

Crossroad Analytics Challenge

Final Presentation
Team Anonymous



COLTS

Meet the Team



Akhil Venu Gopal

Akhil is from Kozhikode, India, and has an undergraduate degree in Electronics and Communication engineering. He worked as a Software Engineer for 4+ years and is currently pursuing a Master's degree in Data Science at Indiana University



Dheeraj jeevagasami

Dheeraj is from Chennai, India, and has an undergraduate degree in Computer Science engineering. He worked in the Data Science field for over a year and is now pursuing a Master's degree in Data Science at Indiana University



Sasidev Mahendran

Sasidev is from Coimbatore, India, and has an undergraduate degree in Production engineering. He has worked as a Dimensional Variation Analysis Engineer for 3 years and is now pursuing a Master's degree in Data Science at Indiana University



Sumitha Vellinalur Thattai

Sumitha is from Chennai, India and has an undergraduate degree in Electronics and Communication engineering. She worked in the Data Science and Analytics field for 4 years and is now pursuing a Master's degree in Data Science at Indiana University

Agenda

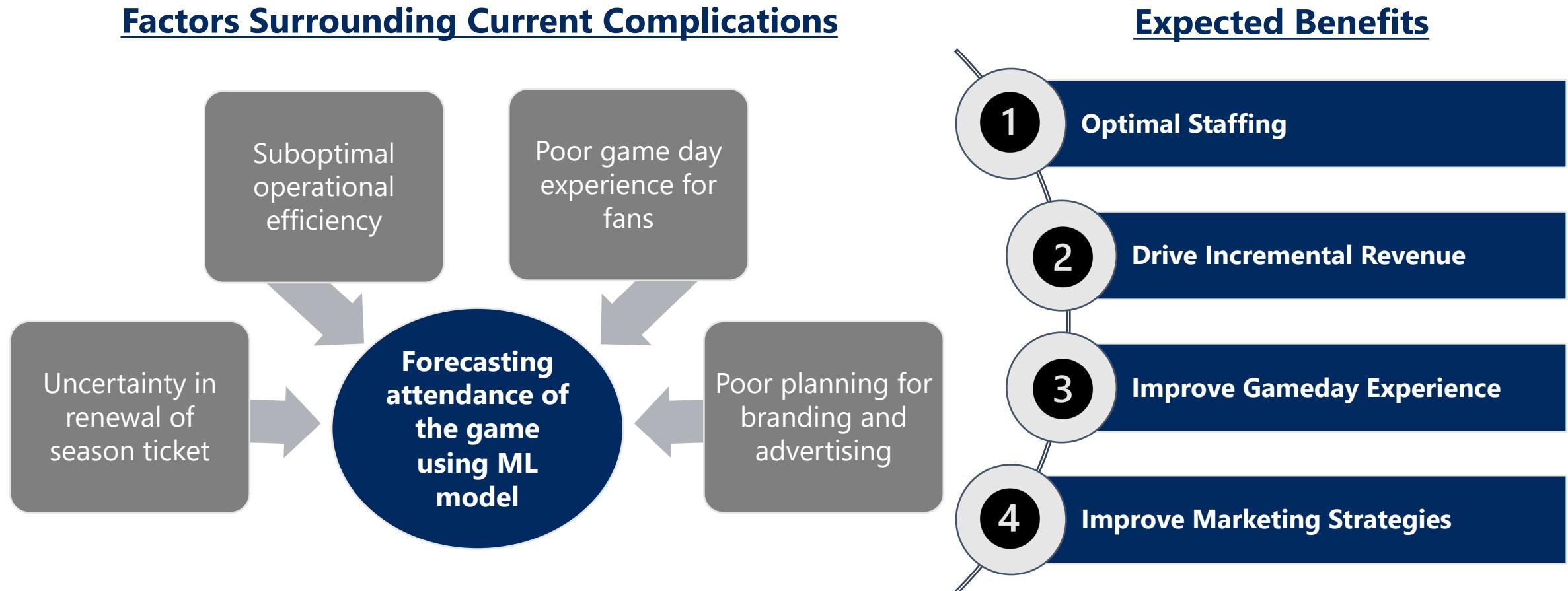
- 1 Current Situation
- 2 Insights
- 3 Solution
- 4 Conclusion
- 5 Product Demo

Lucas oil stadium team struggles to predict the attendance as the pattern is very inconsistent

- Lucas Oil Stadium has a **capacity of 67,000** expandable to over 70,000 with about 250 sections
- There were about 18 events in the last 2 years, with an **average attendance of 77%**
- **Attendance** of each game **fluctuates** rapidly and depends on various factors
 - Min attendance - 49%
 - Max attendance - 89%



COLTS is looking to understand the ticket purchase and attendance pattern for a game day



