Project Report

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department
Thapar Institute of Engineering and Technology
(Deemed to be University), Patiala – 147004

UCS546

Submitted By:
Rimjhim Mittal 102103430
Paridhi Maheshwari 102103491

Submitted To:
Dr. Jasmeet Singh

Index

1. **<u>Data Collection for Laptop Market Analysis:</u>**

Overview

The data collection for this project aims to gather detailed information about laptops available on the market. This data is crucial for performing a comprehensive analysis to understand market trends, consumer preferences, and competitive landscapes within the laptop industry. Using automated web scraping techniques, the data was meticulously extracted from a popular e-commerce platform, Amazon, ensuring a rich dataset representing various brands, specifications, and price points.

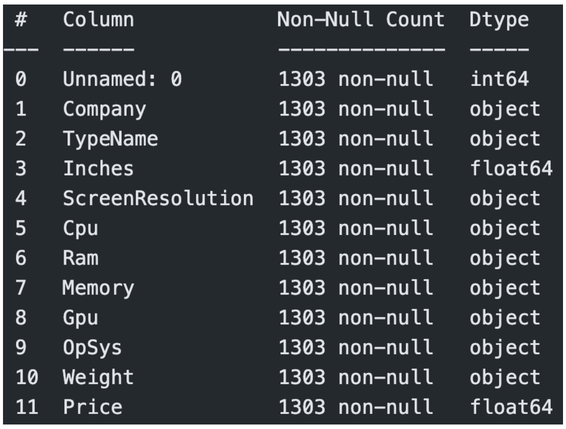
Methodology

The data collection process employed a Python script using the Selenium WebDriver, which simulates user interaction with web browsers. The script navigates through the laptop category on Amazon India, identifying laptop products and extracting relevant information from their detail pages. Each product's specifications, such as brand, model, screen size, resolution, CPU, RAM, storage, GPU, operating system, weight, and price, were collected.

To avoid bot detection and ensure a smooth scraping process, the user-agent was set to mimic a real browser session. The script also handled navigation and pop-ups adeptly, switching between tabs to extract data and handle any exceptions or alerts.

Data Description

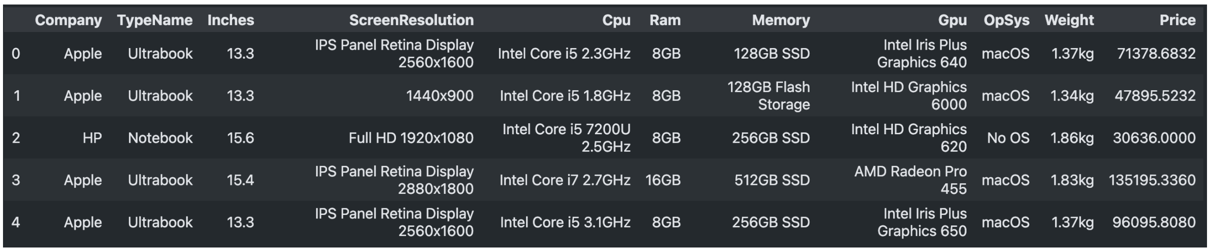
The table below outlines the structure of the collected data :
(1303 rows, 12 columns)



```
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Unnamed: 0        1303 non-null   int64  
 1   Company           1303 non-null   object 
 2   TypeName          1303 non-null   object 
 3   Inches            1303 non-null   float64
 4   ScreenResolution  1303 non-null   object 
 5   Cpu               1303 non-null   object 
 6   Ram               1303 non-null   object 
 7   Memory            1303 non-null   object 
 8   Gpu               1303 non-null   object 
 9   OpSys             1303 non-null   object 
 10  Weight            1303 non-null   object 
 11  Price             1303 non-null   float64
```

2. Data preprocessing

Before preprocessing, the data looks like this:



	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080

1. Standardizing Specification Units

```
df['Ram'] = df['Ram'].str.replace('GB','')
df['Weight'] = df['Weight'].str.replace('kg','')
```

Cleaning the 'Ram' and 'Weight' columns in a dataset. The 'GB' units in the 'Ram' column and the 'kg' units in the 'Weight' column are being removed to convert these columns into numeric data types. This standardization simplifies subsequent data analysis and visualization tasks.

