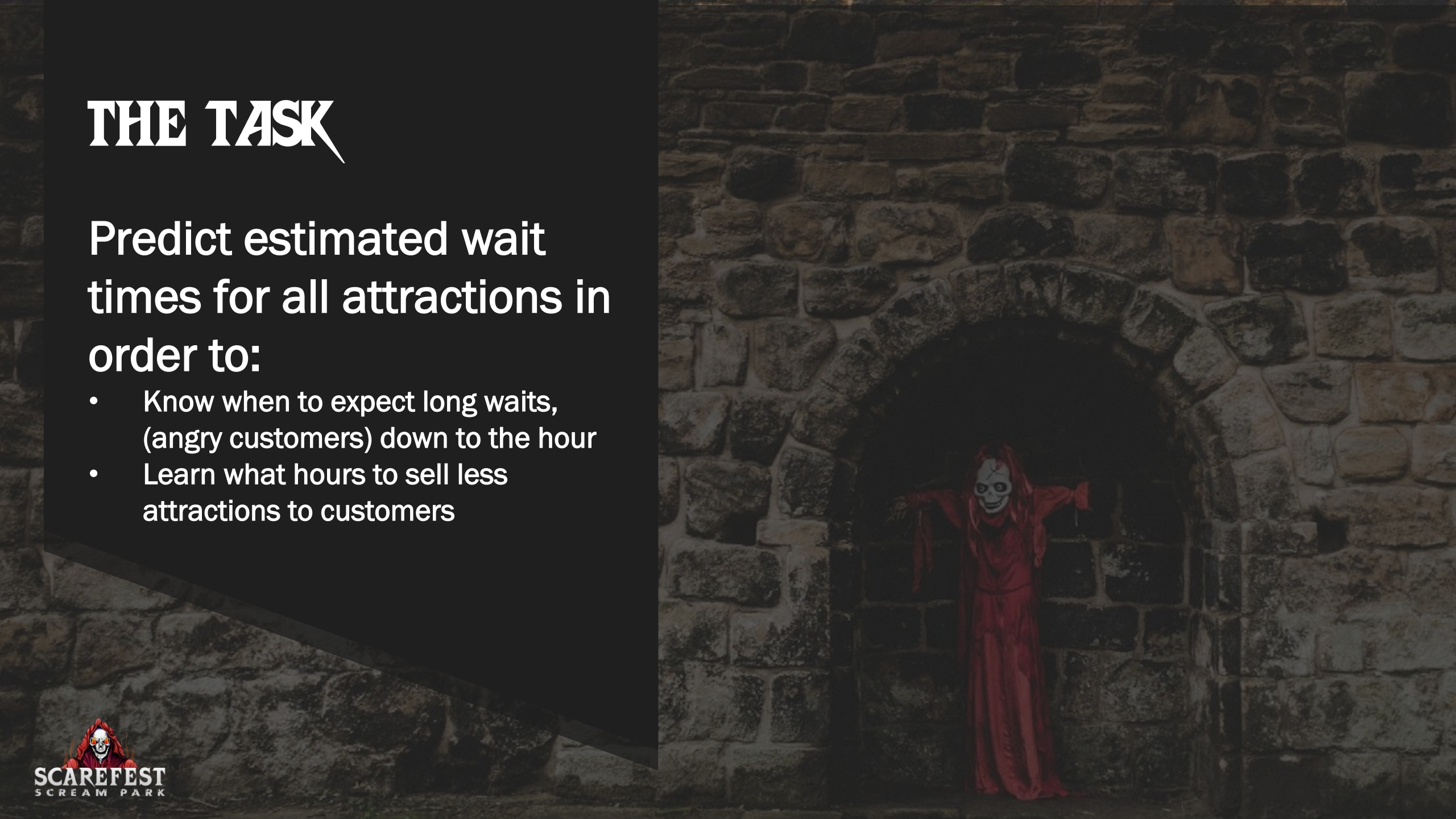


SCAREFEST
SCREAM PARK

THE TASK

Predict estimated wait times for all attractions in order to:

- Know when to expect long waits, (angry customers) down to the hour
- Learn what hours to sell less attractions to customers



SCAREFEST
SCREAM PARK

SOURCES OF DATA (2020-21)

eventbrite

Online Ticket Quantity 2020
Pulled CSV Report

showclix

Online Ticket Quantity 2021
Pulled CSV Report

Waitwhile

Virtual Queue
Pulled CSV Report

REGISCARE

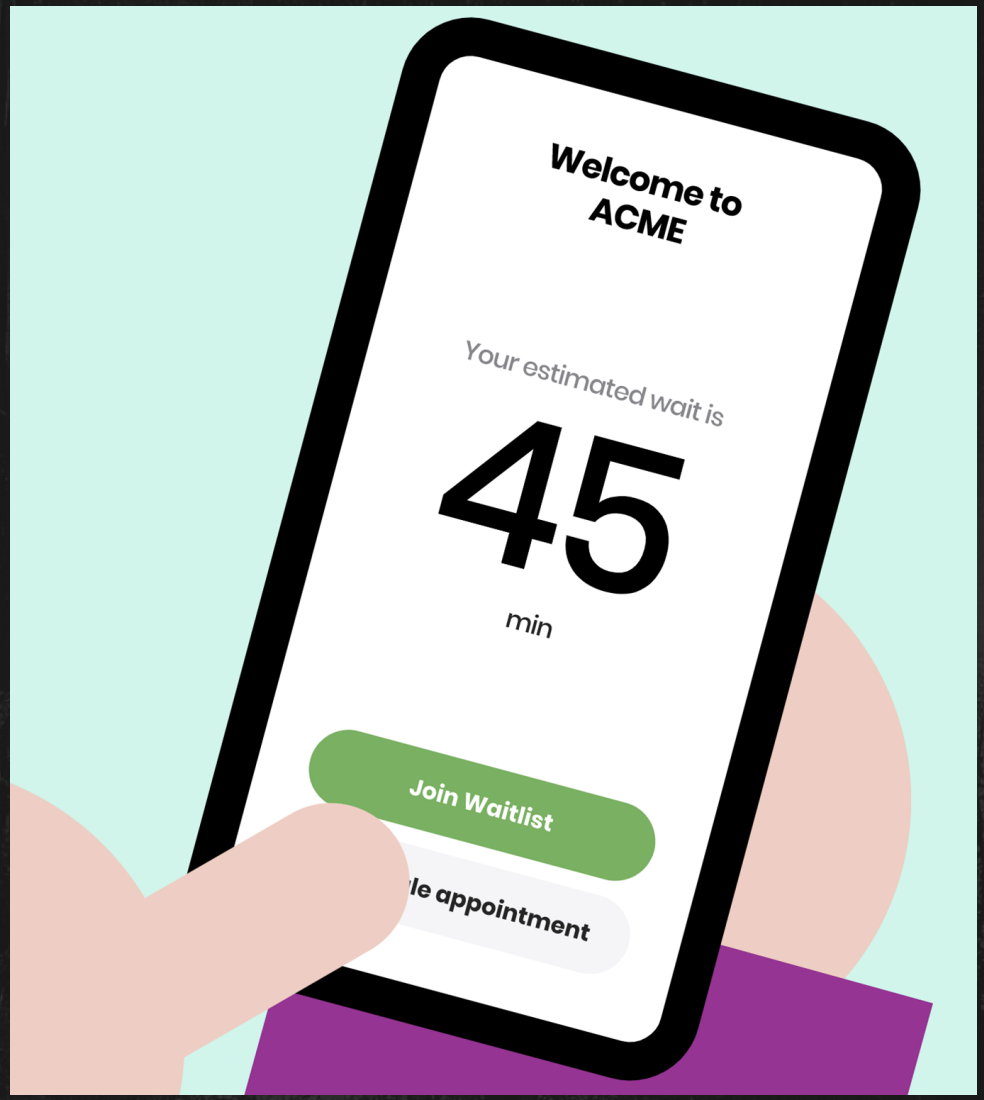
On-site Ticket Quantity
Pulled CSV Reports

The Weather Channel

Rain and Temperature
Manually Collected

SCAREFEST
SCREAM PARK

WHAT IS A VIRTUAL QUEUE?



SCAREFEST
SCREAM PARK

INPUT AND OUTPUT DATA

- People In Line (1 week lag)
- Average Wait Per Step (1 week lag)
- Ticket Count (1 week lag)
- Online Sales Day Before
- Hour of Day
- Hours Open
- Week Number
- Is October (boolean)
- Online Only (boolean)
- Rain (boolean)
- Bad Days (boolean)

