

PROJECT REPORT
ON
IPL AUCTION DATA ANALYSIS

By

Gadde Maruthi Mahesh
AP21110010289

Under Guidance of

K. Meenakshi Mam, APSSDC
K. Narmada Mam, APSSDC

Prepared in the partial fulfillment of the
Summer Internship Course

AT

**Andhra Pradesh State Skill Development
Corporation (APSSDC)**
3rd floor,Infosight Building, survey No. 78/2,Near Pathuru Road Junction,NH-16
Service Road, Tadepalli - 522501,Guntur District,Andhra Pradesh



ACKNOWLEDGEMENTS

We would like to express my special thanks to our trainers Meenakshi and Narmada Mam for their time and efforts that they provided throughout the internship. Your useful advice and suggestions were really helpful to us during the project's completion. In this aspect, We are eternally grateful to you.

We are highly indebted to APSSDC management for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

ABSTRACT

The provided code involves an analysis of the IPL 2022 auction dataset using Python's Pandas, Seaborn, and Matplotlib libraries. The dataset comprises player details, their base prices, roles, costs in Indian rupees and dollars, previous team information, and the team they were picked by. The code begins with data loading and preprocessing steps, addressing null values and transforming data for further analysis. It then proceeds to answer various questions using visualizations and statistics. Key insights include the number of players in different roles, the proportion of players sold in the auction, the count of players acquired by each team, the distribution of players based on retention or draft selection, and the count of players for each role. The code presents a thorough data-driven exploration of the IPL 2022 auction dataset, shedding light on player participation, team choices, and role distributions.

The code snippet presented encapsulates a comprehensive exploration and analysis of the IPL 2022 auction dataset using Python's powerful data manipulation and visualization libraries. The dataset itself encompasses critical information about players, their roles, base prices, costs in both Indian rupees and dollars, as well as team affiliations from the previous year and the current team they were picked by. The code's systematic approach offers valuable insights into the dynamics of player recruitment and team-building strategies in one of the world's most-watched cricket leagues. The initial phase of the code involves data loading using the Pandas library, followed by a series of data preprocessing steps aimed at ensuring the data's quality and readiness for analysis. Notably, the 'Unnamed: 0' column, likely an index column, is removed to streamline the dataset. Null values are efficiently handled in crucial columns such as 'COST IN ₹ (CR.)', 'Cost IN \$ (000)', and '2021 Squad'.

System Requirements

- PC(Minimum 4 GB Ram and 256 SSD)
- OS (Windows , Mac)
- Python(Version 3.10)
- Jupyter Notebook/JupyterLab (IDE)

INTRODUCTION

The Indian Premier League (IPL) stands as one of the most captivating cricket leagues globally, characterized by its intense competition, star-studded lineups, and strategic team compositions. Central to this dynamic is the player auction, a high-stakes event where franchises vie to secure players who can elevate their performance and contribute to their quest for glory. In this context, the provided code offers a comprehensive analysis of the IPL 2022 auction dataset, employing Python's data manipulation and visualization tools to unearth insights that illuminate the intricate dynamics of team-building strategies, player preferences, and role distributions. The IPL 2022 auction dataset encapsulates a wealth of player-related information, including their roles (Batsman, Bowler, All-Rounder, or Wicketkeeper), base prices, costs in Indian rupees and dollars, as well as their previous and current team affiliations. This code not only presents a meticulous exploration of these attributes but also employs data preprocessing techniques to ensure the data's integrity and readiness for analysis. The code embarks on its analysis journey by addressing pertinent questions that unveil crucial insights. The distribution of players based on their roles is depicted through a pie chart, highlighting the prevalence of all-rounders, bowlers, batsmen, and wicket keepers. This initial visualization lays the foundation for understanding how teams seek versatility and balance within their squads. Delving deeper, the code navigates the pivotal question of player acquisition. By distinguishing between sold and unsold players, it quantifies the success rate of securing players in the auction. This insight offers a glimpse into the competitiveness of the auction process and the factors that influence teams' decision-making. A critical aspect of IPL dynamics is the interaction between players and teams, and the code addresses this through a comprehensive visual representation of the number of players acquired by each team. This team-centric analysis unveils the strategies of different franchises, shedding light on their pursuit of specific player profiles to fulfill their strategic objectives. The code's analysis further extends to understanding retention and draft selection strategies, encapsulating the balance between familiar faces and new talents in teams' lineups.

