

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the CSV file
file_path = 'IPL.csv' # Update with the correct file path
ipl_data = pd.read_csv(file_path)

# Data Cleaning: Handling missing values and inconsistencies
ipl_data = ipl_data.copy() # Create a copy to avoid potential issues with chained assignments
ipl_data['Base Price IN ₹'] = ipl_data['Base Price IN ₹'].fillna(0)
ipl_data['COST IN ₹ (CR.)'] = ipl_data['COST IN ₹ (CR.)'].fillna(0)
ipl_data['2022 Squad'] = ipl_data['2022 Squad'].fillna('Not Retained')

# Question 1: Distribution of player types
player_type_distribution = ipl_data['TYPE'].value_counts()

# Question 2: Which team has the highest total expenditure on players in INR?
team_expenditure = ipl_data.groupby('Team')['COST IN ₹ (CR.)'].sum().sort_values(ascending=False)
top_team = team_expenditure.idxmax()
top_team_expenditure = team_expenditure.max()

# Question 3: Correlation between base price and final cost
correlation = ipl_data[['Base Price IN ₹', 'COST IN ₹ (CR.)']].corr().iloc[0, 1]

# Question 4: Most expensive and cheapest player in terms of final cost
most_expensive_player = ipl_data.loc[ipl_data['COST IN ₹ (CR.)'].idxmax(), 'Player']
cheapest_player = ipl_data.loc[ipl_data['COST IN ₹ (CR.)'].idxmin(), 'Player']

# Question 4: Most expensive and cheapest player in terms of final cost
most_expensive_player = ipl_data.loc[ipl_data['COST IN ₹ (CR.)'].idxmax(), 'Player']
cheapest_player = ipl_data.loc[ipl_data['COST IN ₹ (CR.)'].idxmin(), 'Player']

# Question 5: How many players were retained by their 2022 squad in 2023?
retained_players_count = ipl_data[ipl_data['2022 Squad'] != 'Not Retained'].shape[0]

# Question 6: Average cost of players for each role
average_cost_by_type = ipl_data.groupby('TYPE')['COST IN ₹ (CR.)'].mean()

# Question 7 (Updated): What is the average base price of players who were retained by their 2022 squad in 2023?
average_base_price_retained = ipl_data[ipl_data['2022 Squad'] != 'Not Retained']['Base Price IN ₹'].mean()

# Outputs for each question
outputs = {
    "Player Type Distribution": player_type_distribution,
    "Team with Highest Expenditure": top_team,
    "Top Team Expenditure (₹ Cr.)": top_team_expenditure,
    "Correlation (Base Price vs Final Cost)": correlation,
    "Most Expensive Player": most_expensive_player,
    "Cheapest Player": cheapest_player,
    "Retained Players Count": retained_players_count,
    "Average Cost by Player Type (₹ Cr.)": average_cost_by_type,
    "Average Base Price of Retained Players (₹)": average_base_price_retained
}

# Print the outputs
for key, value in outputs.items():
    print(f"{key}: \n{value}\n")

# Visualization for Question 1
plt.figure(figsize=(8, 5))
sns.countplot(data=ipl_data, x='TYPE')
plt.title('Distribution of Player Types')
plt.xlabel('Player Type')
plt.ylabel('Count')
plt.show()

```

```

>>>
= RESTART: C:/Users/gladly/OneDrive/Desktop/DS with Python/Task 6/Task6.py
Player Type Distribution:
TYPE
ALL-ROUNDER    213
BOWLER         189
BATSMAN        91
WICKETKEEPER   75
Name: count, dtype: int64

Team with Highest Expenditure:
Sunrisers Hyderabad

Top Team Expenditure (₹ Cr.):
35.7

Correlation (Base Price vs Final Cost):
-0.10956589267119142

Most Expensive Player:
Sam Curran

Cheapest Player:
Rahul Tewatia

Retained Players Count:
230

Average Cost by Player Type (₹ Cr.):
TYPE
ALL-ROUNDER    0.332160
BATSMAN        0.401099
BOWLER         0.170106
WICKETKEEPER   0.368000
Name: COST IN ₹ (CR.), dtype: float64

Average Base Price of Retained Players (₹):
1002173.9130434783

```

