

## Assignment No: 5

Name: Shinde Shruti Anil

Roll No:759

Div: G(G3)

PRN No: 202201040110

Select any one real-life [dataset](#). Perform data analysis. Identify 10 grains for a given [dataset](#). Develop an interactive dashboard using the matplotlib/Seaborn library.

Code :

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

data=pd.read_csv("/content/sample_data/Batting data set.csv",encoding=
'unicode_escape')
#print("all data:-\n",data)
df=pd.DataFrame(data)

#Line chart of Country VS Matches Played
data=data.sort_values(by='Matches Played')
mat_x=data['Matches Played']
con_y=data['Country']
plt.plot(mat_x,con_y)
plt.title('Country VS Matches Played')
plt.xlabel('Most player played matches from those countries')
plt.ylabel('Country')
plt.show()

#line chart of Country VS Fifities
data=data.sort_values(by='Fifties')
fif_x=data['Fifties']
con_y=data['Country']
plt.plot(fif_x,con_y)
plt.title('Country VS Fifities')
plt.xlabel('Most player Did their fifities from those country')
plt.ylabel('Country')
plt.show()

#line chart of Country VS Centuries
```

```
data=data.sort_values(by='Centuries')
cen_x=data['Centuries']
con_y=data['Country']
plt.plot(cen_x,con_y)
plt.title('Country VS Centuries')
plt.xlabel('Most player Did their Centuries from those country')
plt.ylabel('Country')
plt.show()
```

```
#Bar Plot of Centuries VS Matches Played
```

```
cen_x = data['Centuries']
mat_y = data['Matches Played']

plt.bar(cen_x,mat_y)
plt.title("Bar Plot of Centuries VS Matches Played")
plt.xlabel("Centuries")
plt.ylabel("Match")
plt.show()
```

```
#Bar Plot of Half-Centures VS Matches Played
```

```
fif_x=data['Fifties']
mat_y=data['Matches Played']

plt.bar(fif_x,mat_y)
plt.title("Bar Plot of Half-Centures VS Matches Played")
plt.xlabel("Half-Centuries")
plt.ylabel("Match")
plt.show()
```

```
#Bar plot of Country VS Matches Played
```

```
con_x=data['Country']
mat_y=data['Matches Played']
plt.barh(con_x,mat_y)
plt.title("Bar plot of Country VS Matches Played")
plt.xlabel('Most matches Played by player')
plt.ylabel('Country')
plt.show()
```

```
#Histogram of Debut Year
```

```
deb_his=data['Debut Year']
plt.hist(deb_his,bins=6,edgecolor='black')
plt.xlabel('Debut Year')
plt.ylabel('Frequency Of Player')
plt.title('Histogram of Debut Year')
plt.show()
```

```
#Histogram of Upto Year Played
```

```

up_his=data['Upto year played']
plt.hist(up_his,bins=5,edgecolor='black')
plt.xlabel('Upto year played')
plt.ylabel('Frequency Of Player')
plt.title('Histogram Of Upto Year Played')
plt.show()

#Histogram of Highest Score

hig_his=data['Highest Score']
plt.hist(hig_his,bins=6,edgecolor='black')
plt.xlabel('Highest Score')
plt.ylabel('Frequency of Player who do highest score')
plt.title('Histogram of Highest Score')
plt.show()

