

```
In [2]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
```

Data Load

```
In [198]: df1 = pd.read_csv("IPLPlayerAuctionData.csv")
df1.head()
```

```
Out[198]:
```

	Player	Role	Amount	Team	Year	Player Origin
0	Aaron Finch	Batsman	40000000	Sunrisers Hyderabad	2014.0	Overseas
1	Aaron Finch	Batsman	32000000	Mumbai Indians	2015.0	Overseas
2	Aaron Finch	Batsman	10000000	Gujarat Lions	2016.0	Overseas
3	Aaron Finch	Batsman	62000000	Kings XI Punjab	2018.0	Overseas
4	Aaron Finch	Batsman	44000000	Royal Challengers Bangalore	2020.0	Overseas

```
In [4]: df1.shape
```

```
Out[4]: (970, 6)
```

```
In [5]: df1.columns
```

```
Out[5]: Index(['Player', 'Role', 'Amount', 'Team', 'Year', 'Player Origin'], dtype='object')
```

```
In [6]: df1['Team'].unique()
```

```
Out[6]: array(['Sunrisers Hyderabad', 'Mumbai Indians', 'Gujarat Lions',
        'Kings XI Punjab', 'Royal Challengers Bangalore',
        'Kolkata Knight Riders', 'Gujarat Titans', 'Pune Warriors India',
        'Rajasthan Royals', 'Delhi Daredevils', 'Chennai Super Kings',
        'Rising Pune Supergiant', 'Delhi Capitals', 'Lucknow Super Giants',
        'Punjab Kings'], dtype=object)
```

```
In [86]: dfT=df1['Team'].value_counts()
dfT.head()
```

```
Out[86]: Team
Royal Challengers Bangalore    115
Sunrisers Hyderabad          112
Mumbai Indians                107
Kolkata Knight Riders         107
Rajasthan Royals              100
Name: count, dtype: int64
```

DATA CLEANING

```
In [8]: df1.isnull().sum()
```

```
Out[8]: Player      0
Role      0
Amount     0
Team      0
Year      1
Player Origin  0
dtype: int64
```

```
In [9]: df1.shape
```

```
Out[9]: (970, 6)
```

```
In [199... df2 = df1.dropna()
df2.isnull().sum()
```

```
Out[199]: Player      0
Role      0
Amount     0
Team      0
Year      0
Player Origin  0
dtype: int64
```

```
In [27]: df2.shape
```

```
Out[27]: (969, 6)
```

```
In [12]: df2.loc[30]
```

```
Out[12]: Player      Akash Deep
Role      Bowler
Amount      2000000
Team      Royal Challengers Bangalore
Year      2022.0
Player Origin      Indian
Name: 30, dtype: object
```

```
In [13]: df2['Year'].unique()
```

```
Out[13]: array([2014., 2015., 2016., 2018., 2020., 2022., 2013., 2021., 2017.,
2019.])
```

```
In [14]: df2[(df2['Year']==2014) & (df2['Team']=='Kolkata Knight Riders')]['Player']
```

```

Out[14]: 57      Andre Russell
        175      Chris Lynn
        219      Debabrata Das
        325      Jacques Kallis
        445      Kuldeep Yadav
        493      Manish Pandey
        503      Manvinder Bisla
        570      Morne Morkel
        634      Pat Cummins
        647      Piyush Chawla
        716      Ranganath Vinay Kumar
        744      Robin Uthappa
        757      Ryan Ten Doeschate
        795      Sayan Sekhar Mandal
        808      Shakib Al Hasan
        868      Suryakumar Yadav
        912      Umesh Yadav
        927      Veer Pratap Singh
        957      Yusuf Pathan
Name: Player, dtype: object

```

Feature Engineering

Creating new feature

```

In [15]: def is_year_in_range(year):
        return 1 if( 2014 <= year <= 2020) else 0

```

```

In [200... df5 = df2.copy()

```

```

In [65]: df5['is_in_range'] = df5['Year'].apply(is_year_in_range)

# Use the 'is_in_range' column to conditionally calculate 'price_per_match'
df5['amount_per_match'] = df5.apply(lambda row: row['Amount'] / 14 if row['is_in_range'] else 0, axis=1)
df5['amount_per_match'] = df5['amount_per_match'].round(2)
# Drop the temporary 'is_in_range' column if not needed
df5 = df5.drop(columns=['is_in_range'])

df5.head()

