Attendance Prediction

2022/23 Season

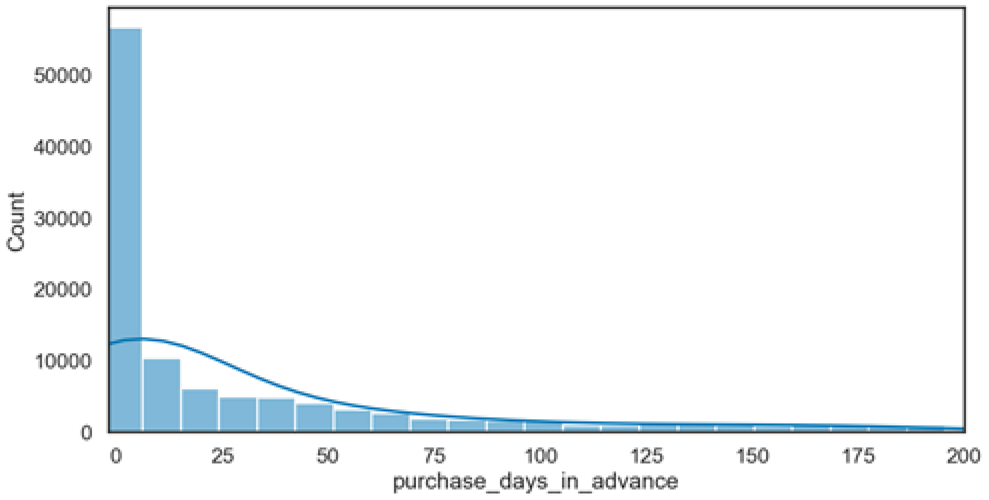
Rikesh Patel

Table of Contents

01

Business Insights

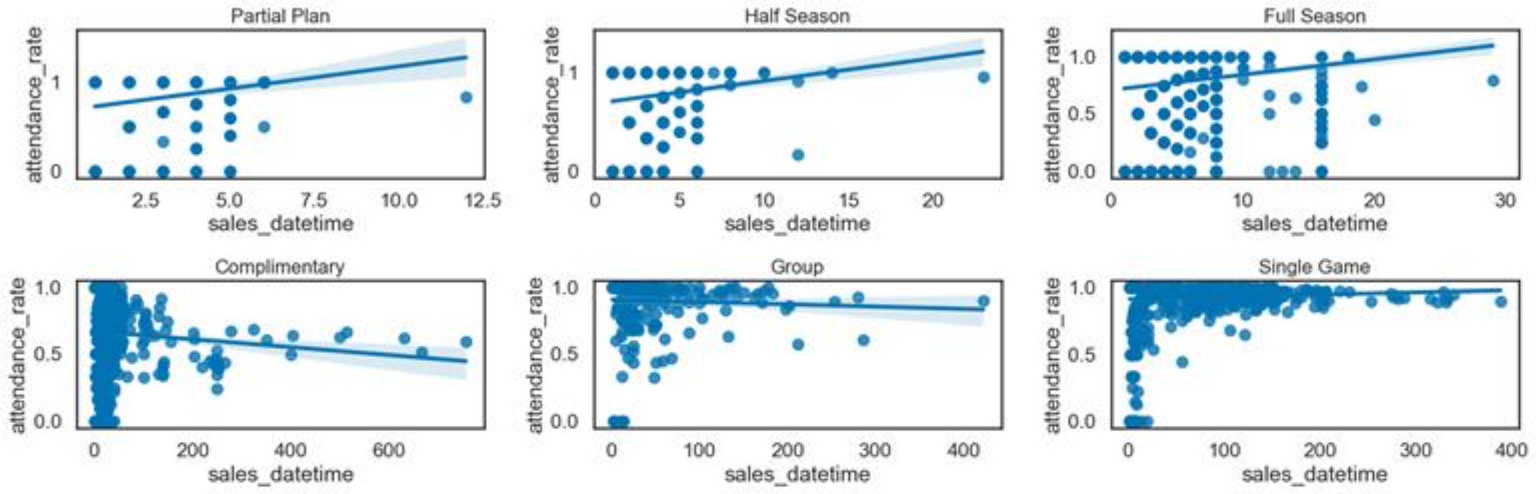
of Days Made Purchase in Advance



Days in Advance	Customer Count
-1 (after game started)	9,655
0 (within one day before game start)	16,154
1	7,023
2	6,846
3	5,409
4	5,941
5	3,367

