

The background of the image is a vibrant green soccer field with white yard lines. A dark green, semi-transparent rectangular overlay covers the center of the field. Inside this overlay, the title and author information are written in white text. The title is in a large, bold, sans-serif font, while the author names and advisor name are in a smaller, regular sans-serif font. At the bottom of the overlay, the journal title is written in a light gray, sans-serif font.

# **Why Did They Succeed? A World Cup 2022 Analysis**

By: Aaron Lim, Harrison Goetz, Fangzhou Dai, Kelly Zhang, Neil Varwandkar  
Advisor: Naomi Korn

Wharton Moneyball Academy Flex 2023 Sports Analytics Student Research Journal

# Introduction

- The World Cup is considered the premier competition in soccer, with the winner claiming the honor of the title of the sport's best.
- Different factors contribute to teams making it all the way to the final and doing well in the competition

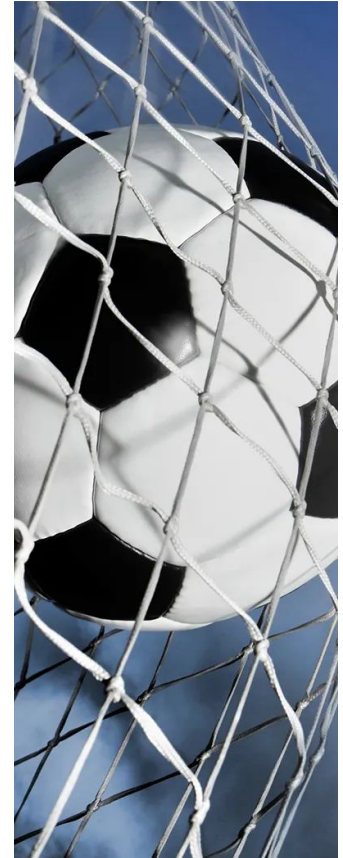
# Hypothesis

The factors salary and goals per 90 are most predictive of a team's World Cup final rank in 2022



