Breaking the ice: Understanding hockey fans through survey data

Kawnain Ahmed Chowdhury Anna Goncharova Qian Zhong Yixin Li

Data

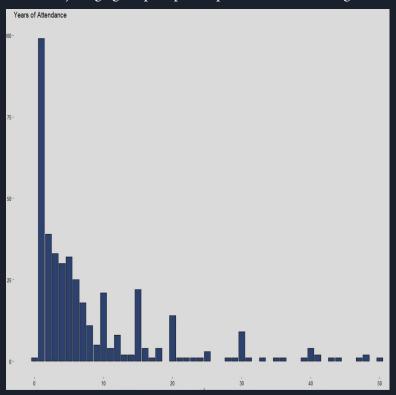
- Raw data: Hockey Survey 2019
 - 516 Observations & 62 Variables
- Cleaned Data:
 - Logistic Regression and xGBoost
 - 414 Observations & 46 Variables
 - Dummy Variables created for both models
 - Sentiment Analysis
 - Removal of stopwords
 - Usage of regular expressions
 - Stem words

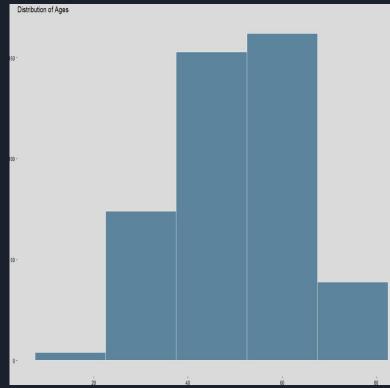


Background Demographic Analysis

The survey participants are following Notre Dame ice hockey for < 8 years.

The major age group of participants are between age 40 - 60.



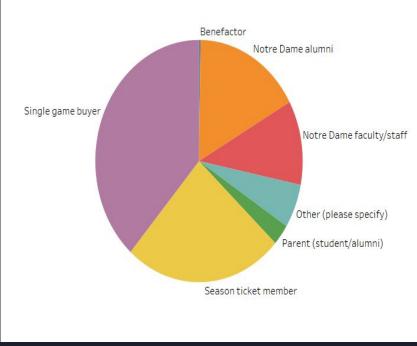


Background Demographic Analysis

Most of the survey participants are from Midwest.

The participants were mostly single game buyer, season ticket member and Notre Dame alumni.





Sentiment Analysis



- Used comments from the survey
- 57 negative comments sentiment comments
- Focus words:
 - Concessions
 - Parking
 - Seats

```
-0.52632652 the bathrooms there was many, but they were dirty. I was not happy that there was no beer available
-0.50000000 Your concessions are terrible
-0.47067872 The concession staffing, menu and availability of the menu don't come close to matching the quality of the venue or the quality elsewhere at Notre Dame.
-0.45786216 Was very disappointed with our seats for the outdoor game. As season ticket holders I felt we should have had alot better seats then where you sat us in section 7.
-0.42137017 Concessions and the souvenir shop had UNMANAGEABLE LINES. Terrible, and it carried over to the Winter Classic the next day. Embarrassing.
```

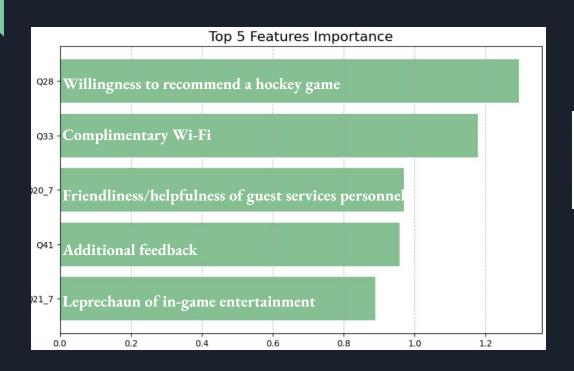
Logistic regression

Dependent variable: Q42(Would you be interested in tickets to future hockey games at Compton Family Ice Arena?)

Data: NA still exist after cleaning, replaced by mean value.

Target: Get the top five variables ranked by importance.

Logistic regression



Evaluation of the model

Accuracy: 0.9588377723970944

Precision: 0.9629629629629

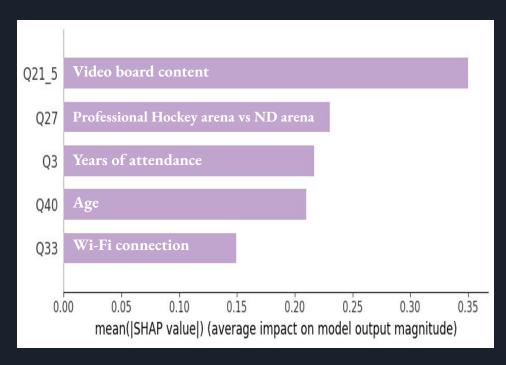
Recall: 0.9948979591836735

F1-score: 0.9786700125470513

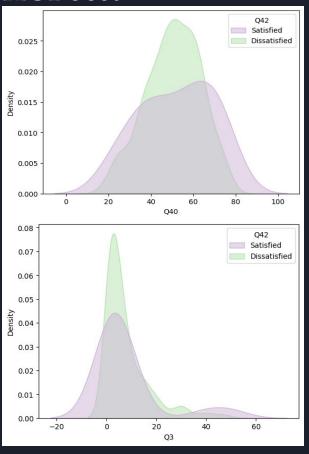
XGBoost

- Main variables driving dissatisfaction:
 - Video Board Content
 - Arena comparison
 - Age of attendee