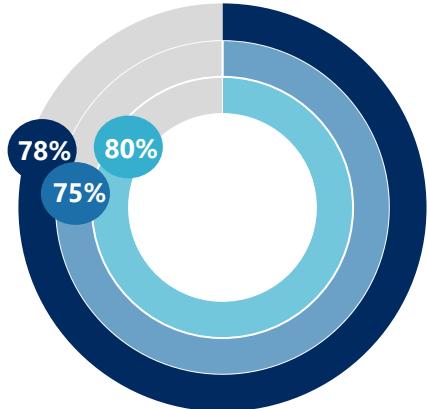
Understanding the important factors driving attendance and taking steps based on the predicted occupancy will help us yield the expected benefits

Insights



Optimizing staffing based on occupancy of a section is recommended; encouraging reselling of tickets can drive incremental revenue

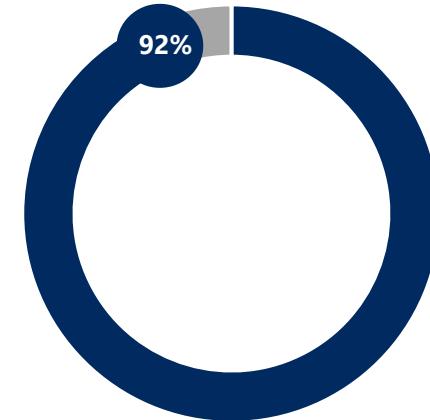
Overall Attendance



There were 18 events in **Total**, with 9 events in **2021** and **2022** each

■ Attended ■ Not Attended

Reselling Ticket Attendance



- 12% of all the tickets are **resold** in the secondary market
- About **50% of house members** tend to resell their tickets

■ Attended ■ Not Attended

- **Outer sections** that are far away from the view are **less likely to be attended** vs. other sections
- **Row number 443 - 446** are **not attended** in any of the matches

Majority of **tickets resold** on secondary market were **attended** by people making it an important predictor

Season ticket holders attending a game are more likely to renew; predicting attendance and appropriate marketing campaigns can help increase season ticket renewals

Tickets by Account Type



About 85% of all season account type attend the game



Season Ticket Renewal

- **Seasonal** ticket holders have the **highest attendance** ratio among all account types

Seasonal Ticket holders 2021

Audience attending > 50%

46% Returning audience in 2022

54% Non-returning audience 2022

Audience attending less than 50%

15 % Returning audience

Forecasting attendance will drive incremental revenue and help improve the gameday experience of fans



Machine learning algorithm to accurately predict the number of actual attendees for Colts games based on a variety of factors

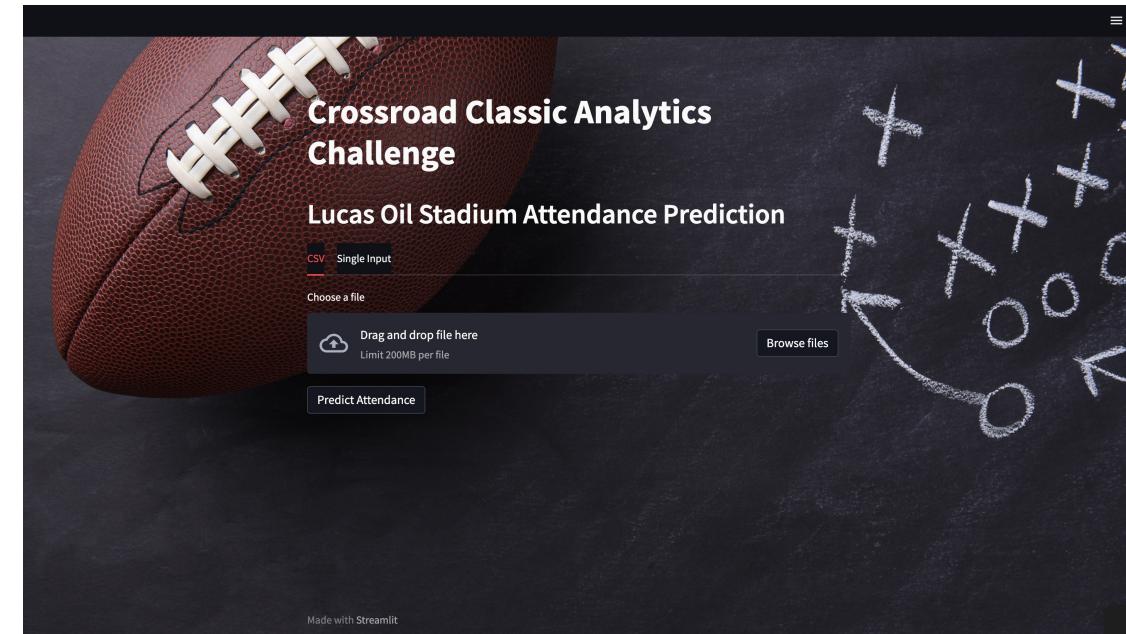


Streamlit dashboard for simple and ease of use for any intended user



Recommendations to improve fan's experience in a game

Product Webpage





Solution – Our Product

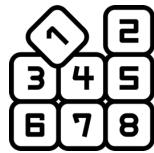


Pruning of existing features and adding derived features helped improve the performance of the model



