Real World Smartphone Dataset

Dataset:

Contains 21 features and 1 output variable – price and have 980 instances.

Dataset does contains some missing values and have both numerical and categorical variables.

Dataset does not contain any duplicate entries.

Imputation:

1. Imputation of avg_rating by mean per Brand

The **average rating** of a smartphone is primarily influenced by the brand's reputation, build quality, software, and customer satisfaction.

2. Imputation of processor_speed by mode per Processor Brand

Each processor brand produces chips with a specific range of speeds, meaning that the most common speed within each brand is a reasonable assumption for missing values. Using the **mode** (most frequent value) is better than the mean here because processor speeds are usually standardized at specific frequencies (e.g., 2.0 GHz, 2.2 GHz, 2.8 GHz).

3. Imputation of fast_charging by mode per Brand

Whether a smartphone has **fast charging** is strongly related to the brand's design decisions.Certain brands (e.g., OnePlus, Realme) frequently include fast charging, while others (e.g., Apple) may not.

4. Imputation of num_cores by mode per Processor Brand

The **number of cores** is defined by the processor manufacturer (e.g., Qualcomm, MediaTek). Most processors from the same brand tend to follow a common pattern (e.g., MediaTek often uses octa-core, while Apple uses hexa-core).

5. Imputation of primary_camera_front by mode per Brand

The **front camera resolution** is often similar across models from the same brand. Brands follow specific design choices for front cameras (e.g., Apple often has 12 MP, budget brands may have 8 MP).

6. Imputation of processor_brand by mode per Brand Name

The **processor brand** (e.g., Qualcomm, MediaTek, Exynos) is typically associated with a smartphone manufacturer. Most smartphone brands use a limited set of processor brands.

7. Imputation of OS by mode per Brand Name

The **operating system (`os`)** is strongly tied to the **smartphone brand**.

EDA:

Give a brief statistics of the each column like mean , median , mode , min , max and percentile at 25 , 50 , 75 % .:

Histograms and Density Plots:
Plotting the distributions of key numerical features (e.g., price, battery capacity) can indicate skewness in the data. For example, a right-skewed price distribution might reveal that a small subset of smartphones is priced significantly higher than the rest.

Correlation Heatmaps:
A correlation matrix visualized via a heatmap (using Seaborn) helps in understanding interdependencies among features. This can highlight, for instance, if higher battery capacity correlates with higher prices or if larger screen sizes associate with better camera resolutions.

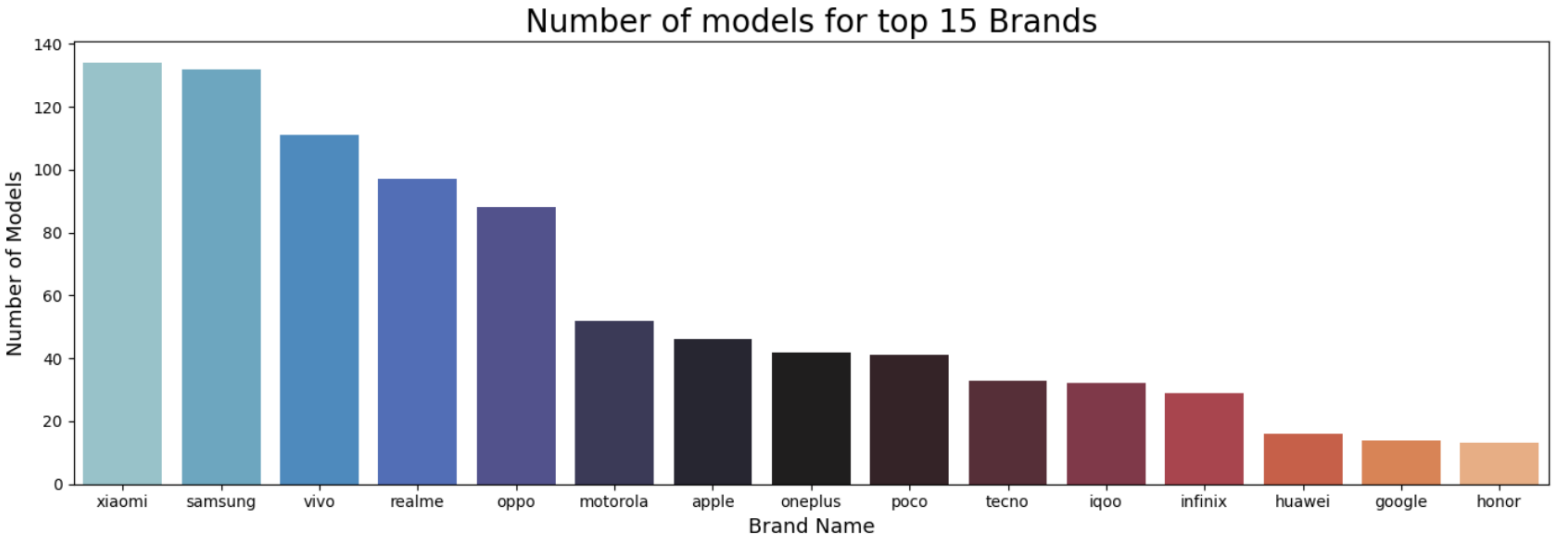
Scatter and Box Plots:
Scatter Plots: Used to observe relationships between two continuous variables (e.g., RAM vs. price) to spot trends or clusters.
Box Plots: Aid in detecting outliers and understanding the spread of the data across different categories (e.g., comparing price distributions across various brands).