```

Out[65]:

	Player	Role	Amount	Team	Year	Player Origin	amount_per_match
0	Aaron Finch	Batsman	40000000	Sunrisers Hyderabad	2014.0	Overseas	2857142.86
1	Aaron Finch	Batsman	32000000	Mumbai Indians	2015.0	Overseas	2285714.29
2	Aaron Finch	Batsman	10000000	Gujarat Lions	2016.0	Overseas	714285.71
3	Aaron Finch	Batsman	62000000	Kings XI Punjab	2018.0	Overseas	4428571.43
4	Aaron Finch	Batsman	44000000	Royal Challengers Bangalore	2020.0	Overseas	3142857.14

```
In [66]: df5_stats = df5['amount_per_match'].describe().round(2)
df5_stats
```

```
Out[66]: count      969.00
mean     1432856.96
std      1884754.37
min       71428.57
25%      142857.14
50%      625000.00
75%      2000000.00
max     11428571.43
Name: amount_per_match, dtype: float64
```

I have done some wrong inputs in Amount lets found them

```
In [22]: def is_float(x):
        try:
            float(x)
        except:
            return False
        return True
```

```
In [77]: df6=df5[~df5['Amount'].apply(is_float)]
df6.head()
```

```
Out[77]:
```

	Player	Role	Amount	Team	Year	Player Origin
5	Abdul Samad	All-Rounder	20L	Sunrisers Hyderabad	2020.0	Indian
14	Abu Nechim Ahmed	Bowler	30L	Royal Challengers Bangalore	2014.0	Indian
67	Anirudha Ashok Joshi	All-Rounder	20L	Royal Challengers Bangalore	2018.0	Indian
100	Ashish Nehra	Bowler	2Cr	Chennai Super Kings	2014.0	Indian
112	Axar Rajesh Patel	All-Rounder	5Cr	Delhi Capitals	2019.0	Indian

```
In [78]: dfA=df6['Amount']
dfA.head()
```

```
Out[78]: 5      20L
14      30L
67      20L
100     2Cr
112     5Cr
Name: Amount, dtype: object
```

```
In [79]: import re
def convert_amount(amount_str):
    if re.match(r'^\d+Cr$', amount_str):
        return int(amount_str[:-2]) * 10000000 # Convert Cr to 10 million
    elif re.match(r'^\d+L$', amount_str):
        return int(amount_str[:-1]) * 100000 # Convert L to 100 thousand
    elif amount_str.isdigit():
        return int(amount_str)
    else:
        # Handle other cases as needed
        print(f"Unrecognized format: {amount_str}")
        return None
```

```
In [80]: df5['Amount'] = df5['Amount'].apply(lambda x: convert_amount(x) if x else x)
```

```
In [85]: result_df = pd.merge(df5, df6, on='Player', how='inner')

# Display the result
result_df.head()
```

```
Out[85]:
```

