ReCell: Dynamic Pricing Strategy

ReCell Used Device Pricing and Supervised Learning - Foundations

Robert Paul Ludwig IV
5/10/2024

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Executive Summary

Project Context

- ReCell aims to establish a competitive edge in the used and refurbished device market by implementing a dynamic pricing model. This approach leverages machine learning to optimize pricing strategies and maximize profitability while meeting consumer demands

Link to Appendix slide on data background check

Executive Summary

Actionable Insights

- Price Sensitivity: Analysis indicates a significant correlation between device features such as camera resolution, battery capacity, and connectivity options with the pricing. Tailor pricing strategies to these features to enhance appeal.
- Market Trends:
  - High Demand for 4G and Emerging 5G Markets: Prioritize stocking devices with these technologies to cater to consumer expectations and command premium prices.
  - Depreciation Factors: Newer devices tend to retain value longer. Adjust prices based on age and technological relevance to maintain competitive positioning.
- Consumer Preferences:
  - Devices with higher RAM and internal memory are trending, suggesting a market preference for performance-oriented features.
  - Large screens and high-resolution cameras are popular, reflecting consumer interest in advanced multimedia capabilities.

Link to Appendix slide on data background check

Executive Summary

Recommendations

- Dynamic Pricing Model Implementation:
  - Integrate the predictive pricing model to continuously update prices based on real-time data and market fluctuations.
  - Use the model's output to adjust prices dynamically, considering factors like release year, tech specs, and days used.
- Inventory Management:
  - Align inventory acquisition with consumer preference trends for specifications like battery life, camera quality, and connectivity.
  - Increase focus on devices with higher resale value and consumer demand based on model insights.
- Marketing and Sales Strategies:
  - Develop targeted marketing campaigns that highlight the premium features of high-demand devices.
  - Offer promotional pricing or bundled offers for older or less popular models to increase turnover.
- Sustainability Focus:
  - Emphasize the environmental benefits of purchasing refurbished devices in marketing efforts to attract eco-conscious consumers.
  - Offer warranties and certifications for refurbished devices to increase consumer trust and perceived value.

Link to Appendix slide on data background check

Executive Summary

Conclusion

- Adopting these insights and recommendations will enable ReCell to optimize its pricing strategy effectively, respond to market trends proactively, and enhance customer satisfaction, thereby securing a stronger market position in the growing sector of used and refurbished devices.

Link to Appendix slide on data background check

Business Problem Overview and Solution Approach

Problem Definition

- Market Context:
  - The used and refurbished phone market is growing rapidly, with an expected value of $52.7 billion by 2023 and a CAGR of 13.6% from 2018 to 2023. This growth is fueled by increased consumer demand for cost-effective alternatives to new devices.
  - COVID-19 has further amplified this trend, pushing consumers to opt for economical options amid financial constraints.
- Business Challenge:
  - ReCell, a startup entering this lucrative market, seeks to capitalize on this opportunity but faces the challenge of accurately pricing used and refurbished devices to maximize sales and profitability.
- Objective:
  - Develop a machine learning-based dynamic pricing strategy that predicts prices of used phones and tablets effectively and identifies key factors influencing these prices.

Link to Appendix slide on data background check

Business Problem Overview and Solution Approach

Solution Approach / Methodology

- Data-Driven Insights:
  - Utilize the provided dataset comprising attributes such as brand name, OS, screen size, connectivity options, camera resolutions, internal memory, RAM, battery capacity, weight, release year, days used, and prices.
  - Perform extensive Exploratory Data Analysis (EDA) to understand distribution, central tendencies, and outlier impacts on prices and device attributes.
- Model Development:
  - Linear Regression Model: Build a model to predict the normalized used price based on relevant features identified during EDA.
  - Feature Engineering: Enhance the model's predictive power by creating new features like "Years Since Release" to capture depreciation effects.
  - Data Preprocessing: Handle missing values through imputation, scale numerical inputs, and encode categorical variables to prepare data for modeling.

Business Problem Overview and Solution Approach

Solution Approach / Methodology

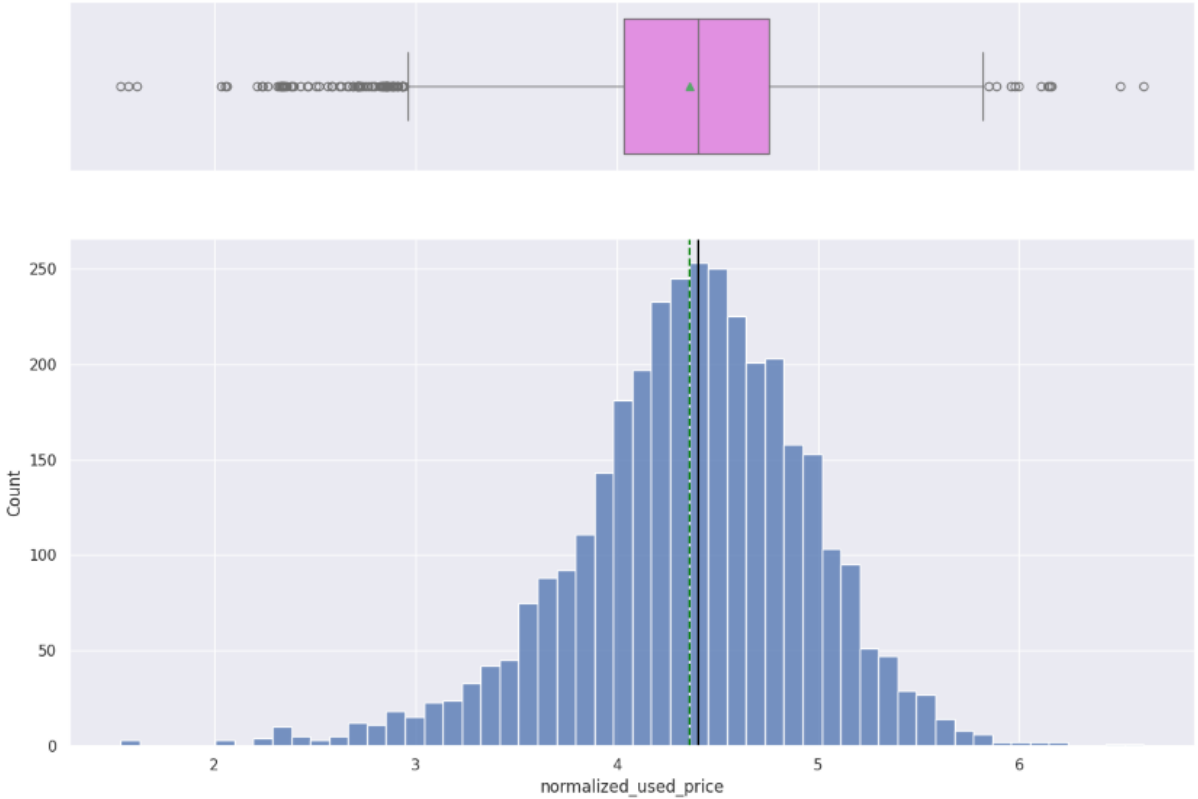
- Evaluation and Refinement:
  - Evaluate the model using metrics like RMSE, MAE, and R-squared values to ensure accuracy and reliability.
  - Refine the model by addressing any potential multicollinearity and ensuring it captures essential price drivers effectively.
- Implementation Strategy:
  - Develop a dynamic pricing tool that ReCell can integrate into their systems to adjust prices based on market data and internal analytics.
  - Continuous monitoring and updating of the model to adapt to market changes and consumer behavior shifts.

Analysis of Normalized Used Price Data

- Key Insights:
  - Distribution: Prices are somewhat normally distributed with a right skew, indicating higher prices are less frequent but notable.
  - Central Tendency: Prices cluster around the median, showing a typical price range for most devices.
  - Outliers: Significant at both low and high ends, suggesting special cases like devices in poor condition or high-demand models.
- Implications for ReCell:
  - Data Handling: Consider data transformation or robust regression to address skewness and outliers.
  - Model Selection: Evaluate models that can manage skew and outliers effectively.
  - Pricing Strategy: Tailor pricing for devices around the median and review strategies for outlier pricing for competitive edge.

Link to Appendix slide on data background check

Analysis of Normalized Used Price Data



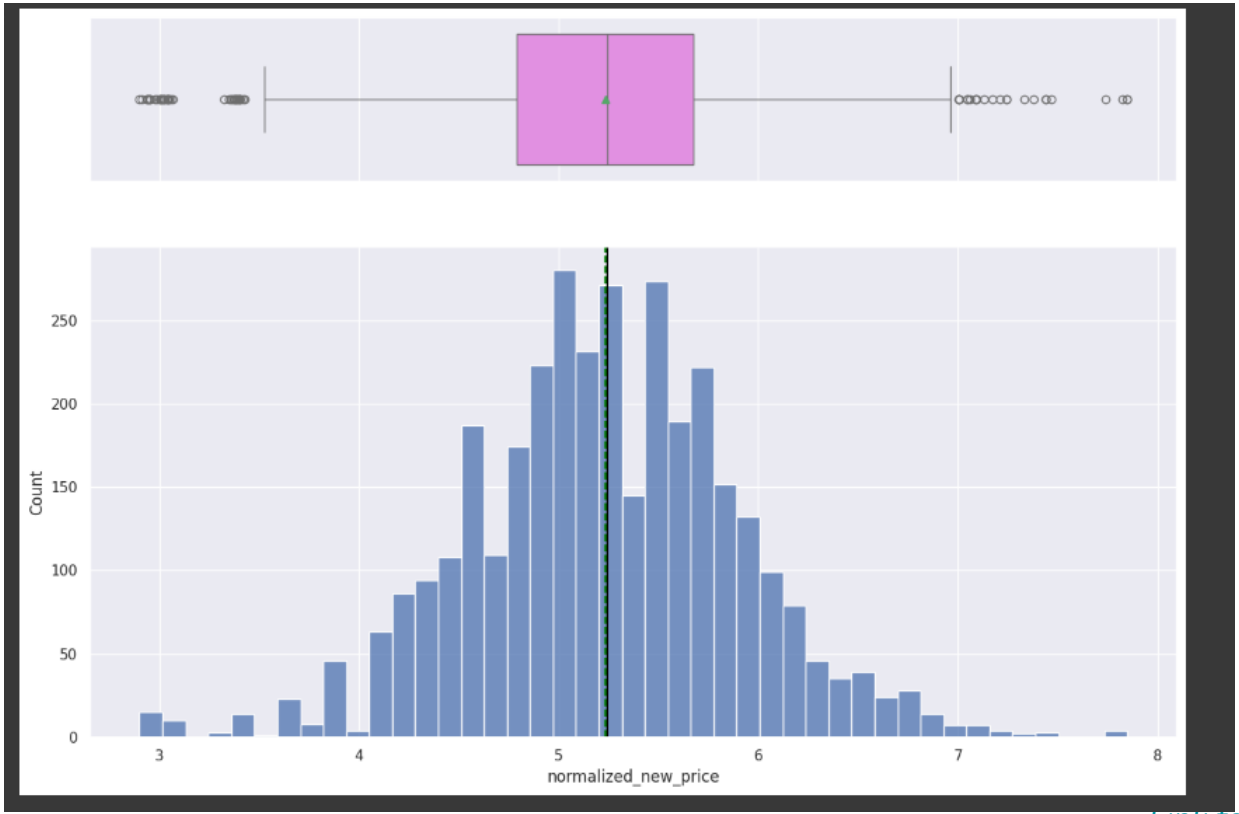
Link to Appendix slide on data background check

Analysis of Normalized New Price Data

- Key Insights:
  - Distribution: Prices generally follow a normal distribution with a noticeable skew toward higher prices.
  - Central Tendency: The median indicates the central typical price, while the mean being higher than the median confirms the right skew.
  - Outliers: Numerous high-price outliers suggest the presence of premium models that exceed typical pricing levels.
- Implications for ReCell:
  - Data Handling: Consider transformations or robust regression to manage skewness and outliers in modeling.
  - Model Selection: Adapt model strategies to effectively handle the distribution characteristics.
  - Pricing Strategy: Align ReCell's pricing not just against average new device prices but also consider positioning against premium models.

