

IPL 2024 SOLD PLAYERS DATA ANALYSIS USING PYTHON

MADHURA VIRKAR





AGENDA

1. Basic Data Exploration
2. Grouping and Aggregation
3. Filtering and Sorting
4. Advanced/Insightful





1. BASIC DATA EXPLORATION

1. How many players are there in total?

```
df.shape
```

```
(332, 6)
```

There are 332 players in total

2. How many players are Indian vs Overseas?

```
# Players Indian vs Overseas
indian_players = df[df["NATIONALITY"] == "Indian"].shape[0]
overseas_players = df[df["NATIONALITY"] != "Overseas"].shape[0]

print(f"Number of Indian players: {indian_players}")
print(f"Number of Overseas players: {overseas_players}")
```

Number of Indian players: 216

Number of Overseas players: 216



3. What is the average price of all players?

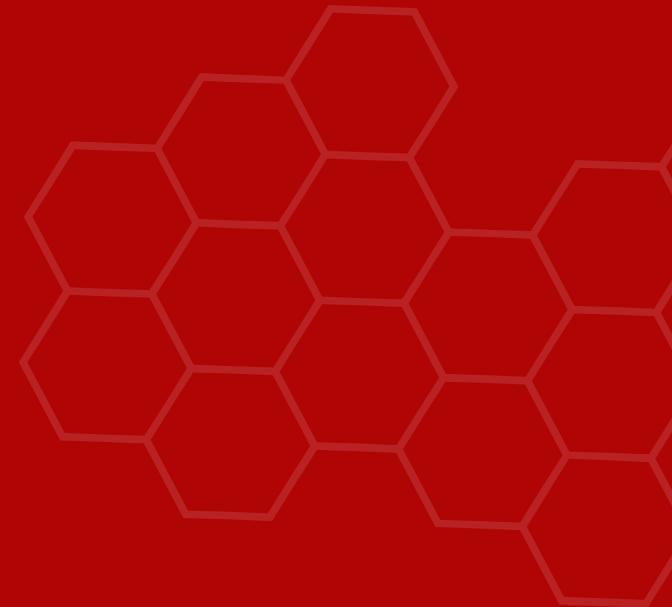
```
avg_price = np.average(df["PRICE"])
print("Average price of players:₹{:.2f}".format(avg_price))
```

Average price of players:₹10,093,373.49



4. List all unique player types.

```
# Unique player types  
# Use set() to get unique types  
unique_types = set(df["TYPE"])  
print(unique_types)  
  
{'All-Rounder', 'Bowler', 'Wicket-Keeper', 'Batter'}
```



5. Which team has the most players?

```
# Team with most players
```

```
players = df.groupby("TEAM").size().nlargest()  
most_player = players.idxmax()  
print(players)
```

TEAM	
Un Sold	260
KKR	10
DC	9
GT	8
MI	8



2. GROUPING AND AGGREGATION

6. What is the total amount spent by each team on players?

```
# Total amount spent per team
from sklearn import datasets
from IPython.display import display
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
total_amount_spent = df.groupby(['TEAM']).agg({'PRICE':['sum']}).style.format("₹{:, .2f}")
display(total_amount_spent)
print(total_amount_spent)
```

TEAM	sum
CSK	₹304,000,000.00
DC	₹190,500,000.00
GT	₹303,000,000.00
KKR	₹313,500,000.00
LSG	₹122,000,000.00
MI	₹167,000,000.00
PK	₹249,500,000.00
RCB	₹204,000,000.00
RR	₹143,000,000.00
SH	₹308,000,000.00

7. Which team spent the most money overall?

```
#Team spent max amount
df1 = df.drop(df[df["TEAM"]=="Un-Sold"].index)
amount_spent = df1.groupby(['TEAM'])['PRICE'].sum().sort_values(ascending=False).head(1)
display (amount_spent)
print(amount_spent)
```

TEAM	PRICE
KKR	313500000



8. What is the average price of Indian players vs Overseas players?

```
avg_price = df.groupby('NATIONALITY').agg({'PRICE':[ 'mean']}).style.format("₹{:, .2f}")  
display (avg_price)  
print(avg_price)
```

