Submitted to
Prof Ashok Harnal and
Prof. Amarnath Mitra

REAL WORLD ANALYTICS CHALLENGE

FIFA ANALYTIC

BY GROUP 6

Name	Roll No.
Arindam Chakraborty	321128
Adhyatik	311063
Aashit Sharma	321148
Subhajit Paul	321173

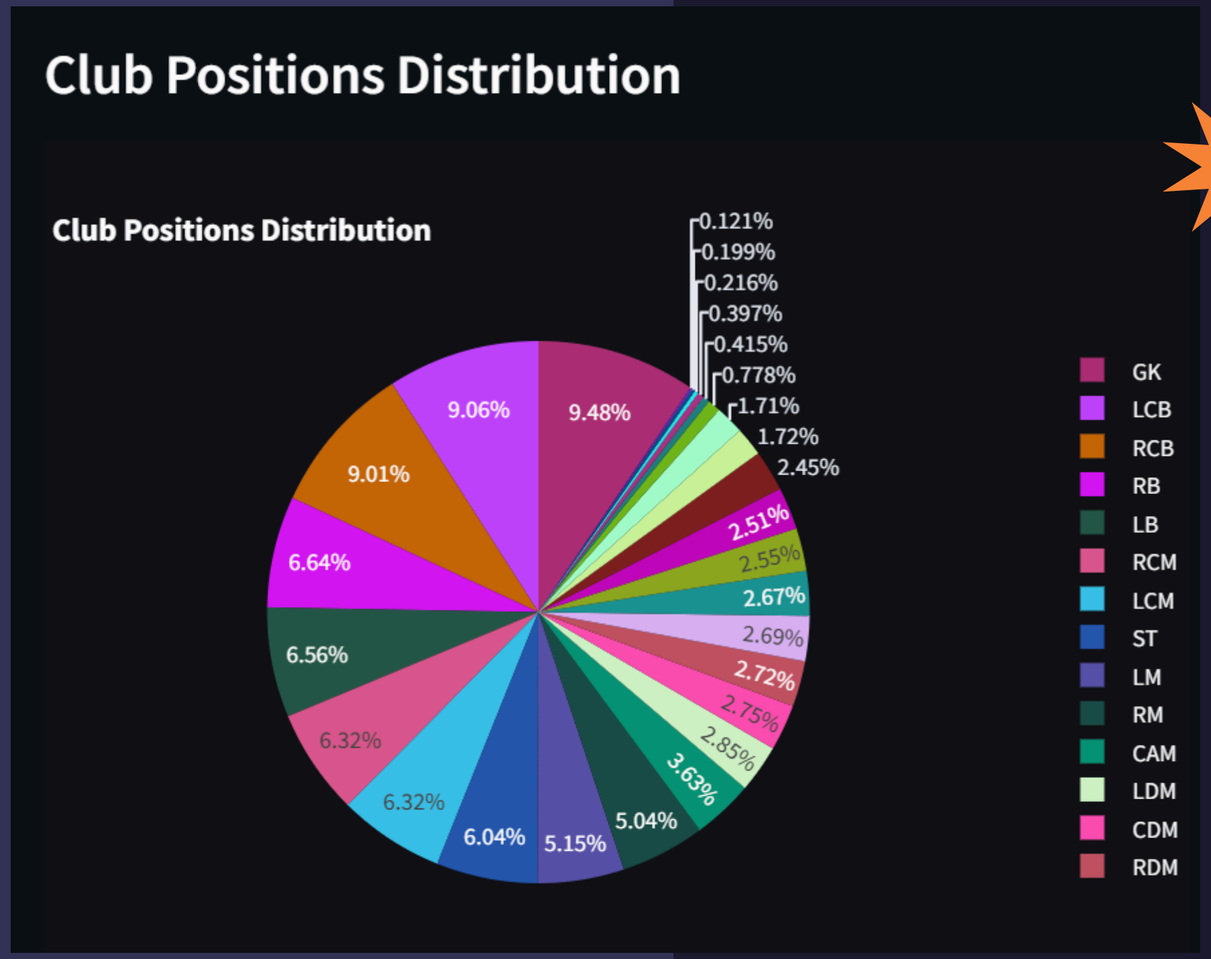
DATASET OVERVIEW

Dataset Description:

- 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.

