LUXURY WATCHES

INTRODUCTION

The dataset Luxury watches which is been collected from Kaggle contains the data of watches and their prices in U.S dollar. The data provides all the external features of the watch like the model,reference value , complication that whether the watch is Automatic,Quartz or Hand-Wound, Case material ,Bracelet material, Dial of the watch, Hour Markings and the Lunnete material of the watch.

OBJECTIVE

The main objective of analysing this dataset is to find some watches with largest prices all over the world and also to find watch with minimum price.

Finding which complication of watch is high in price. The top brand in the Luxury watches etc.

DATA OVERVIEW

- Dataset name: Luxury watches
- Purpose : Dataset contains the Luxury watches their features and prices
- Total Number of Rows: 163598
- Total Number of columns: 10
- Column name and their info:

#	Column	Non-Null Count	Dtype
0	Brand	163598 non-null	object
1	Model	163598 non-null	object
2	Reference	163597 non-null	object
3	Complication	163598 non-null	object
4	Case material	163598 non-null	object
5	Bracelet material	163598 non-null	object
6	Dial	163598 non-null	object
7	Hour Markings	163598 non-null	object
8	Lunette Material	163598 non-null	object
9	Price	163598 non-null	int64

- **Data cleaning**
    1. Missing value handling : 1 Missing value from the table is found and is removes using dropna method.
    2. Duplicate value handling: 30959 duplicate value is found from the dataset and is removes using the drop duplicate method.
    3. Rows after data cleaning is 132638

- **Feature enginnering:**
    1. A new column named Price(Rs) is created by converting the price in usd to Indian Rupees.

DATA ANALYSIS

Univariate Analysis

1. Finding the mean, median ,Standard Deviation, Varience, Min ,Max from the price(Rs)

```
for i in num1:
  print('_____')
  print(f'Mean of price:{df[i].mean()}')
  print(f'Median of price:{df[i].median()}')
  print(f'Standard deviation of price:{df[i].std()}')
  print(f'Varience of price:{df[i].var()}')
  print(f'Max of price:{df[i].max()}')
  print(f'Min of price:{df[i].min()}')
```

Result:

_____

Mean of price:1671024.0913775088

Median of price:757305.12

Standard deviation of price:3742501.596276389

Varience of price:14006318198131.32

Max of price:347482654.09000003

Min of price:4471.610000000001

2. Finding the watch with high price

```
df.loc[df['Price(Rs)'].idxmax()]
```

Result:

	31937
Brand	**Patek Philippe**
Model	**Minute Repeater**
Reference	**6301P-001**
Complication	**Automatic**
Case material	**Platinum**
Bracelet material	**Crocodile-Leather**
Dial	**Black**
Hour Markings	**Arabic**
Lunette Material	**Platinum**
Price	**4118557**
Price(Rs)	**347482654.09**

Insights:

The watch Patek Phillippe Minute Repeater is the Top Luxury watch With price 347482654.09.

3. Finding watch with minimum price

df.loc[df['Price(Rs)'].idxmin()]

Result:

	63795
Brand	Seiko
Model	5
Reference	SNXS77
Complication	Automatic
Case material	Steel
Bracelet material	Steel
Dial	Blue
Hour Markings	No Markings
Lunette Material	Steel
Price	53
Price(Rs)	4471.61

Insights:

The watch Seiko 5 is the budget watch with price(Rs) 4471.61.

4. Finding the Value counts and the number of unique values in all the categorical columns.

```
for i in obj1:
  print('----------------------')
  print(f'{i}: {df[i].value_counts()}')
  print(f'unique count of {i}: {df[i].nunique()}')
```

Result:

```
----------------------
Complication: Complication
Automatic     115364
Quartz          8790
Hand-wound      8484
Name: count, dtype: int64
unique count of Complication: 3
----------------------
Case material: Case material
Steel          84776
Gold/Steel     12474
Rose Gold       9233
Yellow Gold     8358
White Gold      5722
Titanium        5178
Ceramic         3557
Platinum        1191
Plastic          635
Carbon           615
Bronze           587
Silver           205
Aluminium         67
Tantalum          24
Palladium          8
Tungsten           8
Name: count, dtype: int64
unique count of Case material: 16
----------------------
Bracelet material: Bracelet material
Steel              60399
```

Leather 17979
Rubber 13301
Gold/Steel 12236
Crocodile-Leather 9955
Yellow Gold 4870
Textile 3294
Rose Gold 3250
White Gold 2082
Titanium 1674
Cow-Leather 1059
Platinum 632
Plastic 599
Ceramic 580
Silicone 466
Reptile-Leather 84
Ostritch Leather 64
Satin 41
Snake-Leather 31
Shark-Leather 21
Silver 11
Aluminium 10
Name: count, dtype: int64
unique count of Bracelet material: 22

