FIFA World Cup 2022 Data Analysis Dashboard

Problem Statement

The FIFA World Cup 2022 was a highly competitive tournament with diverse team performances. Understanding key performance metrics such as possession, goal-scoring efficiency, defensive pressures, and match outcomes can provide valuable insights into team strategies and success factors. However, raw match data is complex and requires structured analysis to extract meaningful trends.

Objective

Develop an interactive Power BI dashboard with Python-assisted Exploratory Data Analysis (EDA) to uncover patterns in team performance, identify key success factors, and visualize match statistics for better decision-making in football analytics.

Scope

- Data Exploration & Cleaning (Python EDA)
- Handle missing values (if any).
- Convert date/time columns to appropriate formats.
- Derive new metrics (e.g., possession percentage, shot accuracy, pressure success rate).

Key Analysis Areas

- **Team Performance: Compare goals, assists, shots, and defensive actions.**
- **Possession Impact: Does higher possession correlate with winning?**
- **Goal Efficiency: Analyze goals inside vs. outside the penalty area.**
- **Defensive Metrics: Compare tackles, interceptions, and pressures.**
- **Match Trends: Identify patterns based on match category (group stage, knockout, final).**

Power BI Dashboard

- ☑ Interactive filters (team, match stage, date).

Visualizations

- Bar charts (top teams by goals, assists).
- Scatter plots (possession vs. goals scored).
- Heatmaps (defensive pressures vs. tackles).
- Line charts (trends across match stages).

How to Analyze the Dataset

1. Python EDA Steps

```
# Load data
print(df.info())
print(df.describe())

# Convert date & time
df['date'] = pd.to_datetime(df['date'])
df['hour'] = pd.to_datetime(df['hour']).dt.hour

# Possession to numeric
df['possession_team1'] = df['possession_team1'].str.rstrip('%').astype(float)
df['possession_team2'] = df['possession_team2'].str.rstrip('%').astype(float)

# New metrics
df['total_goals'] = df['number of goals team1'] + df['number of goals team2']
df['shot_accuracy_team1'] = df['number of goals team1'] / df['total attempts
team1']
df['shot_accuracy_team2'] = df['number of goals team2'] / df['total attempts
team2']

# Visualizations
# Goals distribution
sns.histplot(df['total_goals'], bins=10)
plt.title("Distribution of Total Goals per Match")

# Possession vs. Goals
sns.scatterplot(x='possession_team1', y='number of goals team1', data=df)
plt.title("Possession vs. Goals Scored (Team 1)")

# Defensive pressures comparison
sns.boxplot(x='category', y='defensive pressures applied team1', data=df)
plt.title("Defensive Pressures by Match Stage")
```

2. Power BI Dashboard Development

Data Import: Load cleaned CSV into Power BI.

Key Visualizations:

Team Performance:

Table: Top 5 teams by goals scored.

Donut chart: Goals inside vs. outside penalty area.

Possession Analysis:

Line chart: Avg. possession % by match stage.

Defensive Metrics:

Stacked bar: Tackles vs. interceptions by team.

Match Insights:

Slicer: Filter by team, date, or category.

3. Insights to Extract

- Which teams had the highest shot accuracy?
- Did teams with higher possession win more often?
- How did defensive strategies vary between group stages and knockouts?
- Were there any outliers (e.g., high possession but low goals)?

Expected Outcome

An interactive Power BI dashboard with Python-driven EDA to help football analysts, coaches, and fans understand key performance trends from the FIFA World Cup 2022.
