What Makes A Good FIFA Player?

SPT Squad: Ariadna Chuaqui, Kevin Ma, Akira Nakamura, Elizabeth Shulman

FIFA

- Best-selling video game franchise in the world, according to the Guinness World Records
- Known for detailed player ratings - a highly debated topic among the FIFA community

PS4
EA SPORTS
FIFA19
FIFA
OFFICIAL LICENSED PRODUCT
EVERYONE
TODOS
E
ESRB
WITH BONUS CONTENT†
Internet connection required to download bonus content. For new FUT players only, see back for details.
CON CONTENIDO ADICIONAL†
Se necesita conexión a Internet para descargar el contenido adicional. Solo para nuevos jugadores de FUT. Consulta el dorso para conocer los detalles.

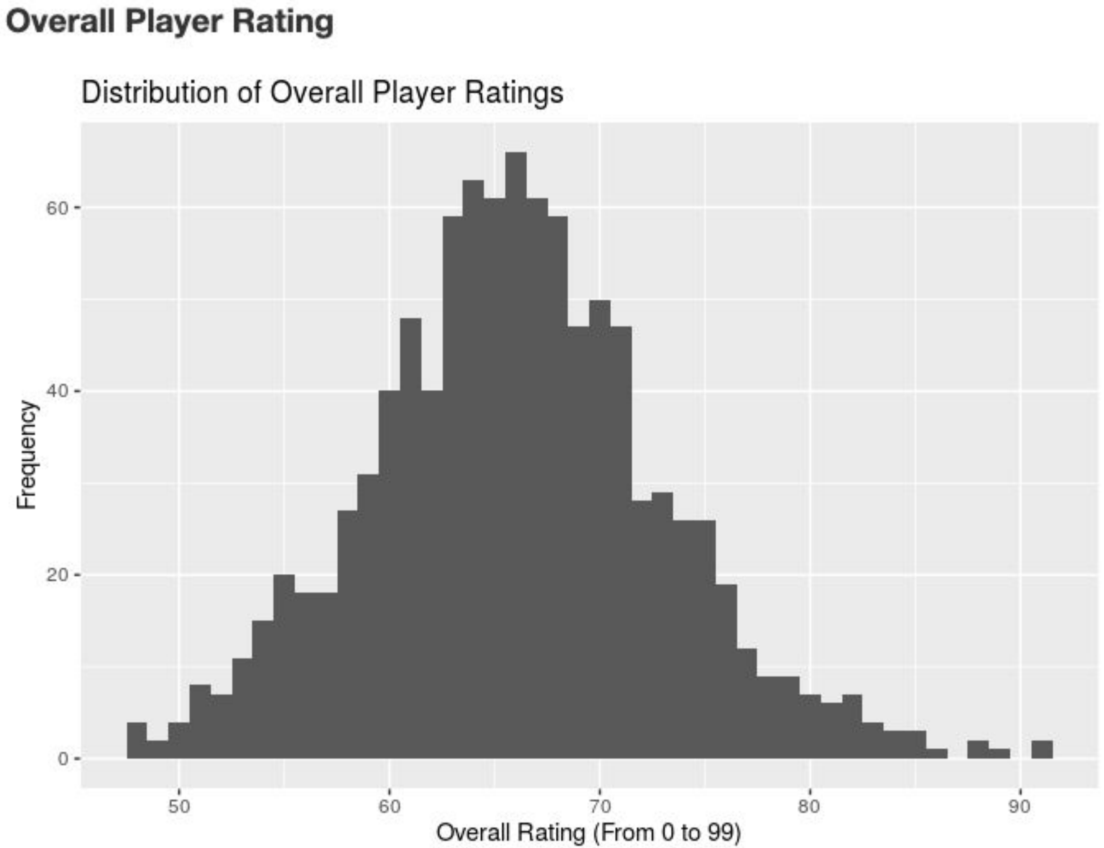
Dataset

- From Kaggle
  - Scraped from a comprehensive online database for FIFA enthusiasts
- 18,000 + players (includes all players in the game)
- **Took a random sample of 1,000 to perform analysis**

Selected Univariate Analysis

Overall Rating

- The distribution of 'Overall' is approximately normal
  - Mean of 66.02
  - Standard deviation of 7.01
- Expected result: the game must have a balance of players with low, medium, and high overall ratings.