- **Result:** Majority of customers made purchases within 5 days before the game start time; around 10,000 customers bought their tickets even after the game start time
- **Insights:** Our promotion for certain game should start at least 5 days before the game starts to draw people's attention; put leftover tickets on secondary markets to attract those who bide their time

of Ticket Purchased vs. Attendance



- **Result:** For Half Season, Group, and Single Game customers, they are more likely to attend games; for Comp customers, the attendance rate tends to be 50% as more tickets given out
- **Insights:** If Half Season, Group, and Single Game customers buy more tickets, we can anticipate the attendance rate to be high; given out free tickets does not necessarily increase attendances

02

Prediction Model

Overview of the Project

Purpose	Develop a model to predict games attendance for 2022/23 season
Data Preparation	Customer attendance history Stadium, customer, and opponent team descriptive information
Metrics	Categorize records that appear in both Sales and Scan Data as 1 (attended), only appear in Sales Data as 0 (did not attend)
Data Cleaning	No NULL values and outliers are acceptable
Feature Engineering	Create new variables based on original features
Feature Selection	Select features that contribute most to prediction of Attendance
Model Tuning	Use machine learning models to make prediction
Results	Fit chosen model on test set to get most accurate prediction

Feature Selection

SequentialFeatureSelection Function

scoring="accuracy"

Ticket Related Features
Date/Time Related Features
Opponent Team Related Features

Feature Importance	Feature Name	Feature Importance	Feature Name
1	purchase_days_in_advance	11	month_4
2	is_primary_market_yes	12	price_code_type_Full Season
3	Playoffs in 2021-2022_Yes	13	month_12
4	dow_Thursday	14	City_Orlando
5	event_hour	15	City_San Antonio
6	price_code_type_Group	16	section_name_107
7	hour_bin_Evening	17	section_name_102
8	price_bin_middle	18	section_name_106
9	price_code_type_Half Season	19	City_Sacramento
10	dow_Saturday	20	City_Memphis

Model Exploration

STEP 1

Split Dataset

70% -> Training

30% -> Testing

STEP 2

Fit Different Variables

Include different numbers of variables

STEP 3

Tune Model

5-fold cross-validation

GridSearchCV -> find best combination

STEP 4

Predict on Testing Set

Predict labels

Calculate accuracy

Model Performance

Model	Criteria	Value
Logistic Regression	# of variables	1
	Accuracy in Training	76.75%
	Accuracy in Testing	**76.75%**
Random Forest	# of variables	15
	Accuracy in Training	77.08%
	Accuracy in Testing	**76.99%**
eXtreme Gradient Boosting	# of variables	20
	Accuracy in Training	77.44%
	Accuracy in Testing	**77.38%**

03

Findings and Recommendations

2022/23 Season Data

Ticket Package

Price Per Game

Group	$35-$595
Mini Plan	$90-$390
Half Season	$55-$371
Full Season	$68-$337

Game Date & Time

Home Game Data

01	42 home games out of 82 games in total
02	9 games in the afternoon 33 games in the evening
03	6 on Mon, 5 on Tue, 7 on Wed, 4 on Thu, 4 on Fri, 8 on Sat, 8 on Sun

Apply Insights on Company Business

Income Source	Recommendation
Ticket Sells	Adjust promotion strategy based on key features Consider overselling tickets to ensure attendance
Advertising	Charge for in-arena advertising based on headcount
Merchandise	Estimate the storage of merchandise and revenue from selling them

04

Limitations and Future Steps

Model Limitations

01

Number of Features

Existed features do not predict attendance well.

