



Seat Analytics

Final Presentation

Capstone Advisor: Arin Brahma
Client Manager: Shareth Hariharan

Team No. 1: Brenda Mas, Mohu Sah & Peng Zhang



AGENDA

- 01. INTRODUCTION**
- 02. OUR RESEARCH**
- 03. METHODOLOGIES**
- 04. MODEL BUILDING**
- 05. CONJOINT ANALYSIS**
- 06. CONCLUSION**



Fandango is the ultimate digital network for all things movies.

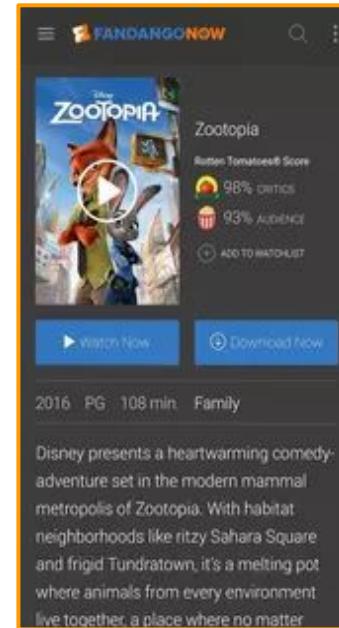
Awareness



Ticketing



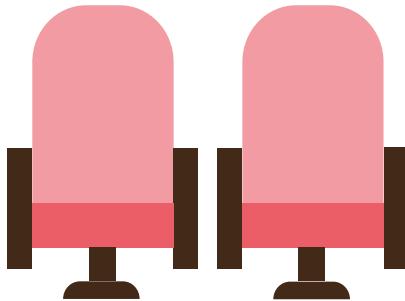
Streaming



Fandango Ticketing

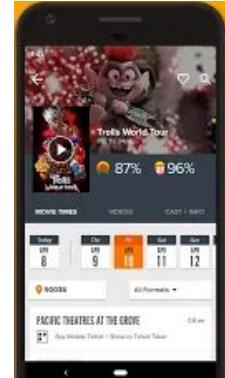
Value Creation Value Capture Sustainability

Theaters



Theaters list their shows & differentiated services on Fandango App

FANDANGO App



Connects theaters & moviegoers in ticket purchasing phase

Moviegoers

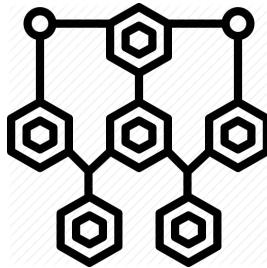


Moviegoers see all the movie theaters & showtimes



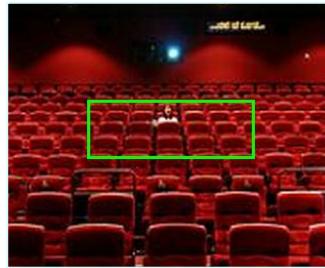
OUR PROJECT

Seat Analytics: Problem Statement



Defining Scoring Framework*

Develop & assign “popularity” score for EACH seat



Defining Zones

Identify popular “zones”



Validating Score Variance

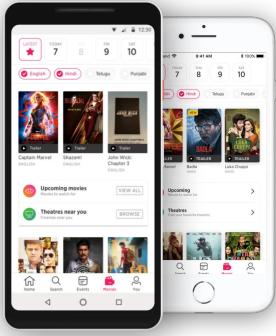
Determine if score varies significantly by features



Score Prediction

Build dynamic predictive model that predicts seat score

How Will This Help Fandango?



Premium/Discount Pricing

- ★ Offer popular seats to “high value” customers & drive ticket sales for least popular seats

Improve Conversion

- ★ Reduce customer flipping between showtimes

Improve Yield

- ★ Help theaters adjust operations strategy based on features with use predictive analytics

Our Dataset



Movie Table

- **Movie ID** 
- Movie Parent ID
- Movie Title
- Movie Parent Title
- Movie Release Date
- Movie Parent Release Date



Seat Availability Data

- Date Name
- Total Capacity
- Reserved Seats
- Time to Showtime
- Seat Data
- Showtime
- **Theater ID** 
- **Movie ID** 



Theater Table

- Theater Name
- **Theater ID** 
- State Code
- Zip Code
- Seat Count
- **Chain Key** 
- Amenities 



Chain Table

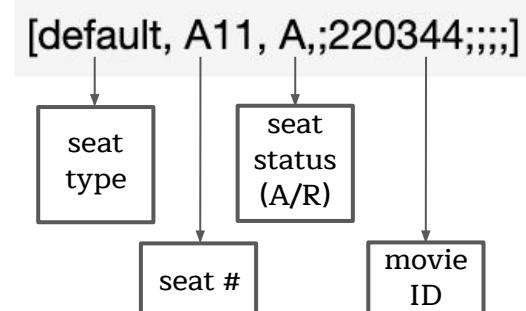
- **Chain Key** 
- Chain name

Seat Availability Data

	date_name	total_capacity	reserved_seats	time_to_showtime	seat_data	theater_id	showtime	movie_id
0	20200111	127	14	386	[default, A1, A,;220344;;;;]	1420	2020-01-11 19:20:00	220344
1	20200111	127	14	386	[default, A10, A,;220344;;;;]	1420	2020-01-11 19:20:00	220344
2	20200111	127	14	386	[default, A11, A,;220344;;;;]	1420	2020-01-11 19:20:00	220344
3	20200111	127	14	386	[default, A12, A,;220344;;;;]	1420	2020-01-11 19:20:00	220344
4	20200111	127	14	386	[default, A13, A,;220344;;;;]	1420	2020-01-11 19:20:00	220344

BIG DATA

- 84,417,456 records
- 20 unique theaters
- 270 unique movies
- Range: Oct 2019 to March 2020

