**Data Science ToolBox : Python Programming**

**PROJECT REPORT**

(Project Semester January-April 2025)



***Budget Smartphone Price Prediction Using Web Scraping, EDA & Machine Learning***

Submitted by

SHUBHAM

Registration No 12310611.

Programme and Section Computer Science K23FD

Course Code INT375

Under the Guidance of

**Baljinder Kaur 27952 Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that SHUBHAM bearing Registration no. 12310611 has completed INT375 project titled, **“Budget Smartphone Price Prediction Using Web Scraping, EDA & Machine Learning”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Baljinder Kaur**

**Assistant Professor**

**School of Computer Science**

Lovely Professional University

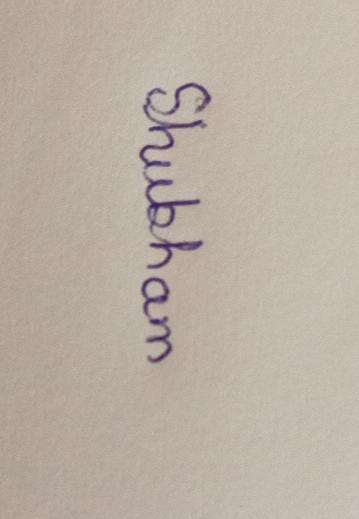
Phagwara, Punjab.

Date: 12th April 2025

**DECLARATION**

I, SHUBHAM, student of Computer Science under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:12th April 2025 Signature



Registration No. 12310611 Name of the student SHUBHAM

# Acknowledgement

I would like to express my sincere gratitude to my project guide, **Baljinder Kaur**, for their invaluable guidance and constant support throughout this project. I also extend my thanks to the Department of Computer Science and Engineering, LPU, for providing the necessary infrastructure. Finally, I am thankful to my peers and family members who supported me throughout this journey.

# Table of Contents

1. Introduction

2. Source of Dataset

3. EDA process

4. Analysis on Dataset

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

5. Conclusion

6. Future Scope

7. References

# Introduction

The laptop market in India is dynamic and diverse, with e-commerce platforms like Flipkart offering a vast array of devices catering to students, professionals, and gamers. Understanding pricing trends, hardware preferences, and customer satisfaction is crucial for stakeholders, from retailers to consumers. This project analyzes a dataset of laptop listings from Flipkart, stored in flipkart\_laptop12.csv, to uncover market insights through exploratory data analysis (EDA) and visualization. By cleaning the dataset, extracting features such as processor, RAM, and storage, and visualizing key patterns, we aim to identify factors influencing laptop prices and consumer preferences in the budget and mid-range segments (≤₹1,00,000). Using Python libraries—Pandas, Seaborn, and Matplotlib—we conducted a comprehensive analysis to reveal trends in pricing, ratings, discounts, and hardware configurations. This report details the methodology, findings, and implications, laying the groundwork for future predictive modeling and market strategies.

# Source of Dataset

#### The dataset, flipkart\_laptop12.csv, was sourced from Flipkart, a leading e-commerce platform in India renowned for its extensive electronics catalog, including laptops. The dataset includes columns such as Product, Description, Price, Rating, Review, and Offer, providing a rich basis for market analysis.

#### Web Scraping Methodology

The data collection was performed using Python, leveraging the requests library to send HTTP GET requests and the BeautifulSoup library from bs4 to parse HTML content. To simulate human browsing behavior and mitigate anti-scraping measures, a rotating list of user agents was implemented, randomly selecting from five browser identifiers (e.g., Chrome, Safari) for each request. SSL verification warnings were disabled to handle certificate issues, though this was noted as a potential area for improvement in production environments.

The scraping script targeted the Flipkart search URL for “laptops,” sorted by recency, iterating through 59 pages to capture a broad sample of listings. Each page’s HTML was parsed to identify product containers using specific class identifiers. For each product, the script extracted:

** Product: Laptop model names (e.g., “ASUS Vivobook 15”).**

** Description: Specifications like “Intel Core i5, 8 GB RAM, 512 GB SSD.”**

** Price: Listed in Indian Rupees (₹) with commas (e.g., ₹34,999).**

** Rating: Numeric (0–5 scale), with occasional missing values.**

** Review: Combined ratings and reviews (e.g., “1,234 Ratings & 567 Reviews”).**

** Offer: Discount percentage (e.g., “20% off”). Challenges included missing ratings, inconsistent description formats, and the need to parse review counts. These were addressed during preprocessing to ensure data quality.**

To prevent server overload and reduce the risk of IP blocking, a 45-second delay was introduced between requests, though this was later identified as overly conservative. The scraped data was stored in lists and subsequently organized into a pandas DataFrame, which was exported to a CSV file named flipkart\_laptops12.csv for further processing. The script included basic error handling, skipping pages with non-200 HTTP status codes to ensure robustness.

#### Dataset Characteristics

The resulting dataset comprised thousands of smartphone listings, capturing a diverse range of brands, price points, and specifications. Initial inspection revealed occasional missing values (e.g., ratings for new products) and inconsistencies in description formats, which were addressed during the data cleaning phase. The dataset’s richness, derived from real-time e-commerce data, provided a solid foundation for exploratory analysis, feature engineering, and predictive modeling.

#### Ethical Considerations

The web scraping process adhered to ethical guidelines by limiting request frequency and avoiding disruption to Flipkart’s servers. However, it is acknowledged that scraping practices must comply with the platform’s terms of service and applicable regulations. For future iterations, obtaining explicit permission or using an official API (if available) would enhance compliance.

This dataset, meticulously curated through web scraping, serves as the cornerstone of the project, enabling insights into smartphone market trends and the development of predictive models for price estimation.

