

FIFA 22-23 RAC PROJECT

Submitted To: Prof. Amarnath Mitra & Prof. Ashok Harnal

Submitted By: Group 6

|  |  |
| --- | --- |
| Name | Roll no |
| Adhyatik | 311063 |
| Aashit Sharma | 321148 |
| Arindam Chakraborty | 321128 |
| Subhajit Paul | 321173 |

|  |  |
| --- | --- |
| TABLE OF CONTENTS | Page No. |
| COVER PAGE | 1 |
| TABLE OF CONTENTS | 2 |
| ACKNOWLEDGEMENT | 3 |
| EXECUTIVE SUMMARY | 4 |
| OBJECTIVES | 5 |
| DESCRIPTION OF DATA | 6 |
| ANALYSIS OF DATA | 14 |
| OBSERVATIONS | 22 |
| MANAGERIAL INSIGHTS | 46 |

EXECUTIVE SUMMARY

*This report presents a comprehensive FIFA players’ analysis for the Real World Analytics project using both supervised and unsupervised machine learning techniques. The objective was to Analyze Football Player Data, statistics, and characteristics, enabling to identify player’ trait and most successful players.*

*The analysis leveraged a dataset encompassing approximately 25000 records, each representing an individual ability.*

*The findings revealed three distinct customer segments:*

**Cluster 0: Established Stars** – Players who are already at peak in their profile

**Cluster 1: Raising Talents** – Players with high potential

**Cluster 2: Veteran Leaders** - Refers to experienced players who has demonstrated exceptional leadership skills, often acquired through years of playing in the league.

**Cluster 3: Mid-Career Players** – Players who are at peak and have long career ahead.

**Cluster 4: International Prospect** – Players with high progress in International career but not in focus in leagues.

**Cluster 5: Domestic Role Player** – Players who excel in specific roles within a team’s formation, particularly in set-piece situations such as free kicks, corners, and penalties.

**Cluster 6: Journeymen -** In FIFA, a journeyman player is a fictional character who has played for multiple teams throughout their career, often without achieving significant success or recognition.

1.) OBJECTIVES

1.1 Creation of the analytical dashboard and analysis of

a. Analyze Football Player Data: Provide a comprehensive view of player attributes, club positions, and league performances.

b. Visualize Key Patterns: Highlight patterns in overall ratings, club position distributions, and league-wise average ratings.

c. Explore Attribute Correlations: Investigate the relationship between player height and overall rating to identify potential performance trends.

d. Player attribute correction and overall player Analysis

e. player value prediction

f. Support Strategic Decisions: Offer valuable insights for player comparison, aiding in informed decision-making for team management and player selection.

1.2.) Developing a Machine learning model using unsupervised K-means and Supervised Decision tree.

1. In k- means clustering we identified the optimal no. of clusters and defined the characteristics of the clusters
2. In decision tree we identified the significant variables from the identified cluster and also found out the accuracy and significance of the model

2.) DESCRIPTION OF DATA

Data Source, Size & Shape

**Data Source (Website Link):-** https://www.kaggle.com/datasets/sabir0000/male-football-players-data

**Data Size:-** 26 MB

**Data Description & Dimensions:-** This dataset, titled "Fifa Footbal players" provides a comprehensive view of football players. It encompasses approximately **100000+** records, each representing an individual player statistics. The primary objective of this dataset is to understand and predict the analysis of football players, a key indicator of each player statistics and player experience in the burgeoning field of sports.

