Amazon Melatonin Products for adults

DAMG-6105
Data Science with Python
Prof. Handan

Team 02

Sunny Wable
Shushil Girish

Topics covered in presentation

01 - Introduction

02 - Data preprocessing

03 - Findings from EDA

04 - N-grams

05 - Word clouds

06 - List of Interesting Insights

07 - Individual contributions

Introduction

Our dataset contains detailed reviews of melatonin supplements for adults from Amazon, categorized into different product dosages, offering following insights:

- **Product Info**: Includes ASIN, brand, product title, and product links.
- **Review Details**: Captures user reviews, scores, titles, and dates for sentiment and efficacy analysis.
- **User Engagement**: Metrics like helpful counts and verification status for credibility insights.
- **Consumer Insights**: Highlights preferences, effectiveness, flavors, and forms (e.g., tablets and gummies)

Features in Dataset

Following were some attributes in the dataset

- ASIN
- Brand
- HelpfulCounts
- PageUrl
- ProductLink
- ProductTitle
- ReviewContent
- ReviewDate
- ReviewScore
- ReviewTitle
- Reviewer
- Verified

Glimpse on data processing

Extraction of dosage from product title

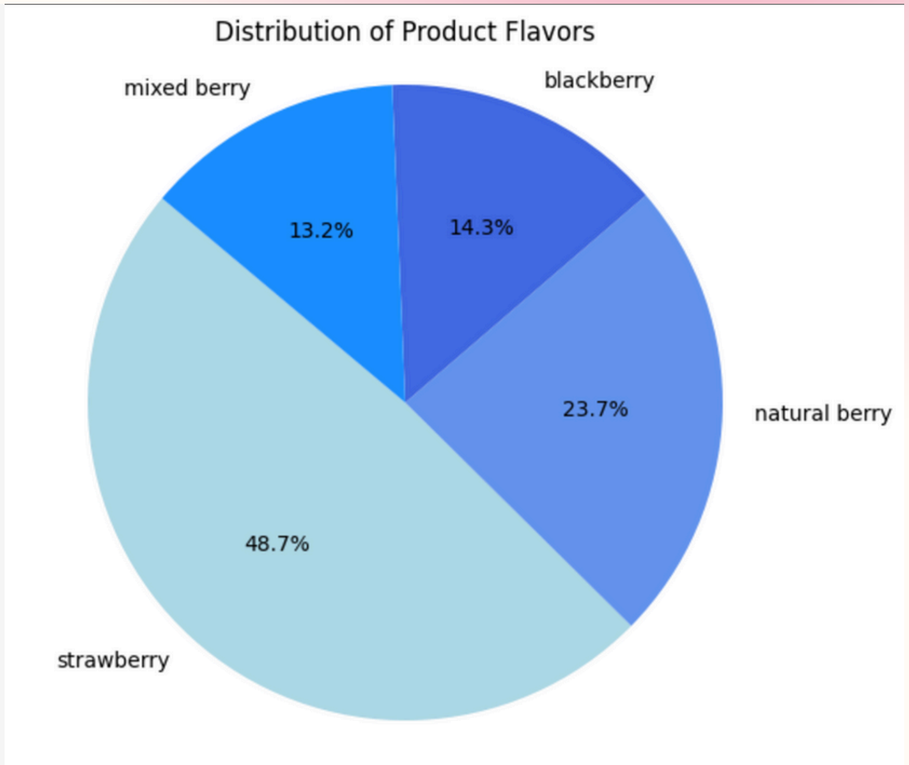
```
def extract_dose(title):
 match = re.search(r'\b(\d+\.?\d*)\s?mg\b', str(title), re.IGNORECASE)
 return float(match.group(1)) if match else None

combined_data['Dose_mg'] = combined_data['ProductTitle'].apply(extract_dose)
```

```
Unique Doses Extracted: [10. 12. 1. 20. 3. 5.]
```

Glimpse on data processing

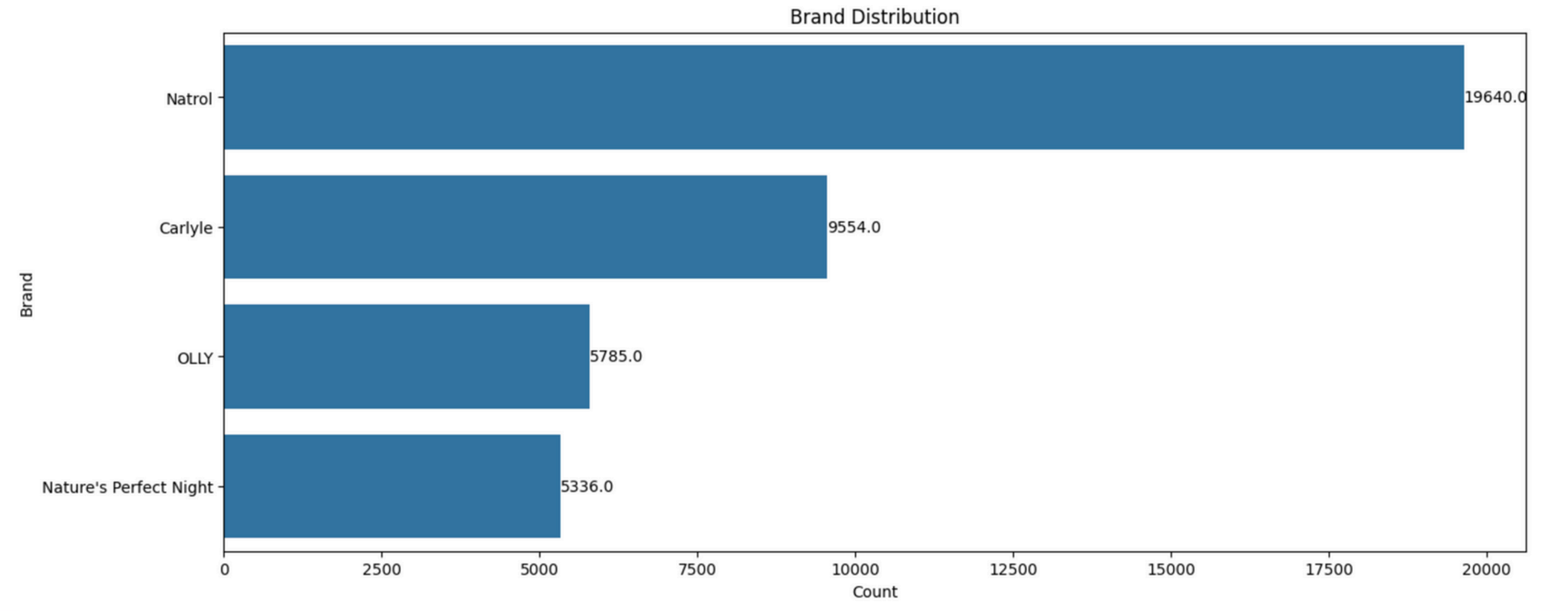
Extraction of flavour and form from product title

```python
def extract_form(title):
    title = str(title).lower()
    if 'gummy' in title or 'gummies' in title:
        return 'Gummy'
    elif 'tablet' in title:
        return 'Tablet'
    else:
        return 'Other'

def extract_flavor(title):
    flavors = ['strawberry', 'mixed berry', 'natural berry', 'blackberry', 'cherry']
    for flavor in flavors:
        if flavor in str(title).lower():
            return flavor
    return 'unknown'


combined_data['Form'] = combined_data['ProductTitle'].apply(extract_form)
combined_data['Flavor'] = combined_data['ProductTitle'].apply(extract_flavor)
```