PRICE	
NATIONALITY	mean
Indian	₹5,400,462.96
Overseas	₹18,831,896.55



9. How many players of each type are there (e.g., how many Bowlers, Batters)?

```
# Number of players per type  
players_per_type = df.groupby('TYPE').size()  
print(players_per_type)
```

TYPE	
All-Rounder	114
Batter	53
Bowler	128
Wicket-Keeper	37

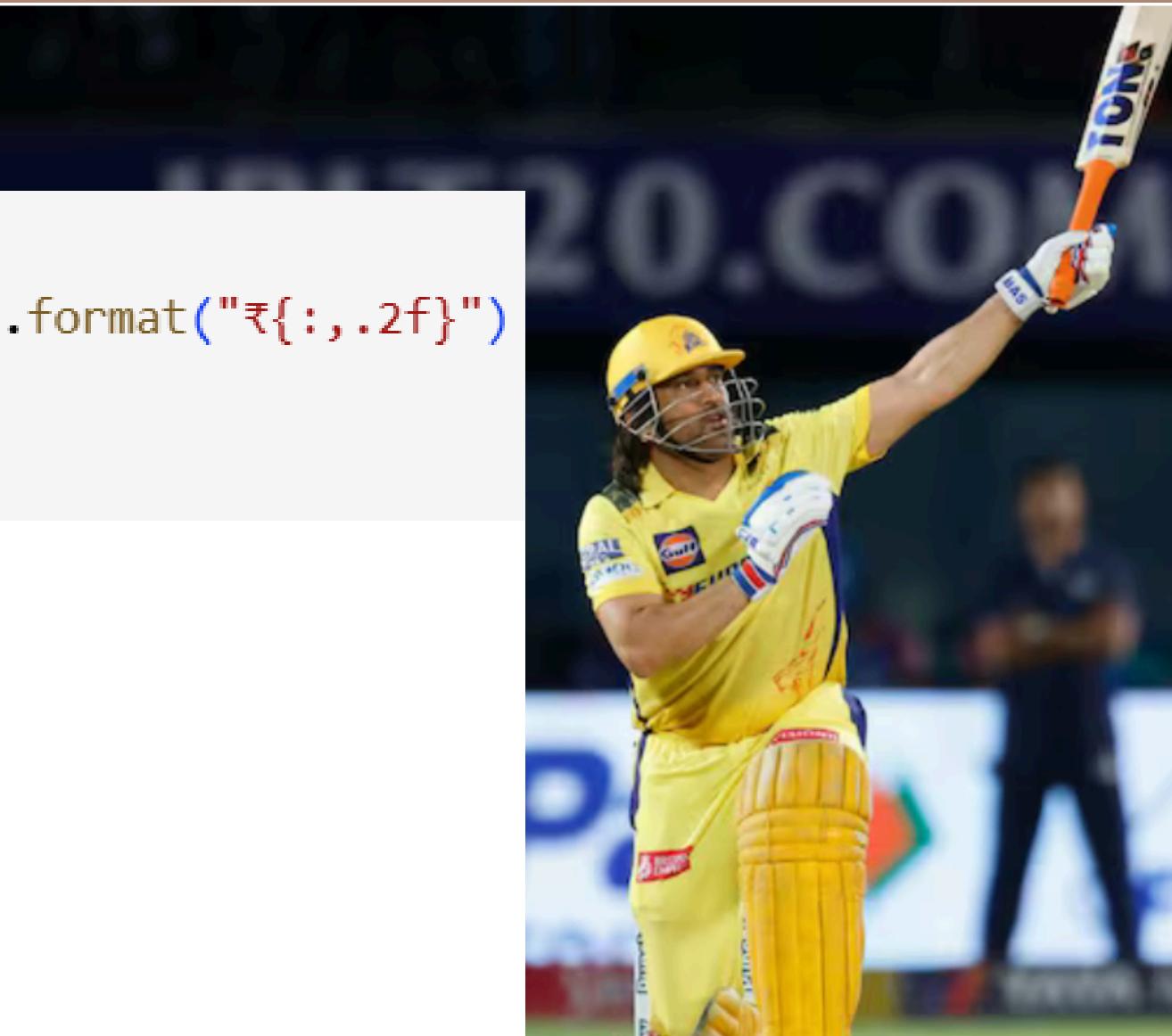


10. What is the average price of each type of player?



```
#Average price fo each palyer type
avg_price = df.groupby('TYPE').agg({'PRICE':['mean']}).style.format("₹{:, .2f}")
display (avg_price)
print(avg_price)
```

TYPE	PRICE
	mean
All-Rounder	₹9,934,210.53
Batter	₹11,669,811.32
Bowler	₹10,511,718.75
Wicket-Keeper	₹6,878,378.38



3. FILTERING AND SORTING

11. List all players who were bought for more than ₹5 crores (50,000,000).

```
# Players who were bought for more than ₹5 crores
high_priced_players = df[df['PRICE'] > 5000000]
print(high_priced_players)
```

PLAYERS	NATIONALITY	TYPE	PRICE	TEAM
Daryl Mitchell	Overseas	All-Rounder	14000000	CSK
Sameer Rizvi	Indian	Batter	84000000	CSK
Kumar Kushagra	Indian	Wicket-Keeper	72000000	DC
Spencer Johnson	Overseas	Bowler	10000000	GT
Shahrukh Khan	Indian	All-Rounder	74000000	GT
Umesh Yadav	Indian	Bowler	58000000	GT
Mitchell Starc	Overseas	Bowler	24750000	KKR
Shivam Mavi	Indian	Bowler	64000000	LSG
Rilee Rossouw	Overseas	Batter	80000000	PK
Harshal Patel	Indian	All-Rounder	11750000	PK