Key Statistics:

- Observations: 12651 players
- Variables: 110 key attributes – Overall, Height (cm), Age, Club Position, League Name, Player ID
- Number Of Observations Taken For Analysis:- 25000



ANALYTICAL DASHBOARD OBJECTIVES

Purpose of the Dashboard:

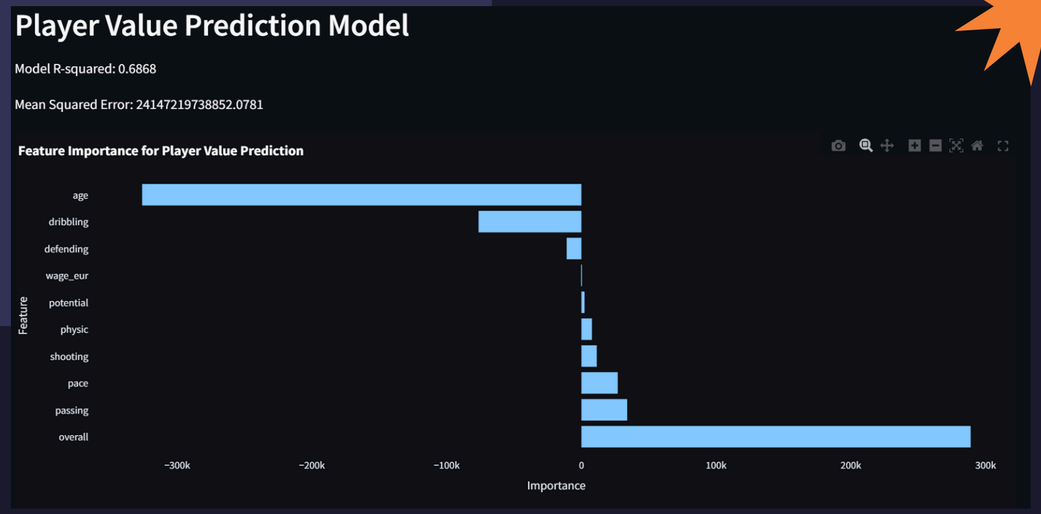
1. Provide a user-friendly interface that offers statistical summaries and insightful visualizations of player data.
2. Highlight relationships between variables through scatter plot matrices and other interactive charts.
3. Analyze Football Player Data: Provide a comprehensive view of player attributes, club positions, and league performances.
4. Visualize Key Patterns: Highlight patterns in overall ratings, club position distributions, and league-wise average ratings.
5. Explore Attribute Correlations: Investigate the relationship between player height and overall rating to identify potential performance trends.
6. Player attribute correction and overall player Analysis
7. Player value prediction

MACHINE LEARNING MODEL OBJECTIVES

MODEL GOAL

- Predict a player's overall rating using attributes like height, age, club position, and league.
- Train the model on a subset of data, ensuring robust evaluation metrics for accuracy.
- In k- means clustering we identified the optimal no. of clusters and defined the characteristics of the clusters
- In decision tree we identified the significant variables from the identified cluster and also found out the accuracy and significance of the model

ANALYSIS OVERVIEW - DASHBOARD



Data Insights:

- Summary Statistics: Quickly assess the distribution of player ratings, height, and age.
- Visual Distributions: Identify trends and outliers across different club positions and leagues.
- Interactive Visuals: Utilize histograms, scatter plots, and pie charts for an immersive analytical experience.

ANALYSIS OVERVIEW - MACHINE LEARNING MODEL

Model Performance:

- **Evaluation Metrics:** Use Mean Squared Error (MSE) and Root Mean Squared Percentage Error (RMSPE) to gauge model accuracy.
- **Insights:** Understand how different player attributes influence overall rating predictions.

cluster	age	league_name	nationality_name	wage_eur	club_name
0	0.334214	20.829992	72.440257	0.204021	555.898156
1	0.565703	20.745566	69.901029	0.255193	557.221590
2	0.823842	21.741784	75.399061	0.236954	591.110329
3	0.240221	20.756079	74.990181	0.106303	557.000000
4	0.701869	21.127620	72.771775	0.251124	570.145785
5	0.446982	23.190882	64.954040	0.247193	543.943106
6	0.840112	25.548387	90.032258	0.226012	595.451613

Cluster 0: Established Stars
Cluster 1: Rising Talents
Cluster 2: Veteran Leaders
Cluster 3: Mid-Career Players
Cluster 4: International Prospects
Cluster 5: Domestic Role Players
Cluster 6: Journeymen

One-way analysis of variance (ANOVA)

