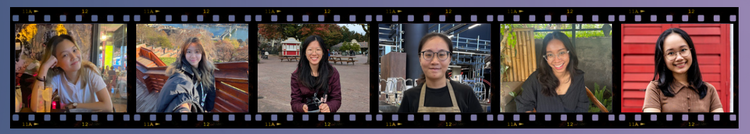
BOX OFFICE FORECASTING & MOVIE SESSION OPTIMIZATION

ACCT414 (SMU-X) GROUP 2
ACCOUNTING ANALYTICS CAPSTONE

CATHAY CINEPLEXES



PROBLEM STATEMENT

How can Cathay leverage their data to better forecast box office demand and revenue, thus optimising session scheduling and aid negotiations with film distributors?

OUR OBJECTIVE

Develop prediction models that:

- Gives deeper insights to **trends**
- **Forecast revenue** through movie attributes
- Aids in **optimising session scheduling**

OUR APPROACH

- Identify **key factors** affecting movie attendance
- Forecast box office revenue and admissions using **machine learning**
- Optimize scheduling with **goal programming** and data **visualization**.

1 FORECASTING NET BOX OFFICE : REGRESSION MODEL

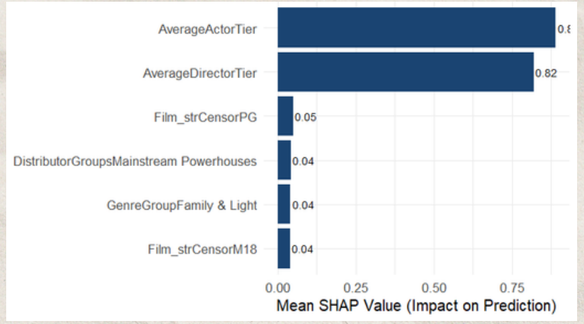
REVENUE REGRESSION

OBJECTIVE:
Predict a film's expected box office revenue & range before release

MODEL INPUTS:

- Actor & Director Tiers *(historical performance)*
- Distributor Group *(K-means on performance & genre mix)* etc

TOP PREDICTORS



PREDICTION EXAMPLE

Feature	Input
Genre Group	Horror&Supernatural
Actor Tier	6
Director Tier	1
Censorship Rating	M18
Language Group	Mainstream
Distributor Group	Mainstream Powerhouse
Release Month Tier	1
Predicted Revenue	**$1,041.93**
Estimated Range	**$403.56 - $2,687.63**

TAKEAWAYS & RECCOMMENDATION

✓ Net box office regression provides some basis for **negotiation**

✓ Admits regression and Optimization model generates predicted weekly demand & allocates **x no. of sessions** for each screen in a week

✓ Data visualisation provides a guideline on popular time slots for **movie scheduling**

✓ Our models incorporates **flexibility** for long term usage

Incorporate media spend, pre-release hype metrics to enhance predictive power

2 OPTIMIZE SESSION SCHEDULING : MACHINE LEARNING MODELS & DATA VISUALISATION

ADMITS REGRESSION

OBJECTIVE:
Provide an estimated weekly demand for a specific movie (used as inputs for the optimisation model)

Category	Specify Input/Values
Previous Week's Admits	0
Weeks since released	0
Genre Group	Drama & Realistic
Language Group	Major Languages
Film Actor Tier	1
Director Tier	1
Distributor Tier	1
Month Tier	1
Holiday week?	1
Cinema branch	5
Predicted Admits	470

OPTIMIZATION MODEL

MODEL CONSIDERATION

- Predicted demand of a movie (from Admits Regression)
- Occupancy rate of each screen
- Number of sessions targetted for each week

OUTPUT EXAMPLE

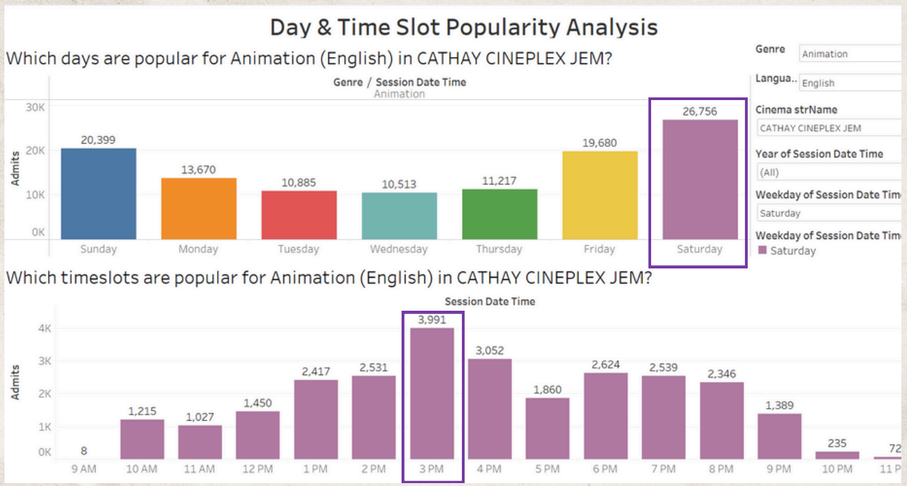
Predicted Demand	Total No. of Sessions	Screen 1	Screen 2
54	3	0	3
104	6	0	6
187	10	7	3
256	14	7	7
294	16	7	7