Dial: Dial
Black 48754
Blue 22127
Silver 15771
White 14560
Gray 8367
Green 5689
Brown 3707
Champaign 3293
Gold 3055
Transparent 2310
Pink 1389
Pearl 1246
Red 667
Yellow 519
Solid Silver 267

Orange 262
Bronze 244
Purple 216
Solid Gold 106
Bordeaux 89
Name: count, dtype: int64
unique count of Dial: 20

Hour Markings: Hour Markings
No Markings 91395
Arabic 27956
Roman 13287
Name: count, dtype: int64
unique count of Hour Markings: 3

Lunette Material: Lunette Material
Steel 64241
Ceramic 21010
Yellow Gold 16060
White Gold 10994
Rose Gold 10155
Titanium 4060
Platinum 1866
Aluminium 1031
Gold/Steel 884
Carbon 596
Bronze 576
Plastic 445
Rose gold 416
Silver 212
Tantalum 49
Tungsten 35
Palladium 8
Name: count, dtype: int64
unique count of Lunette Material: 17

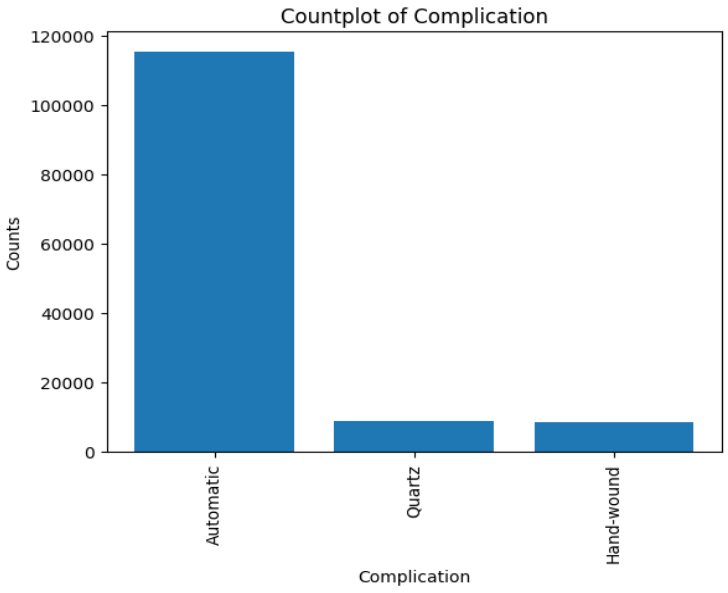
Insights:

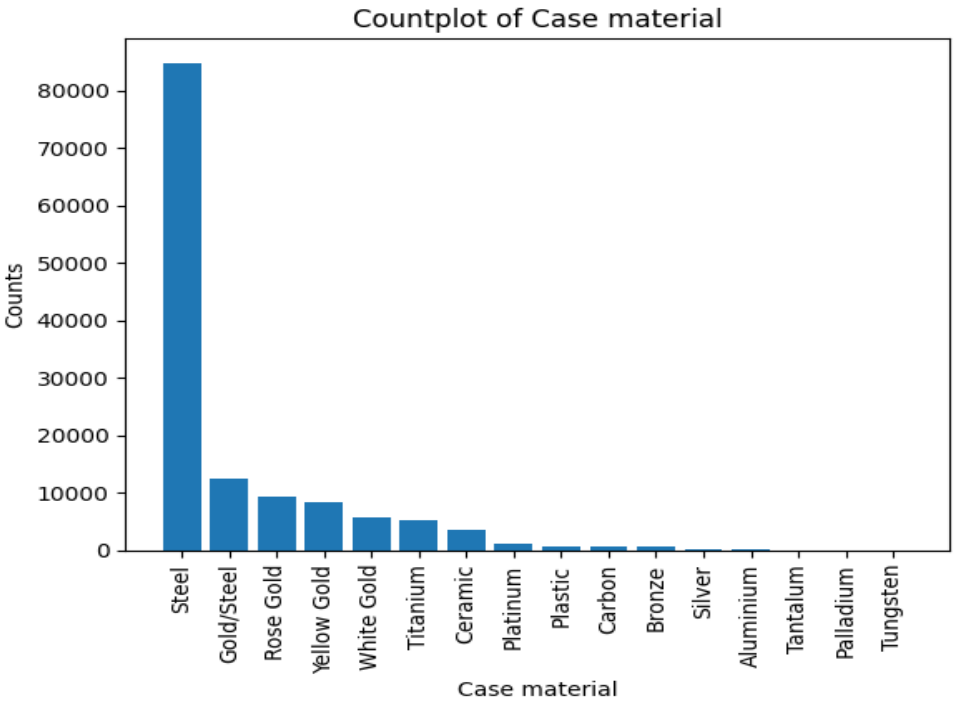
- In the complication Automatic watches are high in number around 115364
- In case material Steel and Gold/Steel is high in number around 84776 and 12474
- In Bracelet material Steel is high in number around 60399
- In Dial black is high in number around 48754
- In Hour Marking No marking watch is high in number around 91395
- In Lunnette Material Steel watch is high around 64241