DATE FEATURES

- Difference in days between event date and sale date
- Features based on Event date:
 - Week number
 - Weekday
 - Month
 - Month end



NUMERICAL FEATURES

- Log transform
 - Price
 - Tenure



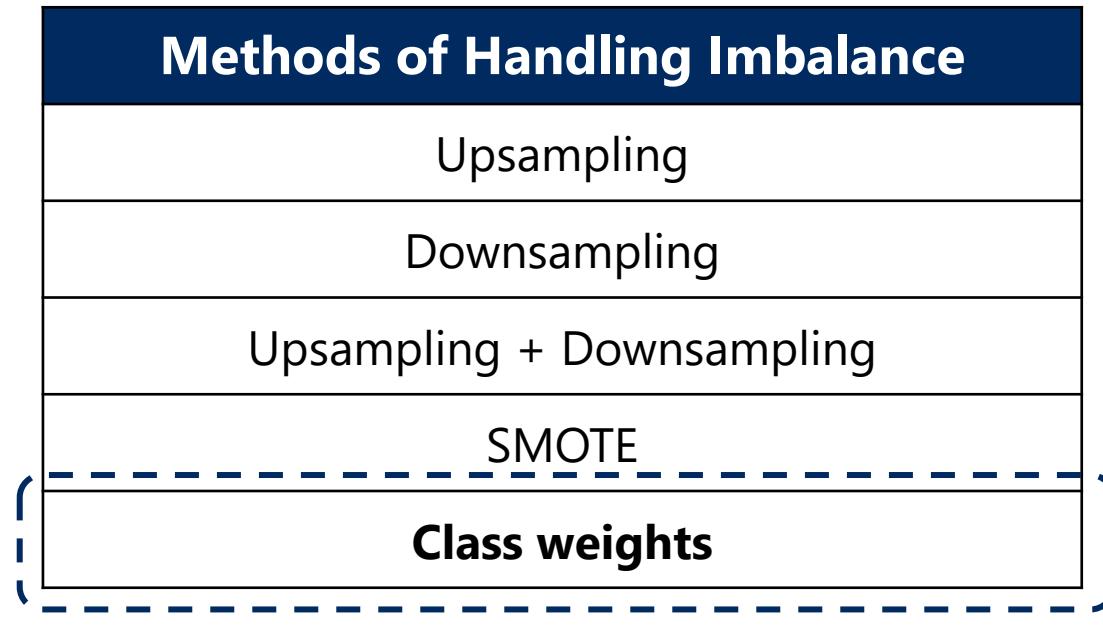
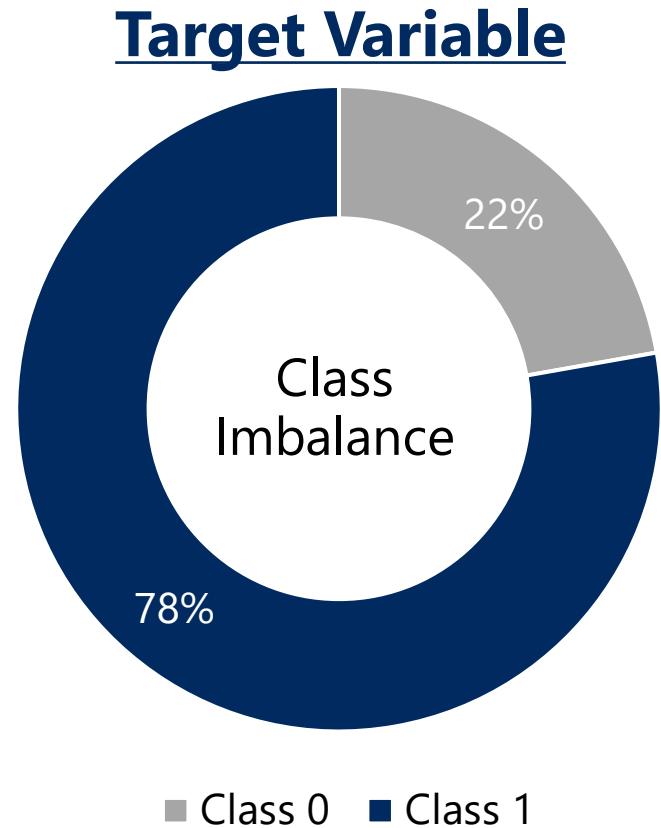
CATEGORICAL FEATURES

- Grouping of section name based on Colt's stadium (price)
- Grouping of ticket type and account type description

Adding new features helped improve the product performance

After **initial feature engineering**, we had **29 features** (5 numerical and 24 categorical)