Exploratory Data Analysis

Brand Distribution over products



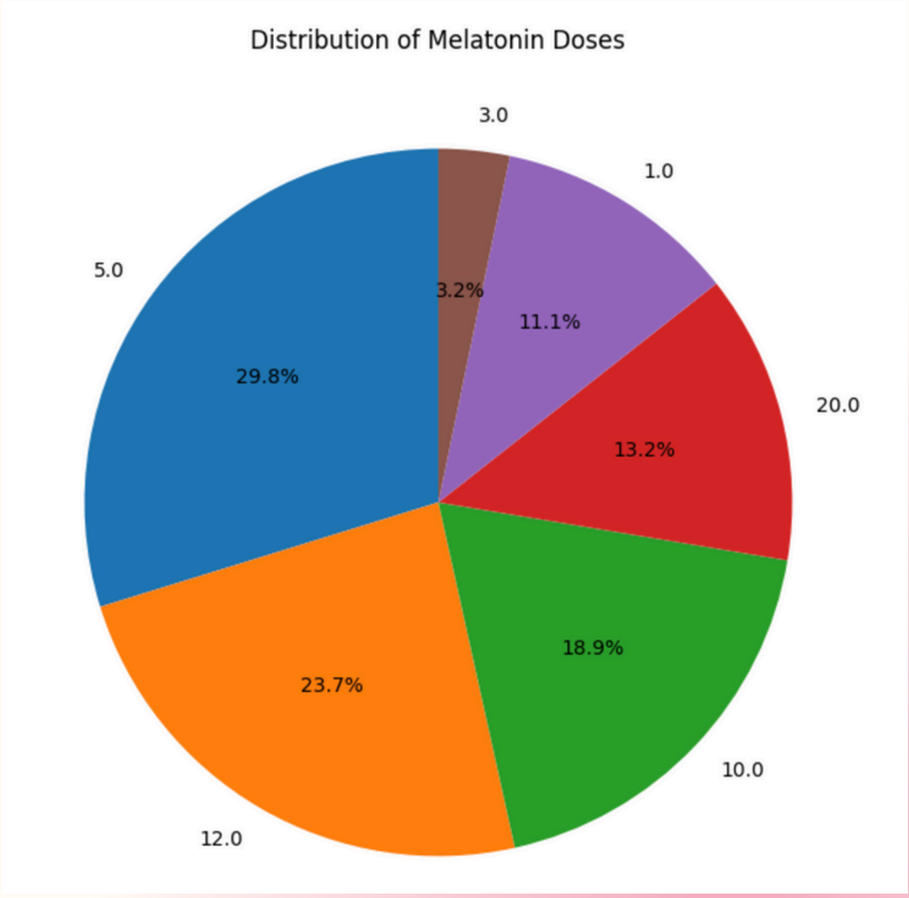
Natrol provides the highest number supplements while **Nature’s Perfect Night** being the lowest.