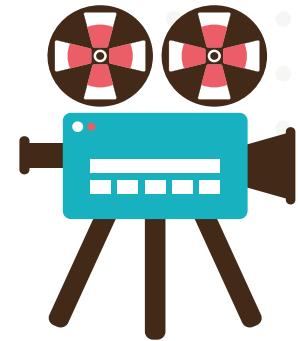


02



OUR RESEARCH

Research 1: Movie Theater Facts



Online Buyers are:

- More likely to consult showtimes on 3rd-party ticketing websites
- Purchase early, on average from 2 hrs to 6 days (Time snapshot)
- Increasingly chose theater for its amenities (Experiences)
- More likely to attend in groups



Research 2: Seat Location as an Attribute

Best Location



- For best sound quality, sit in the center 2/3 back
- Sitting slightly off-center to amplify sound effects
- Last evening show on Monday or Wednesday will be the emptiest

Best seat at the cinema?

Displaying poll results.

Front row

672 votes / 1%

Near the front

3833 votes / 8%

Exactly centered

26194 votes / 55%

Rear section

5336 votes / 11%

Back row

2617 votes / 5%

Anywhere

470 votes / 0%

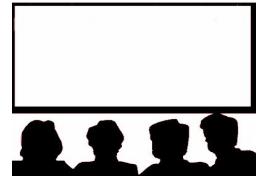
Slashdot Survey Results: 48,000 Respondents

Research 3: Measurable Attributes



Inconvenience

of people you have to cross to get to your seat



Distance To Screen

Radial Distance to the screen



Symmetry

Angle in which you view the screen



Sound

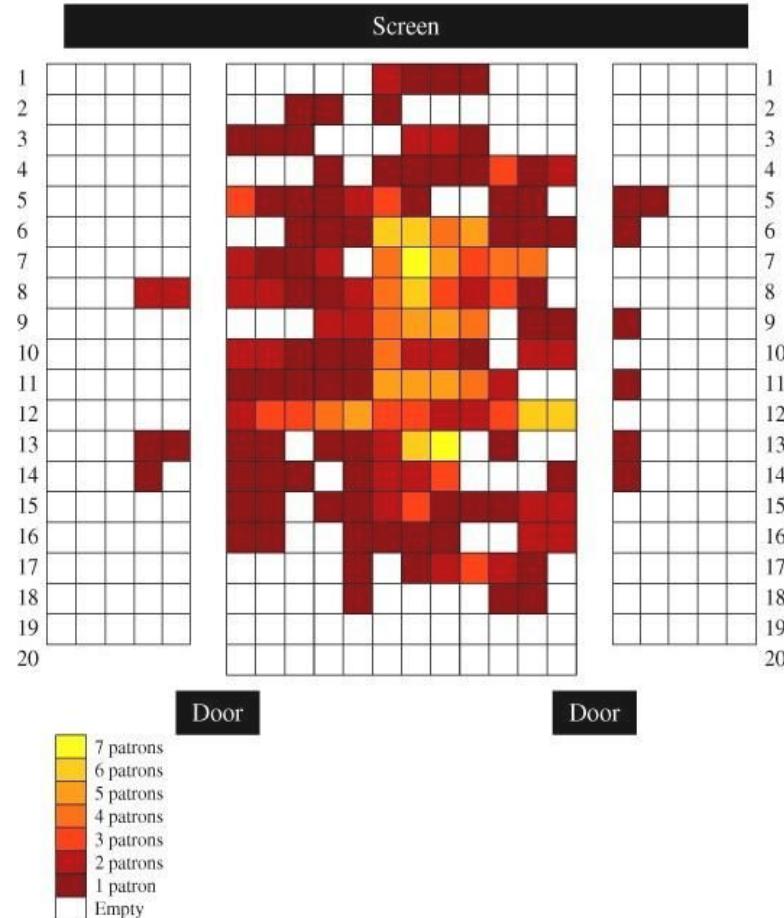
Distance to the speaker



Solo vs Groups

Research 4: Rightward Bias

- ★ Pseudoneglect (classroom/lab setting): leftward bias
- ★ Theater seating influenced by lateral/asymmetrical preferences
- ★ Emotions (Left visual position = right hemisphere processing)
- ★ Exists when choosing in person or on a ‘page’
- ★ Only reliable when relatively empty (approx. 50%)