Shubham Dubey	Indian	Batter	58000000	RR
Rovman Powell	Overseas	Batter	74000000	RR
Alzarri Joseph	Overseas	Bowler	115000000	RCB
Pat Cummins	Overseas	All-Rounder	205000000	SH
Travis Head	Overseas	Batter	68000000	SH

12. Who is the most expensive player in the dataset?

```
#Most expensive player
most_expensive_player = df[df['PRICE'] == df['PRICE'].max()]
print(most_expensive_player)
```

S_No	PLAYERS	NATIONALITY	TYPE	PRICE	TEAM
27	Mitchell Starc	Overseas	Bowler	₹247,500,000.00	KKR





13. Which team has the most overseas players?

```
#Team with most overseas players
overseas_players = df[(df['NATIONALITY'] != 'Indian') & (df['TEAM'] != 'Un-Sold')]
team_with_most_overseas = overseas_players.groupby('TEAM').size().idxmax()
print(team_with_most_overseas)
```

DC





14. List all Wicket-Keepers and their respective teams and prices.

```
# List of Wicket-Keepers and their respective teams and prices.  
wicket_keepers = df[(df['TYPE'] == 'Wicket-Keeper') & (df['TEAM'] != 'Un-Sold')]  
  
list_wicket_keepers = wicket_keepers[['PLAYERS', 'TEAM', 'PRICE']]  
list_wicket_keepers["PRICE"] = df["PRICE"].map("₹{:,.2f}".format)  
print(list_wicket_keepers)
```

PLAYERS	TEAM	PRICE
Avanish Rao Aravelly	CSK	₹2,000,000.00
Ricky Bhui	DC	₹2,000,000.00
Shai Hope	DC	₹7,500,000.00
Kumar Kushagra	DC	₹72,000,000.00
Tristan Stubbs	DC	₹5,000,000.00
Robin Minz	GT	₹36,000,000.00
K.S. Bharat	KKR	₹5,000,000.00
Tom Kohler-Cadmore	RR	₹4,000,000.00



