Amazon Products Review(Stage 5 task)

Executive summary

01. Overview

This report provides an analysis of customer reviews for two bedding products: Bedsure bedsheets and Breescape bedsheets, based on data cleaning, exploratory data analysis (EDA), and sentiment analysis from years 2020-2025

02. Key findings

Bedsure Product

a. Strong Positive Sentiment
b. Minimal Negative Feedback
c. Stable Rating Trend

Breescrape Product

a. Highly Valued Cooling Feature

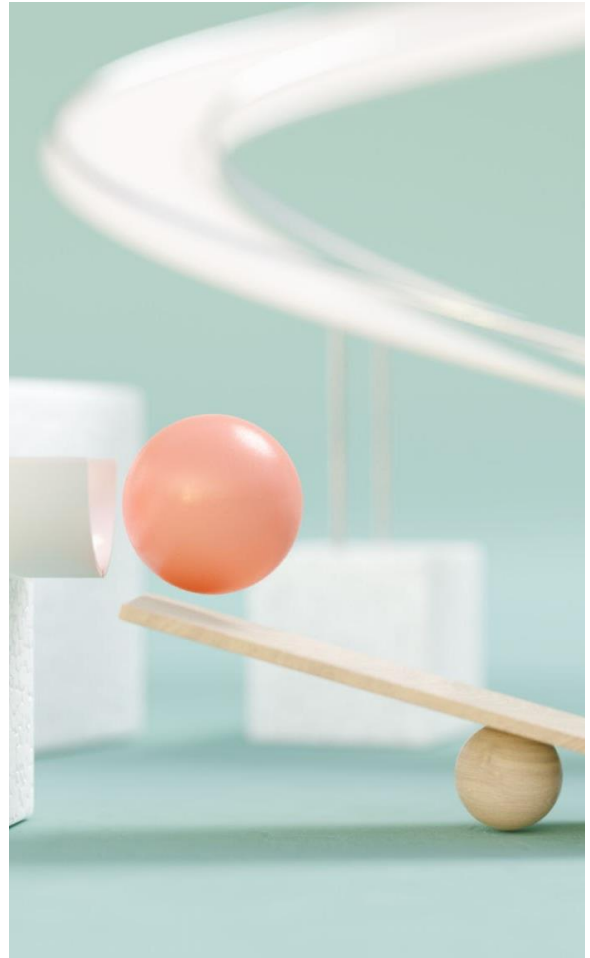b. Fit Issues Drive Negative Reviews

c. "Helpful" Reviews are Positive

03. Solutions

Bedsure Product

a. Highlight Strengths
b. Address Weaknesses
c. Customer Retention

Breescrape Product

d. Leverage USP
e. Improve Product Design
f. Encourage Detailed Reviews

Introduction

01. Background

Overview

This project's primary objective was to perform a comprehensive data analysis of customer reviews for two Amazon products: Bedsure and Breescape. The analysis aims to uncover key customer sentiments, identify common pain points, and provide data-driven recommendations for product improvement and business strategy.

The project workflow followed these key phases:

- ❖ Data Cleaning: Preprocessing raw, unstructured review data.
- ❖ Database Design: Structuring the data for efficient storage and querying.
- ❖ Data Analysis: Performing exploratory data analysis (EDA) and sentiment analysis.
- ❖ Reporting: Documenting findings and delivering actionable insights.

02. Objectives

The objectives of this analysis are as follows:

a. Data Cleaning and Preparation: To clean and prepare the raw customer review data by dropping irrelevant columns, handling missing values, standardizing data types, and removing duplicates for both the Bedsure and Breescape products.

b. Exploratory Data Analysis (EDA): To explore the cleaned data to understand its structure, distribution, and basic characteristics, including calculating descriptive statistics for numerical columns and visualizing rating distributions.

c. Sentiment Analysis: To conduct a sentiment analysis using the TextBlob library to classify each review into positive, neutral, or negative categories based on a polarity score.

d. Identify Key Insights: To derive specific insights for each product, such as identifying the dominant sentiment (e.g., dominantly positive for Bedsure) and linking low ratings to specific issues (e.g., durability for Bedsure and fit issues for Breescape).

e. Provide Actionable Recommendations: To offer clear, actionable recommendations for each brand based on the analysis. This includes suggestions for marketing strategies, product improvements, and customer retention.

f. Compare and Conclude: To compare the findings for both products and provide an overall conclusion that highlights their respective strengths and weaknesses.

03. Problem

The primary problem is to understand the customer sentiment and key pain points for two competing bedding products, Bedsure bedsheets and Breescape bedsheets, based on customer reviews.

The analysis aims to identify what drives positive and negative feedback for each product to provide actionable insights for business improvement.