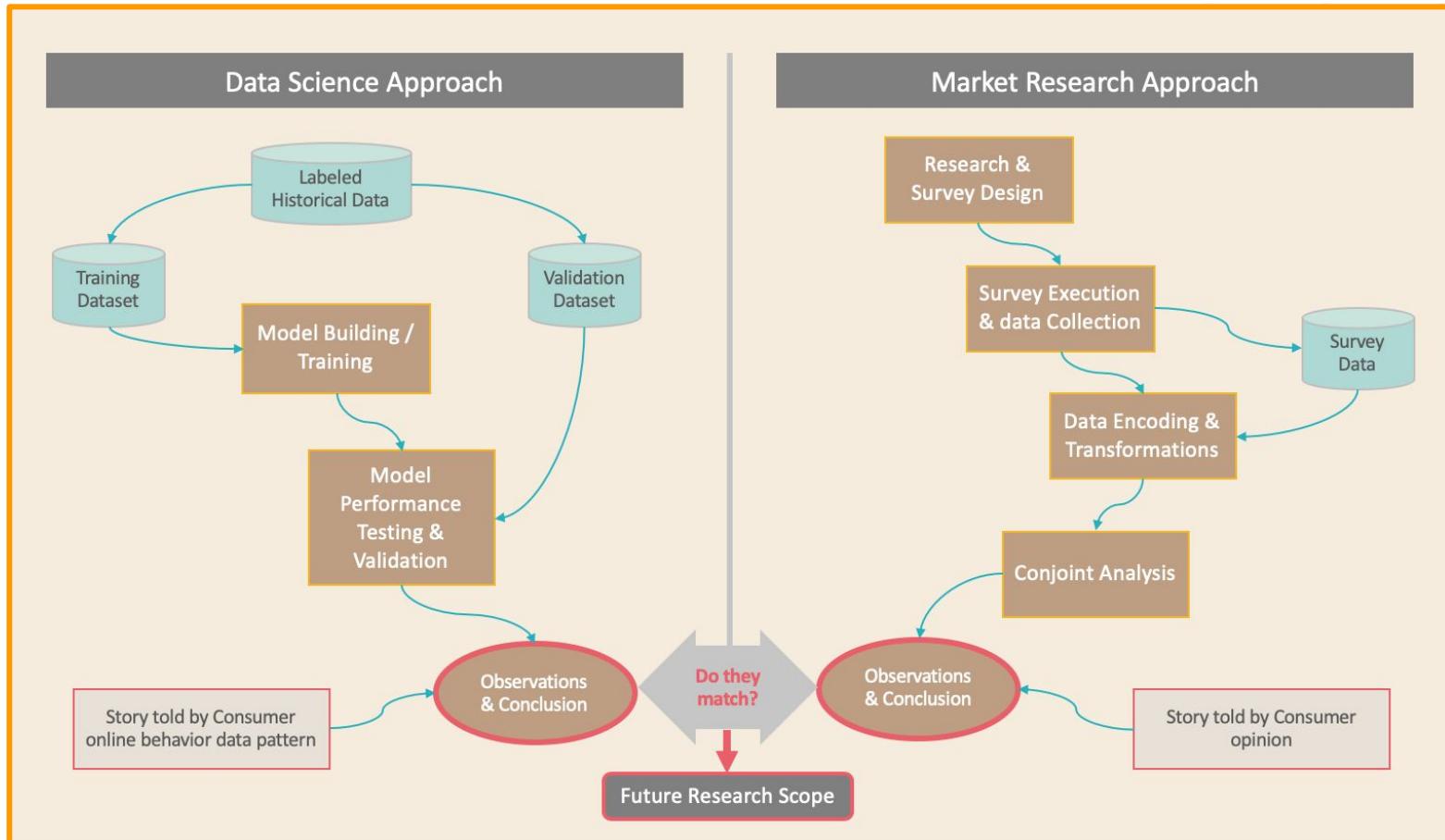


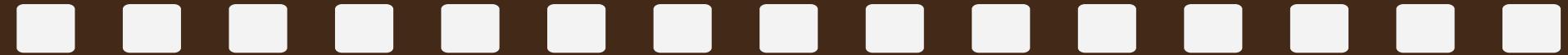
03

METHODOLOGIES



Our Approaches





DATA SCIENCE APPROACH



Scoring Framework

01



Desirability of Seat

- Numerical Score

02



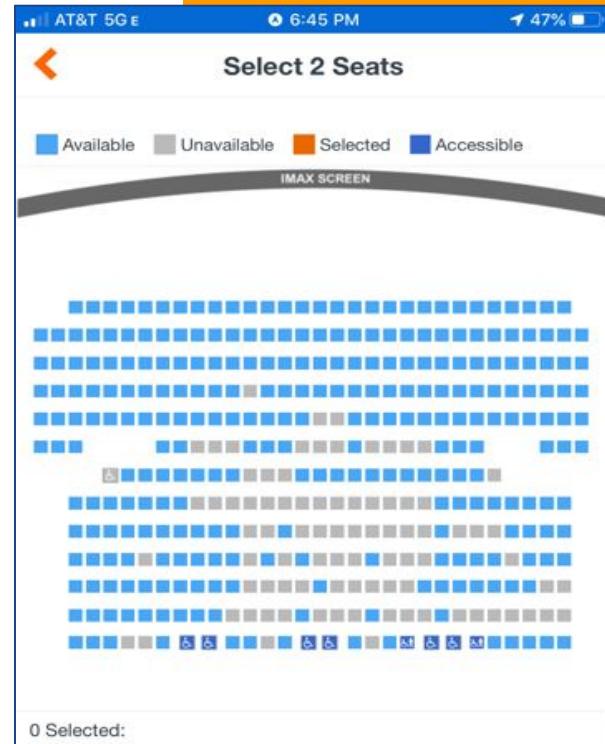
Analytical Model

- Supervised Machine Learning
- Predict Desirability

03

Feature Space Development

- ‘Static’ features
- ‘Dynamic’ features



Seat Heat Maps

of Time Snapshots: 76

One Auditorium • One Movie • One Showtime

COLUMNS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
ROWS	SCREEN																						
A	76	76	76	0	0	0	0	0	0	0	0	76	76	76	76	76	76	76	76	76	76	76	76
B	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76
C	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76
D	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76
E	18	18	74	73	75	75	74	36	36	75	75	51	51	75	75	75	76						
F	76	76	76	41	41	74	74	4	4	4	53	53	53	53	76	76	76						
G	76	76	16	16	16	71	71	76	7	7	52	52	52	52	76	76	19						
H	76	76	28	28	28	28	28	28	13	13	13	13	55	55	39	39	76	76	76				
I	76	76	75	75	52	61	47	48	15	15	15	15	47	47	71	75	75	76	76	76			
J	16	16	76	76	68	68	25	25	32	31	31	32	76	48	48	76	76	76	76	76			
K	76	76	76	76	38	15	15	15	15	15	15	15	15	15	15	15	15	76	76	76			
L	76	76	76	76	76	76	31	21		21	21	21	21	21	22	76	76	76	76	76	14	14	14

Showtime Bin
1 day to 5 days
1 to 5 hours
5 to 24 hours
Less than an hour
more than 5 days
showtime