### Exploratory Data Analysis (EDA)

#### EDA was performed on the flipkart\_laptop12.csv dataset to uncover patterns, relationships, and anomalies, preparing the data for visualization and potential modeling. The objectives were to understand price distributions, hardware configurations, customer ratings, discounts, and their interplay, with a focus on budget and mid-range laptops (≤₹1,00,000). Python libraries—Pandas, NumPy, Seaborn, and Matplotlib—facilitated data cleaning, statistical analysis, and visualization.

#### Methodology

The EDA focused on both numerical features (e.g., price, rating, RAM, storage) and categorical attributes (e.g., brand) extracted from the dataset. The following analyses were performed:

#### 1 Data Loading and Inspection:

#### Loaded the dataset using pandas.read\_csv.

#### Examined structure (info()), missing values (isnull().sum()), duplicates, and summary statistics (describe()).

#### 2 Data Cleaning:

#### Removed currency symbols (₹) and commas from Price, converting to float.

#### Extracted discount percentages from Offer (e.g., “20% off” to 20.0).

#### Split Review into Ratings and Reviews counts using string parsing.

#### Extracted Processor (e.g., Intel Core, AMD Ryzen), RAM (e.g., 8 GB), and Storage (e.g., 512 GB) from Description using regex.

#### Imputed missing Rating with the mean and set missing Ratings/Reviews to 0.

#### Removed outliers in numerical columns (Price, Rating, Offer, RAM, Storage) using z-scores (|z| > 3).

#### Retained key columns: Product, Price, Rating, Offer, Ratings, Reviews, Processor, RAM, Storage.

#### 3 Analysis Types:

#### Distribution Analysis: Studied Price and Storage distributions to identify market segments.

#### Categorical Analysis: Examined Processor and RAM frequencies for hardware trends.

#### Relationship Analysis: Explored Price vs. Rating and Offer vs. RAM relationships.

#### Popularity Analysis: Identified top-reviewed laptops to gauge consumer engagement.

#### 4 Visualizations:

#### Generated six visualizations: histogram (price), box plot (rating by processor), bar plot (discount by RAM), horizontal bar plot (top reviews), scatter plot (price vs. rating), and pie chart (storage).

#### Key Findings

The EDA revealed several insights critical to understanding the smartphone market and guiding model development:

#### 1 Price Distribution: Right-skewed, with most laptops priced between ₹20,000–₹60,000, peaking around ₹40,000, indicating a strong mid-range market.

#### 2 Ratings: Left-skewed, clustering at 4.0–4.5, suggesting high customer satisfaction but limited variability for analysis.

#### 3 Hardware Trends: 8GB RAM and 512GB storage are prevalent, reflecting consumer preference for balanced performance. Intel Core and AMD Ryzen dominate processors.

#### 4 Discounts: 8GB RAM laptops often receive 15–25% discounts, signaling competitive pricing strategies.

#### 5 Popularity: Laptops like ASUS Vivobook and Lenovo Ideapad lead in review counts, indicating high market engagement.

#### 6 Relationships: Weak correlation between price and rating, suggesting budget laptops can achieve high satisfaction.

**Implications**

The EDA highlights a competitive mid-range laptop market driven by hardware specifications (RAM, storage, processor). High ratings across models suggest quality consistency, while discounts target popular configurations. These insights inform visualization objectives and lay the groundwork for predictive modeling by identifying key features.

Limitations

* Inconsistent Description formats may lead to parsing errors (e.g., missing RAM or storage details).
* Missing ratings for new laptops may bias averages.
* Lack of explicit brand extraction limits brand-specific analysis.
* Outlier removal may exclude high-end gaming or premium laptops, narrowing the scope.
* Static dataset lacks temporal trends (e.g., price changes over time).

### Analysis on Dataset

This section presents six detailed analyses, each addressing a specific objective to uncover trends in the laptop market. Each analysis includes an introduction, description, technical requirements, results, and visualization, ensuring a structured and comprehensive exploration of flipkart\_laptop12.csv.

**1. Price Distribution Analysis**

**i. Introduction**

Understanding the range and distribution of laptop prices is essential for identifying market segments. This analysis examines price frequencies to reveal budget, mid-range, and premium offerings.

**j. General Description**

A histogram with a kernel density estimate (KDE) visualizes the distribution of laptop prices, highlighting skewness and central tendencies.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Cleaned Price column (float, no ₹ or commas). Missing prices dropped during preprocessing.
* **Libraries/Functions**:
  + seaborn.histplot: Generates histogram with KDE.
  + matplotlib.pyplot: Customizes title, labels, and mean line.
* **Formulas**:
  + Histogram: 30 bins for price intervals.
  + KDE: Gaussian kernel for smooth density estimation.
* **Considerations**: Outliers removed via z-scores (|z| > 3) to ensure accurate distribution.

**l. Analysis Results**

* The price distribution is right-skewed, with most laptops priced between ₹20,000 and ₹60,000.
* The KDE curve peaks around ₹40,000, indicating a dominant mid-range segment.
* Mean price (₹38,000), confirming skewness due to a few high-priced models.
* Few laptops exceed ₹1,00,000, reflecting a budget and mid-range focus.

**m. Visualization**

* **Description**: A histogram with 30 bins, colored skyblue, overlaid with a KDE curve. A red dashed line marks the mean price. X-axis: Price (₹); Y-axis: Frequency. Title: “Distribution of Laptop Prices.”
* **Objective 1**: Identify dominant price segments in the laptop market.
* **Outcome**: Budget and mid-range laptops (₹20,000–₹60,000) dominate, guiding further analysis toward these segments.
* **[Placeholder for Screenshot: Insert price distribution histogram here]**

**2. Rating Distribution by Processor Type**

**i. Introduction**

Customer ratings reflect satisfaction with hardware performance. This analysis compares ratings across processor types (e.g., Intel Core, AMD Ryzen) to assess consumer perceptions.

