Radiant Rovers Football Academy — Player Performance Analysis

Prepared by: Muddassir Mushtaq Mulla
Tools: Python (Pandas, NumPy), PostgreSQL, Power BI
Dataset: 48 Players | 20 Matches | Player-level performance data

---

1. Project Overview

This project focuses on analysing on-field performance data for **Radiant Rovers Football Academy** using player-level match data collected across **20 competitive matches**. The objective of this analysis is to evaluate individual and team performance, identify top performers, understand fitness and technical trends, and support data-driven coaching and player development decisions.

The complete analytics pipeline follows an end-to-end workflow:
CSV → Python (Jupyter Notebook) → PostgreSQL (SQL KPIs) → Power BI Dashboard

---

2. Dataset Summary

The dataset represents match-wise performance records of academy players.

Dataset characteristics:

- Total rows (player–match records): **960**
- Total players: **48**
- Total matches: **20**

Key features include:

Player Information

- Player ID
- Player Name
- Age
- Playing Position

	player_id text	player_name text	age bigint	position text
1	P1000	Anvay	12	MID
2	P1003	Alqama	12	DEF
3	P1009	Amir dalvi	12	DEF
4	P1014	Dhruvang	12	GK
5	P1019	Tameem Umalkar	12	MID

Match Details

- Match ID
- Match Date
- Minutes Played

	match_id text	match_date date	minutes_played bigint
1	M200	2025-01-01	67
2	M200	2025-01-01	77
3	M200	2025-01-01	20
4	M200	2025-01-01	90
5	M200	2025-01-01	68

Performance Metrics

- Goals
- Assists
- Shots on Target
- Tackles
- Passing Accuracy

	goals bigint	assists bigint	shots_on_target bigint	tackles bigint	passing_accuracy double precision
1	0	0	0	2	81.6
2	0	0	0	6	67.2
3	0	0	0	0	72.4
4	0	0	0	0	71.1
5	0	0	0	2	66.1

Fitness Metrics

- Fitness Score
- Stamina
- Speed

	fitness_score double precision	stamina double precision	speed double precision
1	67.3	62.1	69.5
2	69.4	40.2	53.8
3	79.8	60	66.3
4	82.9	52	75
5	71.1	68.2	77.9

Training Indicator

- Attendance

Data Quality Notes:

- Player goal and assist values were complete after cleaning
- Missing performance metrics were handled logically during preprocessing
- Dataset is suitable for player-level and match-level analysis

---

3. Exploratory Data Analysis using Python

Python (Pandas & NumPy) was used inside a Jupyter Notebook to clean, transform, and prepare the dataset for analysis.

```
Loaded dataframe shape: (960, 32)
Columns: ['match_id', 'match_date', 'player_id', 'player_name', 'age', 'position', 'attendance', 'minutes_played', 'fitness_s
core', 'stamina', 'speed', 'passing_accuracy', 'tackles', 'goals', 'assists', 'shots_on_target', 'saves', 'yellow_card', 'red
_card', 'age_group', 'matches_played', 'total_minutes', 'avg_fitness', 'avg_stamina', 'avg_speed', 'avg_passing', 'total_goal
s', 'total_assists', 'total_saves', 'total_tackles', 'yellow_cards', 'red_cards']
```

	match_id	match_date	player_id	player_name	age	position	attendance	minutes_played	fitness_score	stamina	...	avg_fitness
0	M200	01-01-2025	P1000	Anvay	12	MID	1	67.0	67.3	62.1	...	66.64
1	M200	01-01-2025	P1003	Alqama	12	DEF	1	77.0	69.4	40.2	...	68.91
2	M200	01-01-2025	P1009	Amir dalvi	12	DEF	1	20.0	79.8	60.0	...	70.77
3	M200	01-01-2025	P1014	Dhruvang	12	GK	1	90.0	82.9	52.0	...	72.42
4	M200	01-01-2025	P1019	Tameem Umalkar	12	MID	1	68.0	71.1	68.2	...	70.34
5	M200	01-01-2025	P1037	Arham Tambe	12	DEF	1	89.0	71.1	52.7	...	70.58

avg_fitness	avg_stamina	avg_speed	avg_passing	total_goals	total_assists	total_saves	total_tackles	yellow_cards	red_cards
66.64	59.56	65.36	70.34	2	0	0	22	3	0
68.91	62.12	66.78	58.06	0	0	0	40	1	0
70.77	63.05	68.60	66.79	0	0	0	20	3	0
72.42	64.92	68.35	70.82	0	0	39	0	0	0
70.34	58.92	68.76	69.49	0	1	0	13	1	0

