**Analyzing Popularity Score in Pinterest Dog Design Dataset**

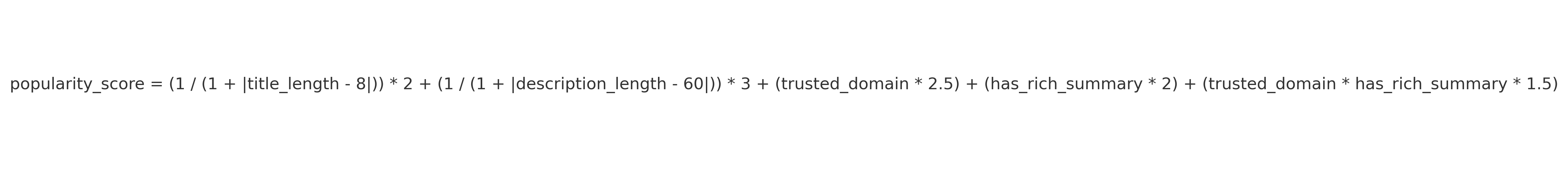
**Overview**

This report outlines insights derived from a dataset of over 20,000 Pinterest posts on dog designs. The primary objective was to determine the popularity of different keywords and styles by creating a custom **Popularity Score** metric.

**Popularity Score Formula**

The **Popularity Score** is a weighted composite metric designed to assess the appeal of posts based on title, description, domain credibility, and summary quality.

The calculation formula is as follows:



**Pinterest Word Analysis: Commercial Report**

**Objective:** To analyze word frequency and sentiment distribution from Pinterest data using a Bag of Words Model

A graph of a bar chart

Description automatically generated with medium confidenceA graph of a bar graph

Description automatically generated

**Findings:**

1. **Top Word Frequency:**
   * The most frequently occurring words included:
     + **"dog"**: Dominating with a frequency exceeding 8000.
     + **"perfect," "pet," "art," "design"**: Frequently observed with values between 3000-5000.
     + Other high-frequency terms: **"stickers," "cat," "size," "make," "gift."**
2. **Sentiment Distribution:**
   * Positive words dominated the frequency counts, especially for terms like:
     + **"dog," "perfect," "pet."**
   * Neutral and negative sentiments were less common but were present in lower frequency words like:
     + **"vintage," "hot," "sale."**

**Topic Modeling Summary**

**Objective**

The purpose of this analysis was to identify dominant themes and key terms within a dataset of textual descriptions using topic modeling techniques.

**Methodology**

1. **Text Vectorization**
   * Technique: TF-IDF (Term Frequency-Inverse Document Frequency) with a maximum of 1000 features.
   * Preprocessing: Removal of stop words and missing data from the descriptions.
2. **Topic Modeling**
   * Model: Latent Dirichlet Allocation (LDA) with 5 topics.
   * Parameters: Random state set to 42 for reproducibility.
3. **Topic Analysis**
   * Each topic is characterized by the top 10 terms contributing to it.
   * Results were visualized in a tabular format.

**Results: Key Topics and Terms**

A screenshot of a computer

Description automatically generated

**Insights**

1. **Recurring Themes**:
   * **Pets**: Dogs and cats feature prominently across multiple topics.
   * **Art and Design**: Creative elements such as art, stickers, and posters emerge as significant themes.
2. **E-commerce Focus**:
   * Terms like *sale*, *Etsy*, and *quality* indicate a connection to online retail or product promotion.
3. **Specialized Products**:
   * Topics highlight niche offerings such as decals, vinyl stickers, and handmade items like crochet patterns.

**Color and Pattern Preferences**

**1. Color Frequency and Popularity**

A graph of color frequency distribution

Description automatically generatedA graph of blue vertical bars with white text

Description automatically generated

* **Key Findings:**
  + The most frequent colors are **black** and **white**, each appearing over 1,000 times, far surpassing other colors.
  + **Blue** follows with approximately 500 occurrences, while **red** and **pink** are moderately frequent at around 400 instances.
  + Least frequent colors include **purple**, **grey**, and **gold**, each below 150 occurrences.
* **Popularity Insight:**
  + Despite its lower frequency, **yellow** ranks as the most popular color with the highest average popularity score.
  + **Silver** and **orange** also score highly in popularity, outperforming more frequent colors like black and white.