Link to Appendix slide on data background check

Analysis of Normalized New Price Data



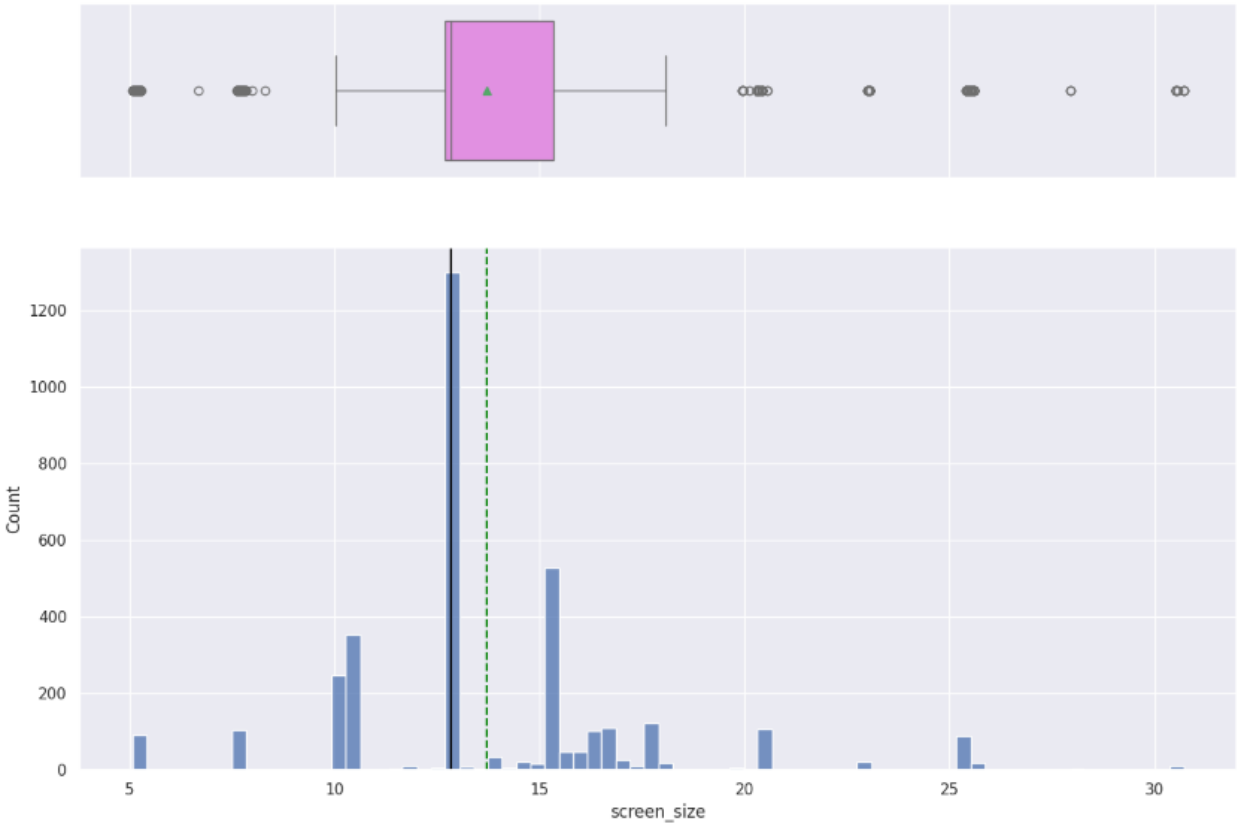
Link to Appendix slide on data background check

Analysis of Screen Size Data

- Key Insights:
  - Distribution: Screen sizes are concentrated around a common standard, indicating a preferred range within the market.
  - Central Tendency: The median and mean are close, suggesting a symmetric distribution with no significant skew.
  - Outliers: Few outliers indicate that most devices fall within a standard screen size range, with rare exceptions being exceptionally small or large.
- Implications for ReCell:
  - Market Trends: The concentration around standard screen sizes reflects prevalent consumer preferences, guiding inventory selection.
  - Data Quality: Uniform distribution ensures consistent data quality with minimal need for outlier management in modeling.
  - Strategic Pricing: Devices with common screen sizes may sell better and could be priced more competitively to enhance turnover.

Link to Appendix slide on data background check

Analysis of Screen Size Data



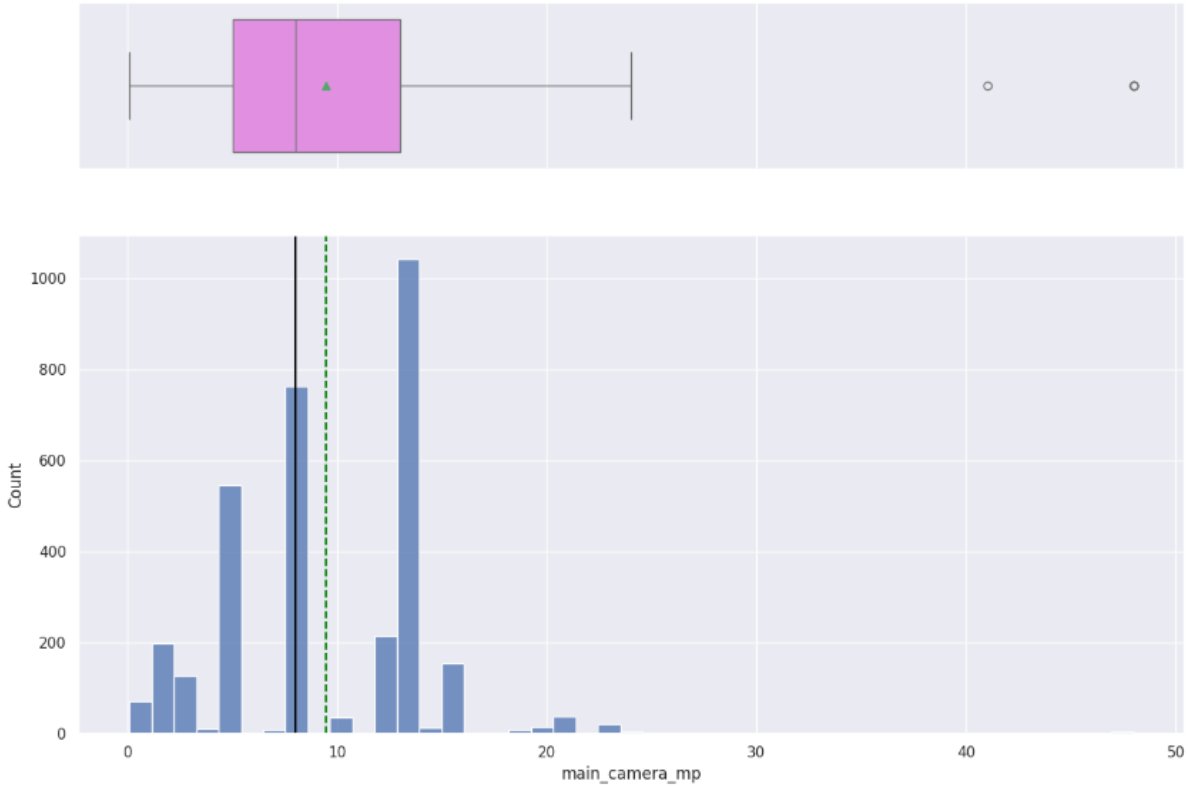
Appendix slide on data background check

Analysis of Main Camera Resolution Data

- Key Insights:
  - Distribution: Concentration around common resolutions like 12 and 48 megapixels, with outliers representing unusually high or low resolutions.
  - Central Tendency: Median and mean resolutions are closely aligned, suggesting a symmetric distribution around popular camera resolutions.
  - Multiple Peaks: Indicates standard camera resolutions in the market, likely driven by consumer preferences and technological trends.
- Implications for ReCell:
  - Consumer Trends: Peaks at popular resolutions highlight consumer preferences, which can guide inventory and pricing strategies.
  - Data Handling: The presence of modes and outliers suggests considering non-linear transformations or categorical treatment for modeling.
  - Strategic Pricing: Emphasize devices with popular resolutions for competitive pricing and market premium resolutions to attract enthusiasts.

Link to Appendix slide on data background check

Analysis of Main Camera Resolution Data



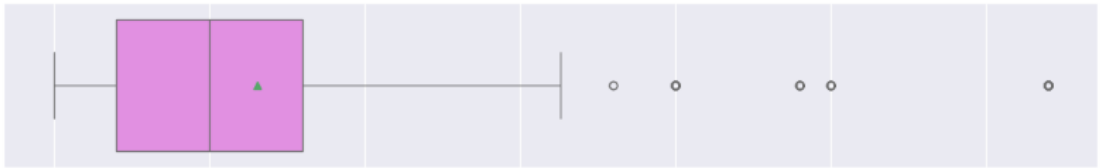
Link to Appendix slide on data background check

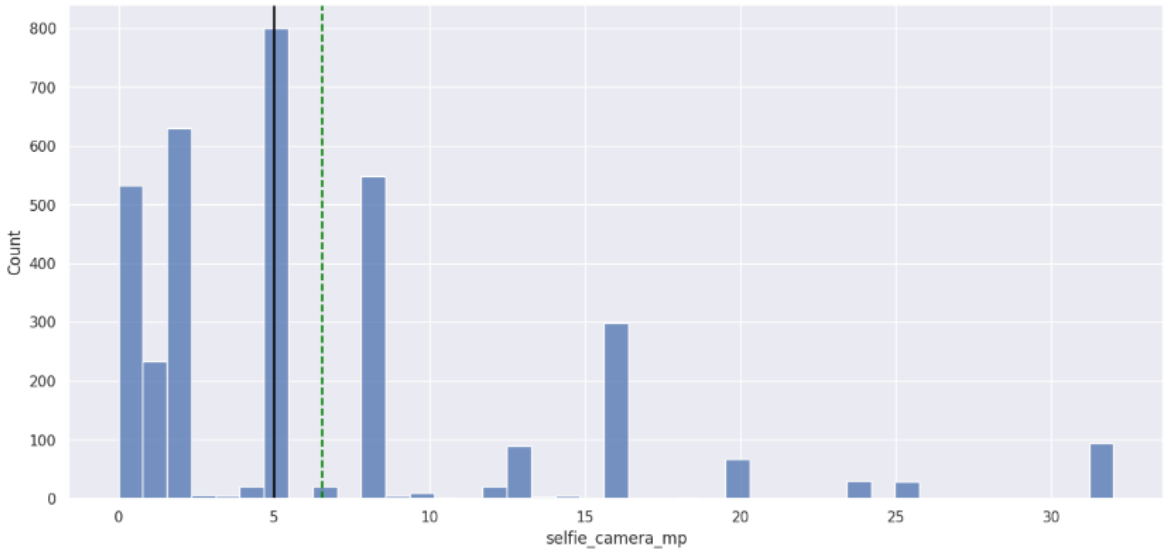
Analysis of Selfie Camera Resolution Data

- Key Insights:
  - Distribution: Multiple peaks in selfie camera resolutions at 5, 8, and 16 megapixels, with outliers indicating very high resolutions.
  - Central Tendency: Median around 8 megapixels, with a mean closely aligned, suggests a balanced distribution around common resolution standards.
  - Variability: The presence of several peaks indicates different popular standards, likely reflecting diverse consumer preferences or technological trends.
- Implications for ReCell:
  - Market Segmentation: The distribution suggests multiple consumer segments, differentiated by selfie camera resolution preferences. This can guide targeted inventory and marketing strategies.
  - Model Adaptation: The multimodal nature and outliers suggest the potential benefit of categorizing resolutions or using non-linear transformations in predictive models to better capture price influences.
  - Strategic Pricing: Understanding which resolutions are most popular or premium can help ReCell strategically price and promote devices, potentially focusing on higher-resolution models for premium pricing.

Link to Appendix slide on data background check

Analysis of Selfie Camera Resolution Data





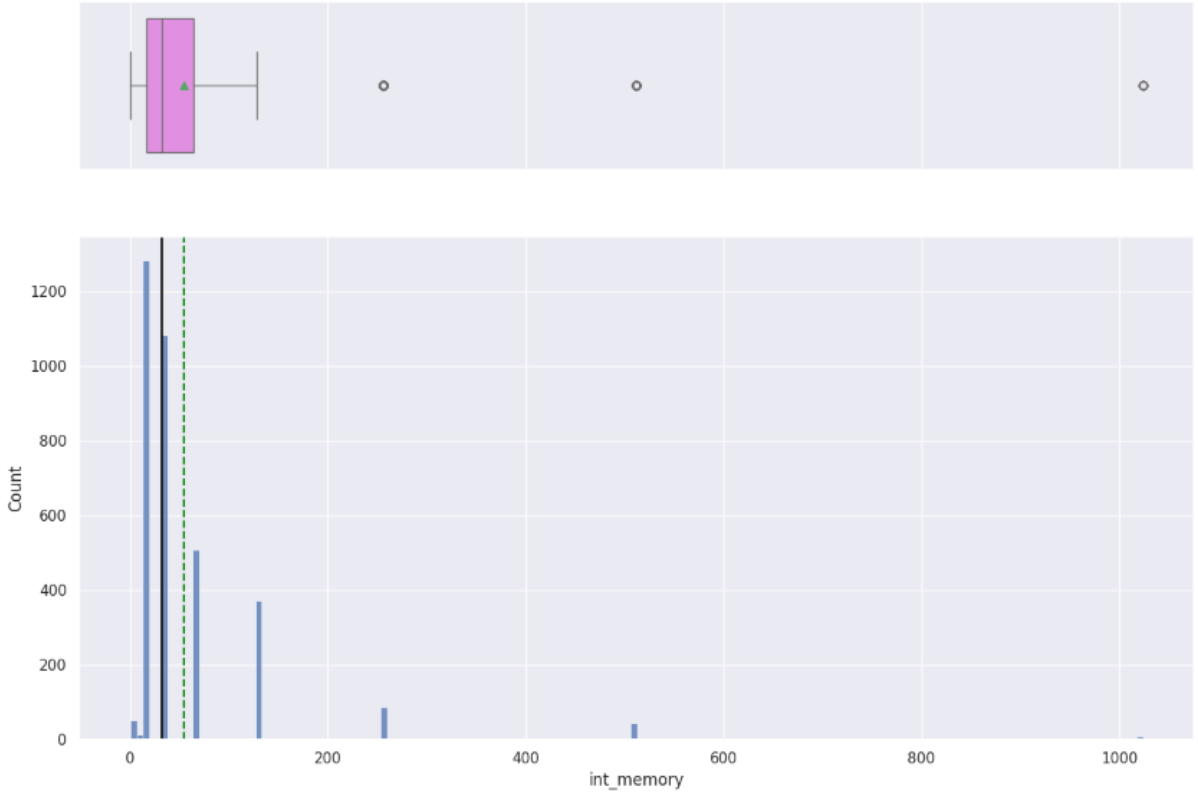
Link to Appendix slide on data background check

Analysis of Internal Memory Data

- Key Insights:
  - Distribution: Majority of devices have internal memory around 64 GB, with a left skew indicating many devices have lower memory capacities.
  - Central Tendency: Median internal memory is 64 GB, while the mean is slightly higher, suggesting higher-memory devices influence the average.
  - Outliers: Devices with unusually high memory are present but rare, suggesting they are premium or specialized models.
- Implications for ReCell:
  - Market Trends: The skew towards lower memory sizes indicates a large segment of the market opts for modest memory, but a niche exists for high-memory devices.
  - Model Adaptation: Due to the skew and outliers, consider data transformations or robust regression models to more accurately predict pricing impacts.
  - Strategic Pricing: Optimize pricing strategies by aligning with common memory sizes for stability and targeting premium pricing for high-memory devices with specialized marketing.