Exploratory Data Analysis

Distribution of melatonin doses over the products

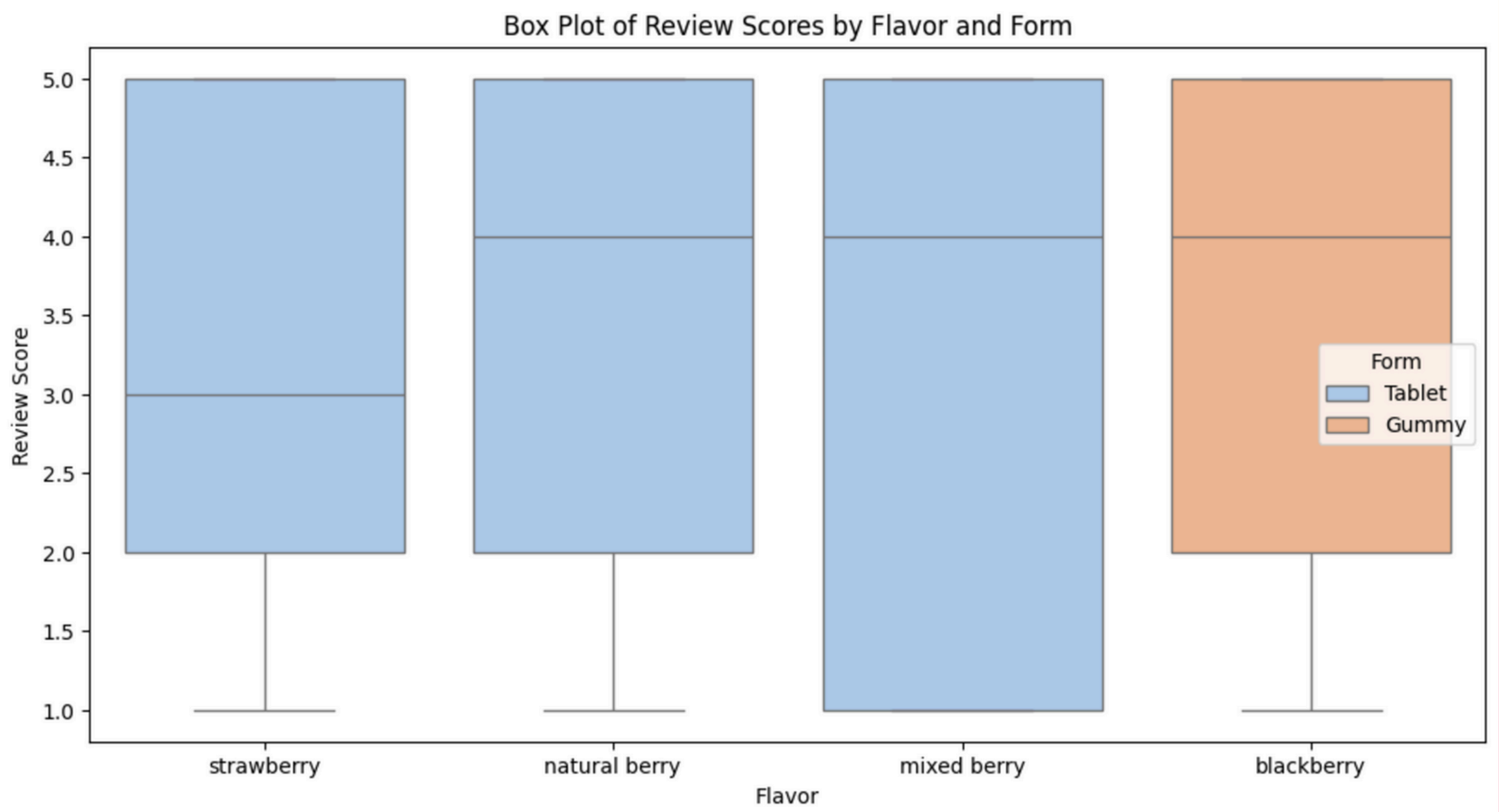
The distribution of melatonin supplements by dosage reveals key consumer preferences.

- The **5 mg dosage is the most consumed**, accounting for nearly 30% of the data
- followed by 12 mg (23.7%) and 10 mg (18.9%).
- Lower and higher dosages (1 mg, 3 mg, 20 mg) show smaller proportions, reflecting specific but niche demand.



Exploratory Data Analysis

Multivariate Analysis: Using Review, Form, and Flavor



The distribution of review scores for melatonin supplements highlights differences in consumer preferences based on flavor and form.

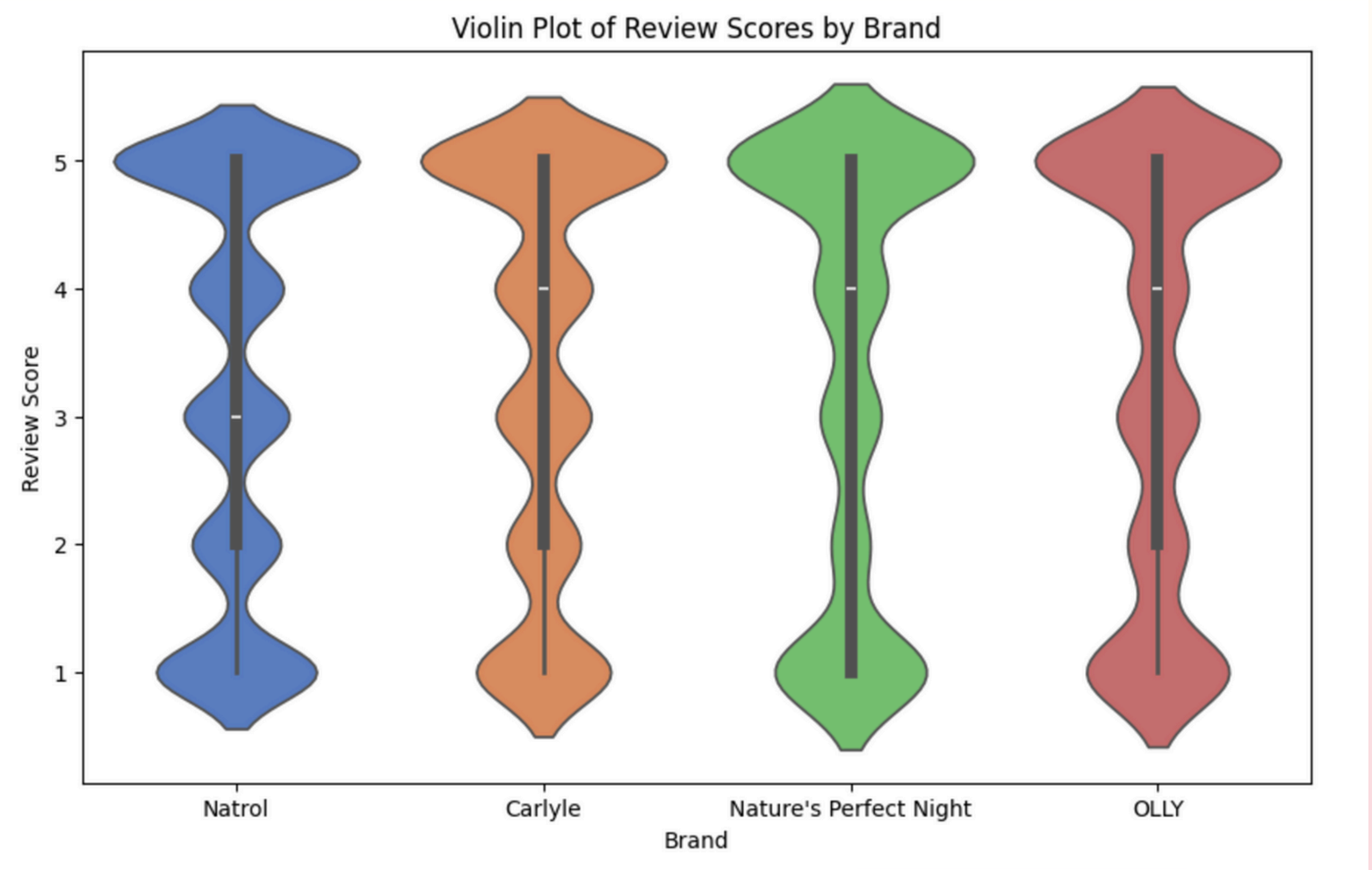
- Tablets dominate strawberry, natural berry, and mixed berry flavors, with consistent scores across these categories.
- Blackberry-flavored gummies show greater score variability, indicating diverse consumer experiences and preferences.