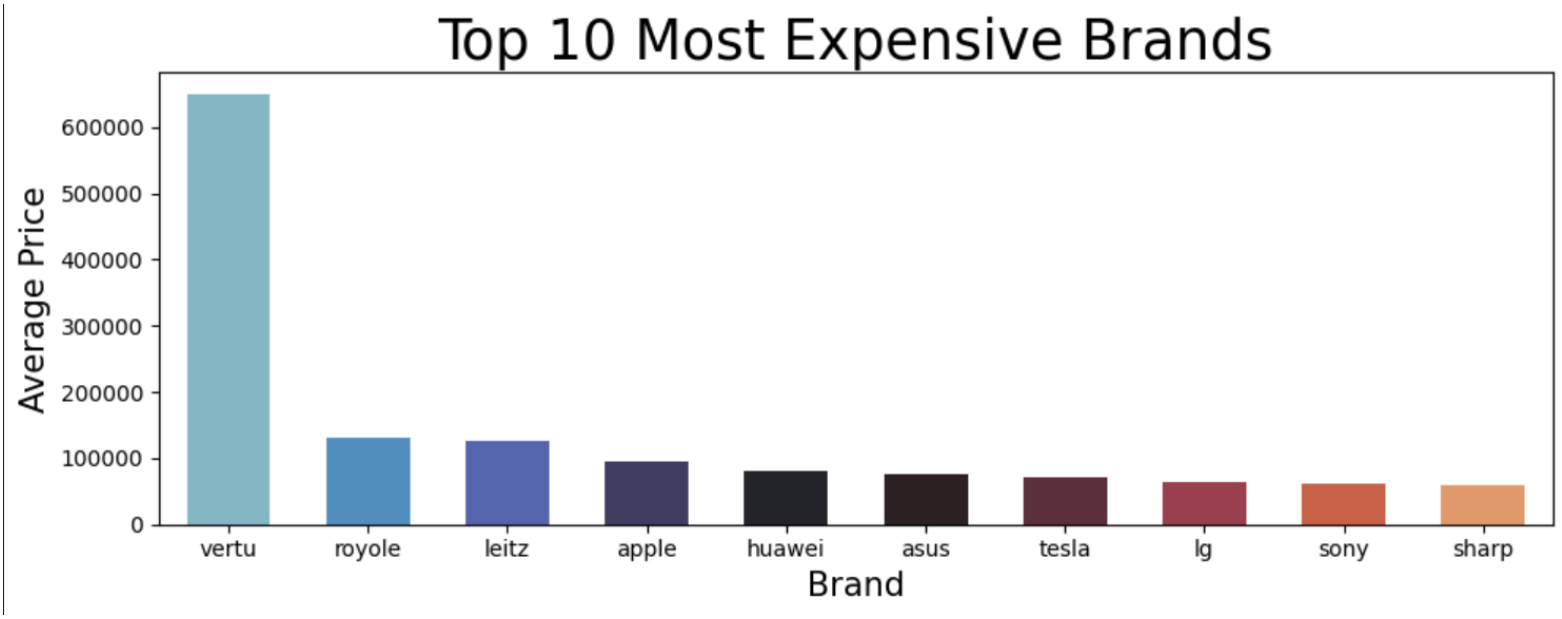
Pair Plots:
Utilizing Seaborn's pairplot can visualize pairwise relationships across multiple features simultaneously, which is particularly useful for high-dimensional datasets.
Feature Interaction:
Examining the combined effect of multiple features, such as creating a composite performance index from CPU speed, RAM, and battery capacity, can reveal deeper insights into what factors drive the pricing or user preference.

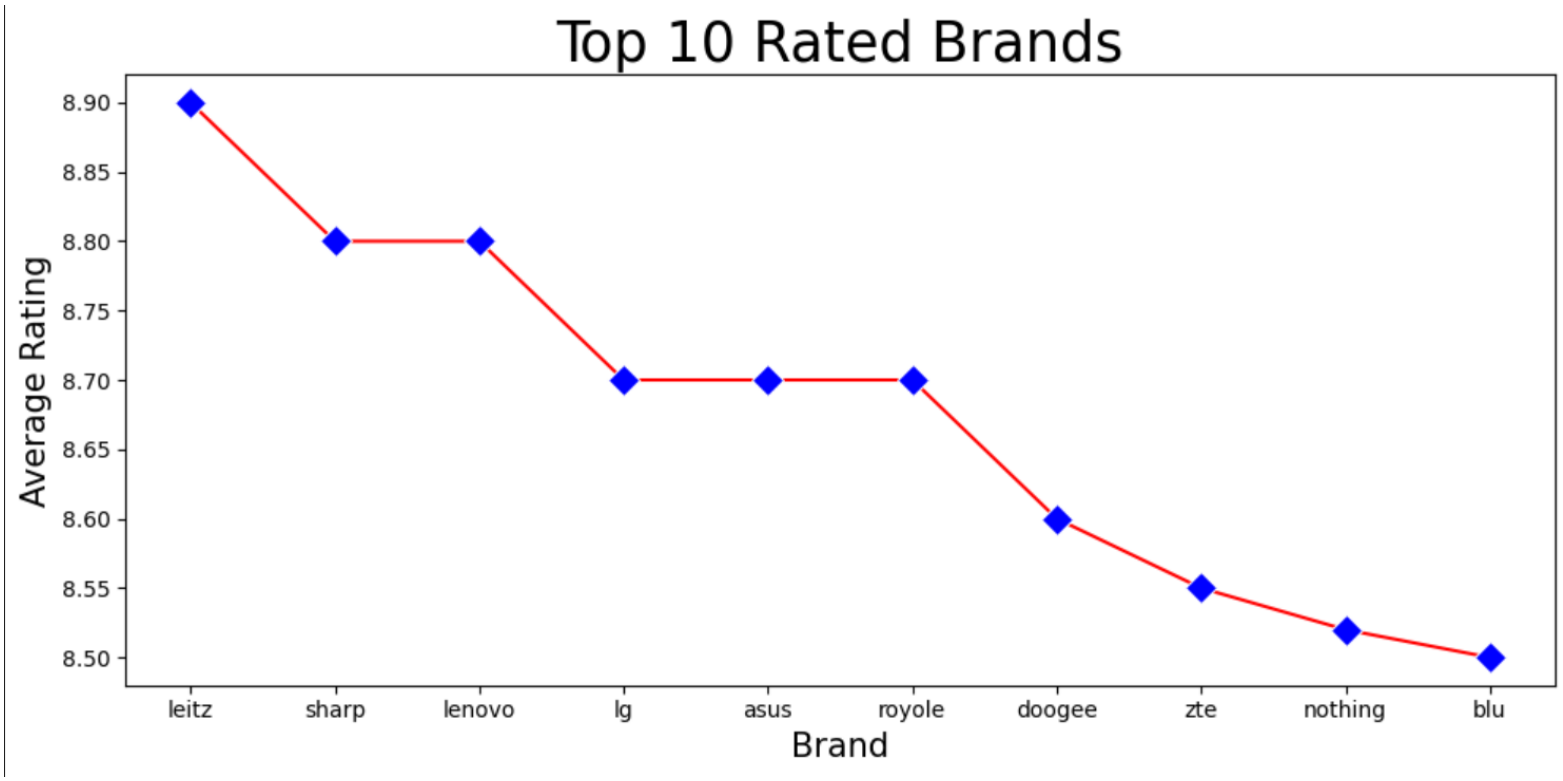
By plotting various histograms of various features price, avg_rating, processor_speed, battery_capacity, screen_size and resolution size we get an idea of distribution of these values for various brand names.

