Avalanche Season Ticket Renewals

Jeff Thomas
General Assembly
DSI CC9

Kroenke Sports & Entertainment

Problem Statement

- The purpose of this project is to create a classification model to predict whether a current Avalanche season ticket holder will renew for next season. I will also aim to identify some of the crucial factors that play into season ticket renewal. I will measure success of the model against the baseline (majority class) of 79.5%.

  - Data was acquired courtesy of Kroenke Sports & Entertainment.

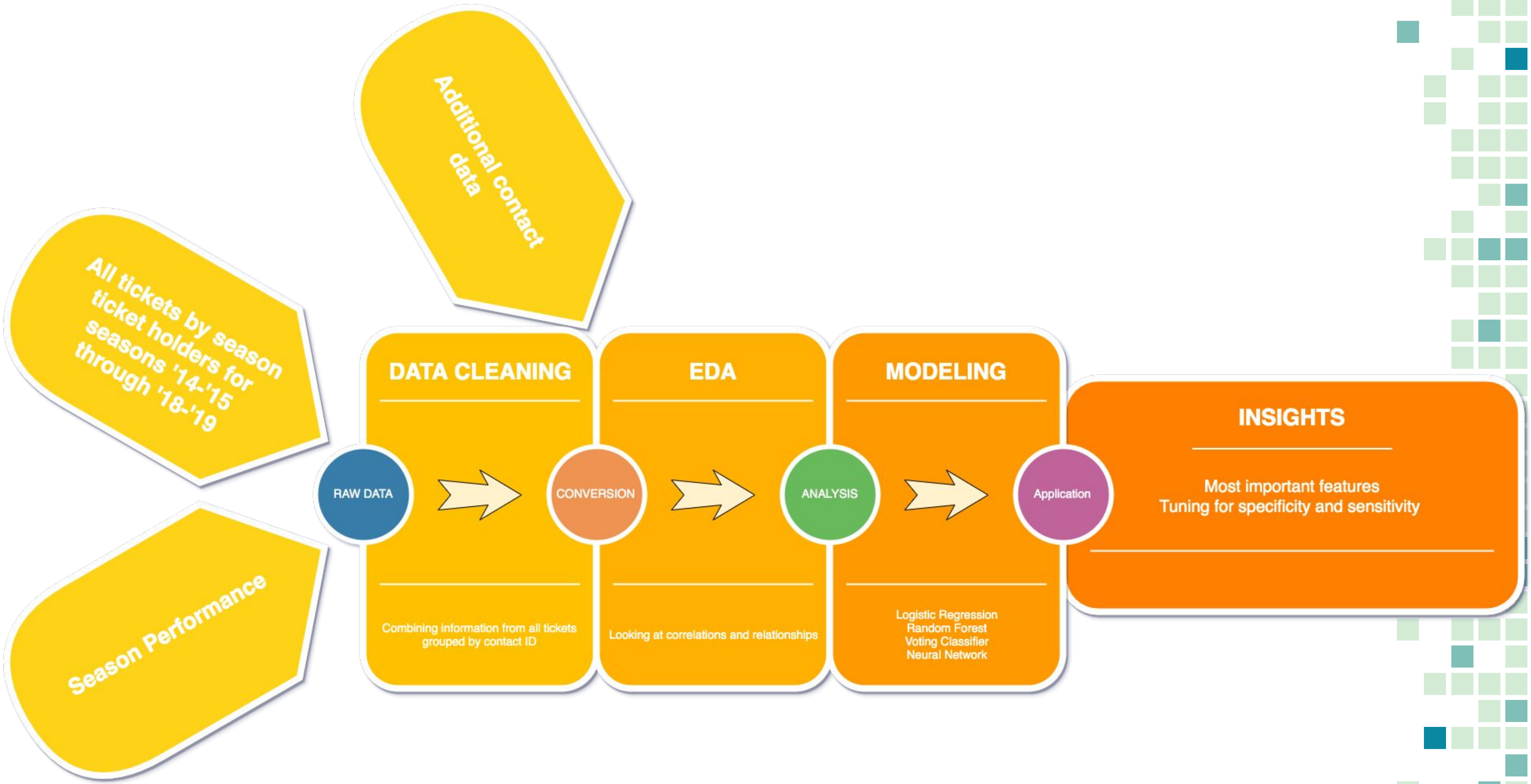
"We are concentrating [on] people spending a lot of money — season ticket holders. That is really a very small number of customers.