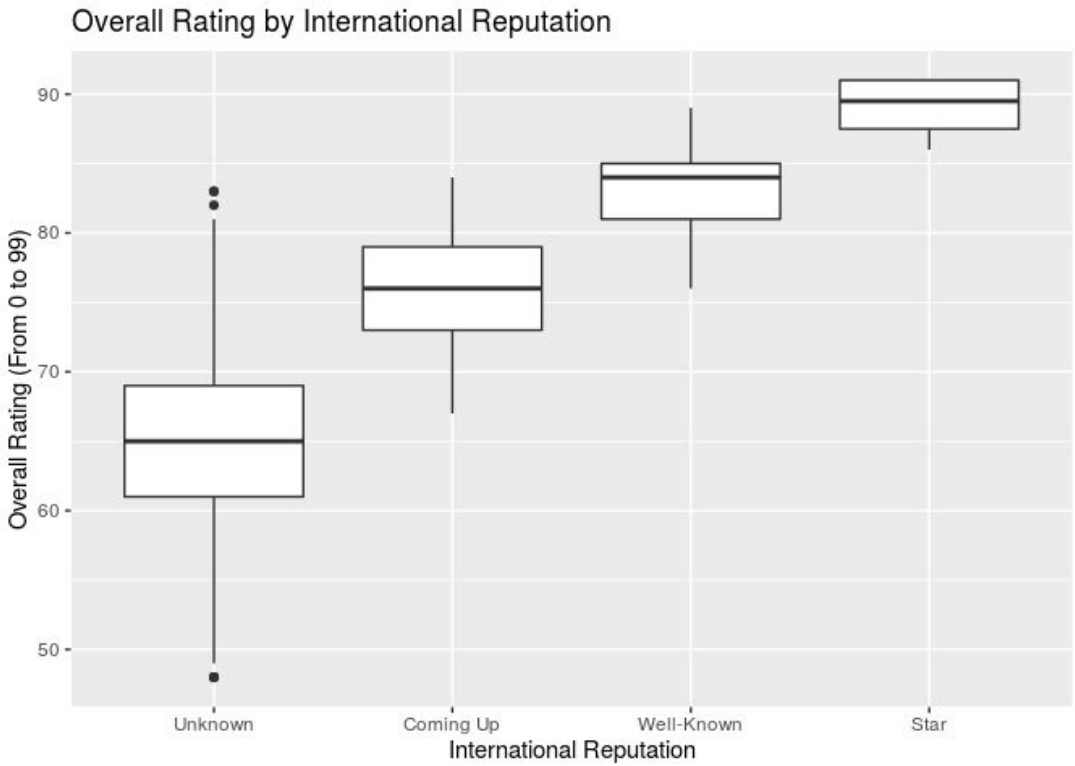


Selected Bivariate Analysis

International Reputation & Overall Rating

- Higher international reputation, higher overall rating
- Expected result: those who have “Star” designation have the highest overall rating.
- Outliers
  - Gerard Moreno (83)
  - Louri Beretta (83)
  - Juiano Mestres (82)*

International Reputation and Overall Rating



What are the characteristics that are important in determining a player’s overall rating?

Dependent Variable: Overall Rating

Selected Independent Variables:

- Age of Player
- Player’s Current Marketing Value
- Current Wage
- International Reputation (rating on scale of 5)
- Player’s Potential Ability (rating on a scale of 100)
- Player’s Ability to successfully complete shots placed (rating on a scale of 100)
- Player’s Dominant Foot (“left” or “right”)
- Player’s Position on the Pitch

Comparison Groups: Nationality of Player, the International Club which the Player plays for, Body Type of Player (“Normal”, “Lean”, “Stocky”, etc.)

Hypothesis:

Out of the mentioned independent variables, we expect ***Value*, *Wage*, *International Reputation*, *Potential Ability***, and ***Ability to successfully complete shots*** to have a statistically significant impact on overall rating based on our exploratory data analysis.