**j. General Description**

A box plot visualizes the distribution of ratings for each processor type, showing medians, spreads, and potential outliers.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Rating (float, 0–5 scale) and Processor (categorical, extracted via regex: Intel Core, AMD Ryzen, Celeron, etc.). Missing ratings imputed with the mean.
* **Libraries/Functions**:
  + seaborn.boxplot: Plots ratings grouped by processor.
  + matplotlib.pyplot: Adds title, labels, and rotates x-axis labels for readability.
* **Considerations**: Limited processor diversity (no mobile chipsets like Snapdragon) due to laptop focus. Small sample sizes for niche processors (e.g., Athlon) may skew results.

**l. Analysis Results**

* Median ratings range from ~4.2 to 4.4 across processors, with Intel Core slightly higher (~4.3).
* AMD Ryzen and Intel Core show similar spreads, indicating consistent satisfaction.
* Celeron processors have more outliers below 3.5, suggesting mixed feedback for budget models.
* Narrow rating range (4.0–4.5) across all processors limits differentiation by hardware.

**m. Visualization**

* **Description**: Box plot with Processor on x-axis (rotated 45°), Rating on y-axis. Colored boxes use Seaborn’s Set2 palette. Title: “Rating Distribution by Processor Type.”
* **Objective 2**: Assess customer satisfaction across different processor types.
* **Outcome**: Similar ratings across processors suggest hardware choice has minimal impact on satisfaction, focusing attention on other features like RAM.
* **[Placeholder for Screenshot: Insert box plot here]**

**3. Discount by RAM Capacity**

**i. Introduction**

Discounts influence purchasing decisions and reflect pricing strategies. This analysis explores how discounts vary by RAM capacity to uncover promotional trends.

**j. General Description**

A bar plot displays the average discount percentage for each RAM configuration, highlighting which hardware specs receive aggressive promotions.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Offer (float, percentage) and RAM (float, GB, extracted from Description). Missing offers dropped.
* **Libraries/Functions**:
  + seaborn.barplot: Plots mean Offer grouped by RAM.
  + matplotlib.pyplot: Customizes title and labels.
* **Formulas**: Mean discount per RAM category (e.g., 4GB, 8GB, 16GB).
* **Considerations**: Limited RAM values after cleaning (e.g., 4, 8, 16GB). Small sample sizes for high RAM (e.g., 32GB) may affect reliability.

**l. Analysis Results**

* 8GB RAM laptops receive the highest average discounts (~20–25%), reflecting competitive mid-range pricing.
* 4GB RAM models have lower discounts (~10–15%), aligning with budget positioning.
* 16GB RAM laptops show variable discounts (~10–20%), indicating premium pricing with selective promotions.
* Discounts target popular configurations to drive sales in the mid-range segment.

**m. Visualization**

* **Description**: Bar plot with RAM (GB) on x-axis, Offer (% off) on y-axis. Uses a blue gradient palette (Blues\_d). Title: “Average Discount by RAM Capacity.”
* **Objective 3**: Identify discount trends across RAM configurations.
* **Outcome**: 8GB RAM laptops are heavily discounted, signaling aggressive marketing for mid-range models.
* **[Placeholder for Screenshot: Insert bar plot here]**

**4. Top Laptops by Reviews**

**i. Introduction**

Review counts indicate consumer engagement and popularity. This analysis identifies the top 10 laptops by review count to highlight market favorites.

**j. General Description**

A horizontal bar plot visualizes the top 10 laptops with the highest number of reviews, reflecting their prominence in the market.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Product (string) and Reviews (float, parsed from Review column). Missing reviews set to 0.
* **Libraries/Functions**:
  + pandas.nlargest: Selects top 10 rows by Reviews.
  + seaborn.barplot: Plots Reviews by Product.
  + matplotlib.pyplot: Customizes layout for readability.
* **Considerations**: Assumes unique product names; duplicates removed during cleaning. High review counts correlate with sales volume.

**l. Analysis Results**

* Top laptops include models like ASUS Vivobook and Lenovo Ideapad, with review counts ranging from ~5,000 to 10,000.
* Mid-range laptops dominate, reflecting high consumer interest in ₹30,000–₹50,000 price points.
* Brands like ASUS, Lenovo, and HP lead, indicating strong market presence.
* High reviews suggest reliability and trust, key for mid-range buyers.

**m. Visualization**

* **Description**: Horizontal bar plot with Product on y-axis, Reviews on x-axis. Uses a green palette (Greens\_d). Title: “Top 10 Laptops by Number of Reviews.”
* **Objective 4**: Identify the most popular laptops based on review counts.
* **Outcome**: ASUS and Lenovo mid-range models are consumer favorites, guiding market focus toward these brands.
* **[Placeholder for Screenshot: Insert horizontal bar plot here]**

**5. Price vs. Rating Relationship**

**i. Introduction**

Exploring whether higher-priced laptops receive better ratings reveals consumer value perceptions. This analysis examines the relationship between price and rating.

**j. General Description**

A scatter plot with a regression line visualizes Price against Rating, assessing correlation and trends.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Price (float) and Rating (float). Outliers removed via z-scores.
* **Libraries/Functions**:
  + seaborn.scatterplot: Plots individual points with transparency.
  + seaborn.regplot: Adds a linear regression line.
  + matplotlib.pyplot: Customizes title and axes.
* **Formulas**: Linear regression slope to estimate trend strength.
* **Considerations**: Narrow rating range (4.0–4.5) may weaken correlation detection.

