Malak Gaballa – 900201683 Masa Tantawy – 900201312

Applied Regression Methods
Dr.Ali Hadi

Data Description

Data Source: Kaggle

https://www.kaggle.com/datasets/rishidamarla/fifa-players-ratings?select=fifa_cleaned.csv

The dataset describes the overall rating for football players worldwide and some attributes. The original dataset contains 92 columns and 17954 observations. However, we will prepare the data beforehand to work with it. Thus we will end up with,

Number of Observations: 15889

Total Number of Variables: 46

Response Variable: Overall Rating (0-100)

Number of Predictor Variables: 44 (Player Name not included)

Variables

	Name	Description	Type	Units of Measurement
1.	age	Player Age	Quantitative	Years
2.	height_cm	Height of Player in centimeters	Quantitative	Centimeters (cm)

3.	weight_kgs	Weight of Player in kilograms	Quantitative	Kilograms (kg)
4.	potential	Absolute maximum rating that the player can reach through his career given his talents	Quantitative	0-100
5.	value_euro	Monetary Value of the player in the market	Quantitative	Euros
6.	wage_euro	Wages that the player earns	Quantitative	Euros
7.	preferred_foot	Player's preferred foot	Categorical 2 levels	Right (R) or Left (L)
8.	international_reputation.1.5.	Player Reputation Worldwide	Ordinal 5 levels	1-5 (very Low to very High)
9.	weak_foot.1.5	Player's shot power and ball control for the other foot of that player than his preferred foot.	Ordinal 5 levels	1-5 (very Low to very High)
10.	skill_moves.1.5.	Player's Skill Variety	Ordinal 5 levels	1-5 (Not Skilled to Very Skilled)

11.	work_rate	the extent to which a player contributes to running and chasing in a match while not in possession of the ball.	Ordinal 9 levels	Low/Low
12.	release_clause_euro	Player's set fee that a buying club can pay a selling club in order to contractually oblige them to offload a player or a coach	Quantitative	Euros
13.	club_rating	Club performance rating	Quantitative	0-100
14.	club_jersey_number	Jersey number worn by the player to represent his club	Quantitative	0-100
15.	national_rating	Rating on the national level	Quantitative	0-100
16.	crossing	Player's ability to cross the ball accurately from wide areas.	Quantitative	0-100
17.	finishing	Ability to put the ball in the back of the net when presented with a chance.	Quantitative	0-100
18.	heading_accuracy	Ability to head the ball with precision and control	Quantitative	0-100

19.	short_passing	Player's accuracy and speed of passing over a short distance.	Quantitative	0-100
20.	volleys	Accuracy and power of volleys at goal. It affects the technique and accuracy of shots taken while the ball is in the air.	Quantitative	0-100
21.	dribbling	Ability to run with the ball and manipulate it under close control.	Quantitative	0-100
22.	curve	Player's ability to curve the ball when passing and shooting	Quantitative	0-100
23.	freekick_accuracy	player's accuracy for taking Free Kicks	Quantitative	0-100
24.	long_passing	Player's ability to perform a long pass in the air to his teammate	Quantitative	0-100
25.	ball_control	ability of a player to control the ball as he receives it.	Quantitative	0-100
26.	acceleration	How quickly a player can get to top speed from a standing start.	Quantitative	0-100

27.	sprint_speed	how fast the player runs while at top speed.	Quantitative	0-100
28.	agility	How well a player can start, stop and move in different directions at varying levels of speed both on and off the ball.	Quantitative	0-100
29.	reactions	How quickly a player responds to a situation happening around him.	Quantitative	0-100
30.	balance	How well a player can stay on his feet, both on and off the ball.	Quantitative	0-100
31.	shot_power	How hard the player hits the ball when taking a shot at goal. It is the amount of power a player can put into a shot while still keeping it accurate.	Quantitative	0-100
32.	jumping	The highest point a player can reach with his head, often influenced by a player's height.	Quantitative	0-100