Uses of Data Analysis library

Data analysis in libraries has various uses that can significantly improve library services, operations, and user experiences. Some of the key uses of data analysis in libraries include User Behavior Analysis, Collection Development, Resource Utilization, Service Improvement, Predictive Analysis, Space Management, Cost Management, Decision Making, Marketing and Outreach. Here are some libraries used in this project

Numpy : Python library for numerical computing and array manipulation.

Pandas: A data manipulation and analysis library that offers easy-to-use data structures and data analysis tools, making it efficient for data preprocessing and cleaning tasks.

Matplotlib: A plotting library that produces publication-quality visualizations, allowing users to create a wide variety of plots and charts.

Seaborn: A statistical data visualization library that builds on top of Matplotlib, providing additional functionality and more visually appealing plots.

SQL : A database library used for storing AQI values

Requests: Used for making HTTP requests to a specified URL.

bs4 BeautifulSoup: To scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.

Descriptive Statistics: To summarize and describe the main characteristics of a dataset, such as measures of central tendency (mean, median, mode) and measures of dispersion (standard deviation, range).

Data Visualization: Data visualization techniques, such as histograms, bar plots, scatter plots, and box plots, are used to visually represent the distribution, relationships, and patterns within the data.

Correlation Analysis: Correlation analysis examines the relationship between variables in a dataset. It helps identify the strength and direction of the relationship using correlation coefficients, such as Pearson's correlation coefficient.

Missing Data Handling Data Transformation and Filtering: Concepts used in exploratory data analysis to address missing values, standardize data, and extract specific subsets, respectively.

PURPOSE OF THE PROJECT

The purpose of this project is to perform a comprehensive analysis of the IPL 2022 auction dataset using data manipulation and visualization techniques. By delving into player roles, acquisition rates, team preferences, and retention strategies, the project aims to uncover valuable insights that shed light on the intricacies of player auctions in the IPL. This analysis serves as a strategic resource for teams, analysts, and cricket enthusiasts, facilitating informed decision-making in team compositions, recruitment strategies, and discussions surrounding the ever-evolving dynamics of the league.

The overarching purpose of this project is to conduct an in-depth and insightful analysis of the IPL 2022 auction dataset. Through meticulous data manipulation and powerful visualization techniques, the project endeavors to unravel hidden patterns and trends within the data, thereby providing valuable insights into the complex world of IPL player auctions.

By categorizing players into distinct roles such as Batsman, Bowler, All-Rounder, and Wicketkeeper, the project aims to highlight the evolving preferences and strategies of teams in the IPL. This analysis can offer a deeper understanding of how teams are adapting to the changing demands of the shortest format of cricket.

The project delves into the success rate of player acquisition by distinguishing between sold and unsold players. This analysis provides a glimpse into the competitiveness of the auction process and potentially uncovers factors that contribute to a player's marketability. Through visualizations illustrating the number of players acquired by each team, the project sheds light on team preferences and strategic inclinations. This insight can be invaluable for team management, enabling them to gauge the overall direction and focus of their rivals.

OBJECTIVE AND SCOPE OF THE PROJECT

The primary objective of this project is to conduct a comprehensive analysis of the IPL 2022 auction dataset to extract valuable insights regarding player roles, acquisition rates, team preferences, and retention strategies. The project aims to provide data-driven insights that contribute to a deeper understanding of the dynamics of player auctions in the IPL, aiding teams, analysts, and cricket enthusiasts in making informed decisions and fostering meaningful discussions. The project's scope includes the acquisition of the IPL 2022 auction dataset and subsequent preprocessing steps. This involves handling missing values, transforming data types, and ensuring the dataset's integrity. **Role-Based Analysis:** The project encompasses an analysis of player roles, including Batsman, Bowler, All-Rounder, and Wicketkeeper. This analysis will provide insights into the distribution of roles and how teams prioritize each role in their strategies. **Auction Success Rate:** The scope extends to analyzing the success rate of player acquisition in the auction. This involves categorizing players as sold or unsold and calculating the percentage of players successfully acquired. The project analyzes the distribution of players among different teams, shedding light on each team's preferences and strategies during the auction. **Retention and Draft Strategies:** The scope includes an exploration of the retention and draft strategies employed by teams. This will unveil the balance between retained players and new acquisitions and reveal the approaches teams take to build their squads. **Data Visualization:** The project's scope encompasses the use of various data visualization techniques, including pie charts, bar plots, and annotations, to effectively convey insights to stakeholders.

LITERATURE REVIEW

1. Data Analysis and Visualization Techniques: The code employs popular Python libraries like Pandas, NumPy, Seaborn, and Matplotlib for data manipulation, cleaning, and visualization. These libraries are widely used in data analysis projects due to their efficiency and versatility.