3.1 Data Loading and Initial Inspection

- Loaded CSV data using - pandas.read_csv()
- Reviewed structure using - df.info()
- Generated descriptive statistics using - df.describe()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 960 entries, 0 to 959
Data columns (total 32 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   match_id          960 non-null    string
 1   match_date        960 non-null    object
 2   player_id         960 non-null    string
 3   player_name       960 non-null    string
 4   age               960 non-null    int64
 5   position          960 non-null    category
 6   attendance        960 non-null    int64
 7   minutes_played    960 non-null    int64
 8   fitness_score     960 non-null    float64
 9   stamina           960 non-null    float64
 10  speed             960 non-null    float64
 11  passing_accuracy  960 non-null    float64
 12  tackles           960 non-null    int64
 13  goals             960 non-null    int64
 14  assists           960 non-null    int64
 15  shots_on_target   960 non-null    int64
 16  saves             960 non-null    int64
 17  yellow_card       960 non-null    int64
 18  red_card          960 non-null    int64
 19  age_group         960 non-null    category
 20  matches_played    960 non-null    int64
 21  total_minutes     960 non-null    int64
 22  avg_fitness       960 non-null    float64
 23  avg_stamina       960 non-null    float64
 24  avg_speed         960 non-null    float64
 25  avg_passing       960 non-null    float64
 26  total_goals       960 non-null    int64
 27  total_assists     960 non-null    int64
 28  total_saves       960 non-null    int64
 29  total_tackles     960 non-null    int64
 30  yellow_cards      960 non-null    int64
 31  red_cards         960 non-null    int64
dtypes: category(2), float64(8), int64(18), object(1), string(3)
memory usage: 227.4+ KB
```

3.2 Handling Missing Values

- Converted numeric columns such as goals, assists, minutes played, fitness scores, and attendance into appropriate data types
- Replaced missing goal- and assist-related values with **0** (logical for match performance)
- Retained NaN values where metrics were not applicable (e.g., passing accuracy when no passes were attempted)

```
match_id            0
match_date          0
player_id           0
player_name         0
age                 0
position            0
attendance          0
minutes_played      0
fitness_score       0
stamina             0
speed               0
passing_accuracy    0
tackles             0
goals               0
assists             0
shots_on_target     0
saves               0
yellow_card         0
red_card            0
age_group           0
matches_played      0          total_goals         0
total_minutes       0          total_assists       0
avg_fitness         0          total_saves         0
avg_stamina         0          total_tackles       0
avg_speed           0          yellow_cards        0
avg_passing         0          red_cards           0
```

Outlier Check (IQR Method)