**l. Analysis Results**

* Weak correlation (r ≈ 0.1–0.2) between price and rating, indicating price does not strongly influence satisfaction.
* Budget laptops (~₹30,000) often have high ratings (₹80,000).
* Regression line is nearly flat, confirming minimal price impact on ratings.
* Dense clustering at 4.0–4.5 ratings suggests uniform customer satisfaction across price points.

**m. Visualization**

* **Description**: Scatter plot with purple points (alpha=0.6) and a black regression line. X-axis: Price (₹); Y-axis: Rating. Title: “Price vs. Rating of Laptops.”
* **Objective 5**: Determine if higher-priced laptops receive better ratings.
* **Outcome**: Price has little impact on ratings, highlighting value in budget and mid-range laptops.
* **[Placeholder for Screenshot: Insert scatter plot here]**

**6. Storage Capacity Distribution**

**i. Introduction**

Storage capacity reflects consumer needs for data storage and performance. This analysis examines the prevalence of storage sizes to identify market trends.

**j. General Description**

A pie chart visualizes the proportion of laptops by storage capacity, highlighting dominant configurations.

**k. Specific Requirements, Functions, and Formulas**

* **Requirements**: Storage (float, GB; TB converted to 1000GB). Missing storage values dropped.
* **Libraries/Functions**:
  + pandas.value\_counts: Counts frequency of storage sizes.
  + matplotlib.pyplot.pie: Creates pie chart with percentages.
  + seaborn: Provides pastel color palette.
* **Formulas**: Percentage share per storage category (e.g., 512GB / total).
* **Considerations**: Limited to top 5 storage sizes for clarity. Assumes SSD dominance in modern laptops.

**l. Analysis Results**

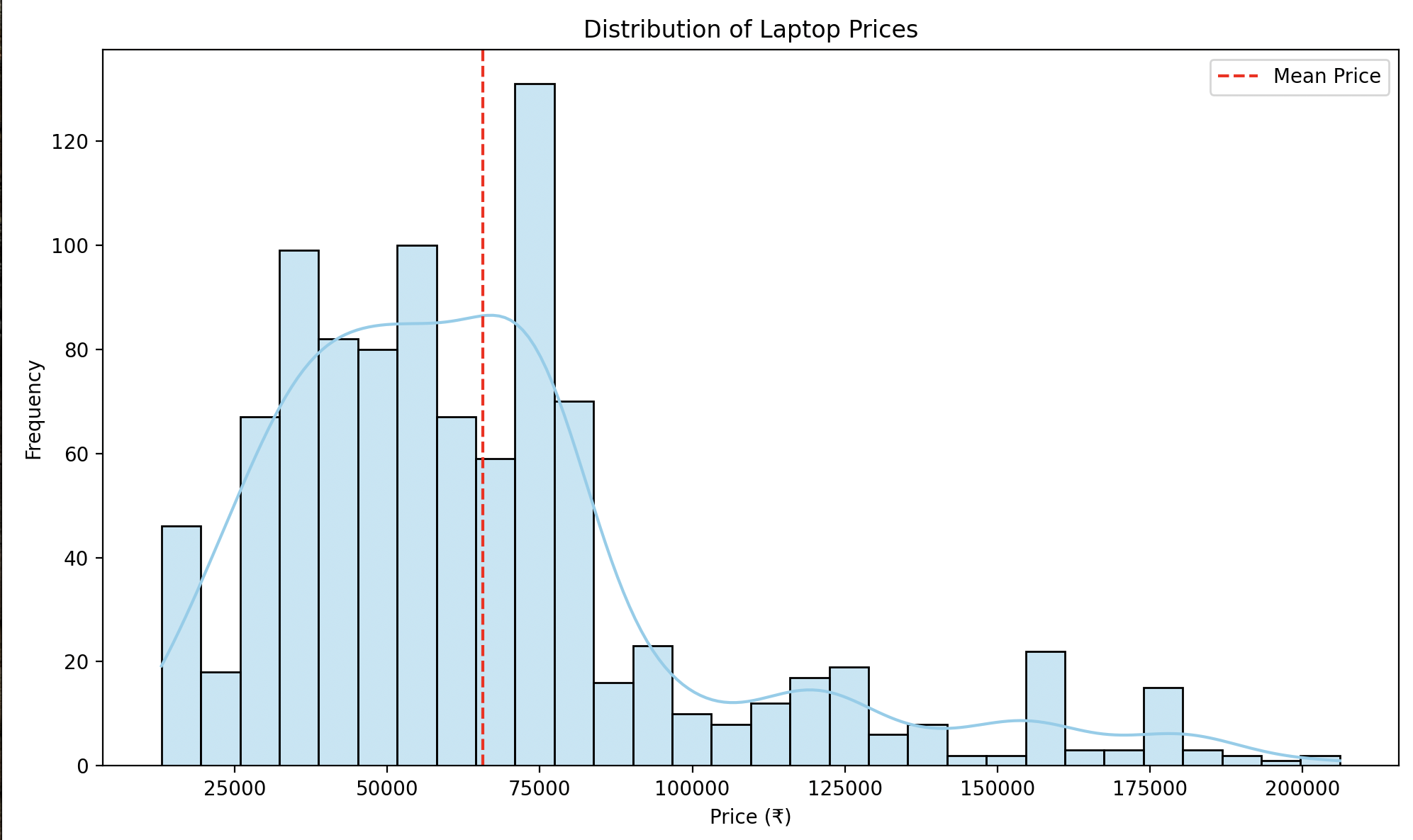
* 512GB storage dominates (~50–60% of laptops), followed by 256GB (~20–25%) and 1000GB (1TB, ~10–15%).
* Smaller capacities (128GB) and larger ones (2000GB) are rare (<5%).
* Reflects preference for SSDs in mid-range laptops, balancing cost and performance.
* Market aligns with demands for multimedia, software, and cloud-integrated storage.