**2. Pattern Frequency and Popularity**

A graph with blue bars

Description automatically generatedA graph of blue bars

Description automatically generated with medium confidence

* **Key Findings:**
  + **Animal prints** dominate with over 1,200 occurrences, significantly ahead of all other patterns.
  + **Floral patterns** are the second most frequent at approximately 400 occurrences.
  + Patterns like **dotted**, **camouflage**, and **checkered** are the least common, with fewer than 50 occurrences each.
* **Popularity Insight:**
  + **Floral patterns** are the most popular, achieving the highest average popularity score.
  + Abstract and solid patterns rank moderately high, while animal prints, despite their frequency, have lower average popularity scores.

**Key Business Implications**

* **Color Recommendations:**
  + Focus marketing and production efforts on black, white, and blue due to their high frequency.
  + Incorporate yellow and silver in targeted collections, as they have strong popularity despite lower frequencies.
* **Pattern Recommendations:**
  + Capitalize on the demand for floral patterns by integrating them into designs.
  + Consider reducing reliance on animal prints despite their frequency, as their popularity is relatively lower.
  + Introduce abstract patterns to appeal to niche markets, leveraging their high popularity scores.

**Identifying Market Opportunities Using NLP Tools**

**Objective:**

Leverage NLP tools to extract and evaluate keywords with high demand and low competition to identify potential market opportunities.

**Methodology:**

**1. Supply Calculation**

* **Process**:
  + Compile all keywords from different sources and "flatten" the data to unify formats.
  + Use the Counter tool to count the occurrences of each keyword across all sources.
  + The result represents the "Supply" of each keyword, indicating how often it appears in the market.

**2. Demand Calculation**

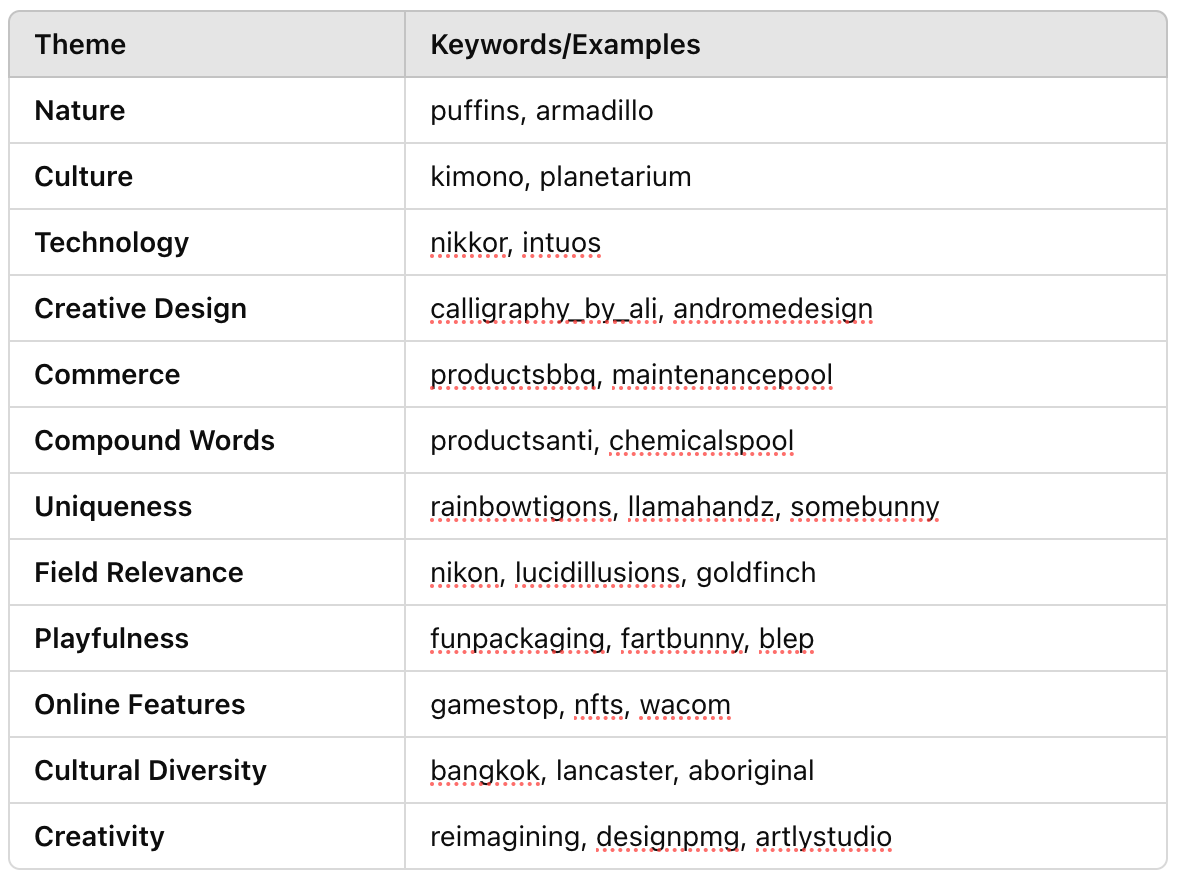
* **Process**:
  + Extract individual keywords and associate them with a "popularity score" sourced from data trends.
  + Calculate the average popularity score for each keyword.
  + The result represents the "Demand" for each keyword, indicating market interest and popularity.