2. Data Preprocessing: Data preprocessing steps, such as handling missing values and transforming data types, are essential to ensure the data's quality and integrity. This aligns with best practices in data analysis, which emphasizes the importance of clean and consistent data.

3. Exploratory Data Analysis (EDA): EDA is a crucial step in understanding the characteristics of the dataset. The code's role-based analysis, distribution of sold vs. unsold players, and team preferences analysis are typical components of EDA. This approach is in line with the practice of diving into data distribution, patterns, and trends before moving on to more advanced analyses.

4. Visualization for Insights: The code employs various types of visualizations, such as pie charts, bar plots, and annotations, to present insights effectively. Effective data visualization is a critical component of any data analysis project, as it makes complex information accessible and understandable to a wide audience.

5. Sports Analytics and Decision-Making: The project's focus on analyzing player roles, team preferences, and acquisition strategies resonates with the growing field of sports analytics. In professional sports, data-driven insights are used to inform decision-making, including player recruitment, team strategies, and performance analysis.

6. Business Intelligence and Strategy: The project's aim to provide strategic insights aligns with the broader field of business intelligence. Organizations, including sports franchises, use data analysis to gain competitive advantages, optimize processes, and make informed decisions.

DATA ANALYSIS OVERVIEW

1. Data Loading and Inspection: The code starts by importing necessary libraries, including Pandas, NumPy, Seaborn, and Matplotlib. The dataset, likely in CSV format, is loaded into a Pandas DataFrame using the `pd.read_csv()` function. Basic information about the dataset is revealed using methods like `.shape` and `.columns`.

2. Data Preprocessing: The code drops the 'Unnamed: 0' column, which appears to be an index column. Null values in columns like 'COST IN ₹ (CR.)', 'Cost IN \$ (000)', and '2021 Squad' are handled using appropriate methods: Missing values in cost-related columns are filled with zeros since they correspond to players who went unsold. Missing values in '2021 Squad' column are replaced with a label indicating that the player didn't participate in IPL 2021.

3. Role-Based Analysis: Players are categorized into roles: Batsman, Bowler, All-Rounder, and Wicketkeeper. A pie chart is created to visualize the distribution of players across these roles. This analysis provides insights into the composition of players in different cricketing roles.

4. Auction Success Rate: Players are categorized into sold and unsold based on their status. A bar plot is generated to visualize the count of sold and unsold players. This analysis showcases the effectiveness of the auction process in acquiring players.

5. Team Preference Analysis: The count of players acquired by each team is visualized using a bar plot. This visualization provides insights into the recruiting strategies of different IPL teams.

6. Retention and Draft Strategies: The code investigates the number of players retained and acquired through the draft for each team. This analysis gives a clear picture of how teams combine retained players and fresh acquisitions to build their squads.

DESIGN OF THE SOLUTION

1. Problem Statement: Analyze the IPL 2022 auction dataset to extract insights into player roles, acquisition rates, team preferences, and retention strategies.

2. Solution Design:

2.1 Data Loading and Inspection: Import necessary libraries: Pandas, NumPy, Seaborn, Matplotlib. Load the IPL 2022 auction dataset using `pd.read_csv()`. Print basic information about the dataset using `.shape` and `.columns`.

2.2 Data Preprocessing: Drop the 'Unnamed: 0' column using `.drop()`. Handle missing values in 'COST IN ₹ (CR.)', 'Cost IN \$ (000)', and '2021 Squad' columns using `.fillna()`. Replace null values in '2021 Squad' with a label using `.fillna()`.

2.3 Role-Based Analysis: Categorize players into roles: Batsman, Bowler, All-Rounder, and Wicketkeeper. Count the number of players in each role. Visualize the distribution using a pie chart.

2.4 Auction Success Rate: Categorize players as sold or unsold based on their status. Count the number of sold and unsold players. Visualize the count using a bar plot.

2.5 Team Preference Analysis: Count the number of players acquired by each team. Visualize the count using a bar plot.

2.6 Retention and Draft Strategies: Group data by 'Curr_team' and 'retention' columns. Count the number of players retained and acquired through the draft for each team. Visualize using a tabular format or bar plot.

2.7 Data Visualization: Use Matplotlib and Seaborn to create visualizations like pie charts and bar plots. Enhance the visualizations with appropriate labels, titles, and annotations.

2.8 Strategic Insights: Translate analysis findings into actionable insights for team management and stakeholders. Provide recommendations for optimizing team composition and recruitment strategies.

