

Lords of the Ring: How to Predict Super Bowl Winner?



Ninad Dixit

Springboard Data Science Career Track

WHY SUPER BOWL?

Players and Fans – It's NEVER enough!



WHAT'S IN IT FOR US?

Tampa Bay Buccaneers vs. Kansas City Chiefs
23.2 million people – 4.3 billion USD bets



PROBLEM

HOW CAN WE PREDICT THE SUPER
BOWL LVI WINNER?

PROBLEM

Teams' performance varies over the season



PROBLEM

Teams' performance varies over the season



DATA

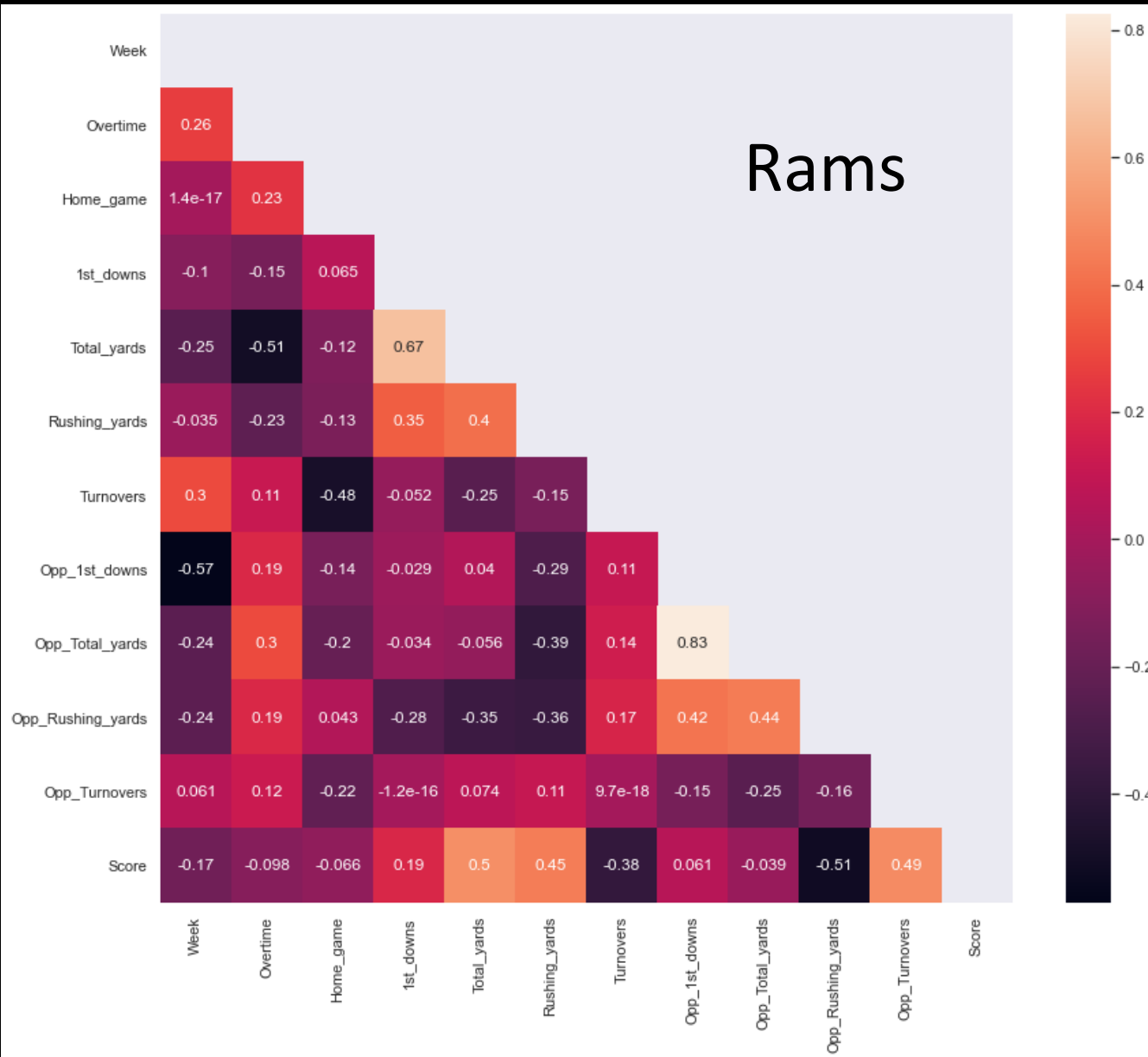
Source: <https://www.pro-football-reference.com/>