RESULTS

No. of sessions allocated to each screen in a week for each movie

DATA VISUALISATION

OUTPUT EXAMPLE:



RESULTS

Saturday 3pm is the most popular time slot for an animation english movie at Jem

1 FORECASTING NET BOX OFFICE : SUPPLEMENTARY

REVENUE REGRESSION

MODEL INPUTS:

- Actor & Director Tiers *(historical performance)*
- Distributor Group *(K-means on performance & genre mix)*
- Censorship Rating
- Grouped Genre, Language & Month *(trends & domain knowledge)*

APPROACH

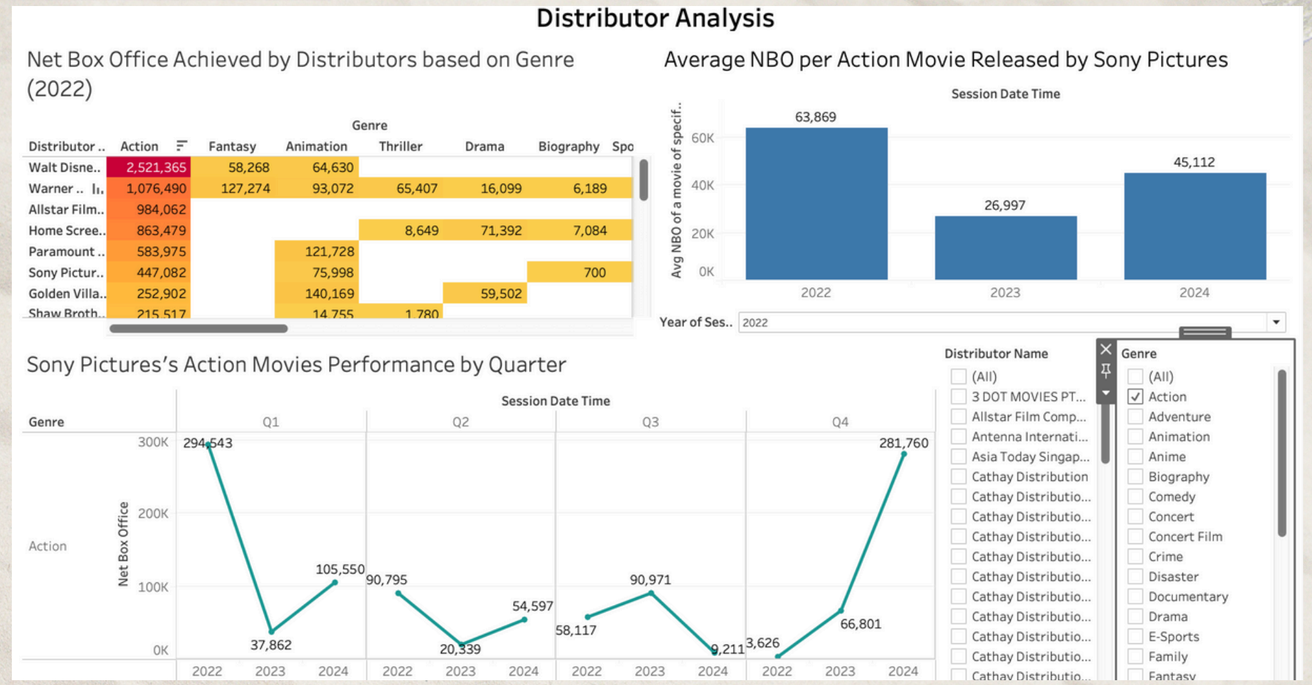
- Select film attributes (genre, actor tier, etc.)
- Predicts box office revenue before release
- Revenue range is provided to reflect uncertainty
- Use results + dashboards to guide planning or distributor negotiation

XGBoost model explains ~83% of box office variations. On average, predictions deviate by ~36% (RMSE = 0.857) from actual outcomes. Prediction ranges are based on the model's error, reflecting the unpredictability of film performances.

MODEL PERFORMANCE

PredictedRevenu	Lower Estima	Upper Estima
$ 234.32	$ 90.28	$ 605.65
$ 234.32	$ 90.28	$ 605.65
$ 150.24	$ 57.67	$ 388.89
$ 234.32	$ 90.28	$ 605.65

DATA VISUALISATION



2 OPTIMIZE SESSION SCHEDULING : MACHINE LEARNING MODELS & DATA VISUALISATION

ADMITS REGRESSION

APPROACH

- 2 prediction models:
  - Predicting first week admits
  - Predicting subsequent weeks admits
- Past admits, movie-specific attributes and time-related features used in predicting admits
- Specify inputs to obtain predicted admits

OPTIMISATION MODEL

	No. of admits
Low-Tier Movies	< 300 admits
Mid-Tier Movies	300 - 1000 admits
High-Tier Movies	> 1000 admits

3 Optimisation Models for different movie tiers

Predicted Demand	Total No. of Sessions	Screen 1	Screen 2	Screen 3	Screen 4	Screen 5	Screen 6
54	3	0	3	0	0	0	0
104	6	0	6	0	0	0	0
187	10	7	3	0	0	0	0
256	14	7	7	0	0	0	0
294	16	7	7	1	0	1	0

NOTES:

- Occupancy rates for each screen in each model calculated from past data
- Flexibility of Model :
  - Different weights are allocated to the different screens
  - Multiplier for occupancy rates
  - Target number of sessions in a week / each screen