

---

---

# Exploratory Data Analysis on Smartphones data

— Dipti Soni —

---

---

# Introduction

Want to buy a smartphone ?

This notebook, is an Exploratory data analysis on smartphone price range and how hardware of a smartphone can affect the price of a smartphone.

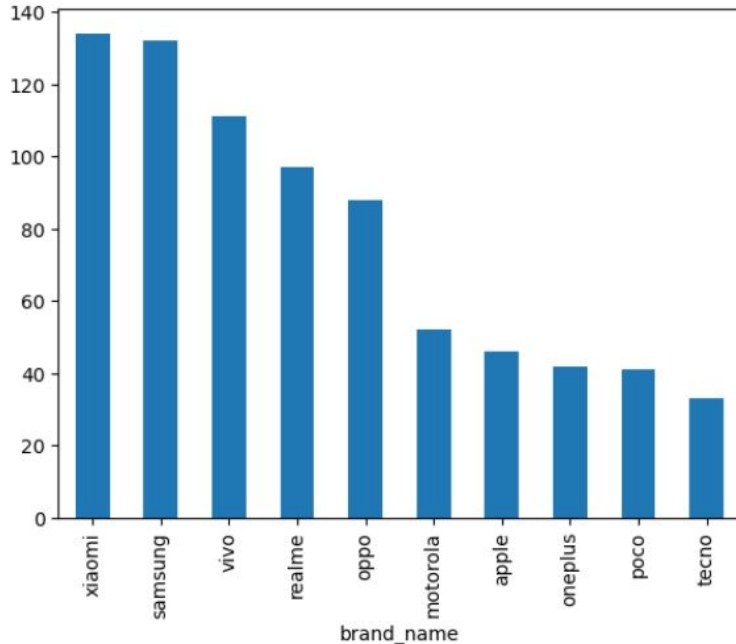
We have collected data of 980 different smartphones available in the market. The dataset talks about 25 attributes of smartphones.

- Brand Name
- Model
- Price
- Rating
- Has\_5g
- Has\_nfc
- has\_ir\_blaster
- Processor\_brand
- Num\_cores
- processor\_speed
- battery\_capacity
- fast\_charging\_available
- fast\_charging
- ram\_capacity
- internal\_memory
- screen\_size
- refresh\_rate
- resolution
- Rear and front cameras
- os
- Extended memory

# Top 10 most used brands and market share in %

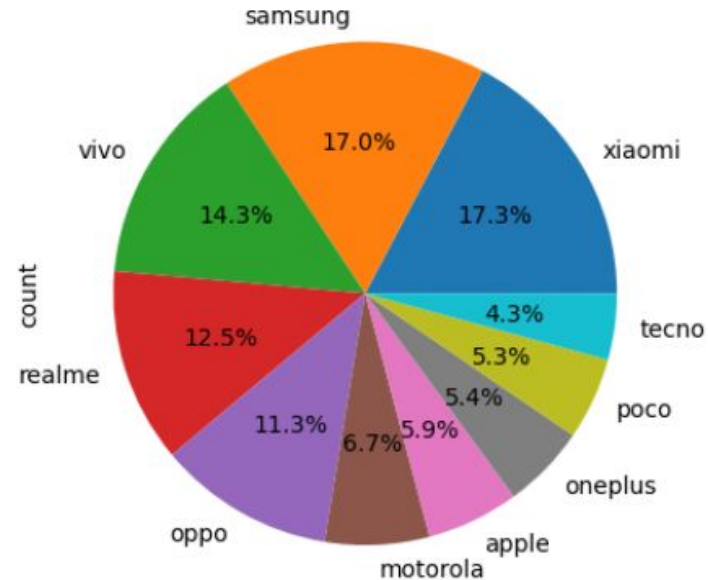
```
# plot a graph of top 10 brands  
df['brand_name'].value_counts().head(10).plot(kind='bar')
```

<Axes: xlabel='brand\_name'>



```
# pie chart  
df['brand_name'].value_counts().head(10).plot(kind='pie', autopct='%0.1f%%')
```