**3. Supply-Demand Ratio Calculation**

* **Process**:
  + Divide the Demand by the Supply for each keyword to calculate the "Demand-Supply Ratio."
  + A higher ratio suggests that a keyword has significant market demand but limited competition, highlighting it as a potential market opportunity.

**Example Keywords and Themes:**

Based on the provided data, thematic keywords were categorized as follows:



**Insights and Recommendations:**

**1. High Potential Keywords:**

* Keywords with a high Demand-Supply Ratio are identified as underrepresented yet highly sought after.
* Example:
  + "llamahandz" and "fartbunny" demonstrate uniqueness and playfulness with niche appeal.
  + "kimono" and "planetarium" indicate cultural and experiential demand.

**2. Emerging Trends:**

* Keywords like "nfts" and "wacom" reflect technological relevance and online market dynamics.
* Creativity-focused terms such as "reimagining" suggest a strong pull towards innovative design services.

**3. Strategic Focus Areas:**

* Invest in niche categories (e.g., "Cultural Diversity" and "Uniqueness") where high Demand-Supply Ratios reveal unmet consumer needs.
* Develop targeted marketing campaigns for terms under "Playfulness" to capture interest among younger audiences.

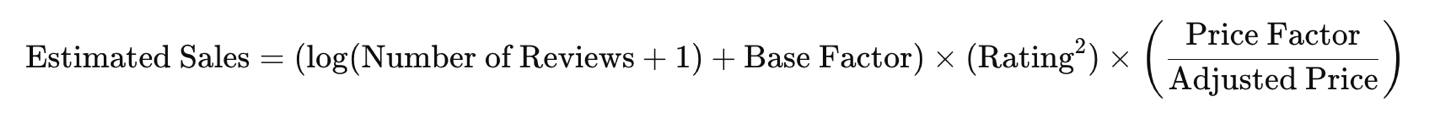
**Etsy Dog Design Product Analysis**

**Objective:**  
Analyze 2,000+ dog design products on Etsy to uncover trends in sales and popularity using advanced metrics and keyword analysis.

**Key Metrics Developed:**

1. **Estimated Sales**:
   * Incorporates a baseline traffic constant, customer reviews, ratings, and price adjustments.
   * Formula factors in price segmentation to account for low, medium, and high price ranges, adjusting the impact accordingly.

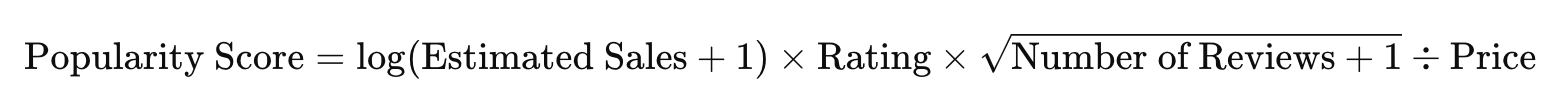
**Formula:**



1. **Popularity Score**:

* Highlights products that balance sales volume, high ratings, and pricing.
* Includes logarithmic transformations for non-linear scaling and adjusts for price sensitivity.