**Number Of Observations Taken For Analysis**:- 25218

**Number Of Variables**:- **110**

**Description of Variables**

**Index Variable(s):** player\_id, fifa\_version, short\_name

**Categorical Variables or Features (CV): -** 09

league\_name—the name of the league in which the game is played

league\_level --

club\_name—The name of the club

club\_position - Position of the club

nationality\_name—the name of the country in which the player belongs

preferred\_foot—The preferred foot of the player

weak\_foot—weak foot of the player

skill\_moves—Skill moves

player\_face\_url-- Url link of the image

player\_traits\_1—traits of the player

player\_traits\_2-- traits of the player

player\_traits\_3—same

player\_traits\_4-- same

player\_traits\_5-- same

player\_traits\_6-- same

player\_traits\_7-- same

player\_traits\_8--same

Categorical Variables or Features - Nominal Type: - 09

league\_name—Name of the league in which the player plays

club\_name-- Name of the club in which the player plays

nationality\_name—Name of the country the player belongs to

preferred\_foot—preferred foot of the player

Categorical Variables or Features - Ordinal Type: -0

league\_level—Level of the league

. club\_position—Position of the club

weak\_foot—Weak foot of the player

Non-Categorical Variables or Features: 8

Overall—rating of the player

Potential—Potentialof the player

value\_eur-- Value of the player in Euros

wage\_eur-- wage of the player in Euros

age—Age of the player

height\_cm – height of the player in cm

weight\_kg—Weight of the player in KG

pace -- Pace of the player

shooting—shooting points of the player

passing—passing points of the player

dribbling—dribbling points of the player

defending – defending points of the player

physic -- Phyics of the plyer

attacking\_crossing -- attacking crossing stats of the player

attacking\_finishing -- attacking finishing of the player

attacking\_heading\_accuracy-- Actta

attacking\_short\_passing

attacking\_volleys

skill\_dribbling

skill\_curve

Skill\_fk\_accuracy- Measures a player's precision in taking free kicks, determining their ability to accurately target the goal or a teammate from a set piece.

Skill\_long\_passing- Represents the player's ability to deliver accurate and effective passes over long distances, typically from deep midfield positions.

Skill\_ball\_control- Indicates how well a player can maintain possession of the ball when receiving it or when under pressure, allowing them to maneuver efficiently in tight Movement\_acceleration- Reflects how quickly a player can reach top speed from a standstill, important for short bursts of pace to beat opponents.

Movement\_sprint\_speed- Measures the top speed a player can maintain during longer sprints, crucial for outrunning defenders or catching up to attackers.

Movement\_agility- Represents a player's ability to change direction quickly and smoothly, helping in dribbling past opponents or adjusting to on-field scenarios.

Movement\_reactions- Indicates how swiftly a player can respond to events on the pitch, such as loose balls, deflections, or changes in play direction.

Movement\_balance- Measures a player's stability and control when dribbling or when challenged by opponents, affecting their ability to stay on their feet.

Power\_shot\_power- Reflects the force behind a player's shots, determining how hard they can strike the ball towards the goal.

Power\_jumping- Indicates how high a player can leap, which is essential for winning aerial duels and heading the ball.

Power\_stamina- Measures a player's endurance, reflecting how long they can perform at their best without tiring during a match.

Power\_strength- Represents the physical power of a player, crucial for shielding the ball, winning physical battles, and holding off opponents.

Power\_long\_shots- Reflects the accuracy and effectiveness of a player's shots from outside the penalty area.

Mentality\_aggression- Measures a player's intensity and willingness to challenge opponents, reflecting their tenacity in duels.

Mentality\_interceptions- Indicates a player's ability to read the game and anticipate passes, enabling them to intercept the ball effectively.

Mentality\_positioning- Represents how well a player positions themselves during offensive play, making them more effective in goal-scoring opportunities.

Mentality\_vision- Reflects a player's ability to spot and execute passes that unlock defenses, contributing to creative playmaking.

Mentality\_penalties- Indicates the player's accuracy and composure when taking penalty kicks.

Mentality\_composure- Measures a player's calmness under pressure, affecting their performance in high-stress situations like 1v1s or critical moments.

Defending\_marking\_awareness- Reflects a player's ability to track and mark opponents, ensuring they stay close to them to prevent scoring chances.

Defending\_standing\_tackle- Indicates the effectiveness of a player's standing tackles, helping them dispossess opponents without committing fouls.