15. Find the top 3 most expensive Indian players.

```
#Expensive indian player
indian_players = df[(df['NATIONALITY'] == 'Indian') & (df['TEAM'] != 'Un-Sold')]
indian_players["PRICE"] = df["PRICE"].map("₹{:,.2f}".format)
most_expensive_indian = indian_players.groupby('PLAYERS')['PRICE'].sum().sort_values(ascending=False).head(3)
print(most_expensive_indian)
```

PLAYERS

Sameer Rizvi	₹84,000,000.00
Shahrukh Khan	₹74,000,000.00
Kumar Kushagra	₹72,000,000.00



4. ADVANCED/INSIGHTFUL

16. Which team spent the most per player on average?

```
# Team spent the most per player on average
avg_price_per_player = df.groupby(['TEAM']).agg({'PRICE':['mean']}).style.format("₹{:, .2f}")
display (avg_price_per_player)
print(avg_price_per_player)
```



TEAM	PRICE
CSK	₹50,666,666.67
DC	₹21,166,666.67
GT	₹37,875,000.00
KKR	₹31,350,000.00
LSG	₹20,333,333.33
MI	₹20,875,000.00
PK	₹31,187,500.00
RCB	₹34,000,000.00
RR	₹28,600,000.00
SH	₹51,333,333.33

• 17. Create a summary showing the number of players by type for each team.

```
# The number of players by type for each team.
```

```
player_count_by_type_and_team = df.groupby(['TEAM', 'TYPE']).size().reset_index(name='Player Count')
print(player_count_by_type_and_team)
```



TEAM	TYPE	Player Count
CSK	All-Rounder	3
CSK	Batter	1
CSK	Bowler	1
CSK	Wicket-Keeper	1
DC	All-Rounder	1
DC	Batter	2
DC	Bowler	2
DC	Wicket-Keeper	4
GT	All-Rounder	2
GT	Bowler	5
GT	Wicket-Keeper	1
KKR	All-Rounder	1
KKR	Batter	3
KKR	Bowler	5
KKR	Wicket-Keeper	1
LSG	All-Rounder	3
LSG	Batter	1
LSG	Bowler	2
MI	All-Rounder	5
MI	Bowler	3
PK	All-Rounder	5
PK	Batter	2
PK	Bowler	1
RCB	All-Rounder	2
RCB	Batter	1
RCB	Bowler	3



18. Calculate the percentage of overseas players in each team.

```
# The percentage of overseas players in each team.  
def calculate_overseas_percentage(df):  
  
    # Calculate the percentage of overseas players for each team  
    overseas_percentage = (  
        df.groupby('TEAM').apply(lambda x: len(x[x['NATIONALITY'] != 'Indian']) / len(x) * 100)  
    )  
    return overseas_percentage.map("{:, .2f}".format)  
  
result = calculate_overseas_percentage(df)  
print(result)
```