02

External Factors

Did not consider factors such as pandemic, which could also influence attendance

03

Model Tuning

Only tuned a few hyperparameter combinations for two ML models.

Model Improvements

Domain Exports

Consult domain exports to have a better understanding of features that might influence attendance

Computing Power

With the help of higher computing power, we can tune the model even further

Outside Data

Include more data sources to potentially increase accuracy

Model Tuning

Try to tune more hyperparameter combinations for our models

Feature Combinations

Explore more in feature engineering, such as days_since_last_purchase

Try different feature selection methods

ML Models

Try more machine learning models, such as K-Nearest Neighbors, Support Vector Machines, etc.

Future Steps

Before Game

Ticket Selling:

- Teaser emails to attract potential customers
- Marketing campaigns to promote selling
- Second market for leftover tickets
- Activation emails before purchasing

Reminders:

- Synchronizing game info to calendars
- Sending reminders at key time points

During & After Game

Entertainment:

- Merchandise stores, snakes and drinks
- Interactions with customers: t-shirts giveaway, DJ, video board

Feedbacks:

- Segment customers and analyze behavior
- Survey for game experience

THANKS!

Open to any feedback!

[redacted]

RESOURCES

Outside Data Source

- Weather Data: https://www.ncei.noaa.gov/cdo-web/
- 2022/23 Season Game Schedule: https://www.cbssports.com/nba/teams/LAC/los-angeles-clippers/schedule/regular/
- 2022/23 Season Ticket Package Price: https://www.nba.com/clippers/seasontickets