Understanding the distribution of key smartphone features is crucial for evaluating market trends, identifying brand-specific patterns, and making informed decisions. By analyzing **histograms of various numerical attributes** such as **price, average rating, processor speed, battery capacity, screen size, and resolution**, we gain valuable insights into how these specifications vary across different smartphone brands.

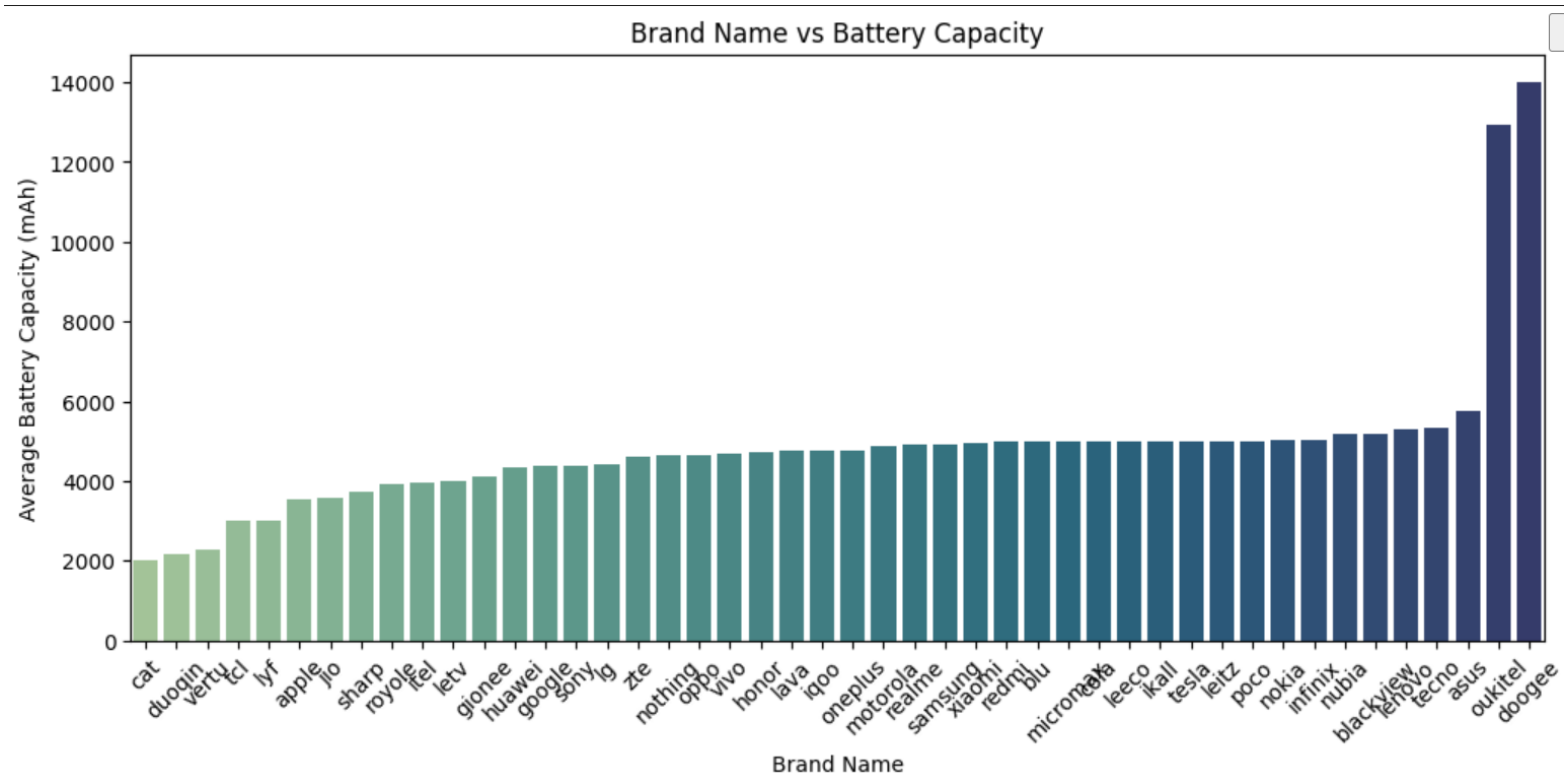
During exploration

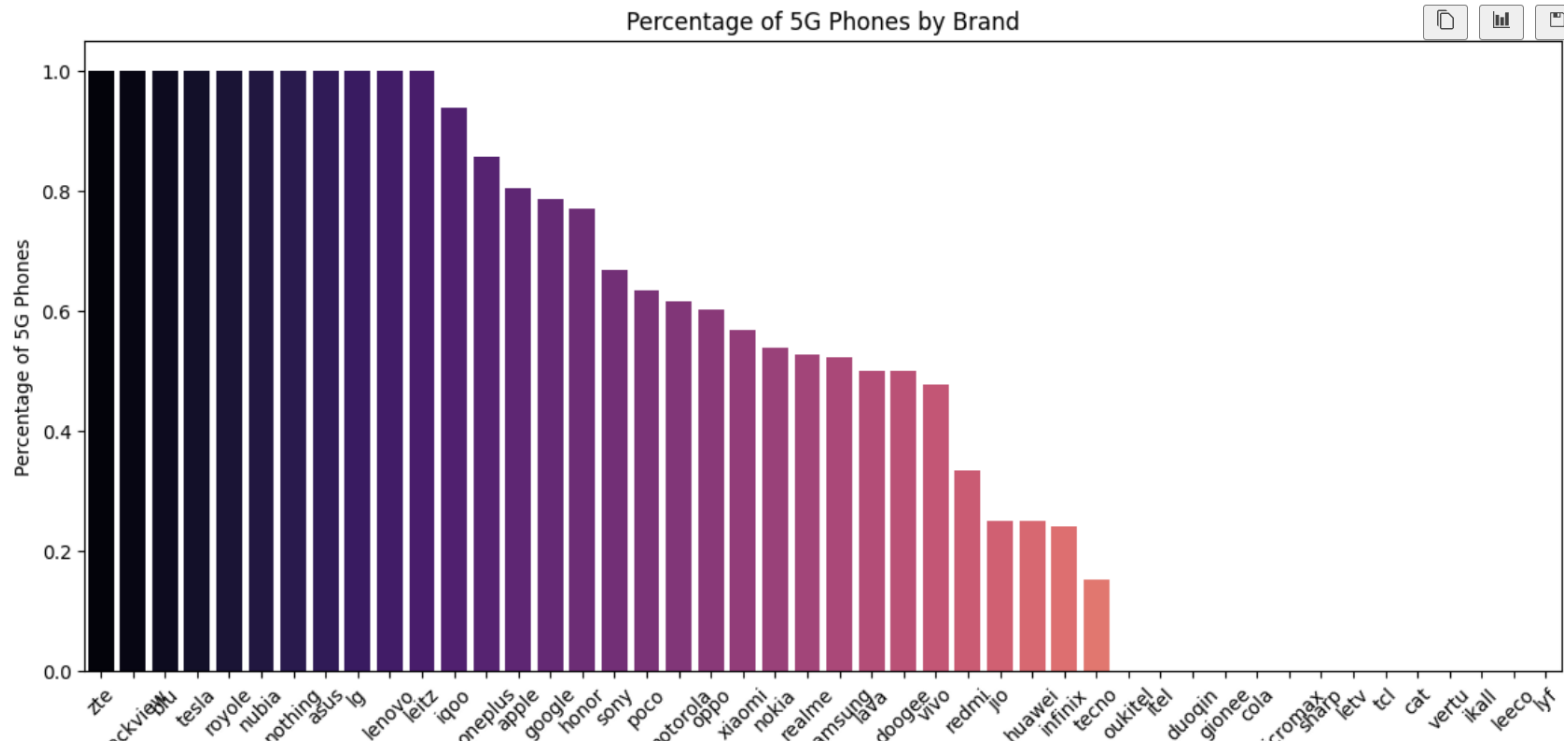

Number of models for top 15 Brands
Number of Models
0
20
40
60
80
100
120
140
xiaomi
samsung
vivo
realme
oppo
motorola
apple
oneplus
poco
tecno
iqoo
infinix
huawei
google
honor
Brand Name


Top 10 Most Expensive Brands
Average Price
0
100000
200000
300000
400000
500000
600000
vertu
royole
leitz
apple
huawei
asus
tesla
lg
sony
sharp
Brand

