

OMKAR BANDU SHINDE

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")
```

```
In [2]: df = pd.read_csv("DataSets/mobile_data.csv")
df.head().T
```

Out[2]:

| | 0 | 1 | 2 | 3 | 4 |
|----------------------|--------|--------|--------|--------|--------|
| battery_power | 842.0 | 1021.0 | 563.0 | 615.0 | 1821.0 |
| blue | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| clock_speed | 2.2 | 0.5 | 0.5 | 2.5 | 1.2 |
| dual_sim | 0.0 | 1.0 | 1.0 | 0.0 | 0.0 |
| fc | 1.0 | 0.0 | 2.0 | 0.0 | 13.0 |
| four_g | 0.0 | 1.0 | 1.0 | 0.0 | 1.0 |
| int_memory | 7.0 | 53.0 | 41.0 | 10.0 | 44.0 |
| m_dep | 0.6 | 0.7 | 0.9 | 0.8 | 0.6 |
| mobile_wt | 188.0 | 136.0 | 145.0 | 131.0 | 141.0 |
| n_cores | 2.0 | 3.0 | 5.0 | 6.0 | 2.0 |
| pc | 2.0 | 6.0 | 6.0 | 9.0 | 14.0 |
| px_height | 20.0 | 905.0 | 1263.0 | 1216.0 | 1208.0 |
| px_width | 756.0 | 1988.0 | 1716.0 | 1786.0 | 1212.0 |
| ram | 2549.0 | 2631.0 | 2603.0 | 2769.0 | 1411.0 |
| sc_h | 9.0 | 17.0 | 11.0 | 16.0 | 8.0 |
| sc_w | 7.0 | 3.0 | 2.0 | 8.0 | 2.0 |
| talk_time | 19.0 | 7.0 | 9.0 | 11.0 | 15.0 |
| three_g | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| touch_screen | 0.0 | 1.0 | 1.0 | 0.0 | 1.0 |
| wifi | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| price_range | 1.0 | 2.0 | 2.0 | 2.0 | 1.0 |

In [3]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   battery_power    2000 non-null   int64  
 1   blue              2000 non-null   int64  
 2   clock_speed      2000 non-null   float64 
 3   dual_sim          2000 non-null   int64  
 4   fc                2000 non-null   int64  
 5   four_g            2000 non-null   int64  
 6   int_memory        2000 non-null   int64  
 7   m_dep             2000 non-null   float64 
 8   mobile_wt         2000 non-null   int64  
 9   n_cores           2000 non-null   int64  
 10  pc                2000 non-null   int64  
 11  px_height         2000 non-null   int64  
 12  px_width          2000 non-null   int64  
 13  ram               2000 non-null   int64  
 14  sc_h              2000 non-null   int64  
 15  sc_w              2000 non-null   int64  
 16  talk_time          2000 non-null   int64  
 17  three_g            2000 non-null   int64  
 18  touch_screen       2000 non-null   int64  
 19  wifi               2000 non-null   int64  
 20  price_range        2000 non-null   int64  
dtypes: float64(2), int64(19)
memory usage: 328.2 KB
```

- No null values

In [4]: df.describe().T

Out[4]:

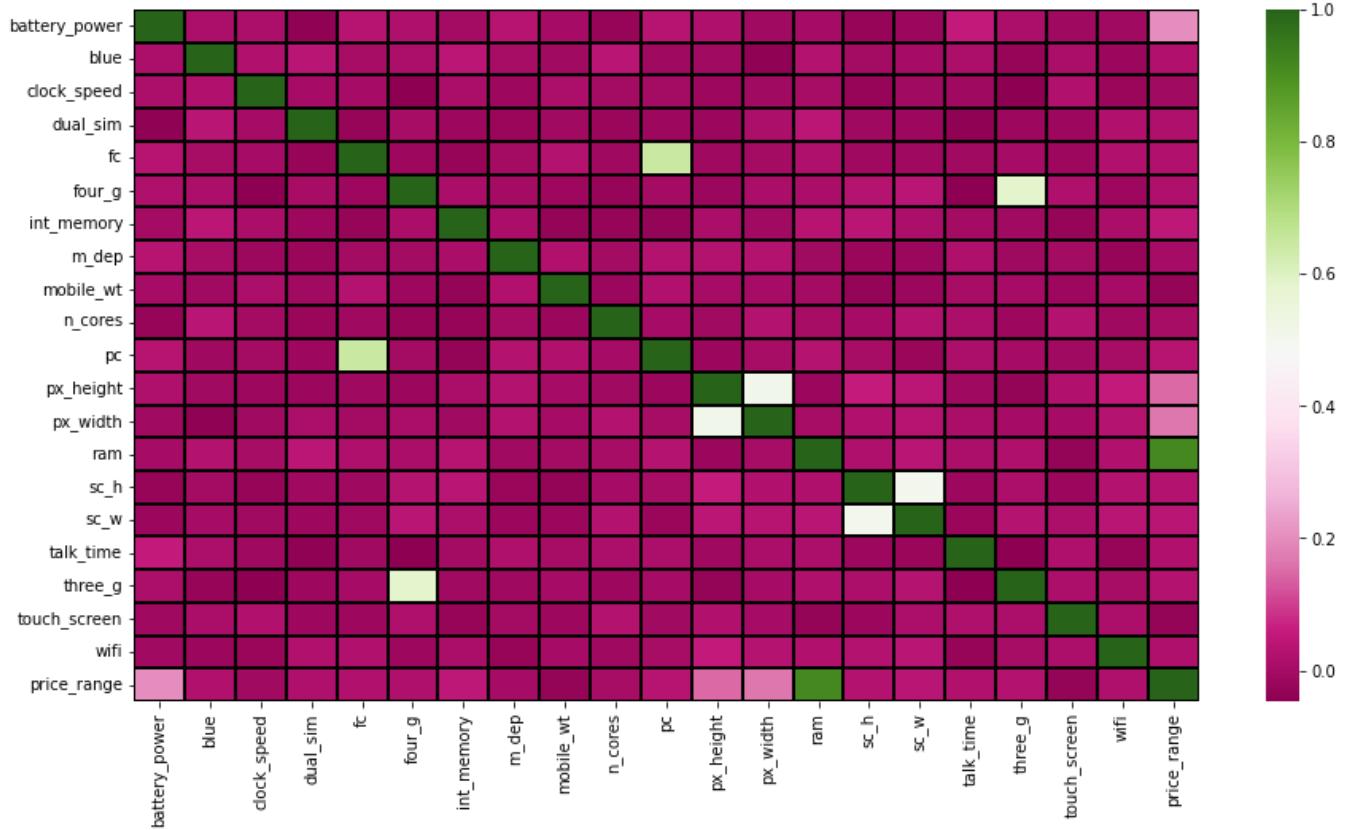
| | count | mean | std | min | 25% | 50% | 75% | max |
|----------------------|--------|------------|-------------|-------|---------|--------|---------|--------|
| battery_power | 2000.0 | 1238.51850 | 439.418206 | 501.0 | 851.75 | 1226.0 | 1615.25 | 1998.0 |
| blue | 2000.0 | 0.49500 | 0.500100 | 0.0 | 0.00 | 0.0 | 1.00 | 1.0 |
| clock_speed | 2000.0 | 1.52225 | 0.816004 | 0.5 | 0.70 | 1.5 | 2.20 | 3.0 |
| dual_sim | 2000.0 | 0.50950 | 0.500035 | 0.0 | 0.00 | 1.0 | 1.00 | 1.0 |
| fc | 2000.0 | 4.30950 | 4.341444 | 0.0 | 1.00 | 3.0 | 7.00 | 19.0 |
| four_g | 2000.0 | 0.52150 | 0.499662 | 0.0 | 0.00 | 1.0 | 1.00 | 1.0 |
| int_memory | 2000.0 | 32.04650 | 18.145715 | 2.0 | 16.00 | 32.0 | 48.00 | 64.0 |
| m_dep | 2000.0 | 0.50175 | 0.288416 | 0.1 | 0.20 | 0.5 | 0.80 | 1.0 |
| mobile_wt | 2000.0 | 140.24900 | 35.399655 | 80.0 | 109.00 | 141.0 | 170.00 | 200.0 |
| n_cores | 2000.0 | 4.52050 | 2.287837 | 1.0 | 3.00 | 4.0 | 7.00 | 8.0 |
| pc | 2000.0 | 9.91650 | 6.064315 | 0.0 | 5.00 | 10.0 | 15.00 | 20.0 |
| px_height | 2000.0 | 645.10800 | 443.780811 | 0.0 | 282.75 | 564.0 | 947.25 | 1960.0 |
| px_width | 2000.0 | 1251.51550 | 432.199447 | 500.0 | 874.75 | 1247.0 | 1633.00 | 1998.0 |
| ram | 2000.0 | 2124.21300 | 1084.732044 | 256.0 | 1207.50 | 2146.5 | 3064.50 | 3998.0 |
| sc_h | 2000.0 | 12.30650 | 4.213245 | 5.0 | 9.00 | 12.0 | 16.00 | 19.0 |
| sc_w | 2000.0 | 5.76700 | 4.356398 | 0.0 | 2.00 | 5.0 | 9.00 | 18.0 |
| talk_time | 2000.0 | 11.01100 | 5.463955 | 2.0 | 6.00 | 11.0 | 16.00 | 20.0 |
| three_g | 2000.0 | 0.76150 | 0.426273 | 0.0 | 1.00 | 1.0 | 1.00 | 1.0 |
| touch_screen | 2000.0 | 0.50300 | 0.500116 | 0.0 | 0.00 | 1.0 | 1.00 | 1.0 |
| wifi | 2000.0 | 0.50700 | 0.500076 | 0.0 | 0.00 | 1.0 | 1.00 | 1.0 |
| price_range | 2000.0 | 1.50000 | 1.118314 | 0.0 | 0.75 | 1.5 | 2.25 | 3.0 |