**Formula:**

****

**Methodology:**

* **Keyword Analysis**:  
  Used a Count Vectorizer to extract high-frequency words from product titles/descriptions, identifying correlations with Estimated Sales and Popularity Score.
* **Data Source**:  
  Scraped from Etsy using Octorparse.
* **Segmentation**:  
  Products categorized into price ranges for tailored insights:
  + Low (< $5)
  + Mid ($5-$15)
  + High (>$15)

**Insights from Visualization:**

A graph with text and numbers

Description automatically generated with medium confidence

1. **High-Performance Keywords**:
   * Words such as *"cute," "puppy," "journal,"* and *"planner"* correlate strongly with both high Estimated Sales and Popularity Score.
   * Products emphasizing *"unique" and "personalized"* designs perform better in terms of niche appeal.
2. **Pricing Insights**:
   * Mid-priced products ($5-$15) balance affordability and quality perception, leading to higher Estimated Sales.
   * Low-priced products often struggle with profitability despite high sales volume.
3. **Ratings and Reviews**:
   * Products with higher ratings (>4.5) and more reviews (>50) exponentially increase their Popularity Score.
4. **Categories to Watch**:
   * *"Dog planners," "cute decals,"* and *"dog lovers"* merchandise dominate in both popularity and sales.

**Sentiment Analysis by Category**

**Objective**

To evaluate sentiment trends across various categories using the NLTK Sentiment Intensity Analyzer.

A graph of different colored lines

Description automatically generated

**Key Metrics**

1. **Positive Sentiment**: Steady values with a slight increase toward the end of the categories, peaking in the "photo" and "using" categories.
2. **Neutral Sentiment**: Gradually decreasing trend across the categories, showing fewer neutral reactions as categories progress.
3. **Negative Sentiment**: Relatively consistent, with a noticeable rise in the final categories ("cars" and "die").
4. **Compound Score**: A clear downward trend, indicating diminishing overall sentiment balance through the categories.

**Insights**

1. **Most Positive Categories**:
   * Higher average positive sentiment in "cute," "lover," and "photo."
2. **Most Negative Categories**:
   * Sharp increase in negative sentiment observed in "cars" and "die."
3. **Neutrality Decline**: As the categories progress, neutral sentiment diminishes, reflecting stronger opinions or emotions in later categories.
4. **Compound Sentiment Decline**: Suggests an overall decrease in sentiment positivity, with "cars" and "die" impacting the trend significantly.

**Recommendations**

1. **Focus on Positive Sentiment**: Highlight and promote categories with positive sentiment, like "cute" and "photo."
2. **Mitigate Negative Sentiment**: Address and investigate the causes of high negativity in the "cars" and "die" categories.
3. **Neutral Sentiment Opportunities**: Leverage decreasing neutrality to create targeted marketing that polarizes positively rather than negatively.

**Analysis of Lowest Competition Stores**

**Objective:**

Identify 20 low-competition stores and discover market gaps using advanced analytical techniques.

**Tools and Methods:**

1. **NetworkX**:
   * **Purpose**: Constructed an undirected graph to model relationships between stores.
   * **Outcome**: Analyzed nodes (representing sellers) and edges (indicating similarity relationships) to determine competitive overlaps.
2. **scipy.spatial.distance.cosine**:
   * **Purpose**: Calculated cosine similarity to measure the similarity between sellers based on product names and characteristics.
   * **Outcome**: Identified stores with minimal similarity, highlighting potential low-competition niches.

**Key Findings:**

1. **Top 20 Low-Competition Stores**:
   * **Identified Sellers**:
     + Estatecleaner: Handmade and Customization, Vintage and Quirky
     + TheStickiShop: High Aesthetic Appeal, Emotion-Driven Products
     + WeirdandWonderful608: Niche and Thematic Products, Vintage and Quirky
     + TheStickerWhisperer: Handmade and Customization, High Aesthetic Appeal
     + FourDaisiesDesigns: Combination of Functionality and Decoration
     + PartyDecalShop: Handmade and Customization, High Aesthetic Appeal
     + RattanbagO: High Aesthetic Appeal, Diverse Material Selection
     + Anagarrandco: Functionality, Decoration, and Diverse Materials
     + Biterswit: High Aesthetic Appeal and Design Sense
     + StickOutLoudDesign: Handmade and Customization, High Aesthetic Appeal
     + vinyldecalsbykristel: Functionality, Decoration, and High Aesthetic Appeal
     + BarkAndMochaCo: Functionality, Decoration, Emotion-Driven Products
     + KMSVintageEclectible: Niche and Thematic Products, Vintage and Quirky
     + VerlenWORLD: Niche and Thematic Products, High Aesthetic Appeal
     + VividPineapple: High Aesthetic Appeal and Design Sense
     + CanisNerdus: Niche and Thematic Products, Emotion-Driven Products
     + StickyFrogDesigns: Niche and Thematic Products, Cartoon or Anime Themes
     + PocketPufff: Niche and Thematic Products, High Aesthetic Appeal
     + ChibiCharmCollective: Niche and Thematic Products, Cartoon or Anime Themes
     + RadRandomStickers: Handmade and Customization, High Aesthetic Appeal
2. **Unique Market Niches**:
   * Handmade and Customization
   * Emotion-Driven Products
   * Functionality and Decoration