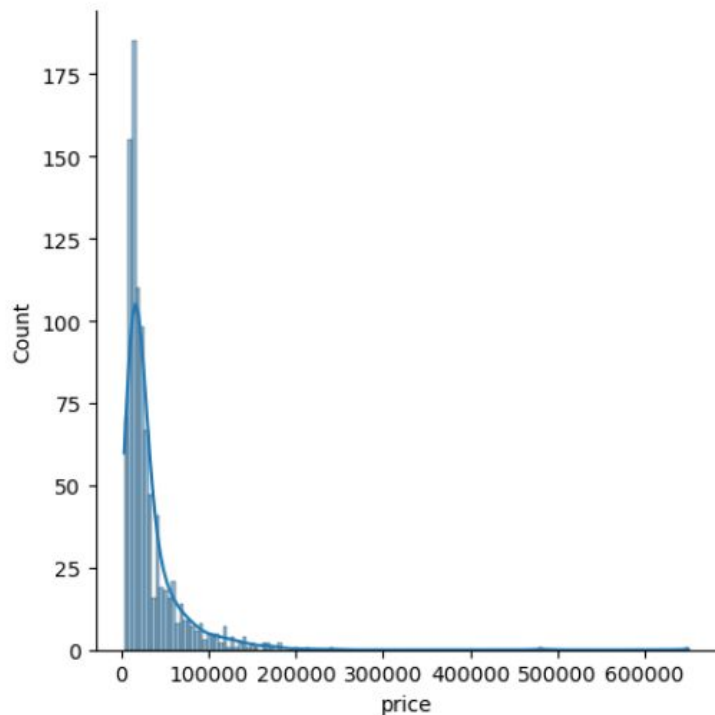
<Axes: ylabel='count'>



# Understanding Price Range

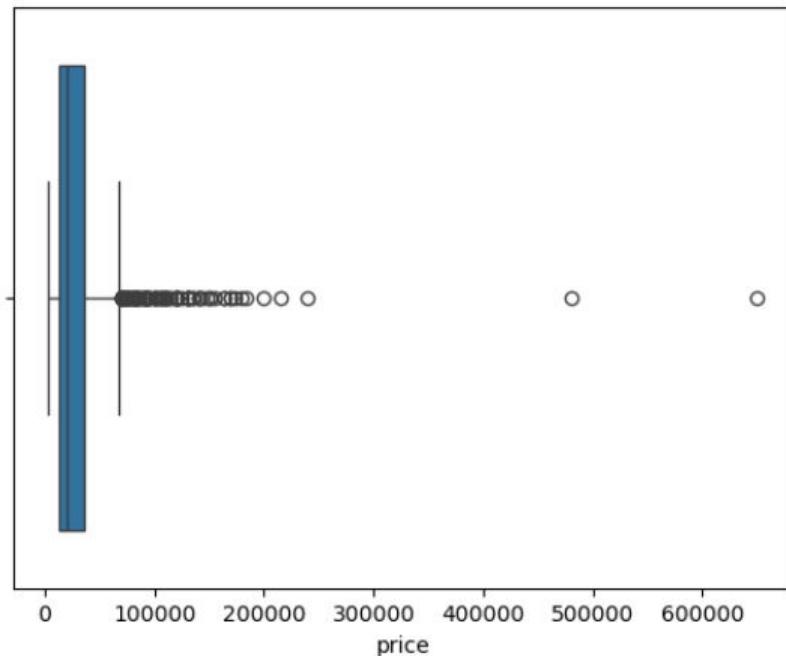
```
sns.displot(kind='hist',data=df,x='price',kde=True)
```

<seaborn.axisgrid.FacetGrid at 0x259aca5b110>



```
sns.boxplot(df['price'],orient="h")
```

<Axes: xlabel='price'>



# Univariate Analysis on price column

- The distribution is not normal, data is skewed
- Most of the phones have very less price
- Less number of phones have higher range of prices
- Definitely there are outliers
- There are no missing values

```
df['price'].isnull().sum()
```