EDA

1 Correlation

In [5]: # Heatmap for corelation

```
plt.figure(figsize=(15,8))
sns.heatmap(df.corr(), annot=False, cmap="PiYG", linewidths=2, linecolor="black")
plt.show()
```

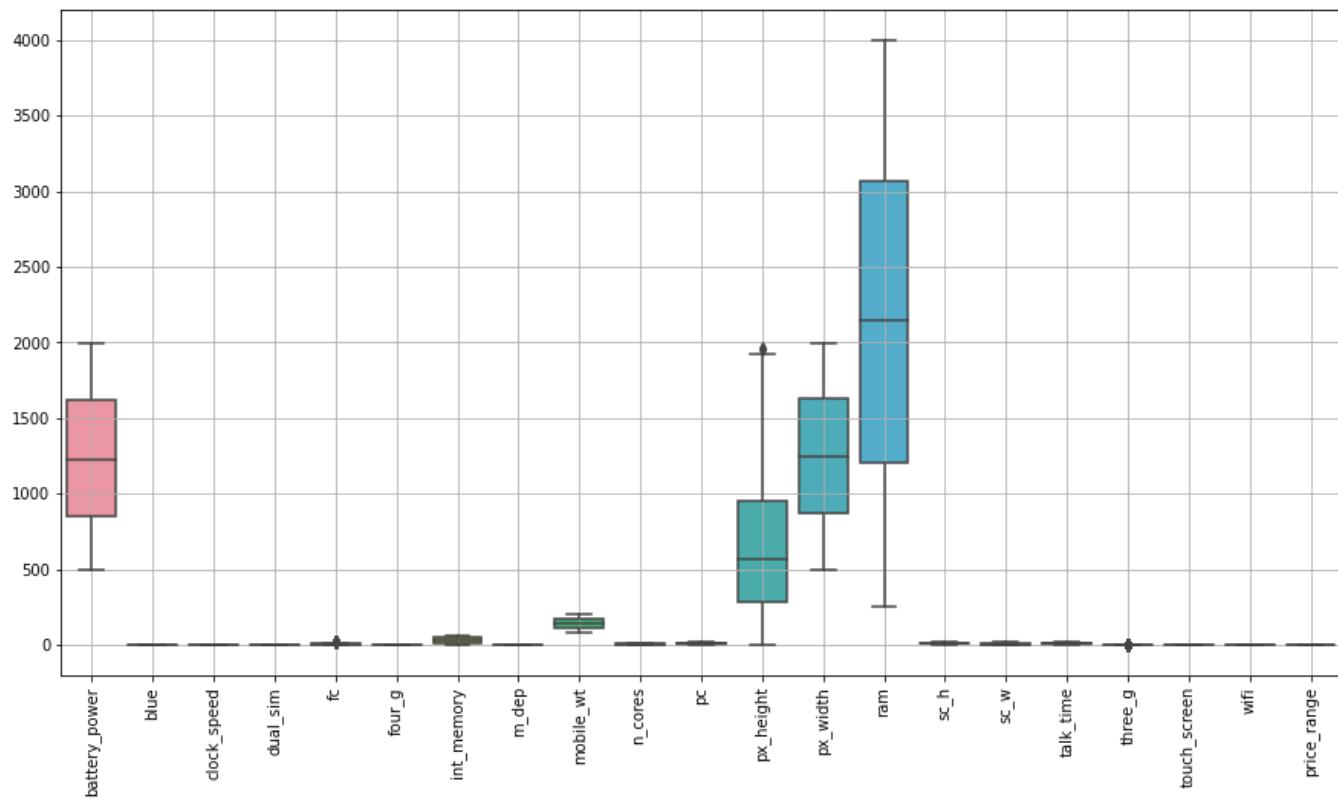


- Ram & Price are Highly correlated with each other (Ram↑ == Price↑) OR (Ram↓ == Price↓)
- PC & FC are nearly (60% - 80%) correlated with each other (PC↑ == FC↑) OR (PC↓ == FC↓)
- four_g & three_G are slightly correlated with each other

2 Outliers

In [6]: # Searching for any outliers

```
plt.figure(figsize=(15,8))
sns.boxplot(data=df)
plt.grid()
plt.xticks(rotation=90)
plt.show()
```



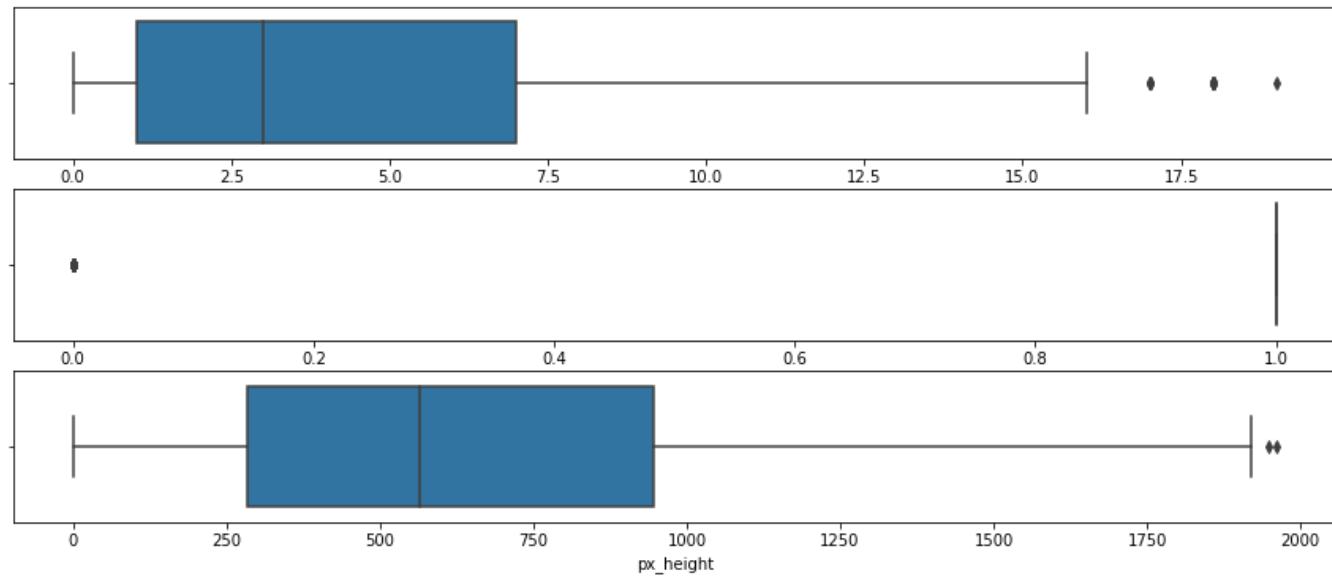
- By seeing the Boxplot i can see there are some outliers in fc , px_height , three_g
- Plotting Individual Boxplot for each and displaying the outling records for the respective columns

```
In [7]: plt.figure(figsize=(15,6))
plt.subplot(3,1,1)
sns.boxplot(df.fc)

plt.subplot(3,1,2)
sns.boxplot(df.three_g)

plt.subplot(3,1,3)
sns.boxplot(df.px_height)

plt.show()
```



```
In [8]: df[df.fc > 17]
```