Target feature: 'Score'

Other features:

- Week
- Date
- Overtime
- Home_game
- Opponent
- 1st_downs
- Total_yards
- Rushing_yards
- Turnovers
- Opp_1st_downs
- Opp_Total_yards
- Opp_Rushing_yards
- Opp_Turnovers

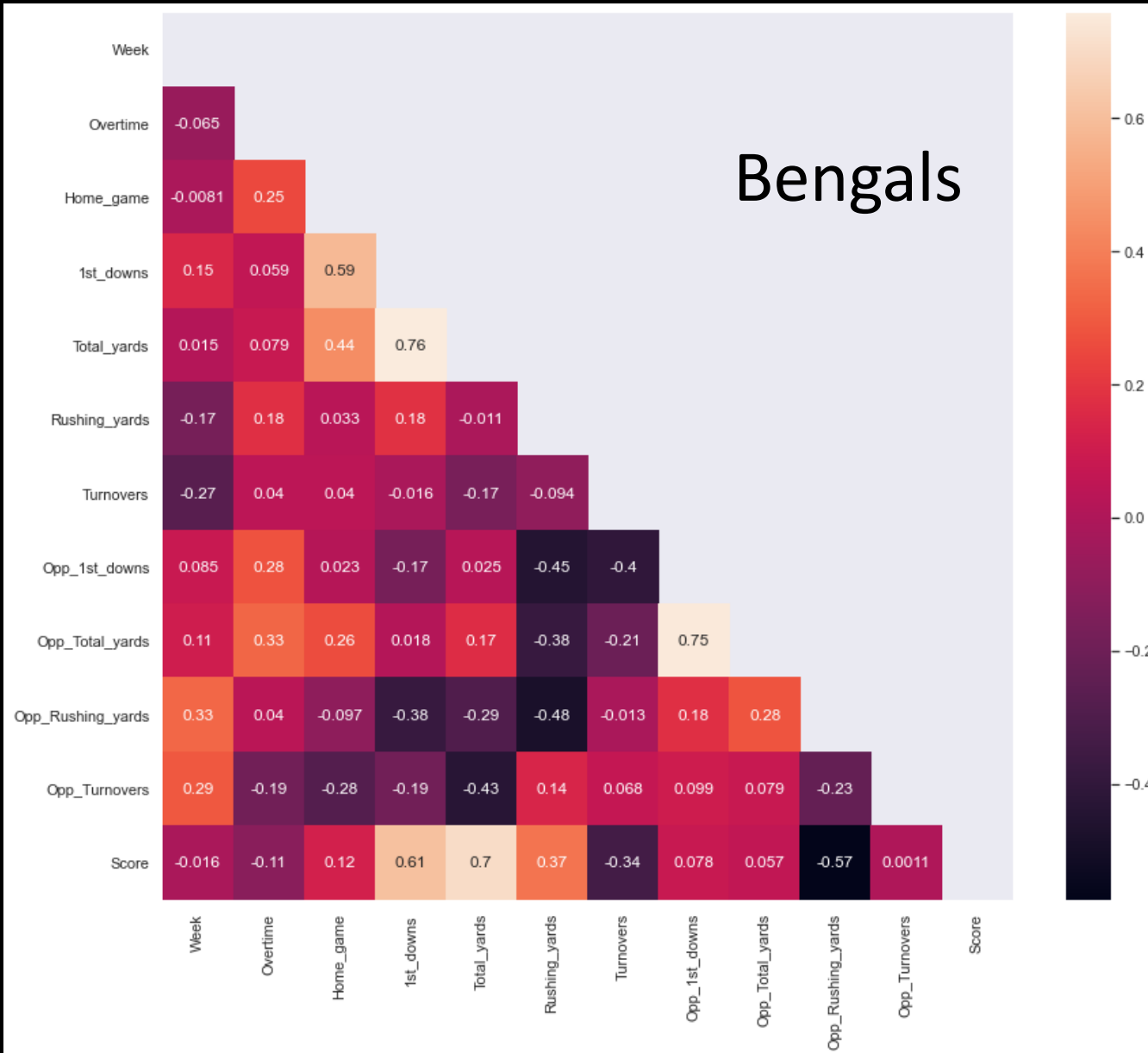
EXPLORATORY DATA ANALYSIS



Correlation with 'Score':

- Total_yards (0.5)
- Opp_Rushing_yards (-0.51)
- Rushing_yards (0.45)
- Opp_Turnovers (0.49)
- Turnovers (-0.38)

EXPLORATORY DATA ANALYSIS



Correlation with 'Score':

- Total_yards (0.7)
- 1st_downs (0.61)
- Opp_Rushing_yards (-0.57)
- Rushing_yards (0.37)
- Turnovers (-0.34)

APPROACH

- Estimate performance statistics
- Linear regression model to predict the scores in previous games