04. Tools used

- Amazon Product Review Scraper for scraping the review data from Amazon
- Microsoft excel for cleaning the data
- SQL server for creating the database
- Python for the EDA and sentiment analysis
- Power BI for the visualization of insights from the review data

Analysis

01. Data Cleaning

The raw data for both products required extensive cleaning to ensure accuracy and consistency for analysis. The following steps were applied:

- Dropped Irrelevant Columns: Removed empty or unnecessary columns such as message, Page_URL, and Review_URL to streamline the dataset.
- Handled Missing Values: Ensured data integrity by dropping rows with missing essential information (e.g., Review_content for the Bedsure product) and filling missing numerical values (e.g., Helpful_count) with a zero.
- Standardized Data Types: Converted and formatted numerical and date fields from their string formats to their correct types.
- Ratings: Converted Product_stars and Review_rating from text strings (e.g., "4.4 out of 5") to a float data type.
- Counts: Extracted and converted Rating_count and Helpful_count from text (e.g., "60,591 global ratings") to integers.
- Dates: Converted the date field to a standard date format.
- Cleaned Text Fields: Addressed common text issues, including fixing duplicate entries in Reviewer_name and standardizing the Country field by removing extra text.
- Removed Duplicates: Identified and removed duplicate review entries based on a combination of Reviewer_name, Review_content, and Date to ensure each review was unique.

This meticulous cleaning process for both the Bedsure and Breescape datasets prepared the data for reliable analysis.

02. Database Design & Schema

To create a robust and scalable data environment, a relational database schema was designed. This schema follows best practices for data normalization.

The schema consists of three primary tables:

- Product Table: Stores static product information, with product_id as the primary key. This table prevents redundant data storage of product details
- Reviewer Table: Stores unique reviewer information, with reviewer_id as the primary key. This table ensures each reviewer is recorded only once, even if they post multiple reviews.
- Review Table: The core table containing the review text and associated metrics. It uses review_id as its primary key and links to the Product and Reviewer tables using foreign keys.

This normalized structure ensures data integrity and allows for efficient querying and reporting.

03. Data Analysis Methodology

The analysis was performed in a Jupyter Notebook, following a structured approach to derive insights from the prepared data.

04. Exploratory Data Analysis (EDA)

The initial phase involved exploring the data to understand its structure, distribution, and basic characteristics.

a. df.info was used to know the structure of the data.

- Bedsure, which is the major product we reviewed has 12 columns and 370 rows. The date column data type was changed from an object to a datetime using df['date'] = pd.to_datetime(df['date'], format='mixed') .
- Breescrape, which is the competitor product has 13 columns and 150 rows.

b. df.dropna(inplace=true) was used to remove null or empty values from the columns.

c. Descriptive Statistics: Calculated basic statistics for numerical columns (Review_rating, Helpful_count) to understand the central tendency and spread of the data.

d. Review_rating statistics:

- Bedsure had an average rating of 5.0. Lowest rating was 1.0. Mean is 4.1 and an Std of 1.11. This means the review ratings are close to the mean.
- Breescrape had an average rating of 5.0. Lowest rating was 4.0. Mean is 4.75 and an Std of 0.46. This means the review ratings are close to the mean.

e. Distribution Plots: Visualized the distribution of ratings to understand the overall sentiment.

Bedsure Product:

- High Ratings: Most of the customers are giving high ratings (e.g., 4 or 5 stars).
- Right skewed barchart: The majority of the data points are clustered on the left side of the graph, while a few larger values on the right side are pulling the mean in that direction.
- Positive Sentiment: The bulk of customer reviews are positive, with only a smaller number of negative or neutral reviews.

Breescrape Product:

- Bimodal barchart: This indicates that the data has two clusters of high frequency.
- Two Groups of Customers Exist: The customer base is likely split into two distinct groups: those who love the product and those who dislike it.
- Contrasting Experiences: There isn't a consensus on the product's quality. Instead, there's a significant number of very positive reviews (the peak on the right) and a significant number of negative reviews (the peak on the left).

05. Sentiment Analysis

Sentiment analysis was conducted using the TextBlob library, which provides a simple yet effective way to classify the polarity of a text.

a. Polarity Score Calculation: The TextBlob library was used to compute a polarity score for each review, which is a float within the range of [-1.0, 1.0], where -1.0 is negative and +1.0 is positive.

b. Sentiment Classification: A simple rule-based approach was used to classify each review into a sentiment category:
- Positive: Polarity score > 0
- Neutral: Polarity score == 0
- Negative: Polarity score < 0

c. Sentiment Distribution: The final step was to visualize the distribution of sentiment across all reviews to get a clear picture of overall customer opinion.