Out[8]:

| | battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep | mobile_wt | n_cores | ... |
|-------------|---------------|------|-------------|----------|----|--------|------------|-------|-----------|---------|-----|
| 95 | 1137 | 1 | 1.0 | 0 | 18 | 0 | 7 | 1.0 | 196 | 3 | ... |
| 226 | 1708 | 1 | 2.4 | 1 | 18 | 1 | 49 | 0.1 | 109 | 1 | ... |
| 305 | 1348 | 0 | 2.0 | 0 | 18 | 0 | 52 | 0.3 | 98 | 3 | ... |
| 1387 | 1533 | 1 | 1.1 | 1 | 18 | 1 | 17 | 0.3 | 160 | 4 | ... |
| 1406 | 1731 | 1 | 2.3 | 1 | 18 | 0 | 60 | 0.5 | 171 | 4 | ... |
| 1416 | 1448 | 0 | 0.5 | 1 | 18 | 0 | 2 | 0.2 | 100 | 5 | ... |
| 1554 | 1957 | 0 | 1.2 | 1 | 18 | 1 | 36 | 0.8 | 151 | 2 | ... |
| 1693 | 695 | 0 | 0.5 | 0 | 18 | 1 | 12 | 0.6 | 196 | 2 | ... |
| 1705 | 1290 | 1 | 1.4 | 1 | 19 | 1 | 35 | 0.3 | 110 | 4 | ... |
| 1880 | 1720 | 0 | 1.6 | 0 | 18 | 1 | 2 | 0.8 | 188 | 5 | ... |
| 1882 | 591 | 0 | 2.1 | 1 | 18 | 1 | 16 | 0.5 | 196 | 7 | ... |
| 1888 | 1544 | 0 | 2.4 | 0 | 18 | 1 | 12 | 0.1 | 186 | 7 | ... |

12 rows × 21 columns

```
In [9]: df[df.three_g == 0]
```

Out[9]:

| | battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep | mobile_wt | n_cores | ... |
|------|---------------|------|-------------|----------|-----|--------|------------|-------|-----------|---------|-----|
| 0 | 842 | 0 | 2.2 | 0 | 1 | 0 | 7 | 0.6 | 188 | 2 | ... |
| 10 | 769 | 1 | 2.9 | 1 | 0 | 0 | 9 | 0.1 | 182 | 5 | ... |
| 19 | 682 | 1 | 0.5 | 0 | 4 | 0 | 19 | 1.0 | 121 | 4 | ... |
| 22 | 1949 | 0 | 2.6 | 1 | 4 | 0 | 47 | 0.3 | 199 | 4 | ... |
| 30 | 1579 | 1 | 0.5 | 1 | 0 | 0 | 5 | 0.2 | 88 | 7 | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1981 | 1454 | 0 | 2.6 | 0 | 8 | 0 | 6 | 0.4 | 199 | 3 | ... |
| 1982 | 1784 | 0 | 1.6 | 0 | 4 | 0 | 41 | 0.4 | 164 | 6 | ... |
| 1983 | 1262 | 0 | 1.8 | 1 | 12 | 0 | 34 | 0.1 | 149 | 5 | ... |
| 1988 | 1547 | 1 | 2.9 | 0 | 2 | 0 | 57 | 0.4 | 114 | 1 | ... |
| 1989 | 586 | 0 | 2.8 | 0 | 2 | 0 | 15 | 0.2 | 83 | 3 | ... |

477 rows × 21 columns

```
In [10]: df[df.px_height > 1900]
```

Out[10]:

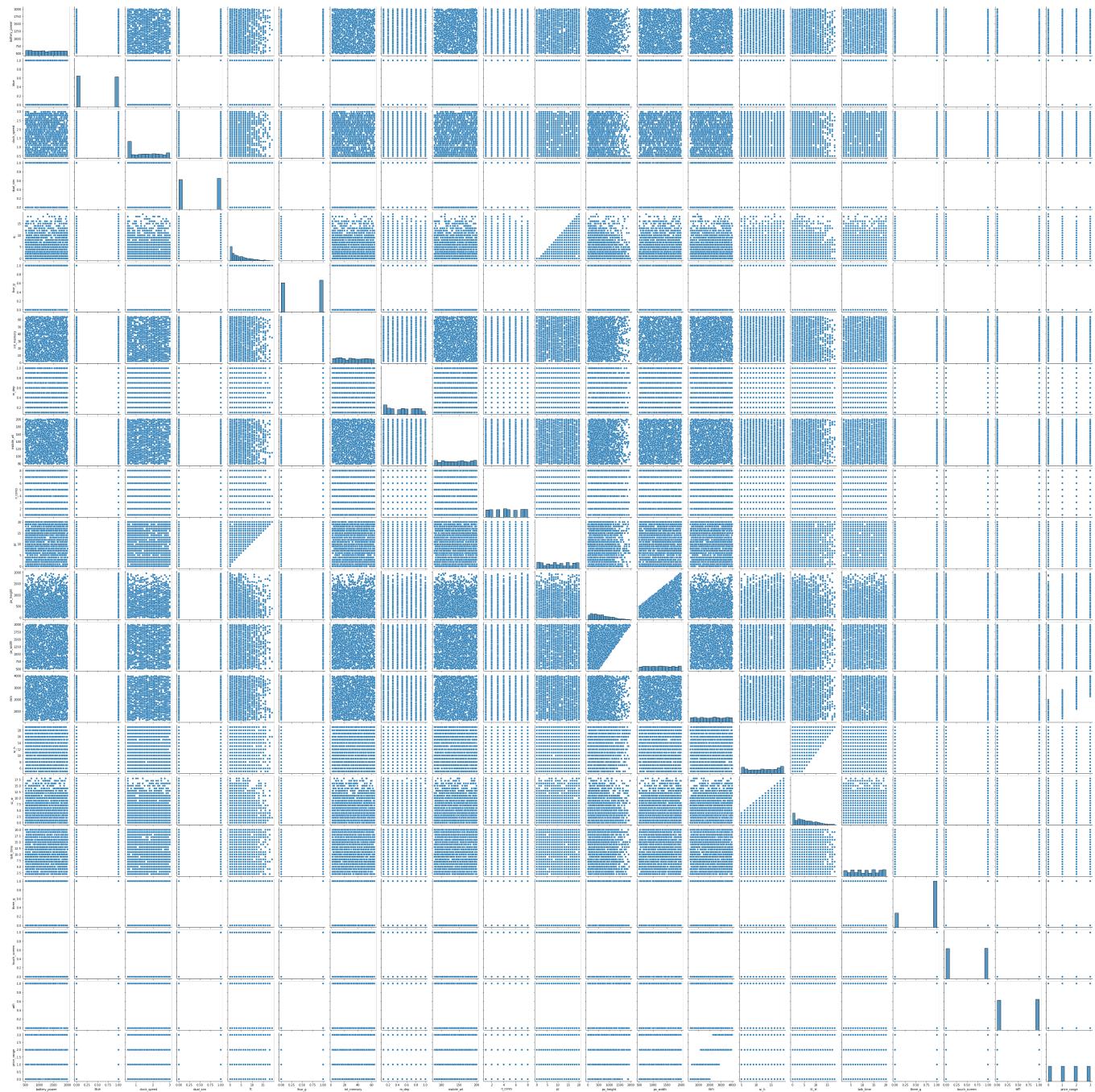
| | battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep | mobile_wt | n_cores | ... |
|------|---------------|------|-------------|----------|----|--------|------------|-------|-----------|---------|-----|
| 260 | 754 | 0 | 0.5 | 1 | 7 | 1 | 59 | 0.7 | 178 | 7 | ... |
| 894 | 1497 | 1 | 0.7 | 0 | 5 | 0 | 32 | 0.7 | 92 | 3 | ... |
| 988 | 1413 | 1 | 0.5 | 1 | 4 | 1 | 45 | 0.4 | 104 | 5 | ... |
| 1163 | 1930 | 1 | 2.0 | 0 | 11 | 0 | 16 | 0.8 | 186 | 8 | ... |
| 1771 | 1230 | 1 | 1.6 | 0 | 0 | 1 | 48 | 0.7 | 111 | 7 | ... |