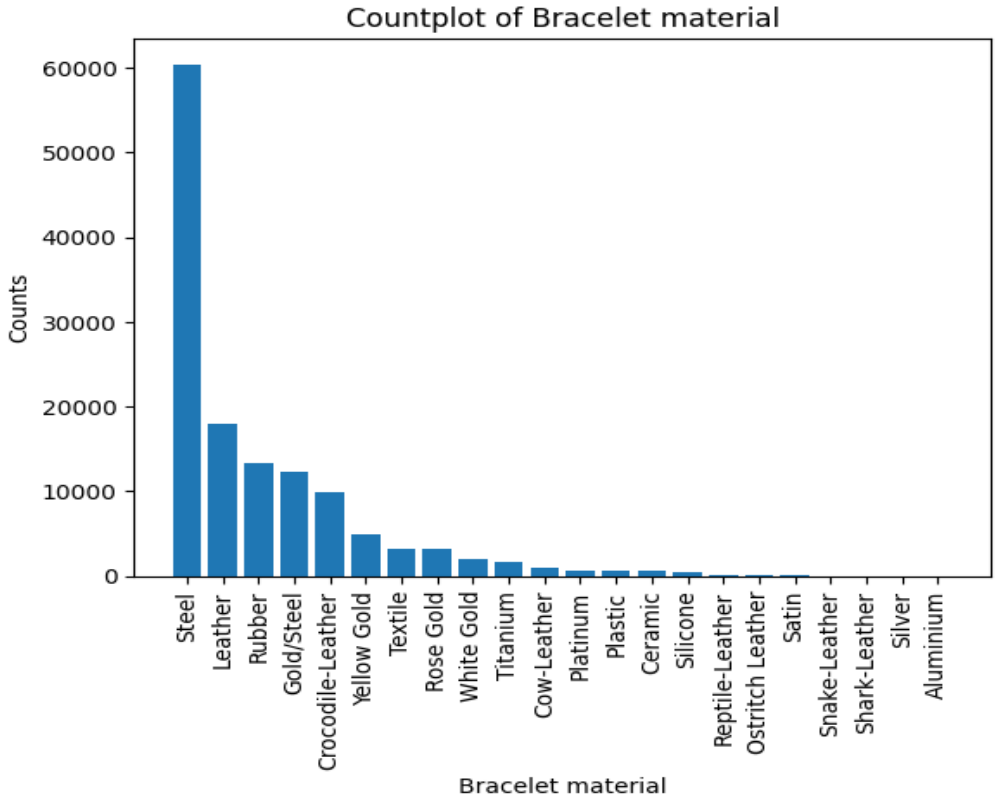
5. Bar plots of the Categorical columns

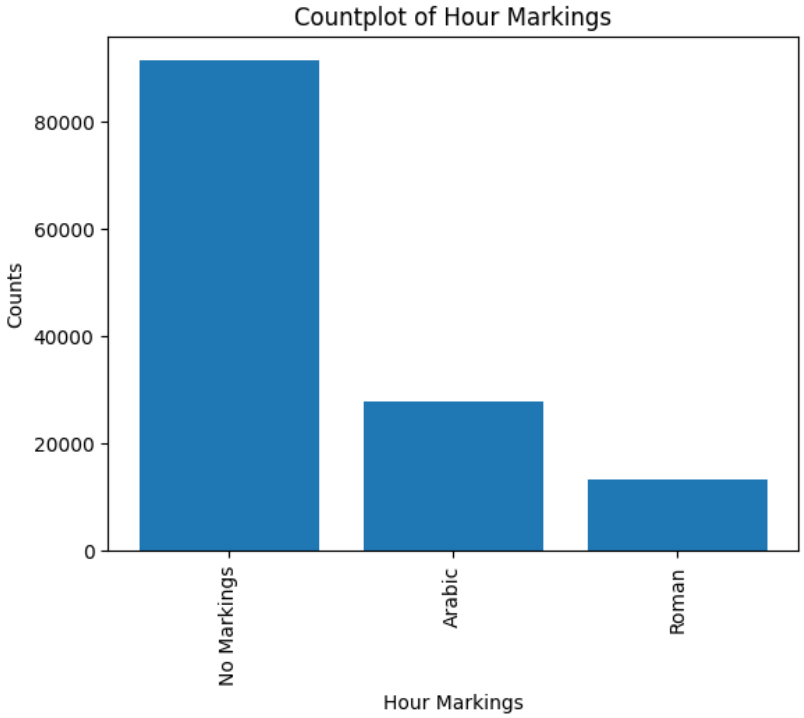
```
import matplotlib.pyplot as plt
import seaborn as sns

for i in obj1:
  if df[i].nunique()<=25:
    plt.bar(x=df[i].value_counts().index,height=df[i].value_counts())
    plt.title(f'Countplot of {i}')
    plt.xticks(rotation=90)
    plt.xlabel(i)
    plt.ylabel('Counts')
    plt.show()
```