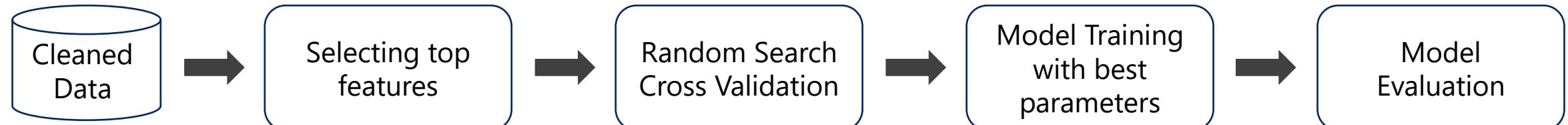
Handling class imbalance using class weights without increasing the size of the data works best for our model



Proposed Method - Using Class Weights

Gives more weightage to minority class;
Penalizes misclassification of minority class

XGBoost model trained with top features achieved the highest performance



Models Used:

- Bagging - Random Forest
- Boosting - XGBoost , CatBoost, LGBM



Metrics for Evaluation:

- AuC, Precision, Recall, F1-Score

Proposed Method:

XGBoost with class weights achieved the **best performance** on both training and test set

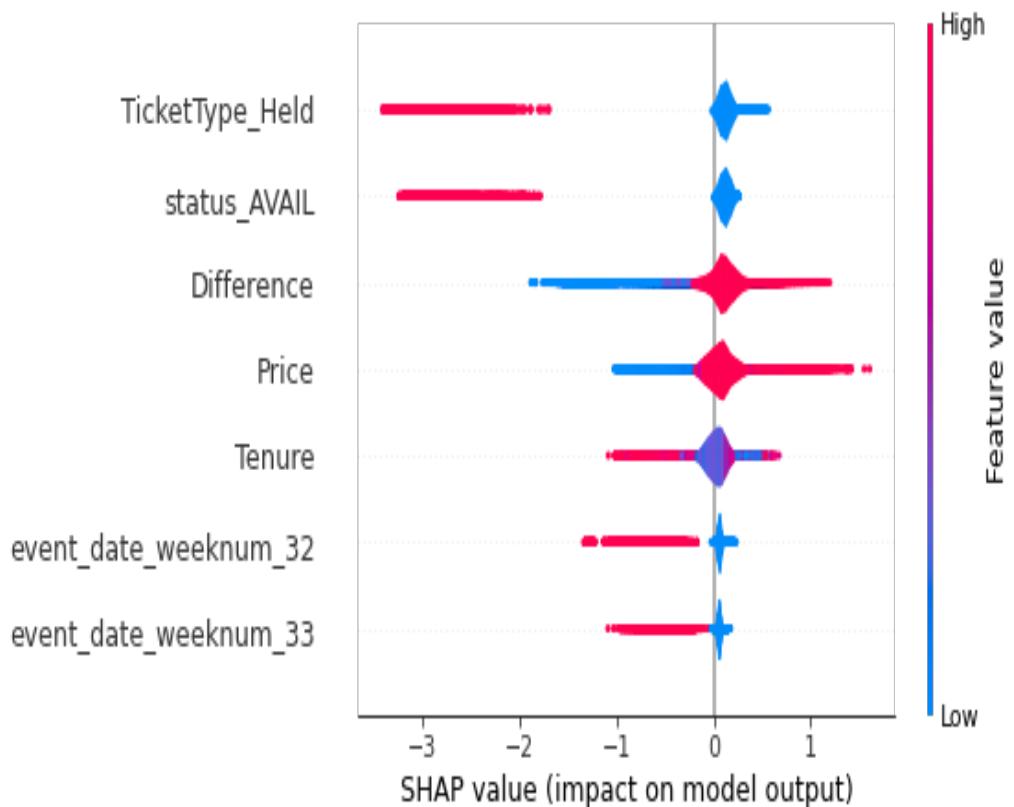


Model	Accuracy		F1-Score	
	Train	Test	Train	Test
XGBoost	0.850	0.840	0.907	0.901

Derived features – difference in event date & ticket selling date and week number of the event were among the top features impacting prediction