**Comparison of Top 500 and Bottom 500 Images by Popularity**

**Objective:**

Analyze and compare characteristics of images sorted by popularity score (top 500 vs. bottom 500).

A screenshot of a group of text

Description automatically generated

**Insights:**

1. **Image Quality:** High clarity drives engagement and popularity.
2. **Emotional Impact:** Strong emotional resonance is a key factor for high popularity.
3. **Professionalism:** Professionalism enhances image appeal significantly.
4. **Composition & Aesthetics:** Strong visual design (composition, color) correlates with high popularity.

**Keyword Extraction and Ranking**

**Objective**  
To identify and prioritize key terms from the dataset based on their importance using the mean TF-IDF scores, and present the top 50 unique keywords for further analysis.

* Use diverse keywords to cover different aspects of user search intent (e.g., broad matches and long-tail keywords)
* Provide a rich variety of content keywords to meet diverse user search needs (e.g., "dog sticker" and "custom pet").

**Key Insights**

**Top 50 Keywords (Ranked by Importance):**

1. **dog**
2. **sticker**
3. **stickers**
4. **pet**
5. **custom**
6. **decal**
7. **vinyl**
8. **dog sticker**
9. **water**
10. **laptop**
11. **gift**
12. **sticker dog**
13. **dog stickers**
14. **custom pet**
15. **waterproof**
16. **cute**
17. **lover**
18. **bottle**
19. **water bottle**
20. **decals**
21. **mom**
22. **dogs**
23. **pet stickers**
24. **cat**
25. **dog mom**
26. **paw**
27. **dog lover**
28. **personalized**
29. **custom dog**
30. **planner**
31. **vinyl sticker**
32. **car**
33. **sticker custom**
34. **stickers pet**
35. **stickers dog**
36. **pet sticker**
37. **puppy**
38. **funny**
39. **print**
40. **lovers**
41. **paw print**
42. **animal**
43. **pack**
44. **sticker water**
45. **bottle sticker**
46. **cat sticker**
47. **photo**
48. **custom cat**
49. **decal dog**
50. **dog decals**

**Recommendations**

* **Marketing Focus**: Highlight terms with the highest relevance (e.g., "dog," "sticker," "pet") in product descriptions, SEO strategies, and advertising campaigns.
* **Product Development**: Consider integrating top themes (e.g., "custom," "vinyl," "waterproof") into product offerings to align with consumer interests.
* **Content Strategy**: Use high-ranking keywords to optimize website content and social media posts for search visibility.

**Impact of Keywords on Product Popularity**

**Objective**:  
Evaluate whether using high-traffic keywords improves product popularity using bootstrapping analysis.

**Keywords Analyzed**:  
50 high-traffic keywords were selected, including terms like "cute," "adorable," "digital," "design," and "decor." These are relevant for products marketed to creative, digital, and home audiences.

**Methodology**:  
Bootstrapping was applied to compare the mean popularity scores of products:

1. **With Keywords**
2. **Without Keywords**  
   Confidence intervals were used to measure reliability.

**Key Results:**

A graph of a chart

Description automatically generated with medium confidence

* **Products with Keywords**:
  + **Higher Mean Popularity Score**: Consistently outperformed products without keywords.
  + **Tighter Confidence Interval**: Indicates greater reliability in popularity improvement.
* **Products Without Keywords**:
  + **Lower Mean Popularity Score**: Displayed weaker performance in comparison.
  + **Wider Confidence Interval**: Reflects less consistent results due to smaller sample size or variability.

**Optimizing Revenue and Sales for Digital Product Designs**

**Objective**

To analyze the relationship between price, popularity, and sales to identify optimal pricing strategies for digital product designs.