Descriptive Statistics

Confidence Interval (CI) Probability: 95.0%

	Group	N	Missing	Missing Group	Mean	Std. Deviation	Std. Error	CI (Lower Bound)	CI (Upper Bound)	Minimum	Maximum
Age	cluster_0	32996	0	0	0.3732	0.2345	0.0013	0.3707	0.3757	0.0	1
Age	cluster_1	32776	0	0	0.3729	0.2337	0.0013	0.3704	0.3754	0.0	1
Age	cluster_2	34420	0	0	0.3729	0.2348	0.0013	0.3704	0.3754	0.0	1
Age	Total	100192	0	0	0.373	0.2343	0.0007	0.3715	0.3744	0.0	1
Distance to Destination (Light-Years)	cluster_0	32996	0	0	0.2999	0.3239	0.0018	0.2964	0.3034	0.0	1
Distance to Destination (Light-Years)	cluster_1	32776	0	0	0.2976	0.3219	0.0018	0.2941	0.3011	0.0006	1
Distance to Destination (Light-Years)	cluster_2	34420	0	0	0.3047	0.3258	0.0018	0.3012	0.3081	0.0006	1
Distance to Destination (Light-Years)	Total	100192	0	0	0.3008	0.3239	0.001	0.2988	0.3028	0.0	1
Duration of Stay (Earth Days)	cluster_0	32996	0	0	0.3222	0.2813	0.0015	0.3192	0.3252	0.0	1
Duration of Stay (Earth Days)	cluster_1	32776	0	0	0.3207	0.2819	0.0016	0.3177	0.3238	0.0	1
Duration of Stay (Earth Days)	cluster_2	34420	0	0	0.3228	0.2813	0.0015	0.3199	0.3258	0.0	1
Duration of Stay (Earth Days)	Total	100192	0	0	0.3219	0.2815	0.0009	0.3202	0.3237	0.0	1
Number of Companions	cluster_0	32996	0	0	0.2219	0.2158	0.0012	0.2195	0.2242	0.0	1
Number of Companions	cluster_1	32776	0	0	0.2206	0.2161	0.0012	0.2182	0.2229	0.0	1
Number of Companions	cluster_2	34420	0	0	0.2207	0.2171	0.0012	0.2184	0.223	0.0	1
Number of Companions	Total	100192	0	0	0.221	0.2163	0.0007	0.2197	0.2224	0.0	1
Price (Galactic Credits)	cluster_0	32996	0	0	0.5632	0.2103	0.0012	0.5609	0.5655	0.1496	1
Price (Galactic Credits)	cluster_1	32776	0	0	0.5038	0.2033	0.0011	0.5016	0.506	0.0	1
Price (Galactic Credits)	cluster_2	34420	0	0	0.5155	0.1983	0.0011	0.5134	0.5176	0.0	1
Price (Galactic Credits)	Total	100192	0	0	0.5274	0.2055	0.0006	0.5261	0.5287	0.0	1

KEY OBSERVATIONS FROM THE DASHBOARD

Data-Driven Insights:

- **Rating Distribution:** Analyze how player ratings vary across leagues and positions.
- **Height & Weight Correlation:** Explore the relationship between player physicality and their performance ratings.
- **Club Position Analysis:** Spot trends in how specific positions affect overall player performance.
- Majority of players have average ratings, with fewer being exceptionally high or low.
- 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.
- Leaguewise graph highlights the varying quality of player talent across different leagues, with top European leagues generally having higher-rated players.
- Majority of players have average ratings, with fewer being exceptionally high or low.
- Dribbling skills boost player value, while passing surprisingly lowers it. Younger players are valued more, as age negatively impacts value. The low R-squared indicates that key factors influencing player value are likely missing from this model.

MANAGERIAL INSIGHTS & RECOMMENDATIONS

Strategic Recommendations:

- **Data Enhancement:** Consider integrating additional variables or refining existing ones for better predictions.
- **Model Optimization:** Adjust hyperparameters or experiment with different algorithms to boost accuracy.
- **Visualization Improvements:** Incorporate more dynamic visual elements for deeper insights.

CONCLUSION & FUTURE DIRECTIONS

Summary:

- This project blends analytical dashboards and machine learning to uncover valuable insights from football player data.
- The findings can inform strategies for player scouting, training, and performance evaluation.

Next Steps:

- Explore additional data sources for a richer analysis.
- Implement advanced machine learning techniques to refine prediction accuracy.
- Expand the dashboard’s interactivity for broader use cases in football management.

THANK YOU!!