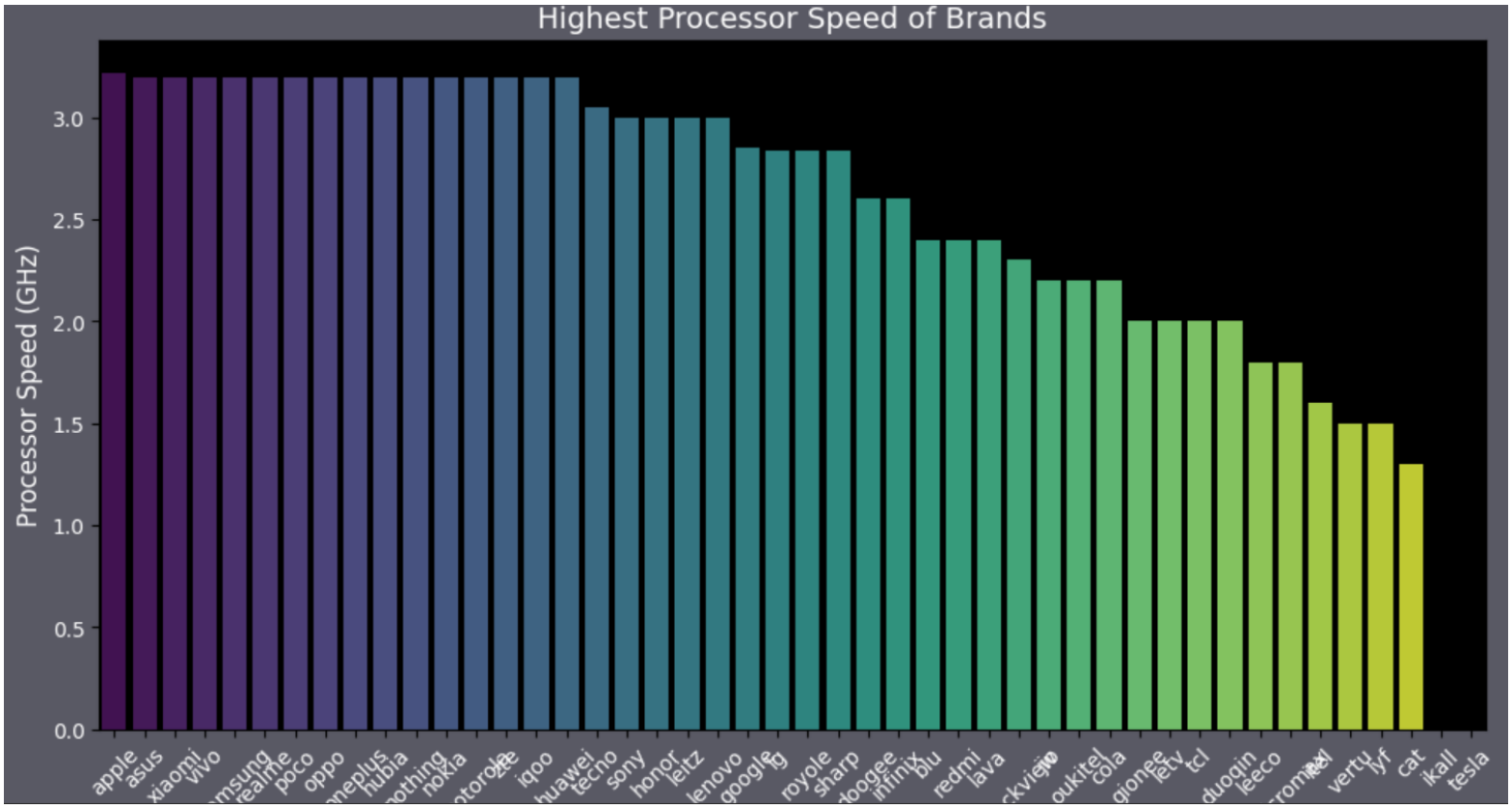


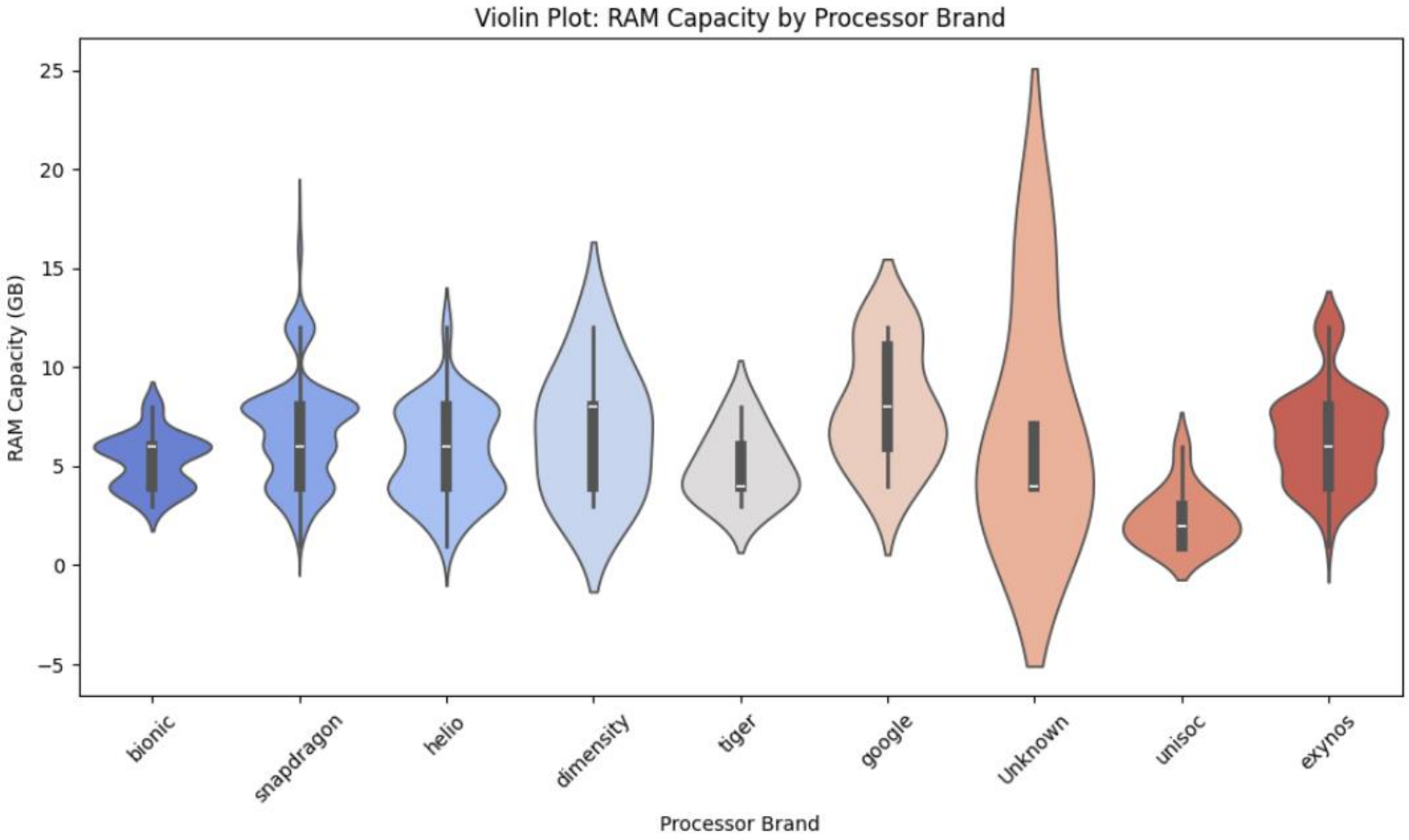
Also performance analysis of different brands