5 rows × 21 columns

3 Distribution of data

Price_Range data

```
In [11]: sns.pairplot(df)
plt.grid()
plt.show()
```



```
In [12]: # Equally distributed data
```

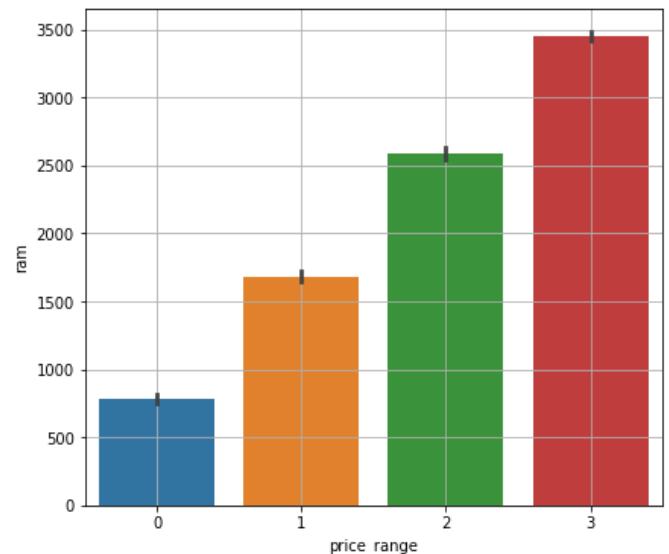
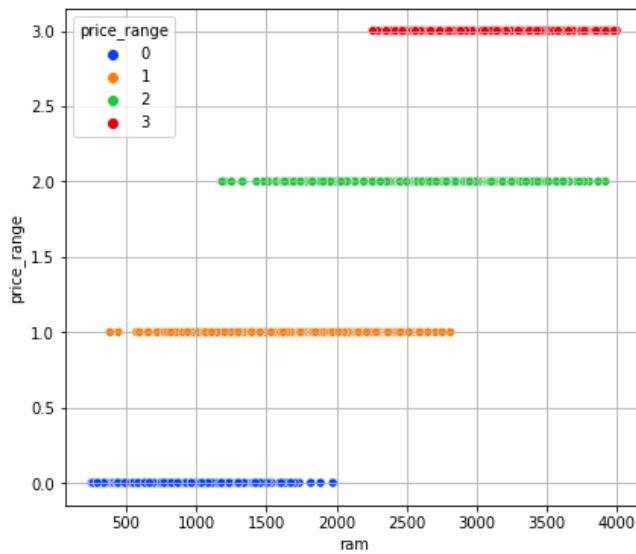
```
plt.pie(df.price_range.value_counts(), autopct="%2f", labels=[0,1,2,3])  
plt.show()
```



Ram vs Price

```
In [13]: plt.figure(figsize = (15,6))
```

```
plt.subplot(1,2,1)  
sns.scatterplot(x=df.ram,y=df.price_range,hue=df.price_range,palette="bright")  
plt.grid()  
  
plt.subplot(1,2,2)  
sns.barplot(x = 'price_range', y = 'ram',data=df)  
plt.grid()  
plt.show()
```



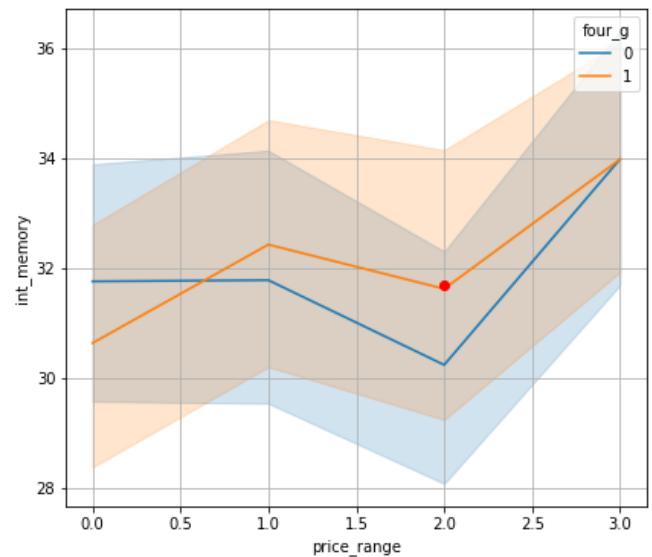
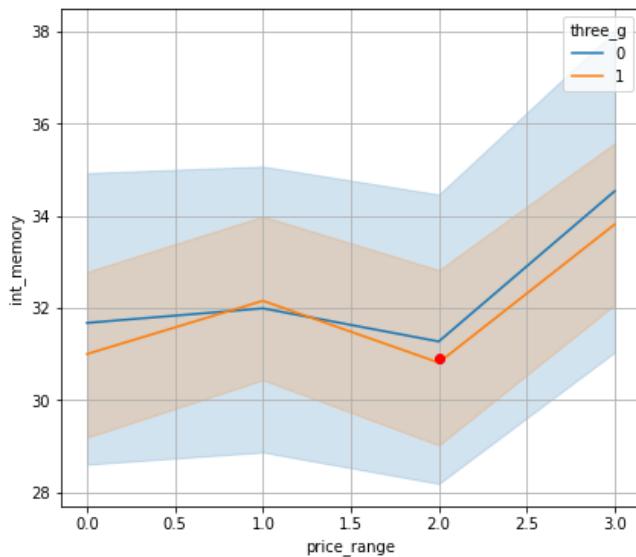
- 0 ==> (0 MB < ram_size < 2000 MB)
- 1 ==> (400 MB < ram_size < 3000 MB)
- 2 ==> (1200 MB < ram_size < 3900 MB)
- 3 ==> (2200 MB < ram_size < 4000 MB)