Theater: Edwards Aliso Viejo Stadium 20 & IMAX

Movie: Star Wars Auditorium Size: 202

Show Date/Time: December 21, 2019 @ 8pm

Static Features

Features	Levels	Score	Weights
Zone	Center	5	60%
	Back	4	
	Center Right	3	
	Center Left	2	
	Front	1	
Sound	Center	3	15%
	Back	2	
	Front	1	
Symmetry	Center	3	15%
	Right	2	
	Left	1	
Seat Type	Default	3	10%
	Companion	2	
	Wheelchair	1	

Dynamic Features

Features	Logic	Range	Weights
Percent Time Remaining	'Time to Showtime' when seat reserved	Higher the Better	60%
Percent Seat Remaining	Seats Remaining/Total Capacity	Higher the Better	20%
Final Status	'Available' or 'Reserved' at last time snapshot	R - 2 A - 1	20%

Seat Popularity Score

at_number	final_status_score	percent_seat_remaining_score	percent_time_remain_score	seat_type_scores	sound_scores	symmetry_scores	zone_scores	final_popularity_score
A1	0.4	0.132426	0.00	0.3	0.15	0.15	0.6	1.73
A2	0.4	0.132426	0.00	0.3	0.15	0.15	0.6	1.73
A3	0.4	0.132426	0.00	0.3	0.15	0.15	0.6	1.73
A4	0.8	0.240099	0.25	0.3	0.15	0.15	0.6	2.49
A5	0.8	0.240099	0.25	0.3	0.15	0.15	0.6	2.49

Dynamic Scores

$$(\%_{\text{time_remaining}} * 0.6 + \%_{\text{seat_remaining}} * 0.2 + \text{final_status} * 0.2)$$

+

Static Scores

$$(\text{zone} * 0.6 + \text{sound} * 0.15 + \text{symmetry} * 0.15 + \text{seat_type} * 0.1)$$

Score Validation: Heat Maps

Based on
Purchase
Behavior



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	SCREEN																						
A	76	76	76	0	0	0	0	0	0	0	0	0	76	76	76								
B	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76				
C	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76					
D	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76			
E	18	18	74	73	75	75	74	36	36	75	75	51	51	75	75	76							
F	76	76	76	41	41	74	74	4	4	4	53	53	53	53	76	76	76						
G	76	76	16	16	16	71	71	76	7	7	52	52	52	52	76	19							
H	76	76	28	28	28	28	28	28	13	13	13	13	55	55	39	39	76	76					
I	76	76	75	75	52	61	47	48	15	15	15	15	47	47	71	75	76	76					
J	16	16	76	76	68	68	25	25	32	31	31	32	76	48	48	76	76	76					
K	76	76	76	76	38	15	15	15	15	15	15	15	15	15	15	15	76	76					
L	76	76	76	76	76	76	31	21		21	21	21	21	21	22	76	76	76	76	76	14	14	14