**Key Findings**

A graph with a red line

Description automatically generated

**1. Sales Regression Analysis**

* **Base Sales**: Products priced at the base level with a zero popularity score are estimated to sell 879 units on average.
* **Price Effect**:
  + For every $1 increase in price, estimated sales decrease by **26.82 units** (p < 0.001).
  + Indicates a strong negative relationship between price and sales.
* **Popularity Score Effect**:
  + For every 1-point increase in popularity score, estimated sales increase by **2.15 units** (p < 0.001).
  + Highlights the positive influence of product popularity on sales.

**2. Revenue Optimization Insights**

* Products with **higher popularity scores** can sustain **higher prices** to maximize revenue.
* Products with **lower popularity scores** should adopt **competitive pricing** to optimize sales and revenue.

**3. Model Performance**

* The regression model explains **37.5% of the variance** in sales (R-squared = 0.375), indicating a moderately strong fit.

**Actionable Recommendations**

**For Pricing Strategy:**

* **High Popularity Products**:
  + Set prices closer to the identified **optimal price** (e.g., $20 for popularity 100 or 500) to maximize revenue.
* **Low Popularity Products**:
  + Lower prices (e.g., $16-$18) to remain competitive and boost sales.

**For Product Development:**

* Focus on improving **popularity metrics** (e.g., engaging product descriptions and attractive visuals) as even small improvements significantly increase sales.

**Keyword-Based Product Popularity Prediction**

A graph with orange and white text

Description automatically generated

1. **Feature Extraction**:
   * Product titles were converted into numerical features using CountVectorizer for analysis.
2. **Model Training**:
   * A **RandomForestRegressor** was trained to predict product popularity scores based on title features.
3. **Insights**:
   * Titles with specific descriptive keywords, such as "Playful Dog Drawing" and "Cartoon Dog Printable File," achieved the **highest predicted scores**, indicating a strong correlation between title specificity and popularity.

**Simulated Strategy Evaluation with Reinforcement Learning**

**Objective**  
To simulate long-term impacts of various strategies (pricing, keywords, advertising) using reinforcement learning, generating data with seasonal and trending variations for analysis.

**Key Simulation Insights**

**1. Simulation Overview**

* **Duration:** Simulates daily data for 2 years (730 days).
* **Variables:** Includes region, product, and retailer characteristics such as inventory levels, price sensitivity, and market share.

**2. Dynamic Influences**

* **Supply Chain Disruptions:** Occur on 5% of days, reducing inventory by 50%.
* **Price Wars:** Market share shifts dynamically based on pricing differences, with a sensitivity factor of 0.2.

**3. Sales Modeling**

* **Growth Trend:** Long-term sales growth follows a logistic growth model.
* **Seasonality:**
  + Annual fluctuations modeled with a sine function.
  + Weekly cyclical variations added for precision.
* **Random Fluctuations:** Gaussian noise simulates unpredictable daily changes.

**4. Marketing Synergies**

* On 10% of days, marketing campaigns generate region- and retailer-specific sales boosts.

**5. Constraints**

* **Inventory Limits:** Sales are capped by inventory availability and supply chain disruptions, ensuring realistic outputs.

**Insights from Chart Analysis**

**Sales Fluctuations**

* **Characteristics:** Significant sales fluctuations influenced by pricing, market share, and marketing efforts.
* **Interpretation:** Unusual peaks or troughs may be linked to price changes, marketing campaigns, or competitive pressures.

**Price Stability**

* **Characteristics:** Prices remain relatively stable with minimal adjustments.
* **Interpretation:** Reflects a conservative pricing approach, potentially avoiding frequent disruptions or customer dissatisfaction.

**Market Share Trends**

* **Characteristics:** Generally stable with minor variations.
* **Interpretation:** Indicates a steady competitive position. Minor declines could signal potential threats or market erosion.

**Marketing Effect Variations**

* **Characteristics:** Marketing impact fluctuates across the simulation period.
* **Interpretation:** Variations align with differing promotional efforts, directly impacting sales in high-marketing periods.

**Combined Insights**

1. **Correlation Between Variables:**
   * Sales are highly correlated with fluctuations in marketing effectiveness. Stronger marketing efforts align with higher sales.
2. **Pricing Impact:**
   * Stable pricing minimizes direct effects on sales, making marketing a more significant driver of performance.
3. **Market Positioning:**
   * Stability in market share underscores the importance of consistent strategies to maintain competitive strength.