Internal_Memory vs Price W.R.T --- (3G / 4G)

```
In [14]: plt.figure(figsize = (15,6))

plt.subplot(1,2,1)
sns.lineplot(y=df.int_memory,x=df.price_range,hue=df.three_g)
plt.plot(2.0,30.9,marker="o",color="r")
plt.grid()

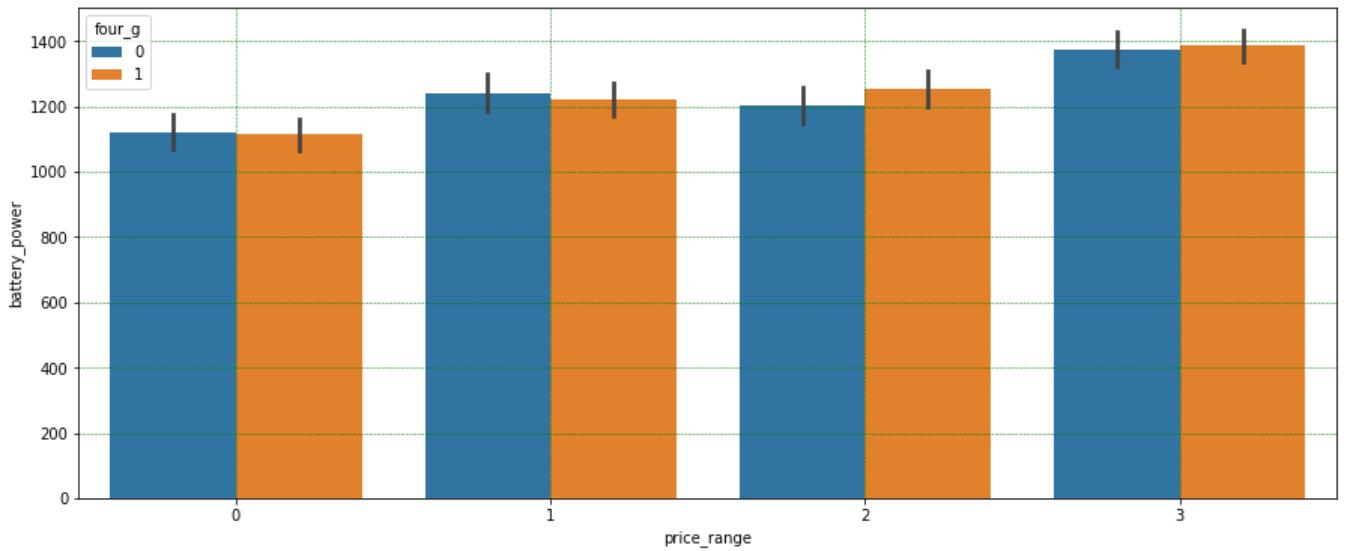
plt.subplot(1,2,2)
sns.lineplot(y=df.int_memory,x=df.price_range,hue=df.four_g)
plt.plot(2.0,31.7,marker="o",color="r")
plt.grid()
```



- 3G phone does have an high drop rate in internal_memory w.r.t to price_range
- 4G phone does not have an high drop rate in internal_memory w.r.t to price_range

Price_range vs battery_power

```
In [15]: plt.figure(figsize = (15,6))
sns.barplot(x = 'price_range', y = 'battery_power', data=df,hue=df.four_g)
plt.grid(color = 'green', linestyle = '--', linewidth = 0.5)
plt.show()
```



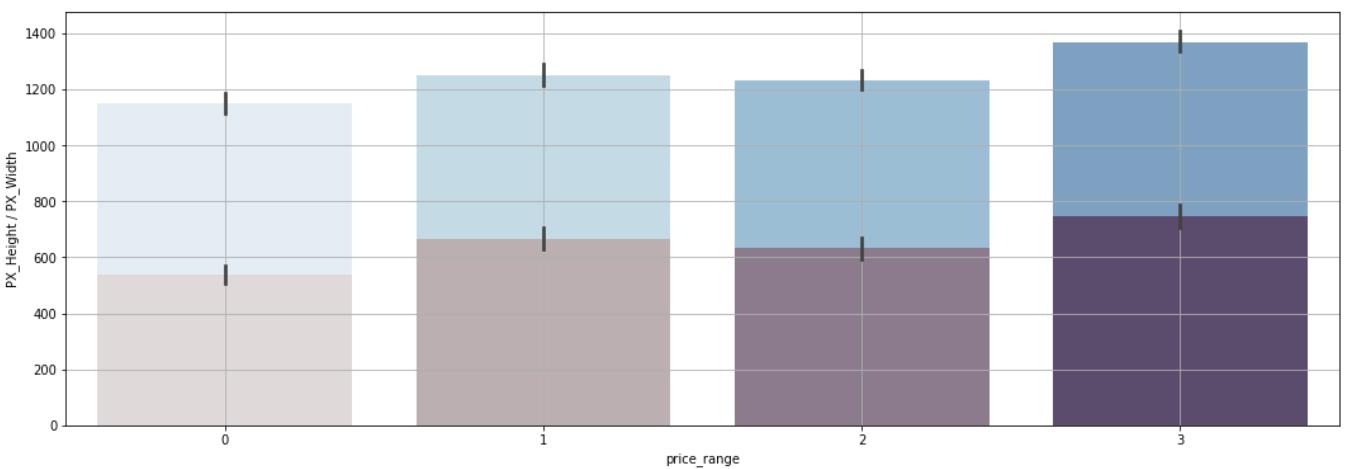
- Costlier phones have higher battery capacity

(PX_Height / PX_Width) vs Price_Range

```
In [16]: plt.figure(figsize = (18,6))

sns.barplot(x = 'price_range', y = 'px_height', data=df, palette = 'Reds')
sns.barplot(x = 'price_range', y = 'px_width', data=df , palette = 'Blues',alpha=0.6)
plt.ylabel("PX_Height / PX_Width")
plt.grid()

plt.show()
```



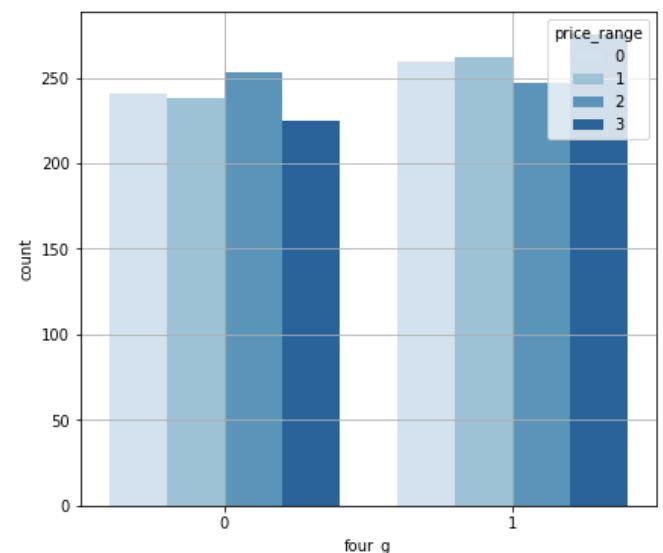
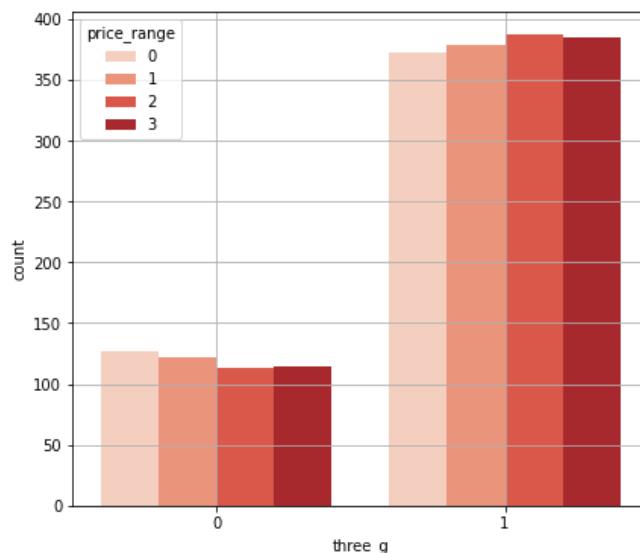
- 0 ==> Smaller Dimensions (H * W)
- 1 & 2 ==> Similar Dimensions (H * W)
- 3 ==> Larger Dimensions (H * W)

Price vs (3G / 4G)

```
In [17]: plt.figure(figsize = (15,6))

plt.subplot(1,2,1)
sns.countplot(df[ 'three_g' ] , hue = df[ 'price_range' ] , palette = 'Reds')
plt.grid()

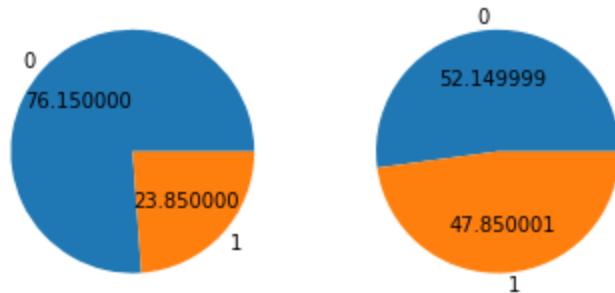
plt.subplot(1,2,2)
sns.countplot(df[ 'four_g' ] , hue = df[ 'price_range' ] , palette = 'Blues')
plt.grid()
plt.show()
```



```
In [18]: plt.subplot(1,2,1)
plt.pie(df.three_g.value_counts(), autopct="%2f", labels=[0,1])
print(df.three_g.value_counts())
plt.subplot(1,2,2)
plt.pie(df.four_g.value_counts(), autopct="%2f", labels=[0,1])
print(df.four_g.value_counts())

plt.show()
```

```
1    1523
0    477
Name: three_g, dtype: int64
1    1043
0    957
Name: four_g, dtype: int64
```