Based on
Popularity
Score

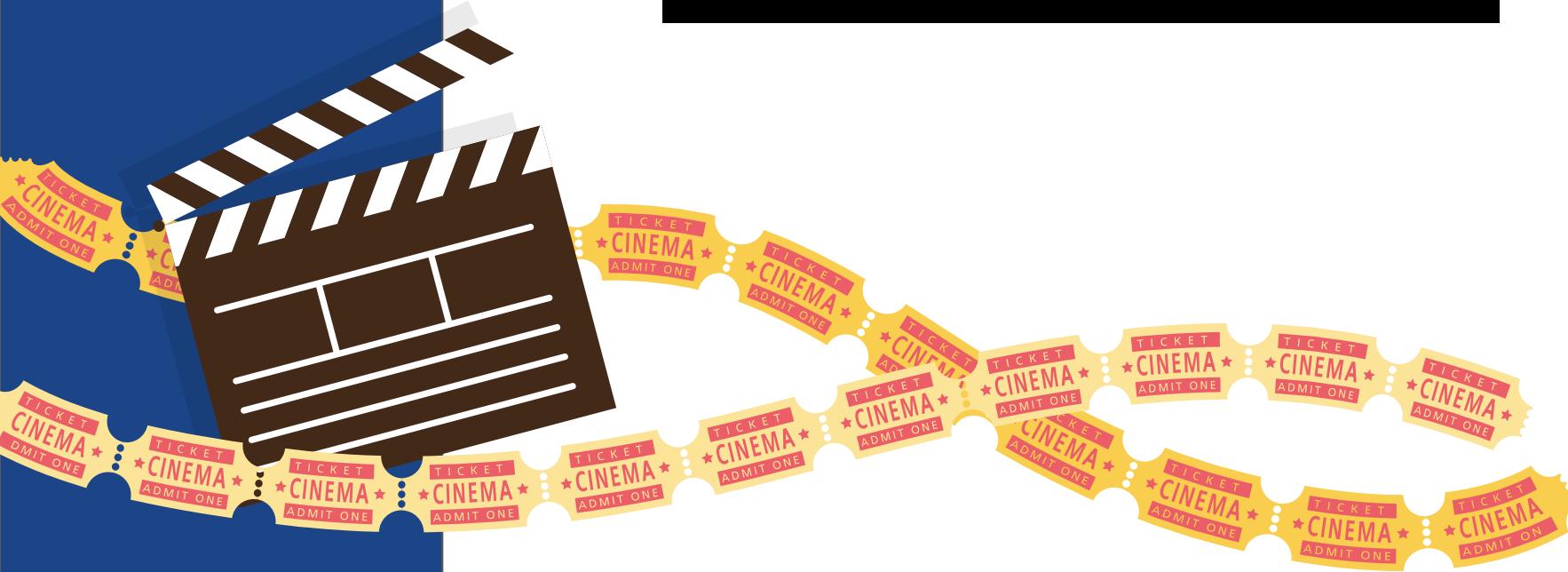


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	SCREEN																						
A	173	173	173	249	279	279	279	279	279	279	203	188	188										
B	173	173	173	173	203	203	203	203	203	203	203	188	188	188	188								
C	153	163	173	173	203	193	183	183	193	203	203	178	168										
D	233	233	233	233	443	443	443	443	443	443	443	308	308	308	308	308							
E	312	312	303	294	464	464	458	508	508	468	468	508	373	333	333	338							
F	263	263	263	307	517	456	456	536	536	536	515	515	380	380	338	338	338						
G	263	263	313	313	523	514	514	473	533	533	515	515	380	380	338	387							
H	263	263	310	310	520	520	520	529	529	529	529	379	379	382	382	338	338						
I	248	248	241	241	375	413	481	491	511	511	511	347	347	307	317	323	323						
J	418	418	368	368	438	438	445	445	443	437	437	443	383	427	427	383	383	383	383				
K	368	368	368	368	443	451	451	451	451	451	451	436	436	436	436	436	383	383	383				
L	368	368	368	368	398	398	444	446		446	446	431	431	383	383	383	383	383	435	435	435		

Analytical Dataset		genre_Action	theater_id	percent_seat_remaining_score
total_capacity	Adventure		showtime	percent_time_remain_score
reserved_seats	Animation		movie_id	seat_type_scores
time_to_showtime	Anime		final_status	sound_scores
audi_size	Comedy		available_seat	symmetry_scores
seat_type	Crime		available_ratio	zone_scores
seat_number	Documentary		max	final_popularity_score
zone	Drama		min	movie_title
sound	dayofweek		first_instance_booking	movie_release_date
symmetry	showweek		percent_seat_remaining	showdate
final_score	is_weekend		percent_time_remaining	showtime_am_pm
movie_format	is_prime		final_seat_status	final_status_score

04

MODEL BUILDING



Models Explored

01. LINEAR MODELS

02. RANDOM FOREST

03. XGBOOST



Model Validation & Performances

Training Data - 80%

Testing Data - 20%

MODEL NAME	Test/Train	MAE	MSE	RMSE	MAPE	RSQR
RidgeCV*	Train	61.45	5954	77.16	21.04	0.478
	Test	61.3	5926.07	76.98	21.07	0.475
Random Forest	Train	40.08	2190.3	46.8	13.25	
	Test	40.01	2189.1	46.78	13.25	
XGBoost	Train	7.09	142.53	11.9	2.61	
	Test	7.11	141.89	11.9	2.62	

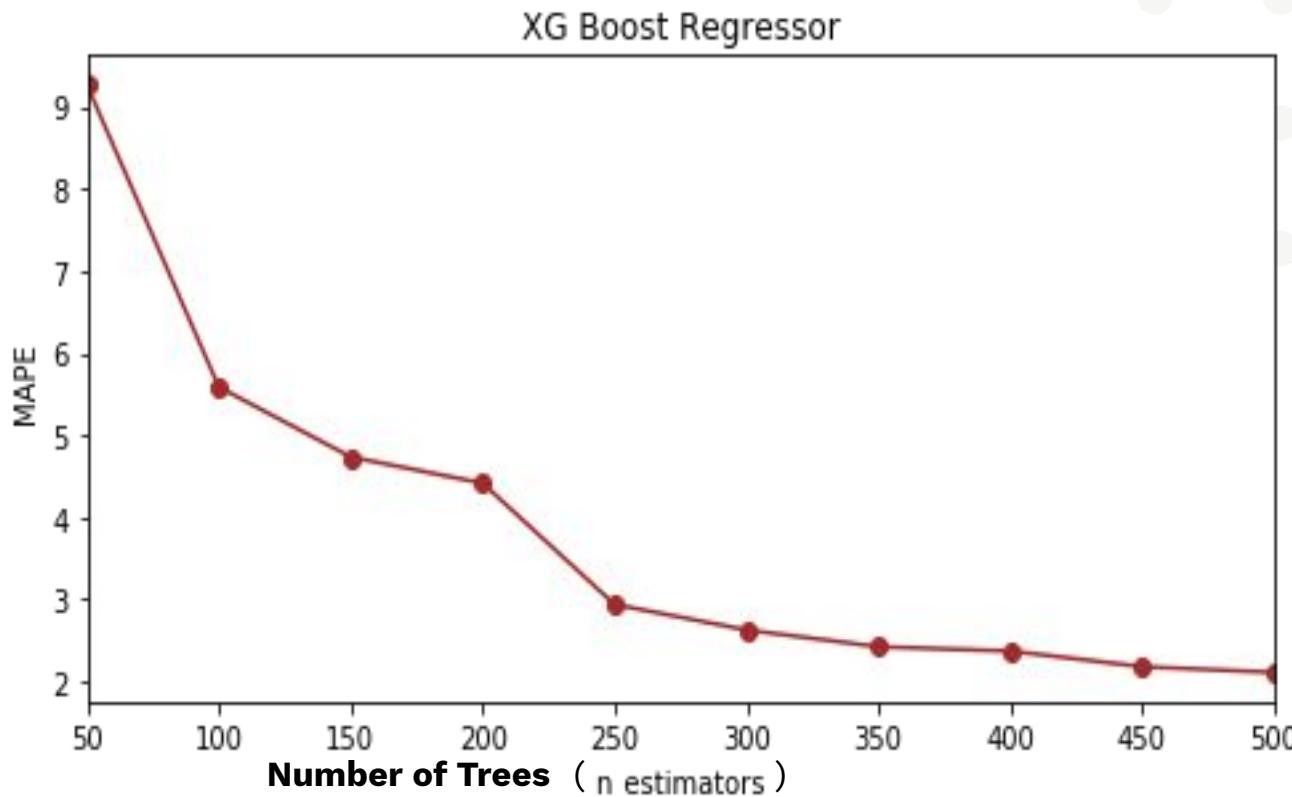
MAE: Mean Absolute Error

MSE: Mean Squared Error

RMSE: Root Mean Squared Error

RQSR: R-Square

XGBoost: Model Performance Tuning



RidgeCV Regression - Feature Coefficients

RidgeCV (built-in):