```
Total Outlier Rows Detected: 505
```

	match_id	match_date	player_id	player_name	age	position	attendance	minutes_played	fitness_score	stamina	...	avg_fitness
1	M200	2025-01-01	P1003	Alqama	12	DEF	1	77	69.4	40.2	...	68.91
3	M200	2025-01-01	P1014	Dhruvang	12	GK	1	90	82.9	52.0	...	72.42
6	M200	2025-01-01	P1038	Abdul Raafe	12	FWD	1	89	64.1	51.6	...	72.14
12	M201	2025-01-08	P1014	Dhruvang	12	GK	1	90	64.0	63.9	...	72.42
15	M201	2025-01-08	P1038	Abdul Raafe	12	FWD	1	71	74.1	83.3	...	72.14
18	M202	2025-01-15	P1000	Anvay	12	MID	1	73	77.7	57.8	...	66.64
20	M202	2025-01-15	P1009	Amir dalvi	12	DEF	1	90	78.9	69.9	...	70.77
21	M202	2025-01-15	P1014	Dhruvang	12	GK	1	90	75.3	66.0	...	72.42
22	M202	2025-01-15	P1019	Tameem Umalkar	12	MID	1	90	76.5	48.9	...	70.34
24	M202	2025-01-15	P1038	Abdul Raafe	12	FWD	1	66	69.5	57.3	...	72.14

3.3 Data Standardization

- Standardized column names for consistency
- Ensured categorical values such as position were consistently formatted
- Parsed match dates into datetime format

3.4 Feature Engineering

To improve analytical depth, the following derived metrics were created:

- **Goals per 90 minutes** – to normalize scoring across different playing times

	player_id text	player_name text	goals_per_90 numeric
1	P1021	Rudransh	0.28
2	P1030	Zaid Shaikh	0.27
3	P1042	Mohmd. Amir	0.24
4	P1044	Saif Solkar	0.21
5	P1038	Abdul Raafe	0.21
6	P1029	Maviya	0.20
7	P1018	Mahir Khanche	0.16
8	P1034	Touhid Qazi	0.15

- **Aggregated player performance metrics**

These features enabled fair comparison between players with varying match minutes.

	position text	avg_fitness numeric	avg_stamina numeric	avg_speed numeric	players bigint
1	DEF	70.07	59.85	63.18	21
2	MID	69.02	60.18	63.88	16
3	FWD	70.79	60.71	65.57	7
4	GK	70.26	60.45	62.92	4

3.5 Database Loading

The cleaned dataset was exported to **PostgreSQL** using SQLAlchemy. A structured table ("FA_P&E_TAB")was created to support KPI calculation and Power BI connectivity.

---

4. Performance Analysis using SQL (PostgreSQL)

SQL queries were used to compute key performance indicators and answer important analytical questions.

Key KPIs and Insights derived:

- **Total matches played: 20**

	total_matches bigint
1	20

- **Total players analyzed: 48**

total_unique_players bigint
48

- **Average goals per match (player-sum): 1.6**

	avg_goals_per_match numeric
1	1.60

- **Identification of top goal scorers**

	player_id text	player_name text	total_goals numeric
1	P1021	Rudransh	4
2	P1030	Zaid Shaikh	4
3	P1038	Abdul Raafe	3
