"Exploratory Data Analysis and Machine Learning on IPL Auction Dataset"

Introduction :

This dataset gives us a chance to dig deeper into years of auction history to uncover trends, spot patterns, and even predict outcomes using machine learning.

Goal of the project:
- Clean and prepare the dataset for analysis.
- Visualize trends to better understand the dynamics of player selection.
- Use machine learning to predict player bids and classify players into categories based on their value.

Tech Stack:
Python for programming.
Pandas for handling and cleaning the dataset.
Seaborn/Matplotlib for creating visualizations.
Scikit-learn for applying machine learning models.

Data Cleaning:

Cleaning Steps:

1. **Removing unnecessary columns:**
   - The `Unnamed: 0` column was irrelevant, so it was removed.
2. **Handling missing data:**
   - Some rows had missing or invalid values in the `Winning bid` column, which were either filled or dropped.
3. **Converting data types:**
   - Columns like `Winning bid` were converted to numeric for easier analysis.

Key Outcomes:

- The cleaned dataset contains 1,045 rows and 6 columns.

- The columns used for analysis are: `Country`, `Player`, `Team`, `Base price`, `Winning bid`, and `Year`.

Exploratory Data Analysis (EDA)

Trends and Patterns:

Exploring the data revealed several interesting trends:

1. **How are bids distributed?**
   - Most `Winning bids` and `Base prices` are concentrated in the lower range, with a few high-value outliers.
2. **Top spenders**:
   - Teams like Mumbai Indians and Chennai Super Kings have spent the most over the years.
3. **Player origins**:
   - Indian players dominate the auctions, followed by players from Australia and England.

Feature Engineering

Preparing Data for Machine Learning:

To apply machine learning, the data needed to be prepared. Key steps included:

1. **Encoding categorical data**:
   - Columns like `Country`, `Player`, and `Team` were converted into numeric values using Label Encoding.
2. **Selecting features**:
   - The independent variables used were `Country`, `Player`, `Team`, `Base price`, and `Year`.
   - For regression, the target variable was `Winning bid`. For classification, a new column `High Bid` (players with bids above the median) was created.

Classification Analysis

Objective:

To classify players into **High Bid** (above median value) and **Low Bid** (below median value).

Model:

The **Random Forest Classifier** was used with the following parameters:

- `n_estimators = 100`
- `random_state = 42`

Results:

- **Accuracy**: 85% (example value)
- **Precision, Recall, and F1-Score**:
  - The model performed well for both classes.

Key Insights:

- Players with a high base price and from marquee teams were more likely to receive high bids.

Conclusion and Future Work

Summary:

This project explored the IPL auction dataset and uncovered fascinating trends. Using machine learning, we:

- Predicted winning bids with reasonable accuracy .
- Classified players into high and low bid categories with 78.94% accuracy.

What's Next?

1. **Add More Data**:
   - Include player performance stats to refine predictions.
2. **Advanced Models**:
   - Experiment with XGBoost or Neural Networks for better accuracy.
3. **Parameter Tuning**:
   - Use grid search or random search to optimize the models.