Photos

- Paul George
- Kawhi Leonard

Appendix

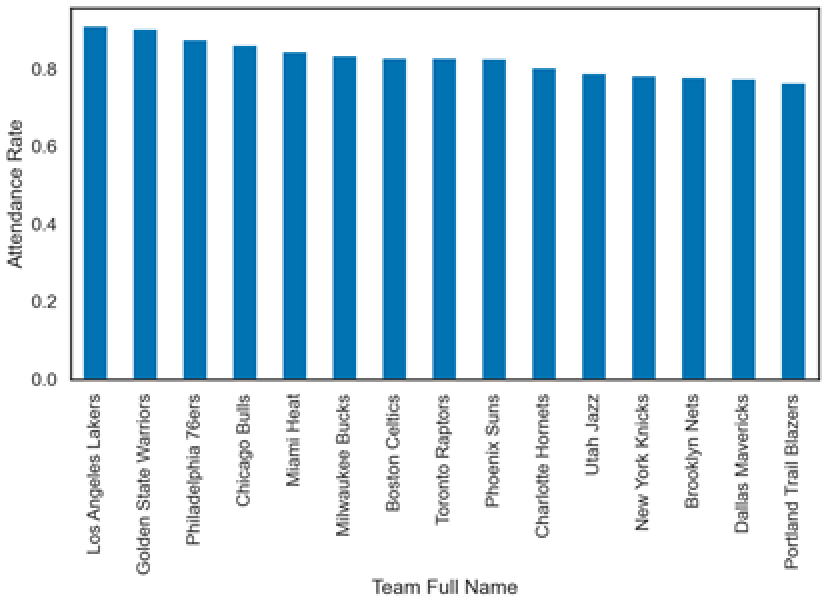
Summary of Numerical Features

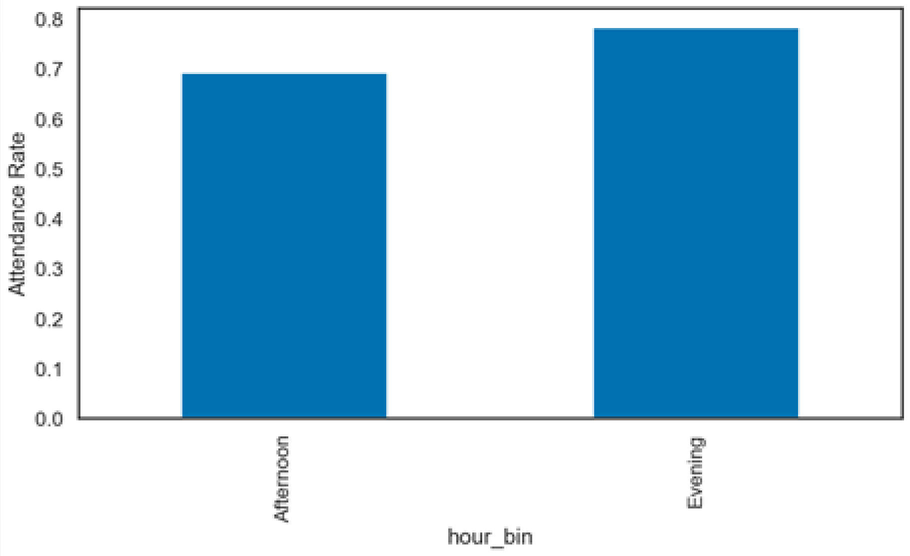
	% Populated	Mean	Min	Max	Standard Deviation	% Zero
price	100	137	0	2,164	166	20
Championships	100	3	0	17	4	34
Vegas Over/Under 21/22	100	41	23	57	10	0
Vegas Over/Under 22/23	100	41	23	54	10	0
seat_count	100	338	126	522	126	0
purchase_days_in_advance	100	113	-114	845	146	5
ticket_bought_num	100	44	1	758	94	0
TMAX	100	72	53	92	10	0

Summary of Categorical Features

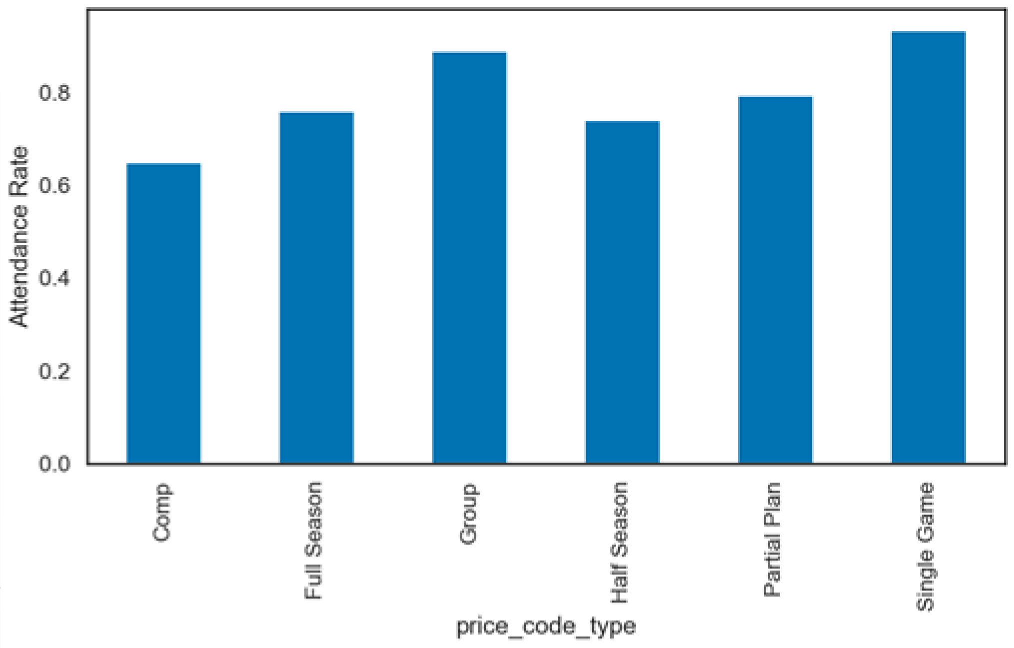
	% Populated	# Unique Values	Most Common Value
Attendance	100	2	1
month	100	7	11 (November)
dow	100	7	Sunday
event_hour	100	3	19 (7pm)
hour_bin	100	2	Evening (after 6pm)
section_name	100	30	116
Conference	100	2	Western
Playoffs in 2020-2021	100	2	Yes
Playoffs in 2021-2022	100	2	Yes
Follow_W3?	100	2	0
City	100	29	Los Angeles
price_code_type	100	6	Full Season
is_primary_market	100	2	Yes
price_bin	100	3	cheap