	Player	Role_x	Amount_x	Team_x	Year_x	Player Origin_x	Role_y	Amount_y	Team_y	Year_
0	Abdul Samad	All-Rounder	2000000	Sunrisers Hyderabad	2020.0	Indian	All-Rounder	20L	Sunrisers Hyderabad	2020
1	Abu Nechim Ahmed	Bowler	3000000	Royal Challengers Bangalore	2014.0	Indian	Bowler	30L	Royal Challengers Bangalore	2014
2	Anirudha Ashok Joshi	All-Rounder	2000000	Royal Challengers Bangalore	2018.0	Indian	All-Rounder	20L	Royal Challengers Bangalore	2018
3	Anirudha Ashok Joshi	All-Rounder	2000000	Rajasthan Royals	2020.0	Indian	All-Rounder	20L	Royal Challengers Bangalore	2018
4	Ashish Nehra	Bowler	20000000	Chennai Super Kings	2014.0	Indian	Bowler	2Cr	Chennai Super Kings	2014

```
In [19]: df5.Player.unique()
```

```
Out[19]: array(['Aaron Finch', 'Abdul Samad', 'Abhijeet Tomar', 'Abhimanyu Mithun',  
              'Abhinav Sadarangani', 'Abhishek Nayar', 'Abhishek Sharma',  
              'Abu Nechim Ahmed', 'Adam Milne', 'Adam Zampa', 'Aditya Garhwal',  
              'Aditya Tare', 'Agnivesh Ayachi', 'Aiden Blizard',  
              'Aiden Markram', 'Ajantha Mendis', 'Ajinkya Rahane', 'Akash Deep',  
              'Akash Singh', 'Akhil Arvind Herwadkar', 'Akila Dananjaya',  
              'Akila Dhananjaya', 'Akshar Rajesh Patel', 'Akshay Karnewar',  
              'Akshay Wakhare', 'Akshdeep Nath', 'Albie Morkel', 'Alex Carey',  
              'Alex Hales', 'Alzarri Joseph', 'Aman Khan', 'Ambati Rayudu',  
              'Amit Mishra', 'Amit Paunekar', 'Andre Russell', 'Andrew Tye',  
              'Aneeshwar Gautam', 'Angelo Mathews', 'Aniket Choudhary',  
              'Anirudha Ashok Joshi', 'Ankeet Bawane', 'Ankit Nagendra Sharma',  
              'Ankit Sharma', 'Ankit Singh Rajpoot', 'Ankush Bains',  
              'Anmolpreet Singh', 'Anrich Nortje', 'Ansh Patel', 'Anuj Rawat',  
              'Anukul Roy', 'Anunay Singh', 'Anureet Singh',  
              'Apoorv Vijay Wankhade', 'Arjun Tendulkar', 'Armaan Jaffer',  
              'Arshdeep Singh', 'Aryaman Vikram Birla', 'Aryan Juyal',  
              'Asela Gunaratna', 'Ashish Nehra', 'Ashish Reddy', 'Ashok Dinda',  
              'Ashok Sharma', 'Ashton Turner', 'Ashwin Hebbur', 'Asif K M',  
              'Atharva Taide', 'Avesh Khan', 'Axar Rajesh Patel', 'Ayush Badoni',  
              'B. Sai Sudharsan', 'Baba Aparajith', 'Baba Indrajith',  
              'Baltej Dhanda', 'Bandaru Ayyappa', 'Barinder Singh Saran',  
              'Barinder Singh Sran', 'Basil Thampi', 'Ben Cutting', 'Ben Dunk',  
              'Ben Dwarshuis', 'Ben Hilfenhaus', 'Ben Laughlin',  
              'Benjamin Stokes', 'Benny Howell', 'Beuran Hendricks',  
