A
Training Report on
Formation of

NEW IPL TEAM

In
Data Science using Python

Submitted To:

Mr Rishu Trivadi

Submitted By:

Garima Kundnani

Problem Statement

You have been hired as a data analyst by a sports management company. They are interested in forming a new team for the upcoming IPL Season 2024 and want your expertise to suggest players that will maximise their chances of winning matches. Your task is to analyse the IPL dataset and recommend the top-performing players in various positions to include in the new team.

Process

Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Load Data

```
df=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowling
df2=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
df3=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
df4=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
df5=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
df6=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
df7=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Bowlin
```

```
df8=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Battin
df9=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Battin
df10=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Batti
df11=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Batti
df12=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Batti
df13=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Batti
df14=pd.read_csv(r"/Users/Garima/Desktop/Python/Project-4/IPL Player Stats/Batti
```

Combining DataFrames

```
df15=pd.concat([df,df2,df3,df4,df5,df6,df7], axis=0, ignore_index=False)
print("Bowling:")
print(df15)
```

```
df16=pd.concat([df8,df9,df10,df11,df12,df13,df14], axis=0, ignore_index=False)
print("Batting:")
print(df16)
```

Extracting Detailed Information

```
print("First rows of Bowling:")
print(df15.head())
print("First rows of Batting:")
print(df16.head())
```

Perform Analysis on the datasets

```
#Check the dimensions of the dataset
print("Formating of Bowlers dataset:")
print(df15.describe())
print(df15.info())
print("Formatting of Batting dataset:")
print(df16.describe())
print(df16.info())

#Identify the different variables/columns available in the dataset
print("The different variables/columns available in the bowling dataset are")
print(df15.columns)
print("The different variables/columns available in the batting dataset are")
print(df16.columns)

#Handle missing values appropriately
print("Missing values in Bowling:")
print(df15.isnull())
print("Missing values in Batting:")
print(df16.isnull())
print("Batting:")
print(df16.corr()['Runs'].sort_values(ascending=False))
print("Bowling:")
print(df15.corr()['Runs'].sort_values(ascending=False))
```

Generating Charts

```
#Total 50s scored by a single player
d=df16.sort_values(by="50",ascending=False)[:20].copy()
d["custom_label"]=d["Player"]+"\n"+d["50"].astype(str)
plt.pie(d["50"],labels=d["custom_label"],radius=1.4, labeldistance=1.1,pctdistan
plt.title("Total 50s scored by a single player")
plt.show()
#Total 100s
e=df16.sort_values(by="100",ascending=False)[:20].copy()
e["custom_label"]=e["Player"]+"\n"+e["100"].astype(str)
plt.pie(d["100"],labels=e["custom_label"],radius=1.4, labeldistance=1.1,pctdista
plt.title("Total 100s scored by a single player")
plt.show()
```

