

IPL 2024 Data Analysis Report

1. Introduction

The IPL 2024 Data Analysis project aims to extract insights from match data, focusing on team performances, player statistics, and predictive analytics for IPL 2025. This report covers data collection, cleaning, exploratory data analysis (EDA), visualization, and advanced machine learning techniques.

2. Data Collection

- **Source:** Kaggle / ESPN Cricinfo
- **Dataset Attributes:**
 - id - Match ID
 - date - Match date
 - team1, team2 - Competing teams
 - toss_winner - Toss-winning team
 - decision - Bat/Field decision after toss
 - first_score, first_wkts - First innings score & wickets lost
 - second_score, second_wkts - Second innings score & wickets lost
 - winner - Match-winning team
 - player_of_the_match - Best performer
 - most_runs, most_wkts - Top scorer and best bowler

3. Data Cleaning

- **Handling Missing Values:**
 - Filled missing values in `first_score`, `second_score` with 0.
 - Replaced missing `first_wkts`, `second_wkts` with 10.
- **Removing Duplicates:** Eliminated duplicate rows to maintain data integrity.
- **Data Formatting:** Standardized date formats and team names.

4. Exploratory Data Analysis (EDA)

- **Most Successful Teams:** Counted match wins per team.
- **Impact of Toss:** Analyzed correlation between toss winners and match outcomes.
- **Top Players:** Identified the most impactful players based on awards.

5. Data Visualization

- **Team Performance:** Bar chart showing number of wins per team.
- **Player Analysis:** Line chart of top run-scorers.
- **Toss Decision Impact:** Box plot comparing toss choices with first innings scores.
- **Winning Trends:** Line chart showing IPL 2024's top-performing teams.

6. Advanced Analysis - Predicting IPL 2025 Winner

- **Machine Learning Model:** Used RandomForestClassifier to predict match winners based on historical data.
- **Visualization:**

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(10,5))
sns.lineplot(x=team_wins["team"], y=team_wins["wins"], marker="o", linestyle="-",
color="b")
plt.xlabel("Teams")
plt.ylabel("Total Wins in IPL 2024")
plt.title("Trend Analysis - Which Team Can Win IPL 2025?")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```

Findings: Teams with higher IPL 2024 wins are strong contenders for IPL 2025.

7. Conclusion & Future Work

- **Key Insights:**
 - Winning teams consistently perform well across seasons.
 - Toss decisions impact first innings scores but don't always decide match outcomes.
 - Certain players dominate run-scoring and wicket-taking metrics.
- **Next Steps:**
 - Enhance predictive accuracy with deep learning.
 - Develop a Power BI/Tableau dashboard for real-time analysis.
 - Expand dataset with player performance trends for better insights.

Project GitHub Repository

GitHub Link : [IPL 2024 Data Analyst Project](#)