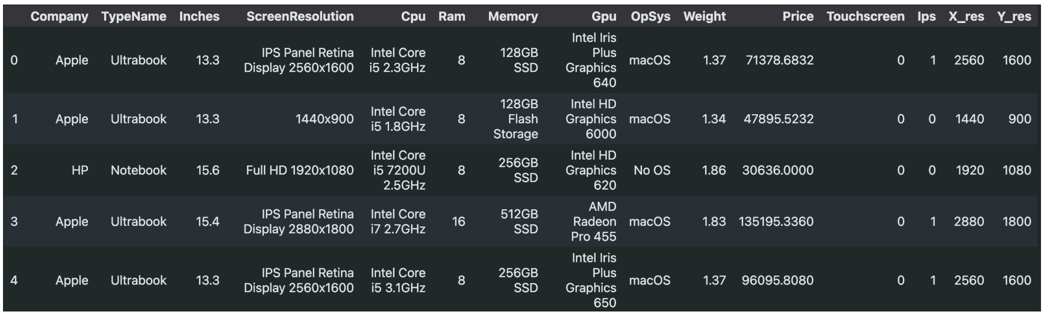
2. Touchscreen and IPS Attributes

```
df['Touchscreen'] = df['ScreenResolution'].apply(lambda x:1 if 'Touchscreen' in x else 0)
```

The dataset is enhanced with two new binary features for display attributes. The 'Ips' column is created by marking entries with 'IPS' in 'ScreenResolution' as 1, signifying an IPS panel, and 0 if not. Similarly, the 'Touchscreen' column is derived by encoding the presence of 'Touchscreen' within 'ScreenResolution' as 1, and its absence as 0. Both transformations convert categorical screen characteristics into a numeric form, streamlining subsequent data analysis.

3. Resolution Split and Cleanup

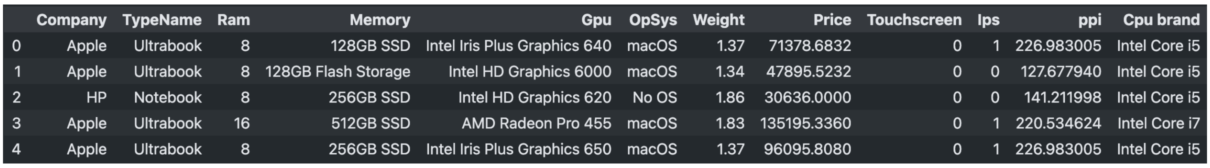
Splits the 'ScreenResolution' column into two new columns, 'X_res' and 'Y_res', representing horizontal and vertical screen resolutions respectively, by separating at the 'x' character and removes commas from the 'X_res' entries and extracts the first set of numerical values to ensure the resolution is in a clean numerical format for analysis. The dataset now looks like:



	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price	Touchscreen	Ips	X_res	Y_res
0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37	71378.6832	0	1	2560	1600
1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34	47895.5232	0	0	1440	900
2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8	256GB SSD	Intel HD Graphics 620	No OS	1.86	30636.0000	0	0	1920	1080
3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16	512GB SSD	AMD Radeon Pro 455	macOS	1.83	135195.3360	0	1	2880	1800
4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37	96095.8080	0	1	2560	1600

4. **Feature Reduction**
   The dataframe is refined by dropping the 'ScreenResolution' column, following the extraction of key features like 'IPS' and 'Touchscreen'. Further reduction is achieved by removing the 'Inches', 'X_res', and 'Y_res' columns to focus on the most relevant features for analysis, thereby streamlining the dataset.

5. **Categorical and Cleanup**
   This creates a simplified 'Cpu Name' feature by extracting the first three words from the 'Cpu' column, which typically represent the CPU's brand and model. It then categorizes the CPU into distinct groups using the fetch_processor function: 'Intel Core i7', 'Intel Core i5', 'Intel Core i3', 'Other Intel Processor', or 'AMD Processor'. Post categorization, it removes the original 'Cpu' and intermediate 'Cpu Name' columns to eliminate redundancy and maintain a tidy dataset focused on relevant processor information.