              'Bhanuka Rajapaksa', 'Bhuvneshwar Kumar', 'Billy Stanlake',  
              'Bipul Sharma', 'Brad Hodge', 'Brad Hogg', 'Brendan Taylor',  
              'Brendon McCullum', 'C Hari Nishaanth', 'C.Hari Nishaanth',  
              'C.M. Gautam', 'Cameron Delport', 'Carlos Brathwaite',  
              'Chaitanya Bishnoi', 'Chama Milind', 'Chamika Karunaratne',  
              'Chetan Sakariya', 'Cheteshwar Pujara', 'Chirag Suri',  
              'Chris Green', 'Chris Jordan', 'Chris Lynn', 'Chris Woakes',  
              'Christopher Barnwell', 'Christopher Gayle', 'Christopher Morris',  
              'Clinton McKay', 'Colin De Grandhomme', 'Colin Ingram',  
              'Colin Munro', 'Corey Anderson', 'Dale Steyn', 'Dan Christian',  
              'Daniel Christian', 'Daniel Sams', 'Darcy Short', 'Darren Bravo',  
              'Darren Sammy', 'Darshan Nalkande', 'Daryl Mitchell',  
              'David Miller', 'David Warner', 'David Wiese', 'David Willey',  
              'Dawid Malan', 'Debabrata Das', 'Deepak Chahar', 'Deepak Hooda',  
              'Deepak Punia', 'Devdutt Padikkal', 'Devon Conway',  
              'Dewald Brevis', 'Dhawal Kulkarni', 'Dhruv Jurel', 'Dhruv Shorey',  
              'Digvijay Deshmukh', 'Dinesh Karthik', 'Dinesh Salunkhe',  
              'Dirk Nannes', 'Dishant Yagnik', 'Dominic Drakes',  
              'Domnic Joseph Muthuswamy', 'Dushmanta Chameera',  
              'Dushmantha Chameera', 'Dwayne Pretorius', 'Dwayne Bravo',  
              'Dwayne Smith', 'Eklavya Dwivedi', 'Eoin Morgan', 'Evin Lewis',  
              'Fabian Allen', 'Faf Du Plessis', 'Farhaan Behardien',  
              'Fazalhaq Farooqi', 'Fidel Edwards', 'Finn Allen',  
              'Gautam Gambhir', 'George Bailey', 'Glenn Maxwell',  
              'Glenn Phillips', 'Gurinder Sandhu', 'Gurkeerat Singh Mann',  
              'Hanuma Vihari', 'Harbhajan Singh', 'Hardik Pandya',  
              'Hardus Viljoen', 'Harpreet Brar', 'Harry Gurney', 'Harshal Patel',  
              'Heinrich Klaasen', 'Himmat Singh', 'Hrithik Shokeen',  
              'HS Sharath', 'Imran Tahir', 'Iqbal Abdullah', 'Irfan Pathan',  
              'Ishan Kishan', 'Ishan Porel', 'Ishank Jaggi', 'Ishant Sharma',  
              'Ishwar Chandra Pandey', 'Isuru Udana', 'J Suchith', 'Jacob Oram',  
              'Jacques Kallis', 'Jagadeesha Suchith', 'Jalaj Saxena',  
              'James Faulkner', 'James Neesham', 'Jaskaran Singh',  
              'Jason Behrendorff', 'Jason Holder', 'Jason Roy', 'Jasprit Bumrah',  
              'Jatin Saxena', 'Javon Searless', 'Jayant Yadav', 'Jaydev Shah',
```