Feature Importance based on SHAP



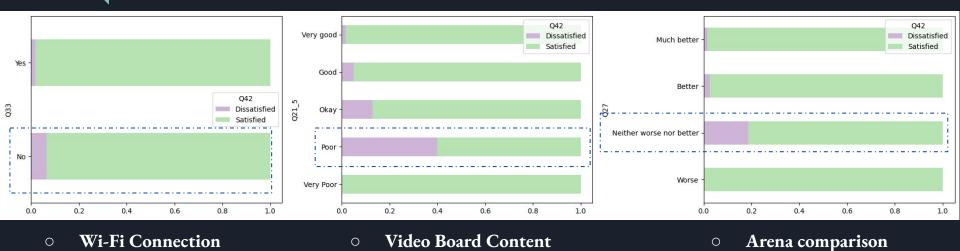
XGBoost

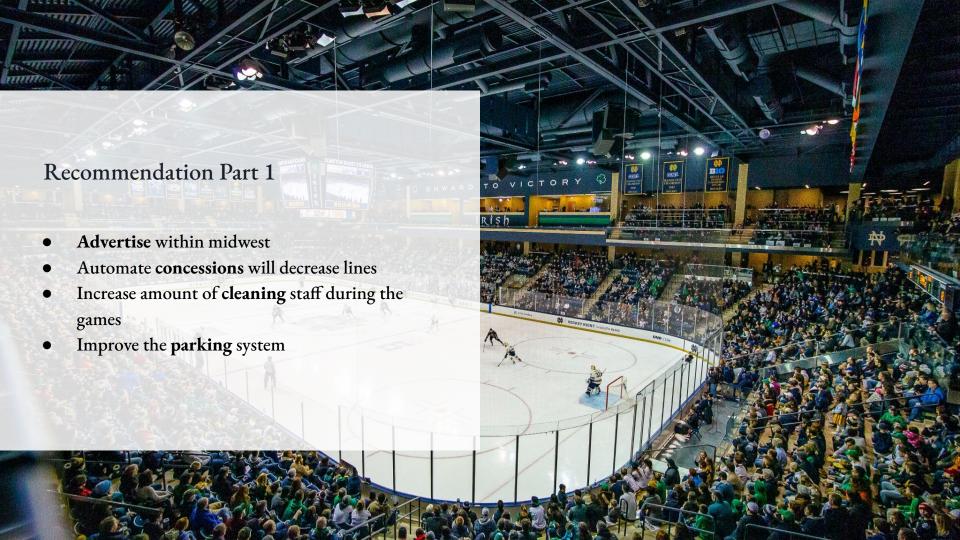


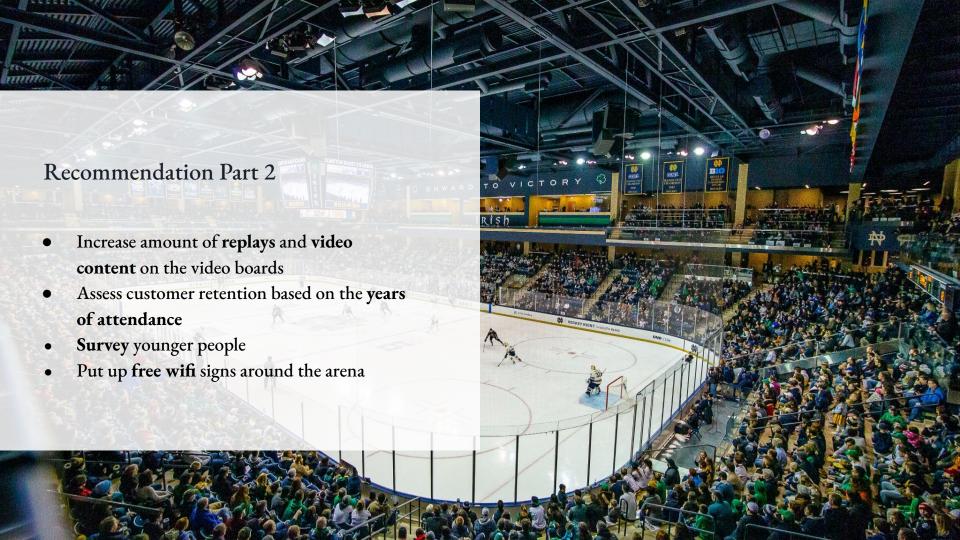
• Age of attendee

• Years of attendance

XGBoost







Questions?