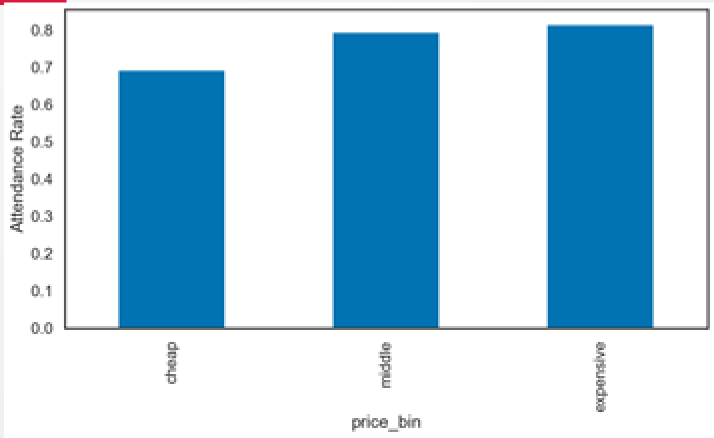
Exploratory Data Analysis

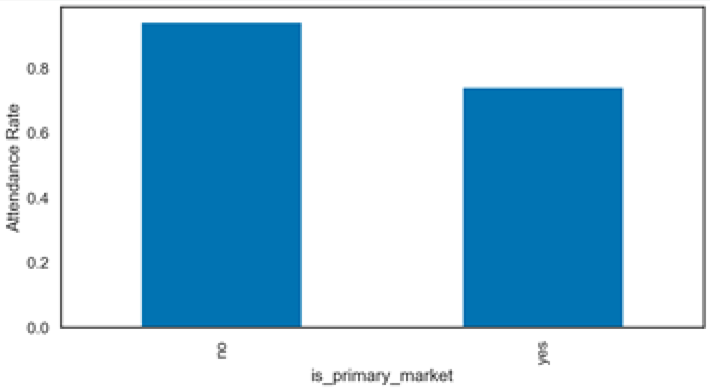




Exploratory Data Analysis







Model Performance

Model	# of Variables	Best Performance Parameters		Train Accuracy	Test Accuracy
Random Forest		n_estimators	max_depth		
	1	150	12	76.89%	76.82%
	2	150	12	76.89%	76.85%
	3	200	12	76.92%	76.88%
	6	300	12	77.02%	76.96%
	10	250	12	77.03%	76.97%
	15	**300**	**12**	**77.08%**	**76.99%**
	20	250	12	76.98%	76.95%

Model Performance

Model	# of Variables	Best Performance Parameters							Train Accuracy	Test Accuracy
		nthread	objective	learning_rate	max_depth	n_estimators	disable_default_eval_metric	eval_metric		
XGBoost	1	4	'binary:logistic'	0.2	5	300	True	'error'	76.90%	76.88%
	2	4	'binary:logistic'	0.2	5	300	True	'error'	76.90%	76.88%
	3	4	'binary:logistic'	0.3	7	100	True	'error'	76.92%	76.89%
	6	4	'binary:logistic'	0.2	7	300	True	'error'	77.06%	76.94%
	10	4	'binary:logistic'	0.2	7	500	True	'error'	77.14%	76.99%
	15	4	'binary:logistic'	0.3	7	500	True	'error'	77.30%	77.17%
	20	**4**	**'binary:logistic'**	**0.3**	**7**	**500**	**True**	**'error'**	**77.44%**	**77.38%**