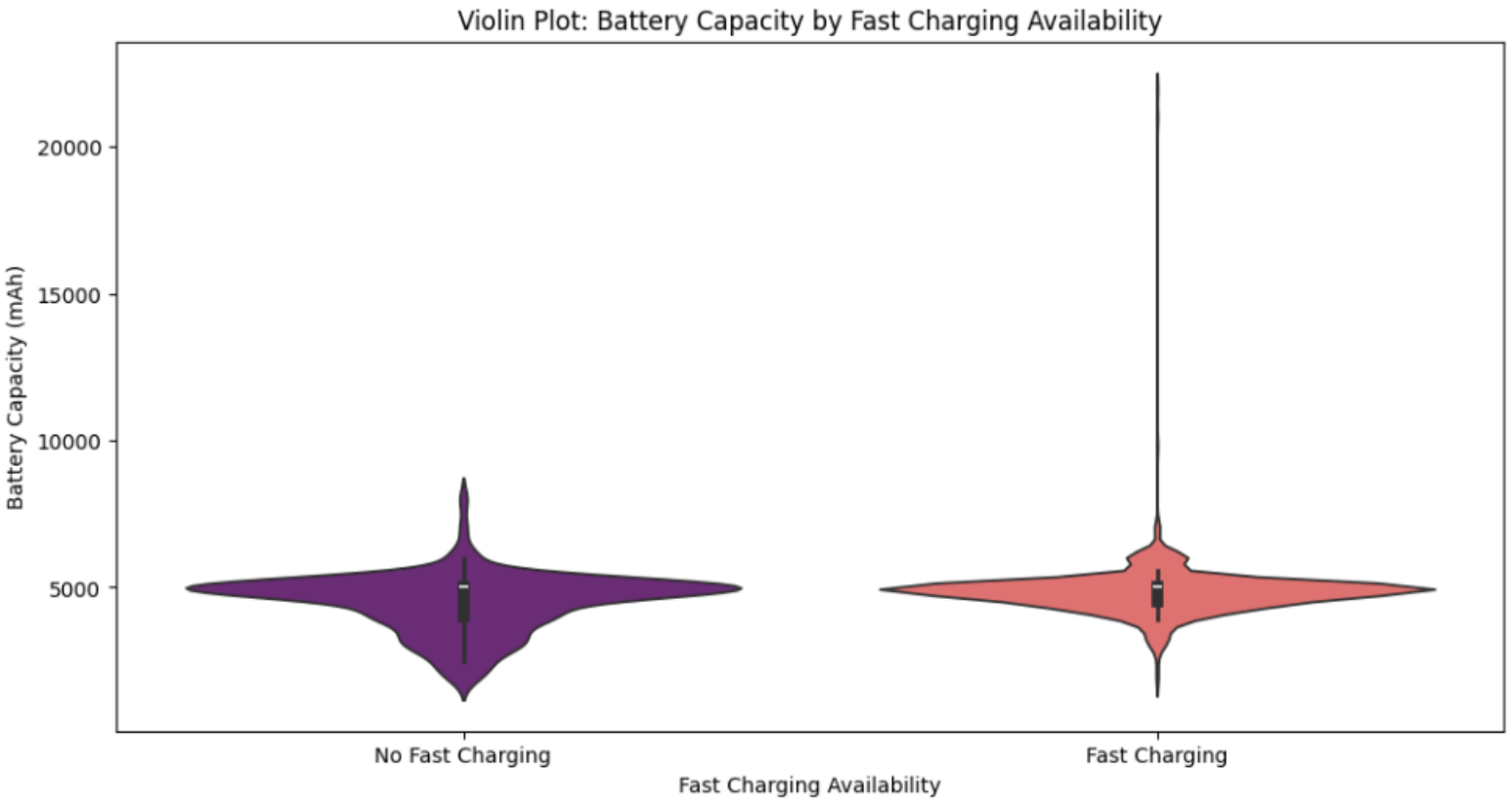

Brand Name vs Battery Capacity
Average Battery Capacity (mAh)
0
2000
4000
6000
8000
10000
12000
14000
cat
duoqin
vertu
tcl
lyf
apple
jio
sharp
royole
itel
letv
gionee
huawei
google
sony
lg
zte
nothing
oppo
vivo
honor
lava
iqoo
oneplus
motorola
realme
samsung
xiaomi
redmi
blu
micromax
cola
leeco
ikall
tesla
leitz
poco
nokia
infinix
nubia
blackview
lenovo
tecno
asus
oukitel
doogee
Brand Name