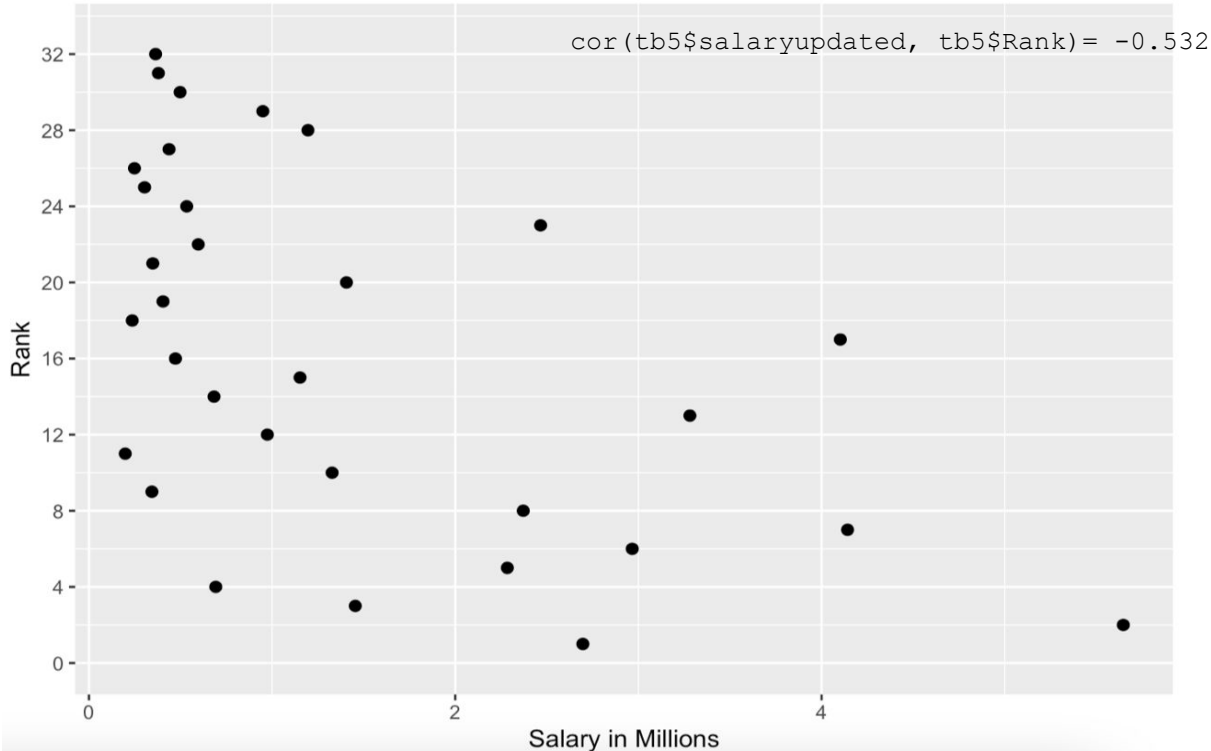
# Procedure

- We took a whole host of data from the 2022 Qatar World Cup, and gave each team a final rank as well as statistics for each team.
- From the data, we compared the rank to many different data categories like salary, goals\_per90, xa\_per90, defensive actions and more, to figure out which data category had the closest correlation to the final standings.
- The data was obtained from Kaggle: Performance Attributes and joined with Fifa World Cup 2022 Rankings.
- Analysis of data and code run in R



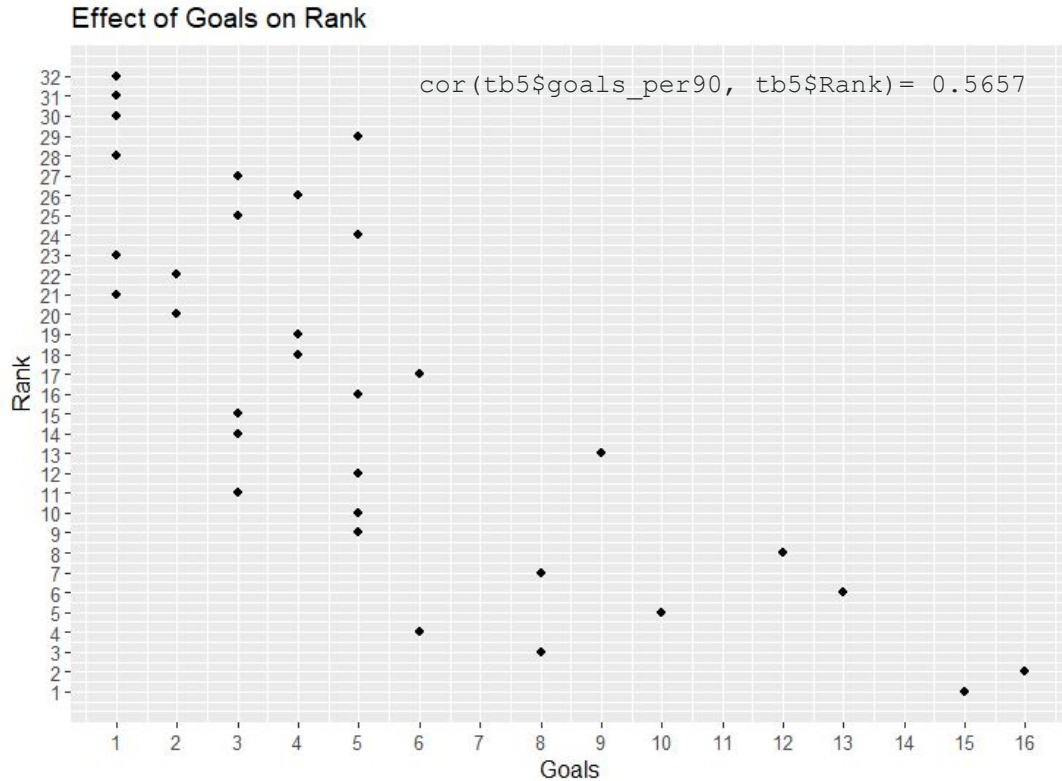
# Results: Effect of Salary on Rank and Stage Reached