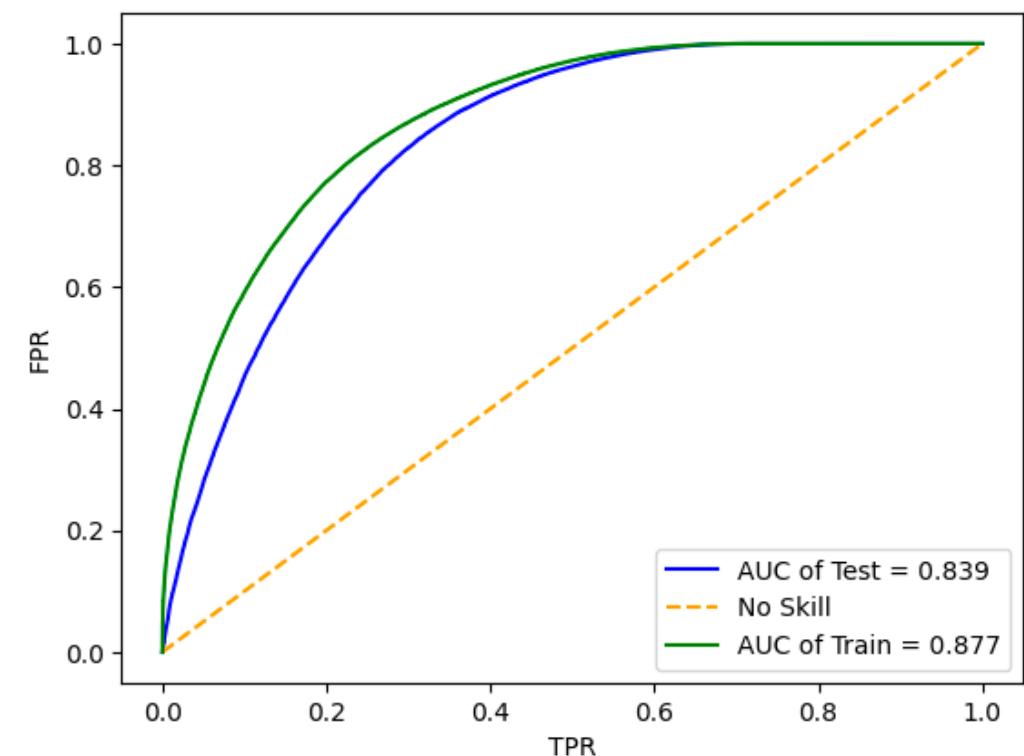
Model Explainability – SHAP Values

SHAP values gives the **contribution of each feature**, either positively or negatively on the prediction of the model



AuC-ROC Curve

AuC scores of training and test set are **similar** indicating the model **did not overfit**



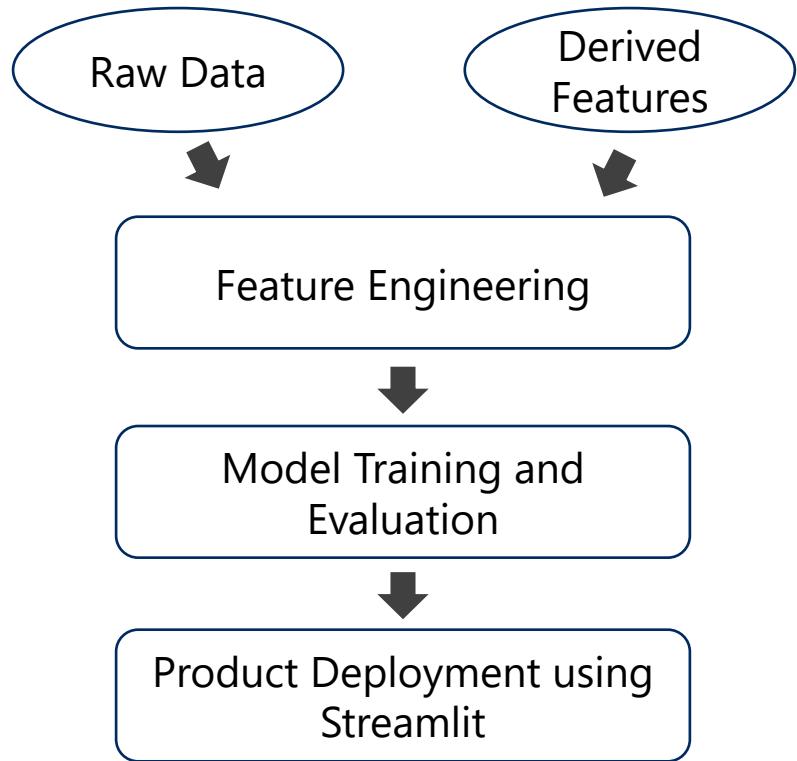


Conclusion



The product has multiple benefits to help COLTs team better understand erratic attendance and plan ahead for the game day

Solution



XGBoost model trained using important features and class weights was able to accurately predict ~80% of the data correctly

Benefits

	Incremental Revenue	Through food and beverage, merchandise sales
	Better Gameday Experience	Plan ahead for better staffing and operational efficiency
	Ease of Use	Simple tool that can be used by anyone