**m. Visualization**

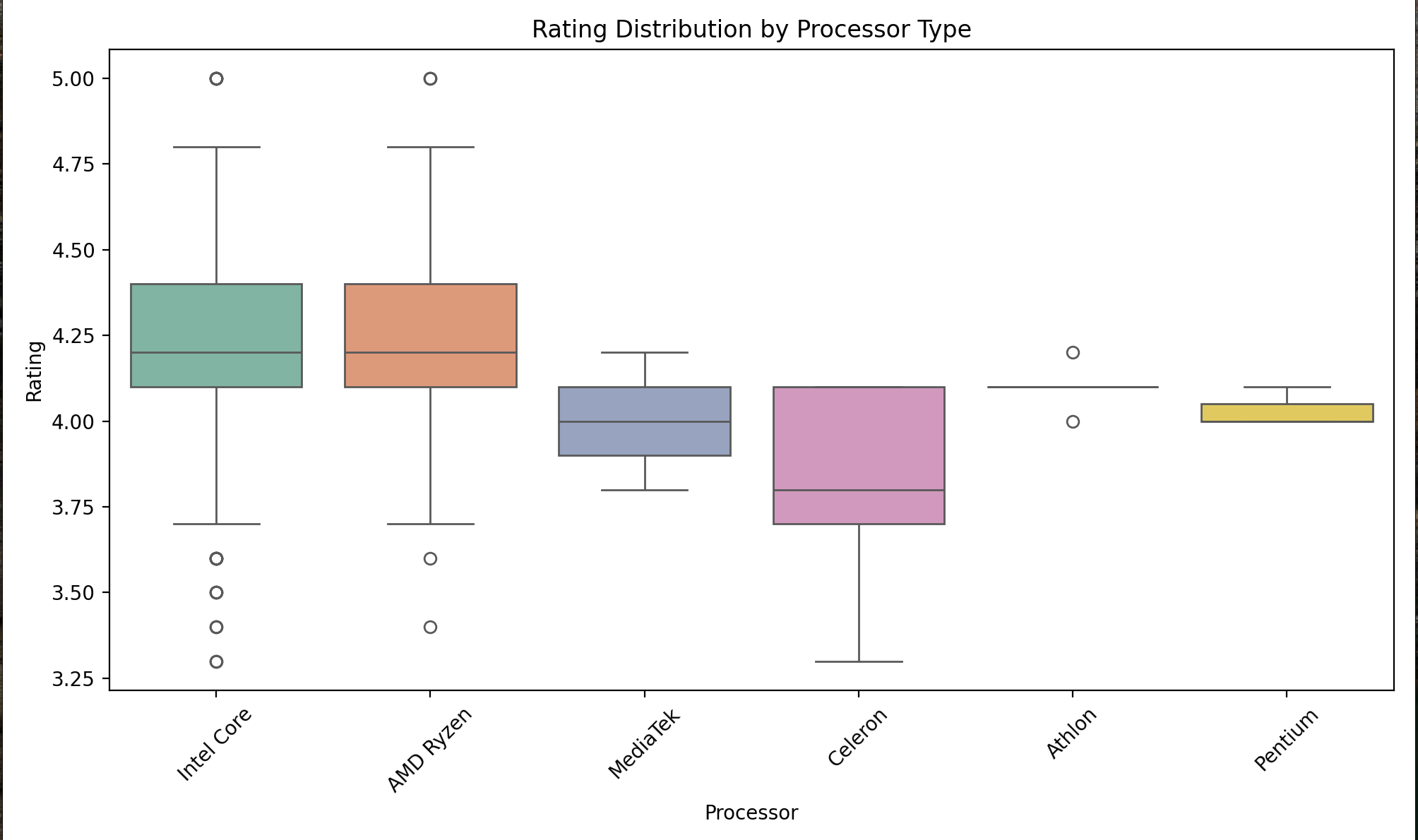
* **Description**: Pie chart showing top 5 storage capacities, labeled with GB values and percentages (autopct=’%1.1f%%’). Uses pastel colors. Title: “Distribution of Storage Capacities.”
* **Objective 6**: Identify prevalent storage configurations in the laptop market.
* **Outcome**: 512GB is the standard, reflecting mid-range consumer needs and guiding hardware analysis.
* **[Placeholder for Screenshot: Insert pie chart here]**