- More no.of counts in 3G phones compared to 4G phones in Market
- 3G ==> Price_Range(2 > 3 > 1 > 0)
- 4G ==> Price_Range(3 > 1 > 0 > 2)

In []:

In []:

In []:

Segregating , Scalling & Splitting

```
In [19]: from sklearn.preprocessing import StandardScaler
SS = StandardScaler()

X = df.iloc[:, :-1]
# X = SS.fit_transform(X)
X
```

Out[19]:

| | battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep | mobile_wt | n_cores | pc |
|------|---------------|------|-------------|----------|-----|--------|------------|-------|-----------|---------|-----|
| 0 | 842 | 0 | 2.2 | 0 | 1 | 0 | 7 | 0.6 | 188 | 2 | 2 |
| 1 | 1021 | 1 | 0.5 | 1 | 0 | 1 | 53 | 0.7 | 136 | 3 | 6 |
| 2 | 563 | 1 | 0.5 | 1 | 2 | 1 | 41 | 0.9 | 145 | 5 | 6 |
| 3 | 615 | 1 | 2.5 | 0 | 0 | 0 | 10 | 0.8 | 131 | 6 | 9 |
| 4 | 1821 | 1 | 1.2 | 0 | 13 | 1 | 44 | 0.6 | 141 | 2 | 14 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1995 | 794 | 1 | 0.5 | 1 | 0 | 1 | 2 | 0.8 | 106 | 6 | 14 |
| 1996 | 1965 | 1 | 2.6 | 1 | 0 | 0 | 39 | 0.2 | 187 | 4 | 3 |
| 1997 | 1911 | 0 | 0.9 | 1 | 1 | 1 | 36 | 0.7 | 108 | 8 | 3 |
| 1998 | 1512 | 0 | 0.9 | 0 | 4 | 1 | 46 | 0.1 | 145 | 5 | 5 |
| 1999 | 510 | 1 | 2.0 | 1 | 5 | 1 | 45 | 0.9 | 168 | 6 | 16 |

2000 rows × 20 columns

```
In [20]: Y = df.price_range
Y
```

```
Out[20]: 0      1
1      2
2      2
3      2
4      1
..
1995    0
1996    2
1997    3
1998    0
1999    3
Name: price_range, Length: 2000, dtype: int64
```

```
In [21]: from sklearn.model_selection import train_test_split, GridSearchCV
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=1)
```

Modelling

```
In [22]: from sklearn.metrics import classification_report,accuracy_score,roc_curve
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC

algo_performance_list = []

def MyModel(model,Name,Params):
    model.fit(X_train,Y_train)
    Y_pred = model.predict(X_test)
    print(classification_report(Y_test,Y_pred))
    algo_performance_list.append([Name,accuracy_score(Y_test,Y_pred),Params,model.score(X_train,Y_train)])
```

KNN

- 1) Default Model
- 2) Hypertuning
- 3) Final Model

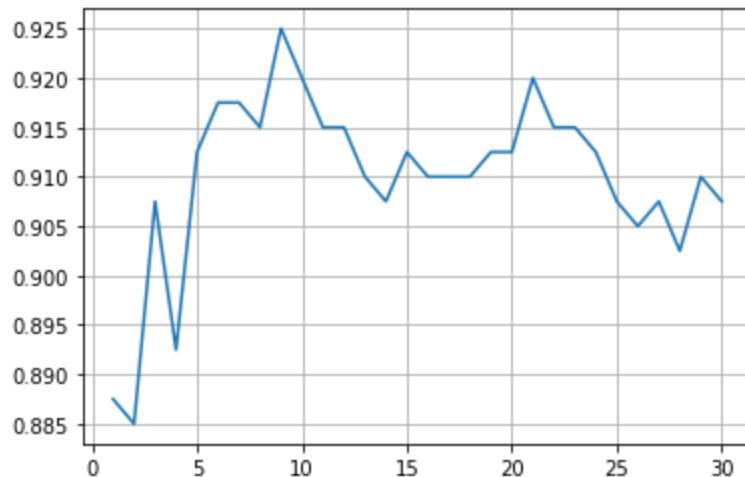
```
In [23]: # 1) Default Model
KNN = KNeighborsClassifier()
MyModel(KNN,"KNN","Default")
print(f"Training Score :- {KNN.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.98 | 0.97 | 0.97 | 92 |
| 1 | 0.90 | 0.94 | 0.92 | 96 |
| 2 | 0.83 | 0.90 | 0.86 | 106 |
| 3 | 0.96 | 0.86 | 0.91 | 106 |
| accuracy | | | 0.91 | 400 |
| macro avg | 0.92 | 0.91 | 0.92 | 400 |
| weighted avg | 0.92 | 0.91 | 0.91 | 400 |

Training Score :- 0.958125

```
In [24]: # 2) Hypertuned KNN
KNN_list = []
for i in range(1,31):
    KNN = KNeighborsClassifier(n_neighbors=i)
    KNN.fit(X_train,Y_train)
    Y_pred = KNN.predict(X_test)
    KNN_list.append(accuracy_score(Y_test,Y_pred))
```

```
In [25]: plt.plot(range(1,31),KNN_list)
plt.grid()
```



```
In [26]: # 3) Final KNN
KNN = KNeighborsClassifier(n_neighbors=9)
MyModel(KNN,"H_KNN","n_neighbors = 9")
print(f"Training Score :- {KNN.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.97 | 0.97 | 0.97 | 92 |
| 1 | 0.90 | 0.95 | 0.92 | 96 |
| 2 | 0.87 | 0.91 | 0.89 | 106 |
| 3 | 0.97 | 0.89 | 0.93 | 106 |
| accuracy | | | 0.93 | 400 |
| macro avg | 0.93 | 0.93 | 0.93 | 400 |
| weighted avg | 0.93 | 0.93 | 0.93 | 400 |

Training Score :- 0.948125

Logistic Regression

- 1) Default Model
- 2) Hypertuning
- 3) Final Model

```
In [27]: # 1) Default Model
LR = LogisticRegression()
MyModel(LR, "LR", "Default")
print(f"Training Score :- {LR.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.82 | 0.77 | 0.79 | 92 |
| 1 | 0.53 | 0.54 | 0.54 | 96 |
| 2 | 0.45 | 0.42 | 0.43 | 106 |
| 3 | 0.68 | 0.75 | 0.71 | 106 |
| accuracy | | | 0.61 | 400 |
| macro avg | 0.62 | 0.62 | 0.62 | 400 |
| weighted avg | 0.61 | 0.61 | 0.61 | 400 |

Training Score :- 0.64125

```
In [28]: # 2) Hypertuning Log_Reg
```

```
params = [ ['lbfgs','l2'], ['lbfgs','none'],
           ['liblinear','l1'],['liblinear','l2'],
           ['newton-cg','l2'], ['newton-cg','none'],
           ['sag','l2'], ['sag','none'],
           ['saga','l1'], ['saga','l2'], ['saga','none'] ]
```

['saga','elasticnet'] ---> Need to perform Scalling on the Data

```
all_combinations = []
```

```
for i in params:
    model = LogisticRegression(solver=i[0] , penalty=i[1])
    model.fit(X_train,Y_train)
    Y_pred = model.predict(X_test)

    acc = accuracy_score(Y_test,Y_pred)
```

```
    print(f"{i} ---> {acc} ")
```

```
    all_combinations.append([i[0],i[1],acc])
```

```
['lbfgs', 'l2'] ---> 0.615
['lbfgs', 'none'] ---> 0.61
['liblinear', 'l1'] ---> 0.86
['liblinear', 'l2'] ---> 0.7725
['newton-cg', 'l2'] ---> 0.955
['newton-cg', 'none'] ---> 0.9625
['sag', 'l2'] ---> 0.6225
['sag', 'none'] ---> 0.6225
['saga', 'l1'] ---> 0.5975
['saga', 'l2'] ---> 0.6025
['saga', 'none'] ---> 0.6
```

