Project – 2: IPL / Sports Data Analysis

BY: SUHEAL AHMAD

Objective:

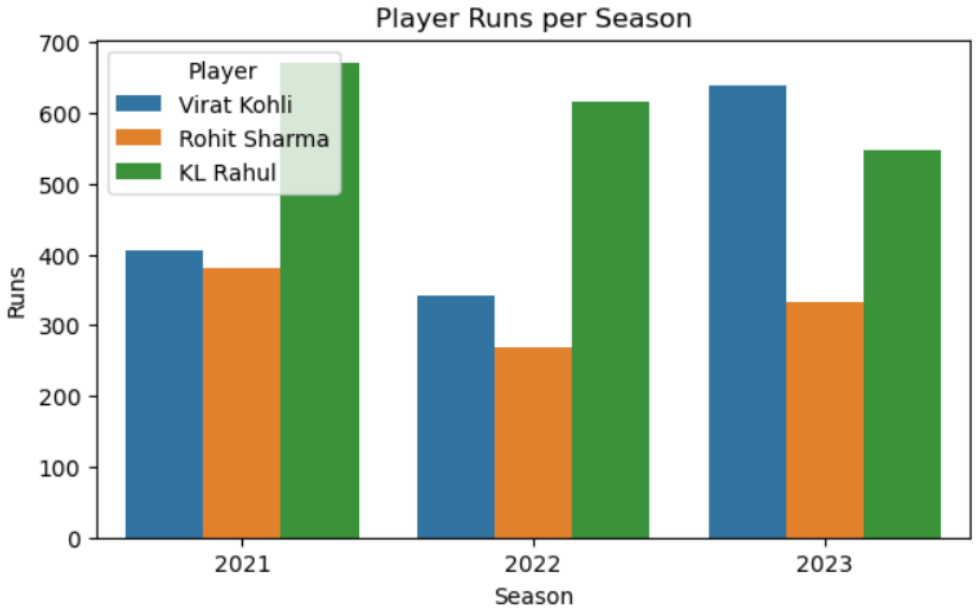
To analyse IPL match and player performance data to identify top scorers, strike rates, and team win rates, compare player performance across seasons, visualize key metrics, and summarize insights for decision-making.

CODES:

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

# --------------------------------
# 1. Create sample IPL dataset
# --------------------------------
data = {
    "Season": [2021,2021,2021,2022,2022,2022,2023,2023,2023],
    "Player": ["Virat Kohli","Rohit Sharma","KL Rahul",
               "Virat Kohli","Rohit Sharma","KL Rahul",
               "Virat Kohli","Rohit Sharma","KL Rahul"],
    "Runs": [405,381,670,341,268,616,639,332,548],
    "Balls": [350,320,520,310,250,480,510,290,430],
    "Team": ["RCB","MI","LSG","RCB","MI","LSG","RCB","MI","LSG"],
    "Matches": [14,14,14,13,13,13,14,14,14],
    "Wins": [8,7,9,7,6,8,9,8,7]
}

df = pd.DataFrame(data)

# --------------------------------
# 2. Calculate strike rate
# --------------------------------
df["Strike_Rate"] = (df["Runs"] / df["Balls"]) * 100

# --------------------------------
# 3. Top scorers per season
# --------------------------------
top_scorers = df.sort_values("Runs", ascending=False)