Increase likelihood of season ticket renewal

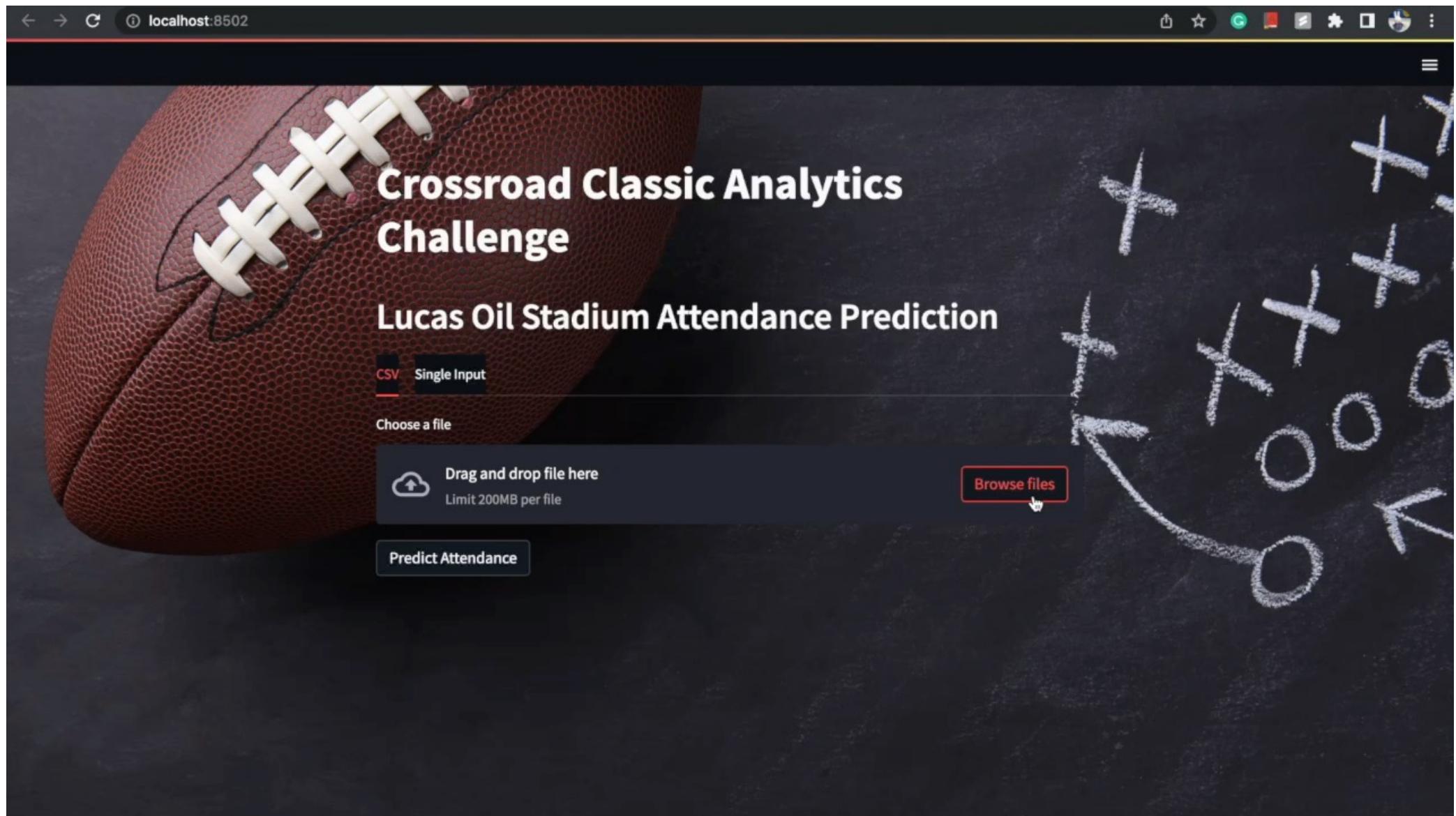
Future Scope

Improve model performance by adding additional information about match start time, weather condition, win-loss record of COLTS before a match