Creating rating system

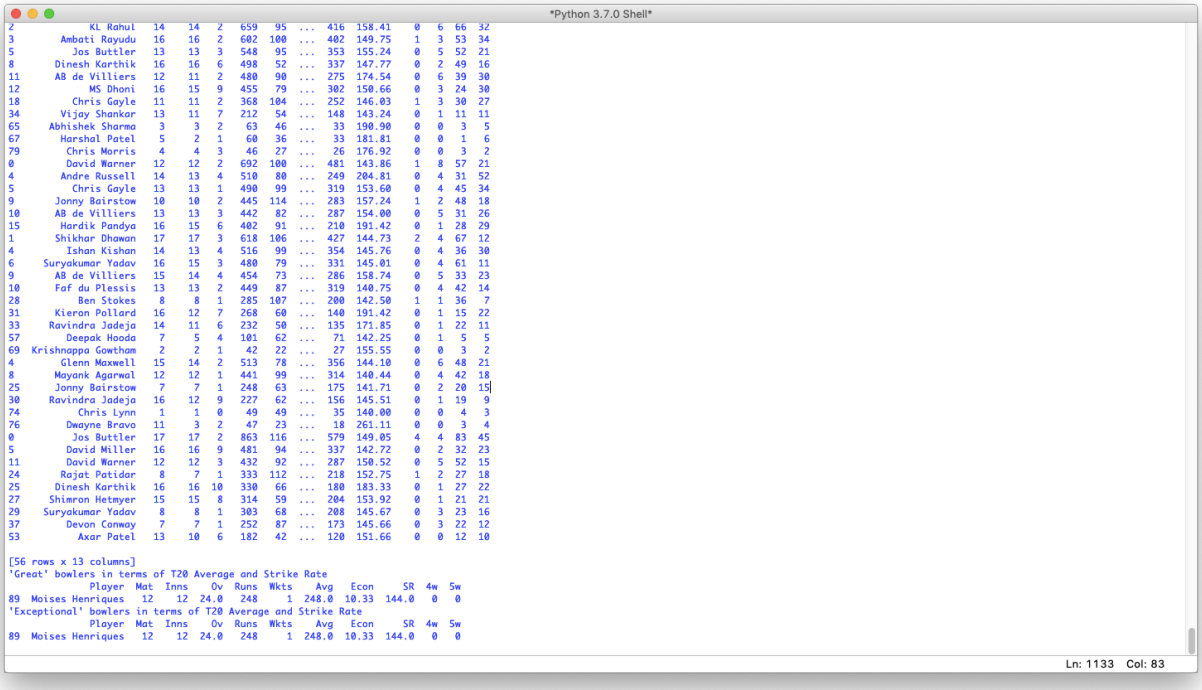
```
#Histograms for Batsman
plt.hist(df16[["Runs"]])
plt.title("Runs Scored by Batsman")
plt.show()

plt.hist(df16[["SR"]])
plt.title("Strike Rate of  Batsman")
plt.show()

plt.hist(df16[["50"]])
plt.title("No of the times 50 scored by Batsman")
plt.show()

plt.hist(df16[["100"]])
plt.title("No of the times 100 scored by Batsman")
plt.show()

plt.hist(df16[["4s"]])
plt.title("Total Fours Scored by Batsman")
plt.show()

plt.hist(df16[["6s"]])
plt.title("Total Sixes Scored by Batsman")
plt.show()
```

```
#Histgram Bowlers
plt.hist(df15[["Wkts"]])
plt.title("Total Wickets taken Bowlers")
plt.show()

plt.hist(df15[["SR"]])
plt.title("Strike Rate Of Bowlers")
plt.show()

plt.hist(df15[["Econ"]])
plt.title(" Bowlers Economy")
plt.show()

plt.hist(df15[["4w"]])
plt.title("4 wickets haul of Bowlers")
plt.show()

plt.hist(df15[["5w"]])
plt.title("5 wickets haul of Bowlers")
plt.show()
```

Target for Aspiring Players

```
#Batsman
df_top_25_batsmen = df16.sort_values(by='Runs', ascending=False)[:25].copy()
df_top_25_batsmen["Runs_In_Boundaries"] =  (df_top_25_batsmen["4s"] * 4) + (df_t
df_top_25_batsmen["Boundary_Percentage"] = round((df_top_25_batsmen["Runs_In_Bou
df_top_25_batsmen["Balls_Per_Innings"] = round(df_top_25_batsmen["BF"] / df_top_
print("The Ideal Percent of total runs scored in boundaries for an aspiring play
print("The Ideal Strike Rate a T20 Batsman should aim for is: " + str(round(df_t
print("The Ideal Average a T20 Batsman should aim for is: " + str(round(df_top_2
print("The Average Balls Faced per Innings Played for a T20 Batsman is: " + str(
print()
#Bowlers
df_top_25_bowlers = df15.sort_values(by="Wkts", ascending=False)[:25].copy()
df_top_25_bowlers["Overs_Per_Inning"] = round(df_top_25_bowlers["Ov"] / df_top_2
print("A top T20 Bowler only allows: " + str(df_top_25_bowlers["Econ"].mean()) +
print("The Number of Balls it takes for a top T20 Bowler to take a wicket is: "
print("The Amount of Runs allowed per each wicket taken for a top T20 Bowler is:
print("The Average Overs Bowled per Innings Played for a top T20 Bowler is: " +
```

