**FIFA World Cup 2022 Data Analysis & Dashboarding Project**

**1. Problem Statement**

**Objective:**
Analyze the FIFA World Cup 2022 dataset to uncover insights about team performance, player statistics, match outcomes, and tournament trends. Build an interactive dashboard to visualize key findings for better decision-making in football analytics.

**Key Questions to Explore:**

- Which teams performed the best (goals, wins, possession, etc.)?
- Who were the top players (goals, assists, passes, etc.)?
- How did home vs. away performance differ?
- What were the most common match outcomes (wins, draws, penalties)?
- Were there any surprising trends (underperforming teams, unexpected stats)?

---

**2. Scope of Analysis**

**Data Exploration (EDA)**

1. **Team Performance Analysis**
    - Win/loss ratios, goals scored/conceded, possession stats.
    - Comparison of group stage vs. knockout stage performance.
    - Fair play rankings (yellow/red cards).
2. **Player Performance Analysis**
    - Top scorers, assist leaders, pass accuracy.
    - Best goalkeepers (saves, clean sheets).
    - Most valuable players (MVPs) based on performance metrics.
3. **Match & Tournament Trends**
    - Most common match results (e.g., 1-0, 2-1).
    - Penalty shootout analysis (success rates).
    - Impact of substitutions on match outcomes.

4. **Advanced Metrics (Optional)**
    - Expected Goals (xG) vs. actual goals.
    - Player heatmaps (positions, passes, shots).
    - Time-based analysis (when most goals were scored).

**Dashboard Development**

- **Tools:**
    - Python (Plotly, Matplotlib, Seaborn) + Dash / Streamlit
    - Power BI / Tableau (for non-coders)
- **Key Visualizations:**
    - Interactive tables (top teams, players).
    - Bar/pie charts (goals, cards, possession).
    - Heatmaps (player positions, pass networks).
    - Geo-maps (team performance by country).

**3. How to Analyze the Dataset**

**Step 1: Data Loading & Cleaning**

```
import pandas as pd


# Load datasets (matches, players, teams)
matches = pd.read_csv("fifa_world_cup_2022_matches.csv")
players = pd.read_csv("fifa_world_cup_2022_players.csv")
teams = pd.read_csv("fifa_world_cup_2022_teams.csv")


# Check for missing values
print(matches.isnull().sum())
print(players.isnull().sum())
```

```
# Handle missing data (fill/drop)

matches.fillna(0, inplace=True)

players.dropna(subset=['goals'], inplace=True)
```

**Step 2: Exploratory Data Analysis (EDA)**

**Team Performance**

```
# Top 5 teams by goals scored

top_teams = matches.groupby('team')['goals'].sum().sort_values(ascending=False).head(5)


# Win/loss/draw ratios

matches['result'].value_counts().plot(kind='bar')  # Win, Loss, Draw
```

**Player Performance**

```
# Top 10 goal scorers

top_scorers =
players.groupby('player_name')['goals'].sum().sort_values(ascending=False).head(10)


# Best passers (accuracy > 90%)

best_passers = players[players['pass_accuracy'] > 90].sort_values('passes',
ascending=False)
```

**Match Trends**

```
# Most common scorelines

matches['scoreline'] = matches['home_score'].astype(str) + "-" +
matches['away_score'].astype(str)

common_scores = matches['scoreline'].value_counts().head(5)


# Penalty success rate

penalty_success = matches['penalties_scored'] / matches['penalties_attempted'] * 100
```

**Step 3: Dashboard Development**

**Using Python (Dash/Streamlit)**

```
import dash

import dash_core_components as dcc

import dash_html_components as html

import plotly.express as px


app = dash.Dash(__name__)


app.layout = html.Div([

  html.H1("FIFA World Cup 2022 Dashboard"),

  dcc.Graph(figure=px.bar(top_teams, title="Top Teams by Goals")),

  dcc.Graph(figure=px.pie(common_scores, names=common_scores.index, title="Most Common Scorelines"))

])


app.run_server(debug=True)
```

**Using Power BI / Tableau**

- Import cleaned data.
- Create visualizations (bar charts, pie charts, maps).
- Add filters (team, player, match stage).

---

**4. Expected Outcomes**

✅ **Insights:**

- Identify best-performing teams & players.
- Detect patterns in match outcomes (e.g., late goals).
- Compare expected vs. actual performance.

📊 **Dashboard Features:**

- Interactive filters (team, player, match stage).
- Dynamic charts (goals, passes, cards).
- Exportable reports for further analysis.

---

**5. Next Steps**

1. **Refine Analysis:**
    - Add advanced stats (xG, player ratings).
    - Compare with past World Cup data.
2. **Enhance Dashboard:**
    - Add geo-visualizations (country performance).
    - Implement machine learning (win prediction).