- Use the regression model and estimated statistics to predict the Super Bowl scores

PERFORMANCE STATISTICS ESTIMATION

Date	2021-09-12	2021-09-19
Week	1	2
Start_Time	13:00:00	13:00:00
Overtime	0.0	0.0
Team	Buffalo Bills	Buffalo Bills
Home_game	1.0	0.0
Opponent	Pittsburgh Steelers	Miami Dolphins
1st_downs	22	21
Total_yards	371	314
Rushing_yards	117	143
Turnovers	1	2
Opp_1st_downs	16	13
Opp_Total_yards	252	223
Opp_Rushing_yards	75	71
Opp_Turnovers	0	3
Score	16	35

PERFORMANCE STATISTICS ESTIMATION

Issues:

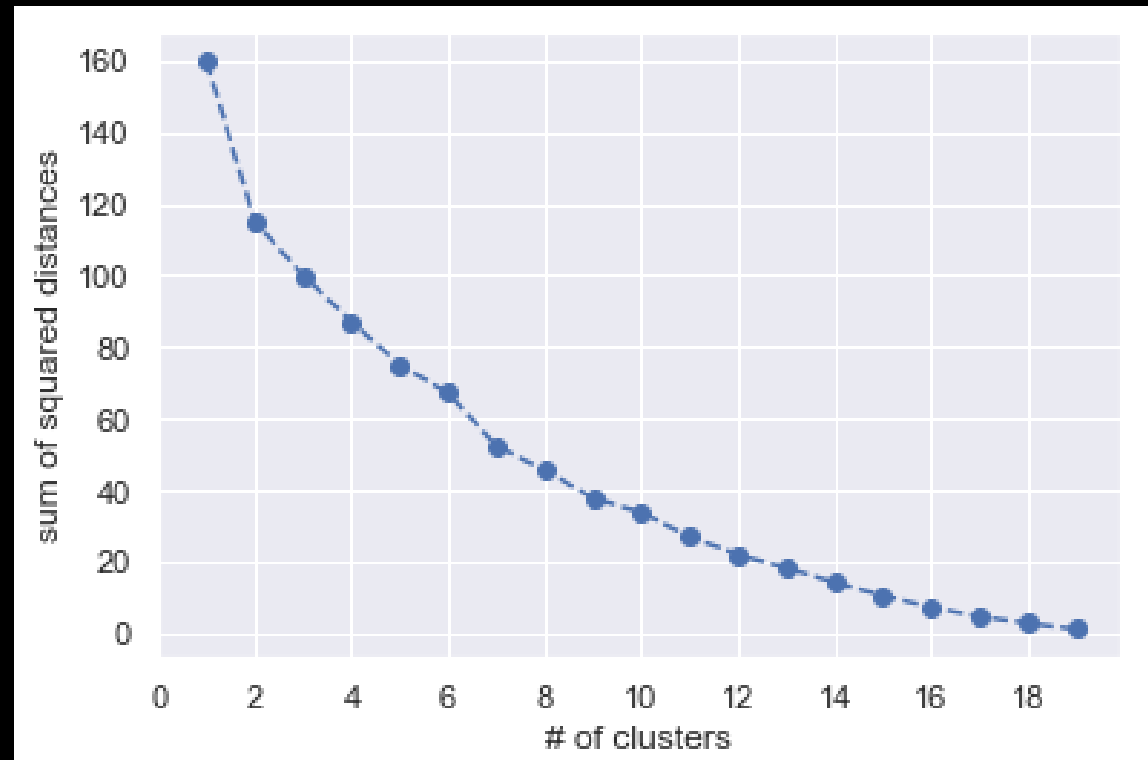
- Rams did not play against the Bengals in 2021
- Teams' performance changes over the season