```
In [29]: def Get_Best_combintion(Values_List):
    max_val = 0
    max_index = 0
    for i in Values_List:
        if max_val < i[2]:
            max_val = i[2]
            max_index = Values_List.index(i)
    return Values_List[max_index]

Get_Best_combintion(all_combinations)
```

Out[29]: ['newton-cg', 'none', 0.9625]

```
In [30]: # 3) Final Log_Reg
H_Log_Reg = LogisticRegression(solver="newton-cg",penalty='none')
MyModel(H_Log_Reg,"H_Log_Reg","solver = newton-cg , penalty = none")
print(f"Training Score :- {H_Log_Reg.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.98 | 0.98 | 0.98 | 92 |
| 1 | 0.96 | 0.96 | 0.96 | 96 |
| 2 | 0.94 | 0.95 | 0.95 | 106 |
| 3 | 0.97 | 0.96 | 0.97 | 106 |
| accuracy | | | 0.96 | 400 |
| macro avg | 0.96 | 0.96 | 0.96 | 400 |
| weighted avg | 0.96 | 0.96 | 0.96 | 400 |

Training Score :- 0.974375

RandomForest Classifier

- 1) Default Model
- 2) Hypertuning
- 3) Final Model

```
In [31]: # 1) Default Model
RFC = RandomForestClassifier()
MyModel(model = RFC , Name = "RFC" , Params = "Default" )
print(f"Training Score :- {RFC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.90 | 0.92 | 0.91 | 92 |
| 1 | 0.81 | 0.84 | 0.83 | 96 |
| 2 | 0.83 | 0.85 | 0.84 | 106 |
| 3 | 0.96 | 0.88 | 0.92 | 106 |
| accuracy | | | 0.87 | 400 |
| macro avg | 0.87 | 0.87 | 0.87 | 400 |
| weighted avg | 0.88 | 0.87 | 0.87 | 400 |

Training Score :- 1.0

In [32]: # 2) Hypertuning RFC

```
D_RFC = RandomForestClassifier()
hyperparameter_dict = {"criterion":['gini','entropy'],
                      "max_depth":range(5,10),
                      "min_samples_split":range(5,10),
                      "min_samples_leaf":range(5,10)}

GS_RFC = GridSearchCV(D_RFC , param_grid = hyperparameter_dict , verbose=3)
GS_RFC.fit(X_train,Y_train)
```

Fitting 5 folds for each of 250 candidates, totalling 1250 fits

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=5;, score=0.869 total time= 0.1s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=5;, score=0.825 total time= 0.1s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=5;, score=0.853 total time= 0.0s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=5;, score=0.863 total time= 0.1s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=5;, score=0.831 total time= 0.1s

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=6;, score=0.863 total time= 0.1s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=6;, score=0.822 total time= 0.1s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=6;, score=0.838 total time= 0.1s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=6;, score=0.844 total time= 0.1s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=6;, score=0.828 total time= 0.1s

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=7;, score=0.850 total time= 0.0s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=7;, score=0.834 total time= 0.0s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=7;, score=0.844 total time= 0.1s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=7;, score=0.841 total time= 0.0s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=7;, score=0.809 total time= 0.0s

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=8;, score=0.853 total time= 0.1s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=8;, score=0.847 total time= 0.1s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=8;, score=0.819 total time= 0.1s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=8;, score=0.853 total time= 0.1s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=8;, score=0.819 total time= 0.1s

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=9;, score=0.866 total time= 0.1s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=9;, score=0.822 total time= 0.1s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=9;, score=0.831 total time= 0.1s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=9;, score=0.850 total time= 0.1s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=5, min_samples_split=9;, score=0.838 total time= 0.1s

[CV 1/5] END criterion=gini, max_depth=5, min_samples_leaf=6, min_samples_split=5;, score=0.866 total time= 0.1s

[CV 2/5] END criterion=gini, max_depth=5, min_samples_leaf=6, min_samples_split=5;, score=0.834 total time= 0.1s

[CV 3/5] END criterion=gini, max_depth=5, min_samples_leaf=6, min_samples_split=5;, score=0.847 total time= 0.1s

[CV 4/5] END criterion=gini, max_depth=5, min_samples_leaf=6, min_samples_split=5;, score=0.856 total time= 0.1s

[CV 5/5] END criterion=gini, max_depth=5, min_samples_leaf=6, min_samples_split=5;, score=0.831 total time= 0.1s


```
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=7;, score=0.831 total time= 0.1s
[CV 1/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.869 total time= 0.2s
[CV 2/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.869 total time= 0.2s
[CV 3/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.866 total time= 0.1s
[CV 4/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.894 total time= 0.1s
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.856 total time= 0.1s
[CV 1/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.894 total time= 0.1s
[CV 2/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.884 total time= 0.1s
[CV 3/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.863 total time= 0.1s
[CV 4/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.869 total time= 0.2s
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.853 total time= 0.1s
```

```
Out[32]: GridSearchCV(estimator=RandomForestClassifier(),
                      param_grid={'criterion': ['gini', 'entropy'],
                                  'max_depth': range(5, 10),
                                  'min_samples_leaf': range(5, 10),
                                  'min_samples_split': range(5, 10)},
                      verbose=3)
```

```
In [33]: GS_RFC.best_params_
```

```
Out[33]: {'criterion': 'entropy',
          'max_depth': 8,
          'min_samples_leaf': 8,
          'min_samples_split': 6}
```

```
In [34]: # 3) Final RFC
H_RFC = RandomForestClassifier( criterion = 'entropy' , max_depth = 9 , min_samples_leaf = 5 , min_samples_split = 6 )
MyModel(H_RFC,"H_RFC","'criterion='entropy','max_depth'= 9,'min_samples_leaf'= 5,'min_samples_split'= 6")
print(f"Training Score :- {H_RFC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.91 | 0.95 | 0.93 | 92 |
| 1 | 0.78 | 0.79 | 0.79 | 96 |
| 2 | 0.79 | 0.80 | 0.80 | 106 |
| 3 | 0.95 | 0.90 | 0.92 | 106 |
| accuracy | | | 0.86 | 400 |
| macro avg | 0.86 | 0.86 | 0.86 | 400 |
| weighted avg | 0.86 | 0.86 | 0.86 | 400 |

```
Training Score :- 0.985625
```

DecisionTree Classifier

- 1) Default Model
- 2) Hypertuning
- 3) Final Model

```
In [35]: # 1) Default Model
DTC = DecisionTreeClassifier()
MyModel(DTC, "DTC", "Default")
print(f"Training Score :- {DTC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.92 | 0.89 | 0.91 | 92 |
| 1 | 0.76 | 0.85 | 0.80 | 96 |
| 2 | 0.81 | 0.76 | 0.79 | 106 |
| 3 | 0.91 | 0.89 | 0.90 | 106 |
| accuracy | | | 0.85 | 400 |
| macro avg | 0.85 | 0.85 | 0.85 | 400 |
| weighted avg | 0.85 | 0.85 | 0.85 | 400 |

Training Score :- 1.0

In [36]: # 2) Hypertuning DTC

```
D_DTC = DecisionTreeClassifier()
hyperparameter_dict = {"criterion":['gini','entropy'],
                      "max_depth":range(5,10),
                      "min_samples_split":range(5,10),
                      "min_samples_leaf":range(5,10)}

GS_DTC = GridSearchCV(D_DTC , param_grid = hyperparameter_dict , verbose=3)
GS_DTC.fit(X_train,Y_train)
```



```
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=7;, score=0.828 total time= 0.0s
[CV 1/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.844 total time= 0.0s
[CV 2/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.838 total time= 0.0s
[CV 3/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.875 total time= 0.0s
[CV 4/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.856 total time= 0.0s
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=8;, score=0.831 total time= 0.0s
[CV 1/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.844 total time= 0.0s
[CV 2/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.844 total time= 0.0s
[CV 3/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.875 total time= 0.0s
[CV 4/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.856 total time= 0.0s
[CV 5/5] END criterion=entropy, max_depth=9, min_samples_leaf=9, min_samples_split=9;, score=0.838 total time= 0.0s
```