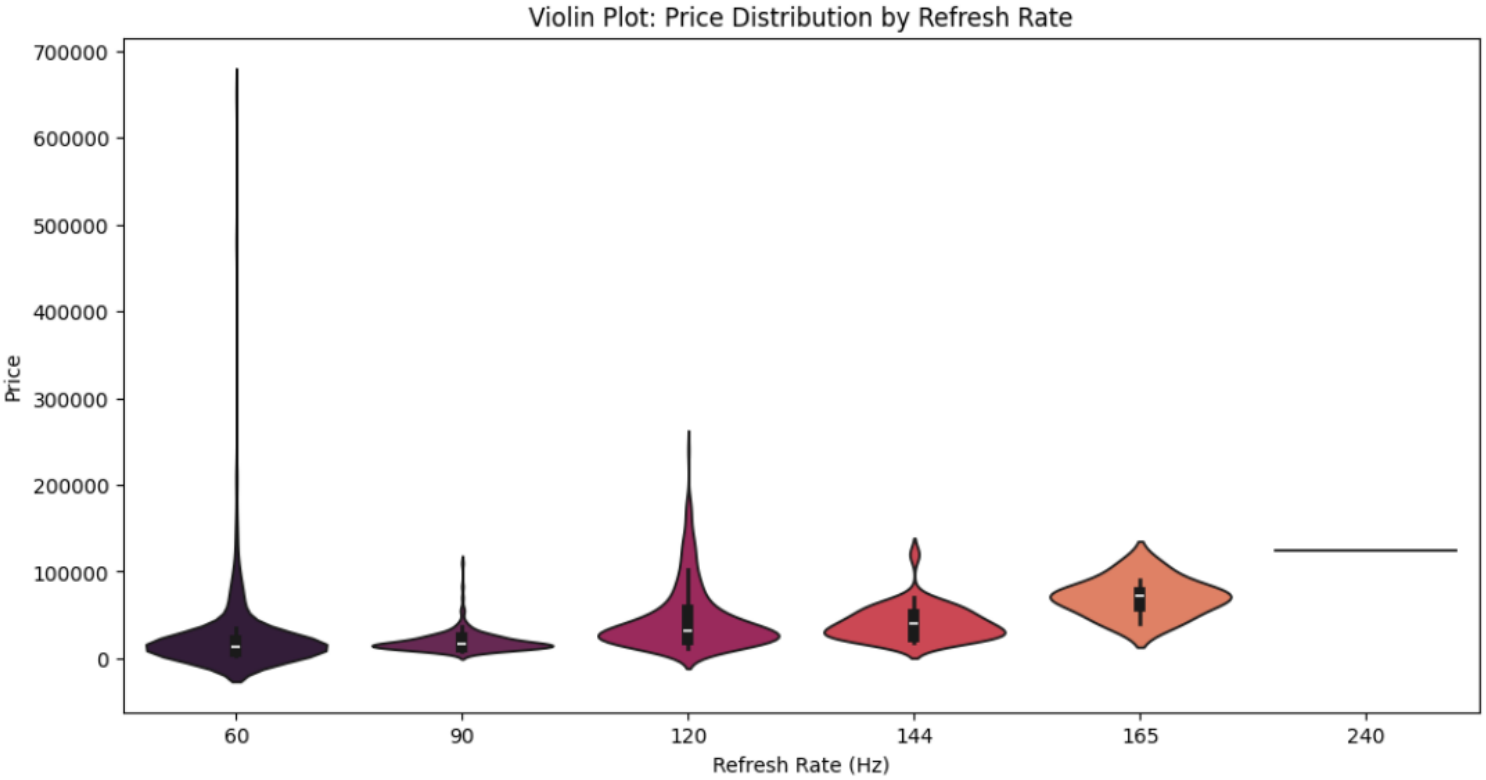

Percentage of 5G Phones by Brand
Percentage of 5G Phones
0.0
0.2
0.4
0.6
0.8
1.0
zte
tesla
royole
nubia
asus
lg
lenovo
leitz
iqoo
apple
google
honor
sony
poco
oppo
xiaomi
nokia
realme
lava
vivo
redmi
jio
huawei
infinix
tecno
oukitel
itel
duoqin
gionee
sharp
letv
tcl
cat
vertu
ikall
leeco
lyf