1. Cross Validation
2. Regularization
3. Normalization

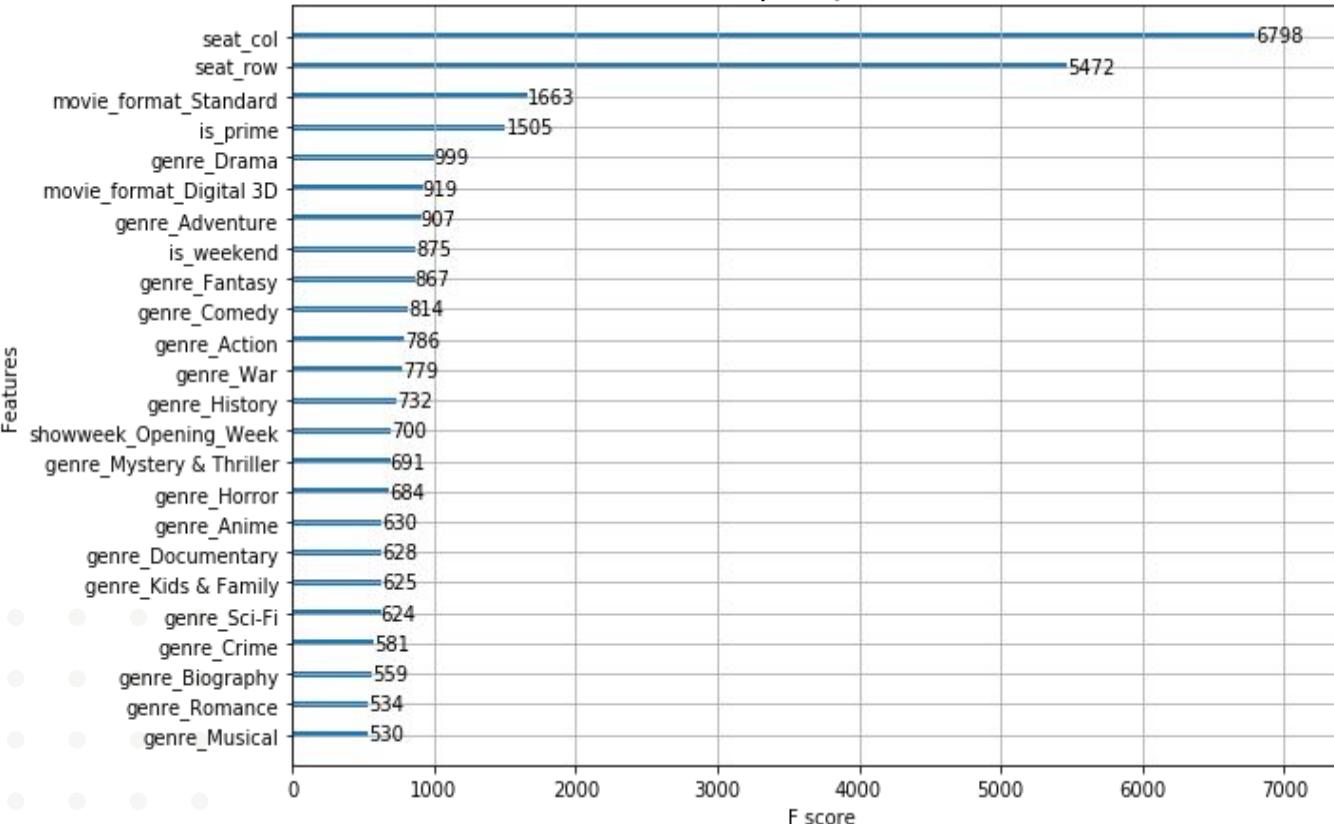
FEATURES	COEF
seat_row	18.4487
genre_Anime	7.5156
genre_Romance	6.3662
seat_col	4.5103
movie_format_Digital 3D	3.4753
genre_Biography	3.1625
genre_Horror	2.9053
movie_format_Standard	1.9016
genre_Mystery & Thriller	1.8713
genre_Kids & Family	1.8364
genre_Action	1.6982
is_weekend	0.6969
genre_War	-1.6543
genre_Documentary	-3.9652
showweek_Opening_Week	-5.9831

Positive Impact

Negative Impact

XGBoost: Feature Importance

Feature Importance (XGBoost)



Feature Importance -
RidgeCV Regression

FEATURES	COEF
seat_row	18.4487
genre_Anime	7.5156
genre_Romance	6.3662
seat_col	4.5103
movie_format_Digital 3D	3.4753
genre_Biography	3.1625
genre_Horror	2.9053
movie_format_Standard	1.9016
genre_Mystery & Thriller	1.8713
genre_Kids & Family	1.8364
genre_Action	1.6982
is_weekend	0.6969
genre_War	-1.6543
genre_Documentary	-3.9652
showweek_Opening_Week	-5.9831

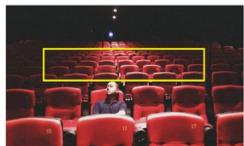
Predicting A Seat Score



Seat: I10
Enter Row: 9
Enter Column: 10



Genre: Drama & Crime



Zone: Center
Sound: Center



When: Opening Week/
Weekday

Sample Movie: Joker



XGBoost

Enter Features



Predict Score



Does Score Change with Features?