# --------------------------------
# 4. Team win rate
# --------------------------------
df["Win_Rate"] = (df["Wins"] / df["Matches"]) * 100

team_winrate = df.groupby("Team")["Win_Rate"].mean()
```

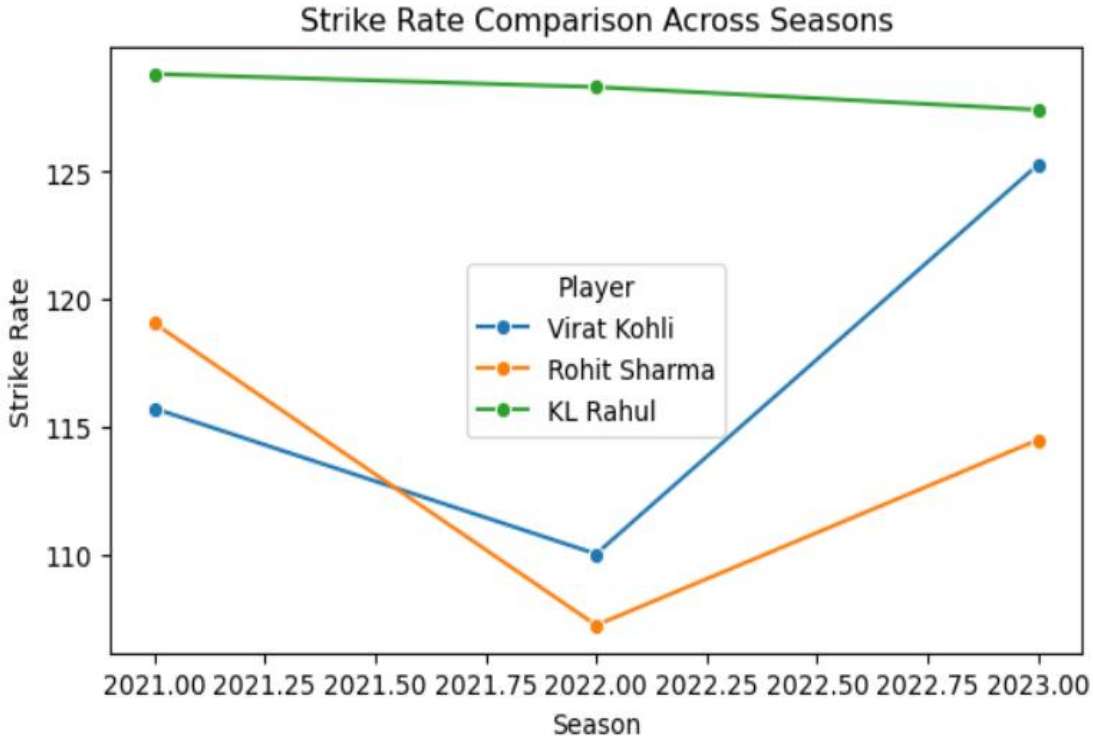
```python
# 5. Visualization: Runs per Season
# --------------------------------
plt.figure(figsize=(7,4))
sns.barplot(data=df, x="Season", y="Runs", hue="Player")
plt.title("Player Runs per Season")
plt.ylabel("Runs")
plt.show()

# --------------------------------
# 6. Visualization: Strike Rate Comparison
# --------------------------------
plt.figure(figsize=(7,4))
sns.lineplot(data=df, x="Season", y="Strike_Rate", hue="Player", marker="o")
plt.title("Strike Rate Comparison Across Seasons")
plt.ylabel("Strike Rate")
plt.show()

# --------------------------------
# 7. Visualization: Team Win Rate
# --------------------------------
plt.figure(figsize=(6,4))
team_winrate.plot(kind="bar")
plt.title("Average Team Win Rate")
plt.ylabel("Win Rate (%)")
plt.show()

