Analysis of FIFA 23 Dataset

1. Introduction:

The FIFA 23 dataset provides comprehensive information on players, clubs, nationalities, and various attributes related to player performance. This report aims to present the key findings and their implications for team selection, player scouting, and identifying areas for improvement within a team. The analysis is based on the FIFA 23 dataset, which includes player information, club information, nationality and national team information, and a wide range of attributes.

* 1. Problem, Goals and Audiences
* Understanding player performance and attribute relationships using the FIFA 23 dataset.
* Highlight the data-driven insights to make informed decisions in team selection and player recruitment.
  1. Goals:
* To gain valuable insights from the FIFA 23 dataset for team selection, player scouting, and identifying areas for improvement.

The analysis was carried out in the part of the Immersive Data Analytics Certificate Programme. The intended audience is both technical and non-technical. This report is supplemented with a data dictionary Excel file, SQL file, Python notebook, and Power Point Presentation.

1. Patterns, trends, and insights

The project was carried out in several stages, including interpreting the data, verifying, cleaning, and preparing the data, summarising and comparing the data, examining correlations and relationships in the data, and evaluation, and finally reporting and recommending insights. To create a database, PostgreSQL server was used followed by a table creation for the FIFA datasets. The FIFA 23 complete player dataset was obtained from Kaggle which was originally scraped from the publicly available website sofifa.com (<https://www.kaggle.com/datasets/stefanoleone992/fifa-23-complete-player-dataset>).

The dataset included files for teams, coaches, and players as well as information for male and female athletes. It contains detailed information about every player in the game, including their attributes, positions, clubs, and national teams. The data was cleaned using PostgreSQL and exported to csv file so that it can be analysed using visualisation tool, Tableau.

In order to detect correlations, the data was then exported again to Python for a more intricate but preliminary visualisation. Additionally, cleaning was done, and a second version (V2) was produced for additional Tableau analysis. Visualisations were created using Tableau, as well as a presentation deck.

**3.1 Exploratory Data Analysis**

From the above EDA analysis few significant correlations between the distinctive variables contained within the dataset. For instance;

Overall Rating and Potential: There was a strong positive correlation (0.74) between a player's overall rating and their potential rating, indicating that players with higher overall ratings tend to have higher potential ratings.

Overall Rating and Value: A moderate positive correlation (0.69) existed between a player's overall rating and their market value in Euros, suggesting that players with higher overall ratings tend to have higher market values.

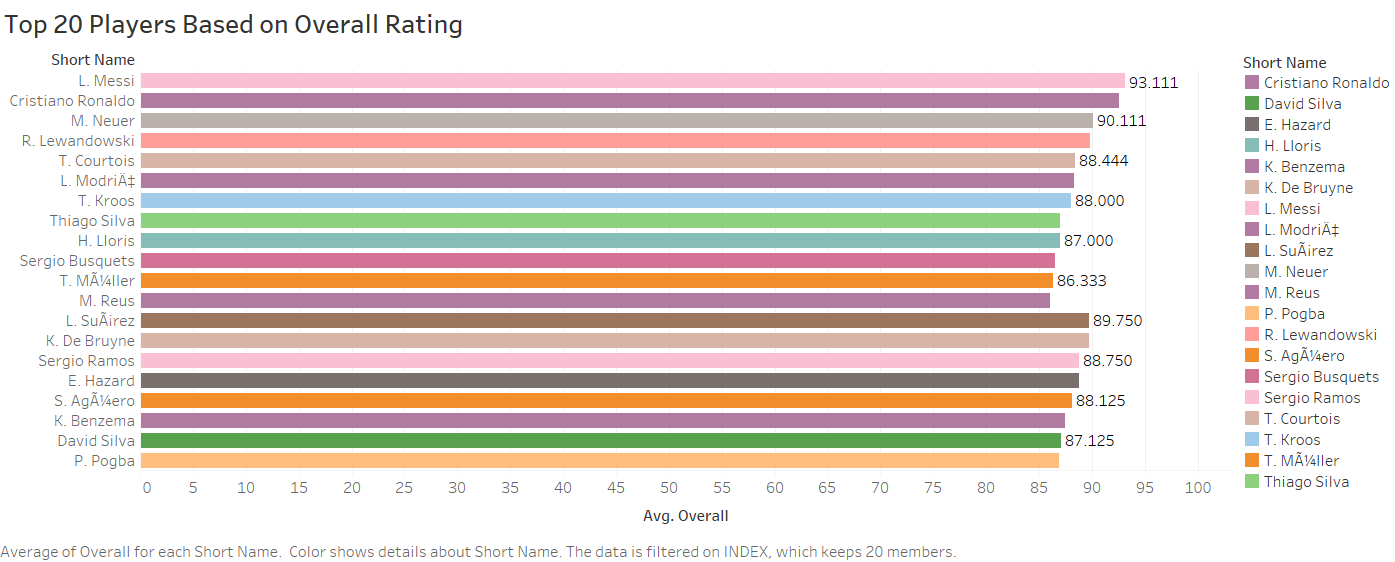
Overall Rating and Wage: A moderate positive correlation (0.71) was observed between a player's overall rating and their wage in Euros, indicating that players with higher overall ratings tend to earn higher wages.