Solution (for the Rams):

- Average last three performances of Bengals and their opponents
- Find Rams' opponent that closely matches with Bengals' average performance

PERFORMANCE STATISTICS ESTIMATION – Los Angeles Rams

Clustering Rams' Opponents



of clusters = 19

PERFORMANCE STATISTICS ESTIMATION – Los Angeles Rams

Bengals' Statistics in Last 3 Games

Team	Opponent	1st_downs	Total_yards	Rushing_yards	Turnovers	Opp_1st_downs	Opp_Total_yards	Opp_Rushing_yards	Opp_Turnovers
Cincinnati Bengals	Las Vegas Raiders	18	308	83	0	23	385	103	2
Cincinnati Bengals	Tennessee Titans	17	345	65	1	16	353	140	3
Cincinnati Bengals	Kansas City Chiefs	21	359	116	1	24	375	139	2

Bengals as Opponent (Flip the Columns)

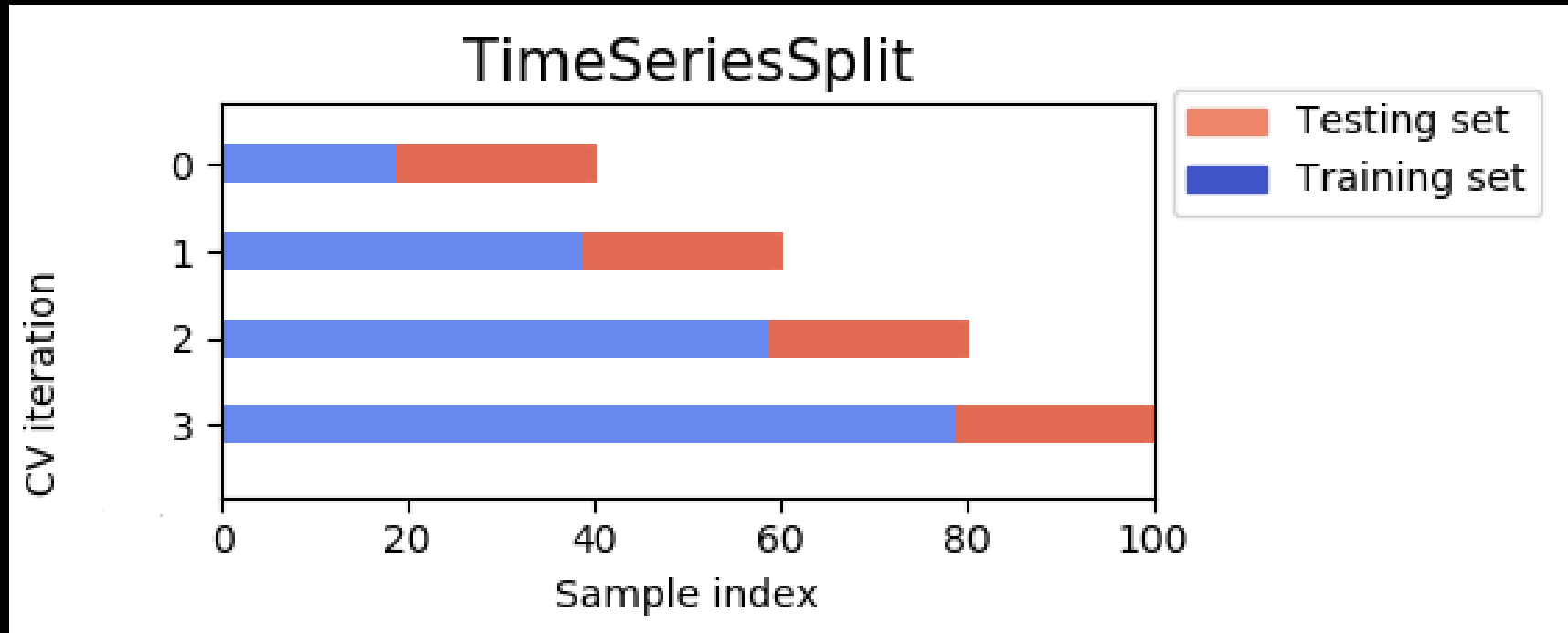
Team	Opponent	1st_downs	Total_yards	Rushing_yards	Turnovers	Opp_1st_downs	Opp_Total_yards	Opp_Rushing_yards	Opp_Turnovers
Imagine as Rams	Cincinnati Bengals	23	385	103	2	18	308	83	0
Imagine as Rams	Cincinnati Bengals	16	353	140	3	17	345	65	1
Imagine as Rams	Cincinnati Bengals	24	375	139	2	21	359	116	1

PERFORMANCE STATISTICS ESTIMATION

Team	Opponent	1st_downs	Total_yards	Rushing_yards	Turnovers	Opp_1st_downs	Opp_Total_yards	Opp_Rushing_yards	Opp_Turnovers
Imagine as Rams	Cincinnati Bengals	23.00	385.0	103.00	2.00	18.00	308.00	83.0	0.0
Imagine as Rams	Cincinnati Bengals	16.00	353.0	140.00	3.00	17.00	345.00	65.0	1.0
Imagine as Rams	Cincinnati Bengals	24.00	375.0	139.00	2.00	21.00	359.00	116.0	1.0
Los Angeles Rams	Arizona Cardinals	24.00	401.0	121.00	2.00	27.00	465.00	216.0	0.0
Rams' Estimated Performance	Cincinnati Bengals	21.75	378.5	125.75	2.25	20.75	369.25	120.0	0.5