'Jaydev Unadkat', 'Jean-Paul Duminy', 'Jeevan Mendis',
'Jesse Ryder', 'Jhye Richardson', 'Jitesh Sharma', 'Joe Denly',
'Joel Paris', 'Jofra Archer', 'Johan Botha', 'John Hastings',
'Jonny Bairstow', 'Jos Buttler', 'Josh Hazlewood',
'Joshua Philippe', 'Juan Theron', 'Krishnappa Gowtham',
'K.Bhagath Varma', 'K.C Cariappa', 'K.C. Cariappa', 'K.K. Jiyaz',
'K.M. Asif', 'K.S. Bharat', 'Kagiso Rabada', 'Kamlesh Nagarkoti',
'Kane Richardson', 'Kane Williamson', 'Kanishk Seth',
'Karan Sharma', 'Karanveer Singh', 'Karn Sharma', 'Kartik Tyagi',
'Karun Nair', 'Kedar Jadhav', 'Keemo Paul', 'Kevin Pietersen',
'Kevon Cooper', 'Kieron Pollard', 'Kishore Pramod Kamath',
'KL Rahul', 'Kona Srikar Bharat', 'Krismar Santokie',
'Krunal Pandya', 'Kshitiz Sharma', 'Kuldeep Sen', 'Kuldeep Yadav',
'Kuldip Yadav', 'Kulwant Khejroliya', 'Kusal Janith Perera',
'Kyle Abbott', 'Kyle Jamieson', 'Kyle Mayers',
'Lakshmipathy Balaji', 'Lalit Yadav', 'Lasith Malinga',
'Laxmi Ratan Shukla', 'Liam Livingstone', 'Lockie Ferguson',
'Luke Pomersbach', 'Lukman Hussain Meriwala', 'Lungisani Ngidi',
'Luvnith Sisodia', 'M Siddharth', 'M. Ashwin',
'M. Harisankar Reddy', 'Washington Sundar', 'Maheesh Theekshana',
'Mahipal Lomror', 'Manan Ajay Sharma', 'Manan Vohra',
'Mandeep Hardev Singh', 'Mandeep Singh', 'Manish Pandey',
'Manjot Kalra', 'Manoj Tiwary', 'Manpreet Gony', 'Manprit Juneja',
'Manvinder Bisla', 'Manzoor Dar', 'Marchant De Lange',
'Marco Jansen', 'Marcus Stoinis', 'Mark Wood', 'Martin Guptill',
'Matt Henry', 'Matthew Wade', 'Mayank Agarwal', 'Mayank Dagar',
'Mayank Markande', 'Mayank Yadav', 'Michael Clarke',
'Michael Hussey', 'Midhun S', 'Milind Kumar', 'Milind Tandon',
'Mitchell Johnson', 'Mitchell Marsh', 'Mitchell McClenaghan',
'Mitchell Santner', 'Mitchell Starc', 'Mithun Manhas', 'Moeen Ali',
'Mohammad Nabi', 'Mohammad Shami', 'Mohammed Azharudeen',
'Mohammed Siraj', 'Mohd. Arshad Khan', 'Mohit Sharma',
'Mohsin Khan', 'Moises Henriques', 'Monu Singh', 'Morne Morkel',
'Mujeeb Zadran', 'Mukesh Choudhary', 'Munaf Patel',
'Murali Kartik', 'Murali Vijay', 'Murugan Ashwin',
'Mustafizur Rahman', 'Muttiah Muralitharan', 'N Jagadeesan',
'N. Jagadeesan', 'N. Tilak Varma', 'Naman Ojha',
'Nathan Coulter-Nile', 'Nathan Ellis', 'Nathan McCullum',
'Nathu Singh', 'Navdeep Saini', 'Nic Maddinson', 'Nicolas Pooran',
'Nidheesh M D Dinesan', 'Nikhil Shankar Naik', 'Nitish Rana',
'Noor Ahmad', 'Obed Mccoy', 'Odean Smith', 'Oshane Thomas',
'Pankaj Jaswal', 'Pankaj Singh', 'Paras Dogra', 'Pardeep Sahu',
'Parthiv Patel', 'Parveez Rasool', 'Parvinder Awana',
'Pat Cummins', 'Pavan Deshpande', 'Pawan Negi', 'Pawan Suyal',
'Peter Handscomb', 'Philip Hughes', 'Piyush Chawla',
'Prabhsimran Singh', 'Pradeep Sangwan', 'Pragyan Ojha',
'Prasanth Padmanabhan', 'Prasanth Parameswaran', 'Prashant Chopra',
'Prashant Solanki', 'Prasidh Krishna', 'Pratham Singh',
'Pratyush Singh', 'Praveen Dubey', 'Praveen Kumar', 'Pravin Dubey',
'Pravin Tambe', 'Prayas Ray Barman', 'Prerak Mankad',
'Prince Balwant Rai Singh', 'Prithvi Raj Yarra', 'Prithvi Shaw',
'Priyam Garg', 'Quinton De Kock', 'R Samarth', 'R. Ashwin',
'R. Sai Kishore', 'R. Sanjay Yadav', 'Rahul Ajay Tripathi',
'Rahul Buddhi', 'Rahul Chahar', 'Rahul Sharma', 'Rahul Shukla',
'Rahul Tewatia', 'Rahul Tripathi', 'Raj Angad Bawa',
'Rajagopal Sathish', 'Rajat Bhatia', 'Rajat Patidar',
'Rajvardhan Hangargekar', 'Ramandeep Singh', 'Ramesh Kumar',