Methodology

- Backwards Selection with AIC
  - A good method to use when narrowing down the appropriate variables for best fitting the data based on a full model
- Recoded Position: Offense vs. Defense
  - There are 24 positions in soccer - too many factor variables for meaningful results

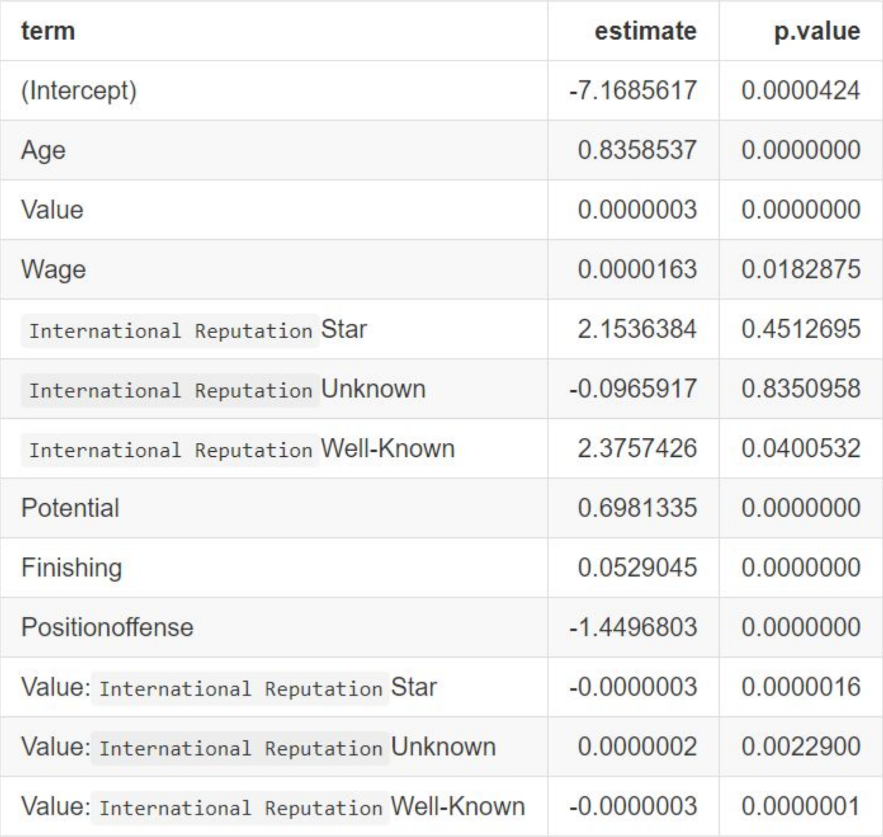
Primary conclusion

The variables that ended up in the selected final model are: ***Age, Value, International Reputation, Potential, Finishing, Position, and Value*International Reputation.***

We were correct about ***Value, Wage, International Reputation, Potential, and Finishing*** being variables that would be significant.

Age was one variable that we did not think would be significant that ended up being significant.