	Company	TypeName	Ram	Memory	Gpu	OpSys	Weight	Price	Touchscreen	Ips	ppi	Cpu brand
0	Apple	Ultrabook	8	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37	71378.6832	0	1	226.983005	Intel Core i5
1	Apple	Ultrabook	8	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34	47895.5232	0	0	127.677940	Intel Core i5
2	HP	Notebook	8	256GB SSD	Intel HD Graphics 620	No OS	1.86	30636.0000	0	0	141.211998	Intel Core i5
3	Apple	Ultrabook	16	512GB SSD	AMD Radeon Pro 455	macOS	1.83	135195.3360	0	1	220.534624	Intel Core i7
4	Apple	Ultrabook	8	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37	96095.8080	0	1	226.983005	Intel Core i5

6. **Normalization:** Memory capacities are standardized by removing the '.0' from numerical entries and converting 'TB' to '000' for consistent units in gigabytes.

7. **Splitting**: The 'Memory' strings are split into two new columns, 'first' and 'second', to separate combined storage specifications (e.g., '128GB SSD + 1TB HDD').

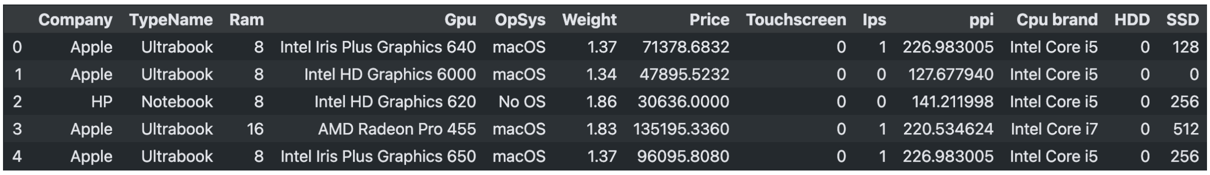
8. **Binary Encoding:** New binary columns are created for each storage type present in the 'first' and 'second' parts of the memory specification, indicating the presence of HDD, SSD, Hybrid, or Flash Storage.

9. **Numeric Extraction:** Non-numeric characters are stripped from the 'first' and 'second' columns to convert them into integer values representing storage size.

10. **Aggregation:** Storage sizes are aggregated into new 'HDD', 'SSD', 'Hybrid', and 'Flash_Storage' columns, multiplying the storage size by its corresponding binary indicator to get total capacities for each storage type.

11. **Dropping Intermediates:** Intermediate columns used for transformations are removed to declutter the dataset.

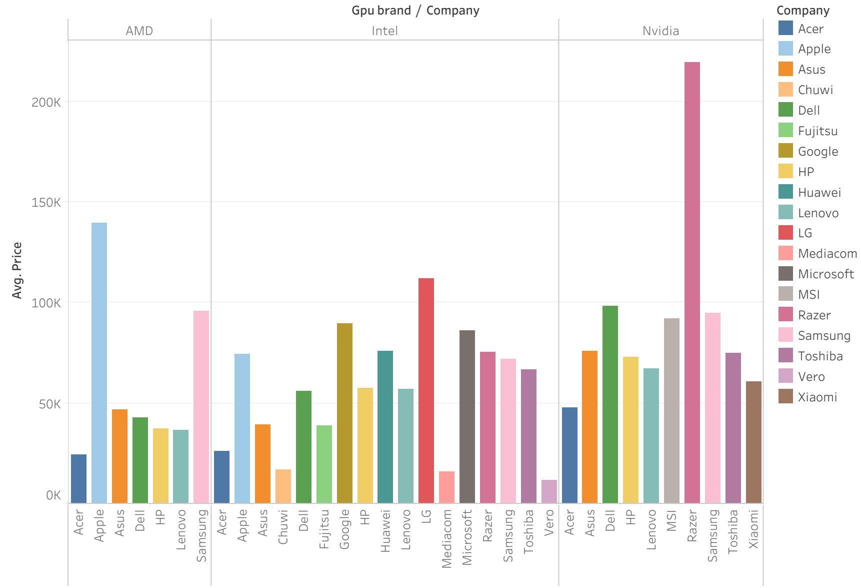
12. **Final Cleanup:** The original 'Memory' column and less common storage types 'Hybrid' and 'Flash_Storage' are dropped, leaving only the more prevalent 'HDD' and 'SSD' features, which simplifies the dataset and focuses on the most impactful features for analysis.