Defending\_sliding\_tackle- Measures a player's ability to successfully execute sliding tackles, allowing them to intercept or clear the ball while on the ground.

Goalkeeping\_diving- Reflects a goalkeeper's ability to make diving saves, crucial for stopping shots aimed at the corners of the goal.

Goalkeeping\_handling- Indicates how well a goalkeeper can catch or hold onto the ball, reducing the chance of rebounds or fumbles.

Goalkeeping\_kicking- Represents the accuracy and distance of a goalkeeper's kicks, important for clearances and starting counter-attacks.

Goalkeeping\_positioning- Measures a goalkeeper's ability to position themselves optimally to save shots and command their area effectively.

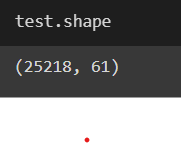
Goalkeeping\_reflexes- Reflects the quickness of a goalkeeper's reactions to make point-blank saves or respond to sudden changes in the trajectory of the ball.

Goalkeeping\_speed- Indicates the goalkeeper's ability to move quickly, essential for rushing off their line or closing down angles in 1v1 situation.

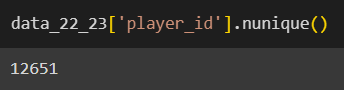
**Descriptive Statistics**

**Descriptive Statistics: Categorical Variables or Features**

**Count & Relative Frequency Statistics**



Count **Unique Player**

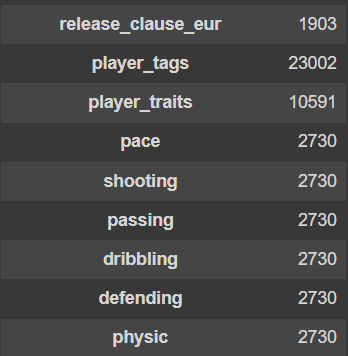
****

**Mean, Standard deviation, max**

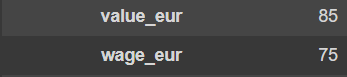
****

3.) ANALYSIS OF DATA

**Missing Data Statistics:**

****





Missing Data Treatment:

The missing values are both Categorical and non-categorical, we used **0 for categorical value by r**eplace missing values with the most frequent value (mode) in that variable.

**For non categorical variables, we used “Not Available” i.e.** Replacing missing values with the text.

Missing Data Exclusion:

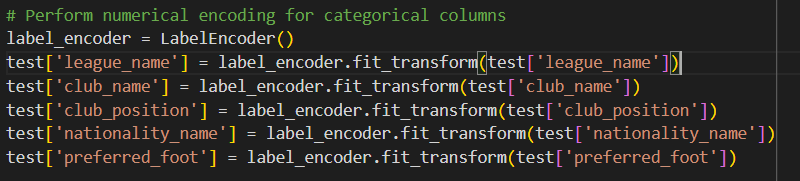
Deletion of Records: If any Record(Rows) contains more than 50% of null values those Records are removed from the dataset.

Variable Deletion: If any Variable(Columns) contains more than 50% of null values persist, the corresponding variables are considered for removal.

Following the implementation of deletion procedures, the dataset achieves completeness.