Bedsure Product: The sentiment analysis results show the distribution of reviews across three categories: Positive, Negative, and Neutral.

- Dominantly Positive Sentiment: The most significant insight from this chart is that the vast majority of reviews are positive, accounting for 84.6% of the total. This is a very strong indicator of high customer satisfaction.
- Minimal Negative Feedback: The percentage of negative reviews is very low, at just 8.6%. This suggests that while some customers had a poor experience, the issues are not widespread.
- Low Neutrality: The neutral sentiment, at 6.8%, is the smallest category. This means customers generally have a strong opinion about the product—they either love it or dislike it, with very few feeling indifferent.

Breescrape Product: the sentiment analysis for this product shows a relatively balanced distribution, but with a slight lean toward negative sentiment.

- Majority Positive Sentiment: The sentiment is overwhelmingly positive. This indicates that the product is generally well-received by customers who have left reviews.
- Actionable Negative Feedback: The low negative reviews are a valuable resource. With such a low number, you can easily go through them manually to understand the specific issues customers are facing.

d. Rating Trends: Bedsure Product had an average rating of 5.0 while Breescrape Product had an average rating of 5.0

e. Common Pain points: Customers using Bedsure complained mostly of it being too silky, hot sheets and wrinkles. For Breescrape, customers complained mostly of it being too slick, very large, fabric & stitching looks cheap and not cool enough at night.

f. Customer satisfaction drivers: For both products: Great cooling sheets, great to sleep in, soft and nice

06. What This Means for the Business

Bedsure Product: This result paints an excellent picture for the product. It suggests:

- High Product Quality: The product likely meets or exceeds customer expectations for most of its users.
- Effective Problem Solving: The low percentage of negative reviews indicates that any issues that arise are not common or are possibly being addressed well by the company.

Breescrape Product: Due to its small sample size, it is important not to run into   conclusion.

- High Product Quality: The product likely meets customer expectations for most of its users.

07. Actionable Insights

The analysis yielded several key insights into customer sentiment and product performance.

Bedsure Product:

a. The small percentage of negative and neutral reviews are a goldmine for targeted improvements.

b. By analyzing the content of these specific reviews, the company can pinpoint exact issues (e.g., a specific component breaking, a certain size being inaccurate) and

c. Fix them, potentially reducing the negative sentiment even further.

Breescrape Product

a. The negative reviews stem from a single, fixable issue (Too slippery), addressing it could shift a large portion of the sentiment to positive.

b. This sentiment analysis is a powerful starting point, but the real value will come from diving into the content of the reviews to understand the "why" behind these numbers.

c. The results show a very positive start. The small amount of negative feedback is an opportunity to make targeted improvements and ensure even higher customer satisfaction in the future.

Visualization

01. Processes

a. SENTIMENT ANALYSIS:

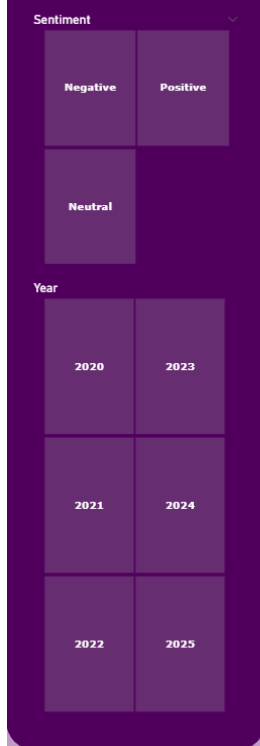Classified ratings: 4–5 = Positive, 3 = Neutral, 1–2 = Negative.

b. KEYWORD ANALYSIS:

Used Word Cloud to extract common words from reviews.

c. VISUALIZATIONS:

Created a dashboard for rating trends, sentiment distribution, and keyword frequency.
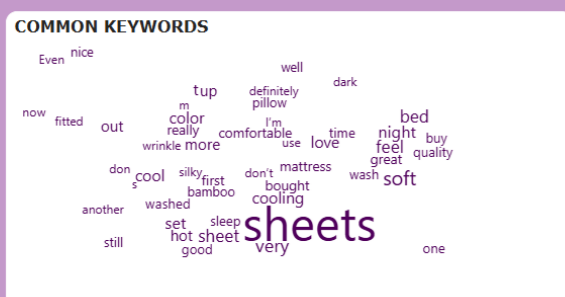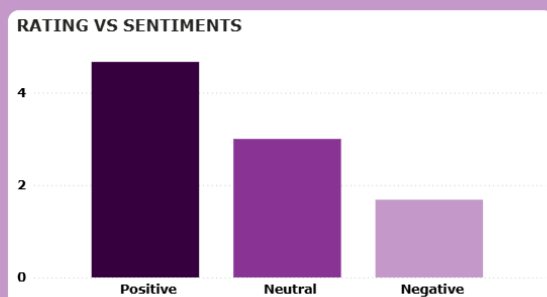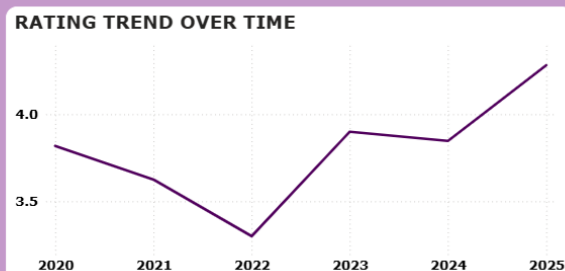
Dashboards