Link to Appendix slide on data background check

Analysis of Internal Memory Data



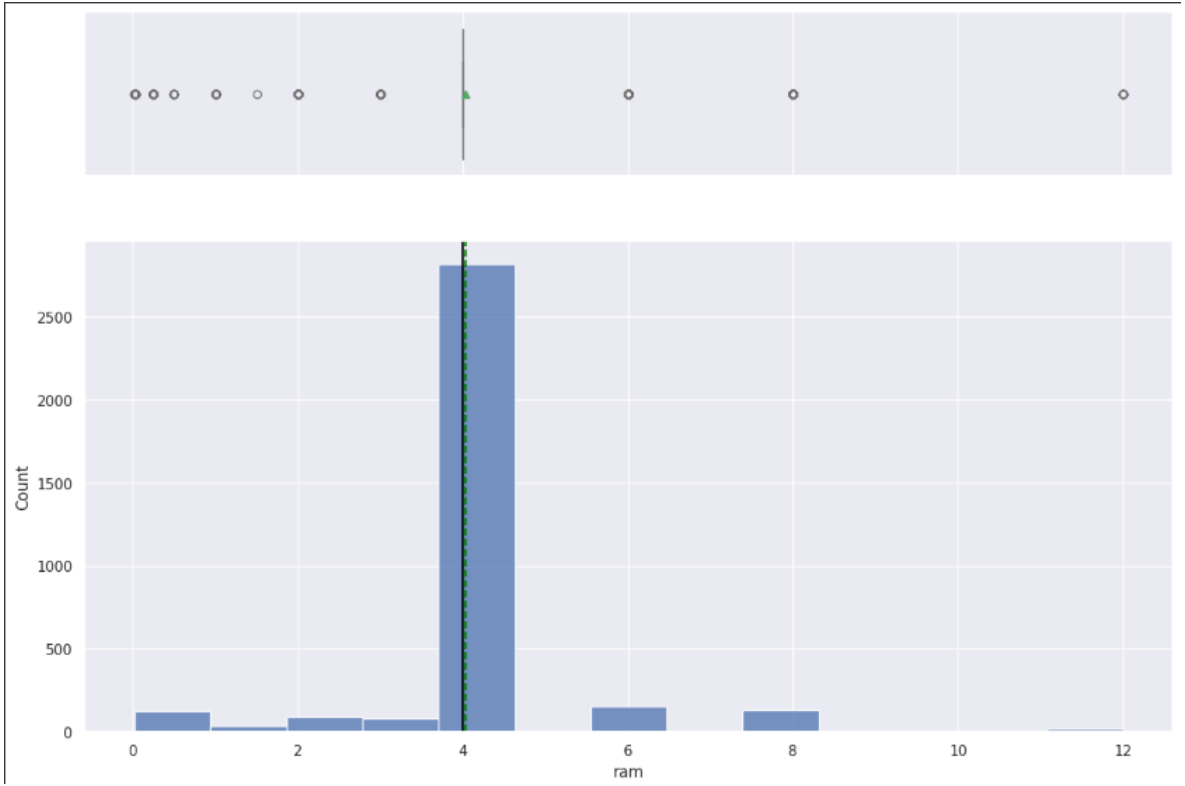
Link to Appendix slide on data background check

Analysis of RAM Size Data

- Key Insights:
  - Distribution: Concentrated around a median of 4 GB, with most devices featuring this RAM size, reflecting a standard market preference.
  - Central Tendency: Median and mean RAM sizes are closely aligned, suggesting a symmetric distribution centered around 4 GB.
  - Outliers: There are outliers at higher RAM sizes (6, 8, 12 GB), indicating the presence of higher-end devices.
- Implications for ReCell:
  - Market Trends: The prevalence of 4 GB RAM in devices suggests it's a common choice among consumers, likely due to balancing performance and cost.
  - Model Adaptation: The strong central peak and presence of outliers suggest that RAM size is a significant predictor of pricing, with deviations potentially commanding premium prices.
  - Strategic Pricing: Focus inventory on popular 4 GB models for stable demand, while strategically pricing and marketing higher RAM devices to target performance-focused consumers.

Link to Appendix slide on data background check

Analysis of RAM Size Data



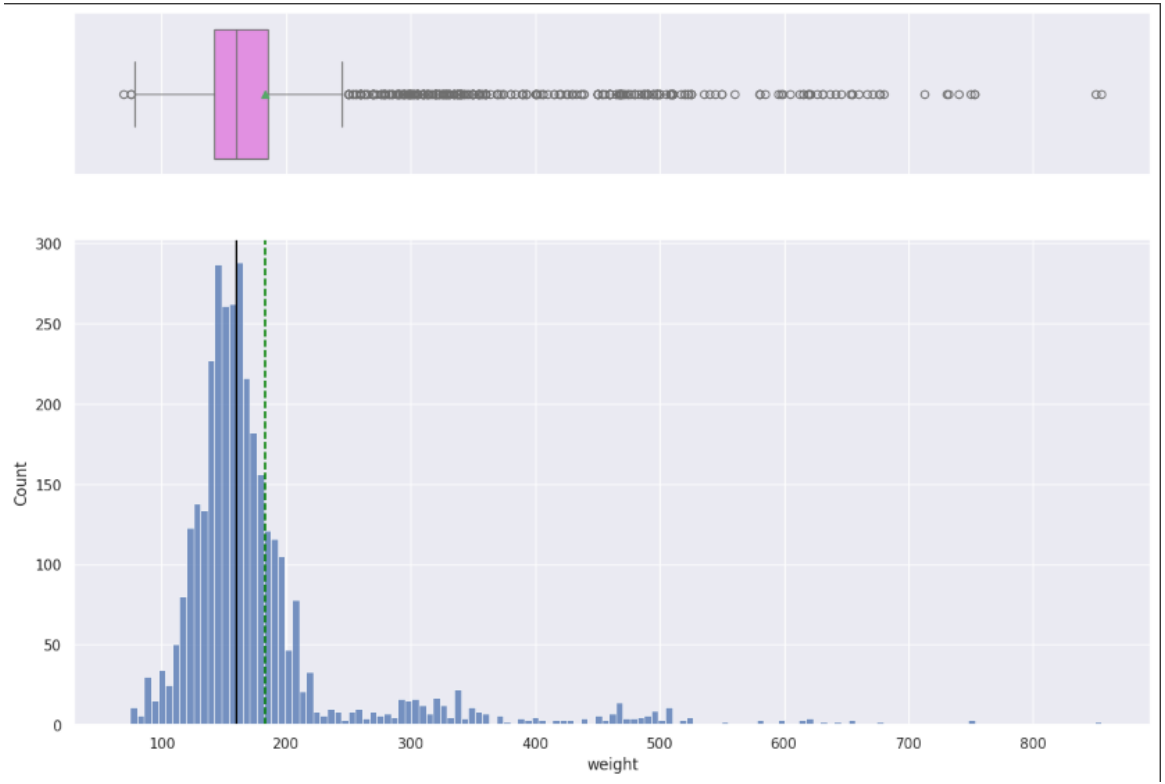
Link to Appendix slide on data background check

Analysis of Device Weight Data

- Key Insights:
  - Distribution: Most devices cluster around a median weight of 150 grams, with a concentration in this central range indicating standardization.
  - Central Tendency: The median and mean weights are closely aligned, reinforcing the prevalence of a common weight standard.
  - Outliers: Significant outliers at both lighter and heavier extremes suggest the presence of compact phones and larger devices like tablets.
- Implications for ReCell:
  - Design Trends: The common weight range likely reflects consumer preferences for portability and ergonomics, important in device design.
  - Model Considerations: The normal-like distribution with outliers suggests the need for careful handling in predictive modeling to account for niche market devices.
  - Strategic Decisions: Focus on devices within the typical weight range for broad appeal, while considering specialized marketing for outliers to target niche markets.

Link to Appendix slide on data background check

Analysis of Device Weight Data



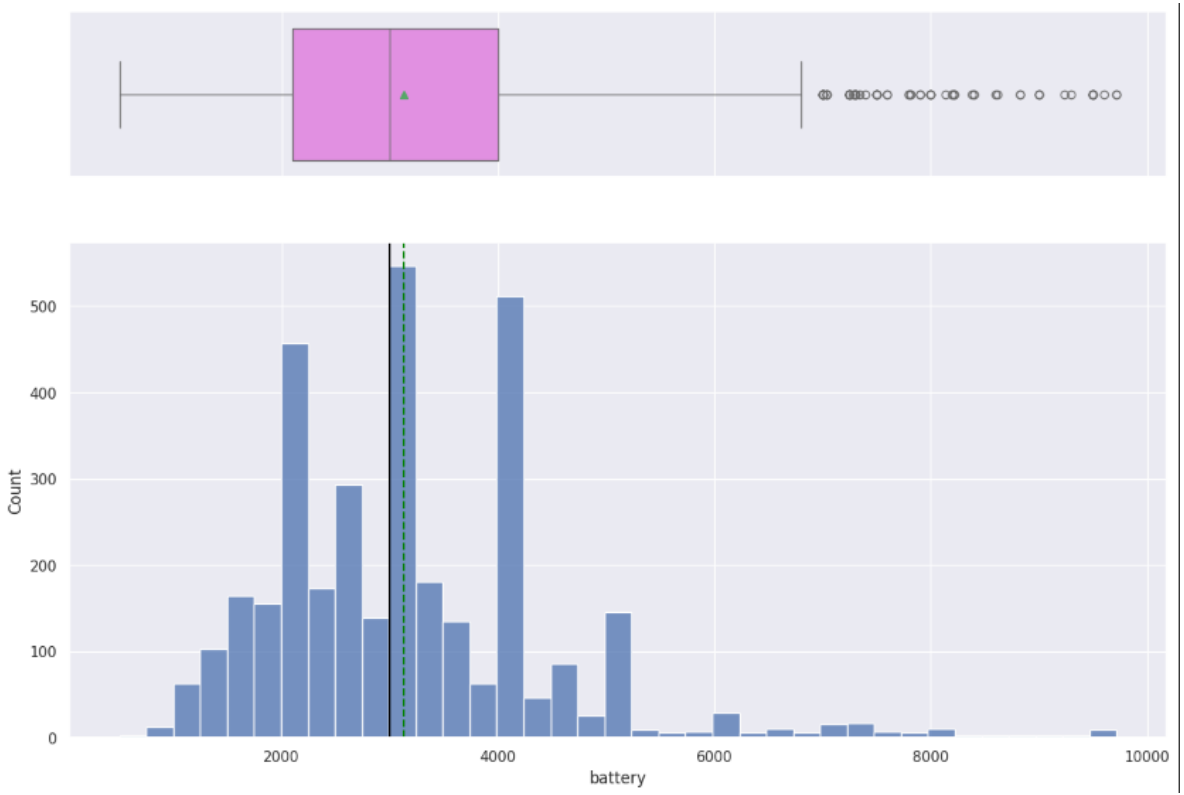
Link to Appendix slide on data background check

Analysis of Battery Capacity Data

- Key Insights:
  - Distribution: Predominant peaks at 3000 mAh and 5000 mAh, with outliers indicating higher capacities typical of premium or high-performance devices.
  - Central Tendency: Median and mean are closely aligned around 3000 mAh, suggesting a balanced distribution around standard battery sizes.
  - Outliers: Presence of higher-capacity outliers highlights the availability of devices with enhanced battery performance.
- Implications for ReCell:
  - Market Trends: The common battery sizes reflect consumer preferences for standard and extended battery life, which can influence buying decisions.
  - Model Adaptation: Consider handling outliers and non-linear relationships in predictive models to better account for the impact of varied battery capacities on pricing.
  - Strategic Pricing: Devices with standard capacities can target a broad market, while high-capacity models might be marketed to power users, potentially at premium prices.

