Exploratory Data Analytics Project

IPL Auction 2013

C. Yashwanth Chowdary

1. Introduction

The Indian Premier League (IPL) is one of the most popular cricket leagues globally. Every year, teams bid for players in an auction, and analyzing this auction data can provide insights into trends, spending patterns, and the value of players. This project focuses on performing Exploratory Data Analysis (EDA) on IPL auction data of 2013 to uncover meaningful patterns and relationships.

2. Objective

The primary objectives of this project are:

- To clean and preprocess the IPL auction data.
- To perform Exploratory Data Analysis (EDA) to understand the dataset.
- To generate meaningful insights using visualizations.
- To identify patterns such as player demographics, team spending, and factors influencing sold price.

3. Data Overview

The dataset used in this analysis consists of data about IPL players and their respective details from past auctions. The columns in the dataset include:

- Sl.NO.: Serial number
- PLAYER NAME: Name of the player
- AGE: Age of the player

- COUNTRY: Player's country

- TEAM: Team the player belongs to

- PLAYING ROLE: The player's role (batsman, bowler, all-rounder)

- T-RUNS: Total runs scored in T20 matches

- T-WKTS: Total wickets taken in T20 matches

- ODI-RUNS-S: Runs scored in ODI matches

- SOLD PRICE: The price at which the player was sold during the auction

- And more...

4. Data Cleaning and Preprocessing

The following steps were performed to clean and prepare the data:

1. Splitting columns: The raw dataset had all the details in a single column, which was split into individual columns using the separate() function.

2. Converting Data Types: Columns with numerical values (e.g., Age, Runs, Wickets, Price) were converted to appropriate numeric types using mutate().

3. Handling Missing Data: Missing values in the "Captaincy Experience" column were replaced with 0 to indicate no captaincy experience.

4. Handling Special Characters: Some columns had commas as thousand separators, which were removed and replaced with periods.

5. Exploratory Data Analysis (EDA)

5.1. Basic Summary

• A summary of the dataset was obtained using the summary() function, providing an overview of the data, including the mean, median, and range for numeric variables.

• The missing data was identified using the colSums(is.na ()) function to ensure no critical data points were left unaddressed.

5.2. Visualisations

The following visualisations were created to provide deeper insights:
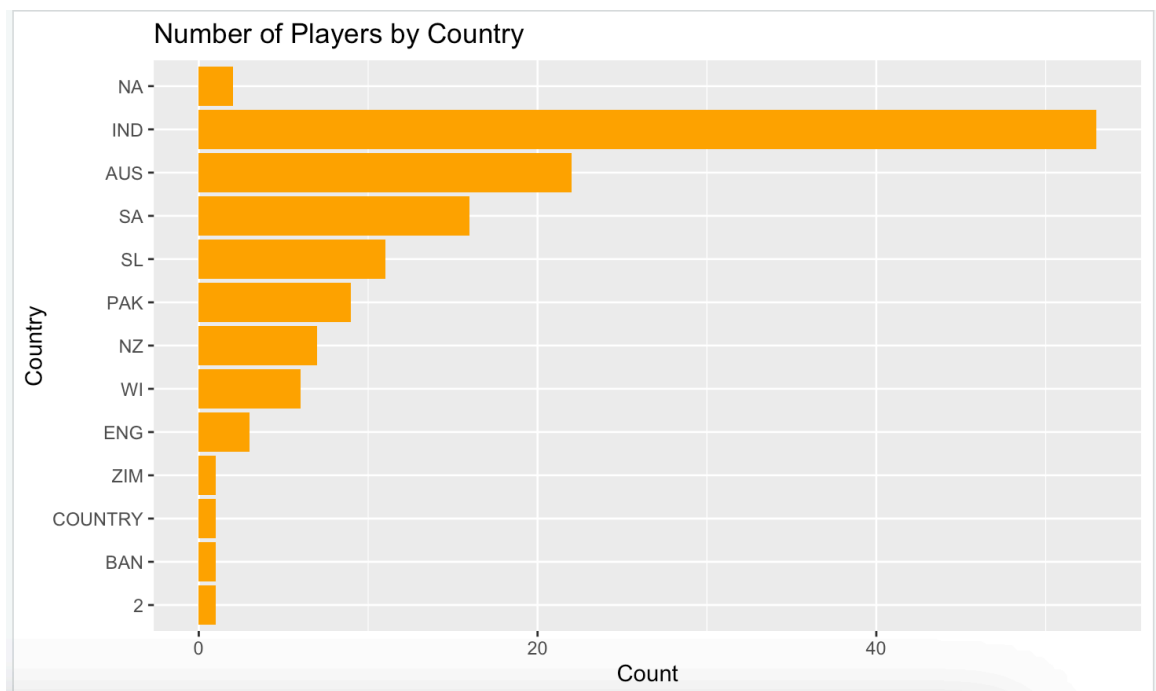
1. Age Distribution:

A histogram was plotted to visualise the age distribution of players, showing that most are between 25 and 30 years old.
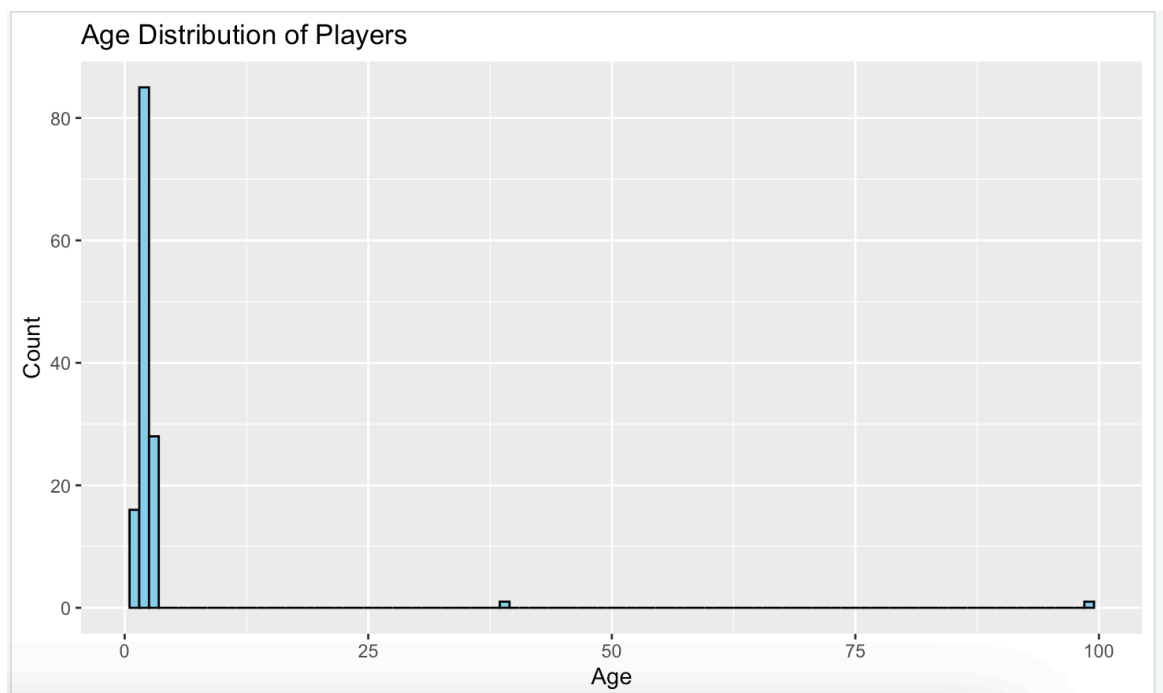
2. Country-wise Player Count:

A bar chart displayed the number of players from each country, highlighting that Indian players dominate the auction.