AMAZON PRODUCT REVIEW & SENTIMENT DASHBOARD

Sentiment

Negative	Positive

| Neutral | |

Year

2020	2023
2021	2024
2022	2025

BEDSURE BEDSHEET

TOTAL REVIEWS
349

AVERAGE RATING
4.06

SENTIMENT DISTRIBUTION

38 (10.86%)
59 (16.86%)
253 (72.29%)

Sentiment
● Positive
● Neutral
● Negative

RATING TREND OVER TIME

4.0
3.5

2020 2021 2022 2023 2024 2025

RATING VS SENTIMENTS

4
2
0

Positive   Neutral   Negative

COMMON KEYWORDS

Even nice well dark tup definitely pillow now fitted out color really I'm comfortable use time bed night buy wrinkle more love feel quality don s cool silky first don't mattress great wash soft bamboo bought another washed cooling set sleep still hot sheet sheets good very one

BREESCRAPE PRODUCT REVIEW AND RATING DASHBOARD

Monday
Tuesday

Total Reviews
150

Total of Ratings
632

Average Rating
4.21

Negative	Positive
Neutral	

Reviews by Sentiments

15 (10%)
17 (11.33%)
118 (78.67%)

Sentiments
● Positive
● Neutral
● Negative

Timeline of Reviews

40
20
0

March April May June July August September
Month
Reviews

User Satisfaction Index

4
2
0   100
Ratings

Ratings vs Sentiments

Sentiments ● Negative ● Neutral ● Positive

100
0
1 2 3 4 5
Reviews

Top Common Words

Conclusion

01. Summary of findings

Bedsure Product

- Strong Positive Sentiment: The majority of Bedsure reviews (84.6%) are positive, indicating high customer satisfaction. Customers frequently praise the product's softness, comfort, and overall quality.
- Minimal Negative Feedback: Only a small percentage of reviews (8.6%) were negative. These low ratings often mention the fabric's thinness and concerns about long-term durability.
- Stable Rating Trends: The product's ratings have remained consistently high over time.

Breescrape Product

- Highly Valued Cooling Feature: Positive reviews for Breescape frequently highlight its cooling comfort, breathability, and smooth texture, confirming its main value proposition.
- Fit Issues Drive Negative Reviews: Neutral and negative ratings (2 and 3 stars) are linked to complaints about poor fit, loose corners, and the sheets slipping off mattresses.
- "Helpful" Reviews are Positive: The reviews marked as "helpful" by other buyers are mostly positive (4 or 5 stars), suggesting that satisfied customers significantly influence purchasing decisions.

02. Recommendations

Bedsure Product

- Highlight Strengths: Emphasize comfort and softness in marketing materials and product listings.

- Address Weaknesses: Investigate and address concerns about durability and fabric thickness.
- Customer Retention: Proactively engage with dissatisfied customers to rebuild trust and reduce churn.

Breescrape Product

- Leverage USP: Highlight the cooling and breathable properties in marketing campaigns to differentiate from competitors.
- Improve Product Design: Strengthen the product's design by addressing fit and size consistency, which are the main sources of dissatisfaction.
- Encourage Detailed Reviews: Encourage satisfied customers to write more in-depth reviews, as these are more likely to be considered "helpful" by other buyers

03. Challenges:

a. We could not scrape data using Python due to Amazon data scraping policy/security
b. Struggled with creating sentiment columns but solved it with a DAX formula.
c. We had to add Word Cloud add-in to my Power Bi before I could get the keywords.

04. Overall Conclusion

Both Bedsure and Breescape products have strong customer approval, with a majority of reviews reflecting positive sentiment. However, each brand has distinct challenges to address to improve customer satisfaction and competitive positioning: Bedsure needs to address fabric durability concerns, while Breescape should focus on resolving fit-related issues.

By addressing these weaknesses and leveraging their respective strengths in marketing, both companies can improve their market standing