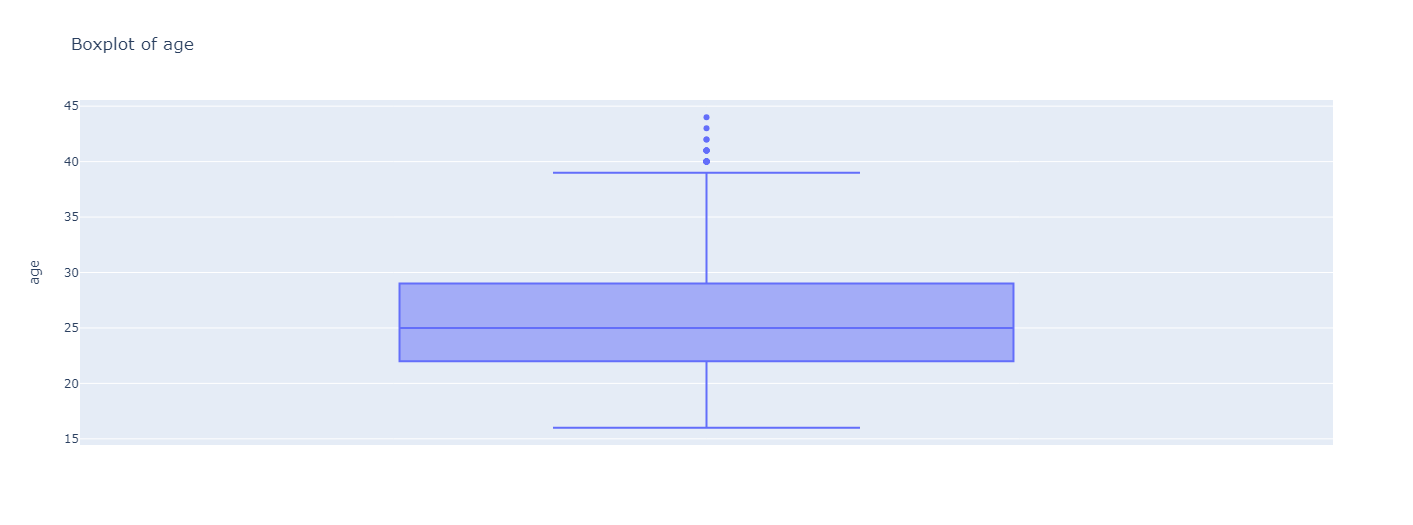
Numerical Encoding of Categorical Variables or Features

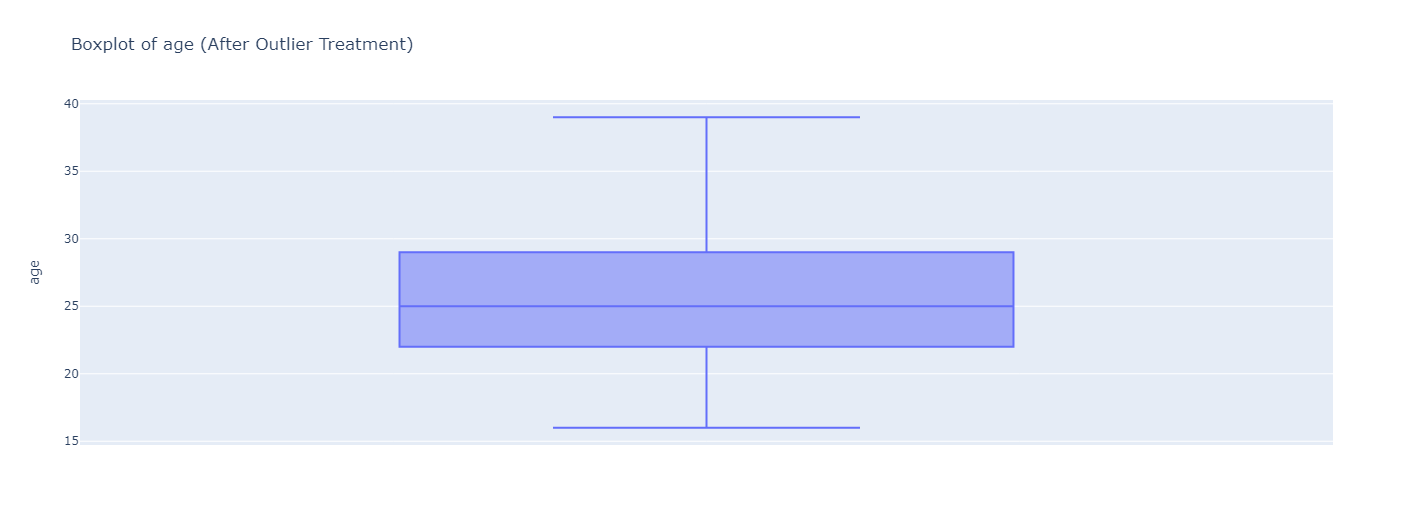
Categorical variables in the dataset which are nominal, we apply label encoding to transform them into numerical representations and the Encoding Schema is Alphanumeric Order.