Number of Players by Country

3. Base Price vs. Sold Price:

A scatter plot was created to show the relationship between base price and sold price, with a linear regression line to show the correlation.
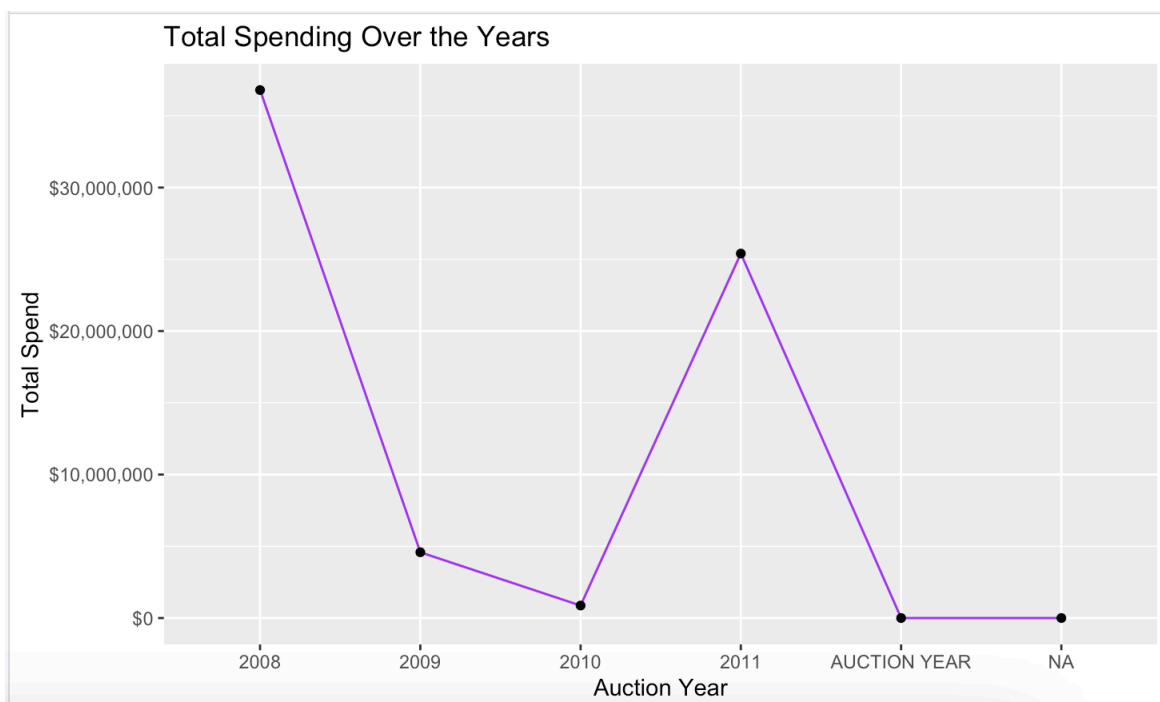

Base Price vs Sold Price

4. Top 10 Highest-Selling Players:

A bar chart displayed the top 10 most expensive players, showing which players fetched the highest prices.

Top 10 Expensive Players
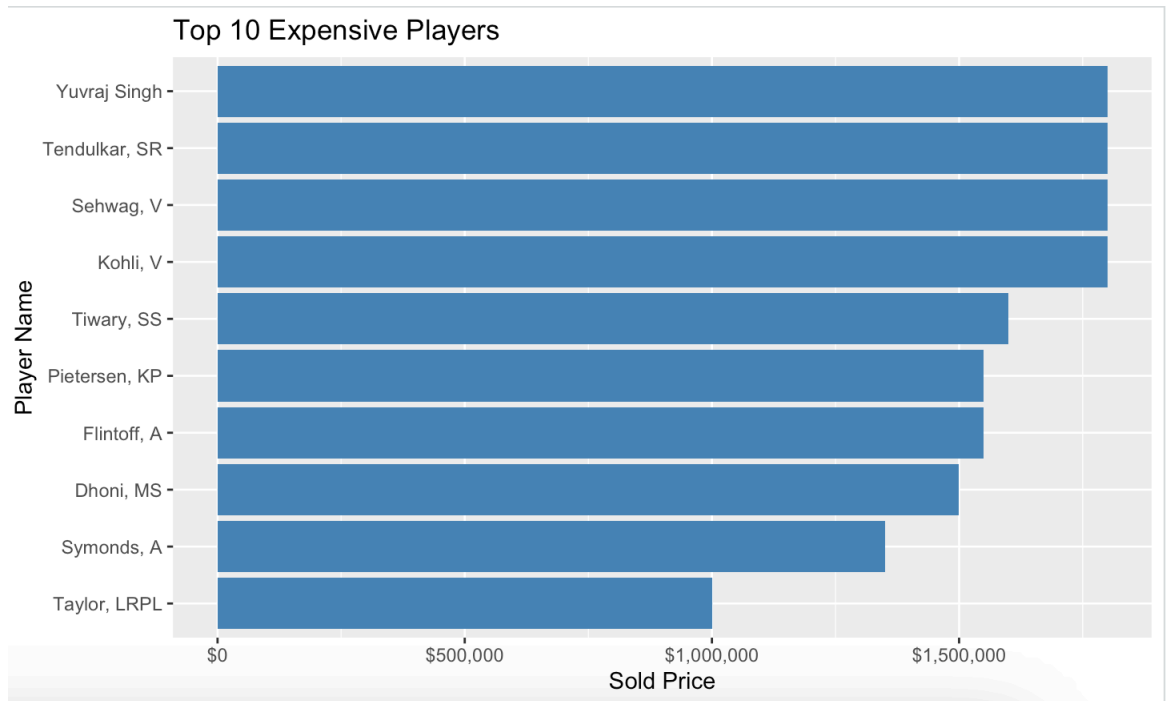
5. Auction Year vs. Total Spend:

A line plot displayed how much was spent in each auction year, showing trends in spending patterns.
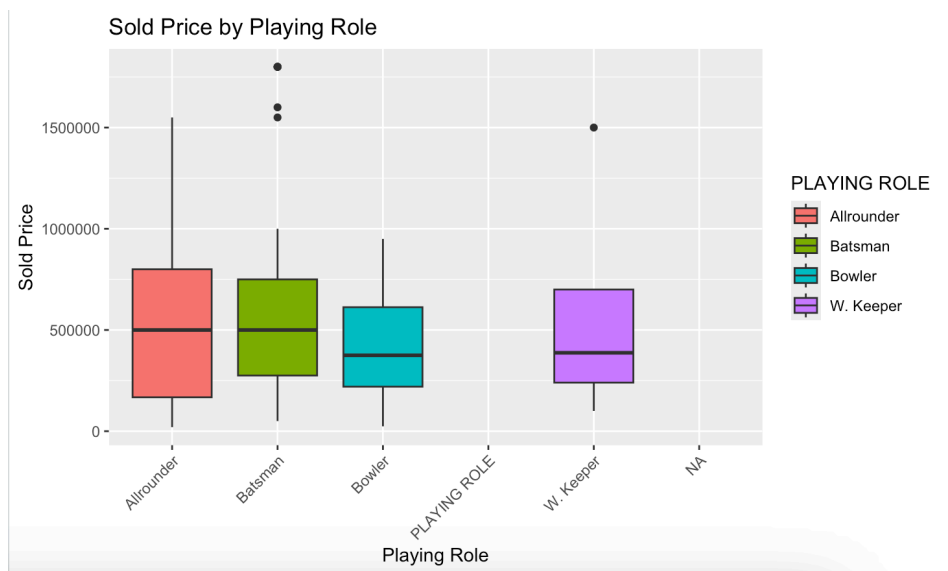

Total Spending Over the Years

6. Team-wise Spending:

A bar chart highlighted the teams that spent the most in the auction, with teams like CSK and MI leading the spending.