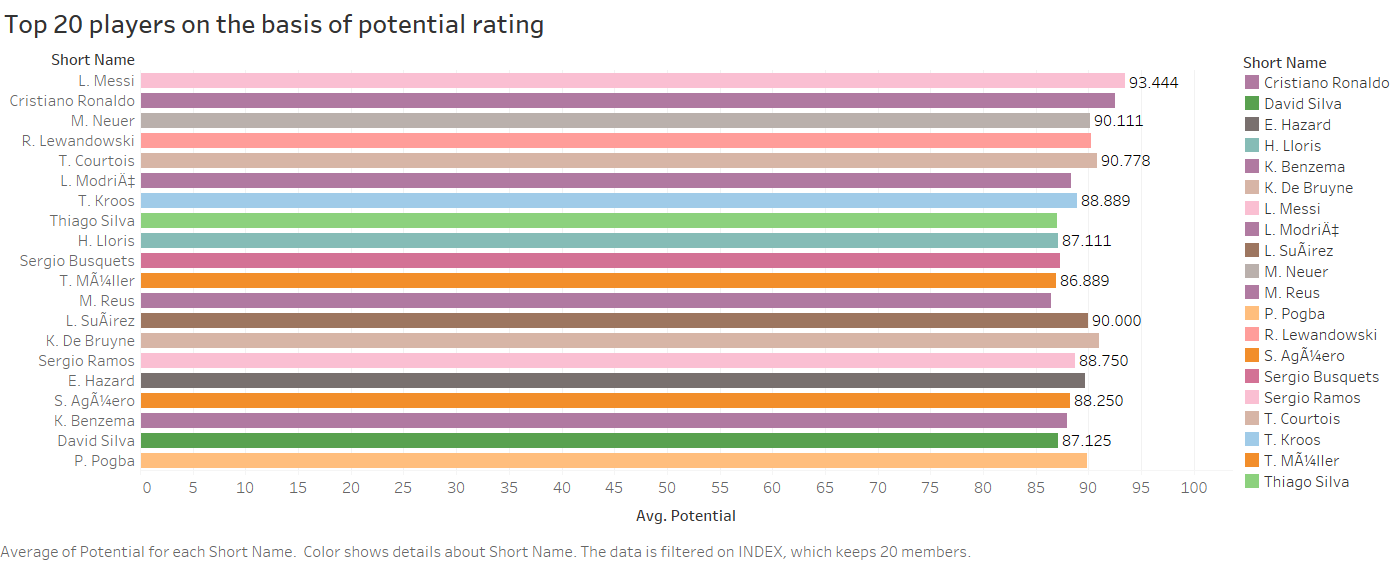
Overall Rating and Age: There was a moderate positive correlation (0.42) between a player's overall rating and their age, suggesting that younger players tend to have higher overall ratings.