33.	stamina	Player's ability to endure high levels of physical activity for extended periods of time.	Quantitative	0-100
34.	strength	The player's ability to exert his physical force on an opponent to his benefit.	Quantitative	0-100
35.	long_shots	Accuracy of shots from outside the penalty area	Quantitative	0-100
36.	aggression	Player's power of will or commitment to a match.	Quantitative	0-100
37.	interceptions	Ability to read the game and intercept passes.	Quantitative	0-100
38.	positioning	The ability of a player to read a situation and maneuver themselves into the best location to deal with unfolding events.	Quantitative	0-100
39.	vision	Ability to see a potential opening and spot an opportunity another player may not have seen.	Quantitative	0-100

40.	penalties	Accuracy of a penalty kick	Quantitative	0-100
41.	composure	The player's steadiness of mind and ability to make intelligent decisions with or without the ball.	Quantitative	0-100
42.	marking	The ability to stick close to his direct opposition in defensive situations.	Quantitative	0-100
43.	standing_tackle	Ability of the player to time standing tackles so that they win the ball rather than give away a foul.	Quantitative	0-100
44.	sliding_tackle	Ability of the player to time sliding tackles so that they win the ball rather than give away a foul	Quantitative	0-100

Correlation Matrix

Overall rating is the response variable. For correlation matrix, index plots of predictors and plots of response variable vs each predictor, please refer to the files attached accordingly.

Applied Regression Methods
Dr.Ali Hadi

Relationship between predictor and response variables

Note that the variables were standardized prior to plotting the data as shown below. However, even if the variables were not standardized in the first place, the plots will still be the same.

	Name	Relationship
1.	age	As the age increases, the overall rating slightly increases since the player gains more experience.
2.	height_cm	The player's height doesn't affect the overall rating.
3.	weight_kgs	The weight has no effect on the overall rating
4.	potential	The higher the player's potential, the greater the overall rating.
5.	value_euro	As the player's value in the market increases, the overall rating increases as well. However, according to the graph, there are some anomalies.
6.	wage_euro	As the player's salary increases, the overall rating increases as well. However, according to the graph, there are some anomalies.

7.	preferred_foot	The player's preferred foot doesn't affect the overall rating.
8.	international_reputation.1.5.	As international reputation increases, overall rating increases.
9.	weak_foot.1.5	The ability to play with one's weak foot has no effect on the player's overall rating.
10.	skill_moves.1.5.	The more skilled a player is, the higher the overall rating.
11.	work_rate	There is almost no relationship between the work rate of a player and their rating.
12.	release_clause_euro	The higher the player's release clause, the higher the overall rating since his value in the market is high. However, according to the graph, there are some anomalies.
13.	club_rating	As the player's rating with the club increases, the overall rating increases as well.
14.	club_jersey_number	A player's jersey number with the club doesn't have an effect on the overall rating.
15.	national_rating	As the player's rating on the national team level increases, the overall rating increases as well.
16.	crossing	The higher the crossing attribute, the higher the overall rating.

17.	finishing	As finishing increases, overall rating increases.
18.	heading_accuracy	The higher the heading accuracy attribute, the higher the overall rating.
19.	short_passing	As short passing increases, overall rating increases.
20.	volleys	The higher the volleys attribute, the higher the overall rating.
21.	dribbling	As dribbling increases, overall rating increases.
22.	curve	The higher the player's curve attribute, the higher the overall rating.
23.	freekick_accuracy	The correlation is very weak. As freekick accuracy increases, overall rating increases.
24.	long_passing	The higher the long passing attribute, the higher the overall rating.
25.	ball_control	As ball control increases, overall rating increases.
26.	acceleration	The player's acceleration doesn't affect the overall rating.
27.	sprint_speed	There is almost no correlation between sprint speed and overall rating.
28.	agility	Agility does not affect overall rating.