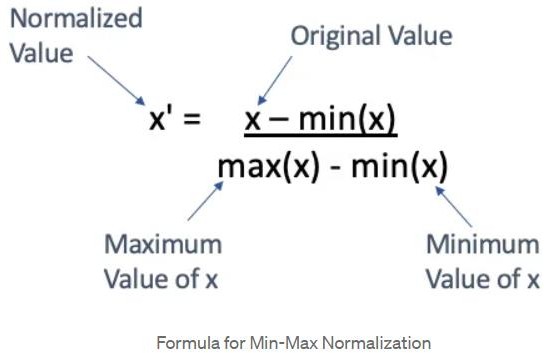
Outlier Statistics and Treatment

Outlier Statistics: Non-Categorical Variables or Features

****

****

The formula for Min-Max Normalization is:



Min-Max Scaling normalizes the data and removes outliers. We utilize the interquartile range (IQR) from the 25th to the 75th percentile.

Data Analysis

Unsupervised Machine Learning Clustering Algorithm: K-Means

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It is an unsupervised learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding

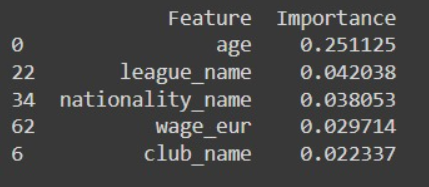
clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

Determines the best value for K center points or centroids by an iterative process.

Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

K- means Clusturing

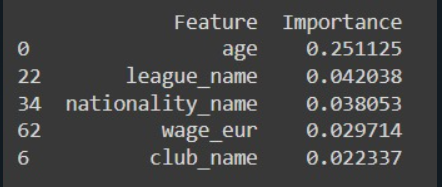




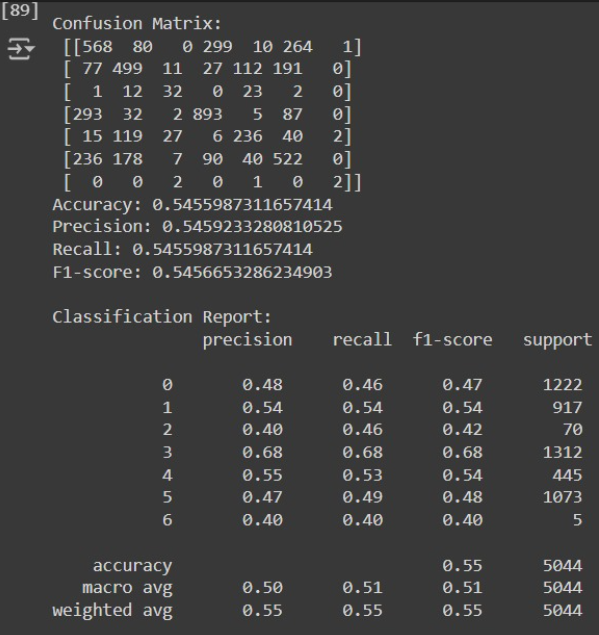
From the cluster analysis we have identified the cluster that is 7 and have identified the main variables (top) from the cluster.

The seven cluster identified are as in the image above.