Spend by Teams
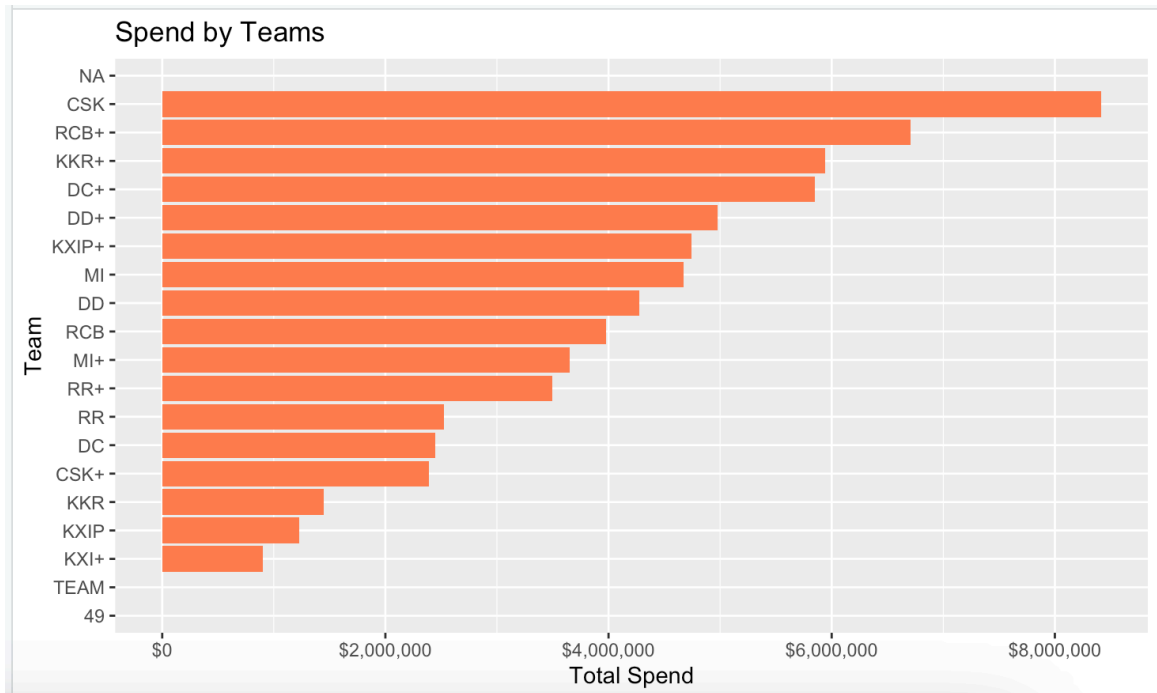
7.   Playing Role vs Sold Price:

A boxplot showed how different playing roles (batsman, bowler, all-rounder) affected the sold price.


Sold Price by Playing Role

8. Correlation Heatmap:

A correlation matrix was visualised using a heatmap to analyse the relationships between various numeric features in the dataset, such as career runs, wickets, age, and auction price.
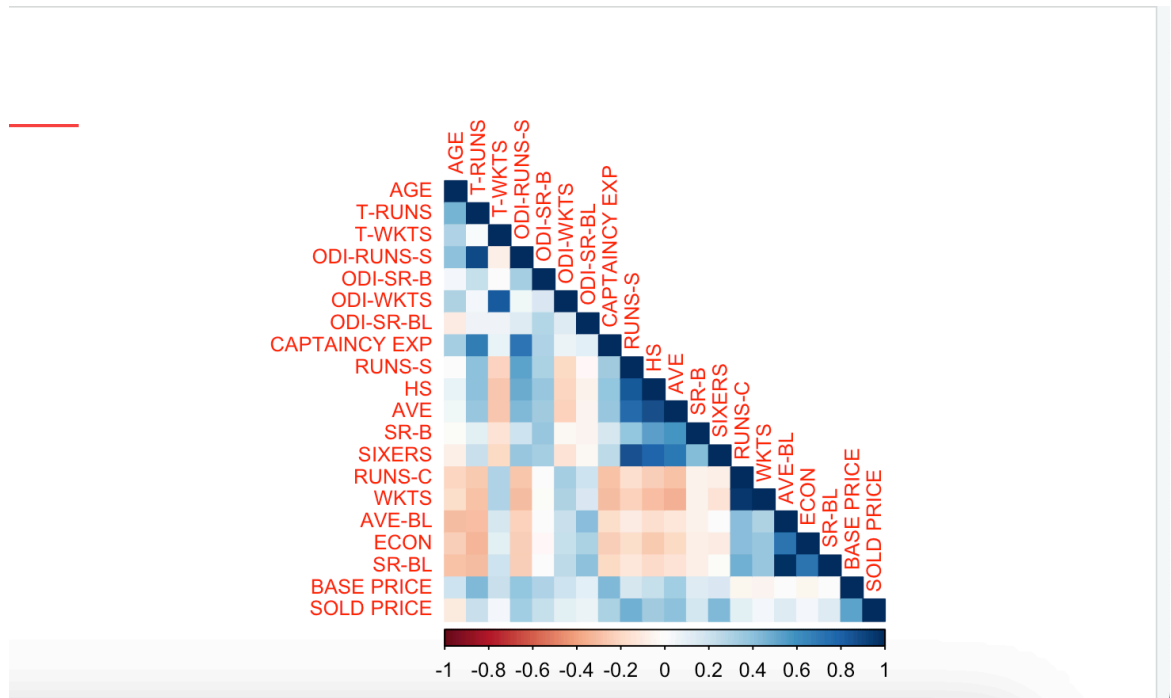

