Predicting NFL Games

Jennifer Stipe, Atul Bachhav, Brandi Berry

July 9, 2024

Project Goal

Analyze historical NFL data - home/away, weather conditions, game location, typical location weather, record against opponent, etc.

- Classify top teams by winning percentage
- Chart impact of total yards on score
- Predict winner of future games

Methodology

Data Clean up

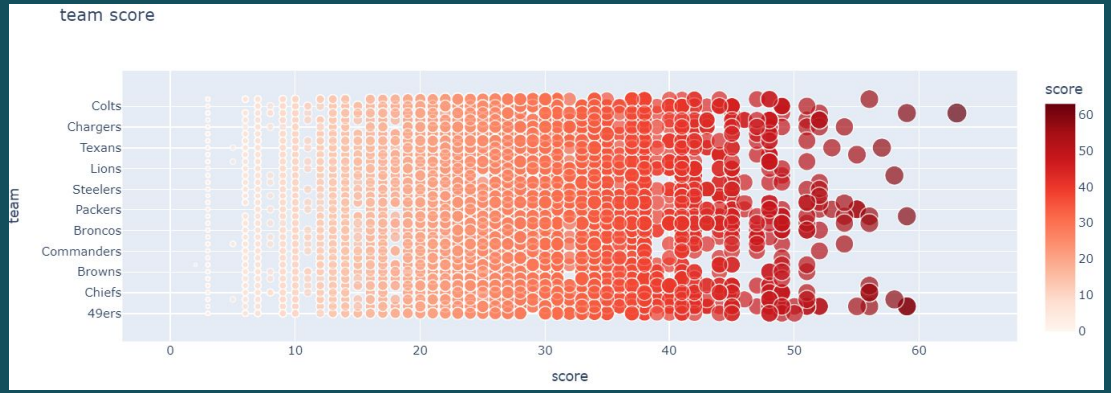
- Find appropriate data sources
- Merge Data sources: NFL game score details, NFL Stadium details, Game weather details
- Seperate home/away columns so that the data can be stacked vertically for easier reporting and modeling
- Clean up team names - Oakland Raiders, LA Raiders, LV Raiders = Raiders
- Clean missing surface data: If indoor stadium replace nulls with turf, if outdoor replace nulls with grass
- Drop remaining rows with Null data leaving 9,260 rows to work with

Classification & Modeling

- Determine the team that posted the highest score, the highest % of wins, and highest % of losses
- Determine if there is a relationship between yards of offense and score
- Use models to predict the winner of a given game

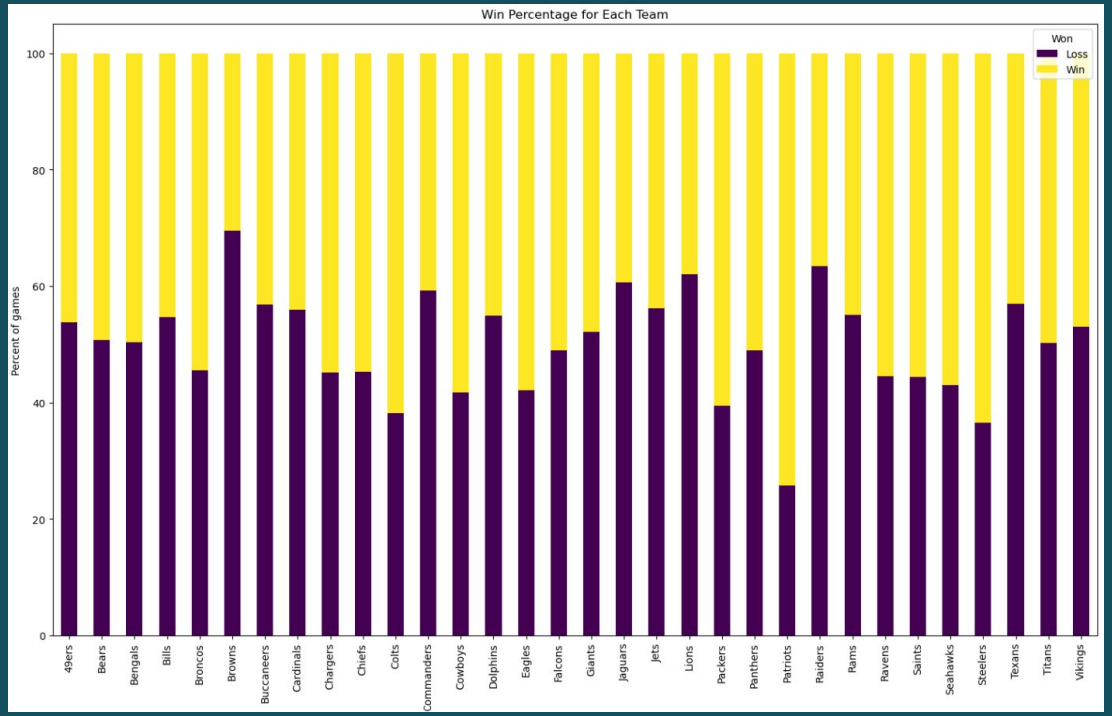
Team/Score Classification

- The highest score recorded by a team in our cleaned dataset was 63 points by the Raiders during a game in Dec 2023
- In actuality, the highest score by a single team from 2002-2023 was 70 by the Miami Dolphins 3 months previously in September of 2023. However, some data points for that game were missing so it was removed. Underscoring the need to understand the data, and the impact of cleaning has on it.