0

```
df['price'].skew()
```

6.591790999665567

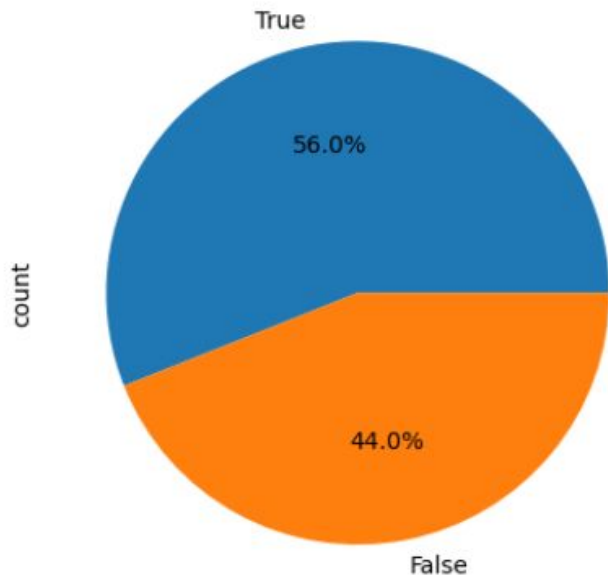
```
# Outliers  
df[df['price'] > 200000]
```

	brand_name	model	price	rating	has_5g	has_nfc	has_ir_blaster
427	vertu	Vertu Signature Touch	650000	62.0	False	True	False
478	huawei	Huawei Mate 50 RS Porsche Design	239999	81.0	False	True	True

# Univariate Analysis on Categorical Columns

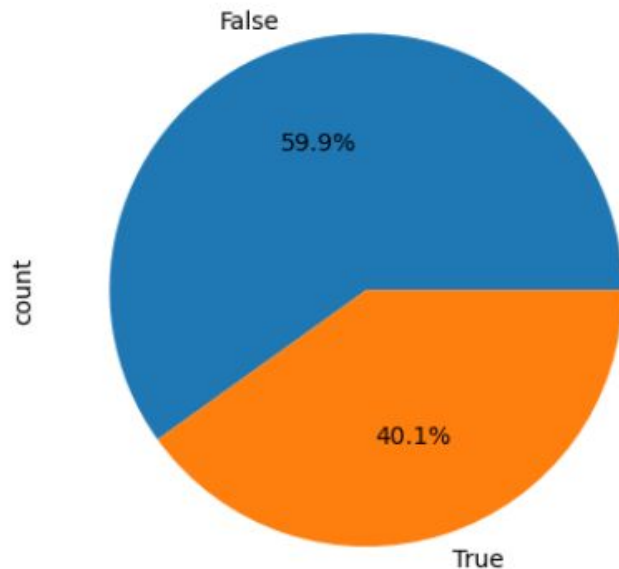
```
# has_5g  
df['has_5g'].value_counts().plot(kind='pie', autopct='%0.1f%%')
```

<Axes: ylabel='count'>



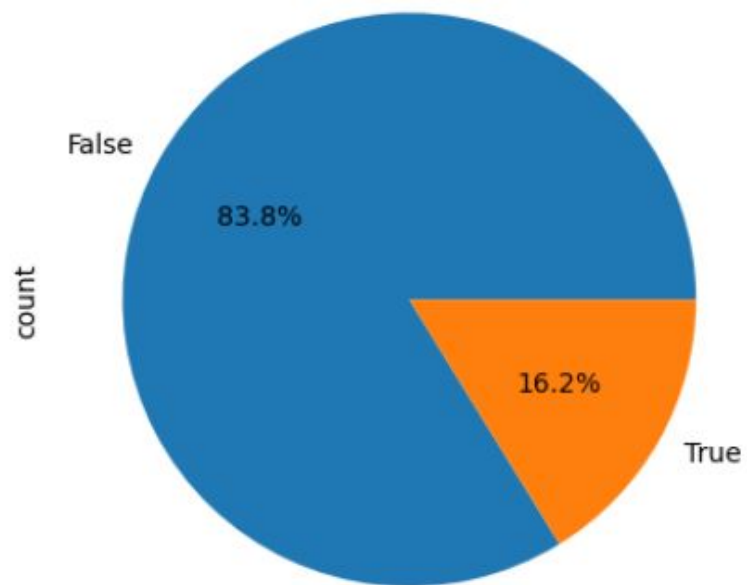
```
# has_nfc  
df['has_nfc'].value_counts().plot(kind='pie', autopct='%0.1f%%')
```

<Axes: ylabel='count'>



```
# ir_blaster  
df['has_ir_blaster'].value_counts().plot(kind='pie', autopct='%0.1f%%')
```