4	P1042	Mohmd. Amir	3
5	P1044	Saif Solkar	3

- **Goals-per-90 efficiency rankings**

	player_id text	player_name text	goals_per_90 numeric
1	P1021	Rudransh	0.28
2	P1030	Zaid Shaikh	0.27
3	P1042	Mohmd. Amir	0.24
4	P1044	Saif Solkar	0.21
5	P1038	Abdul Raafe	0.21

- **Average passing accuracy by playing position**

	position text	avg_passing_accuracy numeric
1	MID	70.31
2	GK	67.98
3	FWD	64.67
4	DEF	64.53

- **Fitness score trends by age**

	age bigint	avg_fitness_score numeric
1	12	70.65
2	13	69.80
3	14	69.40
4	15	69.85
5	16	69.40
6	17	68.84
7	18	70.62

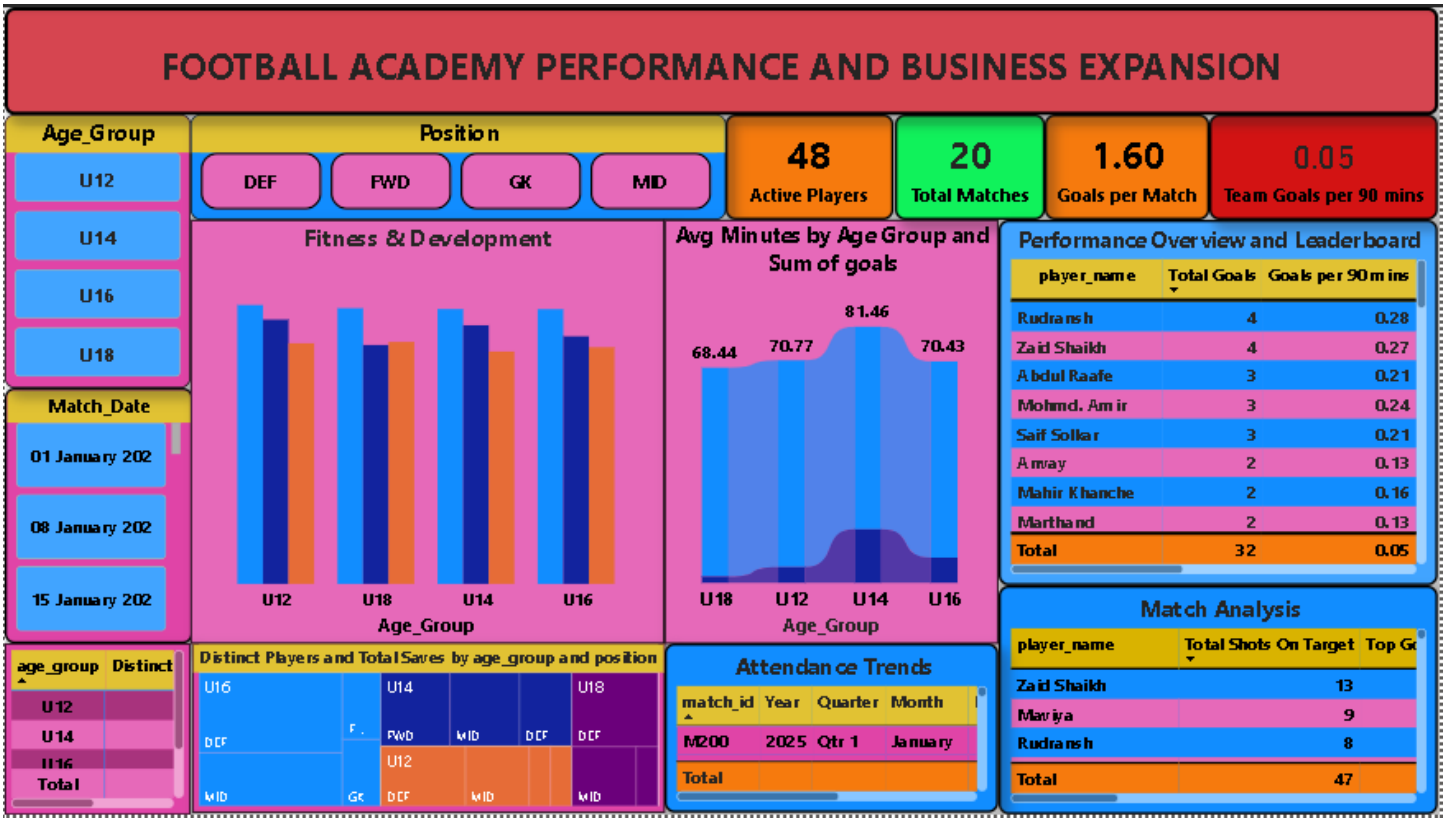
- **Attendance-based consistency indicators**

	player_id text	player_name text	avg_attendance numeric
1	P1008	Agastya Salvi	1.00
2	P1045	Umar	1.00
3	P1030	Zaid Shaikh	1.00
4	P1016	Shlok	1.00
5	P1037	Arham Tambe	1.00

---

5. Power BI Dashboard

An interactive **Power BI dashboard** was developed using the cleaned dataset and SQL-derived KPIs.



Dashboard pages include:

- **Performance Overview** – Total goals, Assists, minutes played, and goals per match
- **Player Leaderboard** – Top scorers, goals per 90, and players conversion percentage
- **Fitness & Development** – Fitness score trend by age
- **Match Analysis** – Shots On Target, Match day Attendance and match contribution metrics
- **Attendance Trends** – Training Consistency Indicators

Slicers were implemented for position, age, and player selection to enable drill-down analysis.

---

6. Coaching & Player Development Recommendations

1. **Improve Finishing Efficiency**
   Players with high shots but lower goals should undergo targeted finishing drills.
2. **Optimize Playing Time**
   Players with high goals-per-90 efficiency should be considered for increased match minutes.
3. **Position-Specific Training**
   Passing accuracy analysis suggests the need for tailored passing drills, especially for defenders.
4. **Fitness Conditioning Programs**
   Lower fitness-score age groups should receive focused conditioning sessions.
5. **Attendance Monitoring**
   Use attendance trends as an early indicator of player engagement and consistency.

---

7. Conclusion

The project demonstrates the practical application of sports analytics in a real academy environment using real match data collected and analyzed by the coach. By integrating Python-based data preparation, SQL-driven KPI analysis, and Power BI visualization, the analysis provides actionable insights for coaching decisions, player development, and performance tracking.

End of Report