- Diny Hurwitz, Milwaukee Brewers data analyst

The Importance of Season Ticket Holders

- Season ticket holders pay a large amount of money up-front, expecting return on value later
  - Not many other markets have such a dedicated customer base that invest thousands of dollars before the product/service reveals it's value
- Season ticket holders should be valued as dedicated consumers

The Workflow


Additional contact data
All tickets by season ticket holders for seasons '14-'15 through '18-'19
Season Performance
DATA CLEANING
EDA
MODELING
INSIGHTS
RAW DATA
CONVERSION
ANALYSIS
Application
Most important features
Tuning for specificity and sensitivity
Combining information from all tickets grouped by contact ID
Looking at correlations and relationships
Logistic Regression
Random Forest
Voting Classifier
Neural Network

Acquiring the Data

- Data was obtained from the Senior Business Intelligence Analyst at Kroenke Sports & Entertainment
- Data came in the form of each individual ticket held by Season ticket buyers for seasons starting '14-'18
- All personal/demographic information was withheld in order to protect privacy
- All revenue/pricing information was also withheld

Cleaning the Data

- Data was given in the form of individual ticket information
- Started with '14-'15 season, then created functions for the rest
- I aggregated based on the contact ID, and extracted information by season
    - Attendance
    - Number of seats
    - General location
    - Number of tickets
    - Number of sections
    - Distance from Pepsi Center
    - Gender
    - Payment in full
    - Renewal

Season Performance

- Obtained season performance data from www.hockey-reference.com
  - Season points
    - Total points divided by total possible points
      - 2 points for wins
      - 1 point for overtime losses
      - 0 points for losses
      - 164 possible points
  - Wins
  - Overtime losses

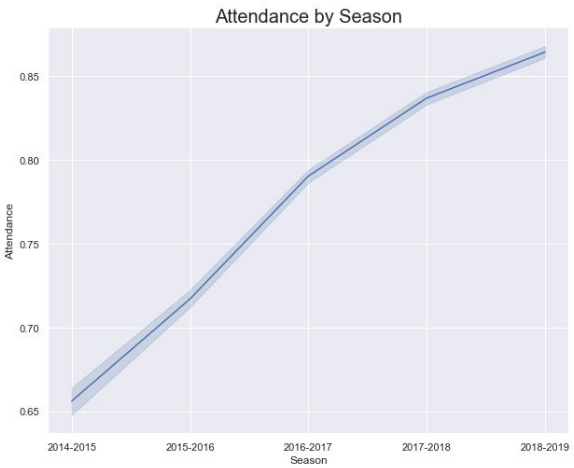
Additional Data

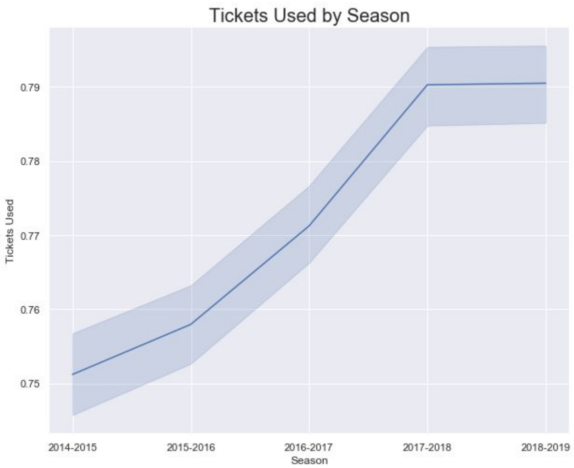
After creating dataset of all season ticket holders information and season performance metrics, I was given additional data

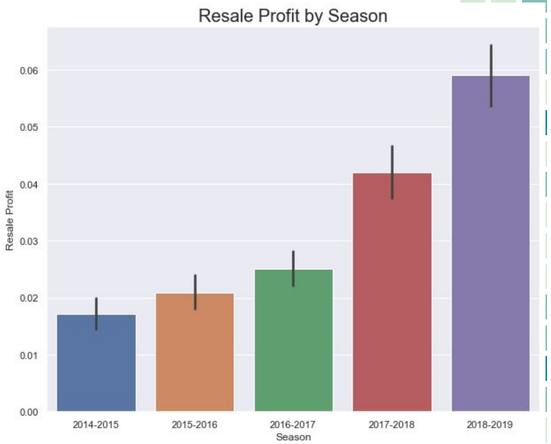
- Tenure for season '14-'15
- Percent total tickets used
- Profit margin from sold tickets
- Renewal rates for '18-'19