<Axes: ylabel='count'>



# Observations on columns has\_5g, has\_nfc, has\_ir\_blaster

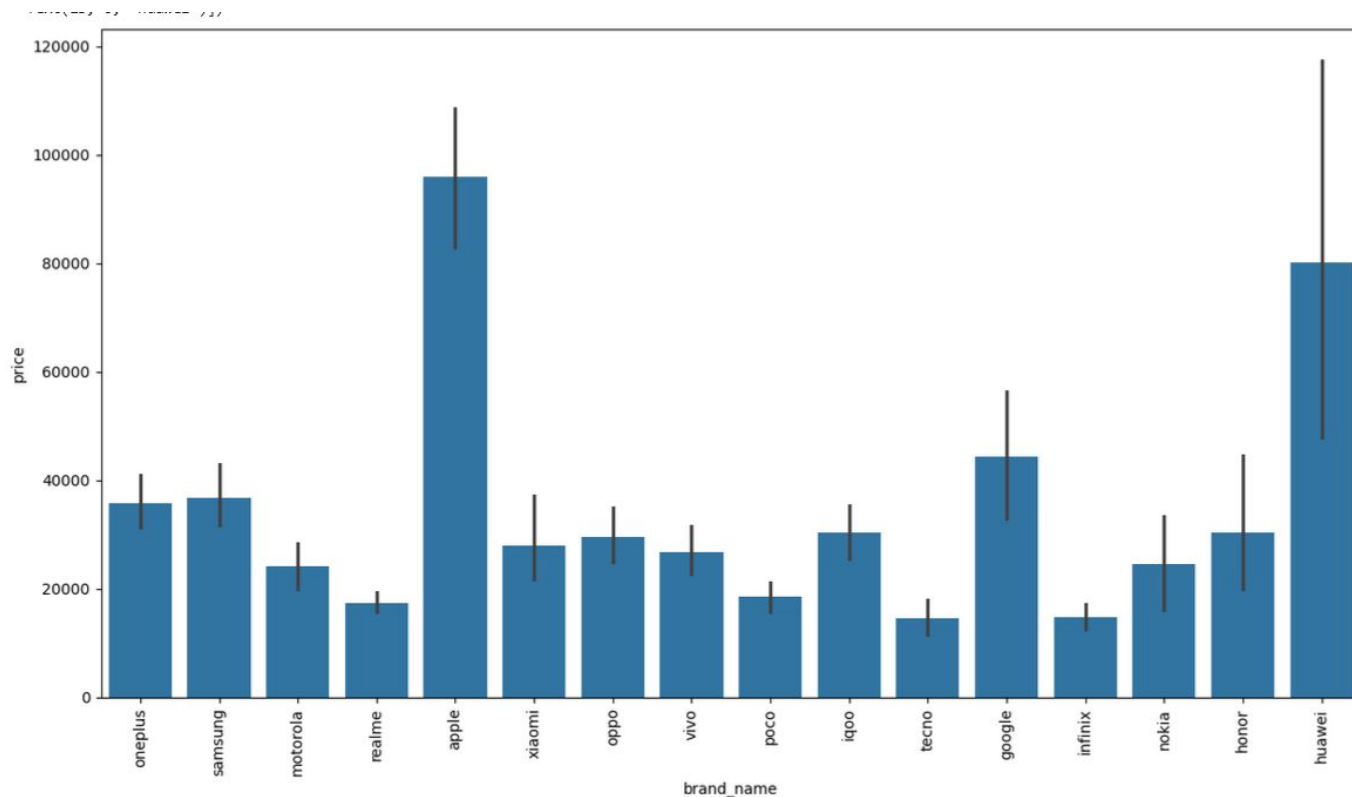
- Approximately 50% phones have 5G enabled
- 40 % phones have nfc
- 16% phones have ir blaster, mostly chinese phones provide this ir blaster feature

```
df[df['has_ir_blaster'] == True]['brand_name'].value_counts()
```

```
brand_name
xiaomi      109
poco        30
iqoo         6
huawei        6
vivo         4
redmi        2
honor        1
samsung      1
Name: count, dtype: int64
```