# --------------------------------
# 8. Export summary
# --------------------------------
df.to_csv("ipl_player_analysis.csv", index=False)
print("Analysis exported successfully!")
```

Results:



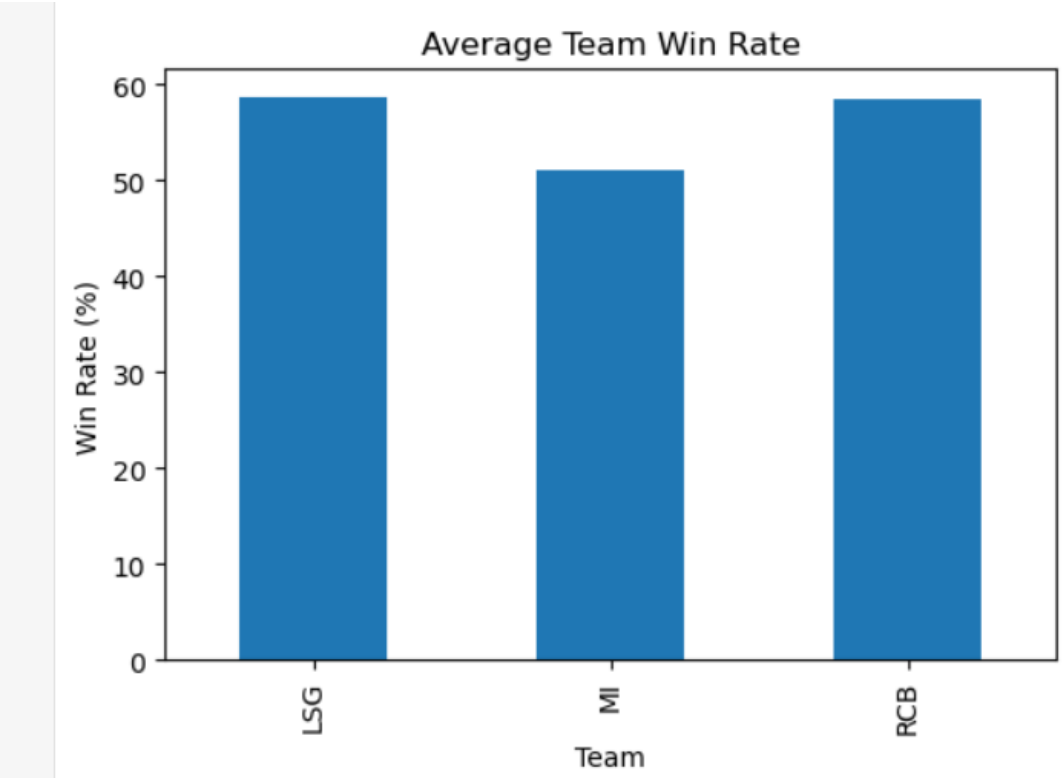
Interpretation:

The chart shows that KL Rahul consistently scored the highest runs in 2021 and 2022, demonstrating strong batting performance and consistency. Virat Kohli experienced a dip in 2022 but made a significant comeback in 2023 with the highest run total. Rohit Sharma maintained relatively stable but lower run totals across all seasons. Overall, the chart highlights variations in player form across seasons and emphasizes the importance of consistency in performance evaluation.



Interpretation:

KL Rahul maintained the highest and most consistent strike rate across all seasons, indicating reliable aggressive batting. Virat Kohli experienced a decline in 2022 but showed a strong improvement in 2023, reflecting a return to form. Rohit Sharma's strike rate dropped in 2022 and partially recovered in 2023. Overall, the plot highlights variations in batting aggression and consistency across seasons.




```
Analysis exported successfully!
```

Interpretation:

The chart shows that RCB and LSG have the highest average win rates, indicating stronger overall team performance during the analysed seasons. MI records a comparatively lower win rate, suggesting less consistency in match outcomes. This comparison highlights how team-level performance varies across franchises and complements individual player analysis.

Conclusion:

This IPL sports data analysis project demonstrates how structured data can be transformed into meaningful performance insights using Python. By analysing runs, strike rates, and team win percentages across multiple seasons, the project highlights individual consistency, seasonal trends, and team-level success. Visual comparisons across seasons reveal changes in player form and playing style, while win-rate analysis connects individual performance with overall team outcomes.

Overall, the project showcases the importance of data-driven evaluation in sports analytics and illustrates how statistical analysis and visualization can support performance assessment, strategy formulation, and fan engagement in professional sports.