- Seat: I10
- Genre: Horror & Thriller
- Format: 3D
- When: Weekday
- **Score - 455**
- Seat: I10
- Genre: Drama & Crime
- Format: Standard
- When: Opening Week
- When: Weekday
- **Score - 466**
- Seat: I10
- Genre: Action, Fantasy, Adventure, Sci-fi
- Format: Standard
- When: Weekend
- **Score - 471**





MARKET RESEARCH APPROACH

Conjoint Analysis

- Trade-offs consumers make during purchasing process
- Rank features (based on Importance)
- Extract consumer preferences



GENRE

Horror
Action
Drama



SEAT LOCATION

Front
Center
Back



MOVIE FORMAT

Standard
IMAX 3D



Profiles/Options: 12

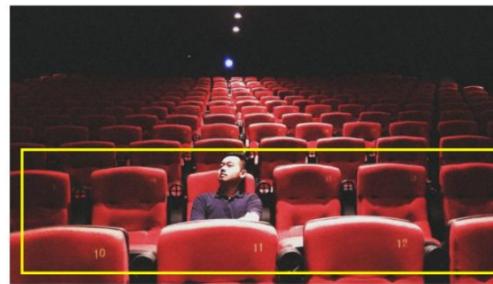
Survey Design

Option 1 *

What is your purchase preference for the following movie option?



Genre: Horror



Seat Location: Front



Format: IMAX 3D

1 2 3 4 5 6 7 8 9

Least Likely to Purchase

Most Likely to Purchase

Linear Regression (SAS)

The SAS System					
The REG Procedure					
Model: MODEL1					
Dependent Variable: Preference Preference					
Number of Observations Read					12
Number of Observations Used					12
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	40.17917	8.03583	87.66	<.0001
Error	6	0.55000	0.09167		
Corrected Total	11	40.72917			
Root MSE		0.30277	R-Square	0.9865	
Dependent Mean		5.20833	Adj R-Sq	0.9752	
Coeff Var		5.81309			
Parameter Estimates					
Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	1.65000	0.31754	5.20
Action	Action	1	2.20000	0.19149	11.49
Drama	Drama	1	1.85000	0.27914	6.63
Center	Center	1	3.40000	0.27914	12.18
Back	Back	1	3.10000	0.27914	11.11
IMAX_3D	IMAX_3D	1	-0.75000	0.21409	-3.50
					0.0128

- **All Features Significant**
- **Positive Impact:** All but one (Most: Center & Back)
- **Negative Impact:** IMAX 3D

Seat Location Driven

Part-Worths

	Description	Part-Worths	Adjusted Part-Worths	Range	Importance
	<i>Intercept</i>	1.65	1.65		
Genre	Horror	0	0	2.2	0.35
	Action	2.2	2.2		
	Drama	1.85	1.85		
Seat Location	Front	0	0	3.4	0.54
	Center	3.4	3.4		
	Back	3.1	3.1		
Movie Format	Standard	0	0.75	0.75	0.12
	IMAX 3D	-0.75	0		
SUM				6.35	1

Feature Importance Ranking:

1) Seat Location

2) Genre

3) Movie Format

Utility

Regression Equation:

$$Y = b_0 + b_1 (\text{Action}) + b_2 (\text{Drama}) + b_3 (\text{Center}) + b_4 (\text{back}) + b_5 (\text{IMAX 3D}) + e$$

Profile	Genre	Seat Location	Format	UTILITY
Option 1	Horror	Front	IMAX 3D	0.9
Option 2	Horror	Back	IMAX 3D	4
Option 3	Horror	Back	Standard	4.75
Option 4	Horror	Center	Standard	5.05
Option 5	Horror	Center	IMAX 3D	4.3
Option 6	Action	Front	IMAX 3D	3.1
Option 7	Action	Back	IMAX 3D	6.2
Option 8	Action	Back	Standard	6.95
Option 9	Action	Center	Standard	7.25
Option 10	Action	Center	IMAX 3D	6.5
Option 11	Drama	Back	Standard	6.6
Option 12	Drama	Center	Standard	6.9

Action
Center
Standard

#2

#1

#3

#12
#10

#11

Horror
Front
IMAX 3D

Share of Preference

LOGIT CHOICE RULE

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Option 9	Option 10	Option 11	Option 12
Utility	0.9	4	4.75	5.05	4.3	3.1	6.2	6.95	7.25	6.5	6.6	6.9
Choice	0%	1%	2%	3%	1%	0%	9%	18%	24%	12%	13%	17%



Option 9
+ **Option 8**
+ **Option 12**
=
59% of Total

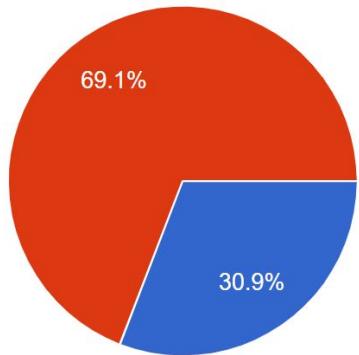
Option 8	Action	Back	Standard
Option 9	Action	Center	Standard
Option 12	Drama	Center	Standard

Student vs Non-Student Behavior

1. Are you currently a student?

256 responses

- Yes
- No



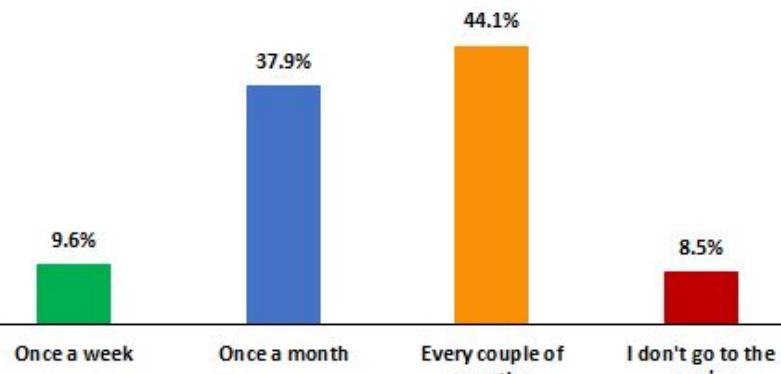
	Description	COMBINED	NON-STUDENTS	STUDENTS
	<i>Intercept</i>			
Genre	Horror	0.35	0.34	0.27
	Action			
	Drama			
Seat Location	Front	0.54	0.62	0.61
	Center			
	Back			
Movie Format	Standard	0.12	0.04	0.11
	IMAX 3D			
SUM		1	1	1