Overall Rating and Height/Weight: The correlations between overall rating and height or weight were very weak (around 0.02 and 0.12, respectively), indicating that there was no substantial relationship between these attributes.

**3.2 Trends and Insights**

These EDA findings were further validated using Tableau to gain more insights on the pattern and trends in the data points. For instances,

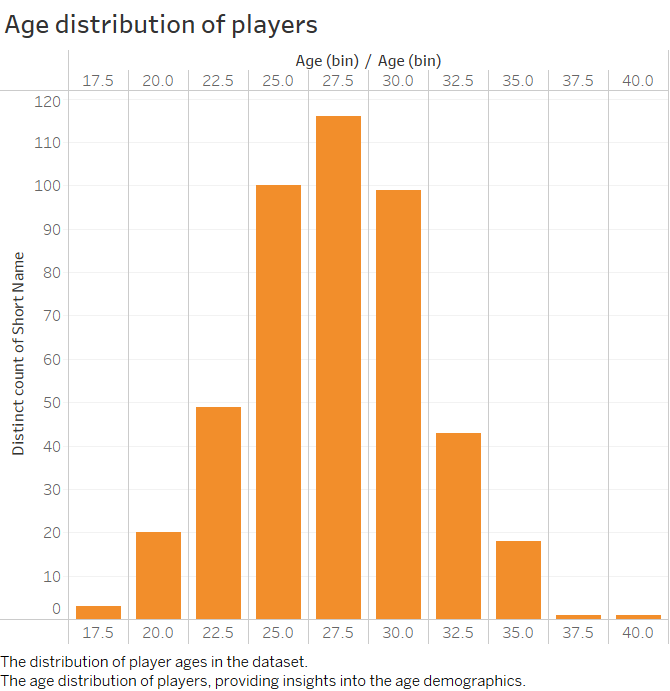




The top 20 players overall and potential ratings showed the same listed athletes. Hence, Overall ratings were considered for the relational analysis. Also,

* The overall rating is a comprehensive assessment of a player's skill level, considering various attributes and performance indicators:
* Technical (pace, shooting, passing, dribbling, defending, and physic)
* Physical (strength, agility, stamina, and speed)
* Mental abilities (positioning, vision, and composure)
* A higher overall rating indicates a more skilled and accomplished player who can make a significant impact on the field.
* The potential rating considers age, skill progression, and performance trajectory to estimate a player's future growth and improvement.
* A higher potential rating suggests a player's potential to become an elite performer, making them an attractive investment for clubs and teams.

The age distribution of players revealed that there were players across a wide range of ages, including both younger players and older players.



**The relationship between overall rating and Age**

A graph with red dots

Description automatically generated

* The overall ratings vary across different ages, indicating that the skill level of players can differ significantly based on age.
* There is no clear linear relationship between age and overall rating. For instance, we can see players at age 21 with overall ratings ranging from 68 to 85, showing that overall rating is influenced by factors other than age alone.

Please note that rest of the analysis has been provided in the presentation slide deck.

1. Conclusion and next steps

* Overall rating serves as a crucial indicator of a player's potential, market value, wage, and age.
* The analysis revealed strong positive correlations between overall rating and potential, market value, and wage.
* Height and weight show weak correlations with overall rating, indicating that skill and performance are not strongly influenced by these physical attributes.
* Skill attributes such as pace, passing, and power have varying impacts on ball control.
* Some physical characteristics may contribute or influence the overall rating based the position of the player.
* Understanding these findings can aid in team selection, player scouting, and decision-making processes in the football industry.
* Clubs and teams can make more informed choices when recruiting and developing players based on these insights.
* Further analysis will be conducted to explore the interrelationships between different attributes and uncover additional insights to enhance player evaluation and performance analysis.
* Model-based approaches, such as player position modelling, will be employed to gain a deeper understanding of the best player positions based on the dataset, which includes multiple positions for players.