Product Demo



Product Demo Video



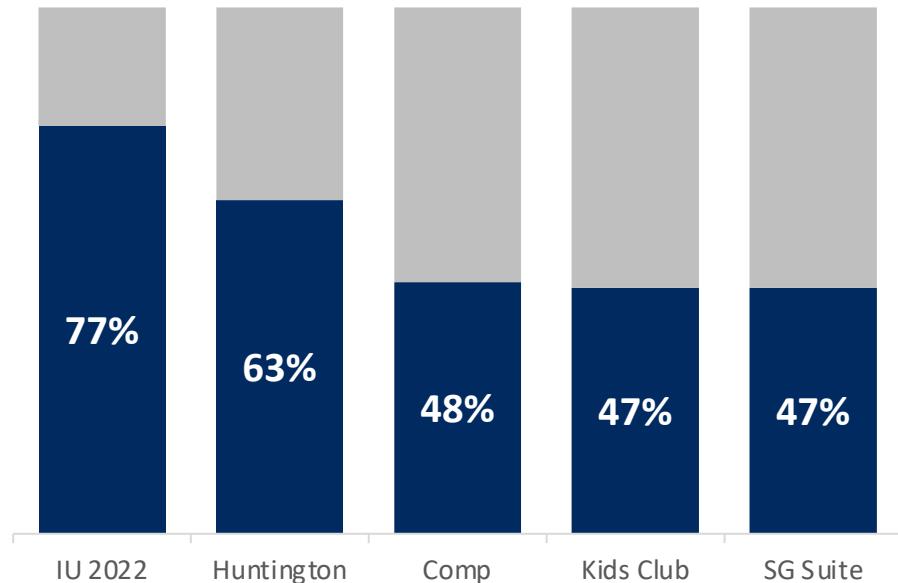


Appendix



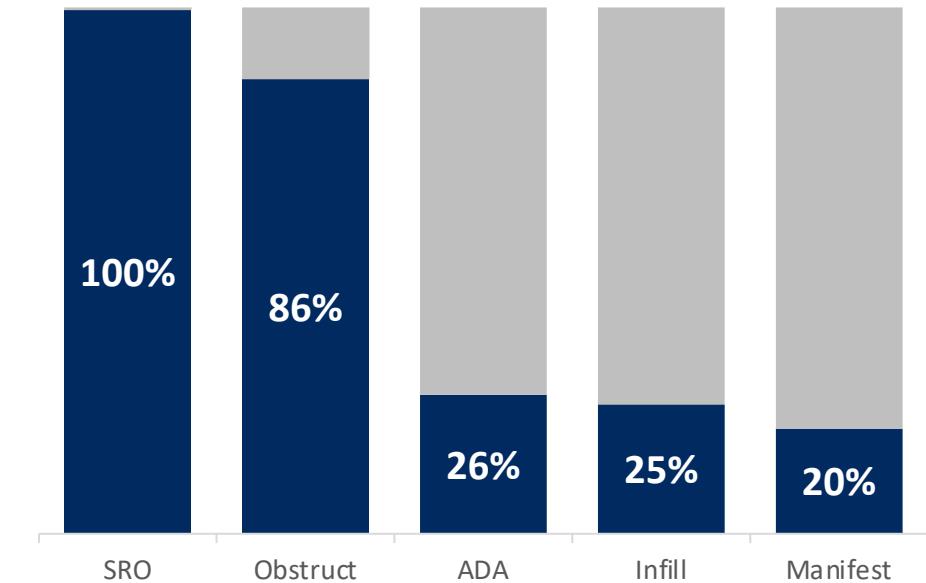
Insights

Account types - IU and Huntington were more likely not to attend the match compared to other description types



■ Not Attended ■ Attended

Majority of SRO and obstruct ticket class are not being occupied in the COLT's stadium making it an important predictor for our model

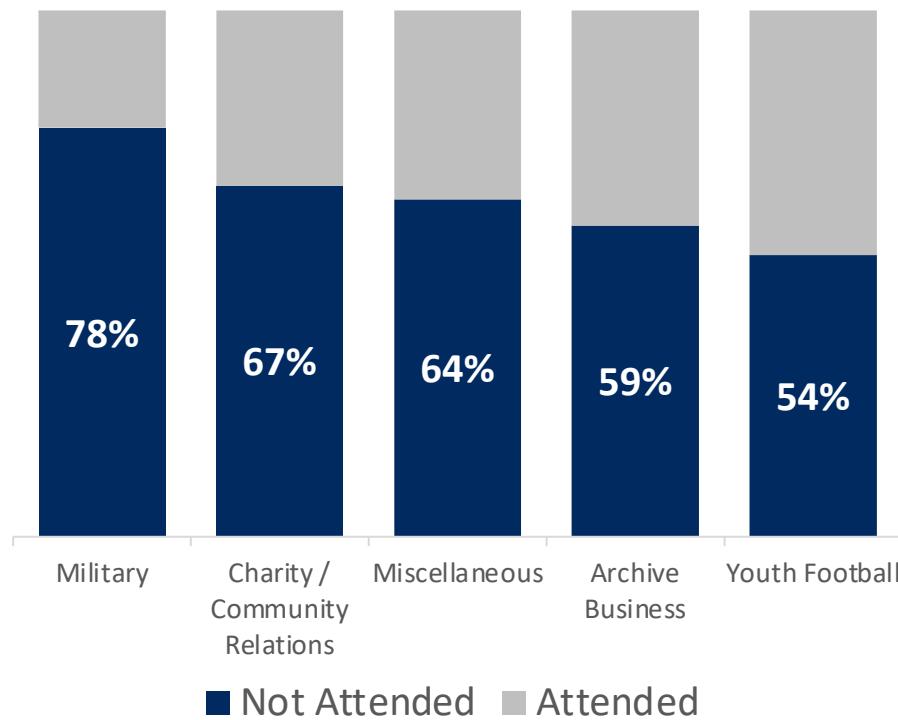


Outer sections that are far away from the view are less likely to be attended vs. other sections