Link to Appendix slide on data background check

Analysis of Battery Capacity Data



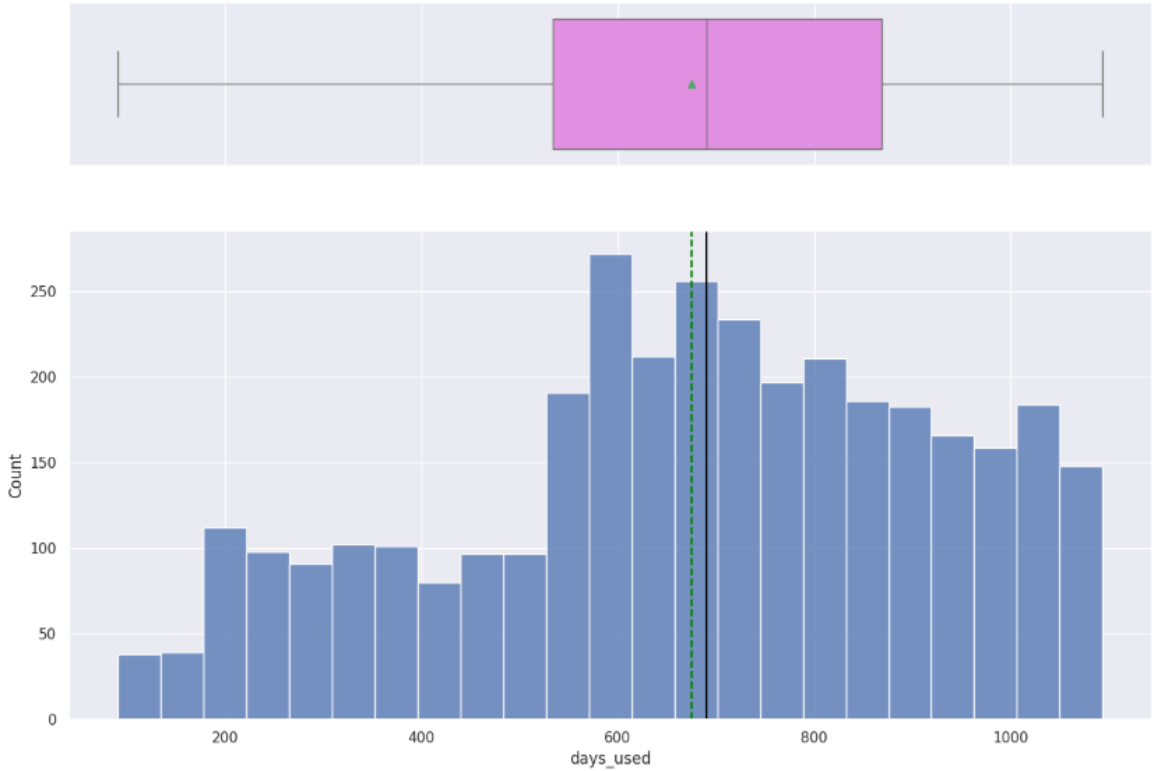
Link to Appendix slide on data background check

Analysis of Days Used Data

- Key Insights:
  - Distribution: Centered around a median of 600 days with a slight right skew, showing most devices are sold after a similar usage period.
  - Central Tendency: Median and mean days used are closely aligned, suggesting a symmetric distribution with some devices used longer than average.
  - Outliers: Indicates that some devices are used significantly longer than typical, pointing to varied device lifecycles.
- Implications for ReCell:
  - Lifecycle Insights: The average usage period around 600 days provides a benchmark for evaluating device condition and pricing accordingly.
  - Model Considerations: The presence of outliers with high days used should be factored into predictive models to enhance pricing accuracy for devices with extended use.
  - Strategic Pricing: Understanding the impact of usage duration on device condition and value can guide ReCell in pricing adjustments and marketing strategies for devices with lower or higher than average days used.

Link to Appendix slide on data background check

Analysis of Days Used Data



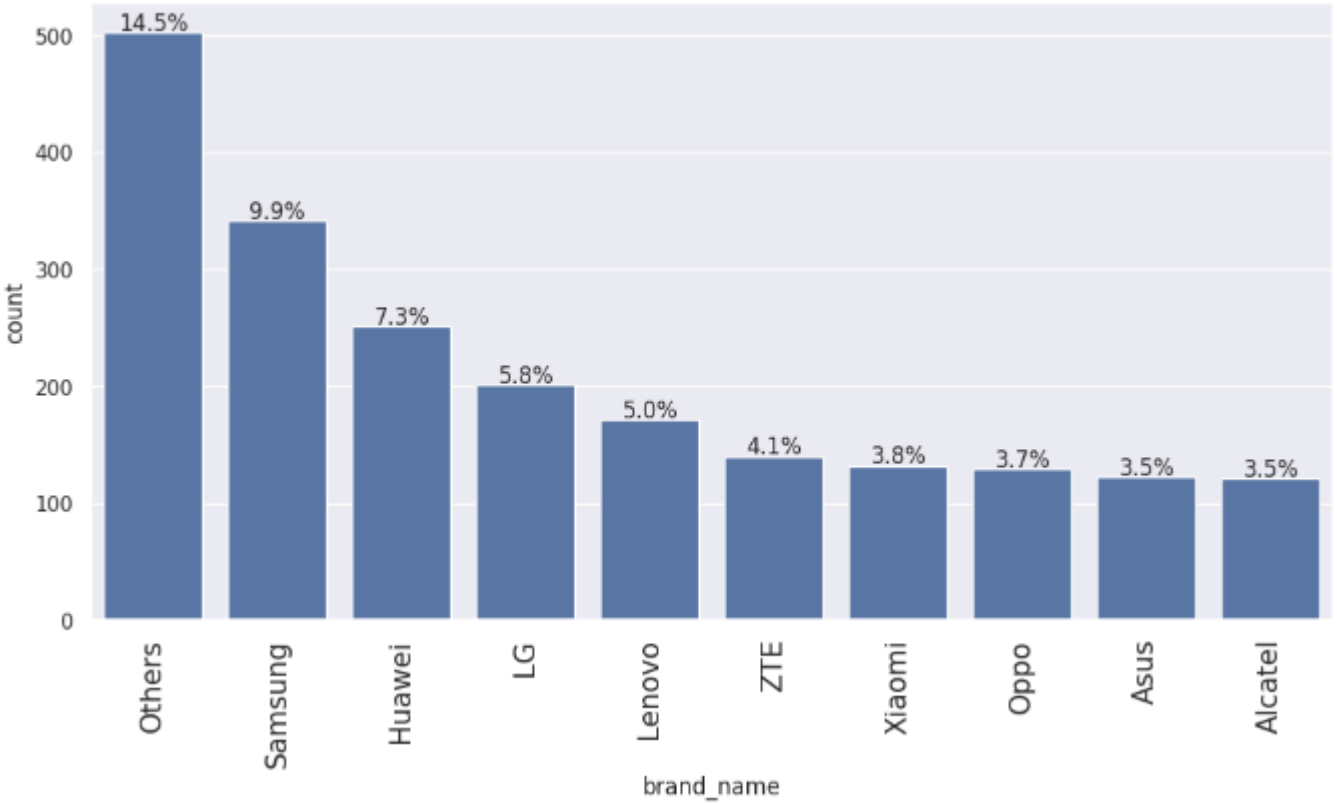
Link to Appendix slide on data background check

Analysis of Device Brand Distribution

- Key Insights:
  - Leading Brands: "Others" leads with 14.5%, followed by Samsung at 9.9%, and Huawei at 7.3%, indicating a fragmented market with no single dominant brand.
  - Brand Diversity: Additional brands like LG, Lenovo, ZTE, Xiaomi, Oppo, Asus, and Alcatel show notable shares, reflecting a competitive and diverse marketplace.
- Implications for ReCell:
  - Market Insights: The high diversity in brands, particularly the significant share of the "Others" category, highlights the need for ReCell to cater to a broad range of consumer preferences.
  - Strategic Inventory: Focus on stocking popular brands to align with market demand, enhancing turnover rates and customer satisfaction.
  - Pricing Strategy: Brand influences pricing significantly due to factors like reputation and quality perception. Prioritize pricing strategies that reflect brand value in the resale market.
  - Targeted Marketing: Tailor marketing efforts to highlight popular and high-value brands to attract the right customer segments.

Link to Appendix slide on data background check

Analysis of Device Brand Distribution



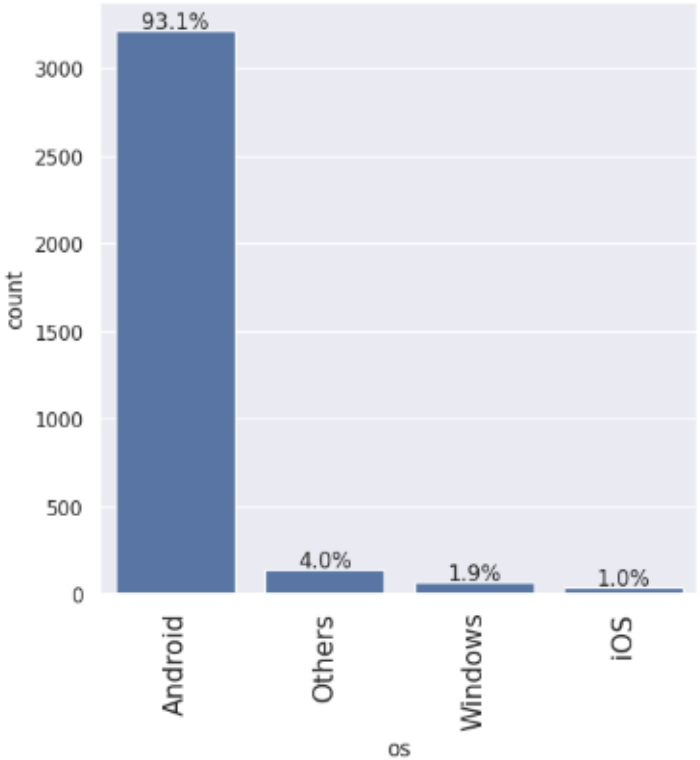
Link to Appendix slide on data background check

Analysis of Operating System Distribution

- Key Insights:
  - Dominant OS: Android dominates the dataset with 93.1%, underscoring its significant presence in the used device market.
  - Other OS Distribution: Minor categories include "Others" at 4.0%, Windows at 1.9%, and iOS at 1.0%, indicating a much smaller market share compared to Android.
- Implications for ReCell:
  - Market Focus: The overwhelming dominance of Android suggests that ReCell's inventory and pricing strategies should primarily cater to Android devices.
  - Strategic Inventory: While focusing on Android, maintaining a selective inventory of iOS and Windows devices could capture niche market segments.
  - Pricing Considerations: The scarcity of iOS and Windows devices in the dataset could justify a premium pricing strategy for these less common OS platforms.
  - Targeted Marketing: Tailor marketing efforts to promote the availability of rare OS platforms like iOS and Windows, potentially attracting consumers seeking these specific devices.

Link to Appendix slide on data background check

Analysis of Operating System Distribution



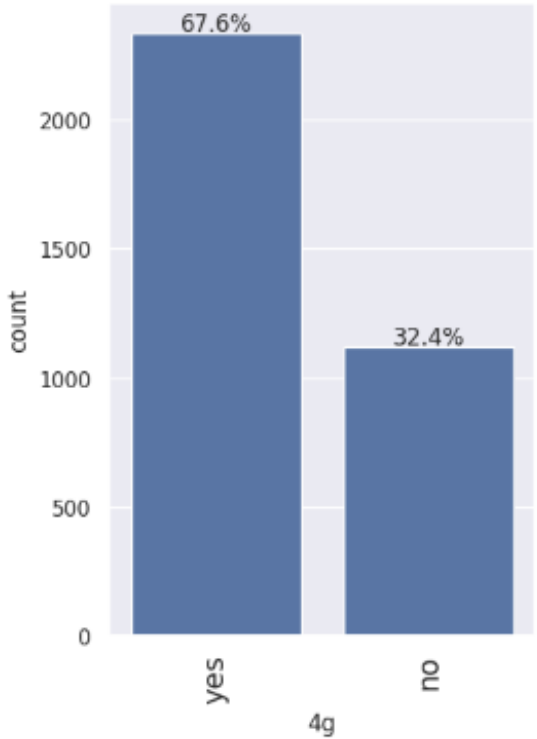
Link to Appendix slide on data background check

Analysis of 4G Connectivity Distribution

- Key Insights:
  - 4G Connectivity: 67.6% of the devices feature 4G connectivity, indicating its prevalence and expected standard among consumers.
  - Lack of 4G Connectivity: 32.4% of devices lack 4G, pointing to a significant segment of the market with different connectivity needs.
- Implications for ReCell:
  - Consumer Expectations: The dominance of 4G-capable devices suggests that most consumers expect this feature, influencing their purchasing decisions.
  - Strategic Inventory: Maintain a strong inventory of 4G devices to meet market demand, while also stocking a variety of non-4G devices for cost-sensitive consumers or less developed areas.
  - Pricing Strategy: 4G devices may command higher prices due to better connectivity and functionality. Pricing for non-4G devices can emphasize affordability.
  - Targeted Marketing: Promote 4G devices for their faster data speeds and reliability, and market non-4G devices as economical alternatives with essential functionalities.