2.9 Stakeholder Engagement: Highlight how the insights generated from the analysis can engage various stakeholders, including team management, analysts, and fans. Emphasize the potential for stimulating discussions and debates within the cricket community.

OUTCOMES

The outcomes of the data analysis conducted in the provided code are valuable insights and information derived from the IPL 2022 auction dataset. These insights provide a deeper understanding of player roles, team strategies, acquisition rates, and more, which can be useful for various stakeholders involved in cricket, including team management, analysts, and enthusiasts. Here are some specific outcomes:

- 1. Role-Based Insights:** Understanding the distribution of players across roles (Batsman, Bowler, All-Rounder, Wicketkeeper). Identifying the most common roles and their proportions in the auction.
- 2. Auction Success Analysis:** Determining the success rate of players' acquisition in the auction. Visualizing the count of sold and unsold players to showcase the effectiveness of the auction process.
- 3. Team Preference Understanding:** Gaining insights into the number of players acquired by each team. Identifying teams' preferences and strategies for recruiting players.
- 4. Retention and Draft Strategies:** Analyzing the balance between retained players and new acquisitions. Understanding how teams approach building their squads through a combination of retention and draft picks.
- 5. Data-Driven Recommendations:** Providing actionable recommendations based on the analysis findings. Offering insights that can help teams optimize player recruitment strategies, achieve a balanced team composition, and enhance overall performance.
- 6. Engagement and Discussion:** Engaging stakeholders including team management, analysts, and cricket enthusiasts in discussions and debates. Stimulating conversations around player recruitment, team strategies, and auction dynamics.

Advantages

- 1. Informed Decision-Making:** The data analysis provides valuable insights that empower teams with information to make informed decisions regarding player acquisitions, team composition, and strategies.
- 2. Optimal Recruitment Strategies:** Teams can use the analysis findings to optimize their recruitment strategies by understanding player roles, team preferences, and retention-draft balance.
- 3. Competitive Advantage:** Teams that leverage data-driven insights are better positioned to gain a competitive edge by making strategic choices aligned with market dynamics.
- 4. Enhanced Performance:** Informed recruitment decisions based on player roles and historical trends can lead to a more balanced team composition and improved overall performance.
- 5. Resource Allocation:** Analysis helps teams allocate their resources effectively by identifying roles that are more prevalent or strategic in the auction.
- 6. Fan Engagement:** Deeper insights into team dynamics and acquisition strategies enhance fan engagement, discussions, and interactions around the sport.
- 7. Evidence-Based Discussions:** Stakeholders, including analysts, commentators, and enthusiasts, can engage in evidence-based discussions backed by data analysis findings.
- 8. Data Literacy Promotion:** The project promotes data literacy within the cricket community, encouraging the understanding and utilization of data in decision-making.
- 9. Continuous Improvement:** Teams can iteratively improve their strategies by incorporating analysis insights into future auctions, leading to more effective decision-making over time.
- 10. Trend Identification:** The analysis may reveal long-term trends in player recruitment, helping teams adapt their strategies to evolving market dynamics.

CONCLUSIONS

The data analysis conducted in the provided code yields several insightful conclusions that shed light on various aspects of the IPL 2022 auction. These conclusions offer valuable insights for team management, analysts, enthusiasts, and the cricketing community as a whole:

1. Role Distribution: The majority of players in the auction are All-Rounders (38.2%), followed by Bowlers (34%), Batsmen (17.7%), and Wicketkeepers (10.1%).

2. Auction Success Rate: Out of 633 players, 237 players (37.5%) were successfully acquired in the auction, while 396 players (62.5%) went unsold. The auction process demonstrates varying success rates for different players based on factors such as skill, reputation, and team requirements.

3. Team Preferences: Teams like Punjab Kings, Kolkata Knight Riders, and Mumbai Indians acquired the highest number of players in the auction, showcasing their active participation and strategic preferences.

4. Retention and Draft Strategies: Teams like Chennai Super Kings, Delhi Capitals, and Rajasthan Royals have adopted a balanced approach, retaining a few players while actively participating in the auction.

5. Data-Driven Insights: The analysis provides actionable insights, enabling teams to make informed decisions regarding player recruitment, team composition, and strategic alignment.

6. Enhanced Fan Engagement: The analysis contributes to a deeper understanding of team dynamics, which can enrich fan discussions, debates, and interactions around the sport.

REFERENCES

[Kaggle: Your Machine Learning and Data Science Community](#)