	Company	TypeName	Ram	Gpu	OpSys	Weight	Price	Touchscreen	Ips	ppi	Cpu brand	HDD	SSD
0	Apple	Ultrabook	8	Intel Iris Plus Graphics 640	macOS	1.37	71378.6832	0	1	226.983005	Intel Core i5	0	128
1	Apple	Ultrabook	8	Intel HD Graphics 6000	macOS	1.34	47895.5232	0	0	127.677940	Intel Core i5	0	0
2	HP	Notebook	8	Intel HD Graphics 620	No OS	1.86	30636.0000	0	0	141.211998	Intel Core i5	0	256
3	Apple	Ultrabook	16	AMD Radeon Pro 455	macOS	1.83	135195.3360	0	1	220.534624	Intel Core i7	0	512
4	Apple	Ultrabook	8	Intel Iris Plus Graphics 650	macOS	1.37	96095.8080	0	1	226.983005	Intel Core i5	0	256

3. Charts Shown on the Dashboard

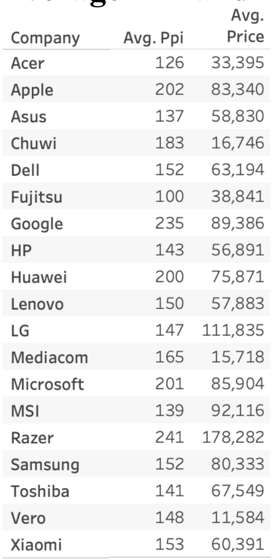
1. How GPU and Company relates with price



Average of Price for each Company broken down by Gpu brand. Color shows details about Company.

The bar chart visualizes the average laptop prices by company, segmented by GPU brand. It highlights price disparities among companies and indicates how GPU brand choice—AMD, Intel, or Nvidia—affects pricing. The chart suggests brand-specific GPU preferences and allows for an assessment of market positioning based on average prices.

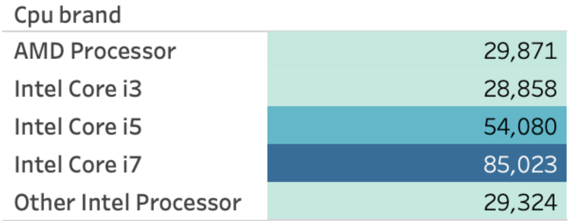
2. Average PPI and Price vs Company



Company	Avg. Ppi	Avg. Price
Acer	126	33,395
Apple	202	83,340
Asus	137	58,830
Chuwi	183	16,746
Dell	152	63,194
Fujitsu	100	38,841
Google	235	89,386
HP	143	56,891
Huawei	200	75,871
Lenovo	150	57,883
LG	147	111,835
Mediacom	165	15,718
Microsoft	201	85,904
MSI	139	92,116
Razer	241	178,282
Samsung	152	80,333
Toshiba	141	67,549
Vero	148	11,584
Xiaomi	153	60,391

This table lists companies alongside their average pixels per inch (PPI) and average laptop prices, providing a quick comparison of display quality and cost across different manufacturers.

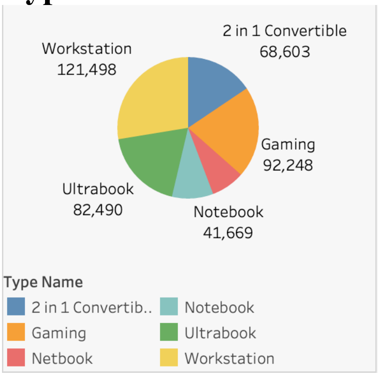
3. CPU Brand VS Price



Cpu brand	
AMD Processor	29,871
Intel Core i3	28,858
Intel Core i5	54,080
Intel Core i7	85,023
Other Intel Processor	29,324

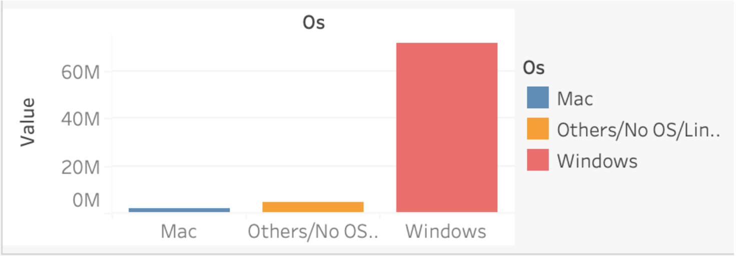
A bar chart that presents the average price of laptops categorized by CPU brand, highlighting the price range associated with different types of processors.