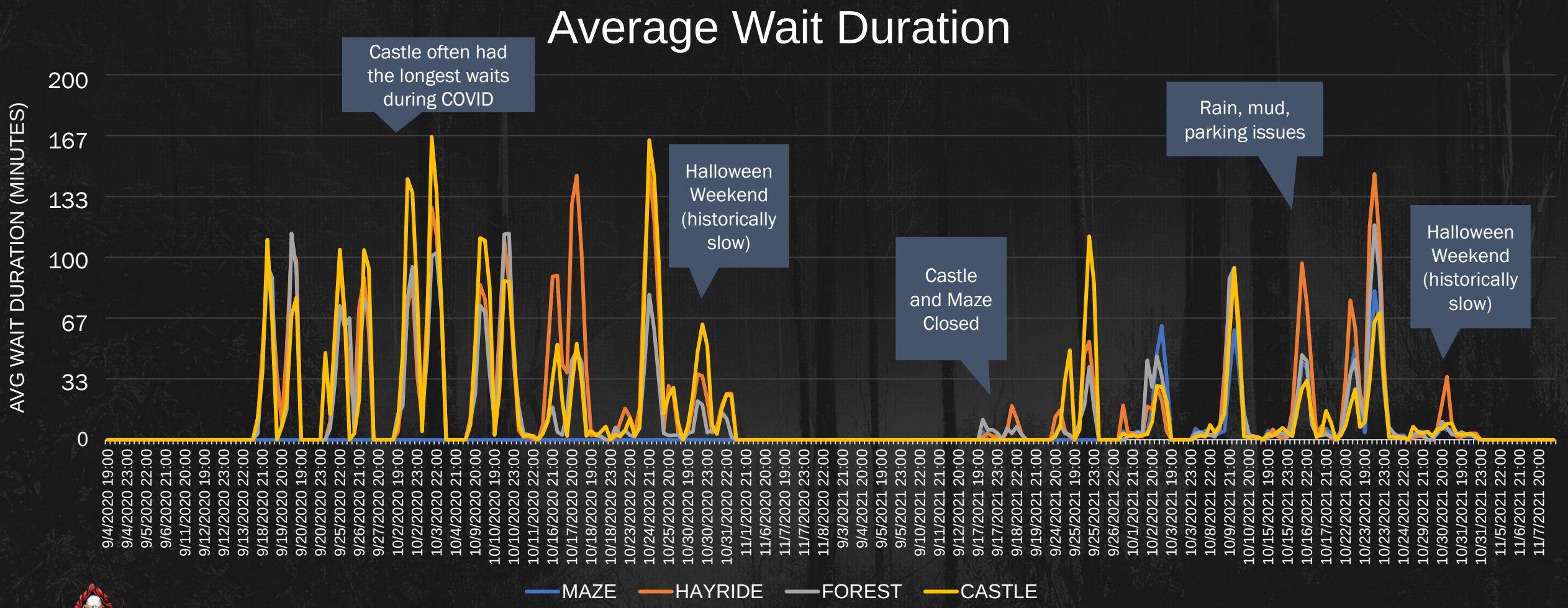
Average Wait Duration

Hourly Level
Non-Uniform
N=300 with padding

SCAREFEST
SCREAM PARK

GETTING TO KNOW SCAREFEST WAIT TIMES

AVERAGE WAIT DURATION TIME SERIES

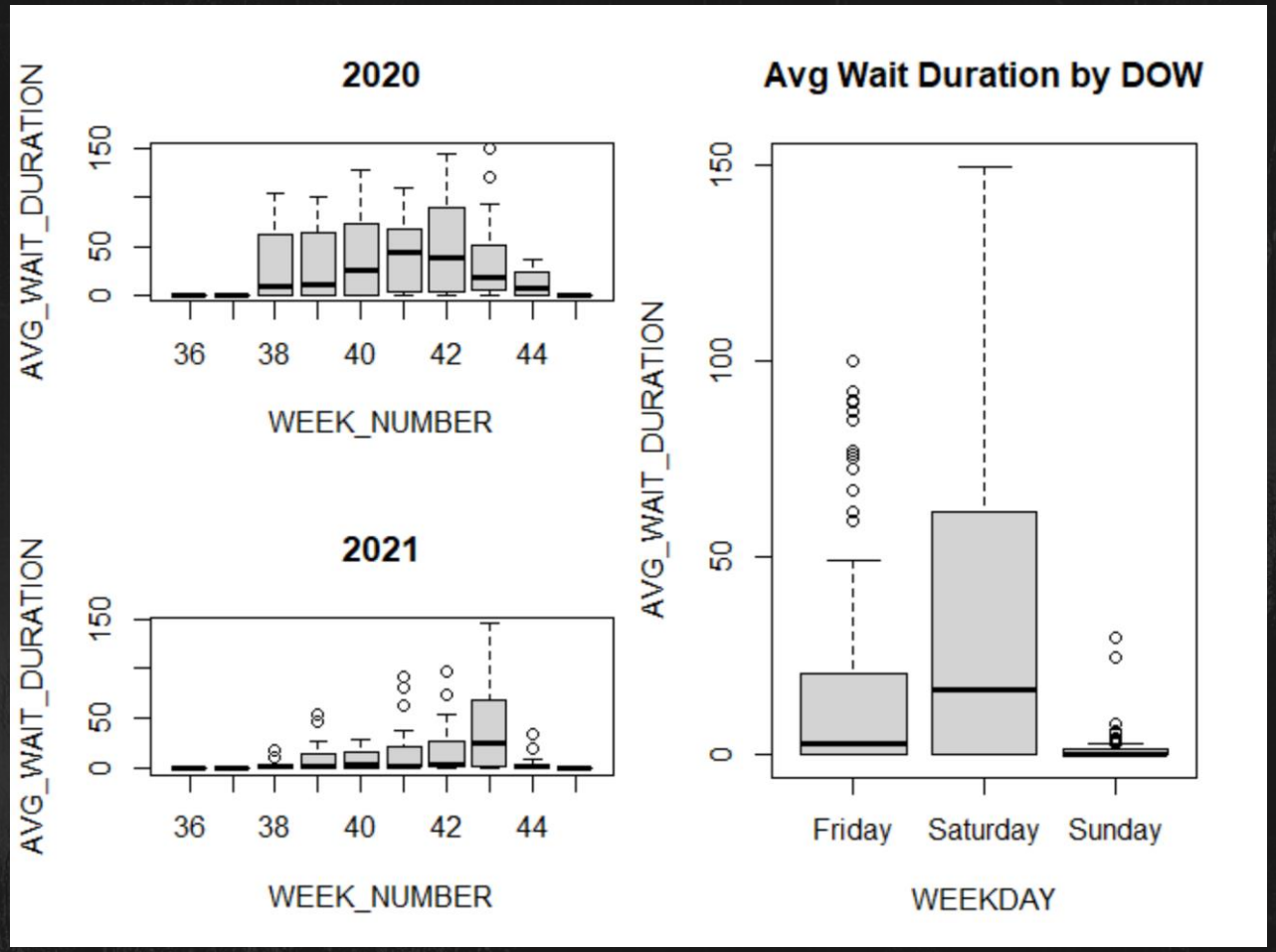


SCAREFEST SCREAM PARK

* Time series axis is the transformed timeseries for uniformity. It includes padded dates and only open days/hours

WAIT TIME BY WEEK AND YEAR

- Wait times were higher and more variable in 2020 due to COVID
- September was unprecedentedly busy in 2020 due to COVID
- Week 43 of 2021 was the largest weekend every in Scarefest's 16 years of operations
- Saturdays are the busiest while Sundays have reduced hours because of low demand

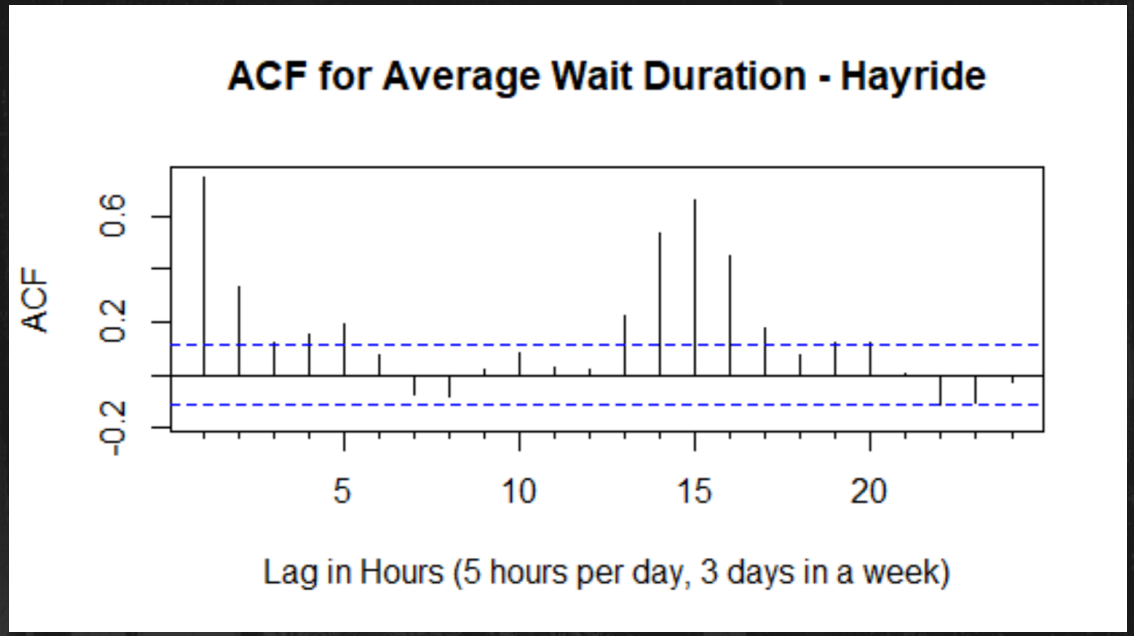


SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

Wait Time Comparison by Month

- Wait times for the next hour are highly correlated with the current hours wait times
- Wait times within a single day are slightly correlated (≤5 hours)
- Wait times of the same hours but one week lagged are highly correlated

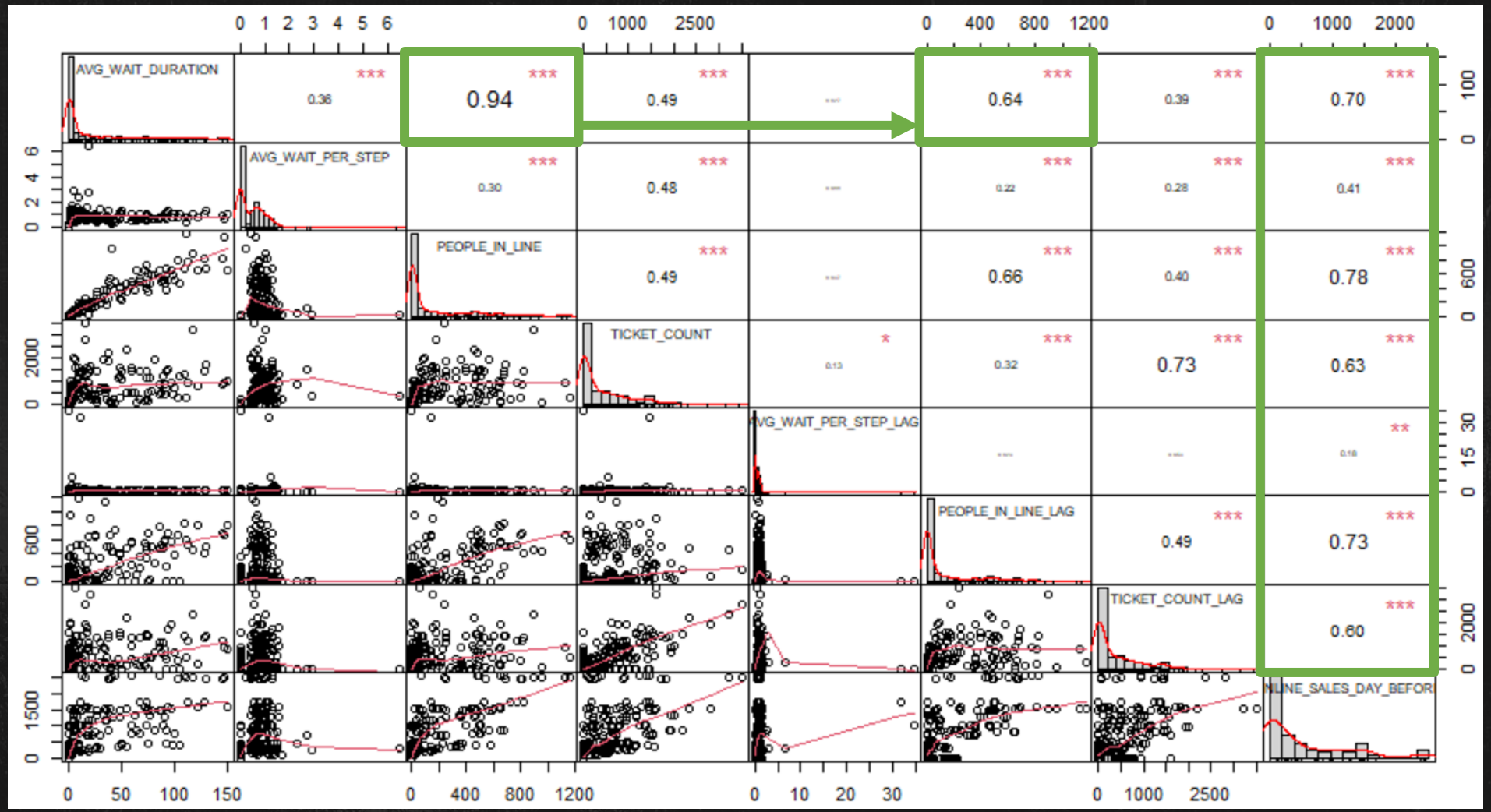


SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

CORRELATION PLOT

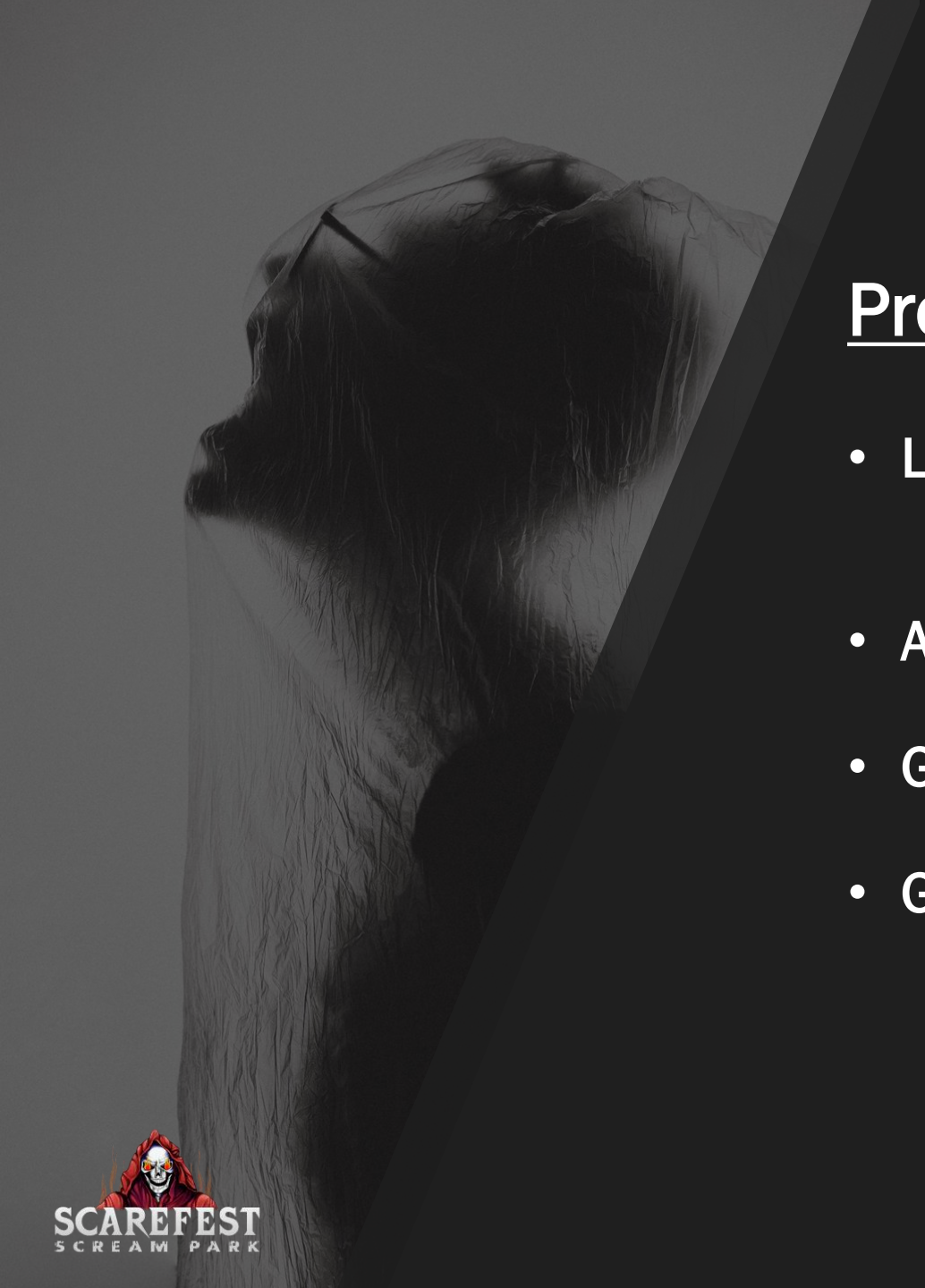
- Intuitively, people in line is highly correlated with the wait duration
- Wait per step isn't highly linearly correlated with wait duration, but seems to have a non-linear shape at low wait durations
- The 1 week lagged variables have less of a correlation to wait times however people in line lagged is still strong at 0.64
- Online sales day before is highly correlated with wait duration