The Effect of Salary on Rank



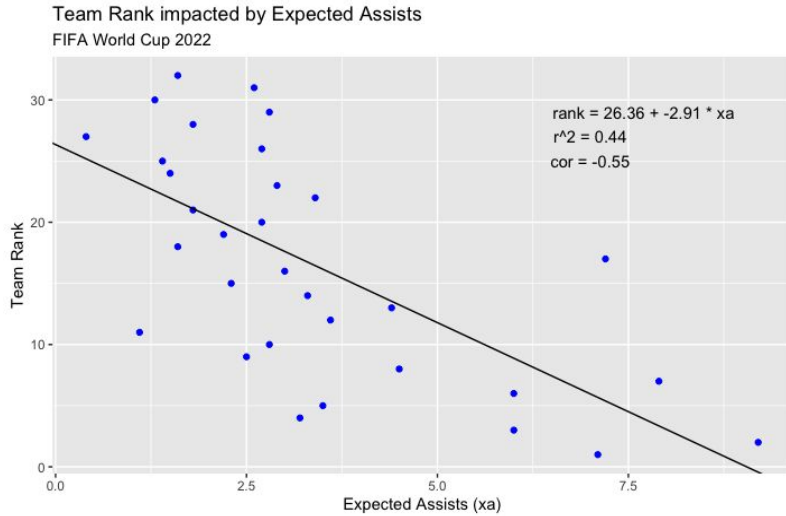
The graph show that teams with lower salaries did worse than teams with higher salary.

# Results: Correlation between Goals Per 90, and Rank

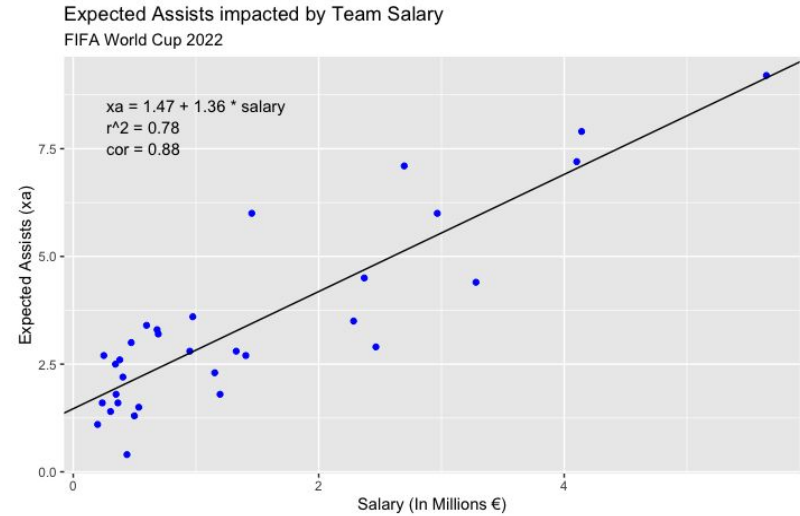


As goals increase, a team's rank gets better. This makes sense as a team's rank is dependent on winning games and in order to win games the team needs to score more goals.

# Results: Further Impact of Salary on Team Rank



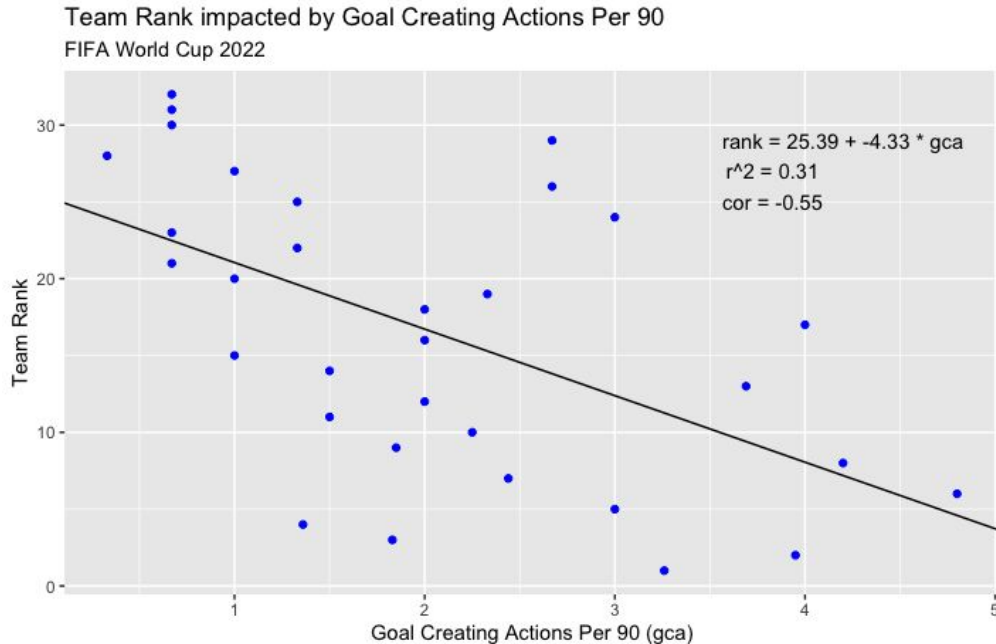
This graph shows a moderately strong correlation between team rank and expected assists. Teams with more expected assists had a better rank. This is confirmed by the linear regression model with a correlation of -0.55 and  $r^2$  of 0.44.