Link to Appendix slide on data background check

Analysis of 4G Connectivity Distribution



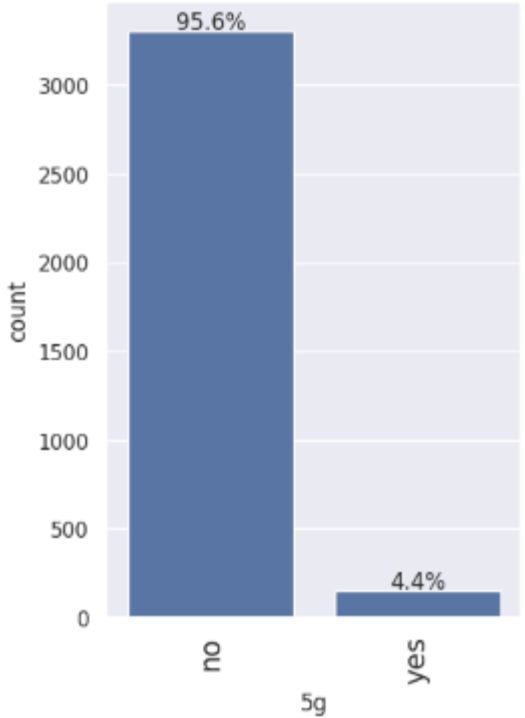
Link to Appendix slide on data background check

Analysis of 5G Connectivity Distribution

- Key Insights:
  - Lack of 5G Connectivity: A vast majority, 95.6%, of devices do not have 5G connectivity, indicating that 5G is still in its nascent stages within this market.
  - Presence of 5G Connectivity: Only 4.4% of devices are equipped with 5G, positioning them as rare and potentially premium offerings.
- Implications for ReCell:
  - Emerging Technology: The minimal presence of 5G devices suggests that the technology is just beginning to penetrate the market, reflecting either an early adoption stage or limited availability during the data collection period.
  - Strategic Inventory: Focus primarily on non-5G devices while including a select inventory of 5G devices to cater to early adopters and tech enthusiasts.
  - Pricing Strategy: Given their scarcity and advanced capabilities, 5G devices could be priced as premium products, appealing to consumers seeking cutting-edge technology.
  - Targeted Marketing: Promote 5G devices as future-proof investments that offer faster data speeds and enhanced network capabilities, appealing to a tech-savvy demographic.

Link to Appendix slide on data background check

Analysis of 5G Connectivity Distribution



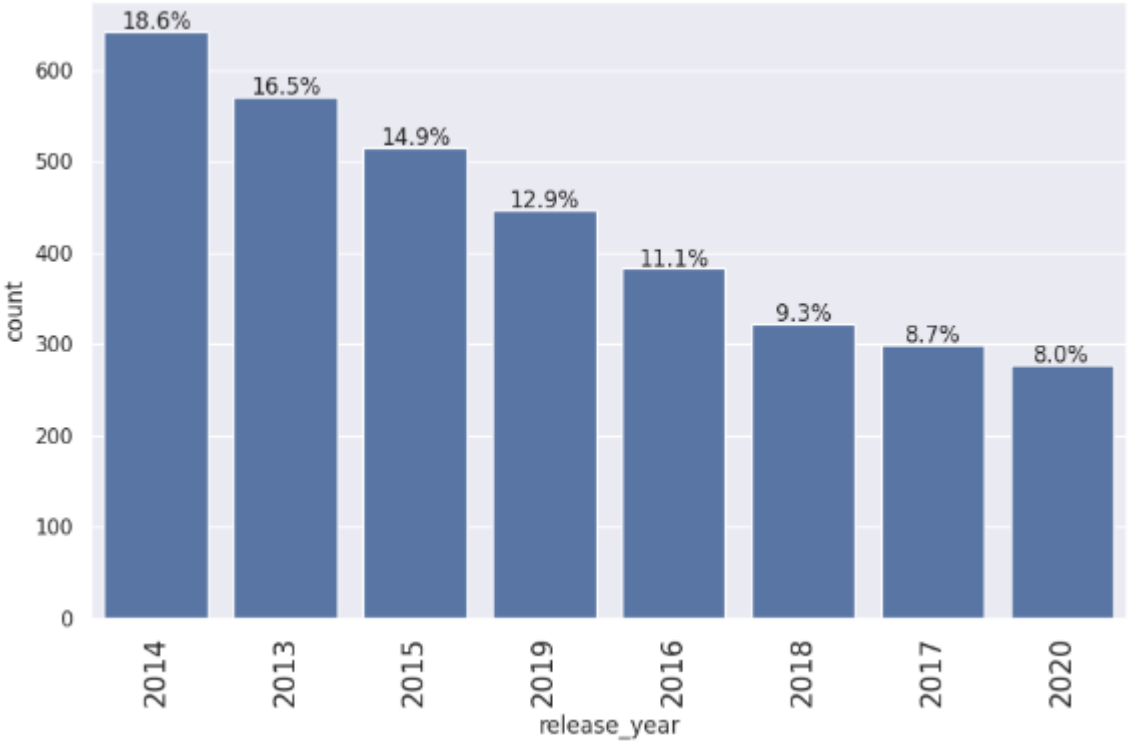
Link to Appendix slide on data background check

Analysis of Device Distribution by Release Year

- Key Insights:
  - Release Year Range: The dataset spans devices released from 2013 to 2020.
  - Highest Representation: 2014 models are the most common, accounting for 18.6% of the dataset.
  - Trend Over Time: There is a decreasing trend in representation from older to newer models, with 2020 models making up only 8.0%.
- Implications for ReCell:
  - Device Age Trends: The prevalence of older models suggests higher durability or prolonged popularity, while fewer recent models may indicate rapid technology turnover or shorter entry times into the secondary market.
  - Pricing Strategy: Older models might be priced lower due to depreciation but could have stable demand, whereas newer models, being rarer, might retain more value and command higher prices.
  - Inventory Management: Optimize stock by focusing on well-represented older models while maintaining a selection of newer models to cater to demand for the latest technology.
  - Market Insights: Analyzing why certain models are more common can offer insights into consumer behavior, usage cycles, and market dynamics, informing strategic decisions in marketing and pricing.

Link to Appendix slide on data background check

Analysis of Device Distribution by Release Year



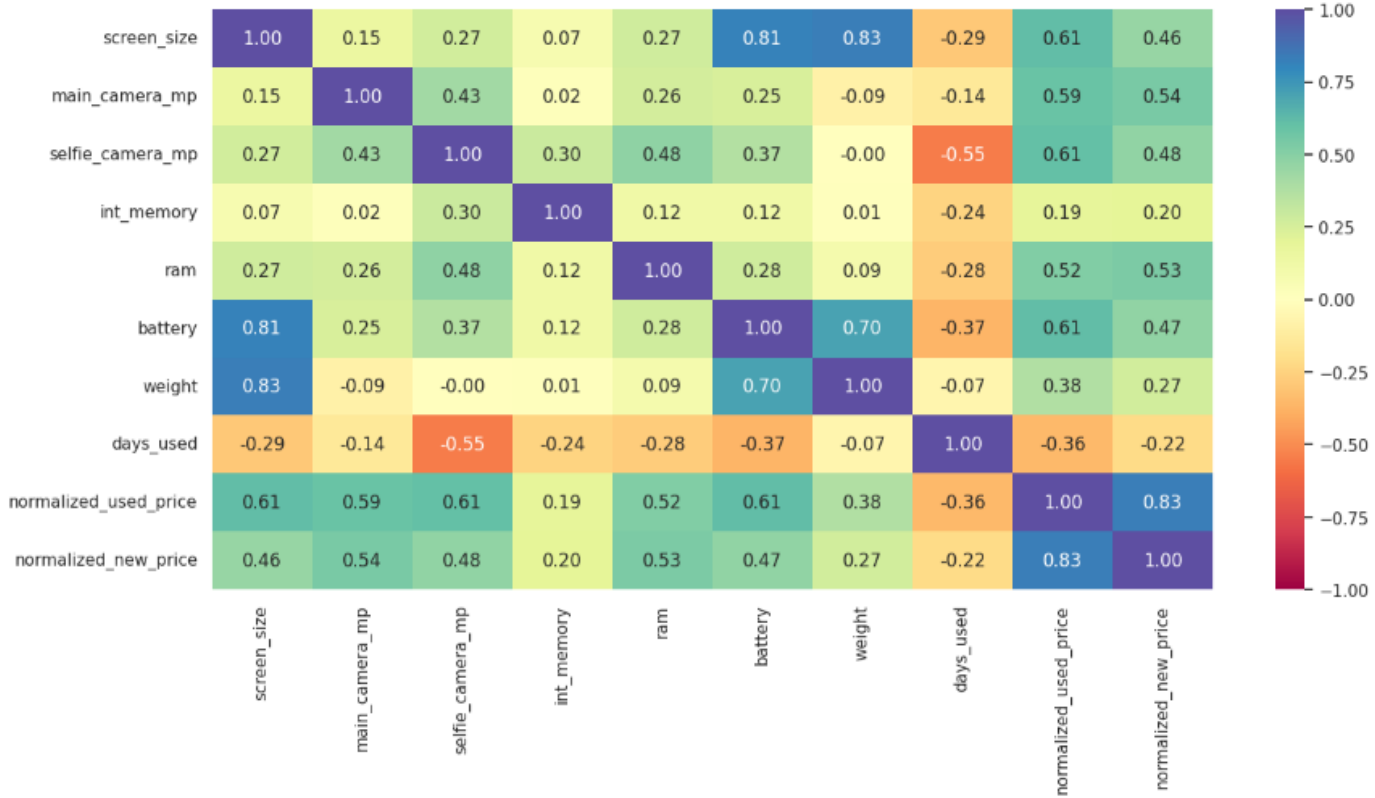
Link to Appendix slide on data background check

Analysis of Correlation Heatmap for Device Attributes

- Key Observations:
  - Strong Positive Correlations:
    - Weight and battery (0.70): Heavier devices typically have larger batteries.
    - Screen size and weight (0.83): Larger devices are heavier.
    - Screen size and battery (0.81): Bigger screens require more battery capacity.
    - Normalized used price and normalized new price (0.83): Higher new prices correlate with higher used prices.
  - Moderate Positive Correlations:
    - Main camera and selfie camera resolutions (0.43): Better main cameras often come with better selfie cameras.
    - Main camera resolution and normalized used price (0.59): Higher camera quality increases used device prices.
    - Battery capacity and normalized used price (0.61): Larger batteries can raise used prices.
  - Negative Correlations:
    - Days used and selfie camera resolution (-0.55): Older devices tend to have lower-resolution selfie cameras.
    - Days used and normalized used price (-0.36): More usage typically lowers resale value.
- Implications for ReCell:
  - Pricing Model Input: Include key features like battery, screen size, and camera quality in the pricing model due to their significant impact on used device prices.
  - Feature Selection: Address potential multicollinearity between highly correlated features to improve model accuracy.
  - Strategic Decisions: Use insights from feature correlations to prioritize device features that enhance value in ReCell's inventory and marketing strategies.

Link to Appendix slide on data background check

Analysis of Correlation Heatmap for Device Attributes



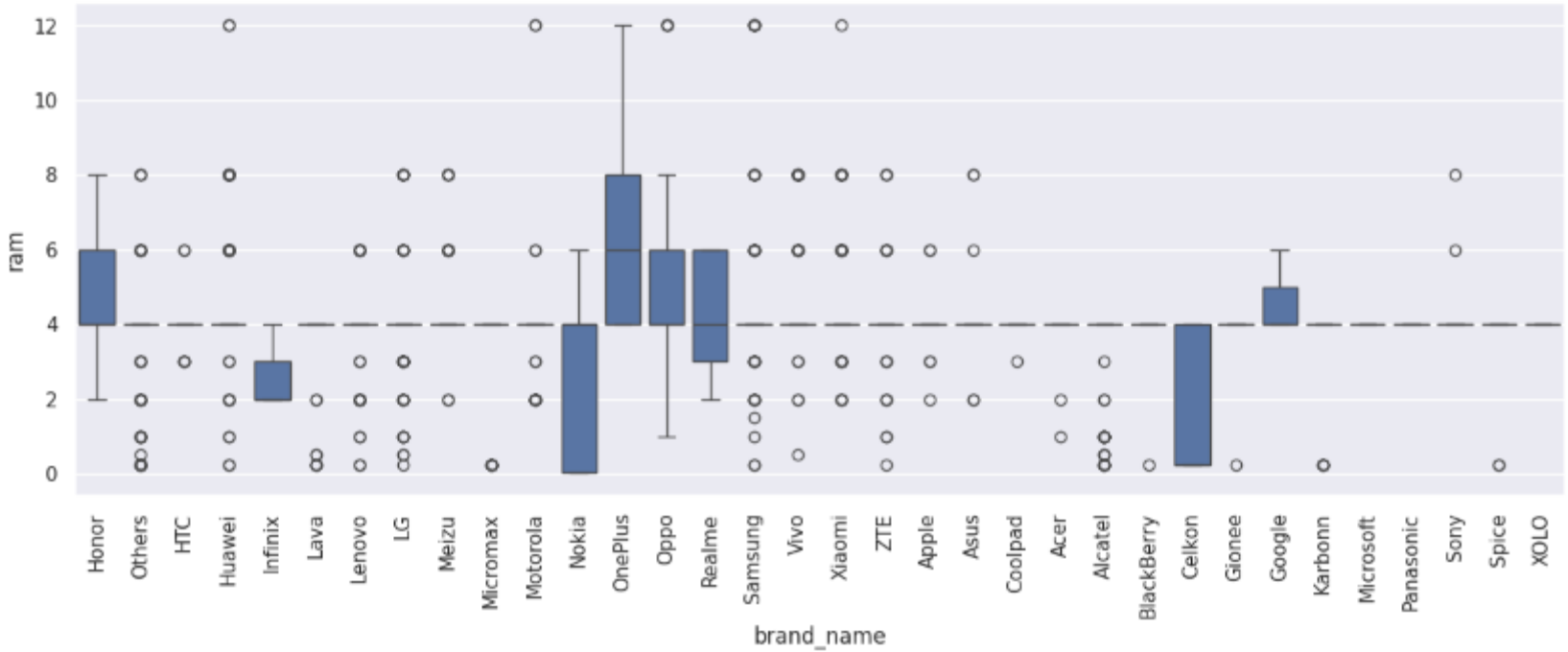
Link to Appendix slide on data background check

Analysis of RAM Distribution by Device Brand

- Key Observations:
  - Variation in RAM: Brands like OnePlus and Samsung show higher median RAM and a broader range, indicative of their focus on high-performance devices.
  - Consistency in Offerings: Brands like Apple display a smaller range but consistent RAM offerings, reflecting a controlled product strategy.
  - Outliers: Presence of outliers in brands like OnePlus suggests some models with exceptionally high RAM, targeting specific high-end market segments.
- Implications for ReCell:
  - Pricing Strategy: Brands with higher RAM typically cater to performance-oriented consumers and can command premium prices due to their advanced specs.
  - Inventory Management: Prioritize stocking devices from brands known for higher RAM, such as OnePlus and Samsung, to attract tech-savvy customers.
  - Marketing Focus: Tailor marketing efforts to highlight devices with higher RAM, appealing particularly to consumers interested in gaming or high-demand applications.