SCAREFEST SCREAM PARK

HAYRIDE OF DOOM


Modelling wait times



COMPARING MODELS

Predicting wait times using:

- **Linear Regression**
  - Was used to find key features for comparison and possible inputs to the other three models
- **ARIMA**
  - Looking at ARIMA, ARIMAX, and SARIMA
- **Gradient Boosting**
  - Measuring the importance of features
- **GAM**
  - Testing spline, loess, quadratic, and interactions

SCAREFEST
SCREAM PARK

LINEAR REGRESSION

- People In Line and Online Sales Day Before are the most significant features in predicting wait times
- Hour 22 and 21 show up frequently as important variables to predict wait times
- Busy days are captured in Online Only and Hours Open 4
- Week numbers, hours and year are all significant however are only needed because the modes is not a time-series based model
- All models have a low DW and R-Squared

Coefficients and Significance				
Features	**Hayride**	**Castle**	**Forest**	**Maze**
PEOPLE_IN_LINE_LAG	0.04 ***	0.04 ***	0.04 ***	0.07 ***
ONLINE_SALES_DAY_BEFORE	0.03 ***	0.01 *	0.02 ***	0.01 ***
HOUR22	15.8 ***	20.56 ***	13.34 ***	
HOUR21	17.34 ***	15.26 ***	9.91 **	
ONLINE_ONLY	-13.77 **		-14.96 ***	-11.18 ***
HOURS_OPEN4	14.97 **	23.67 ***	10.96 *	
YEAR2021	-6.78 **	-8.31 **	-4.48 *	1.52
WEEK_NUMBER43	21.81 *	8.41	16.15 .	12.58 ***
WEEK_NUMBER40	19.76 .	23.75 .	26.07 **	5.13 *
HOURS_OPEN3	10.16	24.27 *	22.1 **	
HOUR23	5.83	13.09 **	6.26 .	
HOURS_OPEN5	7.85	23.83 *	13.05 .	
RAIN	-9.14 *	-7.85	0	3.15 .
IS_OCTOBER	-19.76	-23.75 .	-26.07 *	
HOUR20	7.08 *	3.12	1.54	
AVG_WAIT_PER_STEP_LAG	-0.89 *			
TICKET_COUNT_LAG		0.01 .	0 .	
WEEK_NUMBER41	15.17	8	15.93 .	3.3
WEEK_NUMBER42	18.05 .	-6.19	-2.61	2.01
WEEK_NUMBER38	2.21	-4.51	5.47	-1.02
WEEK_NUMBER39	-2.21	4.51	-5.47	-2.45
WEEK_NUMBER44	-6.6	-9.52	-4.81	-2.38
WEEK_NUMBER45	-1.98	-5.33	-3.43	-0.47
BAD_DAYS		-13.06		
DW	**1.12**	**0.85**	**1.18**	**0.75**
Adjusted R-Squared	**0.67**	**0.51**	**0.6**	**0.44**
AIC	**1,782.80**	**1,878.96**	**1,688.38**	**1,251.69**

SCAREFEST
SCREAM PARK


ARIMA

ARIMA

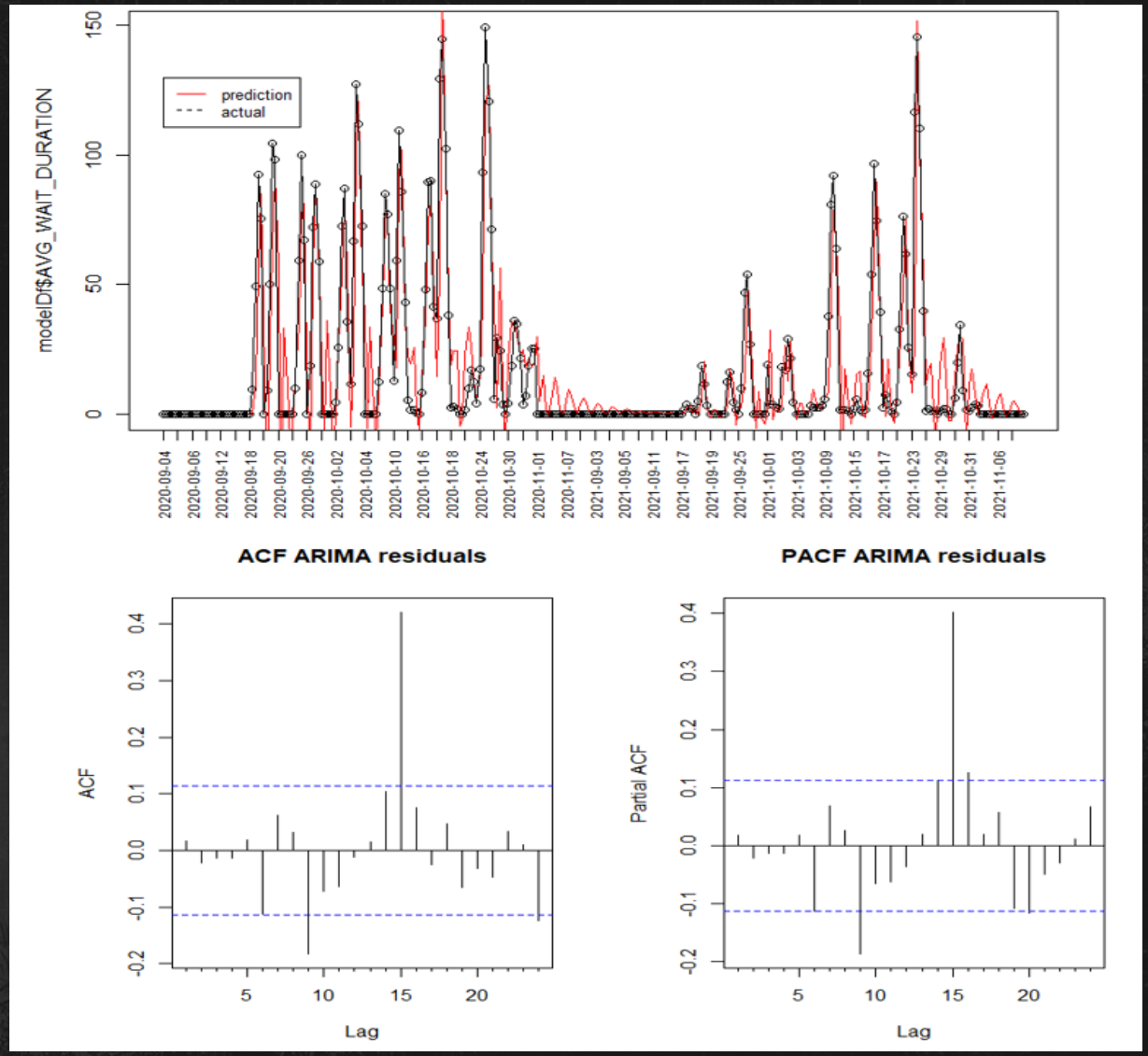
- We compared some handpicked ARIMA models to those suggested by auto.arima in R. Castle and Hayride were the hardest attractions to model.
- Regressors used in fitting the ARIMA are lagged by 1 week. These are People in Line, Avg Wait per Step and Ticket Count.
- Determining (p,d,q). We settle on a lag of 4, a differencing of 1 for stationarity and a moving average of 3.

ARIMA Model Comparison of Results (AIC)						
ARIMA(*p*,*d*,*q*)	Regressors Present	Hayride	Castle	Forest	Maze	Average
(4,1,3)	NO	2517.46	2603.18	2467.62	1794.28	2345.635
(3,1,3)	NO	2543.19	2608.45	2485.75	1792.43	2357.455
(0,1,5)	NO	2548.48	2598.6	2495.84	1793.51	2359.108
(0,1,2)	NO	2648.44	2655.79	2553.73	1901.66	2439.905
(4,1,3)	YES	2469.83	2551.26	2446.53	1766.92	2308.635
(3,1,3)	YES	2479.69	2545.72	2448.01	1767.48	2310.225
(0,1,5)	YES	2483.59	2551.81	2456.13	1766.3	2314.458
(0,1,2)	YES	2545.34	2591.41	2482.71	1879.91	2374.843

SCAREFEST
SCREAM PARK

ARIMA(4,1,3) – HAYRIDE RESIDUALS

- The ARIMA model deals very well with the low peaks of the first and final weeks in 2021
- It does however suffer when predicting high peaks in general, though this is not an easy task for any model.



SCAREFEST SCREAM PARK

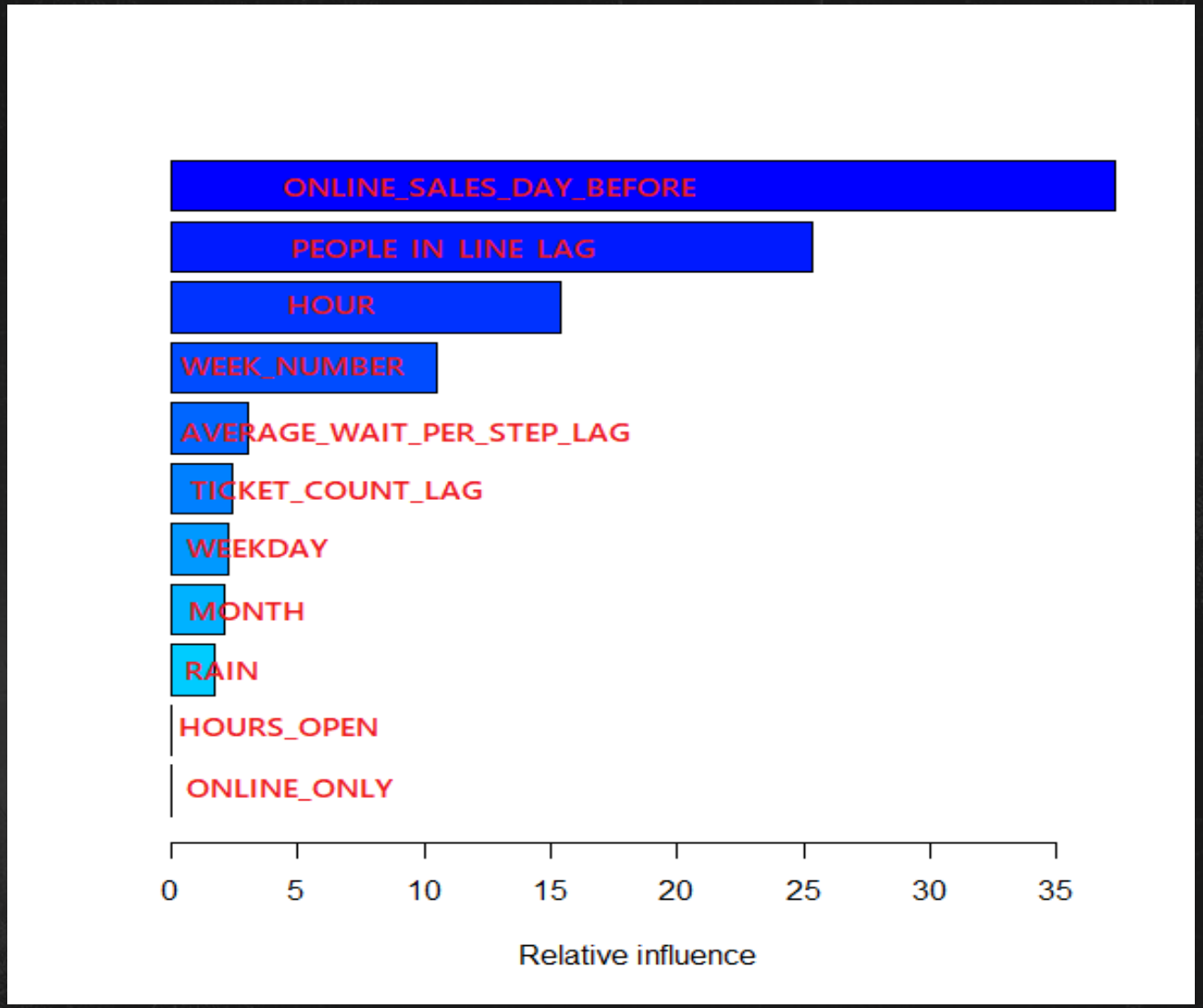
HAYRIDE OF DOOM

GRADIENT BOOSTING

GRADIENT BOOSTING

- Number of Trees = 100
- Shrinkage = 0.1
- Depth = 1

- Results of the most important features align with the Linear Regression
  - Online sales the day before and people in line are the most important predictors
  - Hour and week of the year are the next most important