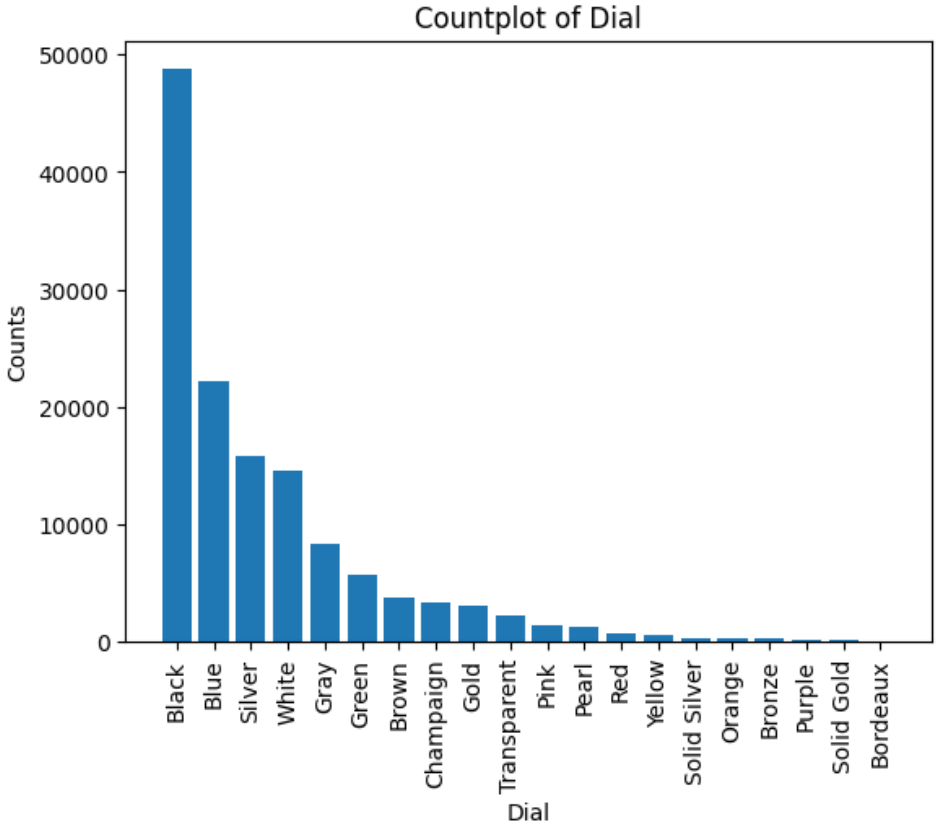
Result:

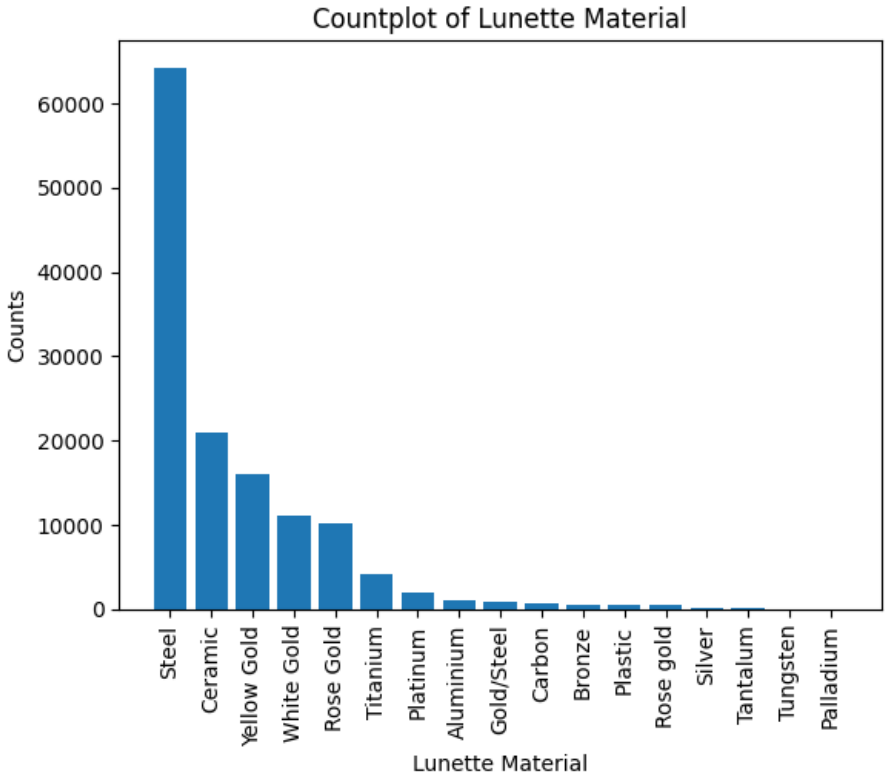



Countplot of Case material
Counts
80000
70000
60000
50000
40000
30000
20000
10000
0
Steel
Gold/Steel
Rose Gold
Yellow Gold
White Gold
Titanium
Ceramic
Platinum
Plastic
Carbon
Bronze
Silver
Aluminium
Tantalum
Palladium
Tungsten
Case material