Exploratory Data Analysis

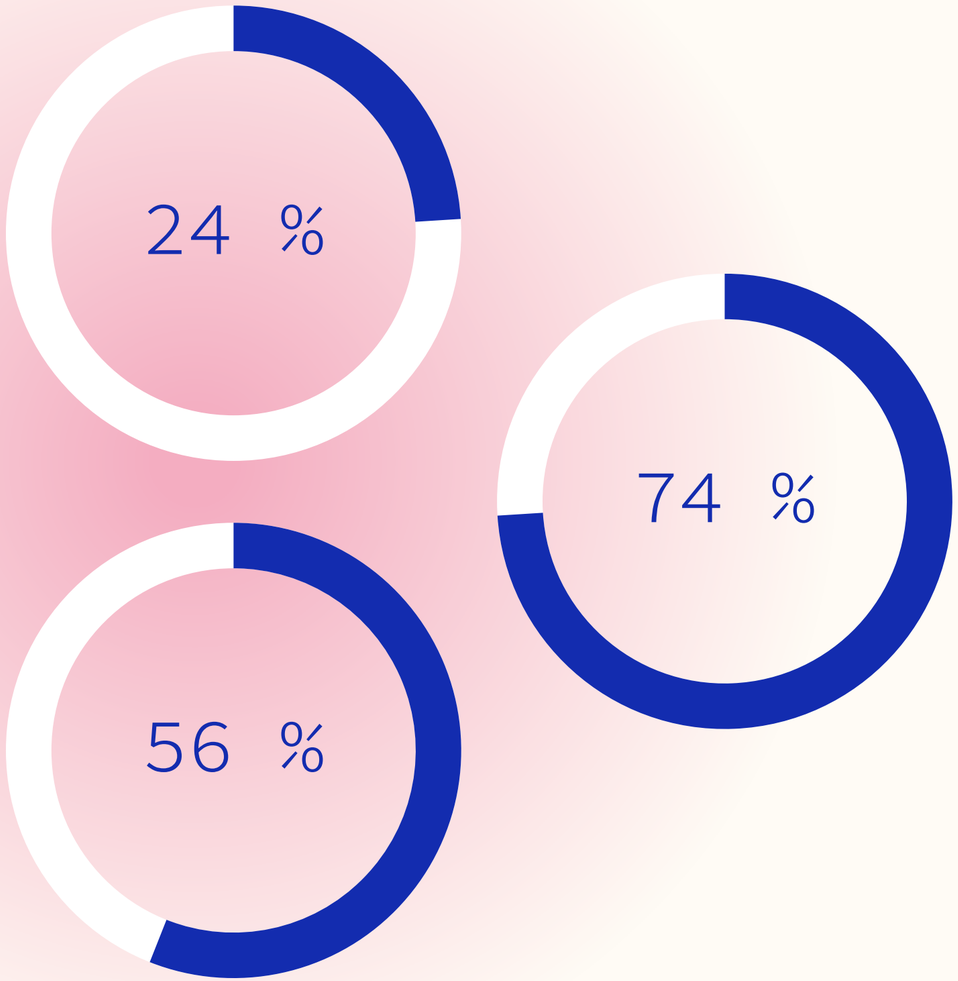
Bivariate Analysis: Using Review Score and Brand

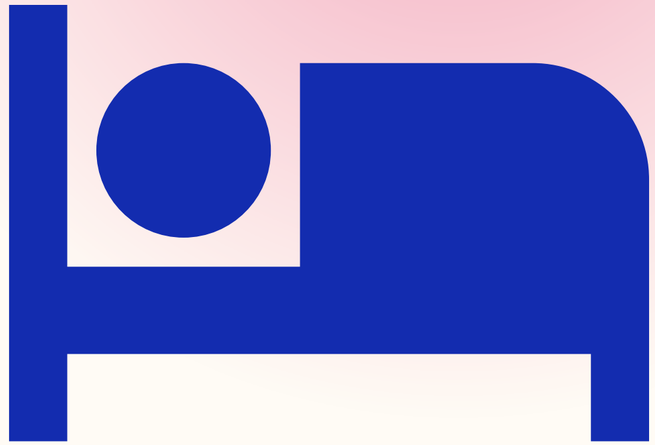
- Natrol Reviews are widely distributed from 1 to 5, but there’s a notable concentration around the higher range (4-5).
- Carlyle distribution is tighter compared to Natrol, with most reviews clustered between 3.5 and 5.
- Nature's Perfect Night Reviews are heavily concentrated between 4 and 5, with almost no scores below 3.

