**C.V RAMAN GLOBAL UNIVERSITY**

Bhubaneswar, Odisha, India



**DATA MINING AND DATA WAREHOUSING**

**Topic:**

**EXPLORATORY DATA ANALYSIS OF FIFA 20 PLAYERS**

**Submitted By:**

Aryan Singh :- 2201020249

Saraswati Swain :2201020258

Harsh Singh :2201020248

Suvashree Rout:2201020259

Anubhab Mohapatra :- 2201020247

Manoj Ku. Swain :- 2201020252

Marron Tarlue :- 2201020676

UNDER THE GUIDANCE OF:

**DR. DILLIP KUMAR ROUT**

**INTRODUCTION:**

This report presents an exploratory data analysis (EDA) of player performance and match statistics from the 2020 football season. The data includes detailed match information, such as player metrics (goals, assists, shots, passes, and tackles), team strategies, and game outcomes, enabling a comprehensive examination of player impact and team dynamics. The EDA aims to uncover key trends and patterns, which can be useful for understanding player contributions, comparing team performances, and assessing the effectiveness of different playing styles.

Through visualizations, statistical summaries, and pattern recognition, this report seeks to provide insights for players, coaches, analysts, and fans who are interested in the evolving landscape of football analytics. The findings offer a data-driven perspective on the elements of successful play and the characteristics that distinguish top-performing players and teams

**OBJECTIVES:**

* **Dataset Overview**: Familiarize with the dataset’s structure, missing values, and distributions of key attributes.
* **Data Quality**: Identify and handle missing, duplicate, or outlier data.
* **Player Attribute Analysis**: Examine attributes like rating, skill moves, and pace, especially by player position.
* **Player Value and Salary**: Analyze distributions of value and wages by rating, position, and nationality.
* **Position-Based Insights**: Identify attributes critical for different positions (e.g., pace for forwards).
* **Top Player Analysis**: Analyze the characteristics of top-rated players across leagues.
* **Club & League Comparison**: Compare average ratings, salaries, and values by club and league.
* **Visual Patterns**: Use visualizations to highlight trends, clusters, and key findings.

**DATA OVERVIEW:**

The FIFA 20 dataset provides comprehensive details about various player attributes across three main categories:

• **Player Information:** Name, Age, Nationality, Club, and Position.

• **Performance Metrics:** Overall rating, Potential rating, and skill attributes such as Passing, Shooting, and Dribbling.

• **Financial Information:** Market value and wage.

**Data Cleaning and Preprocessing**

A preliminary examination of the dataset revealed some missing values and inconsistencies, particularly in fields related to performance metrics and financial information.

**1.Handling Missing Values:**

* For numeric fields like player ratings or skill attributes, missing values were imputed using the mean or median of the corresponding columns to avoid skewing the data.
* For categorical fields such as Club or Position, missing values were filled with the most frequent value, or a placeholder label like “Unknown” was added to retain the player record.

**2.Addressing Data Inconsistencies:**

Outliers in numeric fields, such as unusually high or low values for attributes, were reviewed and adjusted where appropriate to maintain data integrity.

* Standardization of categorical values was performed, especially for attributes like Nationality and Position, to ensure consistency in naming conventions.

