IPL Data Analysis

Exploring IPL seasons 1-9 through data analysis.

TEAM MEMBERS

1. MOHAMMAD ASIF
2. GAYATARI TUMSARE
3. SASWAT RANJAN NAYAK
4. ANBUSELVAN

Overview of the IPL Dataset

1 **Matches**

The dataset contains ball-by-ball data for each match played in the IPL, covering all seasons from 1 to 9.

2 **Player Statistics**

Detailed statistics are available for every player, including batting and bowling performance, runs scored, wickets taken, and more.

3 **Team Data**

Information on team performance, wins, losses, and player rosters is readily accessible.

Exploratory Data Analysis (EDA)

Player Analysis

Analyze individual player performance metrics, including runs scored, wickets taken, and strike rate.

1. Identify top performers across batting, bowling, and all-rounders.
2. Analyze player consistency and form over different seasons.

Team Trends

Explore team performance statistics, such as win-loss records, average scores, and dominant player roles.

1. Analyze how team composition impacts overall performance.
2. Identify any emerging team trends or strategies.

Match Analysis

Examine factors influencing match outcomes, such as venue, weather, and player performance.

1. Identify any correlation between player performance and match results.
2. Analyze the impact of key moments in the match.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import zipfile
import os

# Path to the uploaded ZIP file
zip_file_path = '/Users/saswatranjannayak/Desktop/ASE204 Project/IPL_2016/ipl.zip'

# Extract the ZIP file
extracted_dir = '/Users/saswatranjannayak/Desktop/ASE204 Project/IPL_2016/ipl'
with zipfile.ZipFile(zip_file_path, 'r') as zip_ref:
    zip_ref.extractall(extracted_dir)

# List the extracted files
extracted_files = os.listdir(extracted_dir)
extracted_files
```

```
['Ball_by_Ball.csv',
 '__MACOSX',
 'City.csv',
 'Batsman_Scored.csv',
 'Batting_Style.csv',
 'Bowling_Style.csv']
```

```
# Removing the problematic feature engineering part for 'Batsman_Scored'
# Proceeding with the rest of the EDA and data processing

# Load the CSV files into pandas DataFrames
dataframes = {}
for file in extracted_files:
    if file.endswith('.csv'):
        df_name = os.path.splitext(file)[0]
        file_path = os.path.join(extracted_dir, file)
        dataframes[df_name] = pd.read_csv(file_path)
```

```
# Display basic information and first few rows for each DataFrame
for name, df in dataframes.items():
    print(f"\n--- {name} ---")
    print(df.info())
```

SQL in Data Analysis

Data Management

SQL is the backbone for managing relational databases.

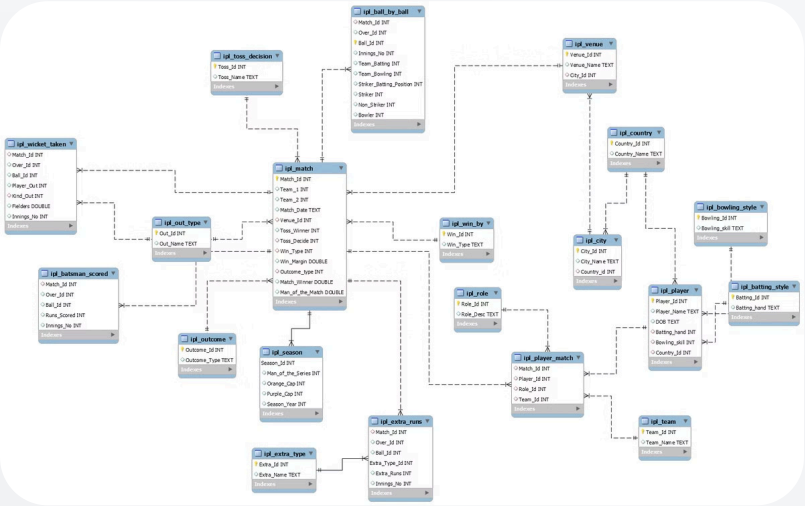
1. SELECT
2. JOIN
3. UPDATE

Data Analysis

Filtering, sorting, and grouping data for insights.

1. INSERT
2. DELETE
3. GROUP BY

Data Overview & Preparation Using SQL



1

Data Acquisition

Gathering ball-by-ball data from IPL seasons 1-9.

2

Data Cleaning

Addressing missing values, duplicates, and inconsistencies.

3

Make Relationship

We also use SQL to create relationships between different tables in the dataset. and also SQL help us structure the data and perform complex queries to analyze the data effectively.

Team Balancing

1

Team 1

Balanced team with a mix of top batsmen, bowlers, and a wicketkeeper.

2

Team 2

Similar strategy, focusing on different players for optimal performance.

3

Team 3

Utilizing data to build a balanced team with a winning edge.

4

Team 4

Strategic selection based on player performance and consistency.

5

Team 5

Completing the lineup with a team that stands out in the league.

Q&A

Open for questions and further discussion.