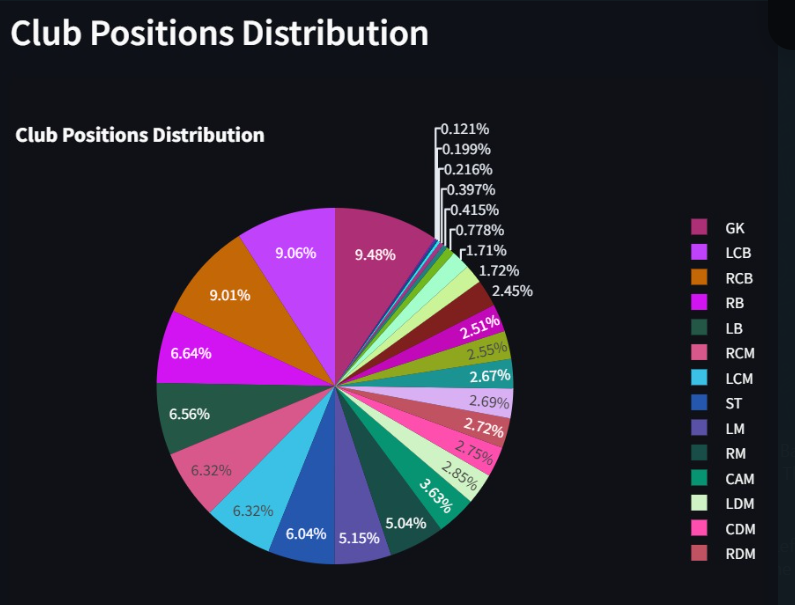
Decision Tree



These are the top 5 important variables

 Confusion Matrix

4.) OBSERVATIONS

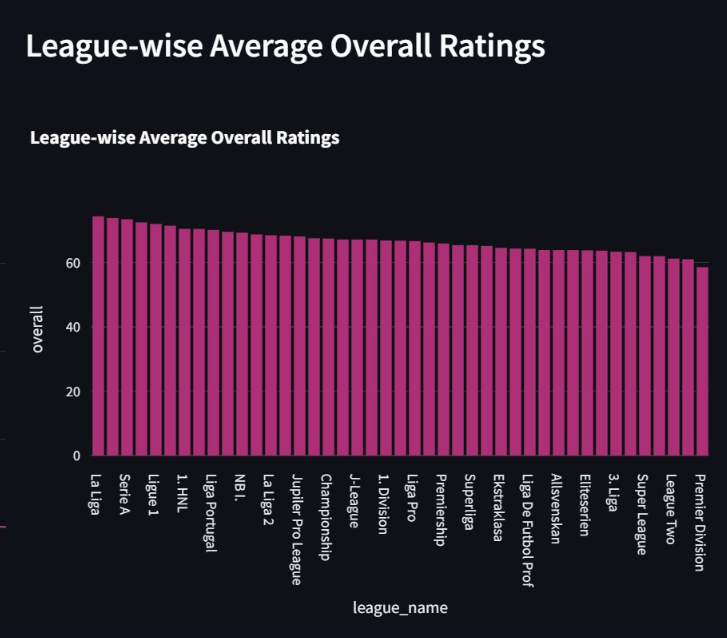
\

Key Positions: The most common positions are Goalkeeper (GK), Left Center Back (LCB), and Right Center Back (RCB), each representing around 9% of players. This suggests a strong emphasis on defensive roles in team compositions.

Midfield Roles: Midfield positions such as Right Center Midfield (RCM) and Left Center Midfield (LCM) each make up around 6-7% of players, highlighting the central role of midfielders in team dynamics.

Forwards: Strikers (ST) and attacking midfielders (CAM) are less common, each constituting around 2.5-3% of the distribution, indicating a more specialized and selective allocation of these roles.

Insight: The distribution emphasizes the critical importance of defensive and midfield positions in team structures, with a balanced but slightly lesser focus on attacking roles. This insight can guide managerial decisions on player recruitment and team strategy.



Top Leagues:

La Liga, Serie A, and Ligue 1 have the highest average player ratings, indicating stronger overall player quality.

Mid-Tier Leagues:

Leagues like Liga Portugal, Eredivisie, and J-League have slightly lower but comparable average ratings.