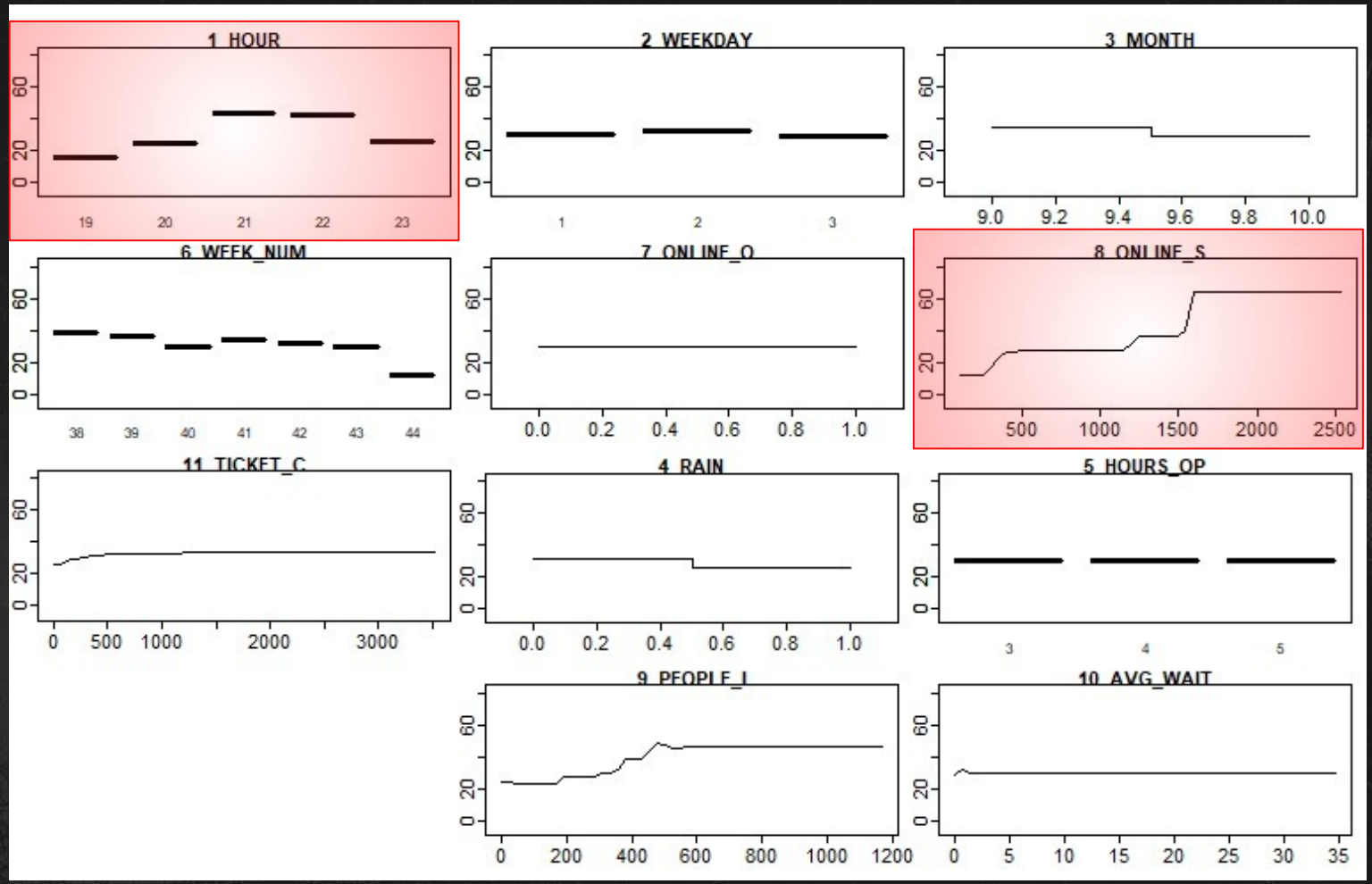


SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

GRADIENT BOOSTING - PARTIAL EFFECTS

- Wait times increase as online sales day before increases
- Online sales the day before is only available at a day level so it is setting an average where hour and other variables are creating the hourly variance

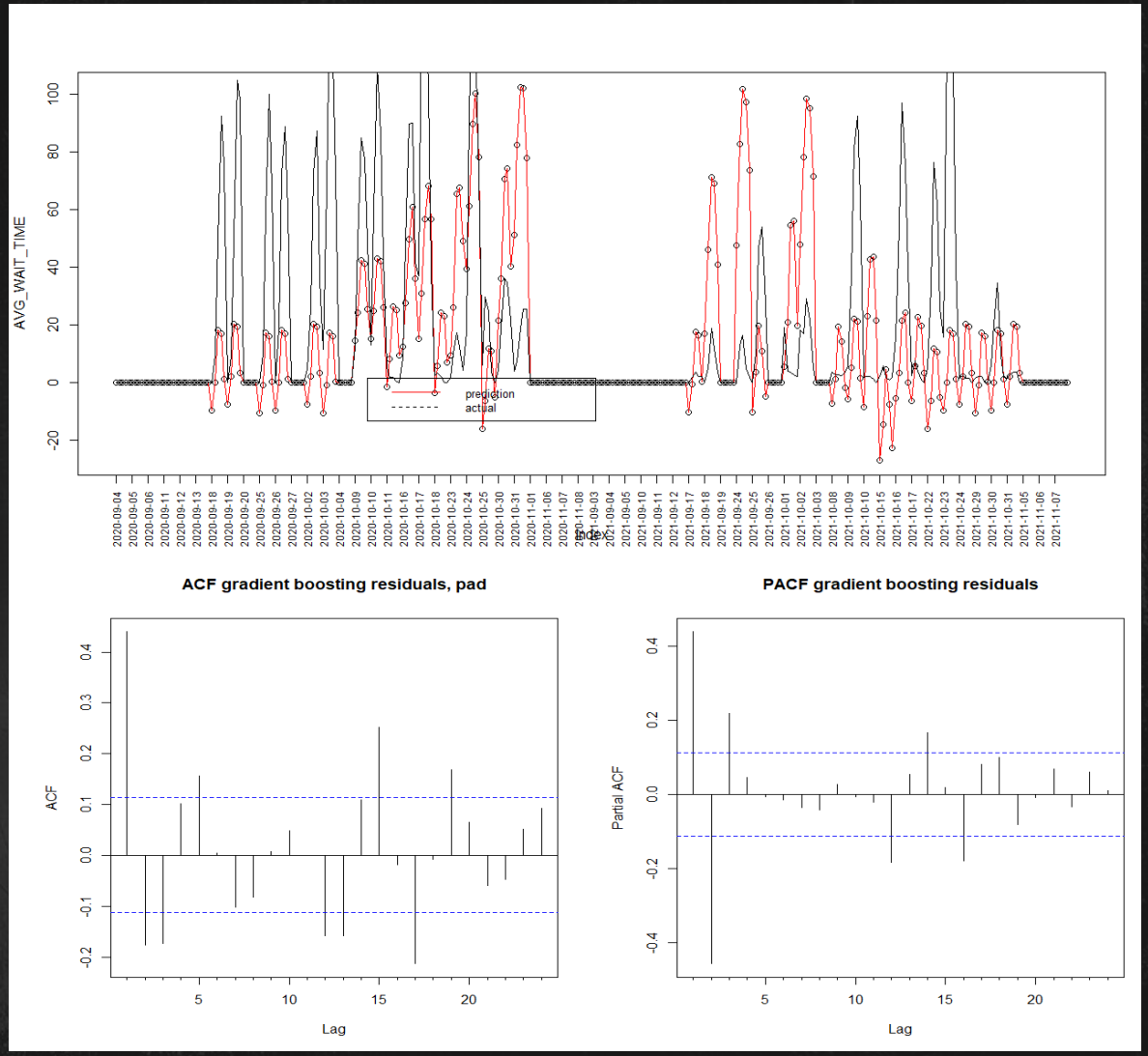


SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

GRADIENT BOOSTING VISUAL FIT

- GBM often predicts negative values
- The model underfits the beginning of 2020 but overfits the beginning of 2021 (likely due to strong September during COVID)
- The model has noisy predictions towards the end of the 2021



SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

Generalized Additive Models

GAM

- Smoothing splines gam
- Model selection done with step Gam;
- Hour, weekday, week number, number of online sales the day before and people in lag the day before are the most significant predictors;
- Durbin-Watson value of 1.11, with a p-value of 0.762, Aic = 1607.075
- Of all these variables, the organizers have only direct control on the number of sold tickets the day before