Link to Appendix slide on data background check

Analysis of RAM Distribution by Device Brand



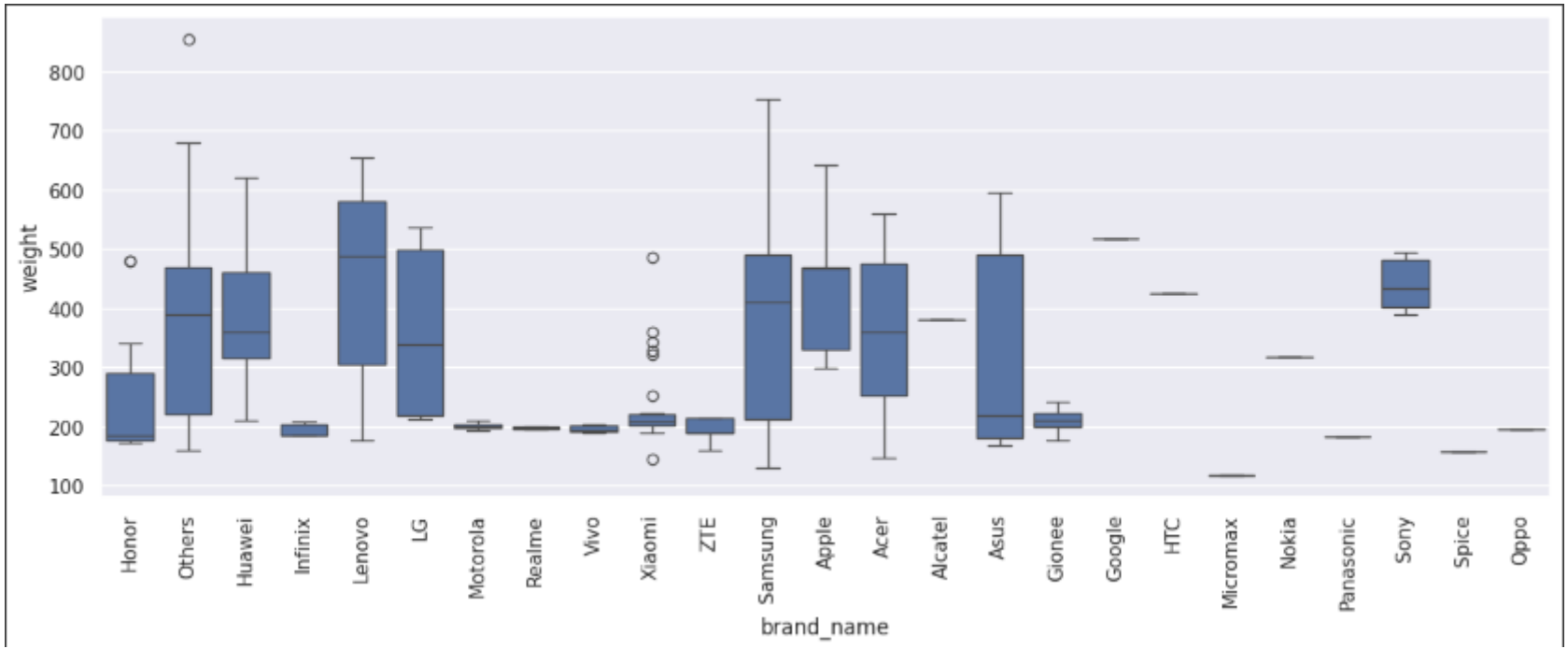
Link to Appendix slide on data background check

Analysis of Device Weight Distribution for Large Battery Devices

- Key Observations:
  - Variation in Weight: Significant variation among brands, with Xiaomi and Samsung showing higher median weights and greater variability, indicative of larger or feature-rich models.
  - Lighter Devices: Apple and Oppo show lower median weights, suggesting the use of compact or lightweight materials despite large battery capacities.
  - Outliers: Presence of outliers highlights models that are exceptionally heavier or lighter than typical offerings from the brands.
- Implications for ReCell:
  - Pricing Considerations: Heavier devices might offer more advanced features, potentially commanding higher prices in the used market.
  - Consumer Targeting: Recognizing weight preferences linked to large battery capacities can help tailor ReCell's inventory to meet diverse consumer needs—some preferring portability, others durability and enhanced features.
  - Strategic Inventory: Balance the stock between lightweight and heavier models to cater to different market segments effectively.
  - Marketing Strategy: Promote the benefits of each weight category—portability for lighter devices and extended features and durability for heavier models.

Link to Appendix slide on data background check

Analysis of Device Weight Distribution for Large Battery Devices



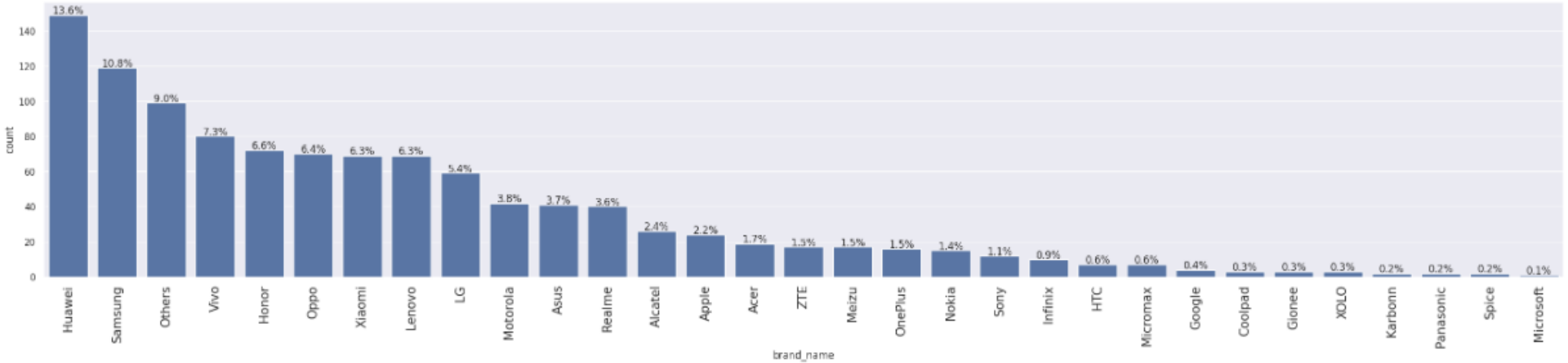
Link to Appendix slide on data background check

Analysis of Brand Distribution for Large Screen Devices

- Key Observations:
  - Leading Brands: Huawei leads with 13.6% of large-screen devices, followed by Samsung (10.8%) and Vivo (9.0%), indicating strong market presence in this segment.
  - Brand Diversity: A wide variety of brands offer large-screen devices, with varying market shares, highlighting diverse consumer offerings.
  - Market Strategies: The significant shares of certain brands suggest targeted strategies towards consumers favoring large displays for enhanced multimedia experiences.
- Implications for ReCell:
  - Market Positioning: Brands like Huawei, Samsung, and Vivo are key players in the large-screen market, appealing to users interested in premium visual experiences.
  - Pricing Strategy: Devices from dominant brands in this segment may command higher prices due to their perceived value for entertainment, productivity, and gaming.
  - Inventory Decisions: Prioritize stocking devices from popular large-screen brands to align with market demand and consumer preferences.
  - Targeted Marketing: Focus marketing on the advantages of large-screen devices, especially those from leading brands, to attract consumers seeking superior multimedia and productivity capabilities.

Link to Appendix slide on data background check

Analysis of Brand Distribution for Large Screen Devices



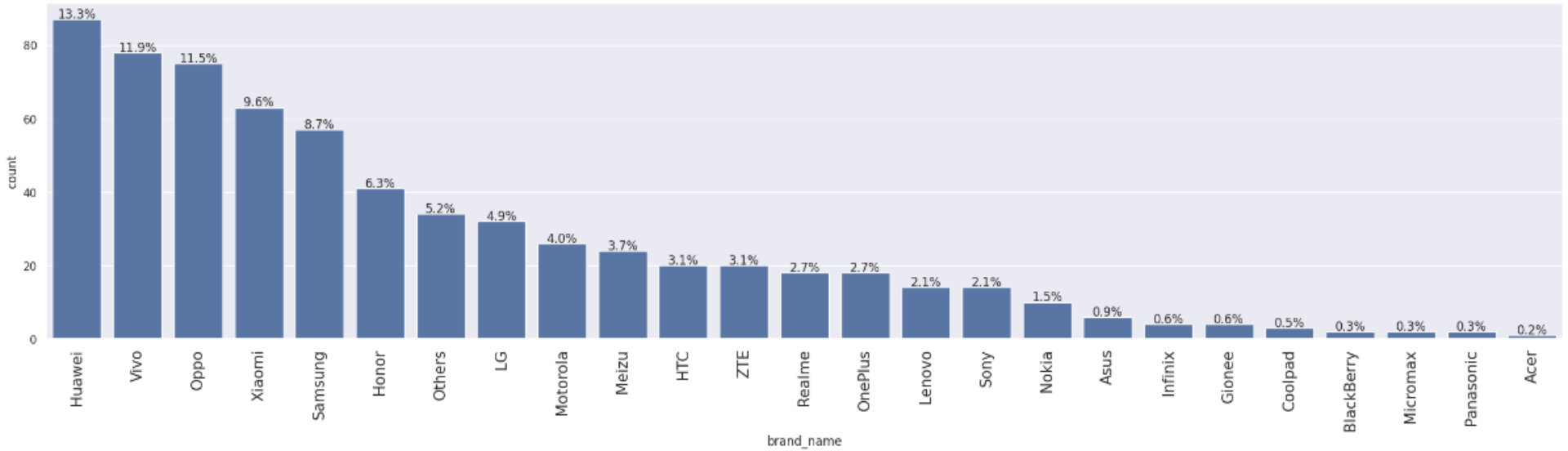
Link to Appendix slide on data background check

Analysis of Brand Distribution for High-Resolution Selfie Cameras

- Key Observations:
  - Leading Brands: Huawei (13.3%), Vivo (11.9%), and Oppo (11.5%) lead in offering devices with high-resolution selfie cameras, indicating a strong focus on front camera quality.
  - Competitive Offerings: Xiaomi (9.6%) and Samsung (8.7%) also demonstrate substantial presence, showcasing competitive selfie camera features.
  - Market Diversity: Lesser shares for brands like Sony, Nokia, and Asus may reflect varying market strategies or focus areas.
- Implications for ReCell:
  - Consumer Demand: Strong demand for high-quality selfie cameras, especially in markets dominated by social media and selfie culture.
  - Pricing Strategy: Brands known for superior selfie cameras may command premium prices, using camera quality as a competitive edge.
  - Inventory and Marketing Strategy: Stock devices from brands popular for their selfie cameras and highlight these features in marketing efforts to attract target customers.
  - Brand Positioning: Position leading brands prominently in marketing and retail spaces to appeal to consumers seeking advanced selfie capabilities.