9. Total Career Runs vs Sold Price:

A scatter plot with a regression line was created to examine how a player's total career runs impact their auction sold price, identifying the correlation between batting performance and market value.
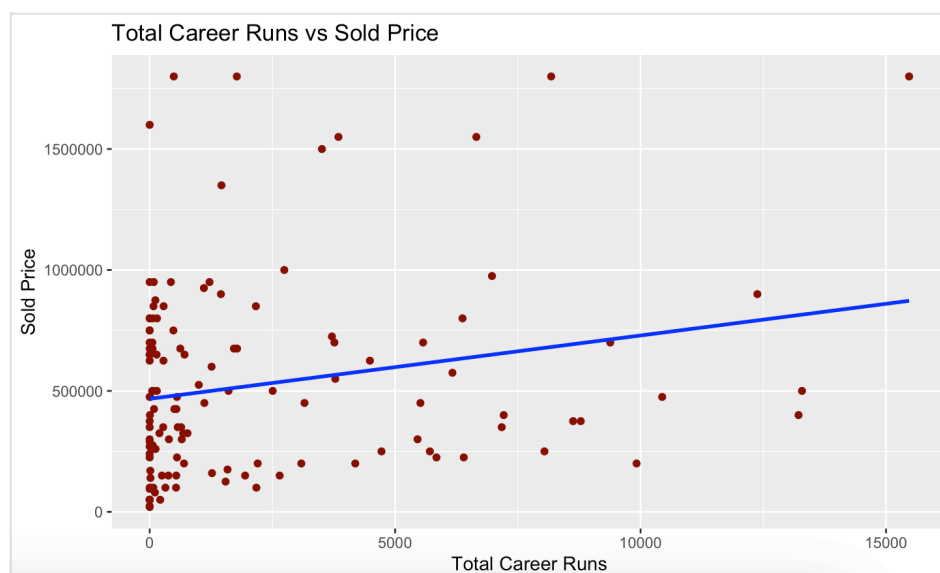
10. Total Career Wickets vs Sold Price:

A scatter plot with a trend line was plotted to explore the relationship between a player's total career wickets and their sold price, highlighting how bowling achievements affect auction demand.

Total Career Wickets vs Sold Price
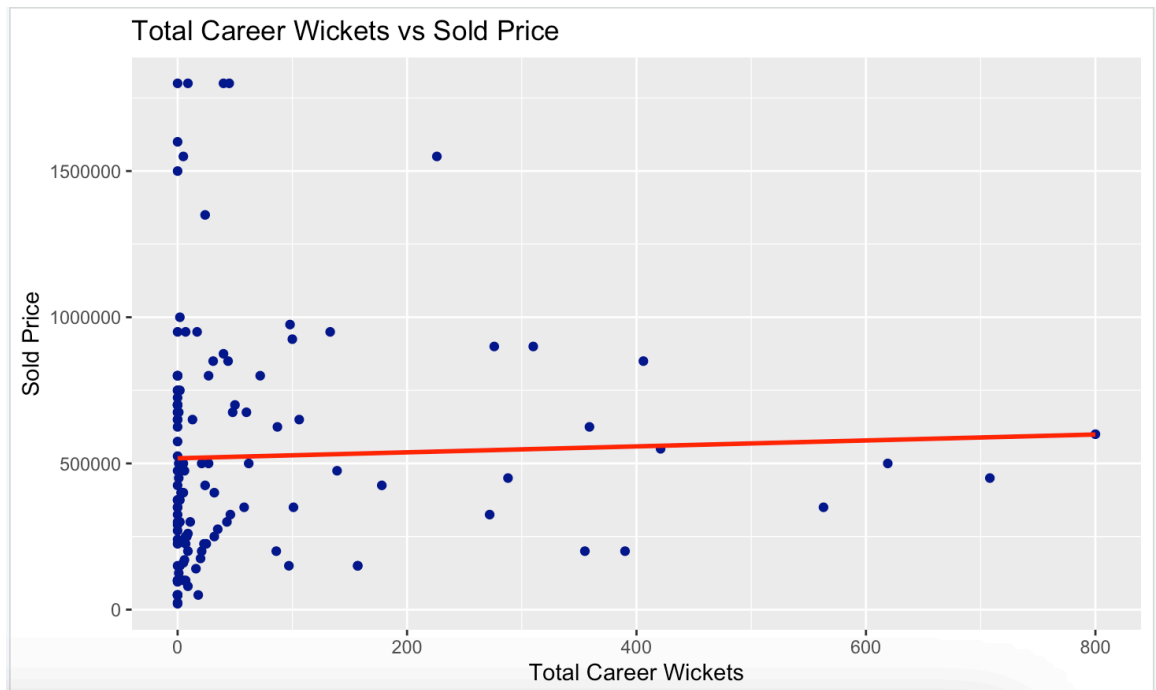
11. Captaincy Experience Impact:

A scatter plot was drawn to study the effect of captaincy experience on auction price. A trendline was added to observe whether leadership experience significantly boosts a player's auction value.

Captaincy Experience vs Sold Price
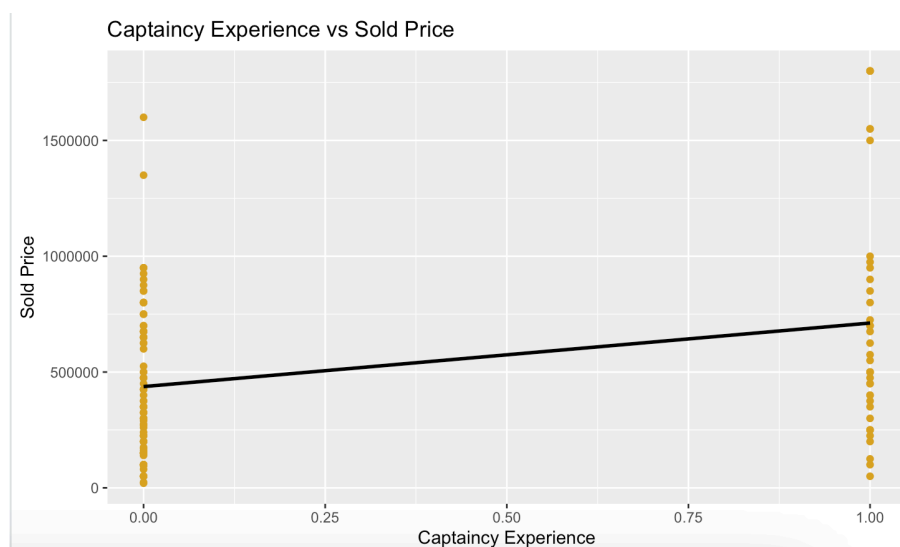
Trend of Sold Price with Age:

A scatter plot with a LOESS smooth curve was used to visualise how a player's age influences their auction sold price, revealing non-linear trends between experience and market value.

Trend: Age vs Sold Price

5.3. Insights from EDA

1. Age Distribution: The majority of players are between 25 and 30 years old, which is the prime age for players to perform at their best.

2. Country-wise Analysis: Indian players dominate the auction, with a significant representation from countries like Australia, South Africa, and the West Indies.

3. Auction Price Influences: A strong correlation between the base price and sold price was observed. Higher base prices generally resulted in higher selling prices.

4. Team Spending: Teams like CSK and MI have the highest spending in the auctions, indicating their large player investments.

5. Role-based Auction Price: Batsmen and all-rounders generally fetched higher prices compared to bowlers.

6. Captaincy Experience: Players with captaincy experience tended to fetch slightly higher prices in the auction.

7. Career Runs and Wickets: Both total career runs and wickets have a noticeable impact on the sold price, with players having higher numbers in these categories being valued higher.

6. Git Repo link:

https://github.com/mintu3770/ipl_auction_EDA

7. Conclusion

The EDA on the IPL auction data reveals several trends and patterns:

• Age and Player Value: Players in the prime age range (25-30 years) are highly sought after.

• Country Influence: Indian players have a significant representation in the auction, likely due to the popularity of IPL in India.

• Role and Value: All-rounders and batsmen fetch higher prices than bowlers.

• Captaincy: Players with prior captaincy experience have an edge in terms of pricing.

• Team Spending: Teams like CSK and MI dominate in terms of total spending at the auctions. These insights can help teams make more informed decisions during the auction process, especially when considering player demographics and roles.