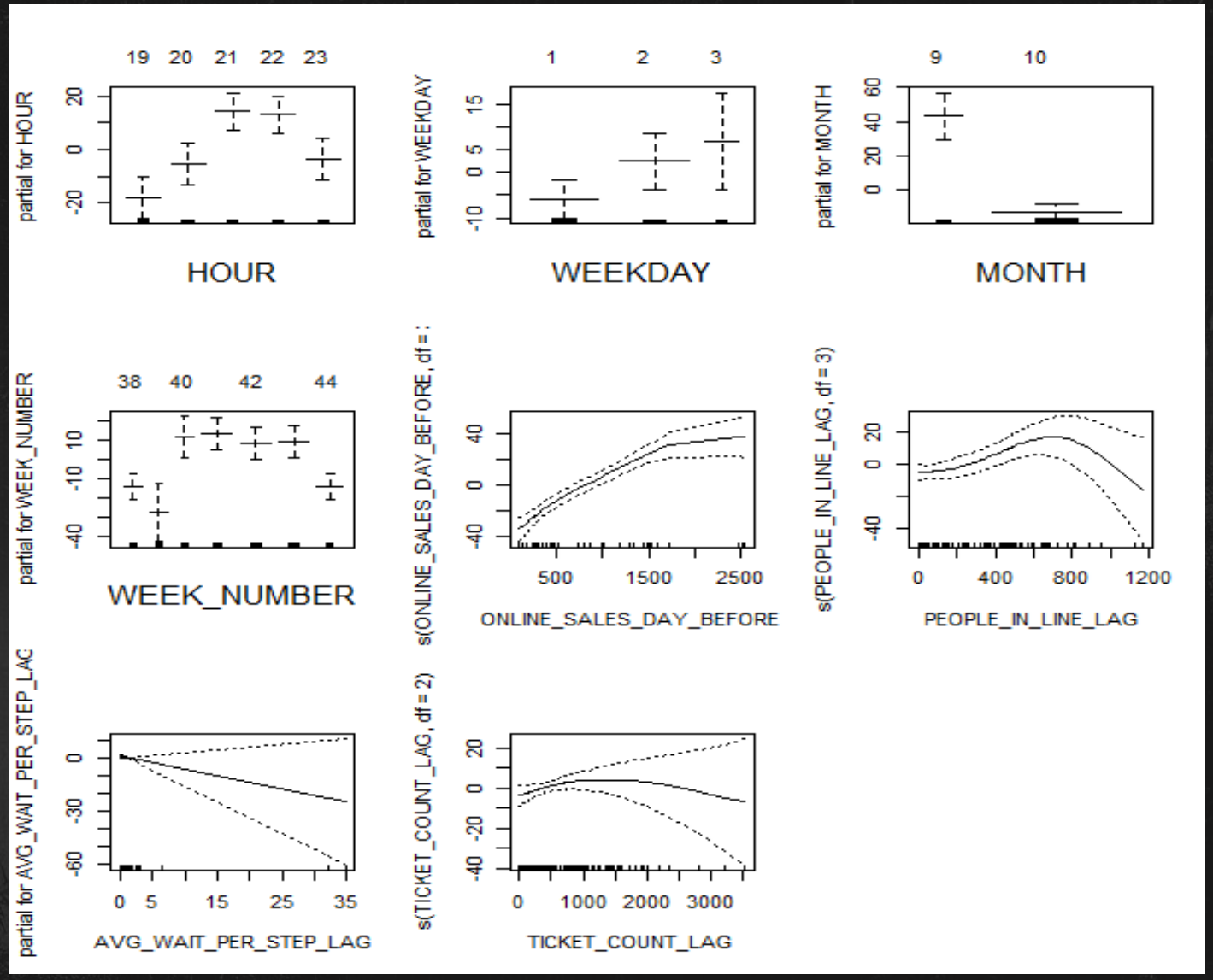
```
(Dispersion Parameter for gaussian family taken to be 501.1479)

    Null Deviance: 249465.2 on 174 degrees of freedom
Residual Deviance: 76675.61 on 153 degrees of freedom
AIC: 1607.075

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
                                    Df Sum Sq Mean Sq F value    Pr(>F)
HOUR                                 4  35898    8974 17.9078 4.428e-12 ***
WEEKDAY                              2  52736   26368 52.6152 < 2.2e-16 ***
MONTH                                1    977     977  1.9503   0.16457
WEEK_NUMBER                          5  17460    3492  6.9680 6.844e-06 ***
s(ONLINE_SALES_DAY_BEFORE, df = 3)   1  42070   42070 83.9464 3.143e-16 ***
s(PEOPLE_IN_LINE_LAG, df = 3)        1   2686    2686  5.3593   0.02194 *
AVG_WAIT_PER_STEP_LAG                1   1302    1302  2.5973   0.10911
s(TICKET_COUNT_LAG, df = 2)          1    275     275  0.5480   0.46025
Residuals                          153  76676     501
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
                                   Npar Df Npar F    Pr(F)
(Intercept)
HOUR
WEEKDAY
MONTH
WEEK_NUMBER
s(ONLINE_SALES_DAY_BEFORE, df = 3)       2 5.6206 0.004411 **
s(PEOPLE_IN_LINE_LAG, df = 3)            2 6.1636 0.002664 **
AVG_WAIT_PER_STEP_LAG
s(TICKET_COUNT_LAG, df = 2)              1 3.6162 0.059098 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

GAM PARTIAL EFFECTS

- Partial effect on hour as expected, predicted picks of average wait time around 9/10 pm
- Reasonable expected partial effect on online sales
- Strange final behaviour of people in line, but confidence intervals allow the possibility of an asymptotic effect
- As seen before, low significance for wait per step, week number and ticket count lag

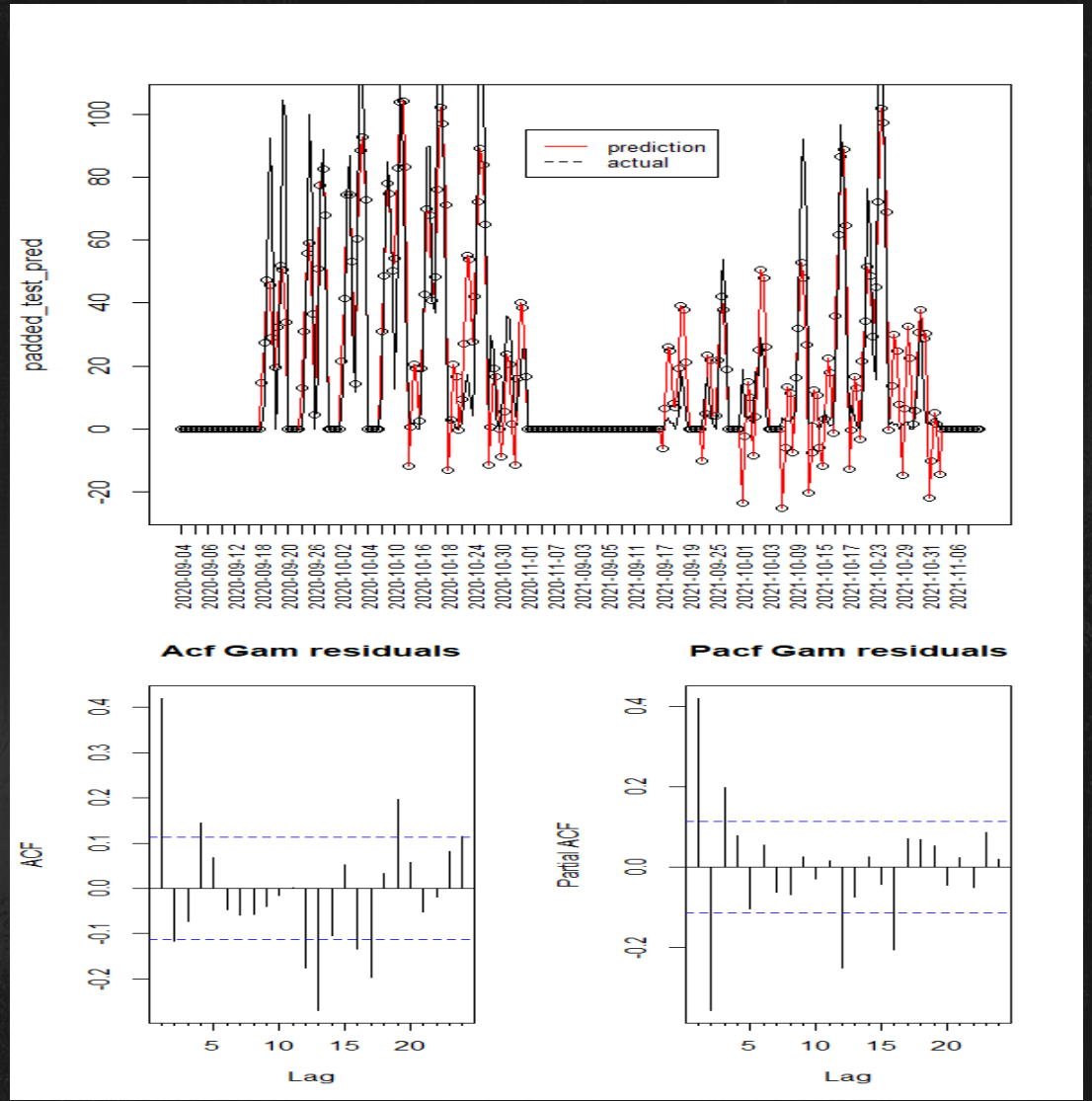


SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

GAM VISUAL FIT

- Model fitted on 2020 and then incremental predictions done over 2021 as a test for goodness.
- As expected, used as a static model day by day, gam is not able to capture significant correlations over time.