Overall, Nature's Perfect Night demonstrates the most consistent satisfaction, while Natrol has the widest range of experiences.



Data Visualizations

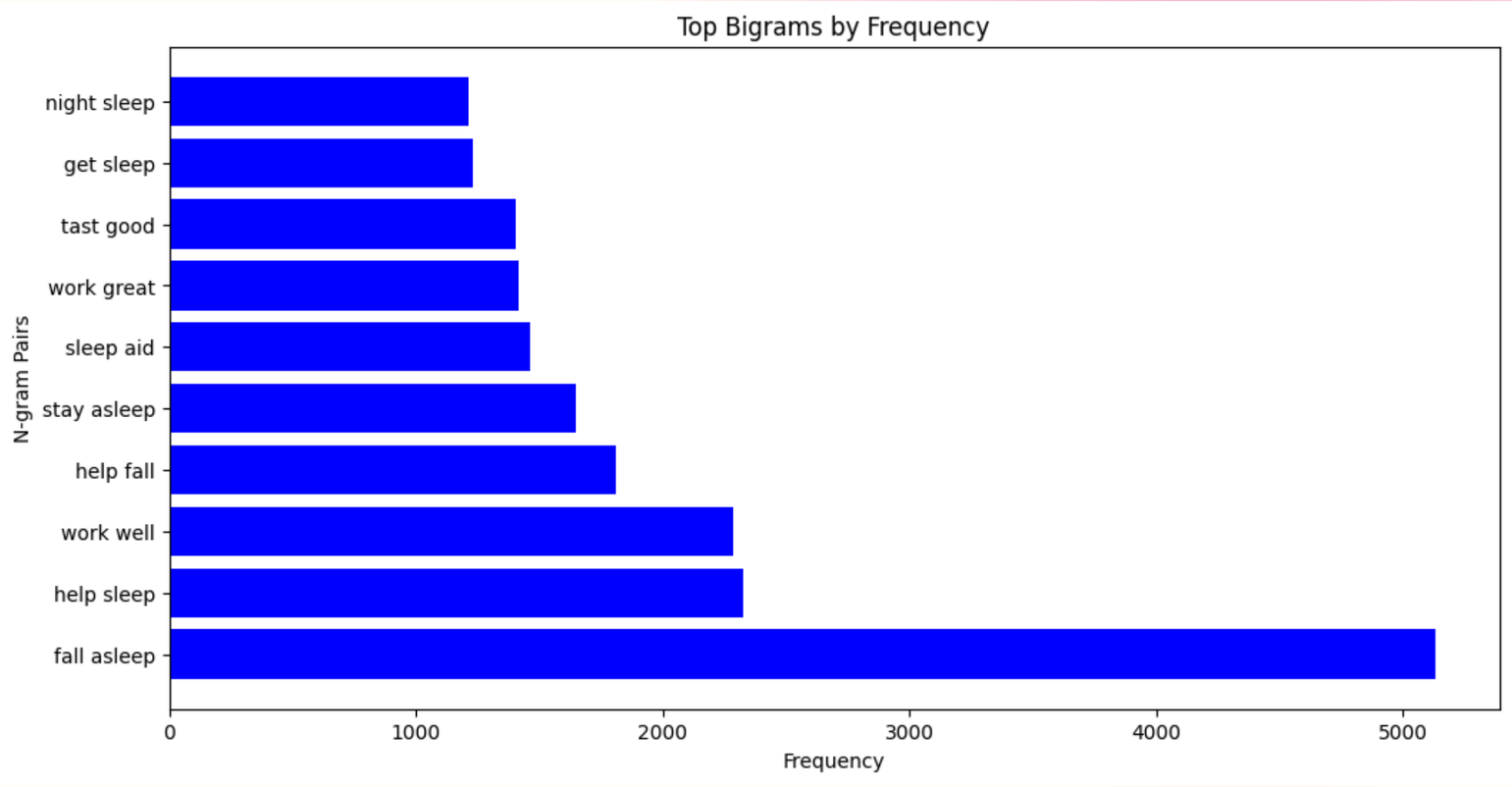




NLP from Review: Analysis and Insights

Ngrams

Bigrams



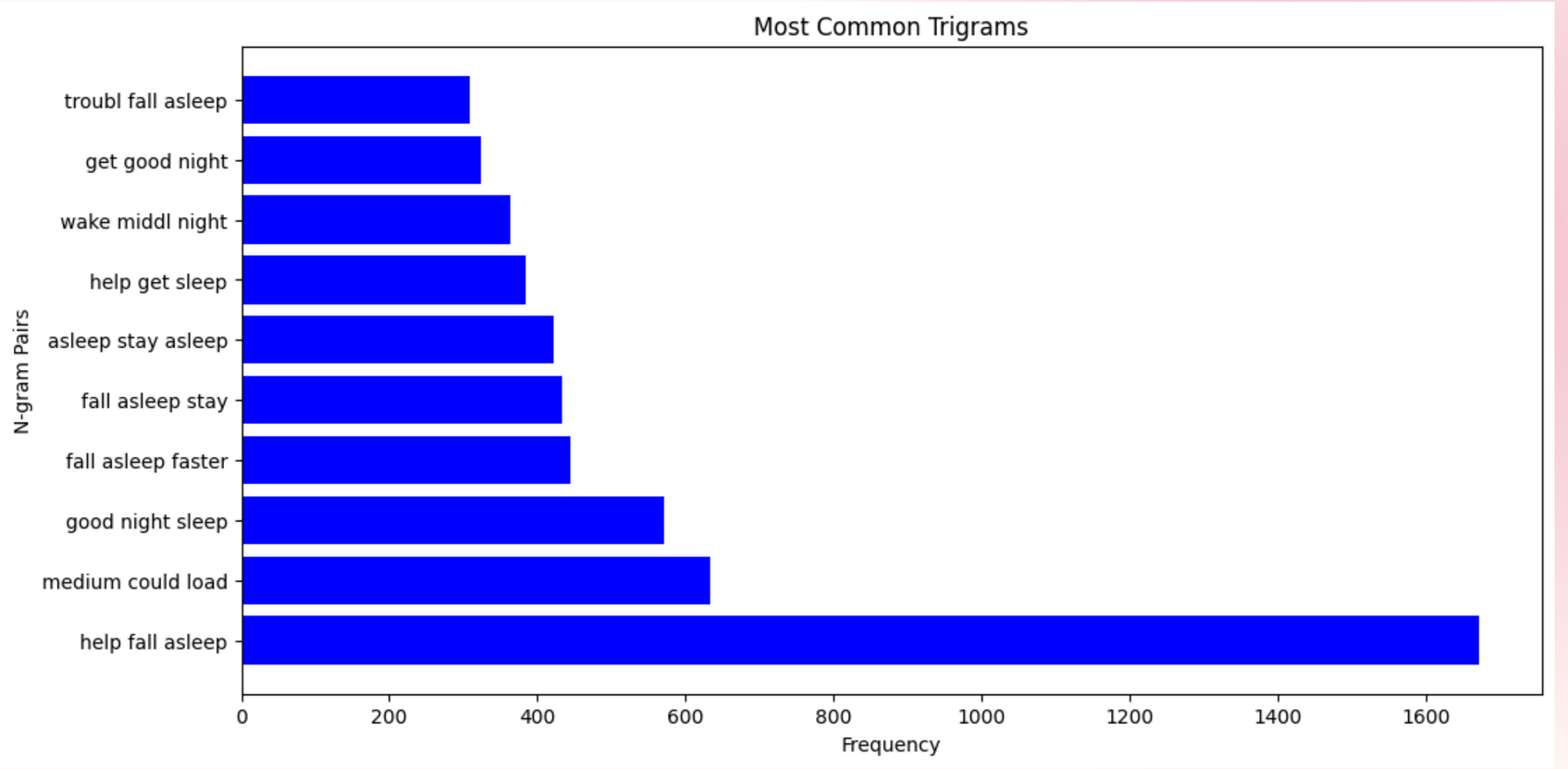
The analysis shows how customers naturally describe their experiences, focusing on product effectiveness, timing, and overall satisfaction with melatonin products