**Summary of Objectives and Outcomes**

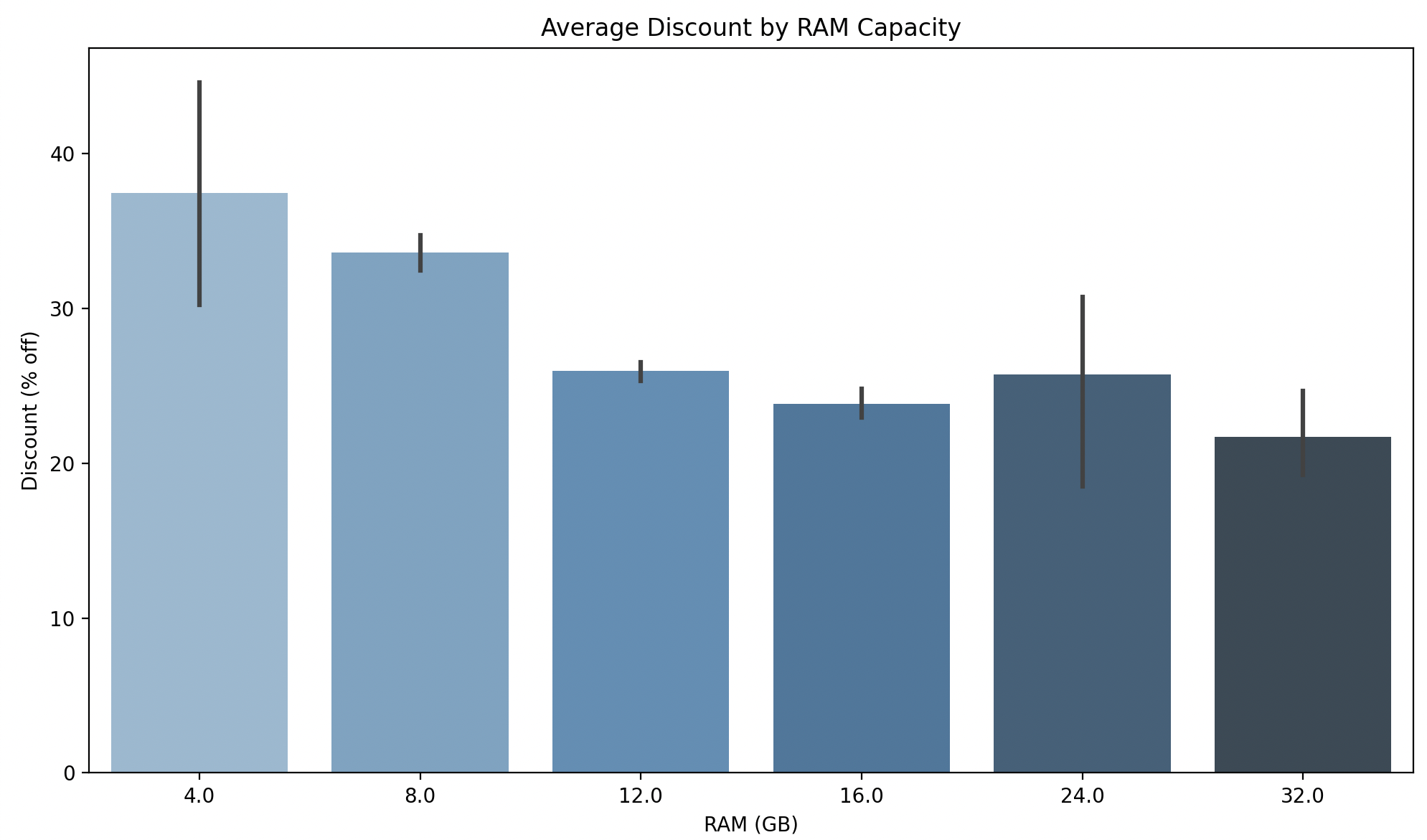
1. **Objective 1**: Identify dominant price segments.
   * **Outcome**: Budget/mid-range laptops (₹20,000–₹60,000) dominate, focusing analysis on these segments.
   * **Graph**: Price distribution histogram.



1. **Objective 2**: Assess satisfaction across processor types.
   * **Outcome**: Similar ratings (~4.2–4.4) across processors, suggesting minimal impact on satisfaction.
   * **Graph**: Rating by processor box plot.



1. **Objective 3**: Identify discount trends by RAM.
   * **Outcome**: 8GB RAM laptops receive the highest discounts (~20–25%), indicating mid-range promotions.
   * **Graph**: Discount by RAM bar plot.