- Bad fit on days after closure cause available data are older in time.



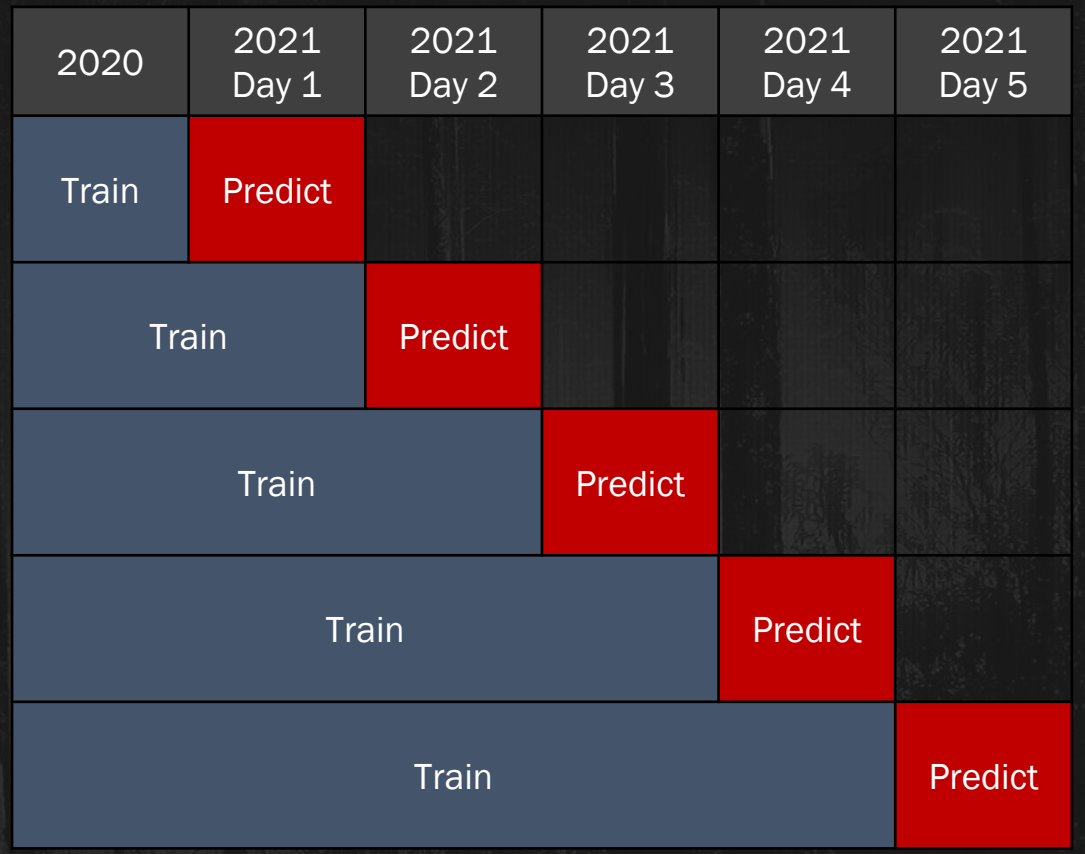
SCAREFEST SCREAM PARK

HAYRIDE OF DOOM


INCREMENTAL HOLDOUT
SIMULATION
SCAREFEST
SCREAM PARK

INCREMENTAL HOLDOUT

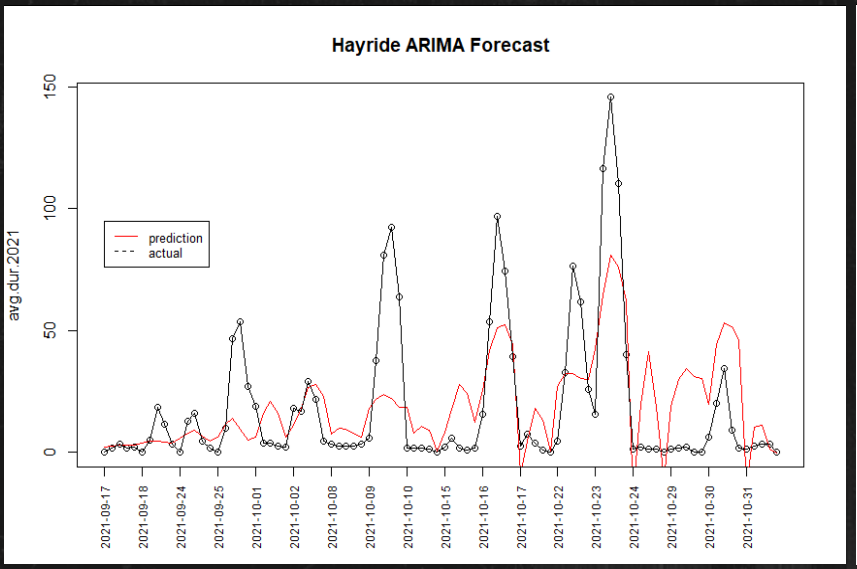
- Base train data 2020
- Each fold add one day (in sequence) to the training data
- Each fold predict one day in the future

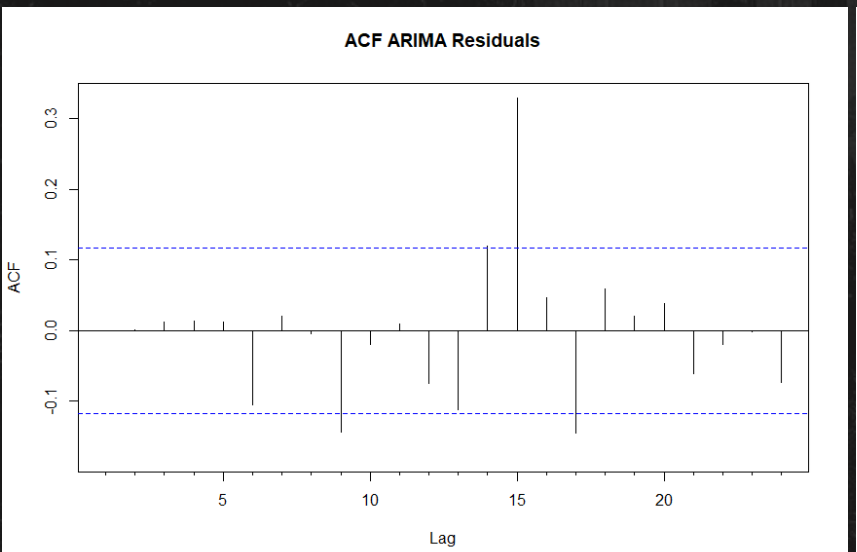


SCAREFEST
SCREAM PARK

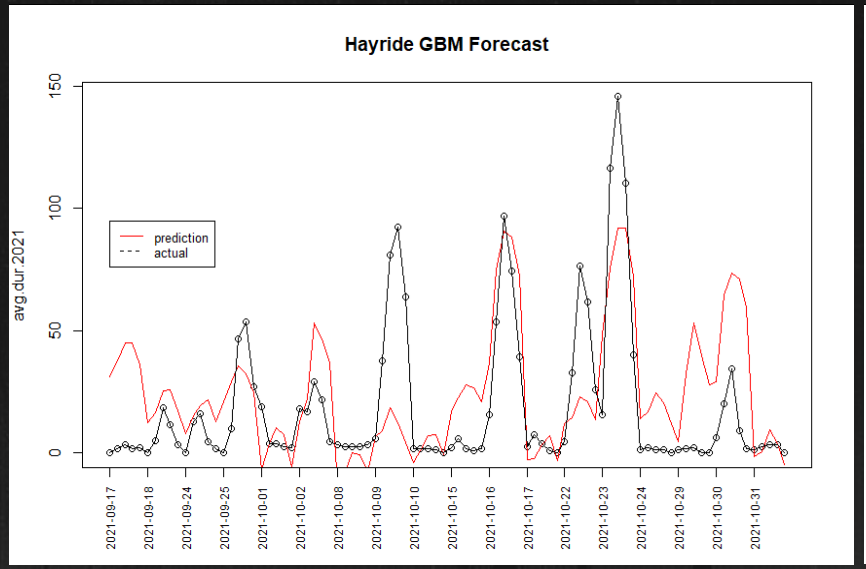
2021 FORECAST FOR BEST MODELS

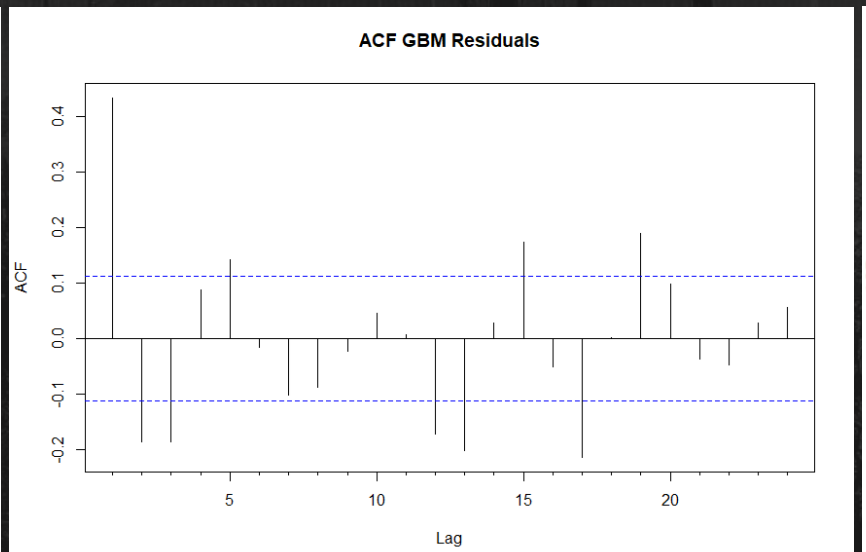
ARIMA



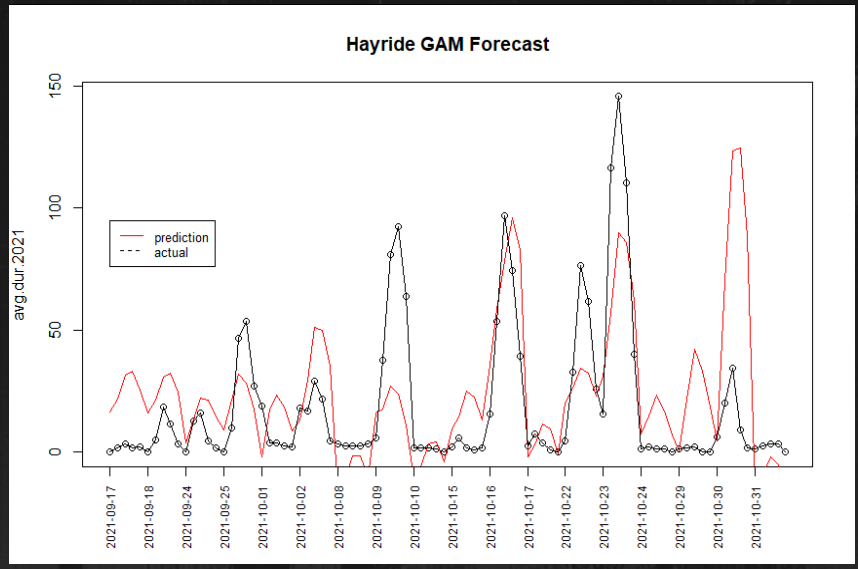


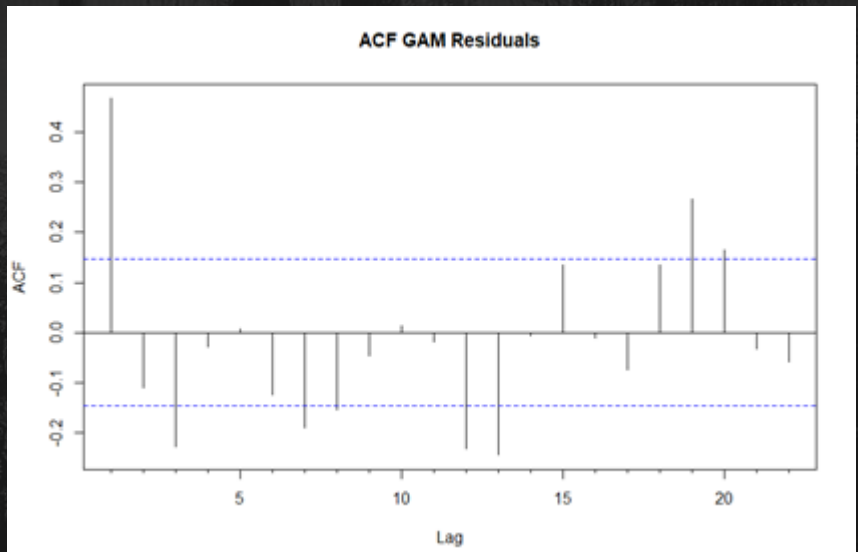
Gradient Boosting





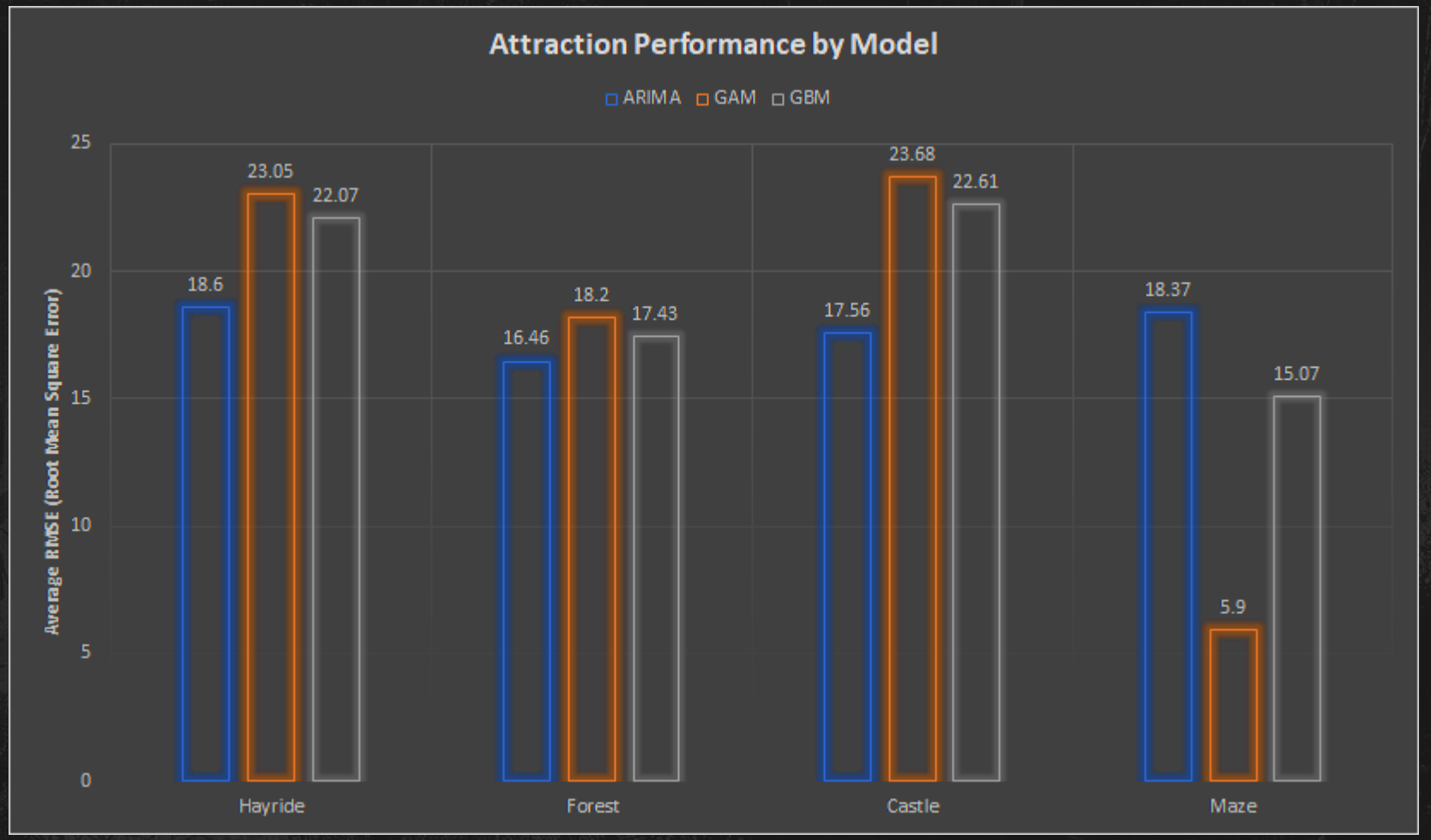
GAM



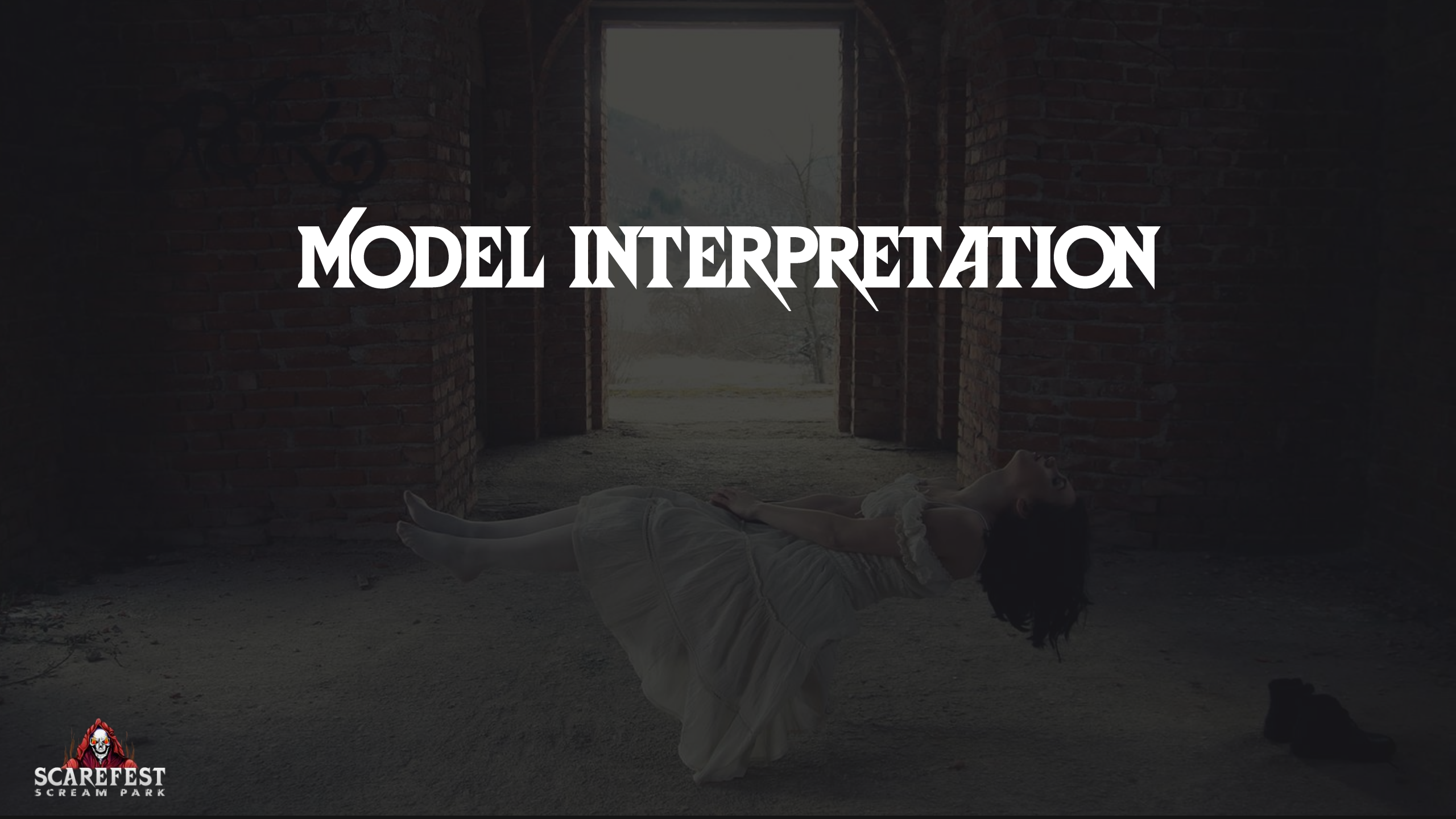


APPLYING BEST MODELS TO ALL ATTRACTIONS

- We compared the RMSE performance of our 3 best models (ARIMA, GAM and GBM) across all attractions.
- Deviation between model performance was consistent across all park attractions.
- Only exception to the above is the maze which had less data than the other attractions (only 2021). This made it harder for ARIMA to forecast



Model interpretation



SIMULATE 2021

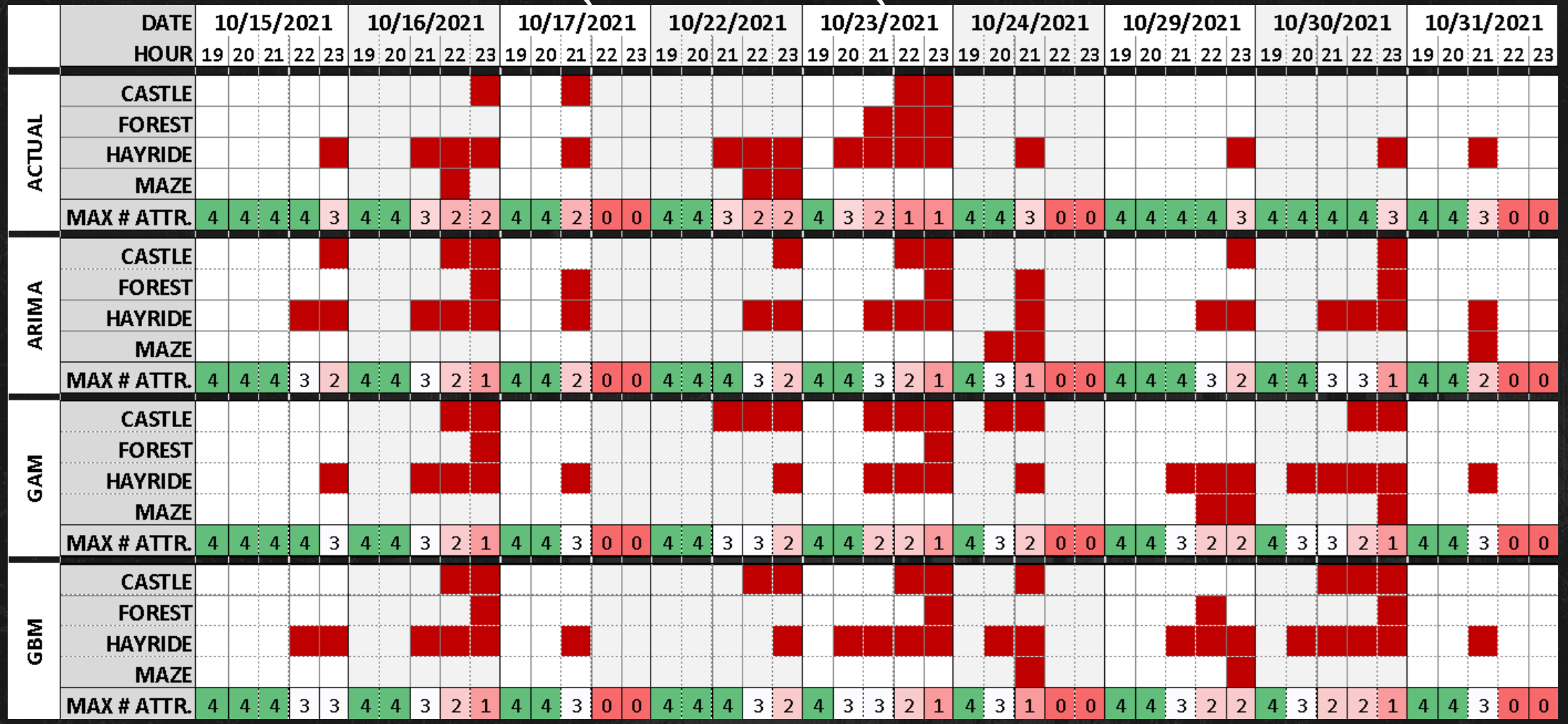
- Using the incremental holdout strategy predict the wait times for each attraction one day in advance
- For each hour in the day see if the cumulative wait time of attractions exceed the park close time
- Mark the attraction that people will not have time to attend

Attraction	Hour 22 Estimated Wait (minutes)	Attraction Duration (minutes)	Total Time	Reverse Cum. Time Remaining
Hayride	60	25	85	-65
Castle	30	15	45	20
Forest	20	10	30	65
Maze	10	15	25	95
Total	120	65	**185**	**120**

SCAREFEST SCREAM PARK

HAYRIDE OF DOOM

Max Number of Attractions Simulation



	DATE	10/15/2021					10/16/2021					10/17/2021					10/22/2021					10/23/2021					10/24/2021					10/29/2021					10/30/2021					10/31/2021				
	HOUR	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23	19	20	21	22	23
ACTUAL	CASTLE																																												
	FOREST																																												
	HAYRIDE																																												