29.	reactions	As reactions increase, overall rating increases.
30.	balance	Player's balance doesn't affect the overall rating
31.	shot_power	As shot power increases, overall rating increases.
32.	jumping	There is almost no correlation between jumping and overall rating.
33.	stamina	The higher the stamina, the higher the overall rating.
34.	strength	There is almost no correlation between strength and overall rating.
35.	long_shots	As long shots increase, overall rating increases.
36.	aggression	The more aggressive the player is, the higher the overall rating.
37.	interceptions	If the player is able to intercept passes, the overall rating increases.
38.	positioning	As positioning increases, overall rating increases.
39.	vision	As vision increases, overall rating increases.
40.	penalties	As the player's penalty kick accuracy increases, the overall rating increases.
41.	composure	As composure increases, overall rating increases.

42.	marking	The higher the player's ability to stick close to his direct opposition in defensive situations, the greater the overall rating.
43.	standing_tackle	As the standing tackle attribute increases, the overall rating slightly increases.
44.	sliding_tackle	As the sliding tackle attribute increases, the overall rating slightly increases.

Summary of each variable

age

Mean= 25.45006 Median= 25

Min= 17 Max= 42

Standard Deviation= 4.580574

height_cm

Mean= 173.4328 Median= 175.26

Min= 152.4 Max= 203.2

Standard Deviation= 13.81423

weight_kgs

Mean= 74.44273 Median= 73.9

Min= 49.9 Max= 110.2

Standard Deviation= 6.718296

potential

Mean= 71.63918 Median= 71

wage_euro

Mean= 10548.63 Median= 3000

Min= 1000 Max= 565000

Standard Deviation= 23468.58

release_clause_euro

Mean= 5036639 Median= 1400000

Min= 18000 Max= 226500000

Standard Deviation= 11805275

club_rating

Mean= 69.36665 Median= 69

Min= 54 Max= 86

Standard Deviation= 5.106742

club_jersey_number

Mean= 20.15983 Median= 17

Min= 2 Max= 99

Standard Deviation= 16.06168

Min= 51 Max= 95

Standard Deviation= 6.073455

value_euro

Mean= 2668088 Median= 775000

Min= 10000 Max= 110500000

Standard Deviation= 6099490

national_rating

Mean= 72.2738 Median= 71.92036

Min= 63 Max= 85

Standard Deviation= 2.786233

crossing

Mean= 54.30562 Median= 57

Min= 11 Max= 93

Standard Deviation= 14.11943

finishing

Mean= 49.70835 Median= 52

Min= 10 Max= 95

Standard Deviation= 16.40468

heading_accuracy

Mean= 57.07187 Median= 58

Min= 18 Max= 94

Standard Deviation= 11.5839

volleys

Mean= 46.72113 Median= 47

Min= 10 Max= 90

Standard Deviation= 14.79635

dribbling

Mean= 60.67682 Median= 63

Min= 16 Max= 97

Standard Deviation= 12.49505

curve

Mean= 51.325 Median= 52

Min= 11 Max= 94

Standard Deviation= 15.18741

freekick_accuracy

Mean= 46.37082 Median= 44

Min= 10 Max= 94

Standard Deviation= 15.107

short_passing

Mean= 62.65133 Median= 64

Min= 20 Max= 93

Standard Deviation= 9.825192

long_passing

Mean= 56.16603 Median= 58