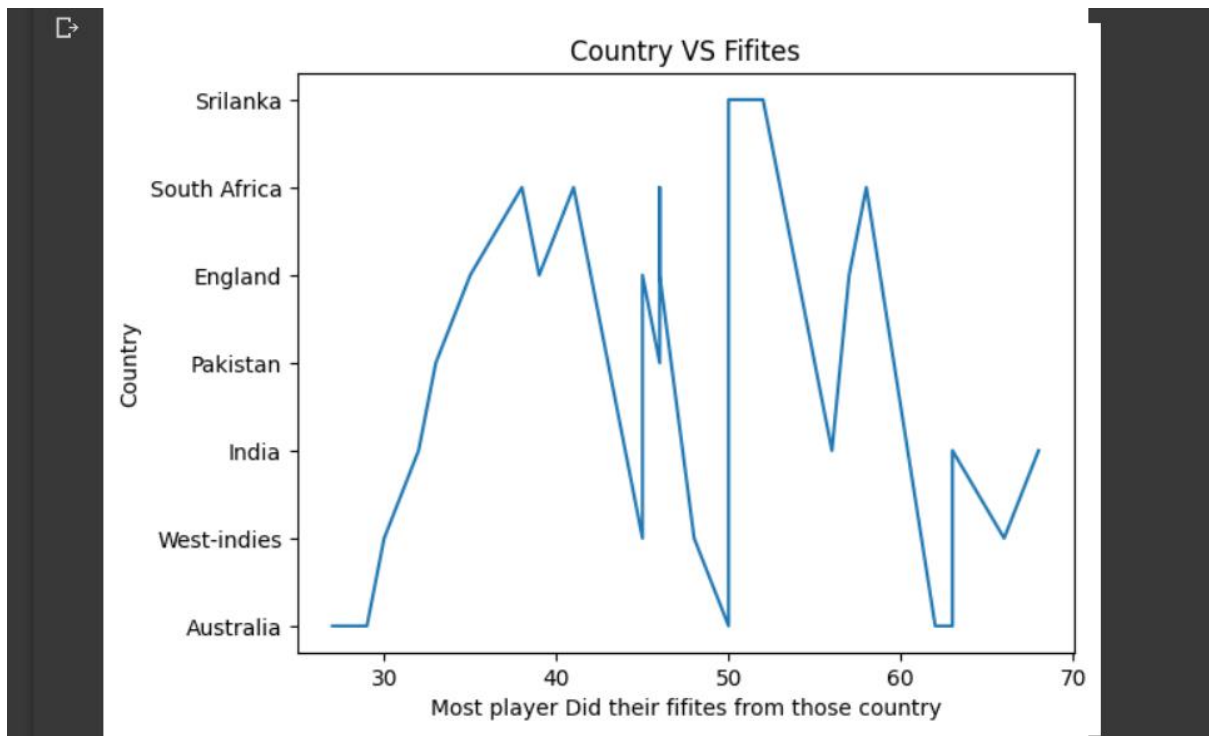
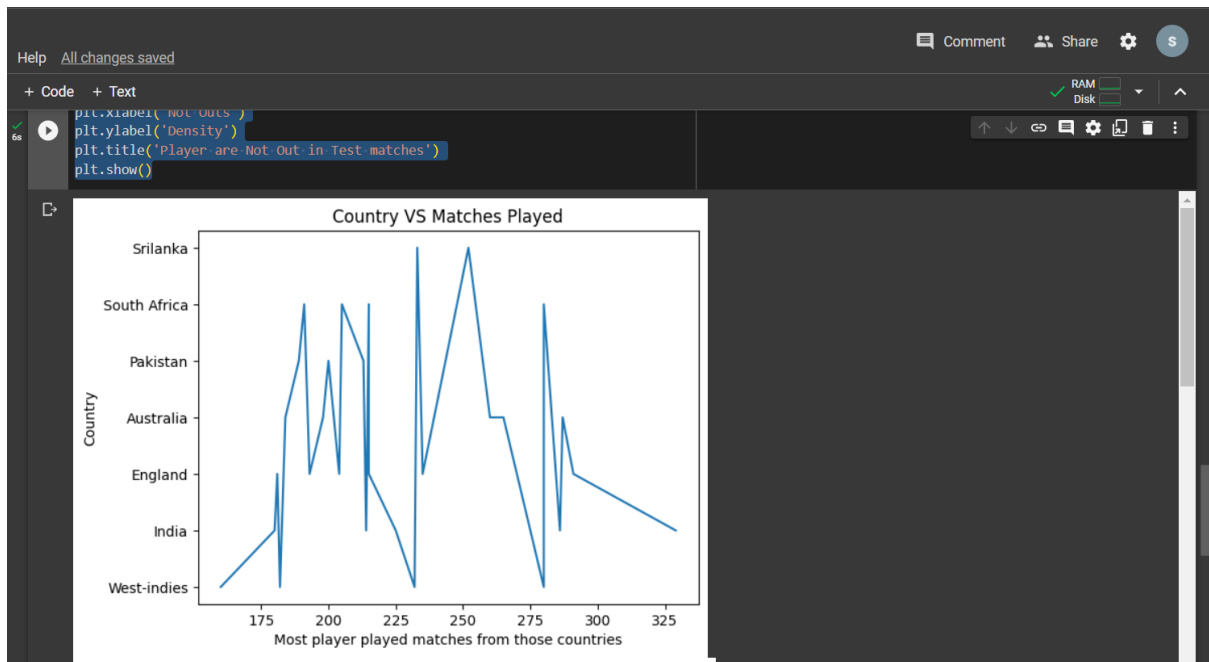
#Density Plot of Batting Average
Bat_avg=data['Batting Average']
dens1=np.random.normal(Bat_avg)
plt.hist(dens1,density=True,bins=10,alpha=0.5)
plt.xlabel('Batting Average')
plt.ylabel('Density')
plt.title('Batting Average of player')
plt.show()