	MAZE																																												
	MAX # ATTR.	4	4	4	4	3	4	4	3	2	2	4	4	2	0	0	4	4	3	2	2	4	3	2	1	1	4	4	3	0	0	4	4	4	4	3	4	4	4	4	3	4	4	3	0	0
ARIMA	CASTLE																																											
	FOREST																																												
	HAYRIDE																																												
	MAZE																																												
	MAX # ATTR.	4	4	4	3	2	4	4	3	2	1	4	4	2	0	0	4	4	4	3	2	4	4	3	2	1	4	3	1	0	0	4	4	4	3	2	4	4	3	3	1	4	4	2	0	0
GAM	CASTLE																																											
	FOREST																																												
	HAYRIDE																																												
	MAZE																																												
	MAX # ATTR.	4	4	4	4	3	4	4	3	2	1	4	4	3	0	0	4	4	3	3	2	4	4	2	2	1	4	3	2	0	0	4	4	3	2	2	4	3	3	2	1	4	4	3	0	0
GBM	CASTLE																																											
	FOREST																																												
	HAYRIDE																																												
	MAZE																																												
	MAX # ATTR.	4	4	4	3	3	4	4	3	2	1	4	4	3	0	0	4	4	4	3	2	4	3	3	2	1	4	3	1	0	0	4	4	3	2	2	4	3	2	2	1	4	4	3	0	0

Comparison of model to actual

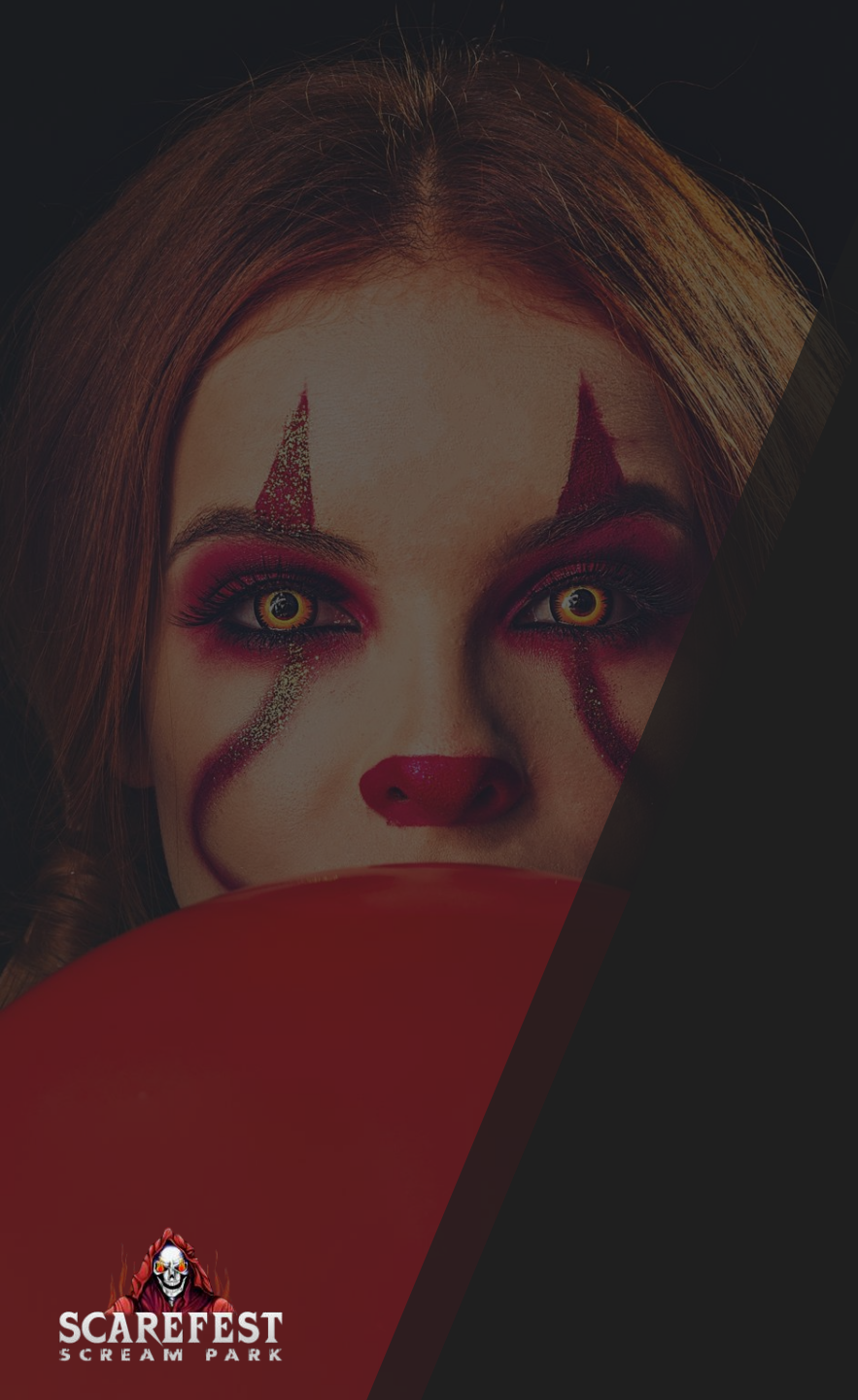
Correct: **156**
% Correct: **87%**

Correct: **153**
% Correct: **85%**

Correct: **153**
% Correct: **85%**

SCAREFEST SCREAM PARK

= not enough time to complete attraction



KEY TAKEAWAYS

- People In Line from last week and Online Sales the Day Before are consistently the most important predictors of wait time
- 2020 COVID oddities are likely throwing off 2021 predictions
- The wait times have patterns that ARIMA can capture well, but the large spikes are harder for the model to predict
- GAM and Gradient Boosting can sometimes predict large spikes, but they often are poor at consistent predicting
- Scarefest stops selling All Inclusive tickets at 21 but the model suggests on Friday that they can be sold until 22
- All models consistently say to stop selling 4 attraction by 21 on Saturday in late Oct specifically for the Hayride

SCAREFEST
SCREAM PARK

SCAREFEST
SCREAM PARK

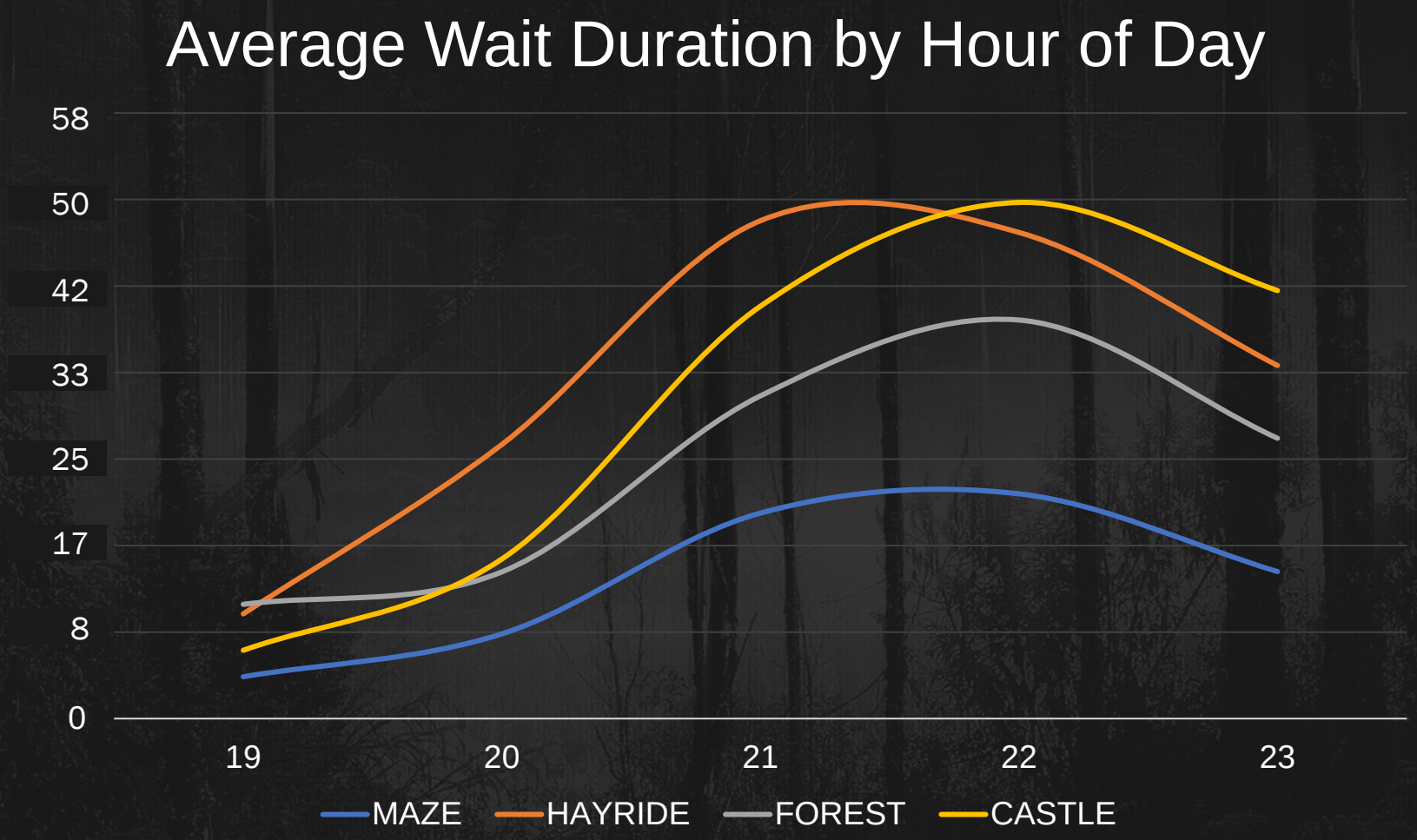
DATA FIELDS