Link to Appendix slide on data background check

Analysis of Brand Distribution for High-Resolution Selfie Cameras



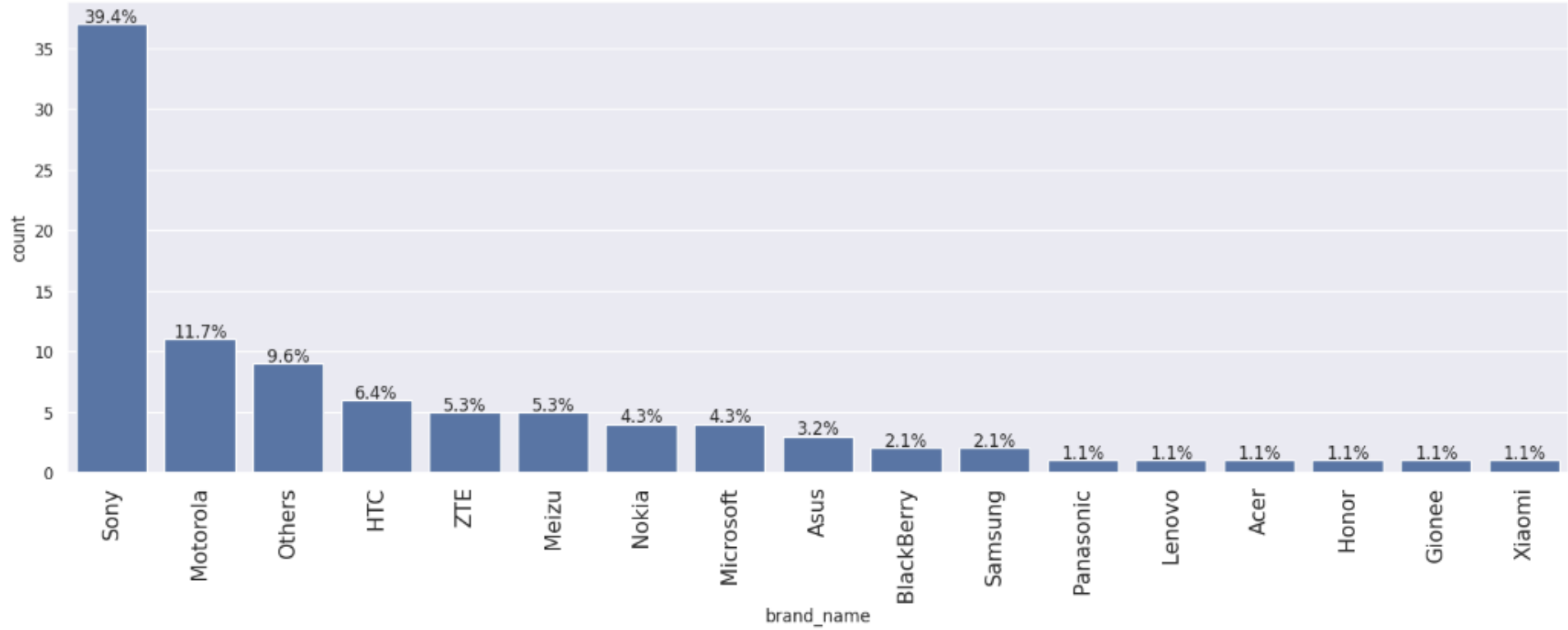
Link to Appendix slide on data background check

Analysis of Brand Distribution for High-Resolution Rear Cameras

- Key Observations:
  - Sony Dominance: Sony captures 39.4% of devices with high-resolution rear cameras, indicating a strong focus on premium camera quality.
  - Other Key Players: Motorola holds 11.7%, with other brands like HTC, ZTE, and Meizu showing smaller shares, reflecting diverse priorities in camera technology.
  - Market Diversity: Despite a wide range of brands, some like Samsung, Lenovo, and Xiaomi have smaller representations in this high-resolution segment.
- Implications for ReCell:
  - Consumer Demand: Sony's lead suggests a significant market for high-quality photography equipment, appealing particularly to photography enthusiasts.
  - Pricing Strategy: High-resolution camera devices, especially from Sony, may be priced higher due to their advanced camera capabilities and brand reputation.
  - Inventory Management: Prioritize stocking brands known for superior rear cameras to meet the expectations of targeted consumer segments.
  - Marketing Focus: Highlight the advanced camera features of these brands in marketing campaigns to attract consumers seeking top-tier photography performance.

Link to Appendix slide on data background check

Analysis of Brand Distribution for High-Resolution Rear Cameras



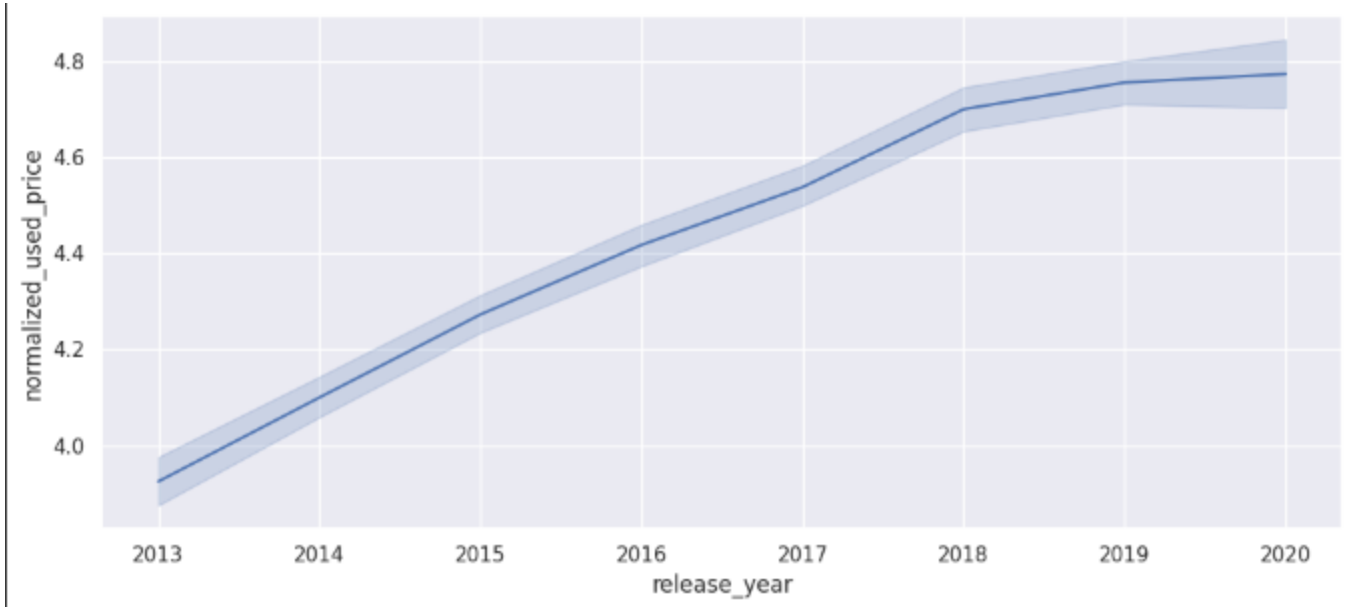
Link to Appendix slide on data background check

Analysis of Normalized Used Prices by Release Year

- Key Observations:
  - Increasing Trend: A clear upward trend in normalized used prices from 2013 to 2020, indicating that newer devices retain more value.
  - Steady Growth: The increase in used prices over the years suggests improvements in technology, increased original pricing, and greater desirability of newer models.
- Implications for ReCell:
  - Value Retention: Newer devices depreciate less quickly, reflecting their advanced features and sustained consumer interest.
  - Pricing Strategy: Implement a pricing model that reflects the higher retained value of newer devices, potentially commanding premium prices.
  - Inventory Focus: Prioritize acquiring and selling newer models, which are likely to have higher demand and faster turnover due to their perceived value.
  - Market Forecasting: Use insights from the pricing trend to enhance market forecasting and strategic planning, aligning inventory and marketing efforts with consumer preferences for newer technology.

Link to Appendix slide on data background check

Analysis of Normalized Used Prices by Release Year



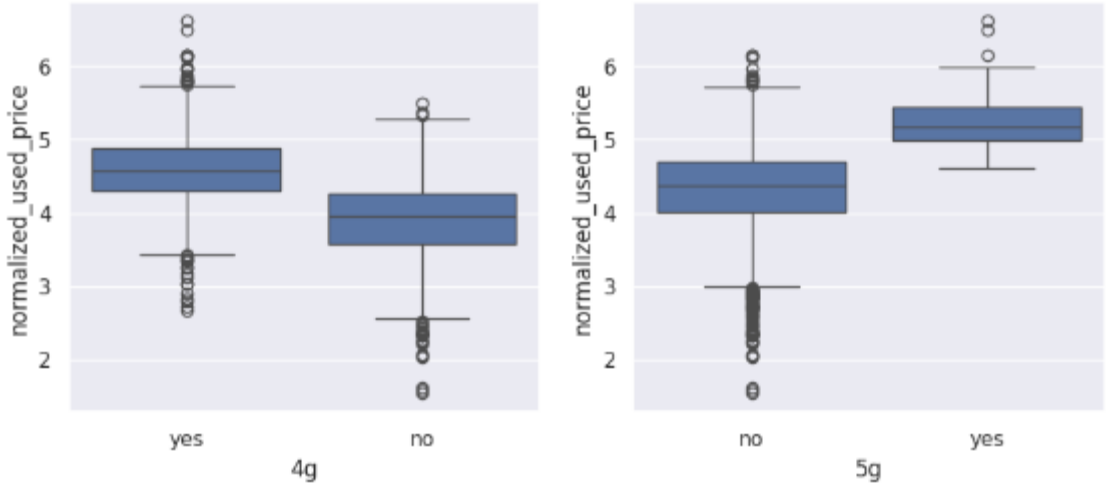
Link to Appendix slide on data background check

Analysis of Price Distribution by Connectivity (4G and 5G)

- Key Observations:
  - 4G Connectivity: Devices with 4G show higher median prices than those without, indicating its value in the resale market. The price variability among 4G devices reflects differences in brands, models, or additional features.
  - 5G Connectivity: Devices with 5G connectivity command significantly higher median prices and show less price variability, suggesting a premium for newer technology and limited model availability.
- Implications for ReCell:
  - Pricing Strategy: Incorporate 4G and 5G connectivity as key factors in pricing, recognizing their influence on resale value. Price 5G devices at a premium due to their advanced technology.
  - Inventory Management: Focus on stocking more 5G devices to leverage their higher resale value and anticipated demand. Maintain a robust selection of 4G devices to cater to a broader market.
  - Marketing and Consumer Targeting: Emphasize the benefits of 4G and 5G connectivity in marketing campaigns. Highlight 5G's advanced capabilities to attract tech-savvy consumers seeking the latest technology.

Link to Appendix slide on data background check

Analysis of Price Distribution by Connectivity (4G and 5G)



Link to Appendix slide on data background check

Missing Value Treatment

- Missing Values
  - Identified missing data in key features including camera resolution, internal memory, battery capacity, and device weight.
- Imputation Techniques Employed
  - Median Imputation: Used the median values for continuous variables like camera megapixels and battery capacity. The medians were calculated within groups segmented by device 'release_year' and 'brand_name' to reflect technological trends and brand-specific standards accurately.
- Rationale for Median Imputation
  - Chosen for its robustness against outliers and its ability to preserve the typical value within each subgroup, ensuring technological relevancy and brand specificity.
- Grouping Strategy
  - Grouped data by 'release_year' and 'brand_name' for imputation to maintain accuracy reflective of technological advancements and consistency within brand models.
- Handling Sparse Data
  - For attributes with persistent missing values after initial imputation, applied a general median fill to ensure completeness of the dataset.
- Validation and Impact
  - Conducted thorough checks to ensure zero missing values post-imputation. Assessed the impact on data distribution to verify that the imputation did not significantly alter the characteristics of the dataset.

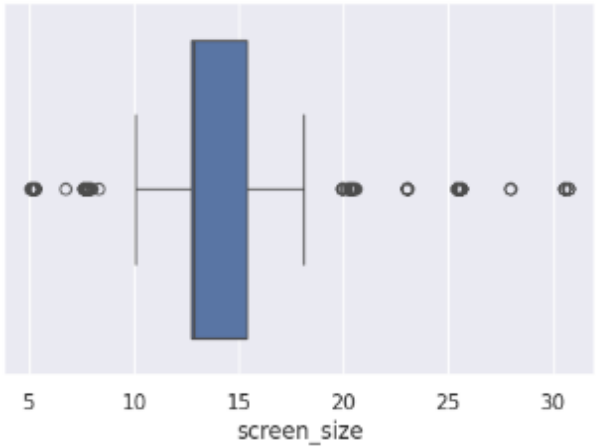
Outlier Check

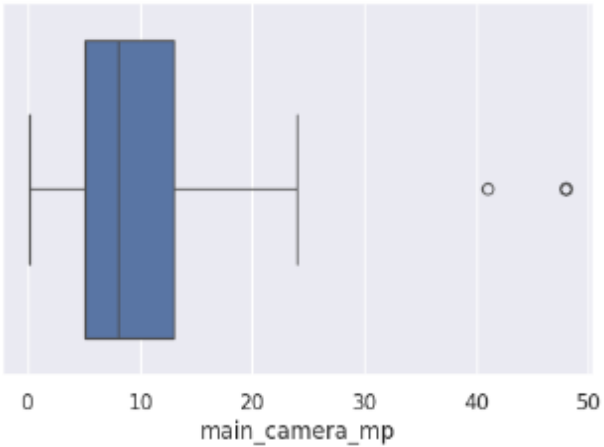
- Outlier Impact
  - Initial boxplot analysis revealed significant outliers in several attributes including days_used and normalized_new_price, critical for pricing strategy.
- Significance of Outliers
  - Outliers can drastically skew predictive model accuracy and misrepresent the data's central tendency, leading to flawed business decisions.
- Detection Techniques
  - Utilized boxplots to visually identify outliers in continuous data variables such as days_used and normalized_prices. This method helps pinpoint values that are substantially different from the rest of the data distribution.