- Common combinations: "fall asleep," "help sleep"
- Time indicators: "next day," "night sleep"
- Effect descriptions: "work great," "sleep well"

Ngrams

Trigrams

The analysis shows how customers naturally describe their experiences, focusing on product effectiveness, timing, and overall satisfaction with melatonin products



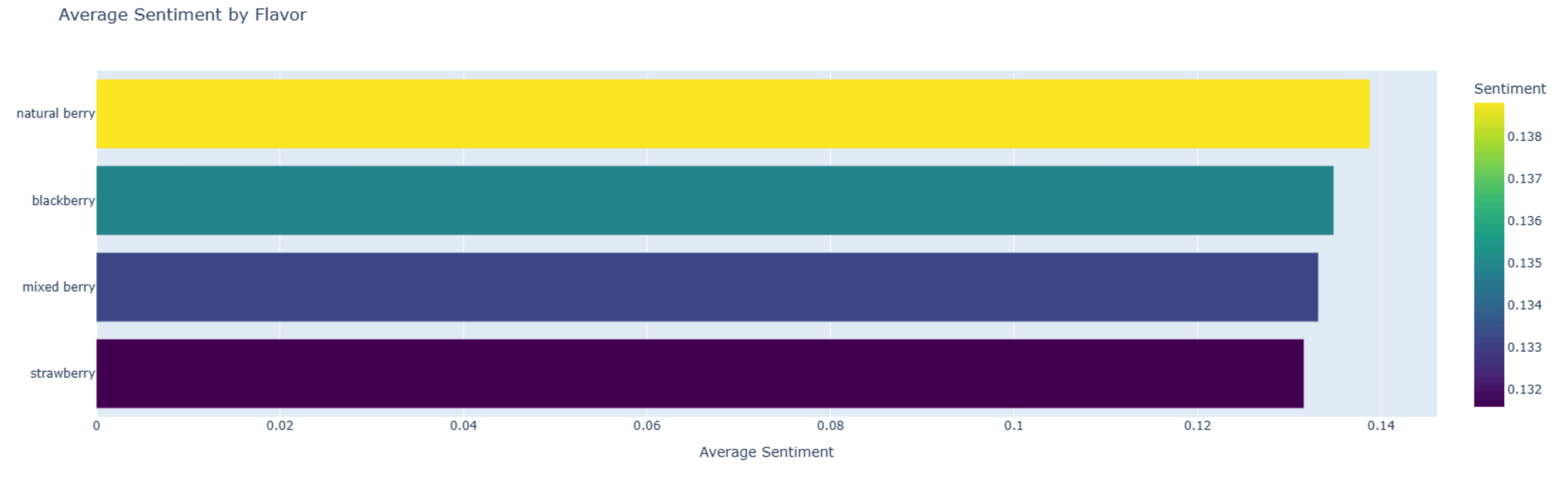
- Usage patterns: "help fall asleep"
- Time-related phrases: "fall asleep fast"
- Experience descriptions: "work really well"