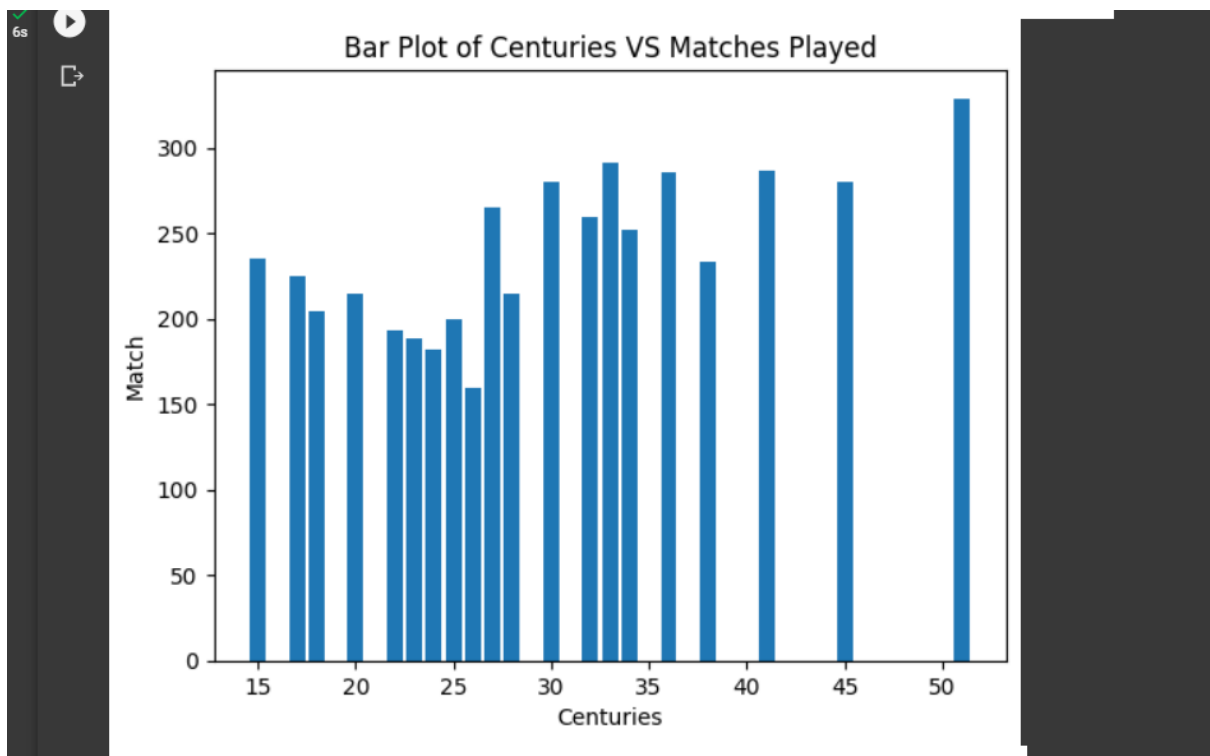
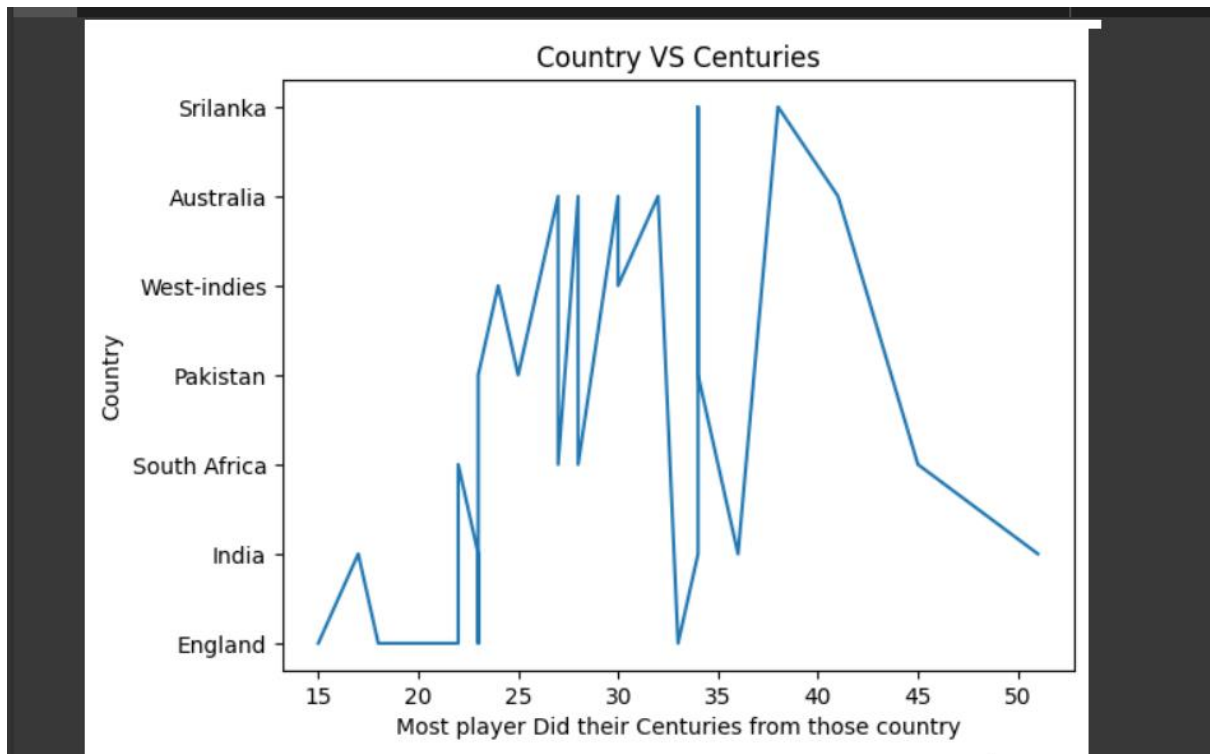
#Density plot of Runs Scored
run=data['Runs Scored']
dens2=np.random.normal(run)
plt.hist(dens2,density=True,bins=10,alpha=0.5)
plt.xlabel('Batting Average')
plt.ylabel('Density')
plt.title('Runs Scored By player')
plt.show()