TEAM	
CSK	50.00
DC	44.44
GT	25.00
KKR	40.00
LSG	33.33
MI	50.00
PK	25.00
RCB	50.00
RR	60.00
SH	50.00

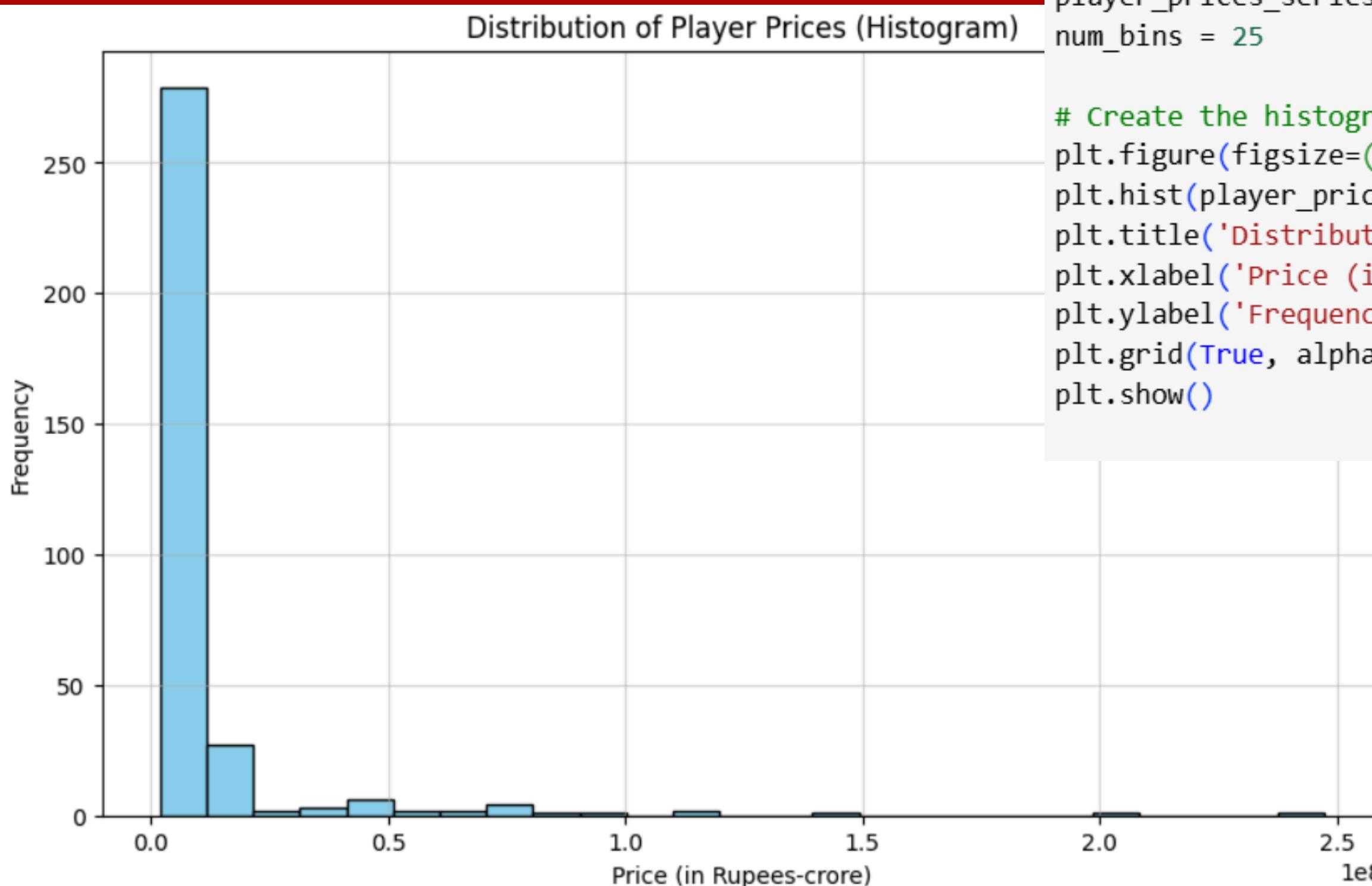
💡 19. Who is the cheapest overseas player?

```
# The cheapest overseas player
overseas_players = df[(df['NATIONALITY'] != 'Indian') & (df['TEAM'] != 'Un-Sold')]
cheapest_overseas_player = overseas_players.loc[overseas_players['PRICE'].idxmin()]
print(cheapest_overseas_player)
```

S_No	59
PLAYERS	Tom Kohler-Cadmore
NATIONALITY	Overseas
TYPE	Wicket-Keeper
PRICE	4000000
TEAM	RR



20. Visualize the distribution of player prices using a box plot or histogram.



```
#20.Distribution of player prices using a box plot or histogram.  
player_prices = df['PRICE']  
  
player_prices_series = pd.Series(player_prices)  
num_bins = 25  
  
# Create the histogram  
plt.figure(figsize=(10, 6))  
plt.hist(player_prices, bins=num_bins, color='skyblue', edgecolor='black')  
plt.title('Distribution of Player Prices (Histogram)')  
plt.xlabel('Price (in Rupees-crore)')  
plt.ylabel('Frequency')  
plt.grid(True, alpha=0.5)  
plt.show()
```

Maximum price range falls between 1-5crore

THANK YOU