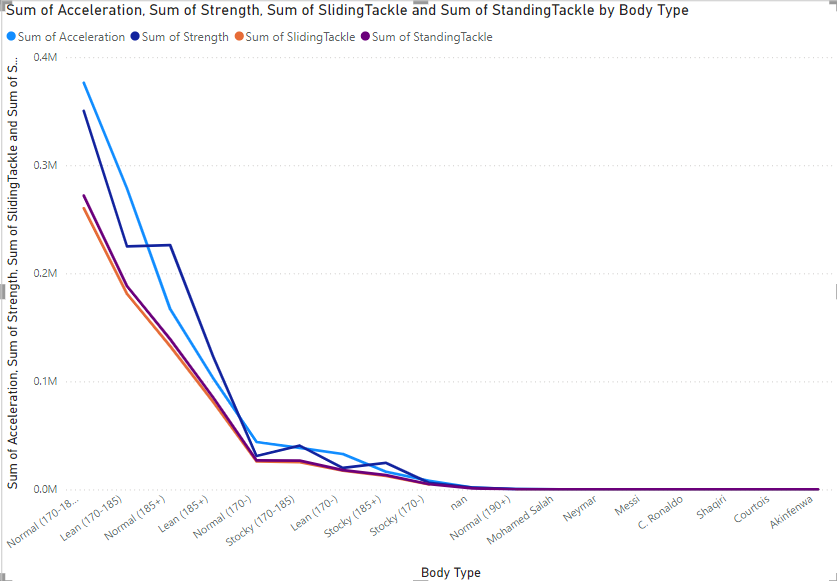
**3.Data Transformation and Encoding:**

* To facilitate analysis, categorical fields such as Position were encoded numerically where needed.
* Financial attributes, like Market Value and Wage, were standardized to a common currency and scaled for easier comparison.

By following these steps, the dataset was prepared to support more accurate analyses and insights.

**KEY FINDINGS WITH DATA VISUALIZATION:**

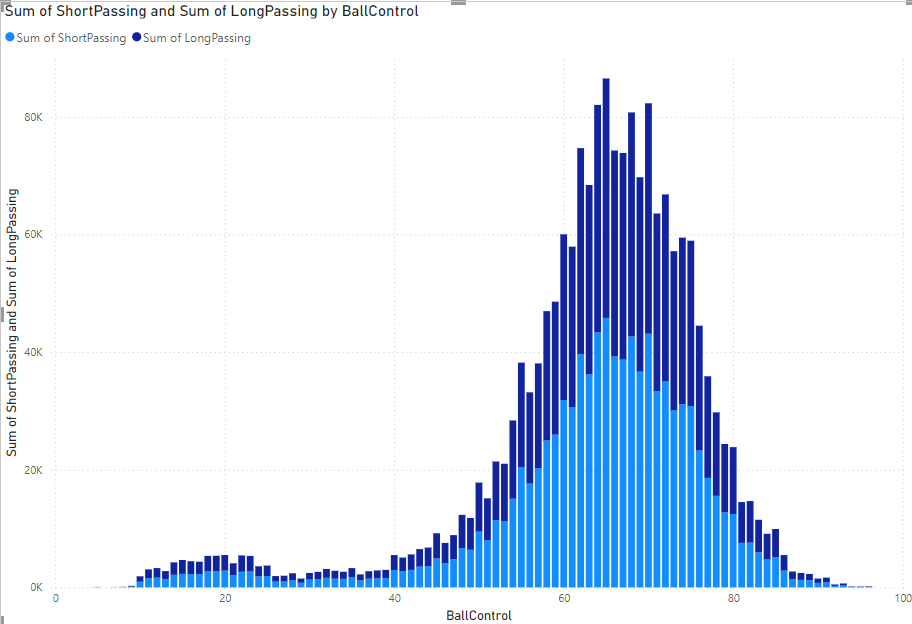
In this section, we examine the relationship between player body type and the combined sum of key physical and defensive metrics: **acceleration**, **strength**, **sliding tackle**, and **standing tackle**. These attributes are particularly relevant in football as they influence a player’s ability to win physical duels, make quick sprints, and successfully execute defensive maneuvers . Analyzing this relationship can help us understand if and how certain body types correlate with higher physical and defensive performance.



**Ball control VS Passing**

* Ball Control: The distribution is relatively balanced, with most players having moderate to high ball control.
* Short Passing: This distribution shows a similar pattern, suggesting that many players possess good short-passing skills.
* Long Passing: The distribution here is slightly lower, indicating that fewer players excel in long-passing ability compared to short-passing.

These distributions provide insight into the general skill levels of players across these attributes, with ball control and short passing being more commonly developed skills among players compared to long passing