#Density plot of Not Outs
not_out=data['Not Outs']
dens3=np.random.normal(not_out)
plt.hist(dens3,density=True,bins=10,alpha=0.5)
plt.xlabel('Not Outs')
plt.ylabel('Density')
plt.title('Player are Not Out in Test matches')
plt.show()

```

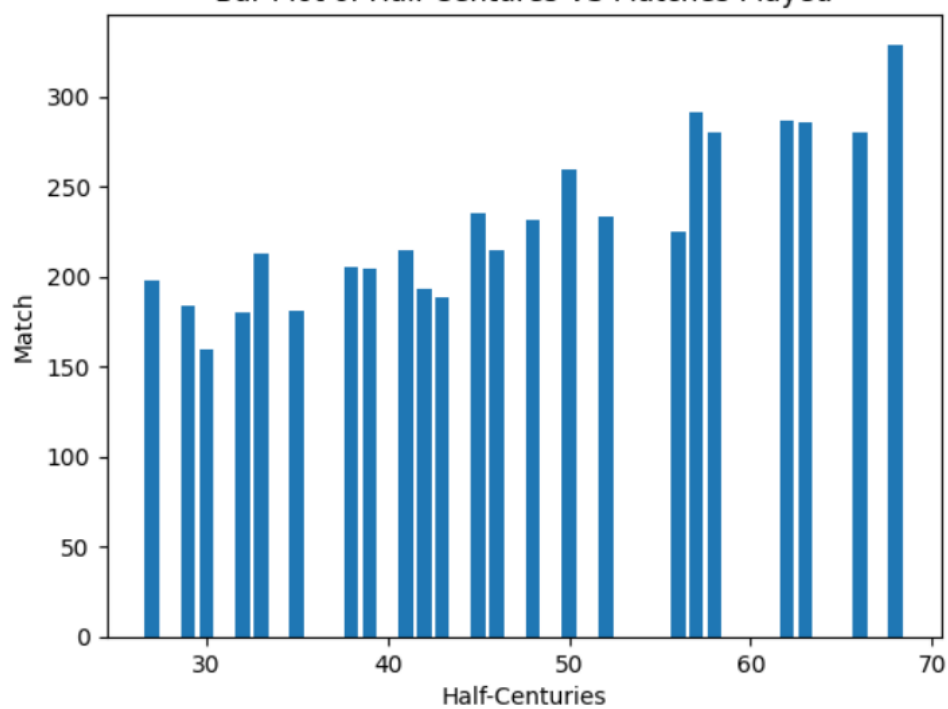
Output:



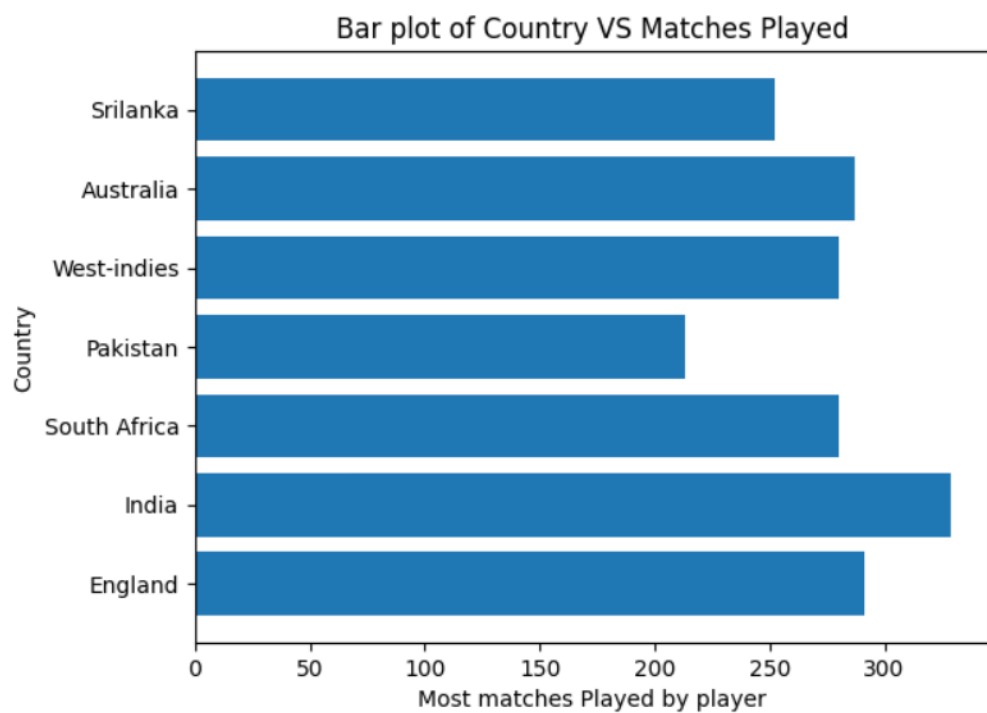




Bar Plot of Half-Centuries VS Matches Played

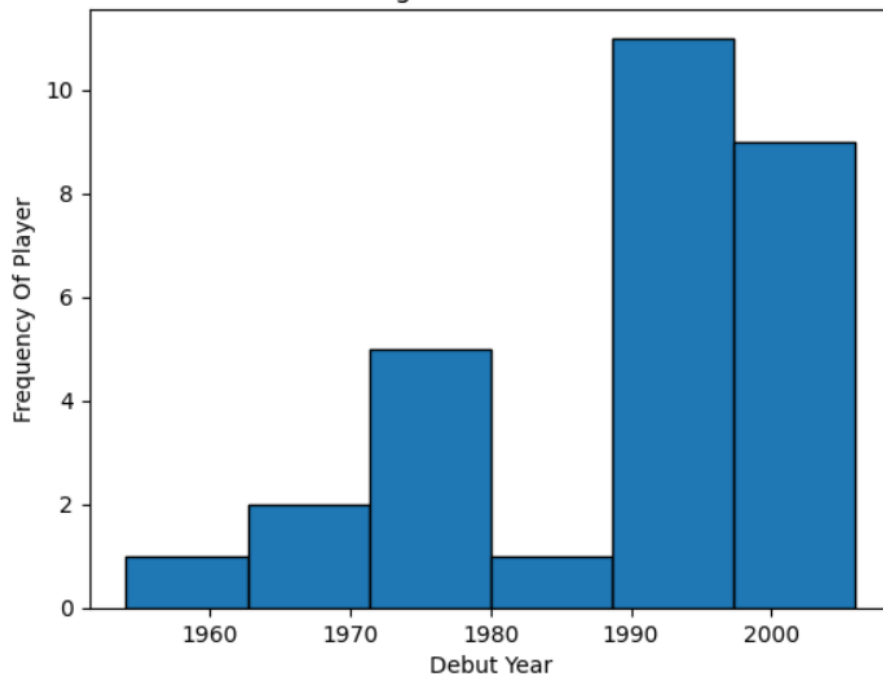


Bar plot of Country VS Matches Played