Countplot of Bracelet material
Counts
60000
50000
40000
30000
20000
10000
0
Steel
Leather
Rubber
Gold/Steel
Crocodile-Leather
Yellow Gold
Textile
Rose Gold
White Gold
Titanium
Cow-Leather
Platinum
Plastic
Ceramic
Silicone
Reptile-Leather
Ostritch Leather
Satin
Snake-Leather
Shark-Leather
Silver
Aluminium
Bracelet material


Countplot of Hour Markings
Counts
80000
60000
40000
20000
0
No Markings
Arabic
Roman
Hour Markings

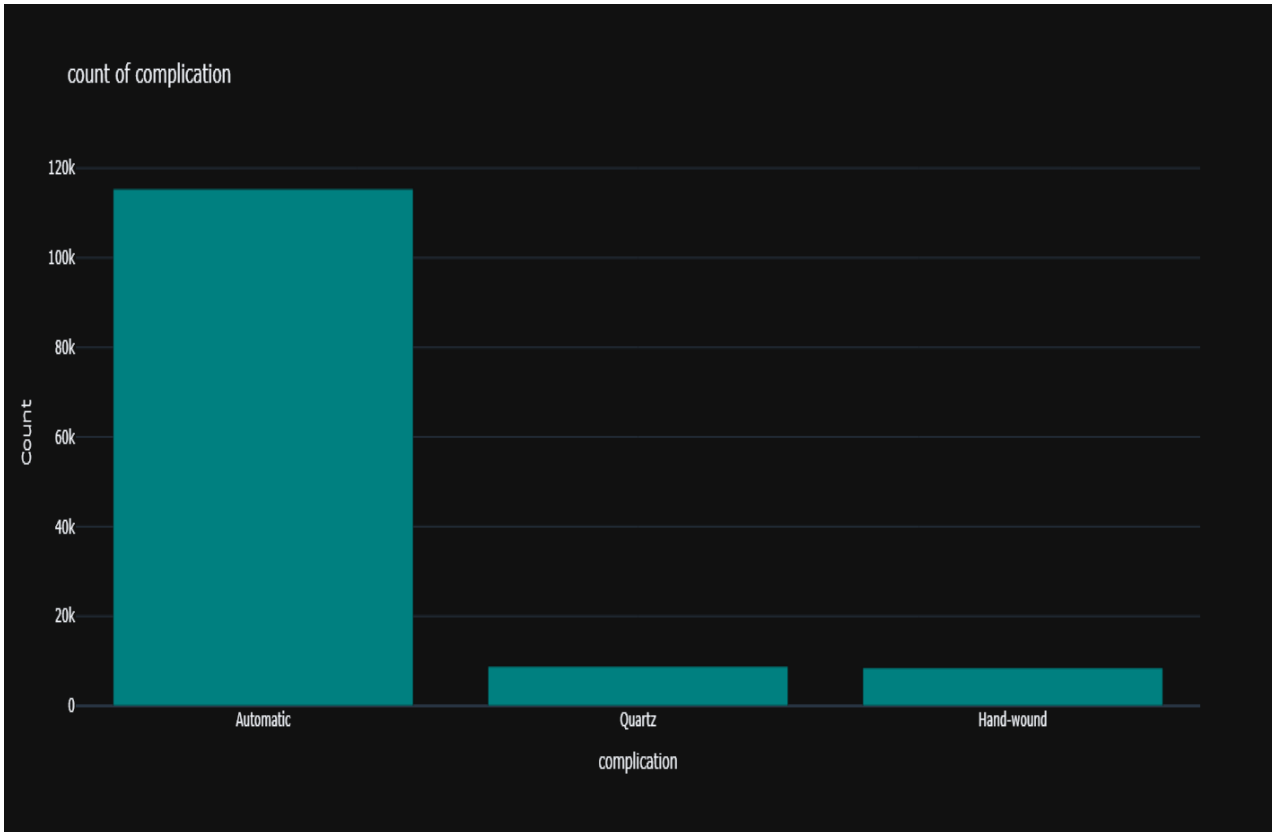

Countplot of Dial
Counts
50000
40000
30000
20000
10000
0
Black
Blue
Silver
White
Gray
Green
Brown
Champaign
Gold
Transparent
Pink
Pearl
Red
Yellow
Solid Silver
Orange
Bronze
Purple
Solid Gold
Bordeaux
Dial