Lower-Tier Leagues:

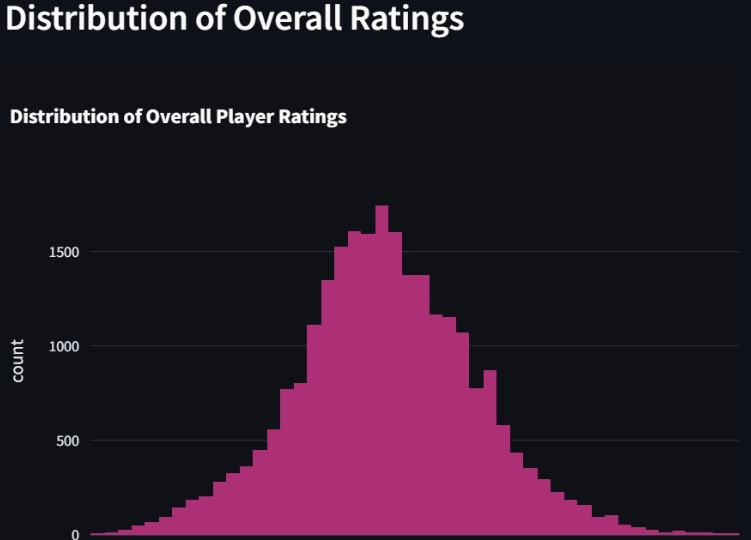
Leagues such as Premier Division, League Two, and Super League have the lowest average ratings, suggesting weaker player quality.

Overall Trend:

There’s a gradual decline in average ratings as we move from top-tier to lower-tier leagues.

Insight:

This graph highlights the varying quality of player talent across different leagues, with top European leagues generally having higher-rated players.



5. MANAGERIAL INSIGHTS

Cluster Segmentation and Targeting:

**Cluster 0**: **Alpha Centauri (Destination), December (Month), Medium Distance to Destination, Highest Price, Medium Customer Satisfaction Score**

This segment represents customers willing to invest in premium space travel experiences to the Alpha Centauri destination, potentially during the month of December.

Marketing campaigns should emphasize the exclusivity and unique experiences offered at Alpha Centauri, aligning with the higher price point preferred by this segment.

Promotional efforts can highlight special events, luxury accommodations, or curated activities tailored to the preferences of this segment during the December travel period.

Cluster 1: Tau Ceti (Destination), January (Month), Lowest Distance to Destination, Lowest Price, Lowest Customer Satisfaction Score

This segment represents price-sensitive customers seeking affordable space travel experiences to nearby destinations like Tau Ceti, potentially in January.

Marketing strategies should focus on value-driven offerings, budget-friendly packages, and competitive pricing to attract this cost-conscious segment.

Emphasize the convenience and accessibility of the Tau Ceti destination, catering to the preference for shorter travel distances within this segment.

Implement targeted initiatives to address and improve customer satisfaction, as this segment exhibits the lowest satisfaction levels.

Cluster 2: Proxima Centauri (Destination), July (Month), Highest Distance to Destination, Medium Price, Highest Customer Satisfaction Score

This segment represents customers seeking immersive space travel experiences to distant destinations like Proxima Centauri, potentially during the month of July.

Marketing campaigns can highlight the adventure and exploration aspects of traveling to this distant destination, appealing to the sense of curiosity and adventure within this segment.

Offer moderately priced packages or flexible pricing options to cater to the medium price preference of this segment.

Leverage the high customer satisfaction levels within this segment by showcasing positive reviews, testimonials, and exceptional service standards.

Destination Management and Experience Curation:

Collaborate with destination management organizations and local tourism authorities to curate tailored experiences and activities at each destination, catering to the preferences of the respective customer segments.

Develop immersive educational programs, guided tours, or interactive exhibits at Alpha Centauri to align with the preferences of the premium segment (Cluster 0).

Enhance the on-site facilities and amenities at Tau Ceti to improve customer satisfaction levels for the price-sensitive segment (Cluster 1).