'Ranganath Vinay Kumar', 'Rashid Khan Arman', 'Rasikh Dar',
'Rassie Van Der Dussen', 'Ravi Bishnoi', 'Ravi Bopara',
'Ravi Rampaul', 'Ravichandran Ashwin', 'Ricky Bhui',

```
'Ricky Ponting', 'Riley Meredith', 'Rinku Singh', 'Ripal Patel',
'Rishabh Pant', 'Rishi Dhawan', 'Riyan Parag', 'Robin Uthappa',
'Romario Shephard', 'Ronit More', 'Ross Taylor', 'Rovman Powell',
'Rudra Pratap Singh', 'Ruturaj Gaikwad', 'Ryan McLaren',
'Ryan Ten Doeschate', 'Sachin Baby', 'Sachin Rana',
'Sachithra Senanayaka', 'Sagar Trivedi', 'Sam Billings',
'Sam Curran', 'Samuel Badree', 'Sandeep Bavanaka',
'Sandeep Lamichhane', 'Sandeep Sharma', 'Sandeep Warrier',
'Sanjay Yadav', 'Sanju Samson', 'Sarabjit Ladda', 'Sarfaraz Khan',
'Sarfaraz Naushad Khan', 'Saurabh Dubey', 'Saurabh Kumar',
'Saurabh Tiwary', 'Sayan Ghosh', 'Sayan Sekhar Mandal',
'Scott Boland', 'Sean Abbott', 'Shadab Jakati', 'Shahbaz Ahamad',
'Shahbaz Nadeem', 'Shahrukh Khan', 'Shakib Al Hasan',
'Shane Watson', 'Sharad Lumba', 'Shardul Thakur', 'Shashank Singh',
'Shaun Marsh', 'Sheldon Cottrell', 'Sheldon Jackson',
'Shelley Shaurya', 'Sherfane Rutherford', 'Shikhar Dhawan',
'Shimron Hetmyer', 'Shishir Bhavane', 'Shivam Dube', 'Shivam Mavi',
'Shivam Sharma', 'Shivil Kaushik', 'Shreevats Goswami',
'Shreyas Gopal', 'Shreyas Iyer', 'Shrikant Mundhe',
'Shubam Agrawal', 'Shubham Garhwal', 'Shubham Ranjane',
'Shubman Gill', 'Siddarth Kaul', 'Siddharth Kaul',
'Siddhesh Dinesh Lad', 'Simarjeet Singh', 'Srikanth Anirudha',
'Steven Smith', 'Stuart Binny', 'Subhanshu Senapati',
'Subramaniam Badrinath', 'Sudeep Tyagi', 'Sumit Narwal',
'Suryakumar Yadav', 'Sushant Marathe', 'Suyash Prabhudesai',
'Suyash Prabhudessai', 'Swapnil Singh', 'Syed Khaleel Ahmed',
'Syed Mehdi Hasan', 'T Natarajan', 'T. Natarajan',
'Tajinder Dhillon', 'Tanmay Agarwal', 'Tanmay Mishra',
'Tejas Baroka', 'Tejas Singh Baroka', 'Thisara Perera',
'Tim David', 'Tim Seifert', 'Tim Southee', 'Tirumalasetti Suman',
'Tom Banton', 'Tom Curran', 'Travis Head', 'Trent Boult',
'Tushar Deshpande', 'Tymal Mills', 'Umang Sharma', 'Umesh Yadav',
'Unmukt Chand', 'Utkarsh Singh', 'Vaibhav Arora', 'Vaibhav Rawal',
'Varun Aaron', 'Varun Chakaravarthy', 'Veer Pratap Singh',
'Venkatesh Iyer', 'Venugopal Rao', 'Vicky Ostwal', 'Vijay Shankar',
'Vijay Zol', 'Vikas Tokas', 'Vikramjeet Malik', 'Virat Singh',
'Virender Sehwag', 'Vishnu Vinod', 'Wanindu Hasaranga',
'Wayne Parnell', 'Wriddhiman Saha', 'Writtick Chatterjee',
'Yash Dayal', 'Yash Dhull', 'Yashasvi Jaiswal', 'Yogesh Gowalkar',
'Yogesh Takawale', 'Yudhvir Charak', 'Yusuf Pathan',
'Yuvraj Singh', 'Yuzvendra Chahal', 'Yuzvendra Singh Chahal',
'Zaheer Khan', 'Zahir Khan Pakteen'], dtype=object)
```

```
In [91]: dfT.head()
```

```
Out[91]: Team
Royal Challengers Bangalore    115
Sunrisers Hyderabad           112
Mumbai Indians                 107
Kolkata Knight Riders          107
Rajasthan Royals               100
Name: count, dtype: int64
```

```
In [119... dfTA=df1.groupby('Team')['Amount'].sum()
dfTA.head()
```



```
Out[119]: Team
Chennai Super Kings    1570875000
Delhi Capitals          953000000
Delhi Daredevils       1945500000
Gujarat Lions          205000000
Gujarat Titans         518500000
Name: Amount, dtype: int64
```

```
In [118... len(dfTA)
```

```
Out[118]: 15
```

```
In [121... len(dfTA[dfTA>200000000])
```

```
Out[121]: 4
```