Word Cloud

Insight Interpretation

- Consumer focus on sleep effectiveness
- Product taste experience importance
- Timing patterns of usage
- Overall positive sentiment indicators through terms like "good," "great," "help"
- Usage behavior patterns through action words like "take," "work," "help"

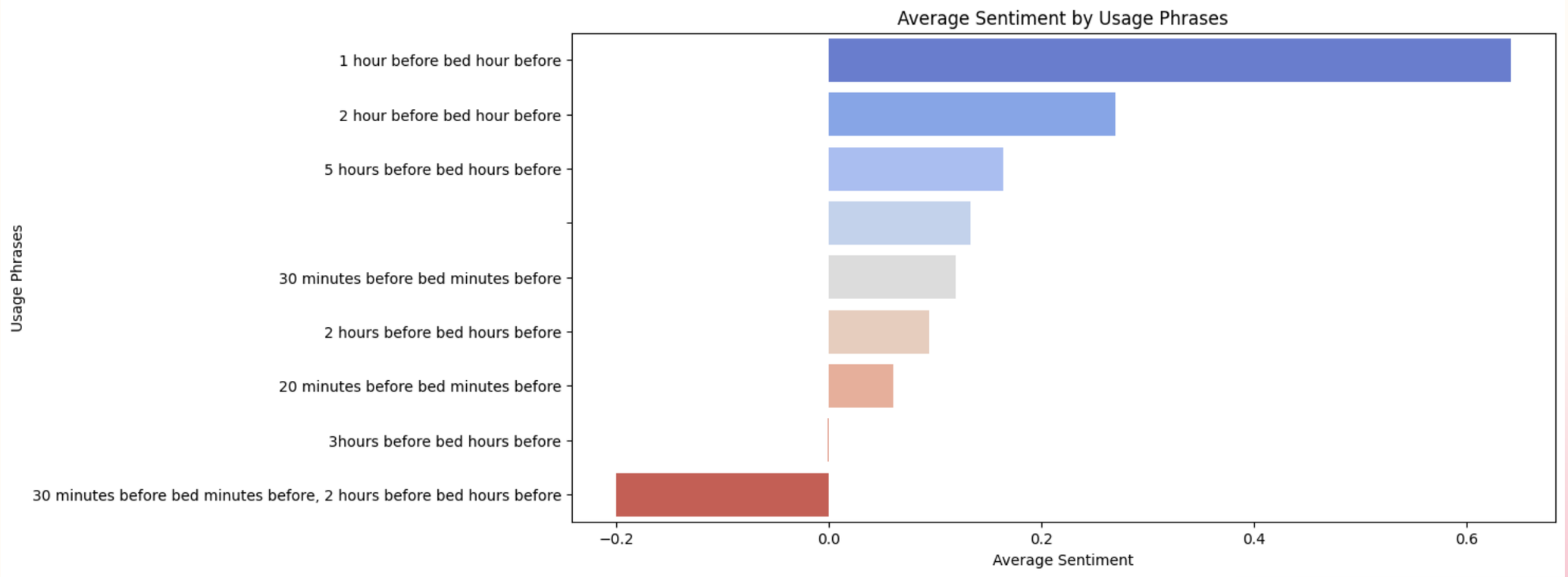
Review Insights



Average Sentiment by Flavor:

- Strawberry leads with highest sentiment score
- Natural berry, blackberry, mixed berry follow closely
- All flavors maintain positive sentiment (0.12-0.14 range)