DATA FIELDS	DEFINITION
WAITLIST_TIME_HOUR	The day and hour guests entered the virtual line
AVG_WAIT_DURATION	The average time all guests had to wait within the hour (measured in seconds)
AVG_WAIT_PER_STEP	The rate at which parties moved one position in line (seconds per 1 position in line)
PEOPLE_IN_LINE	The greatest number of guests that were in line within the hour
HOUR_NUMBER	The hour guests entered the virtual line (integer)
YEAR	The year guests entered the virtual line
WEEKDAY	The weekday guests entered the virtual line (1 - Friday, 2 - Saturday, 3 - Sunday)
MONTH	The month guests entered the virtual line (integer)
HOURS_OPEN	The number of hours the park is open that day (0 - Closed, 3 - Sundays, 4 – Sep, 5 - Oct)
IS_OPEN_DAYS	Binary classification if the park was open that day (1 - open, 0 - closed)
WEEK_NUMBER	Week number of the year (adjusted to align first weekend of October)
ONLINE_ONLY	Binary classification if tickets had to be purchased online that day (1 - online only, 0 - online and onsite)
BAD_DAYS	Binary classification of days that had major operational issues (1 - bad day, 0 - normal day)
RAIN	Binary Classification of days where rain or bad weather impoacted sales (1 - rain, 0 - no rain)
NUM_ATTRACTIONS	The number of attractions that were open that day (integer)

SCAREFEST
SCREAM PARK

AVERAGE WAIT DURATION BY HOUR OF DAY

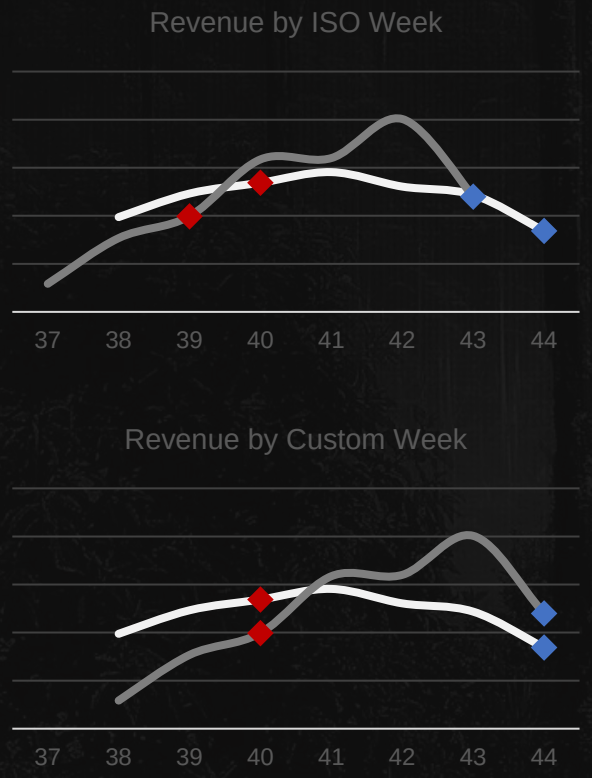
- The most common customer behavior is to go straight to the Hayride and end with the Castle



SCAREFEST
SCREAM PARK

DATA TRANSFORMATIONS

Adjusted Week Number
(Halloween Alignment)



Padded Date/Hours with 0
(Non-Uniform Time Series)

Every year has 30 days
Every week has 3 days
Every day has 5 hours

Added 1 week lag to features
(Only features unknown in future)

PEOPLE IN LINE

AVERAGE WAIT PER STEP

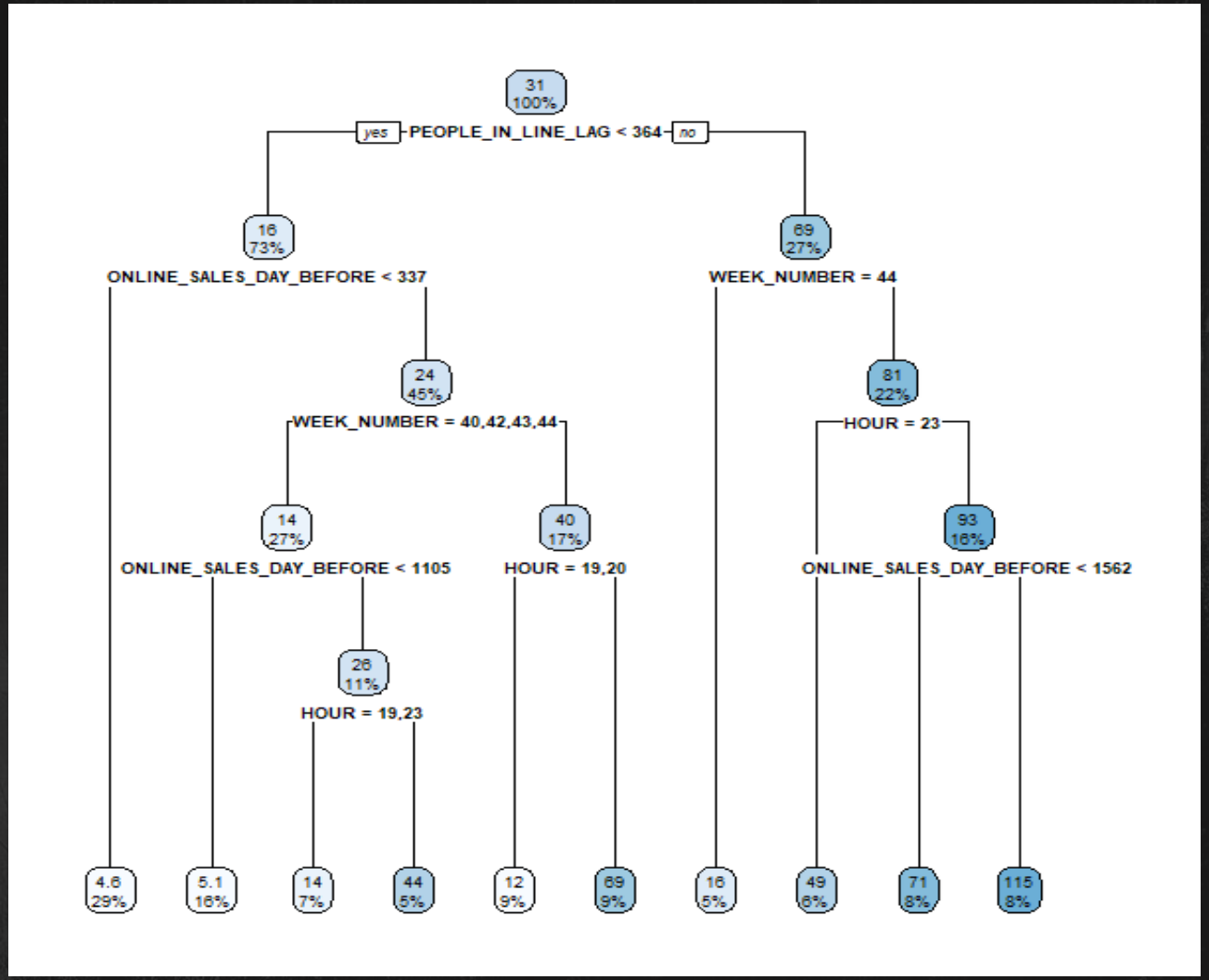
TICKET_COUNT

Allowing for a 1 week forward forecast if features are used in models

SCAREFEST
SCREAM PARK

REGRESSION TREE

- Completely interpretable decision rule
- As expected again the online sales the day before is part of the decision tree



SCAREFEST SCREAM PARK

HAYRIDE OF DOOM