Invest in infrastructure and resources to support long-haul travel and extended stays at Proxima Centauri, catering to the preferences of the adventure-seeking segment (Cluster 2).

Pricing and Revenue Management:

Implement dynamic pricing strategies and yield management techniques to maximize revenue potential from the premium segment (Cluster 0) while maintaining exclusivity.

Offer bundled packages, discounts, and loyalty programs to attract and retain the price- sensitive segment (Cluster 1).

Explore flexible pricing options or modular offerings for the medium-priced segment (Cluster 2), allowing customers to customize their experiences based on preferences and budget.

Transportation and Logistics:

Optimize transportation routes and fleet allocation based on the varying distance preferences across the clusters, ensuring efficient travel to nearby destinations like Tau Ceti (Cluster 1) and longer-range travel capabilities for destinations like Proxima Centauri (Cluster 2).

Explore partnerships with transportation manufacturers and research institutions to develop advanced spacecraft or propulsion systems tailored to the different distance requirements of each segment.

Product Development and Innovation:

Invest in research and development to enhance existing spacecraft and transportation modes, catering to the diverse distance preferences across all clusters.

Explore the development of new propulsion systems, fuel-efficient engines, or sustainable transportation technologies to support long-haul travel to destinations like Proxima Centauri (Cluster 2) while minimizing environmental impact.

Collaborate with renowned designers, architects, and hospitality experts to develop innovative and immersive accommodation options at each destination, ranging from budget- friendly to luxury, catering to the preferences of all clusters.

Leverage emerging technologies, such as virtual reality and augmented reality, to create unique and engaging pre-travel experiences, enabling customers to explore their destinations virtually before their actual visit.

Strategic Partnerships and Collaborations:

Establish strategic partnerships with transportation manufacturers, aerospace companies, and research institutions to drive innovation in transportation modes and propulsion systems, aligning with the varying distance requirements of each cluster.

Collaborate with destination management organizations, local tourism authorities, and educational institutions to curate authentic and enriching experiences at each destination, tailored to the interests and preferences of different customer segments.

Partner with luxury hospitality brands and activity providers to create exclusive packages and personalized experiences for the premium segment (Cluster 0) visiting Alpha Centauri.

Foster collaborations with sustainable tourism organizations and eco-friendly initiatives to align with the preferences of environmentally conscious segments and promote responsible space travel practices.

Service and Operational Optimization:

Implement efficient check-in/check-out processes, streamlined services, and seamless transportation coordination for clusters with shorter travel distances (Cluster 1) to ensure a convenient and hassle-free experience.

Develop comprehensive training programs for staff and service personnel to deliver exceptional customer experiences tailored to the unique preferences and expectations of each cluster, addressing the low satisfaction levels within Cluster 1.

Leverage advanced data analytics and customer relationship management (CRM) systems to gain insights into customer preferences, enabling personalized recommendations and tailored services across all segments.

Optimize resource allocation, staffing, and on-site operations at each destination based on the varying duration preferences and peak travel periods identified within each cluster.

Continuous Improvement and Customer Experience:

Establish robust feedback mechanisms and customer satisfaction monitoring systems to continuously gather insights and preferences from customers across all clusters.

Implement targeted initiatives to address the low customer satisfaction levels within Cluster 1, such as service quality improvements, enhanced on-site experiences, or personalized attention.

Leverage the high customer satisfaction levels within Cluster 2 by encouraging positive word-of-mouth, testimonials, and referrals to attract new customers to the Proxima Centauri destination.

Foster a culture of continuous improvement and innovation within the organization, encouraging employees to contribute ideas and embrace change to maintain a competitive edge in the dynamic space travel and tourism industry.

By adopting a holistic approach that encompasses targeted marketing, destination management, pricing strategies, transportation logistics, and continuous improvement, space travel and tourism companies can effectively cater to the diverse preferences and characteristics of their customer segments. This comprehensive approach will enable the delivery of exceptional experiences, foster customer loyalty, and drive sustainable growth within the industry.