Exploratory Data Analysis

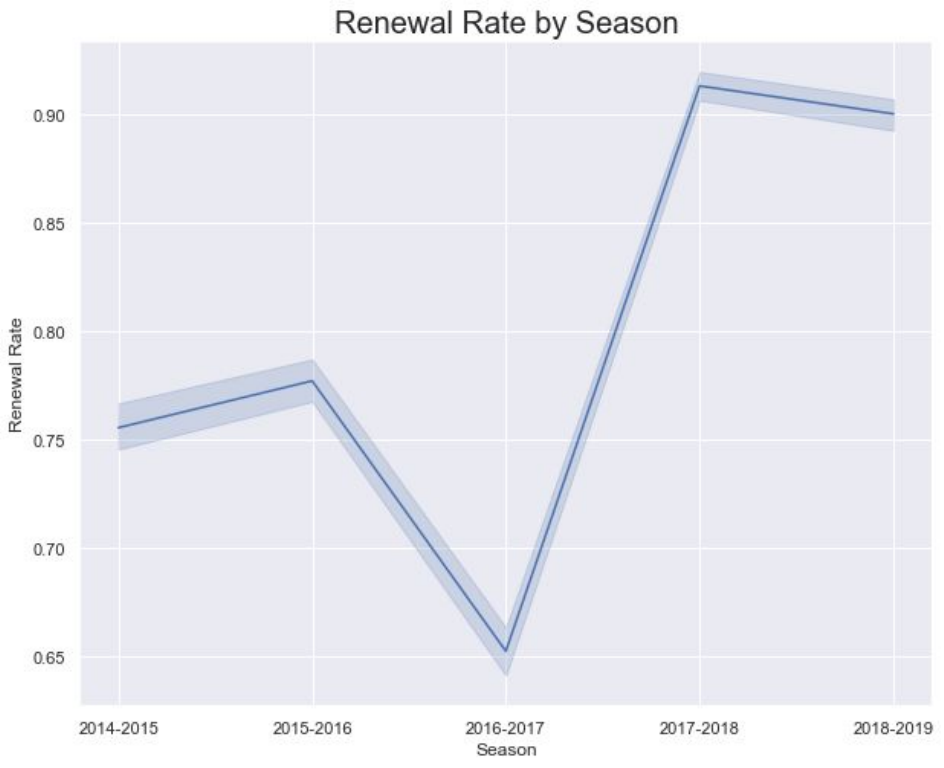
After combining all seasons, I visualized relationships in the data to obtain initial insights





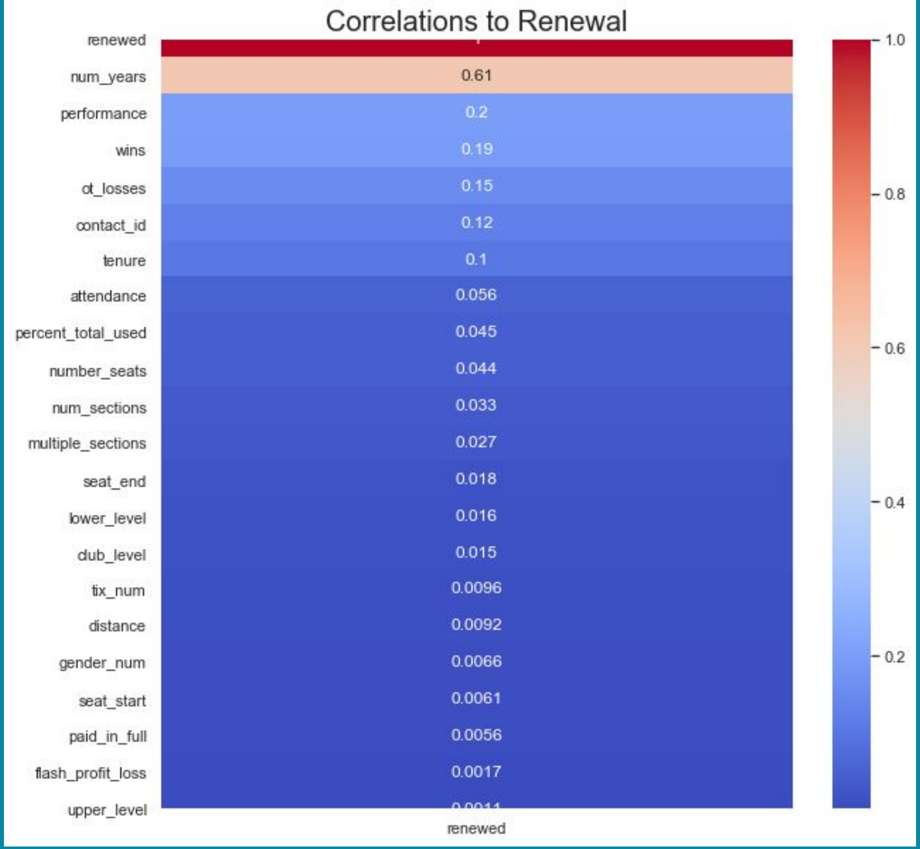


Renewal



Renewal Rate by Season
0.90
0.85
0.80
0.75
0.70
0.65
Renewal Rate
2014-2015
2015-2016
2016-2017
2017-2018
2018-2019
Season

