Usman Institute of Technology

V1 BS CS

BIG DATA ANALYTICS ,CSC-316

#### 

#### 

Topic: Analysis of Cricket Match Outcomes

#### 

#### 

#### 

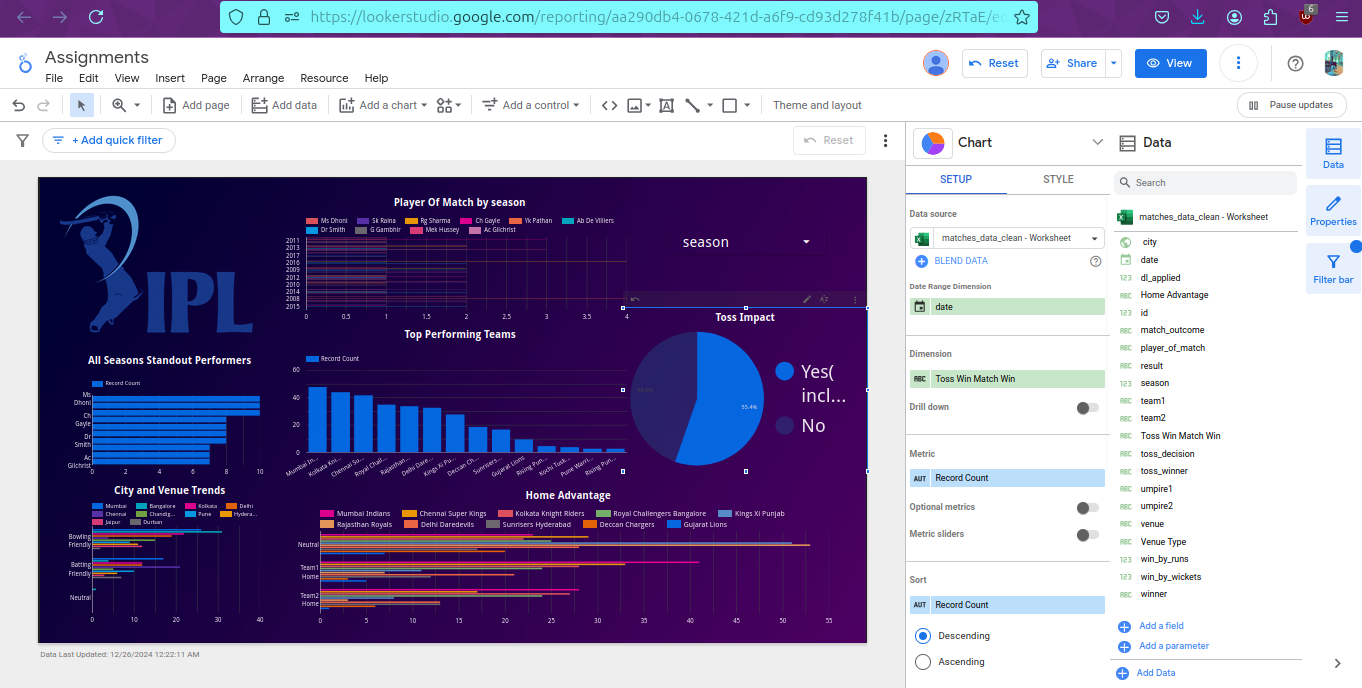
#### **Module\_3**

#### 

#### 

#### **Group Details**

* + Member 1: Syed Saad Akhtar 22sp-029-cs
  + Member 2: Kinza Fatima 22sp-032-cs
  + Member 3: Syeda Akfa Feroz 22sp-046-cs



**Report link:** <https://lookerstudio.google.com/s/nWqVw-Znh64>

### **1. Match Outcome Analysis**

* **Toss Impact**:
  + Target Variable: winner (whether the toss-winning team wins the match).
  + Features to analyze: toss\_winner, toss\_decision, and match\_outcome.

### **2. Team Performance**

* **Top Performing Teams**:
  + Target Variable: winner (to identify the team with the most wins).
  + Features to analyze: season, team1, and team2.

### **3. Player Performance**

* **Seasonal Standouts**:
  + Target Variable: player\_of\_match (to find standout players in specific seasons).
  + Features to analyze: season and player\_of\_match.

### **4. Venue and Location Insights**

* **City and Venue Trends**:
  + Target Variable: match\_outcome (to analyze if a venue favors batting or bowling).
  + Features to analyze: venue, city, win\_by\_runs, and win\_by\_wickets.
* **Home Advantage**:
  + Target Variable: winner (to check if home teams perform better).
  + Features to analyze: city, team1, team2, and toss\_winner.

Code:

import pandas as pd

data = pd.read\_csv('matches\_data\_clean.csv')

# function to calculate "home advantage"

def determine\_home\_advantage(row):

city = row['city'].strip().lower()

team1 = row['team1'].strip().lower()

team2 = row['team2'].strip().lower()

if city in team1:

return "Team1 Home"

elif city in team2:

return "Team2 Home"

else:

return "Neutral"

# function to create new "home advantage" field

data['home\_advantage'] = data.apply(determine\_home\_advantage, axis=1)

# Save updated dataset to a new CSV file

updated\_file\_path = 'matches\_data\_updated.csv' data.to\_csv(updated\_file\_path, index=False)

print(f"Updated dataset saved as {updated\_file\_path}")

**Here’s a brief description of the changes made to the dataset:**

New Field Added: Introduced a new column named home\_advantage.

Logic for the Field:

If the city name matches any word in team1, the value is "Team1 Home".

If the city name matches any word in team2, the value is "Team2 Home".

Otherwise, the value is "Neutral".

Transformation: The comparison was case-insensitive, and whitespace was trimmed for accurate matching.

Saved Output: The updated dataset with the home\_advantage field was saved as a new CSV file named matches\_data\_updated.csv.