Min= 19 Max= 93

Standard Deviation= 12.39853

ball_control

Mean= 63.21197 Median= 64

Min= 25 Max= 96

Standard Deviation= 10.04894

acceleration

Mean= 68.17622 Median= 69

Min= 20 Max= 97

Standard Deviation= 11.54564

sprint_speed

Mean= 68.2537 Median= 69

Min= 25 Max= 96

Standard Deviation= 11.22298

agility

Mean= 66.41815 Median= 68

Min= 23 Max= 96

balance

Mean= 66.59211 Median= 68

Min= 21 Max= 96

Standard Deviation= 12.12374

shot_power

Mean= 59.57971 Median= 61

Min= 15 Max= 95

Standard Deviation= 13.27369

jumping

Mean= 65.92731 Median= 67

Min= 25 Max= 95

Standard Deviation= 11.37413

stamina

Mean= 67.41412 Median= 68

Min= 28 Max= 97

Standard Deviation= 10.91952

sliding_tackle

Mean= 49.83039 Median= 56

Min= 10 Max= 90

Standard Deviation= 19.04257

Standard Deviation= 12.26463

reactions

Mean= 62.22789 Median= 62

Min= 30 Max= 96

Standard Deviation= 8.800826

strength

Mean= 65.77815 Median= 67

Min= 25 Max= 97

Standard Deviation= 12.57631

long_shots

Mean= 51.32557 Median= 54

Min= 11 Max= 94

Standard Deviation= 15.81702

aggression

Mean= 59.667 Median= 61

Min= 13 Max= 95

Standard Deviation= 14.33014

interceptions

Mean= 50.53005 Median= 56

Min= 10 Max= 92

Standard Deviation= 18.75376

positioning

Mean= 54.89452 Median= 57

penalties

Mean= 52.05897 Median= 52

Min= 11 Max= 92

Standard Deviation= 12.53304

composure

Mean= 60.52653 Median= 61

Min= 30 Max= 96

Standard Deviation= 10.15364

marking

Mean= 51.35125 Median= 56

Min= 10 Max= 94

Standard Deviation= 17.23745

standing_tackle

Mean= 52.10825 Median= 59

Min= 10 Max= 93

Standard Deviation= 19.05165

vision

Mean= 55.51847 Median= 57

Min= 12 Max= 94

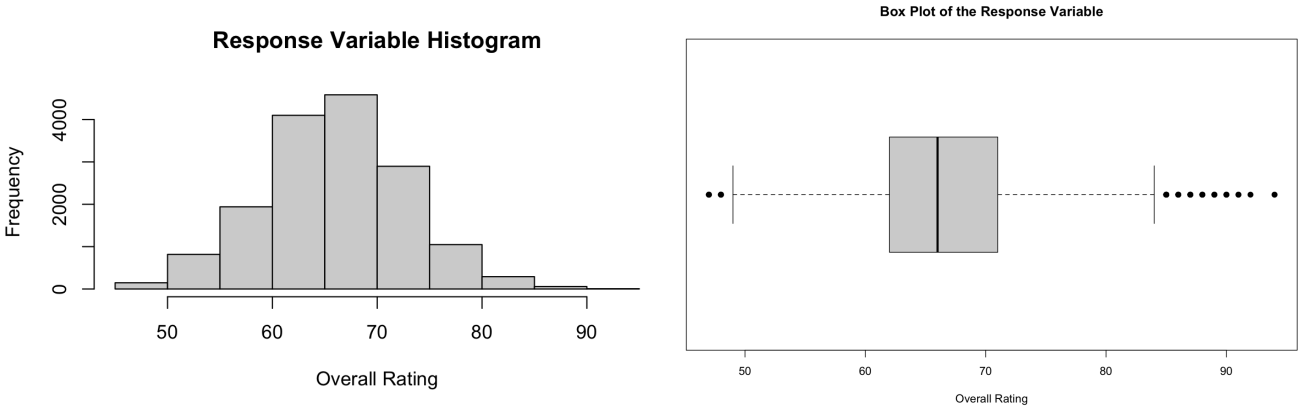
Standard Deviation= 12.90704

Applied Regression Methods
Dr.Ali Hadi

Min= 11 Max= 95

Standard Deviation= 14.67517

Histogram and Box Plot of the Response Variable



According to the histogram above, the overall rating almost follows a normal distribution and is approximately symmetric. The majority of the players (almost 4000-5000 players) have ratings between 60 and 70 while very few players are rated below 50 and even fewer players are rated above 85. There are no players with ratings below 40 but there is a very minute number of players with a rating above 90.

As for the box plot, the ratings fall within the range of 47 to 94; half of the ratings are greater than approximately 66 (the median) while the other half is less than 66. Also, there are outliers after both whiskers, which indicates that some players have very low or very high ratings compared to the other player's, making them outliers. The majority of the outliers are the players with very high ratings (83+) while very few have ratings that are below 50. Since the median is in the middle of the box, and the whiskers are of about the same length on both sides of the box, so the distribution of the response variable is almost symmetric as seen in the histogram.