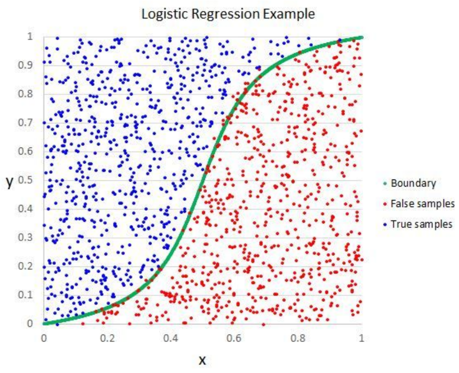
Correlation

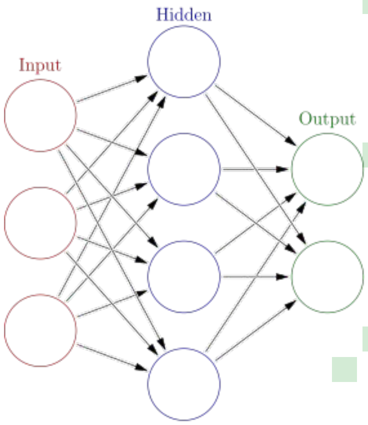


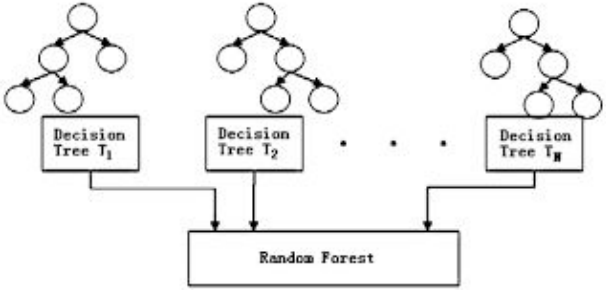
Statistical Modeling

Models used:

- Logistic Regression
- Random Forest
- Voting Classifier
  - AdaBoostClassifier
  - GradientBoostingClassifier
  - DecisionTreeClassifier
  - RandomForestClassifier
- Neural Network







Results

- Train on ‘14 through ‘17, test on ‘18
  - Logistic regression was highest performing
    - 90% test accuracy
  - Most other models yielded around 80-83%
  - Neural network also has 90%
    - Too time consuming
- Random train test split
  - Yielded much higher results across the board
    - Up to 98% test accuracy with random forest
    - Offers less information for the stakeholders

Most Important Factors

Some of the highest coefficients:

- Number of wins
- Number of years holding season tickets
- Distance from Pepsi Center
- Tenure
- Overtime Losses

Tuning the predictions

Accuracy, Sensitivity, or Specificity?

While accuracy is important, we may want to tune towards false negatives or false positives

- The cost of focusing on retaining a contact less likely to renew may be high
- The benefit may be higher

Base predictions

```
true negatives: 40
false positives: 582
false negatives: 51
true positives: 5552
```

Requiring 75% confidence (70% accuracy)

```
true negatives: 363
false positives: 259
false negatives: 1624
true positives: 3979
```

Key Takeaways

Season performance is of the highest importance to both attendance and renewal

Loyal fans will typically stay loyal

Distance also has a high impact on renewal

Conclusion

- Highest accuracy score for season '18-'19 was from a logistic regression at ~90%
- Randomly splitting testing and training data improved accuracy by up to ~95%
  - While not as applicable, this model can still be used to gain insights into important factors for season ticket holder retention
- The most important factors:
  - Season performance metrics
  - Distance from Pepsi Center
  - Tenure/years holding season tickets

Recommendations

- Focus on obtaining more season ticket holders
  - They tend to stay loyal
- Find a way to decrease the burden of commuting
- Continue to focus on building a high performing team

Further Steps

- Access to more demographic data
- Access to revenue/pricing data
- Include weather data
- Spend more time feature engineering
  - Combining contact data, revenue, and weather

Thank you for your time!

Any questions?

You can find me at:

https://www.linkedin.com/in/jeffreywthomas/

Special thanks to Kroenke Sports & Entertainment for providing the data