4. Type vs Price



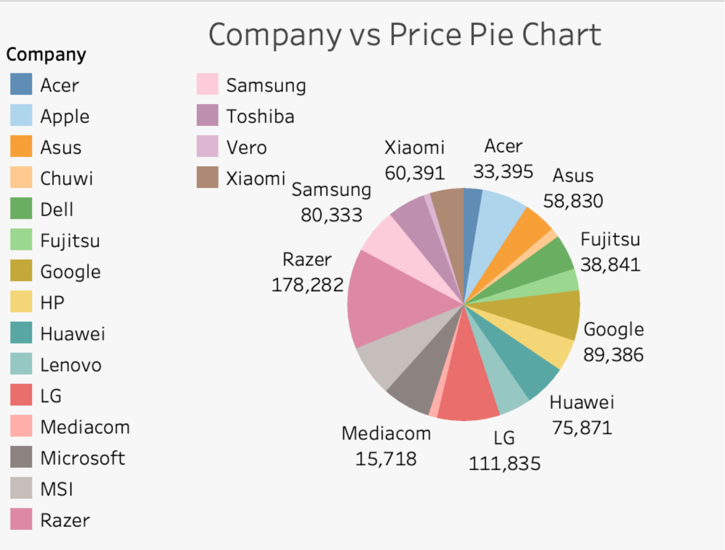
A colored bar chart displaying the average price of laptops by type, such as workstation, ultrabook, and gaming, offering insight into the cost associated with each category.

5. Most Used OS



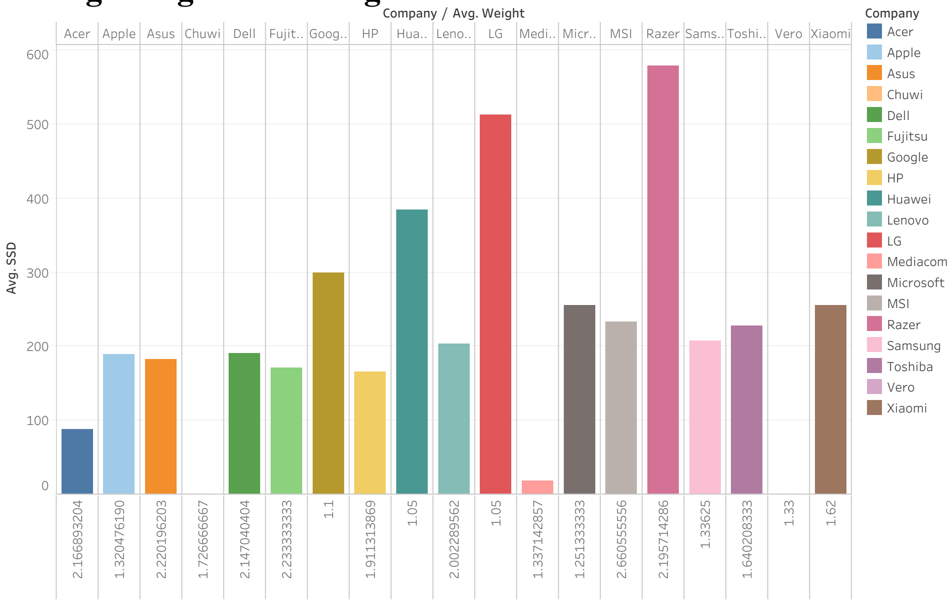
This bar plot indicates that Windows is the dominant OS, with a significantly larger number of laptops using it compared to Mac and other operating systems, implying that Windows-based laptops have a larger market presence or availability in the dataset.

6. Company vs Price Pie Chart



A pie chart providing a visual comparison of the average laptop prices among different companies, illustrating market positioning based on price.

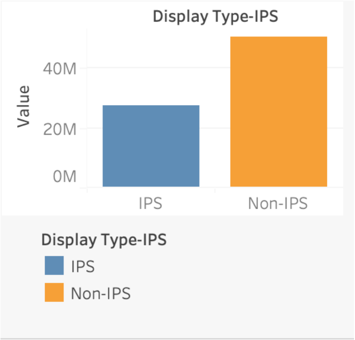
7. Average weight vs Average SSD



Average of SSD for each average of Weight broken down by Company. Color shows details about Company.

A bar chart comparing the average weight and SSD storage capacity of laptops from various companies, potentially indicating a relationship between laptop portability and storage options.