3. Prior to COVID-19, how often did you visit the movies?

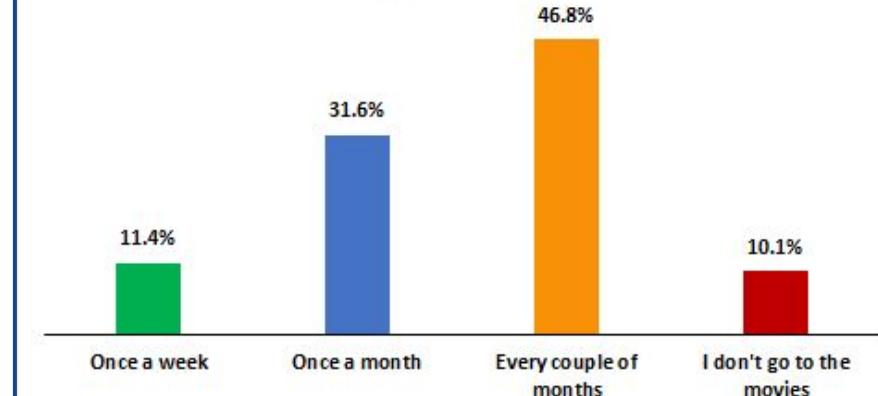
Non-Student = 177

Student = 79

**Non-Student: Movie Visit Frequency
(Pre-COVID)**



**Student: Movie Visit Frequency
(Pre-COVID)**



Similar Behavior

4. Which best describes your feelings about returning to the movies if theaters are following social distancing guidelines?

Non-Student: Movie Visit Sentiment
(Post-COVID)



■ I can't wait to go back! ■ Undecided ■ I won't be returning any time soon

Student: Movie Visit Sentiment
(Post-COVID)



■ I can't wait to go back! ■ Undecided ■ I won't be returning any time soon

06

CONCLUSION



Data Science & Market Research

- XGBoost Best Predictive Model
- Performance Consistent For Different Theaters
- Seat Location & Genre Have Biggest Impact

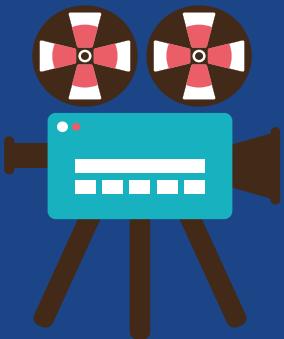
VALIDATED BY BOTH APPROACHES

- Seat Location Have Biggest Impact
- Most Preferred Zone: Center & Back
- Least Preferred Zone: Front
- Preferred Movie Format: Standard

Option 9 [Action, Center & Standard]
Predicted Seat Score = **464**

Consumer Behavior = Consumer Preference
(Data Science) (Market Research)





Limitations & Future Scope

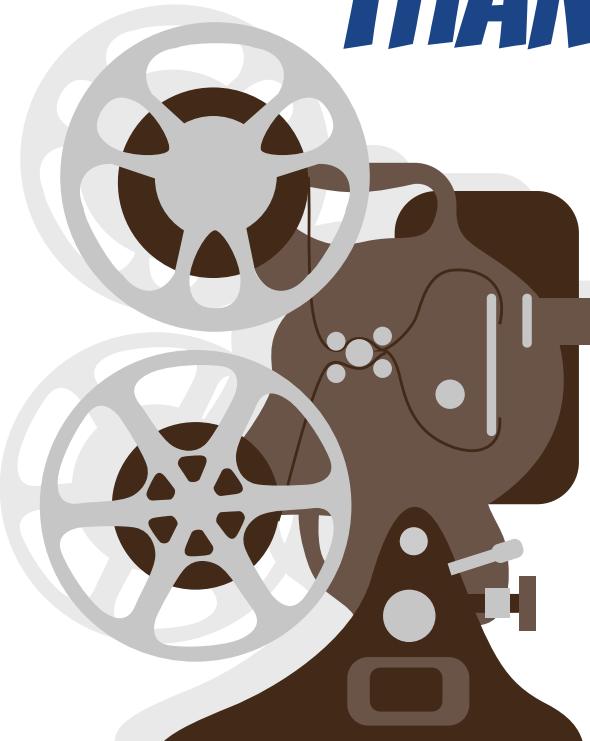
- Framework: Model Not Perfected
- Remote Work/COVID
- Big Data
- Seat Map Accuracy
- No Auditorium IDs



MSBA: Course Knowledge Used

- **Capstone Workshops (Lontok):** Python (Big Data) & Project Management (Jira)
- **Data Modeling/Business Intelligence(Seal):** Python, Advanced Excel & Visualization
- **Data Visualization (Vo):** Tableau
- **Machine Learning (Brahma):** XGBoost, Random Forest, Feature Importance
- **Programming (Mesgari):** Python (SKLearn, Pandas, Matplotlib)
- **Strategic Integrated Analysis (Namin):** Conjoint Analysis & SAS
- **Marketing Analytics (Tang, Haley):** Survey Design

SPECIAL THANKS



ARIN BRAHMA
SHARETH HARIHARAN
LEO ANGHELESCU
REETO MOOKHERJEE
INDRANEEL SHEOREY
GREG LONTOK
AIDIN NAMIN
KALA SEAL
LINDA LEON
SIJUN WANG
MSBA FACULTY & CLASSMATES

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.