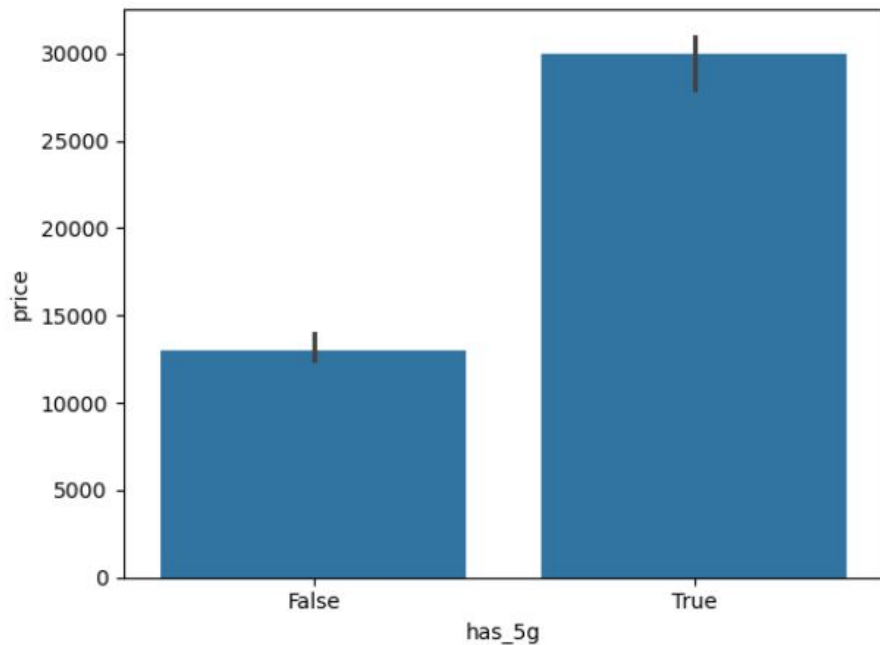
# Price vs Brand name (Bivariate Analysis)



# 5g enabled / nfc phones are costlier

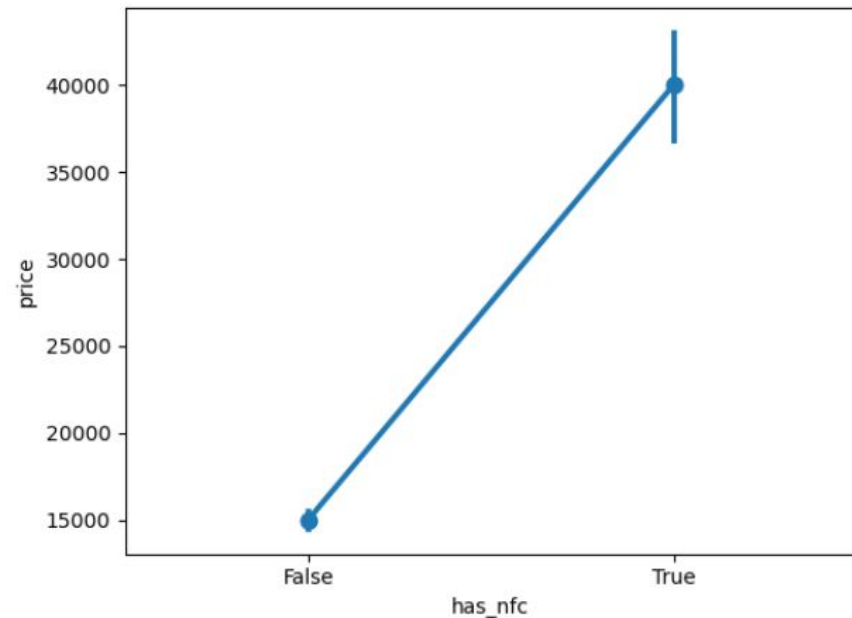
```
# price vs has_5g
sns.barplot(data=temp_df, x='has_5g', y='price', estimator=np.median)
```

<Axes: xlabel='has\_5g', ylabel='price'>



```
# price vs has_nfc
sns.pointplot(data=temp_df, x='has_nfc', y='price', estimator=np.median)
```

<Axes: xlabel='has\_nfc', ylabel='price'>



# Correlation of price with other columns

```
df.corr(numeric_only=True)
```

	price
price	1.000000
rating	0.283504
has_5g	0.305066
has_nfc	0.470951
has_ir_blaster	-0.015807
num_cores	-0.048561
processor_speed	0.474049
battery_capacity	-0.159232

fast_charging_available	0.116739
fast_charging	0.277591
ram_capacity	0.386002
internal_memory	0.557168
screen_size	0.113253
refresh_rate	0.244115
num_rear_cameras	0.125330
num_front_cameras	0.115228
primary_camera_rear	0.092095
primary_camera_front	0.162995
extended_memory_available	-0.448628
extended_upto	0.091945