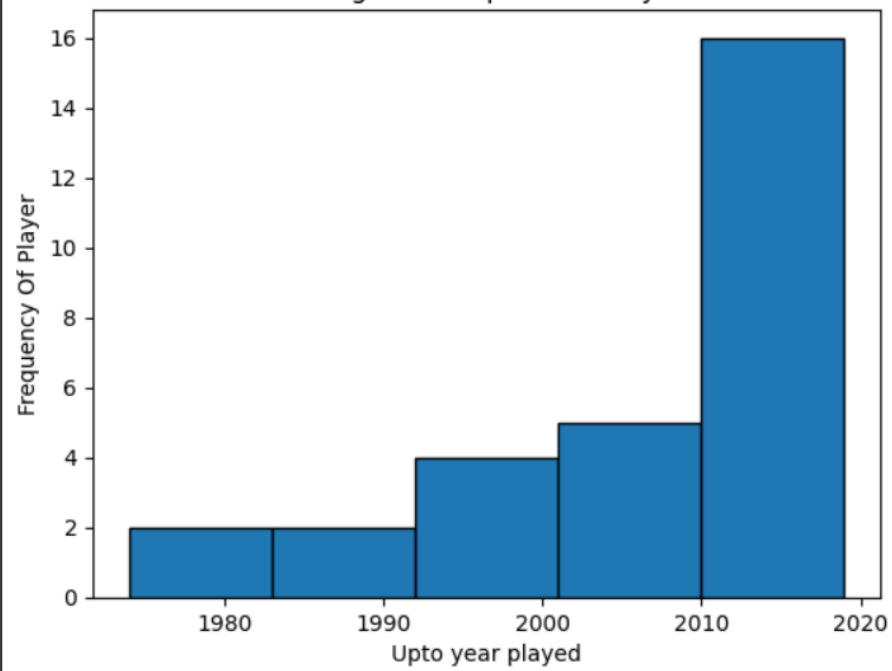




Histogram of Debut Year

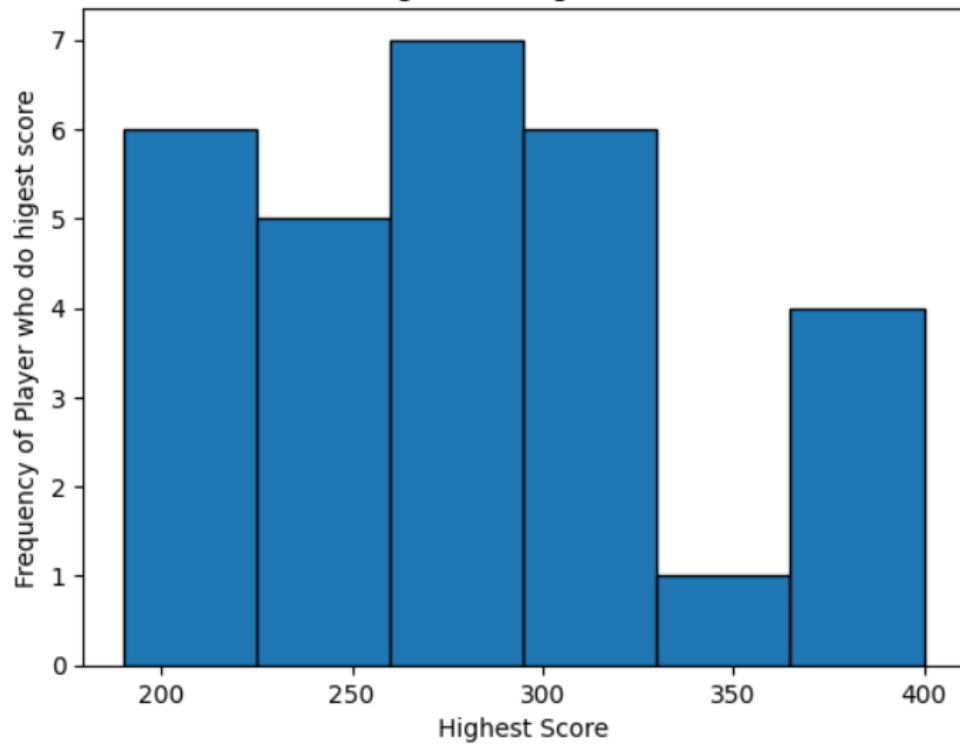


Histogram Of Upto Year Played

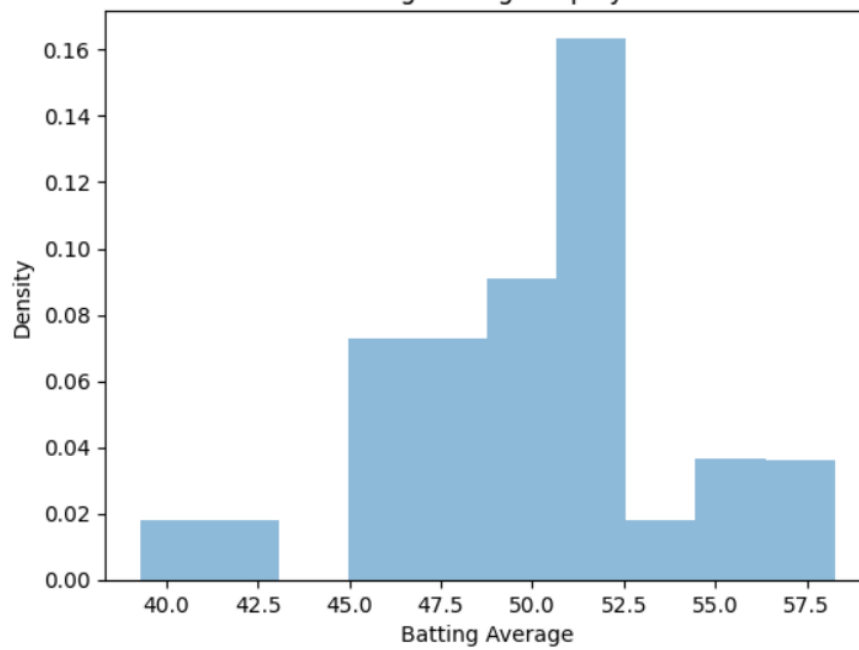




Histogram of Highest Score



Batting Average of player





6s