Dimensionality Reduction

```
In [122... df5.head()
```

```
Out[122]:
```

	Player	Role	Amount	Team	Year	Player Origin
0	Aaron Finch	Batsman	40000000	Sunrisers Hyderabad	2014.0	Overseas
1	Aaron Finch	Batsman	32000000	Mumbai Indians	2015.0	Overseas
2	Aaron Finch	Batsman	10000000	Gujarat Lions	2016.0	Overseas
3	Aaron Finch	Batsman	62000000	Kings XI Punjab	2018.0	Overseas
4	Aaron Finch	Batsman	44000000	Royal Challengers Bangalore	2020.0	Overseas

```
In [124... df5.shape
```

```
Out[124]: (969, 6)
```

```
In [201... df100=df5[df5['Amount']>=10000000]
df100.shape
```

```
Out[201]: (483, 6)
```

```
In [202... len(df100.Player.unique())
```

```
Out[202]: 268
```

Outlier Removal

```
In [203... df100stats=df100['Amount'].describe().round(2)
```

```
In [204... df100stats.head()
```

```
Out[204]: count      483.00
mean    39105175.98
std     30352904.54
min     10000000.00
25%     18500000.00
Name: Amount, dtype: float64
```

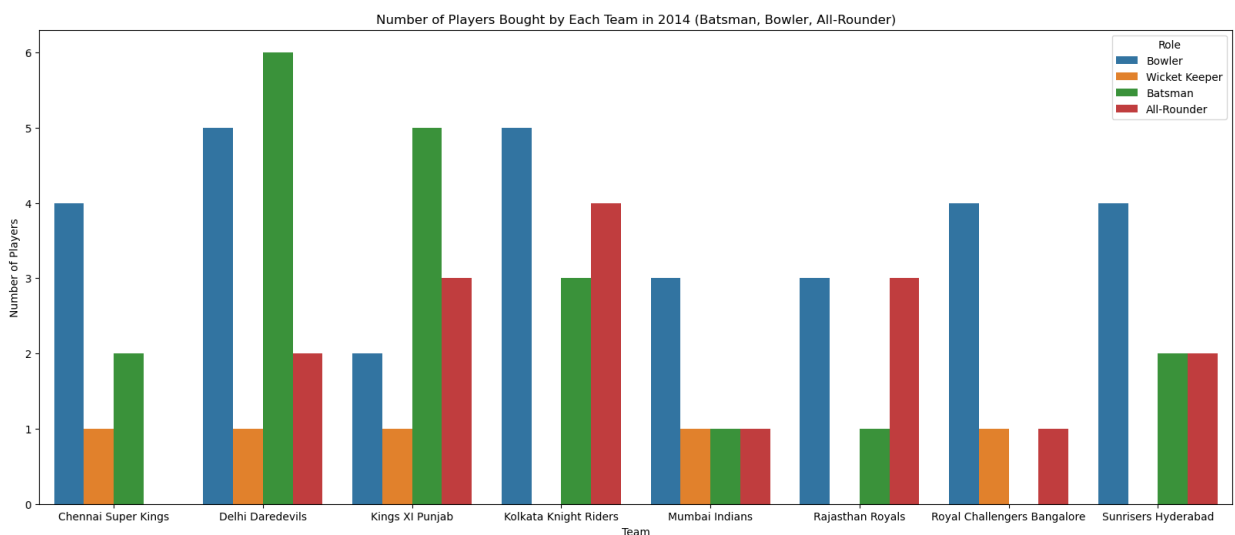
```
In [205... def remove_pps_outliers(df):
df_out = pd.DataFrame()
for key, subdf in df.groupby('Team'):
    m = np.mean(subdf.Amount)
    st = np.std(subdf.Amount)
    reduced_df = subdf[(subdf.Amount>(m-st)) & (subdf.Amount<=(m+st))]
    df_out = pd.concat([df_out,reduced_df],ignore_index=True)
return df_out
df7 = remove_pps_outliers(df100)
df7.shape
```

```
Out[205]: (387, 6)
```