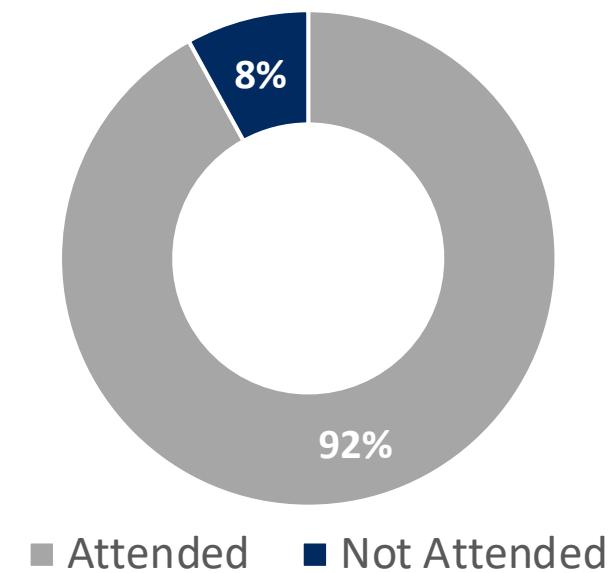
■ Not Attended ■ Attended

Insights

Major **complimentary ticket categories** of people who **don't attend** the games are Military personnel, Business professionals, and members of the Youth Football Club



Majority of **tickets resold** on secondary market were **attended** by people making it an important predictor

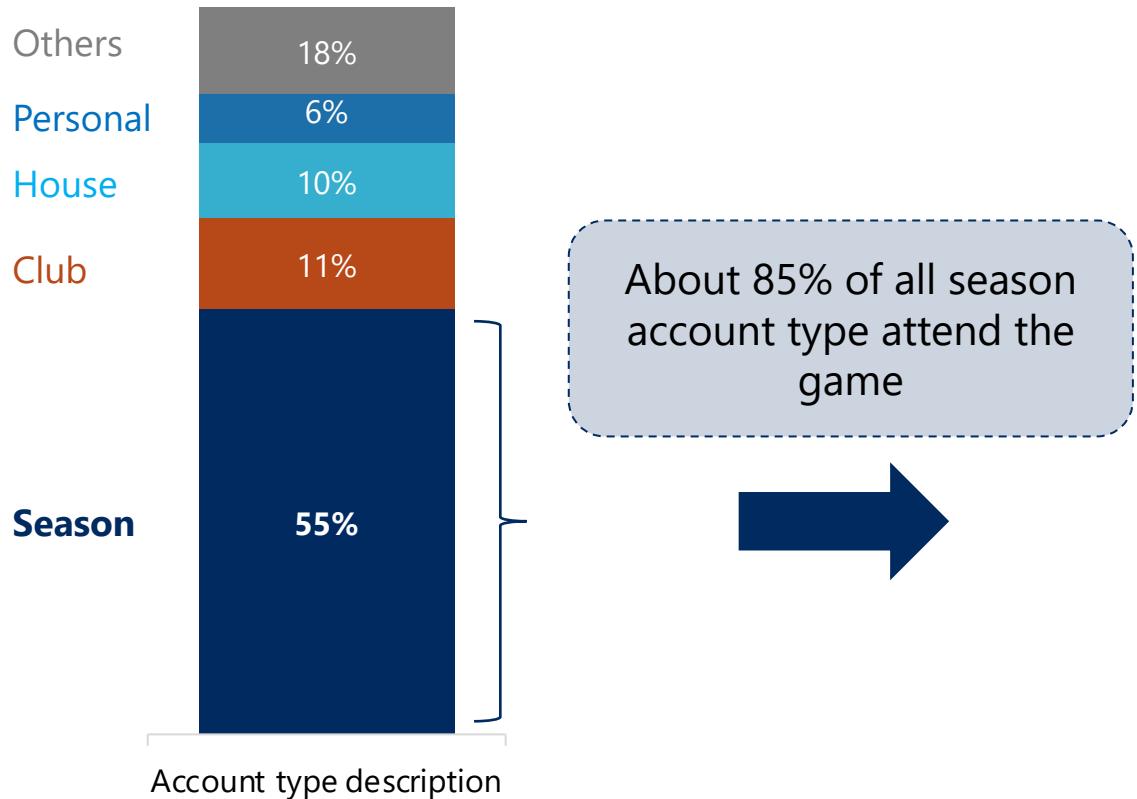


- Overall, 78% of people attended the game
- 12% of tickets are resold

Row number 443 - 446 are not attended in any of the matches

Season ticket holders attending a game are more likely to renew; predicting attendance and appropriate marketing campaigns can help increase season ticket renewals

Tickets by Account Type



Season Ticket Renewal

- **Seasonal** ticket holders have the **highest attendance** ratio among all account types
- About 46% of **seasonal ticket holders** who attended more than half games in 2021 **returned to a game** in 2022 as well
 - Of the season ticket holders who **did not attend more than half the games** in 2021, only 15% returned to watch a game in 2022

References

- https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html
- <https://xgboost.readthedocs.io/en/stable/parameter.html>
- <https://www.lucasoilstadium.com/tours-stadium-info/>
- <https://docs.streamlit.io/>