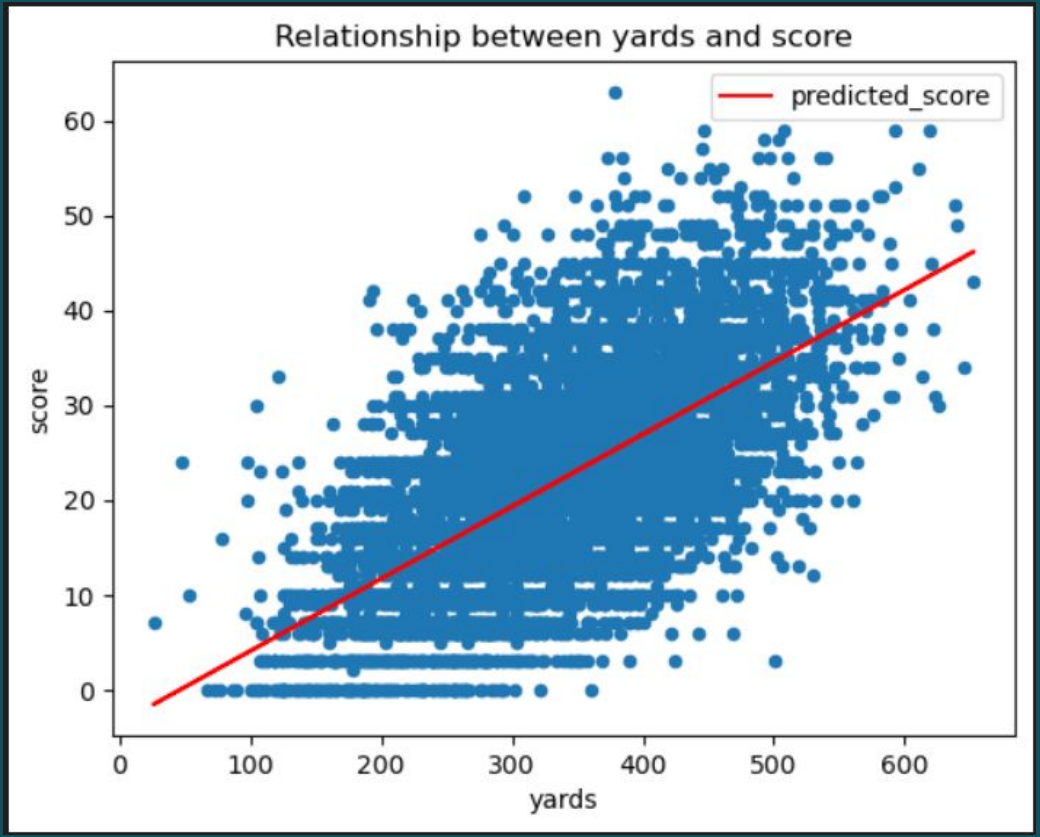
Wins/Loss by Team

- The chart shows the Patriots have the highest winning percentage, while the Browns have the least, by quite a margin



Total yards impact on Score

- A positive linear relationship exists between yards of offense and score
- This information could be helpful in estimating scores of each team in a game based on their average yards for last several games
- However, the r^2 value is only .398, meaning the two metrics have a moderate, but not strong relationship



Modeling a Winner

- Given the extensive number of attributes in our data, only the more complex models were attempted
- Random Forest had the best model result, with test score of 81.1%
- The variable contributing the most to the outcome were Rushing attempts, time of possession, and passing attempts
- Given the extensive game details in the data set, there is definitely an X-factor in determining a win

```python
# Train the Random Forest modelw/ n_estimators = 100
RF_Model = RandomForestClassifier(random_state=1, n_estimators=100)

# Fit the model
RF_Model.fit(X_train, y_train)


# Score the model
print(f'Training Score: {RF_Model.score(X_train, y_train)}')
print(f'Testing Score: {RF_Model.score(X_test, y_test)}')
```

✓ 1.4s

```
Training Score: 1.0
Testing Score: 0.811231101511879
```

```
[(0.10756367812738386, 'rush_att'),
 (0.055239070592370945, 'possession'),
 (0.048519658448996354, 'pass_att'),
 (0.04661891227088974, 'rush_yards'),
 (0.04389010543997839, 'interceptions'),
 (0.037075956208523095, 'yards'),
```

Predicting a Future Winner: NFL Week 1 2023

- We trimmed our modeled data to remove 2023 data so we could use that see how well the model could prediction without the game data

- Results of prediction vs what happened were 57.1%, which is not as good as current Vegas standards.

- The accuracy of the prediction would increase as game inputs are entered into the model

Game	Precicted winner	Acutal
Dolphins @ Chargers	Dolphins	Dolphins
Texans @ Ravens	Texans	Ravens
Panthers @ Falcons	Panthers	Falcons
Bengals @ Browns	Bengals	Browns
Bills @ Jets	Bills	Jets
Jaguars @ Colts	Jaguars	Jags
Packers @ Bears	Packers	Packers
Eagles @ Patriots	Eagles	Patriots
Lions @ Chiefs	Lions	Lions
Cardinals @ Commanders	Cardinals	Commanders
Buccaneers @ Vikings	Buccaneers	Buccaneers
Raiders @ Broncos	Raiders	Raiders
Cowboys @ Giants	Cowboys	Cowboys

Next Steps

- Determine the ‘X’ factor in determining a win to increase the accuracy rate
  - Quarterback
  - Head Coach
  - Defensive rank
  - # of Take-Aways
  - # of starters on IR
  - Weighting the most recent season results
- Create a user interface to allow for entry of future games and provide an predictive outcome

*This team is not responsible for any financial loss should individuals choose to utilize this model to place bets.