```
Out[36]: GridSearchCV(estimator=DecisionTreeClassifier(),
                      param_grid={'criterion': ['gini', 'entropy'],
                                  'max_depth': range(5, 10),
                                  'min_samples_leaf': range(5, 10),
                                  'min_samples_split': range(5, 10)},
                      verbose=3)
```

```
In [37]: GS_DTC.best_params_
```

```
Out[37]: {'criterion': 'gini',
          'max_depth': 8,
          'min_samples_leaf': 7,
          'min_samples_split': 5}
```

```
In [38]: # 3) Final DTC
H_DTC = DecisionTreeClassifier( criterion = 'entropy' , max_depth = 8 , min_samples_leaf = 6 , min_samples_split = 7 )
MyModel(H_DTC,"H_DTC","'criterion='entropy','max_depth'= 8,'min_samples_leaf'= 6,'min_samples_split'= 7")
print(f"Training Score :- {H_DTC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.90 | 0.87 | 0.88 | 92 |
| 1 | 0.74 | 0.81 | 0.78 | 96 |
| 2 | 0.80 | 0.76 | 0.78 | 106 |
| 3 | 0.90 | 0.90 | 0.90 | 106 |
| accuracy | | | 0.83 | 400 |
| macro avg | 0.84 | 0.84 | 0.84 | 400 |
| weighted avg | 0.84 | 0.83 | 0.84 | 400 |

```
Training Score :- 0.939375
```

SVC

- 1) Default Model
- 2) Hypertuning
- 3) Final Model

```
In [39]: # 1) Default Model
D_SVC = SVC()
MyModel(D_SVC, "SVC", "Default")
print(f"Training Score :- {D_SVC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.97 | 0.99 | 0.98 | 92 |
| 1 | 0.92 | 0.95 | 0.93 | 96 |
| 2 | 0.91 | 0.91 | 0.91 | 106 |
| 3 | 0.97 | 0.93 | 0.95 | 106 |
| accuracy | | | 0.94 | 400 |
| macro avg | 0.94 | 0.94 | 0.94 | 400 |
| weighted avg | 0.94 | 0.94 | 0.94 | 400 |

Training Score :- 0.95375

In [40]: # 2) Hypertuning SVC

```
SVC_Rbf = SVC()
hyperparameter_dict = {"kernel":['poly', 'rbf', 'sigmoid'],
                      "C":[0.1,0.01,0.001,0.0001],
                      "gamma": [0.1,0.01,0.001,0.0001]}

GS = GridSearchCV(SVC_Rbf , param_grid = hyperparameter_dict , verbose=3)
GS.fit(X_train,Y_train)
```



```
Out[40]: GridSearchCV(estimator=SVC(),
                      param_grid={'C': [0.1, 0.01, 0.001, 0.0001],
                                  'gamma': [0.1, 0.01, 0.001, 0.0001],
                                  'kernel': ['poly', 'rbf', 'sigmoid']},
                      verbose=3)
```

```
In [41]: GS.best_params_
```

```
Out[41]: {'C': 0.0001, 'gamma': 0.0001, 'kernel': 'poly'}
```

```
In [42]: # 3) Final SVC
```

```
H_SVC = SVC( kernel = "poly" , C = 0.0001 , gamma = 0.0001 )
MyModel(H_SVC,"H_SVC","'C': 0.0001, 'gamma': 0.0001, 'kernel': 'poly'")
print(f"Training Score :- {H_SVC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 0.97 | 0.98 | 92 |
| 1 | 0.95 | 0.98 | 0.96 | 96 |
| 2 | 0.96 | 0.96 | 0.96 | 106 |
| 3 | 0.98 | 0.97 | 0.98 | 106 |
| accuracy | | | 0.97 | 400 |
| macro avg | 0.97 | 0.97 | 0.97 | 400 |
| weighted avg | 0.97 | 0.97 | 0.97 | 400 |

```
Training Score :- 0.998125
```

Conclusion

```
In [43]: Algo_df = pd.DataFrame(algo_performance_list,columns=[ "ALgo", 'Accuaracy', 'Params', 'Training Score'])
Algo_df
```

```
Out[43]:
```

| | ALgo | Accuaracy | Params | Training Score |
|---|-----------|-----------|---|----------------|
| 0 | KNN | 0.9125 | Default | 0.958125 |
| 1 | H_KNN | 0.9250 | n_neighbors = 9 | 0.948125 |
| 2 | LR | 0.6150 | Default | 0.641250 |
| 3 | H_Log_Reg | 0.9625 | solver = newton-cg , penalty = none | 0.974375 |
| 4 | RFC | 0.8725 | Default | 1.000000 |
| 5 | H_RFC | 0.8575 | 'criterion'='entropy','max_depth'= 9,'min_samp... | 0.985625 |
| 6 | DTC | 0.8475 | Default | 1.000000 |
| 7 | H_DTC | 0.8350 | 'criterion'='entropy','max_depth'= 8,'min_samp... | 0.939375 |
| 8 | SVC | 0.9425 | Default | 0.953750 |
| 9 | H_SVC | 0.9700 | 'C': 0.0001, 'gamma': 0.0001, 'kernel': 'poly' | 0.998125 |

Best Algo

```
In [44]: Algo_df[Algo_df.Accuaracy == Algo_df.Accuaracy.max()]
```

Out[44]:

| ALgo | Accuracy | Params | Training Score |
|---------|----------|--|----------------|
| 9 H_SVC | 0.97 | 'C': 0.0001, 'gamma': 0.0001, 'kernel': 'poly' | 0.998125 |

Random Sample Testing

```
In [45]: H_SVC = SVC( kernel = "poly" , C = 0.0001 , gamma = 0.0001 )
MyModel(H_SVC,"H_SVC","'C': 0.0001, 'gamma': 0.0001, 'kernel': 'poly'")
print(f"Training Score :- {H_SVC.score(X_train,Y_train)}")
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.99 | 0.97 | 0.98 | 92 |
| 1 | 0.95 | 0.98 | 0.96 | 96 |
| 2 | 0.96 | 0.96 | 0.96 | 106 |
| 3 | 0.98 | 0.97 | 0.98 | 106 |
| accuracy | | | 0.97 | 400 |
| macro avg | 0.97 | 0.97 | 0.97 | 400 |
| weighted avg | 0.97 | 0.97 | 0.97 | 400 |

Training Score :- 0.998125

```
In [46]: def predict_range(value):
    if value == 0:
        print("Price Is Low")

    elif value ==1:
        print('Price is Medium')

    elif value == 2:
        print('Price is High')

    else:
        print('Price is Very High')
```

```
In [47]: # PR-0
s0 = np.array([df.loc[7].values[:-1]])
sa0 = H_SVC.predict(s0)
predict_range(sa0[0])
```

Price Is Low

```
In [48]: # PR-1 (Testing on record present in Dataset)
s1 = np.array([df.loc[0].values[:-1]])
sa1 = H_SVC.predict(s1)
predict_range(sa1[0])
```

Price is Medium

```
In [49]: # PR-2 (Testing on record present in Dataset)
s2 = np.array([df.loc[1].values[:-1]])
sa2 = H_SVC.predict(s2)
predict_range(sa2[0])
```

Price is High

```
In [50]: # PR-3 (Testing on record present in Dataset)
s3 = np.array([df.loc[6].values[:-1]])
sa3 = H_SVC.predict(s3)
predict_range(sa3[0])
```

Price is Very High

```
In [ ]:
```

```
In [51]: # (Testing on Random Values)
mobile=[842,0,2.2,0,1,0,7,0.6,188,2, 2,20,756,2549,9,7,19,0, 0, 1]
mobile = np.array([mobile])
value = H_SVC.predict(mobile)
predict_range(value)
```

Price is Medium

```
In [ ]:
```

```
In [ ]:
```