Data Visulaization

```
In [159... import seaborn as sns
df_2014 = df7[df7['Year'] == 2014]

# Plot the histogram
plt.figure(figsize=(20, 8))
sns.countplot(x='Team', hue='Role', data=df_2014)
plt.title('Number of Players Bought by Each Team in 2014 (Batsman, Bowler, All-Rounder)')
plt.xlabel('Team')
plt.ylabel('Number of Players')
plt.show()
```



```
In [166... def plot_pie_chart(df, team, year):
# Filter data for the specified team and year
df_filtered = df[(df['Team'] == team) & (df['Year'] == year)]

if df_filtered.empty:
    print(f"No data available for {team} in {year}")
    return
```

```

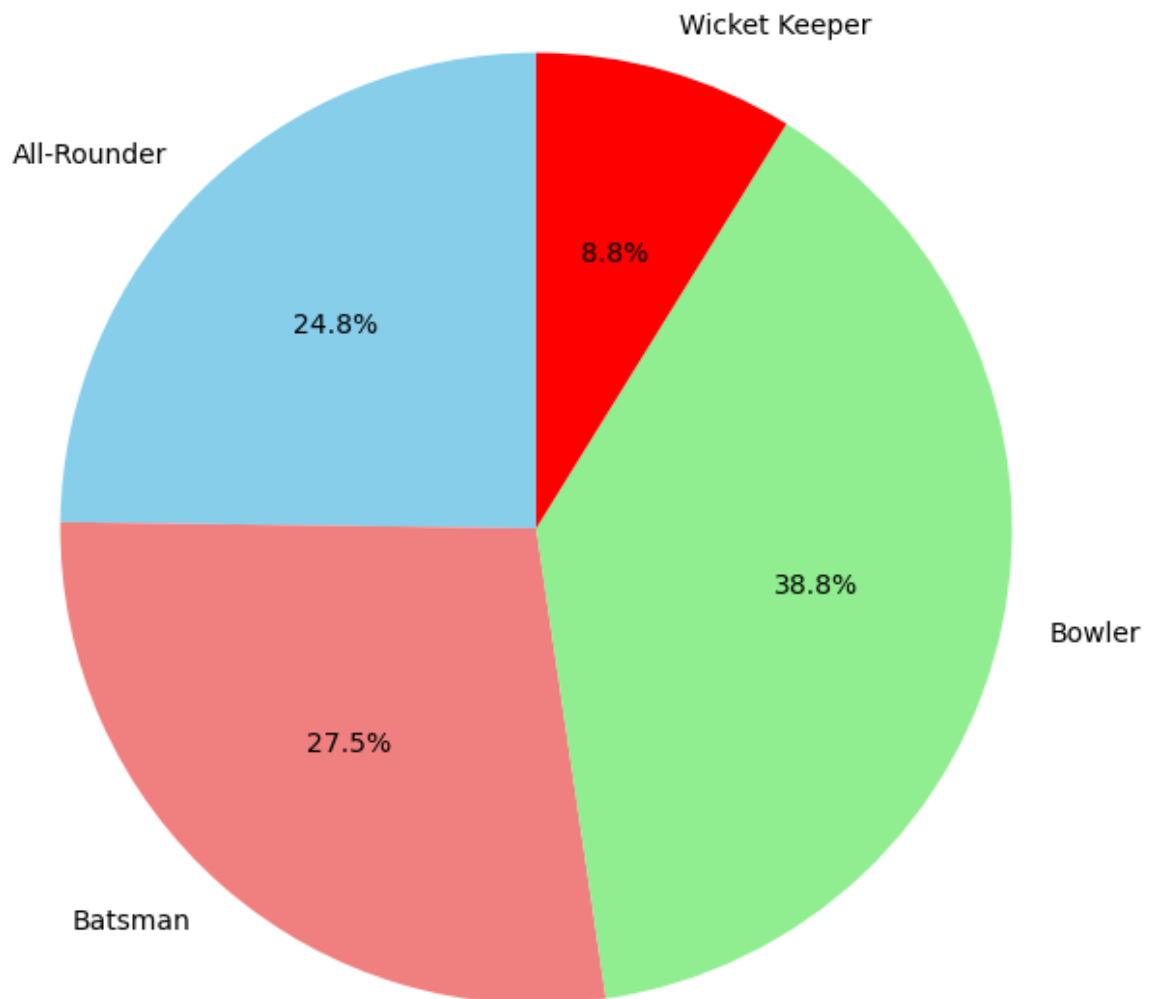
role_amounts = df_filtered.groupby('Role')['Amount'].sum()

# Plot a pie chart
plt.figure(figsize=(8, 8))
plt.pie(role_amounts, labels=role_amounts.index, autopct='%1.1f%%', startangle=90,
plt.title(f'Sum of Amounts Distribution in {team} for {year}')
plt.show()

plot_pie_chart(df7, 'Mumbai Indians', 2014)

```

Sum of Amounts Distribution in Mumbai Indians for 2014



In [169...

```

import pandas as pd
import matplotlib.pyplot as plt

def scatter_plot(df, team, year):
    # Filter data for the specified team and year
    df_filtered = df[(df['Team'] == team) & (df['Year'] == year)]

    if df_filtered.empty:

```

```

print(f"No data available for {team} in {year}")
return

# Group by player and calculate total amount
player_amounts = df_filtered.groupby('Player')['Amount'].sum()

# Get the number of players
num_players = player_amounts.count()

# Plot a scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(range(1, num_players + 1), player_amounts.values, color='blue')
plt.title(f'Scatter Plot of Amount vs. Number of Players for {team} in {year}')
plt.xlabel('Number of Players')
plt.ylabel('Total Amount')
plt.show()

scatter_plot(df5, 'Mumbai Indians', 2014)

```