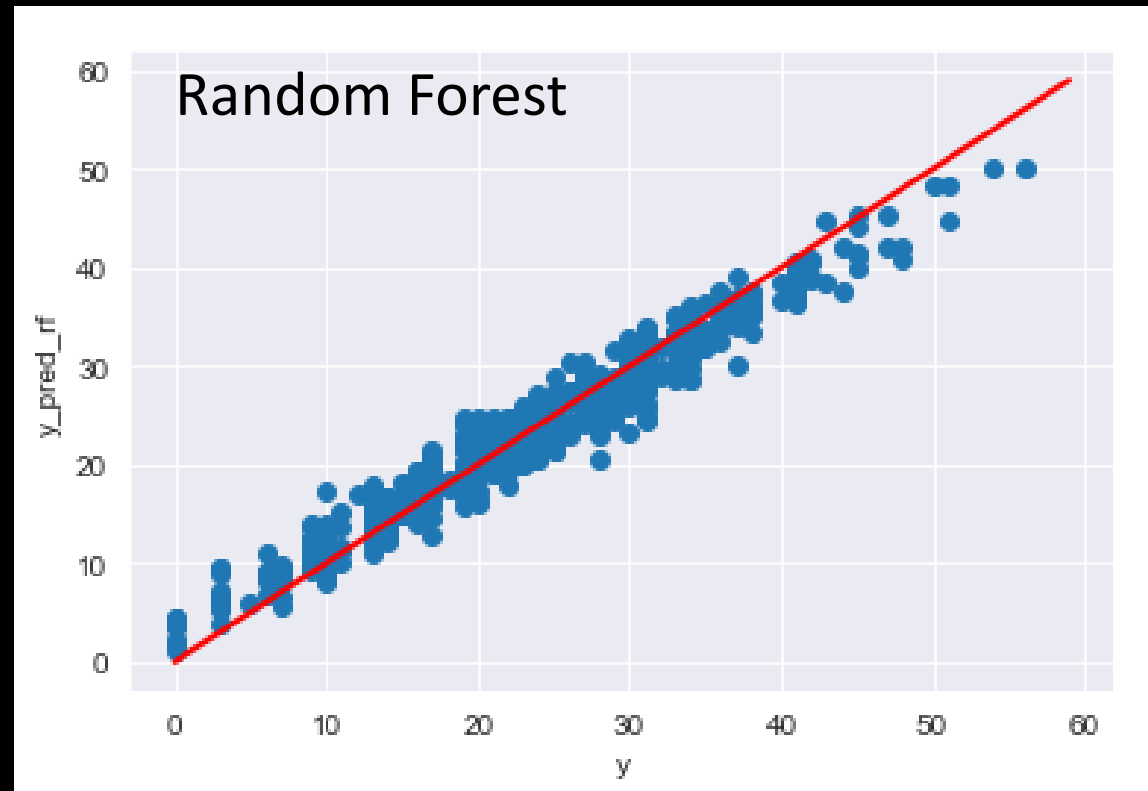
Team	Opponent	1st_downs	Total_yards	Rushing_yards	Turnovers	Opp_1st_downs	Opp_Total_yards	Opp_Rushing_yards	Opp_Turnovers
Imagine as Bengals	Los Angeles Rams	14.00	183.0	61.00	2.0	16.00	375.0	140.0	0.00
Imagine as Bengals	Los Angeles Rams	20.00	359.0	51.00	2.0	24.00	428.0	73.0	4.00
Imagine as Bengals	Los Angeles Rams	16.00	282.0	50.00	1.0	25.00	396.0	70.0	1.00
Cincinnati Bengals	Baltimore Ravens	19.00	520.0	111.00	1.0	20.00	393.0	115.0	0.00
Bengals' Estimated Performance	Los Angeles Rams	17.25	336.0	68.25	1.5	21.25	398.0	99.5	1.25

CROSS-VALIDATION



- Training set: 'K' folds
- Testing set: $(K+1)^{\text{th}}$ fold
- Entries: 568
- Min. # Games / Week: 26
- # Slices: 22

MODEL SELECTION & PERFORMANCE



Regressor	Mean Squared Error	Tuned Hyperparameters
Ridge	34.74	$\alpha: 1$
Elastic Net	36.53	$\alpha: 0.1, l1_ratio: 0.05$
Random Forest	5.71	$n_estimators: 500$

SUPER BOWL SCORE PREDICTION

Regressor	Los Angeles Rams	Cincinnati Bengals
Ridge	25	23
Elastic Net	23	21
Random Forest	23	20
<i>Actual score</i>	<i>23</i>	<i>20</i>



RAMS WIN!!!

CONCLUSIONS

- Linear regression model developed to predict Super Bowl score
- Random forest regressor – best performer

SCOPE FOR IMPROVEMENT

Include

- Special teams performance
- Penalties
- Time of possession
- Injuries / notable absences

Acknowledgements

- <https://www.pro-football-reference.com/> for making the data available publicly
- Ramkumar Hariharan for providing valuable advice

Thank You!