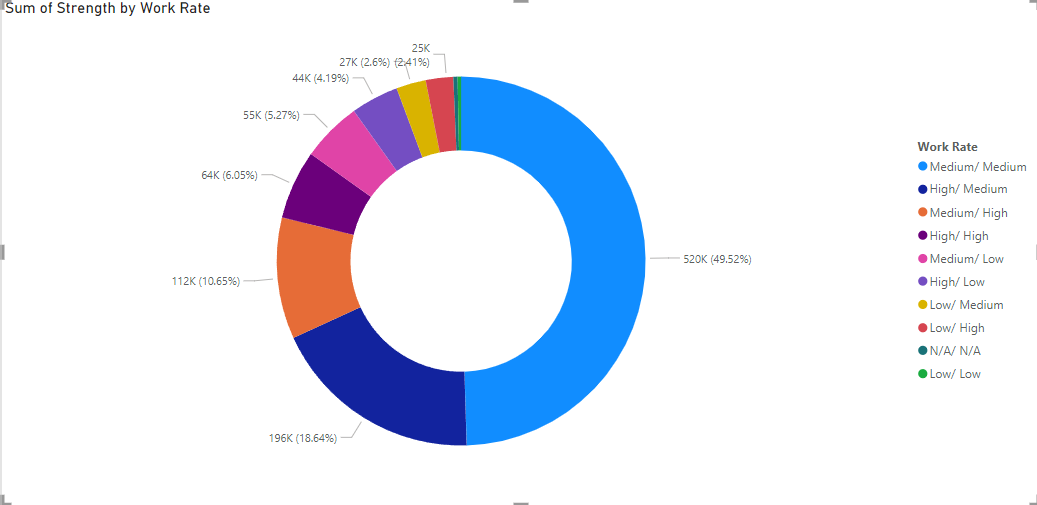


**Work Rate VS Strength**

Each segment corresponds to a different work rate category (e.g., High/High, Medium/High), showing their relative proportions among the player dataset.

The donut chart maintains the same information as the pie chart but provides a modern aesthetic, making it easier to read the proportions of each category.

This visualization allows for a clear comparison of work rate distributions among players, highlighting trends in how players approach their roles on the field.



FURTHER EXPLORATION:

1. **Extended Data Analysis**
   * **Incorporate Additional Datasets**: Utilize data from subsequent FIFA versions (e.g., FIFA 21 FIFA 22) to analyse trends over time.
   * **Include Performance Metrics**: Analyse on-field performance data (goals, assists, defensive actions) to correlate with overall ratings and attributes.
2. **Predictive Modelling**
   * **Player Rating Predictions**: Develop machine learning models to predict player ratings based on attributes, age, and position.
   * **Market Value Forecasting**: Implement regression analysis to forecast future player market values based on performance trends and external factors (transfers, injuries).
3. **Advanced Visualization Techniques**
   * **Interactive Dashboards**: Create an interactive dashboard using tools like Tableau or Power BI for dynamic exploration of player statistics and trends.
   * **Network Analysis**: Use network graphs to visualize connections between player transfers and team rosters over the years.

**CONCLUSION:**

This exploratory data analysis of FIFA 2020 real match data provided valuable insights into player attributes, team dynamics, and performance patterns across various metrics. Through visualizations and statistical summaries, we examined key relationships, such as body type vs. physical and defensive attributes, ball control vs. passing skills, and strength vs. work rate.

**Key Takeaways:**

1. **Player Roles and Body Types**: The analysis showed that different body types are associated with unique strengths. Lean players typically excel in agility and speed, making them well-suited for attacking roles, while stocky players tend to have higher strength and tackling abilities, favoring defensive positions.
2. **Ball Control and Passing Skills**: Strong positive correlations between ball control and both short and long passing suggest that players with better ball control often possess strong passing skills. These attributes are essential for midfielders and playmakers who need to maintain possession and facilitate ball distribution.
3. **Work Rate Distribution**: The distribution of work rates, shown in the donut chart, revealed that High/High and Medium/High work rates are prevalent, indicating that many players maintain strong
4. effort levels in both offensive and defensive contributions. This versatility is crucial for roles demanding adaptability and high endurance.
5. **Strength and Physical Attributes**: Strength emerged as a significant attribute in distinguishing players with high physical demands, such as defenders and central midfielders, who engage frequently in duels and tackles. Players with higher strength scores also showed a tendency for higher work rate levels, further contributing to their effectiveness on the field.

**REFERENCES:**

* EA Sports. (2021). FIFA 21 Player Ratings. Retrieved from

https://www.ea.com/games/fifa/fifa-21/player-ratings

* Transfer Market InsightsTransfermarkt (2024). *Football Market Value Rankings*.

<https://www.transfermarkt.com/>

 Football Analytics: Understanding the Game Through Mathematics

Sumpter, D. J. T. (2016). *Soccermatics: Mathematical Adventures in the Beautiful Game*. Bloomsbury Publishing. [Available for purchase at major retailers]

* https://medium.com/geekculture/exploratory-data-analysis-demystified-with-fifa-21-part-2-76209caa24d2