Output



```
========== RESTART: /Users/Garima/Desktop/Python/Project-4/IPL.py ==========
Bowling:
     POS                Player  Mat  Inns    Ov  ...    Avg   Econ     SR  4w  5w
0      1     Bhuvneshwar Kumar   17    17  66.0  ...  21.30   7.42  17.21   1   0
1      2      Yuzvendra Chahal   13    13  49.0  ...  19.09   8.15  14.04   1   0
2      3          Shane Watson   16    16  56.0  ...  24.25   8.58  16.95   1   0
3      4       Dhawal Kulkarni   14    14  49.0  ...  20.22   7.42  16.33   1   0
4      5  Mitchell McClenaghan   14    14  53.0  ...  25.64   8.17  18.82   1   0
..   ...                   ...  ...   ...   ...  ...    ...    ...    ...  ..  ..
98    99         Shreyas Gopal    1     1   3.0  ...  34.00  11.33  18.00   0   0
99   100          Fabian Allen    1     1   4.0  ...  46.00  11.50  24.00   0   0
100  101           Sean Abbott    1     1   4.0  ...  47.00  11.75  24.00   0   0
101  102           Riyan Parag   17     4   4.0  ...  59.00  14.75  24.00   0   0
102  103         Dewald Brevis    7     1   0.3  ...   8.00  16.00   3.00   0   0

[615 rows x 13 columns]
Batting:
     POS              Player  Mat  Inns  NO  ...      SR 100 50  4s  6s
0      1         Virat Kohli   16    16   4  ...  152.03   4  7  83  38
1      2        David Warner   17    17   3  ...  151.42   0  9  88  31
2      3       AB de Villiers   16    16   3  ...  168.79   1  6  57  37
3      4       Gautam Gambhir   15    15   2  ...  121.89   0  5  54   6
4      5       Shikhar Dhawan   17    17   4  ...  116.78   0  4  51   8
..   ...                  ...  ...   ...  ..  ...     ...  .. ..  ..  ..
157  158     Fazalhaq Farooqi    3     1   1  ...   25.00   0  0   0   0
158  159   Jagadeesha Suchith    5     2   0  ...   25.00   0  0   0   0
159  160          Tim Southee    9     5   1  ...   16.66   0  0   0   0
160  161  Nathan Coulter-Nile    1     1   1  ...   50.00   0  0   0   0
161  162        Anrich Nortje    6     1   1  ...   16.66   0  0   0   0

[1005 rows x 14 columns]
First rows of Bowling:
   POS                Player  Mat  Inns    Ov  ...    Avg  Econ     SR  4w  5w
0    1     Bhuvneshwar Kumar   17    17  66.0  ...  21.30  7.42  17.21   1   0
1    2      Yuzvendra Chahal   13    13  49.0  ...  19.09  8.15  14.04   1   0
2    3          Shane Watson   16    16  56.0  ...  24.25  8.58  16.95   1   0
3    4       Dhawal Kulkarni   14    14  49.0  ...  20.22  7.42  16.33   1   0
4    5  Mitchell McClenaghan   14    14  53.0  ...  25.64  8.17  18.82   1   0

[5 rows x 13 columns]
First rows of Batting:
   POS          Player  Mat  Inns  NO  Runs  ...   BF      SR  100  50  4s  6s
0    1     Virat Kohli   16    16   4   973  ...  640  152.03    4   7  83  38
1    2    David Warner   17    17   3   848  ...  560  151.42    0   9  88  31
2    3  AB de Villiers   16    16   3   687  ...  407  168.79    1   6  57  37
3    4  Gautam Gambhir   15    15   2   501  ...  411  121.89    0   5  54   6
4    5  Shikhar Dhawan   17    17   4   501  ...  429  116.78    0   4  51   8

[5 rows x 14 columns]
```