Regression Output

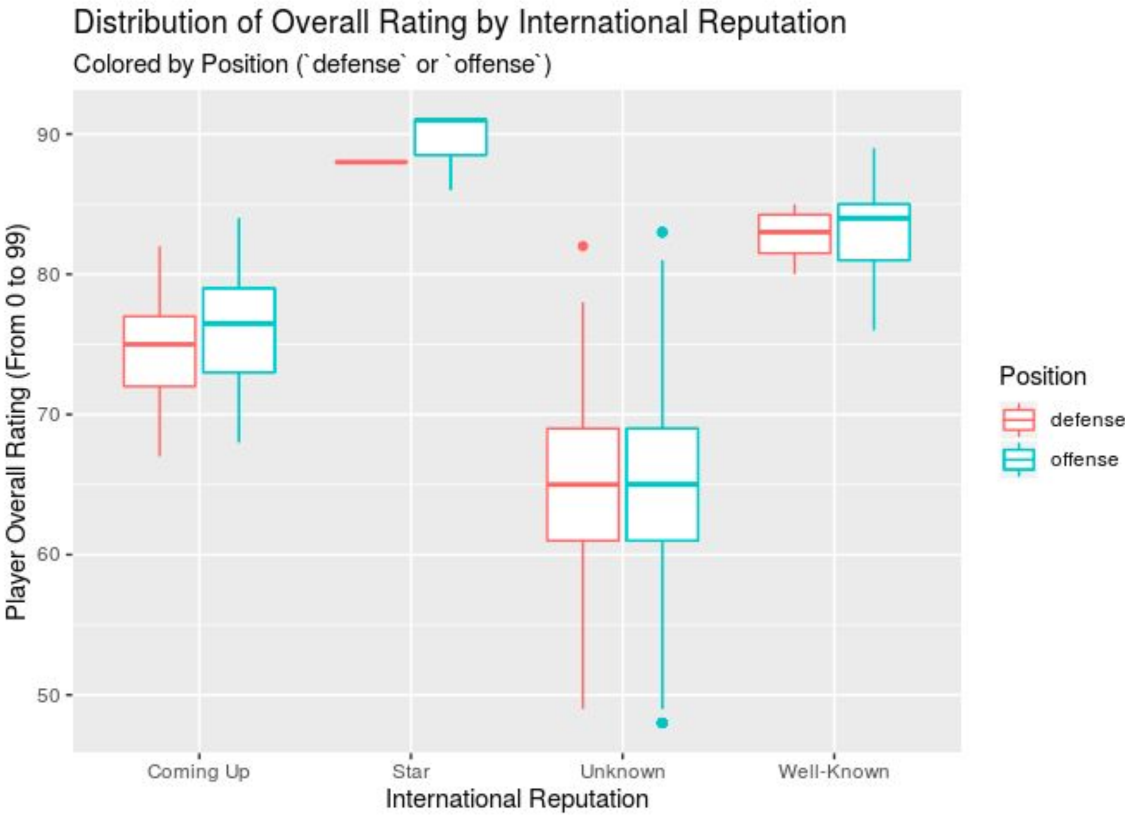


term	estimate	p.value
(Intercept)	-7.1685617	0.0000424
Age	0.8358537	0.0000000
Value	0.0000003	0.0000000
Wage	0.0000163	0.0182875
`International Reputation`Star	2.1536384	0.4512695
`International Reputation`Unknown	-0.0965917	0.8350958
`International Reputation`Well-Known	2.3757426	0.0400532
Potential	0.6981335	0.0000000
Finishing	0.0529045	0.0000000
Positionoffense	-1.4496803	0.0000000
Value:`International Reputation`Star	-0.0000003	0.0000016
Value:`International Reputation`Unknown	0.0000002	0.0022900
Value:`International Reputation`Well-Known	-0.0000003	0.0000001

Visualizations

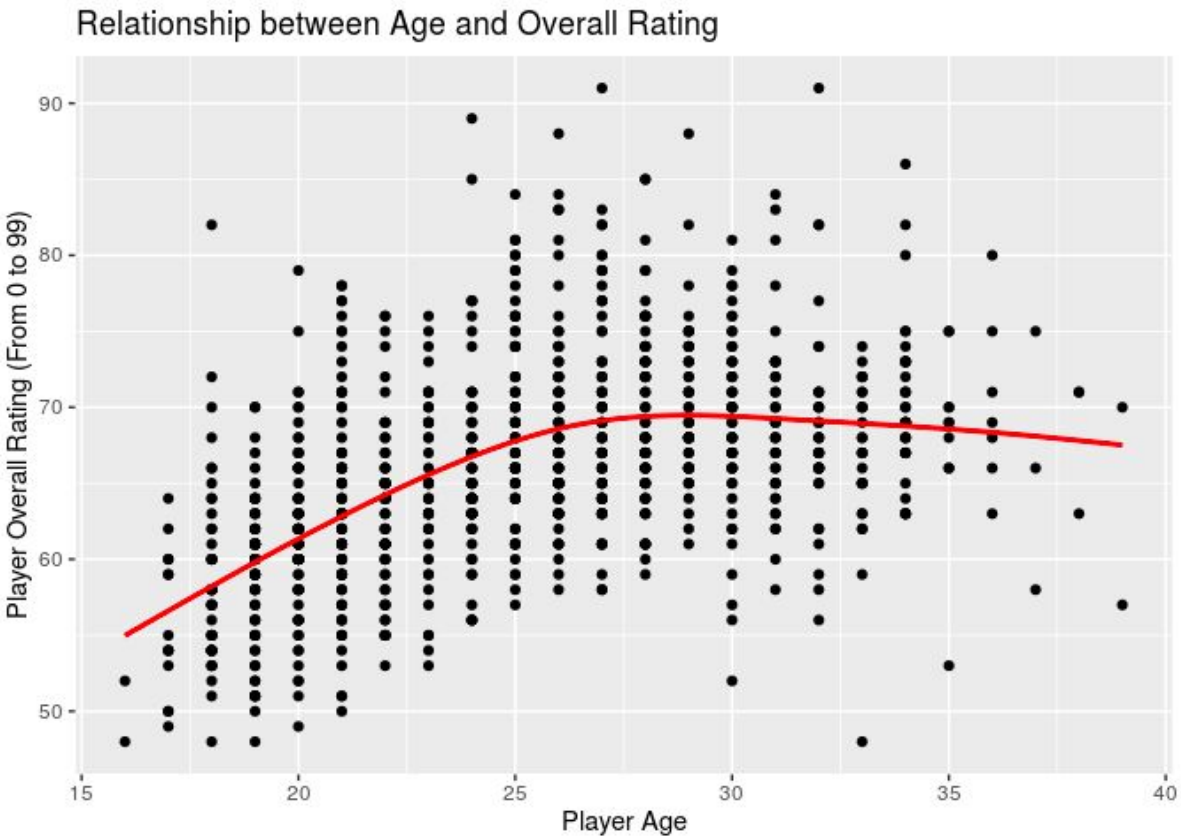
Overall Rating, International Reputation & Position