6. Bar plot using plotly.graph_objects

```
import plotly.graph_objects as go
v1=df['Complication'].value_counts()
fig=go.Figure([go.Bar(x=v1.index,y=v1,marker=dict(color='teal'))])
fig.update_layout(
    title='count of complication',
    xaxis_title='complication',
    yaxis_title='Count',
    template='plotly_dark'

)

fig.show()
```

Result:



Bivariate Analysis

1. Finding Top 10 watches with low price

   df[[‘Brand’,’Price(Rs)’]].sort_values(by='Price(Rs)').head(10)

   Result:

	Brand	Model	Reference	Price(Rs)
63795	Seiko	5	SNXS77	4471.61
90946	Casio	G-Shock	G-2900F1V	4555.98
84672	Casio	G-Shock	GA-2100-1A1ER	5737.16
91550	Casio	G-Shock	GA-100BP-1A	5737.16
90942	Casio	G-Shock	DW9052-1BCR	5821.53
44967	Locman	Sport Tonneau	487	6159.01
84671	Tissot	Seastar	315T	6327.75
104784	Casio	Edifice	EFV-100D-2AVUEF	6412.12
132318	Casio	G-Shock	DW-5600LS-7DR	6412.12
58400	Seiko	Kinetic	5M43-0A70	6496.49

2. Finding Top 10 watches with high prices

df[['Brand','Price(Rs)']].sort_values(by='Price(Rs)',ascending=False).head(20

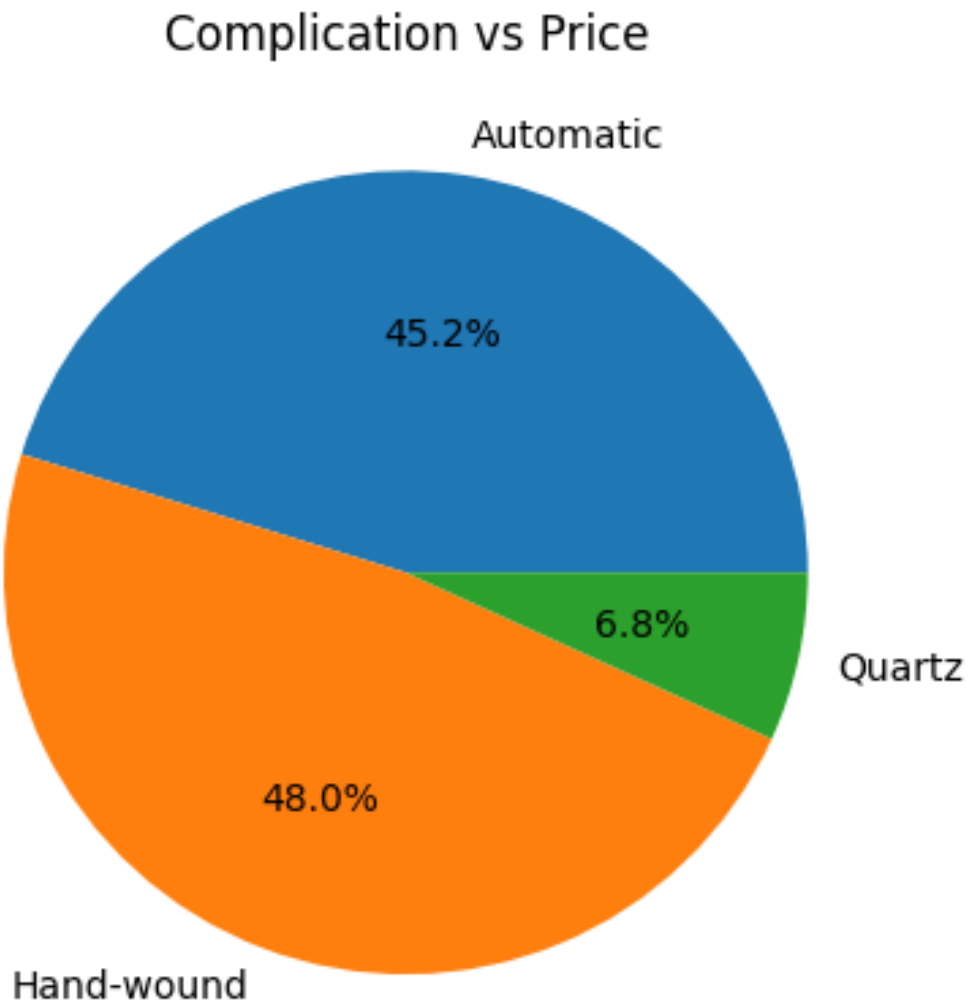
Result:

	Brand	Model	Reference	Price(Rs)
31937	Patek Philippe	Minute Repeater	6301P-001	347,482,700
62520	Patek Philippe	Minute Repeater Perpetual Calender	5304301R-001	2283,257,500
31888	Patek Philippe	Nautilus	57111A-018	263,607,600
131622	Patek Philippe	Nautilus	57111A-018	230,677,200
87346	Patek Philippe	Nautilus	57111A-018	215,417,500
126665	Patek Philippe	Nautilus	57111A-018	214,200,300
61520	Patek Philippe	Nautilus	57111A-018	205,862,800
62519	Patek Philippe	Minute Repeater	5303R-001	158,363,000
31338	Patek Philippe	Nautilus	59901400G-011	127,018,800
62514	Audemars Piguet	Royal Oak Offshore	26582CB.OO.A010CA.01	104,486,900

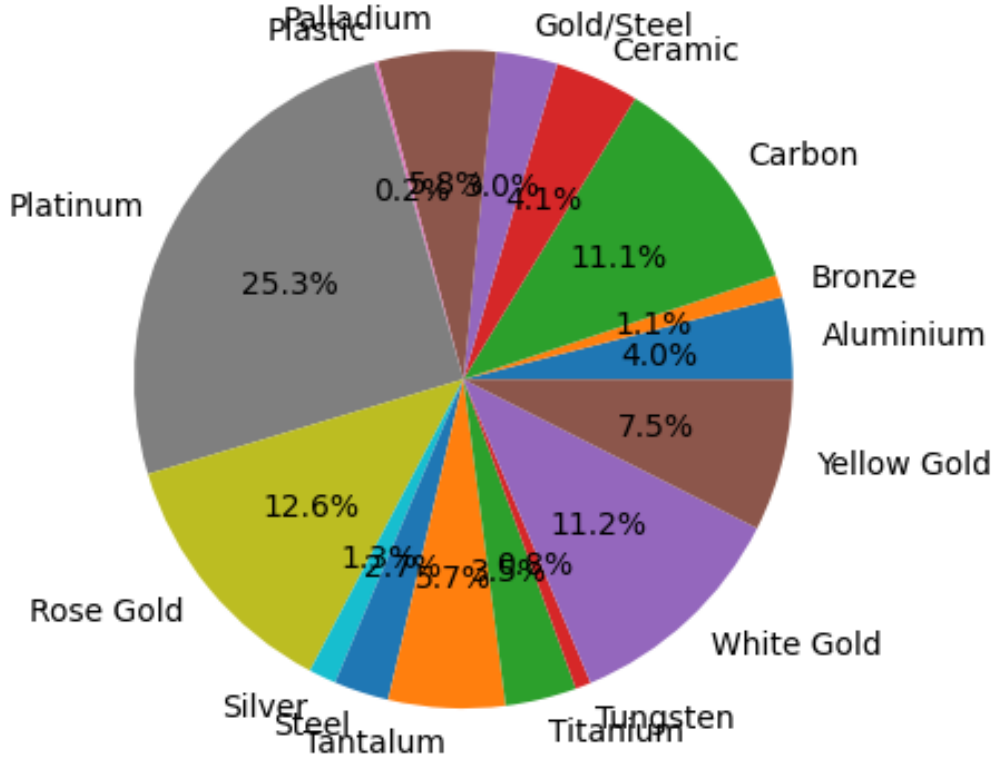
3. Pie plot of Price(Rs) and Categorical columns

```
for i in obj1:
  if df[i].nunique()<=17:
    plt.pie(df.groupby(i)['Price(Rs)'].mean(),labels=df.groupby(i)['Pric
e(Rs)'].mean().index)
    plt.title(f'{i} vs Price(Rs)')
    plt.show()
```

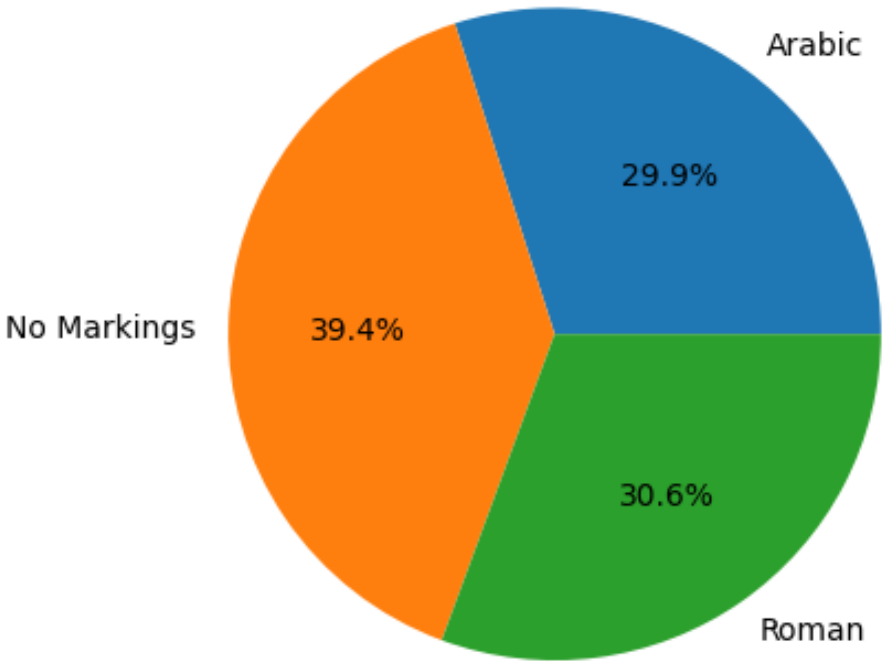
Result:

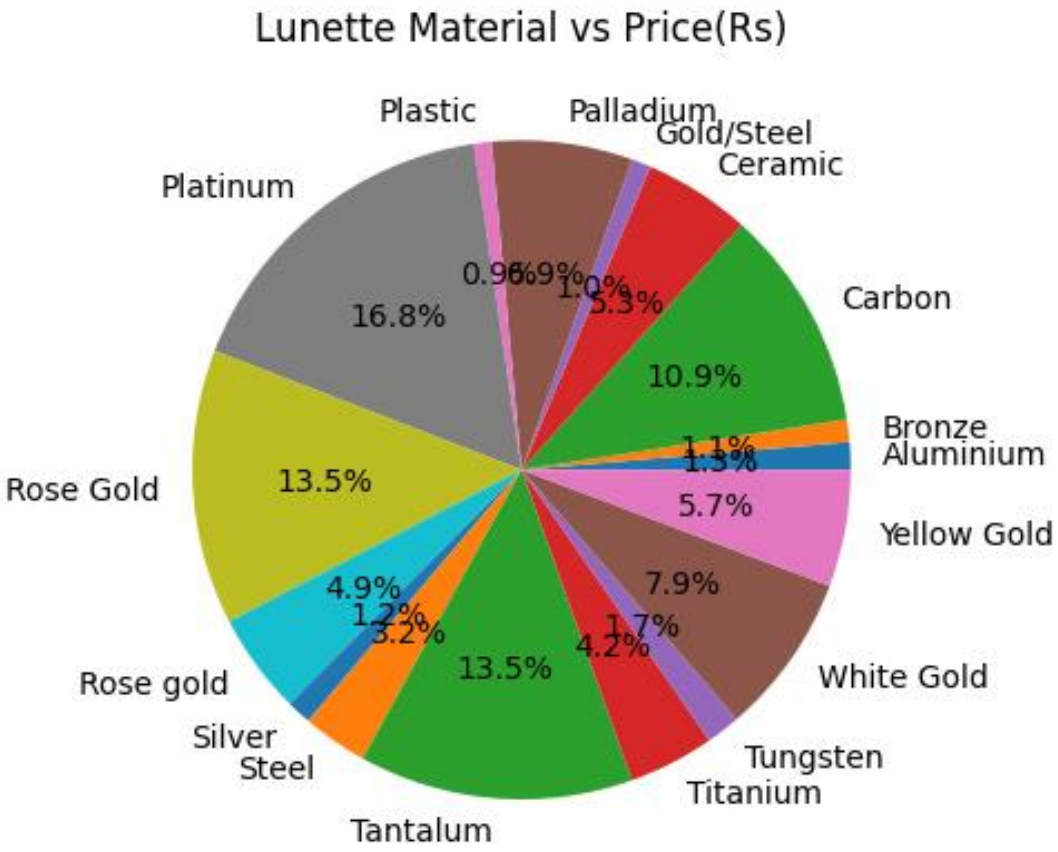


Case material vs Price(Rs)



Hour Markings vs Price(Rs)





Insights:

- In complication the Price of Hand Wound watch is high in number and then Automatic watch.
- In Case material Platinum watch has high rate and then White gold.
- In Hour marking No marking watch has higher rates then Roman.
- In Lunnette material Platinum has higher rates and then Tantalum.

CONCLUSIONS

1. The Top 10 Watches with High rates in this 9 of them are of the Brand **Patek Phillippe** and the 10th one is **Audemars Piguet.**
2. The 10 watches with low rates the main brand which provide Budget Watches are **Casio** with 6 watches and then the **Seiko** with 2 watches And there is also two watches of brands **Locman** and **Tissot**.

3. Patek Phillippe,Minute Repeater ,6301P-001,Automatic watch with Platinum Case ,Crocodile Leather ,Arabic Hour Marking is the watch with 347482654 Indian Ruppes.
4. Seiko 5,SNXS77,Automatic watch with Steel case and Bracelet with no Hour Marking has the Price 4471.61 which is the lowest price watch in this dataset.
5. Comparing the price the watches which are Hand wounded ,with case Platinum,No Marking and Lunnete Material Platinum watch has highest rates .

REFERENCES

Link of the Dataset : https://www.kaggle.com/datasets/yoerireumkens/timepiece-treasures-a-luxury-watches-dataset