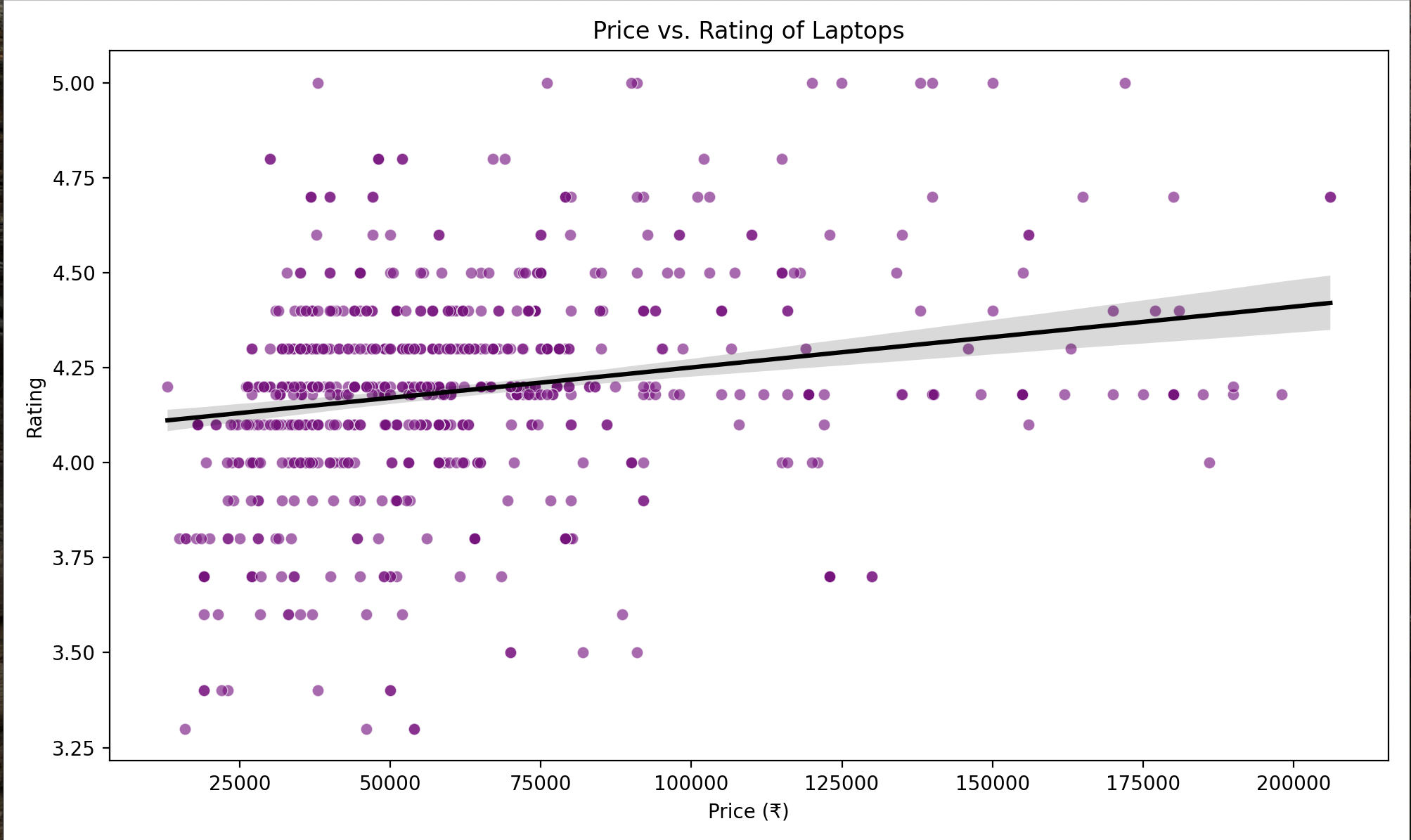


1. **Objective 4**: Identify popular laptops by reviews.
   * **Outcome**: ASUS and Lenovo mid-range models lead, reflecting consumer preference.
   * **Graph**: Top 10 laptops by reviews bar plot.

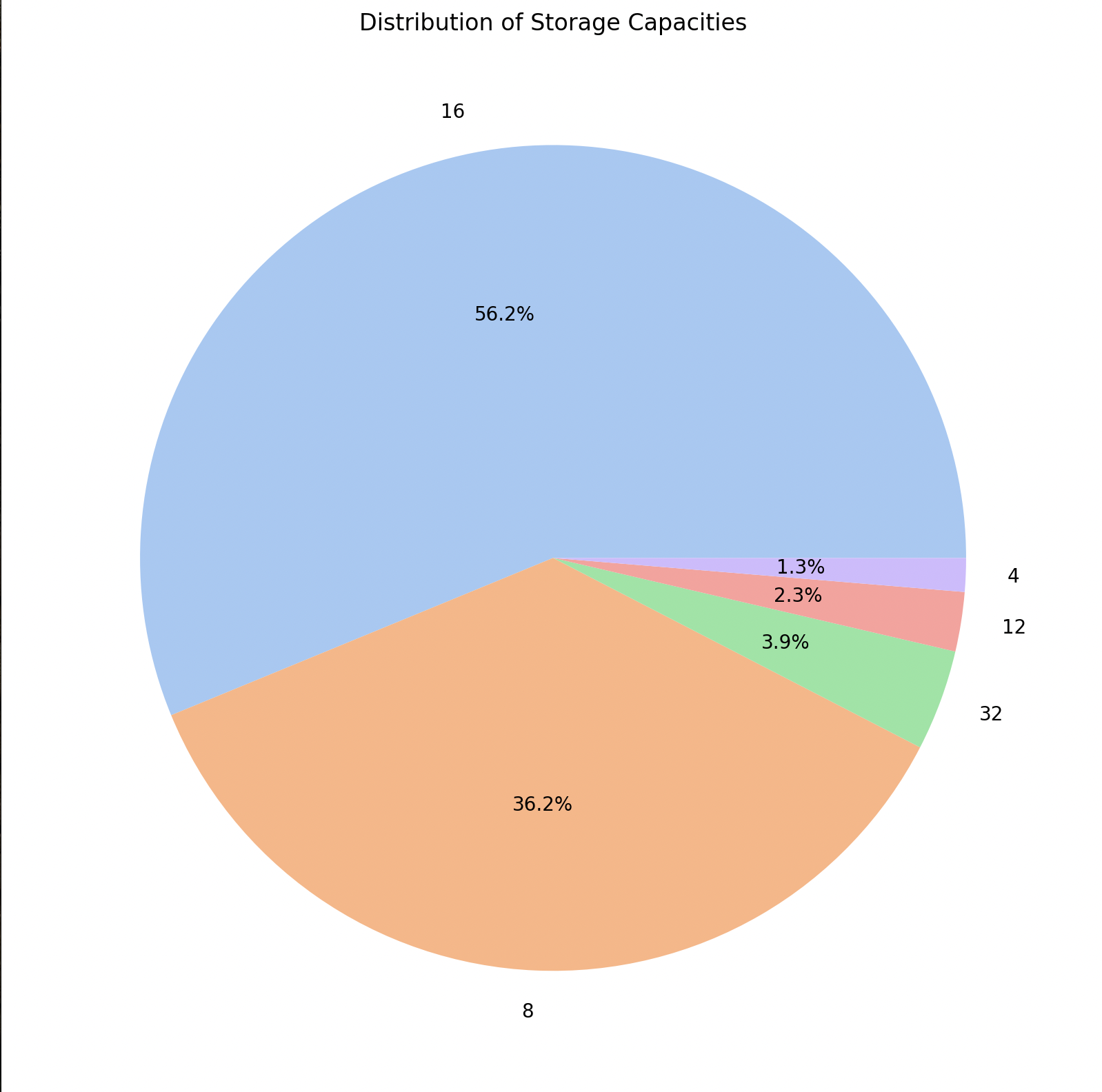
A screenshot of a computer

Description automatically generated

1. **Objective 5**: Determine price-rating relationship.
   * **Outcome**: Weak correlation; budget laptops achieve high ratings, highlighting value.
   * **Graph**: Price vs. rating scatter plot.