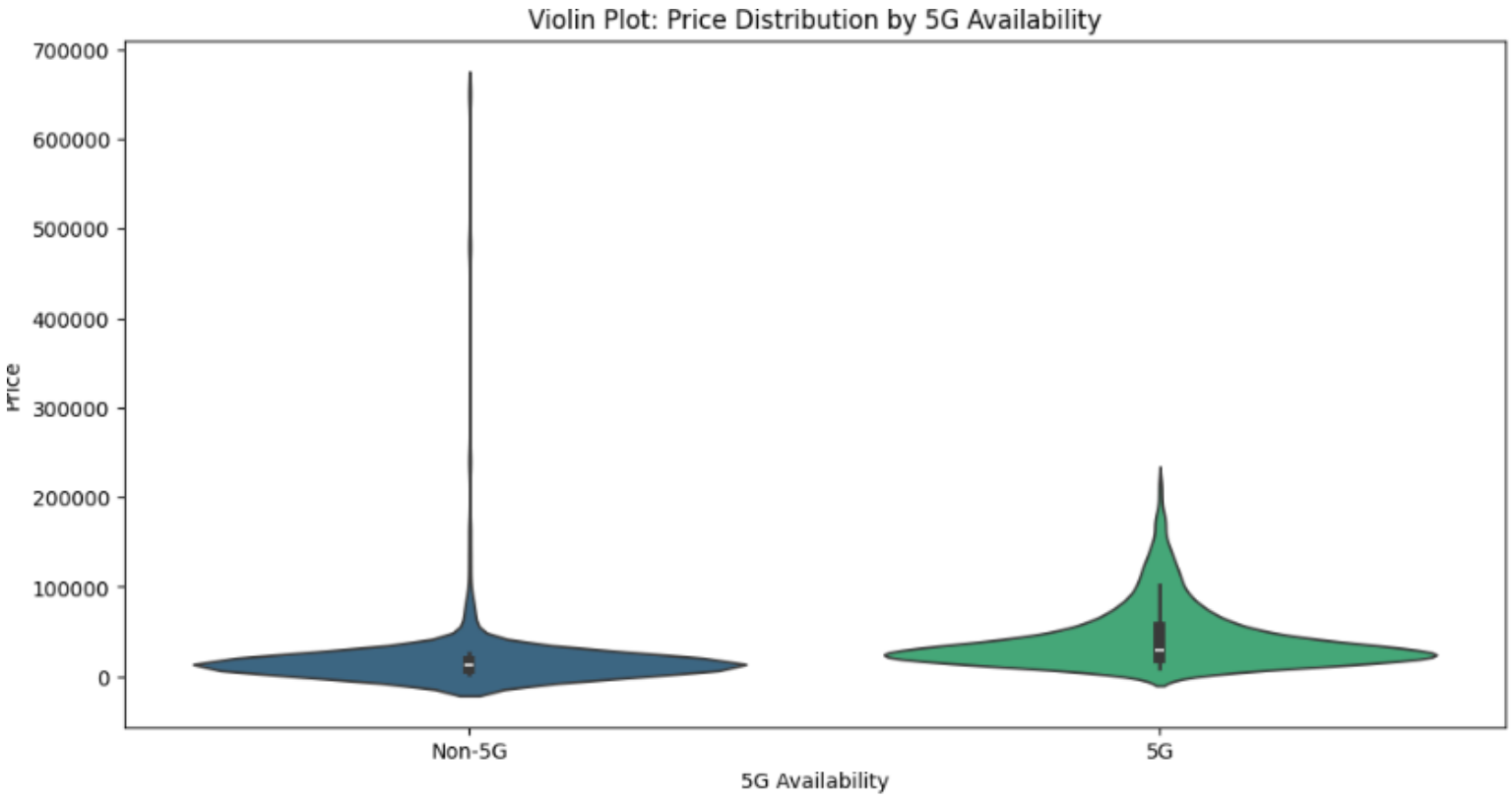
some


Highest Processor Speed of Brands
Processor Speed (GHz)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
apple
asus
xiaomi
vivo
realme
poco
oppo
hubia
nokia
iqoo
huawei
tecno
sony
honor
leitz
lenovo
google
lg
royole
sharp
infinix
blu
redmi
lava
oukitel
cola
letv
tcl
duoqin
leeco
vertu
lyf
cat
ikall
tesla


Violin Plot: RAM Capacity by Processor Brand
RAM Capacity (GB)
25
20
15
10
5
0
−5
bionic
snapdragon
helio
dimensity
tiger
google
Unknown
unisoc
exynos
Processor Brand


Violin Plot: Battery Capacity by Fast Charging Availability
Battery Capacity (mAh)
20000
15000
10000
5000
No Fast Charging
Fast Charging
Fast Charging Availability


Violin Plot: Price Distribution by Refresh Rate
Price
700000
600000
500000
400000
300000
200000
100000
0
60
90
120
144
165
240
Refresh Rate (Hz)



1. RAM Capacity by Processor Brand

The violin plot displays RAM distribution across various processor brands. Key observations include:

- **Bionic, Snapdragon, and Helio** processors tend to have a **moderate range of RAM**, typically between 4GB and 10GB.
- **Dimensity and Google** processors exhibit a **wider RAM distribution**, with values reaching **up to 16GB**.
- **Unisoc and some "Unknown" brands** tend to have lower RAM capacities.
- **Exynos and Tiger** processors show moderate RAM capacities but with a few high-end models reaching above 10GB.

Insights:

- Flagship processors (Google, Dimensity) generally support **higher RAM capacities**.

- Budget-oriented processors (Unisoc, some Unknown brands) tend to have **lower RAM limits**.
- Snapdragon and Helio appear to target a balance between mid-range and high-end configurations.

---

2. Battery Capacity by Fast Charging Availability

This plot compares battery capacities between devices with and without fast charging.

- **Devices with fast charging** show a **wider distribution**, with some extreme outliers exceeding **20,000 mAh**.
- **Devices without fast charging** typically range between **3000mAh and 6000mAh**, with fewer extreme values.

Insights:

- **Higher battery capacities** are more common in fast-charging devices, suggesting they are designed to handle **power-hungry features**.
- Some non-fast-charging devices still have **large batteries**, likely targeting long battery life instead of quick recharging.
- The extreme outliers in the **fast-charging group** may represent **power banks or rugged phones**.

---

3. Price Distribution by Refresh Rate

This visualization shows how smartphone prices vary based on screen refresh rates.

- **60Hz models** display the **widest price range**, including an extreme outlier above **600,000**.
- **Higher refresh rates (120Hz, 144Hz, 165Hz, 240Hz)** show a trend of increasing median prices.
- **90Hz and 120Hz models** are the most balanced in terms of price distribution.
- **240Hz has a single data point, likely a high-end gaming phone**.

Insights:

- The presence of high-end **60Hz phones** suggests that some flagship models prioritize other features (e.g., **cameras, design**) over refresh rates.
- **120Hz and 144Hz screens dominate the mid-to-high-end segments**.
- **165Hz and 240Hz models are niche products**, likely gaming-focused.

---

4. Price Distribution by 5G Availability

This plot compares smartphone prices based on **5G capability**.

- **Non-5G phones have a wider price range**, with a few extreme outliers.
- **5G phones tend to be more expensive**, but there is a **concentration around mid-to-high price points**.

Insights:

- **5G capability is generally found in mid-to-premium smartphones**.
- Some **budget devices still lack 5G**, indicating that **affordable 5G adoption is still evolving**.
- The **outliers in the non-5G category** may be premium 4G phones that focus on **other premium features like build quality or camera technology**.