The graph above displays a strong positive correlation between expected assists and team salary. Teams with a higher salary typically have better expected assists. This is confirmed by a linear regression model with a satisfactory  $r^2$  of 0.78.

Salary impacts additional team rank factors, making it a significant indicator of team rank.

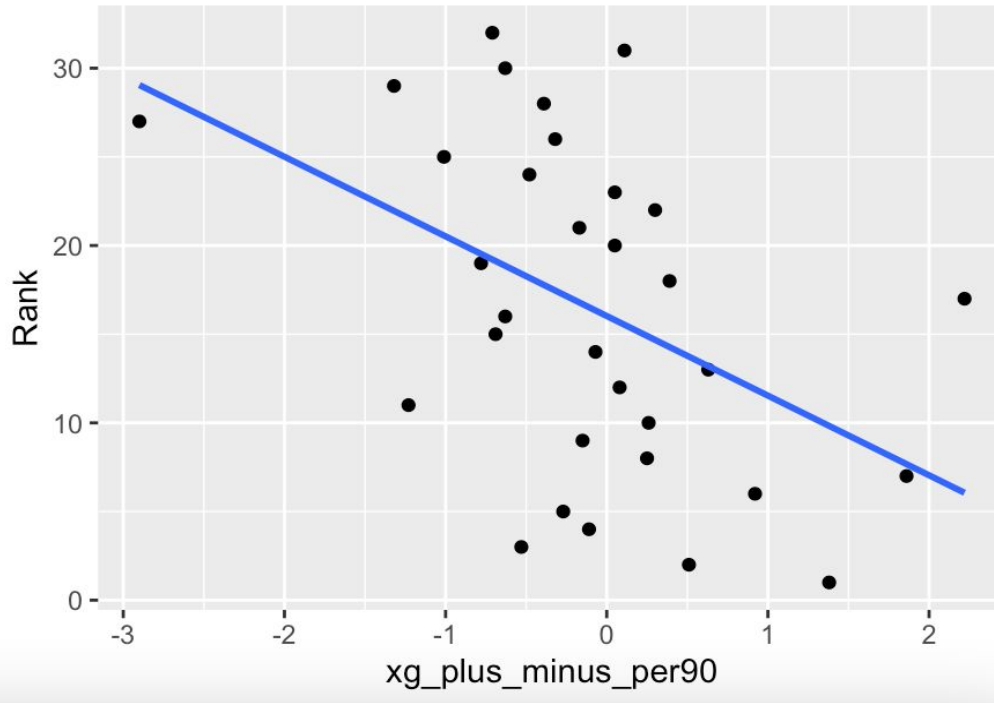
# Results: Correlation between Team Rank and Goal Creating Actions per 90



This graph shows a moderately strong correlation between team rank and goal creating actions per 90. Teams with more goal creating actions typically have a better rank. This is confirmed by the linear regression model with a correlation of -0.55 and a  $r^2$  of 0.31.

# Results: Correlation between Team Rank and Expected Goals Plus Minus Per 90

Team Rank based on xg plus minus per 90



This graph displays that the best teams in this competition ended up having a higher overall expected goals per 90 over their counterparts. This is established by the negative linear regression correlation of  $-0.4534295$ .



## *Results:* Predictor variable correlations with rank Compared

<b>Variable Compared with Rank</b>	<b>Correlation</b>
Salary	-0.53
Goals per 90	-0.57
Expected Assists	-0.55
Goal Creating Actions	-0.55
Expected Goals +/- per 90	-0.45

Note, lower rank is better. So a correlation with a negative value is a variable with a positive correlation to a world cup success.

# Conclusion

- While “goals per 90” and “expected assists” showed the highest correlation, all of the other variables showed a strong correlation with rank
- Since all of these stats (bar salary) measure ‘attacking stats’, the strong correlation showed that more of a attacking style of soccer is favored.
- External factors like injuries, team chemistry, and strength of opposition were all components that skewed certain data points, hence the anomalies.