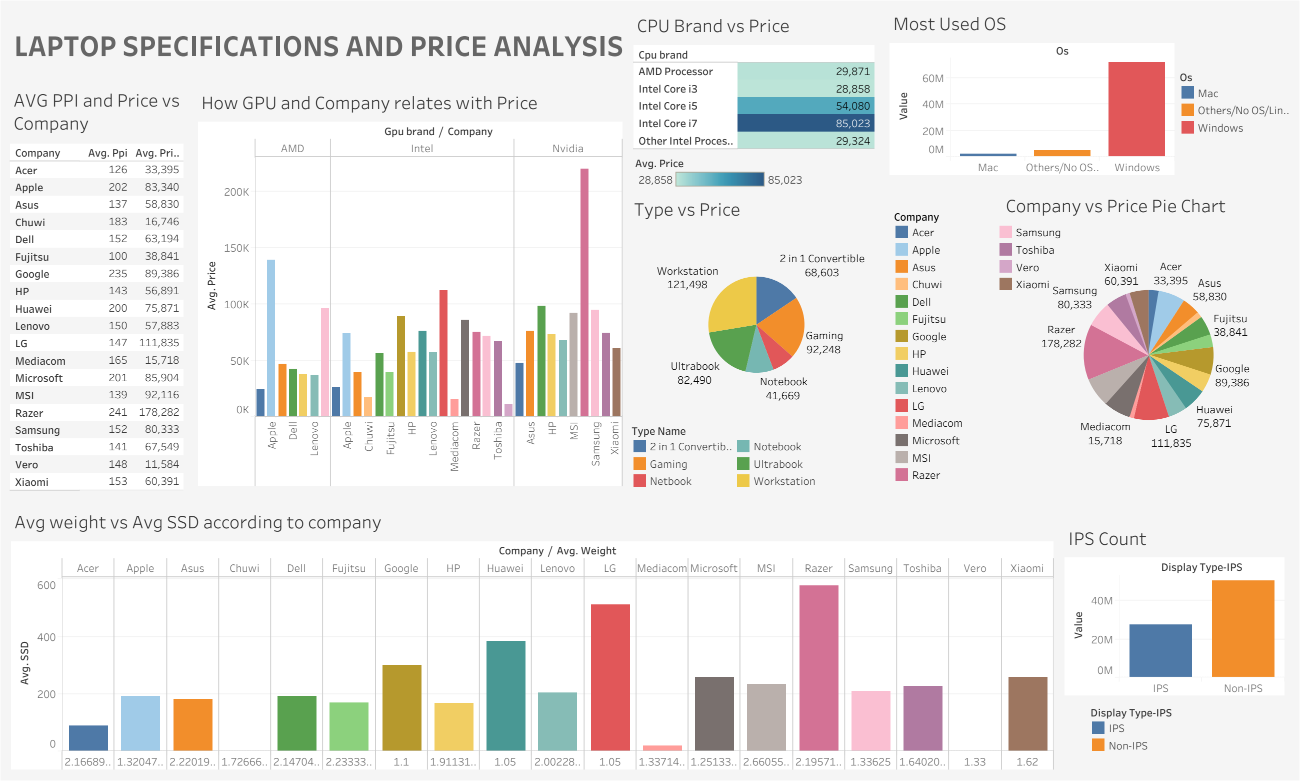
8. IPS Count



A simple bar chart showing the count of laptops with IPS displays versus non-IPS displays, reflecting the prevalence of high-quality screens in the dataset.

4. Snapshot of entire dashboard and link to that dashboard

Link to the dashboard:
https://public.tableau.com/app/profile/rimjhim.mittal/viz/LAPTOP-SPECS/Dashboard2?publish=yes



LAPTOP SPECIFICATIONS AND PRICE ANALYSIS

AVG PPI and Price vs Company

Company	Avg. Ppi	Avg. Pri..
Acer	126	33,395
Apple	202	83,340
Asus	137	58,830
Chuwi	183	16,746
Dell	152	63,194
Fujitsu	100	38,841
Google	235	89,386
HP	143	56,891
Huawei	200	75,871
Lenovo	150	57,883
LG	147	111,835
Mediacom	165	15,718
Microsoft	201	85,904
MSI	139	92,116
Razer	241	178,282
Samsung	152	80,333
Toshiba	141	67,549
Vero	148	11,584
Xiaomi	153	60,391

CPU Brand vs Price

Cpu brand	
AMD Processor	29,871
Intel Core i3	28,858
Intel Core i5	54,080
Intel Core i7	85,023
Other Intel Proces..	29,324

Avg. Price
28,858 85,023