1. **Objective 6**: Identify storage trends.
   * **Outcome**: 512GB dominates (~50–60%), aligning with mid-range needs.
   * **Graph**: Storage distribution pie chart.



# 5. Conclusion

# This project successfully analyzed the flipkart\_laptop12.csv dataset to uncover key trends in Flipkart’s laptop market. The EDA and visualizations revealed a competitive mid-range segment, with significant findings:

# Pricing: Most laptops are priced between ₹20,000 and ₹60,000, peaking at ~₹40,000, catering to budget-conscious and mid-range buyers.

# Hardware: 8GB RAM and 512GB storage are market standards, with Intel Core and AMD Ryzen processors leading, reflecting balanced performance demands.

# Ratings: High ratings (4.0–4.5) across models indicate strong customer satisfaction, though limited variability suggests uniform quality.

# Discounts: 8GB RAM laptops receive competitive discounts (~20–25%), driving mid-range sales.

# Popularity: ASUS and Lenovo models, like Vivobook and Ideapad, dominate reviews, signaling consumer trust and engagement.

# The six objectives—price distribution, rating by processor, discount by RAM, top reviews, price vs. rating, and storage distribution—provided a comprehensive view of market dynamics. The cleaned dataset, free of outliers and enriched with extracted features, is well-prepared for advanced analyses, such as price prediction. This project demonstrates proficiency in Python-based data analysis, leveraging Pandas, Seaborn, and Matplotlib to deliver actionable insights. It serves as a strong foundation for further exploration, including predictive modeling and interactive tools to benefit consumers and retailers.

**Future Scope**

The project opens several avenues for enhancement, building on its robust EDA and visualization framework:

1. **Expanded Data Collection**:
   * Scrape additional e-commerce platforms (e.g., Amazon, Croma) to compare laptop pricing and availability.
   * Include features like screen size, GPU type, or battery life for deeper analysis.
2. **Predictive Modeling**:
   * Apply machine learning models like Random Forest or XGBoost to predict laptop prices based on RAM, Storage, Processor, and Rating.
   * Use clustering to segment laptops into categories (e.g., gaming, ultrabooks, budget).
3. **Sentiment Analysis**:
   * If review text is available, analyze customer sentiments to identify strengths and weaknesses of popular models.
4. **Interactive Dashboard**:
   * Develop a Plotly Dash application with filters for Processor, RAM, or price range, and visualizations like price trends or feature comparisons.
   * Deploy the dashboard online to assist buyers and retailers in decision-making.
5. **Practical Applications**:
   * Build a laptop recommendation system (e.g., “Best laptop under ₹50,000 with 16GB RAM”).
   * Provide pricing insights to retailers, optimizing discount strategies for competitive configurations.

These extensions would enhance the project’s utility, transforming it into a comprehensive tool for market analysis and consumer guidance.

Linkedlin : <https://www.linkedin.com/in/shubham-kumar-46422128a/>

### References

The following resources and tools were instrumental in the development and execution of this project:

1. Flipkart. (n.d.). Online Shopping Site for Mobiles, Electronics, Furniture, Grocery, Lifestyle, Books & More. Retrieved from <https://www.flipkart.com>
   * Primary data source for web scraping smartphone product details.
2. Python Software Foundation. (2025). Python 3.11 Documentation. Retrieved from <https://docs.python.org/3/>
   * Official documentation for the Python programming language used throughout the project.
3. Requests. (n.d.). Requests: HTTP for Humans. Retrieved from <https://requests.readthedocs.io/en/latest/>
   * Documentation for the requests library used for HTTP requests in web scraping.
4. Beautiful Soup. (n.d.). Beautiful Soup Documentation. Retrieved from <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
   * Documentation for the BeautifulSoup library used for HTML parsing.
5. Pandas. (n.d.). pandas: Python Data Analysis Library. Retrieved from <https://pandas.pydata.org/docs/>
   * Documentation for the pandas library used for data manipulation and storage.
6. Scikit-learn. (n.d.). scikit-learn: Machine Learning in Python. Retrieved from <https://scikit-learn.org/stable/>
   * Documentation for the scikit-learn library used for predictive modeling and evaluation.
7. Seaborn. (n.d.). seaborn: Statistical Data Visualization. Retrieved from <https://seaborn.pydata.org/>
   * Documentation for the seaborn library used for creating visualizations in exploratory data analysis.
8. Matplotlib. (n.d.). Matplotlib: Visualization with Python. Retrieved from <https://matplotlib.org/>
   * Documentation for the matplotlib library used for plot customization.
9. Plotly. (n.d.). Dash: Analytical Web Apps for Python, R, Julia, and Jupyter. Retrieved from <https://dash.plotly.com/>
   * Documentation for the dash and plotly libraries used for building the interactive dashboard.
10. WordCloud. (n.d.). WordCloud for Python Documentation. Retrieved from <https://amueller.github.io/word_cloud/>
    * Documentation for the WordCloud library used for text visualization.