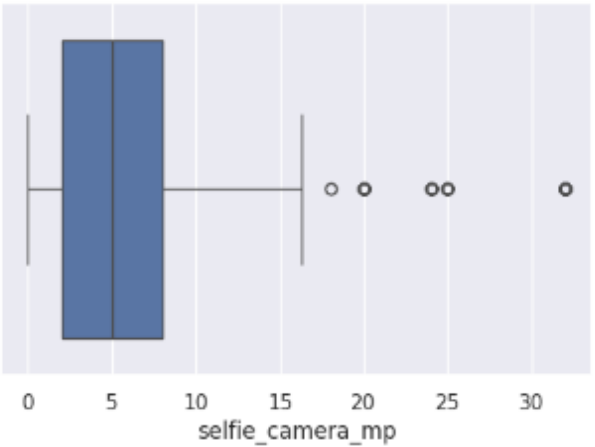
Outlier Check

- Examples of Outlier Impact
  - Days Used: Devices with unusually high days of usage might indicate exceptional durability or neglect in upgrading, impacting resale value.
  - Normalized New Price: Extreme values may reflect premium models or pricing errors, which could distort average price calculations.
- Treatment Methods
  - Adjusted outliers by capping them at specific percentiles (e.g., 1st and 99th) to minimize distortion in model predictions without losing valuable data.
  - In cases where outliers were deemed errors or non-representative, data points were either corrected or removed based on additional information or thresholds established from industry standards.
- Strategic Insights for ReCell
  - Understanding and adjusting for outliers in days_used and normalized_prices ensures more accurate pricing models, reflecting true market conditions and device desirability.
  - Provides a foundation for dynamic pricing strategies that reflect actual consumer perception and device worth.

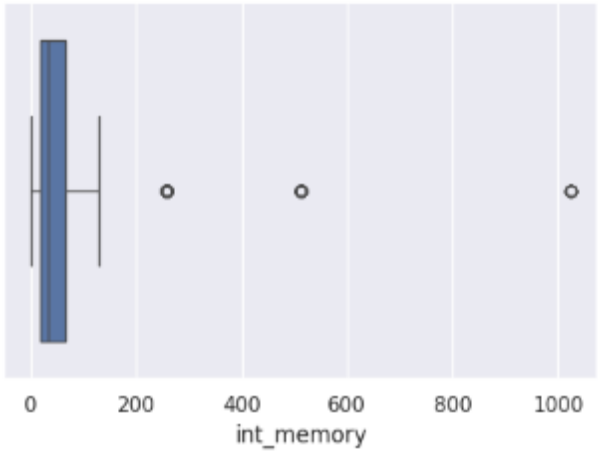
Outlier Check

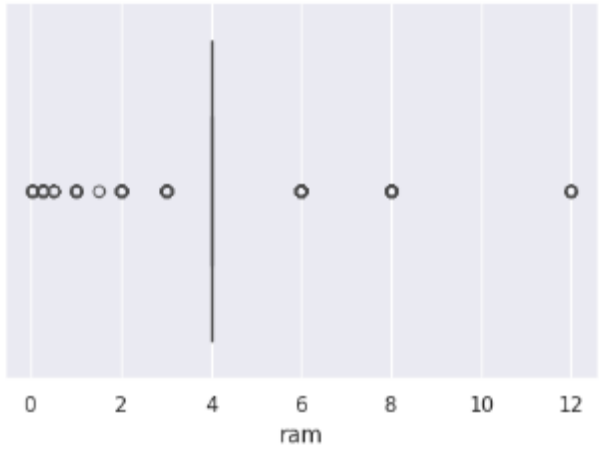


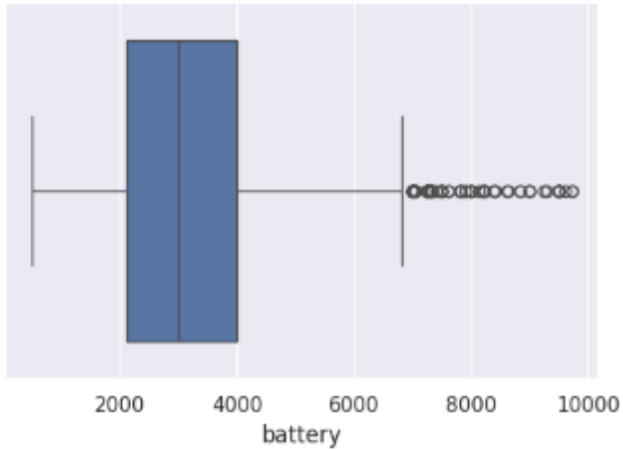




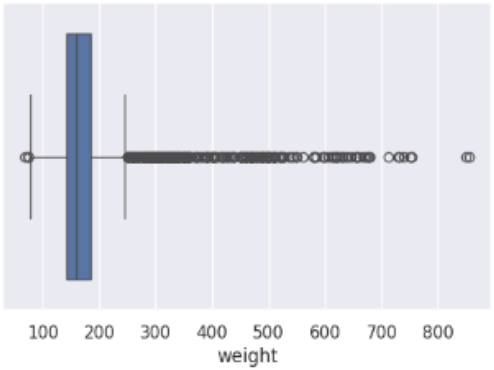
Outlier Check

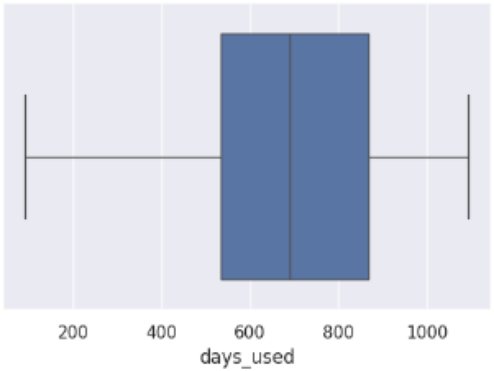


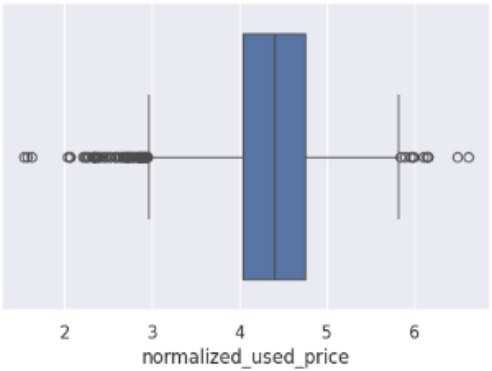


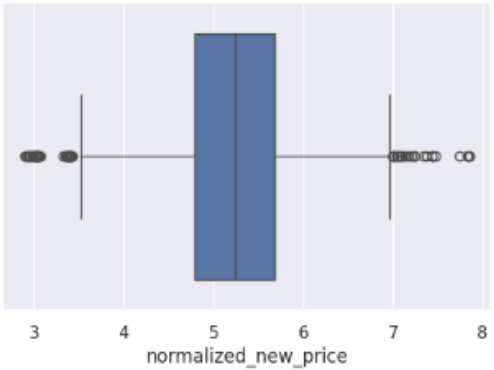


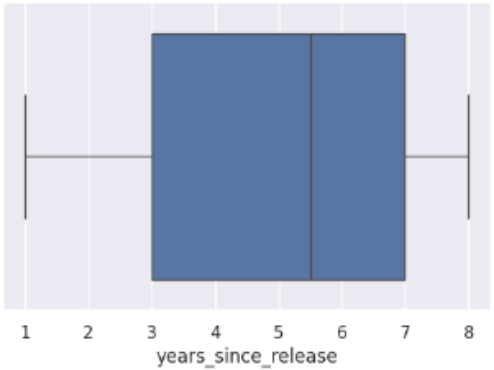
Outlier Check











Feature Engineering

- Newly Created Features
  - Years Since Release: Calculated as the difference between the year of data collection (2021) and the device release year. This feature helps capture the depreciation effect and technological outdating on pricing.
- Rationale for New Features
  - Older devices generally depreciate in value, and 'Years Since Release' quantifies this effect, providing a straightforward measure of age that can be directly related to pricing dynamics.
- Additional Feature Ideas
  - Interaction between Release Year and Battery: Considering creating a feature that combines the age of the device with its battery capacity. This could help capture how older devices with still robust battery capacities might retain more value.
  - Combination of RAM and Internal Memory: Proposing a feature that aggregates RAM and internal storage, which could better represent the device's performance capabilities, a key factor in pricing.
- Impact of Feature Engineering
  - These engineered features are expected to provide deeper insights into what drives device prices, potentially leading to more accurate predictions and enabling dynamic pricing strategies that reflect real market values.

Data Preperation for Modeling

- Encoding Categorical Variables
  - Categorical Features: Encoded categorical variables such as 'brand_name' and 'os' using dummy variables. This transformation is necessary to include these categorical attributes in the linear regression model.
- Scaling Numerical Inputs
  - Numerical Features: Though not detailed in your provided steps, typically, scaling numerical features to a similar scale can prevent attributes with larger ranges from dominating the model. Consider standardizing features like 'days_used' and 'battery' for consistency.
- Splitting the Dataset
  - Split the data into training (70%) and testing (30%) sets. This separation allows the model to be trained on a diverse subset of data and then validated against a set that it hasn't seen, ensuring that the model can generalize to new data effectively.

Model Performance Summary

- Model Overview:
  - Dependent Variable: Normalized used price.
  - R-squared: 0.847, indicating 84.7% of the variability in price is explained by predictors.
  - Adjusted R-squared: 0.846, confirming the appropriateness of the predictor count.
  - Significance: F-statistic confirms model significance.
  - Coefficients: Positive impacts from screen size and main camera megapixels; negative impacts from battery and weight.

Link to Appendix slide on model assumptions

Model Performance Summary

- Performance Metrics:
  - Training Data:
    - RMSE: 0.23114
    - MAE: 0.175033
    - R-squared: 0.847102
    - Adjusted R-squared: 0.840802
    - MAPE: 4.322087%
  - Test Data:
    - RMSE: 0.238225
    - MAE: 0.188703
    - R-squared: 0.833349
    - Adjusted R-squared: 0.830374
    - MAPE: 4.513082%

Link to Appendix slide on model assumptions

Model Performance Summary

- Interpretation and Implications:
  - Model Fit: Excellent predictive power with slight variation between training and test data, suggesting good generalization and minimal overfitting.
  - Multicollinearity: High condition number suggests possible multicollinearity, though performance metrics indicate stable prediction.
  - Decision-Making Tool: Reliable for predicting used prices, aiding ReCell's pricing strategy decisions.

Link to Appendix slide on model assumptions

APPENDIX

Data Background and Contents

Python Notebook for Data Analysis:

- Title: ReCell Used Device Pricing Analysis Notebook
- Description:
  - This Jupyter notebook encompasses the entire data analysis workflow tailored for ReCell's needs. It includes data preprocessing, exploratory data analysis (EDA), feature engineering, and model development steps.
  - The notebook also conducts extensive statistical analysis to determine the significant predictors affecting the prices of used phones and tablets.

Dataset Used in Analysis:

- Title: ReCell Used Device Pricing Data
  - File Name: used_device_data.csv
  - Description:
    - The dataset provides detailed attributes of used and refurbished phones and tablets collected in 2021.
    - Features include brand name, operating system (OS), screen size, 4G/5G capability, camera resolutions (main and selfie cameras), internal memory, RAM, battery capacity, weight, release year, days used, and normalized prices for new and used devices.
    - It aims to help ReCell develop a data-driven, dynamic pricing strategy by identifying factors that significantly influence the pricing of used devices.

Link to Appendix slide on data background check

Model Assumption

- Multicollinearity Check (VIF)
  - Definition: Variance Inflation Factor (VIF) tests for multicollinearity among predictors.
  - Thresholds:
    - VIF = 1: No correlation.
    - VIF > 5: Moderate multicollinearity.
    - VIF > 10: High multicollinearity.
  - Action Taken: Variables with VIF > 5 were iteratively removed, reassessing the model until all VIF values were acceptable.
- High p-value Variables
  - Strategy: Variables with p-values > 0.05 were considered for removal to enhance model significance.
  - Process:
    - Iteratively build the model, remove the variable with the highest p-value, and reassess.
    - Continue until all p-values are below 0.05.

Link to Appendix slide on data background check

Model Assumptions

- *Linearity and Independence*
  - *Method: Residual plots analyzed to check for linearity and independence of residuals.*
  - *Visual Aid: Example plot showing residuals vs. fitted values.*
    - *Ideal Outcome: No discernible pattern (indicates linearity and independence).*
- *Normality of Residuals*
  - *Tests Used:*
    - *Histogram of residuals.*
    - *Q-Q plot comparison with a normal distribution.*
  - *Shapiro-Wilk Test: Conducted to formally assess normality.*
    - *Criteria: p-value > 0.05 suggests normality.*
- *Homoscedasticity Test*
  - *Method: Goldfeld-Quandt test used to evaluate the constancy of variance in residuals.*
  - *Result Interpretation:*
    - *p-value > 0.05: Residuals are homoscedastic.*

Link to Appendix slide on data background check

Great Learning

Happy Learning !