- Across International Reputations: “Better” International Reputation → higher Overall Rating
- Little variation between offensive and defensive players within each group



Age & Overall Rating

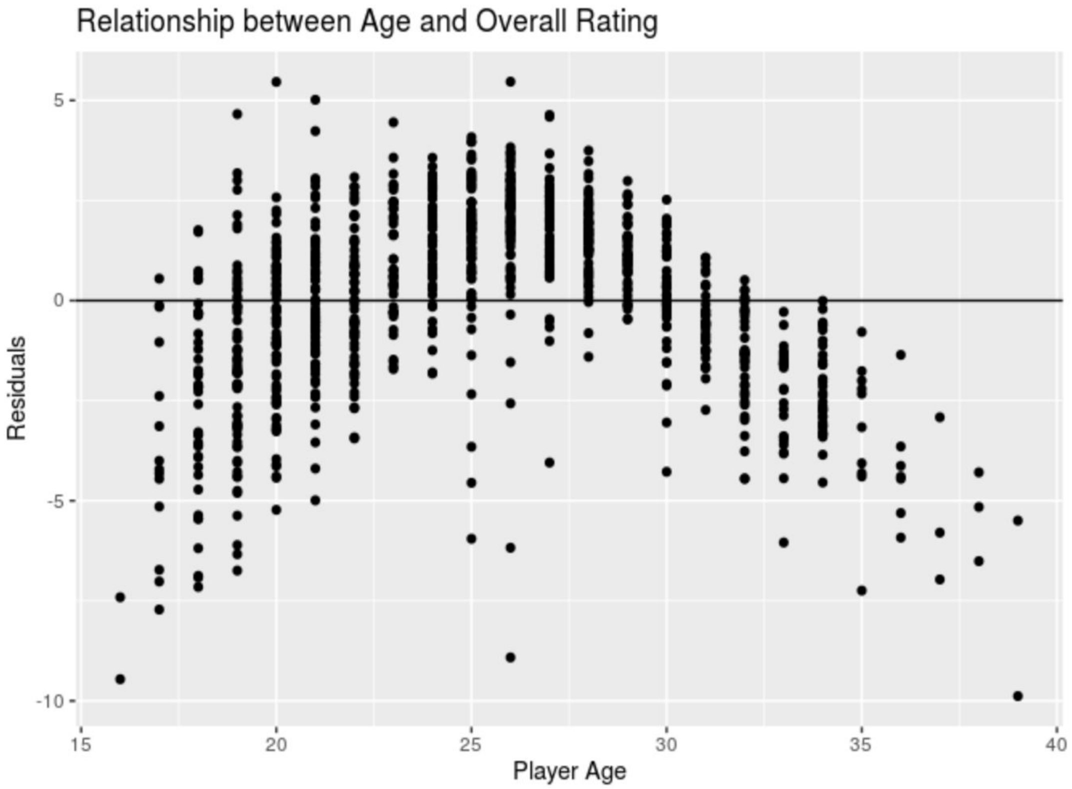
- Mostly-linear, positive relationship between age and overall rating until Age ~ 27 → after which age has a slightly negative effect on overall rating
- Younger players → lack the experience, consistency, and legacy
- Oldest players → subject to injuries and lose athleticism



- The selection model had high predictive power:
    - 89.08% of the variability in overall rating can be explained by the covariates in our selected model (R-Squared = 0.8908)

...However, the R-Squared value does not determine whether the coefficient estimates and predictions are biased

Let's take a look at a residuals plot against one of our independent variables, ***Age***



- An unbiased model has residuals that are randomly scattered around 0
  - Our residuals follow some sort of pattern

- While backwards model selection is a good method to find models with low AIC and high R-Squared, there are clearly limitations to this method

Selecting and fine-tuning new regression models would be a worthwhile next step for our project!