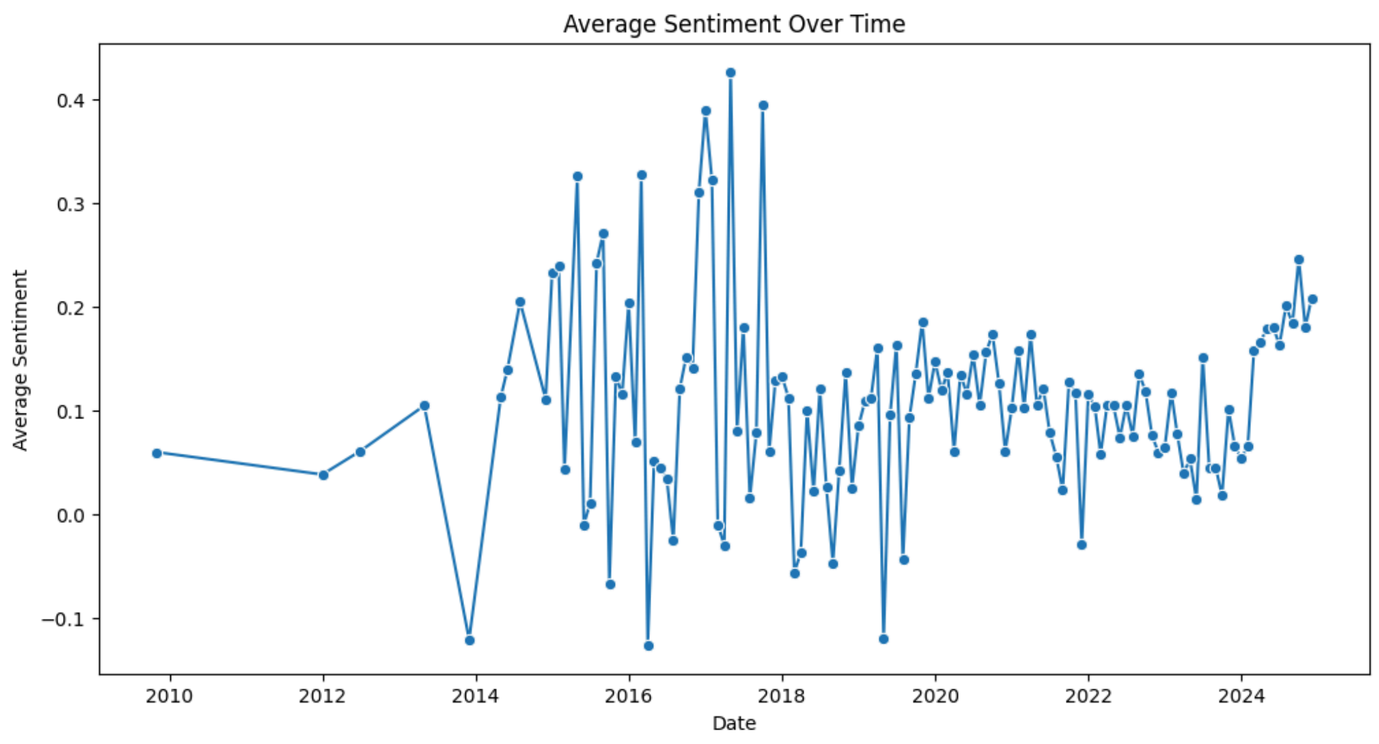
Review Insights



Average Sentiment by Usage Phrases:

- 1 hour before bed shows highest sentiment (0.6)
- 2-5 hours before bed shows moderate positive sentiment
- 20-30 minutes before bed shows lower effectiveness

Review Insights



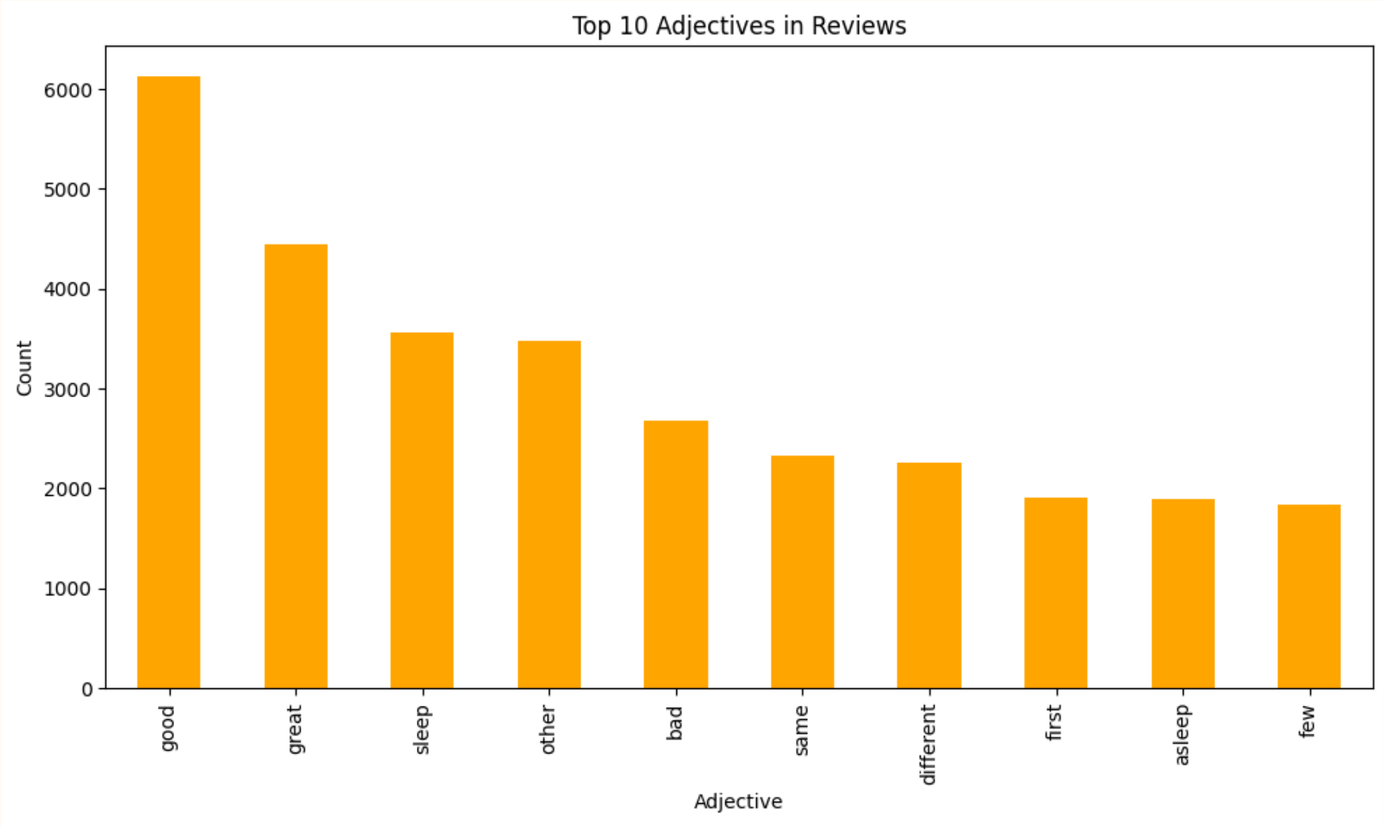
- **Evolution from volatile sentiment (-0.1 to 0.4) to stable positive sentiment (0.1-0.2)**
- **Sentiment stabilization post-2020 indicates product maturity**
- **Overall positive trajectory despite fluctuations**

Average Sentiment over Time

The visualization reveals that while melatonin product sentiment has fluctuated over the years, it maintains a generally positive trajectory. The stabilization in recent years suggests consistent product satisfaction, with fewer extreme variations in user sentiment compared to earlier periods.

Review Insights

- **"Good" dominates with 6,000+ mentions**
- **"Great" follows with 4,500+ mentions**
- **Positive descriptors lead the top 3 positions**

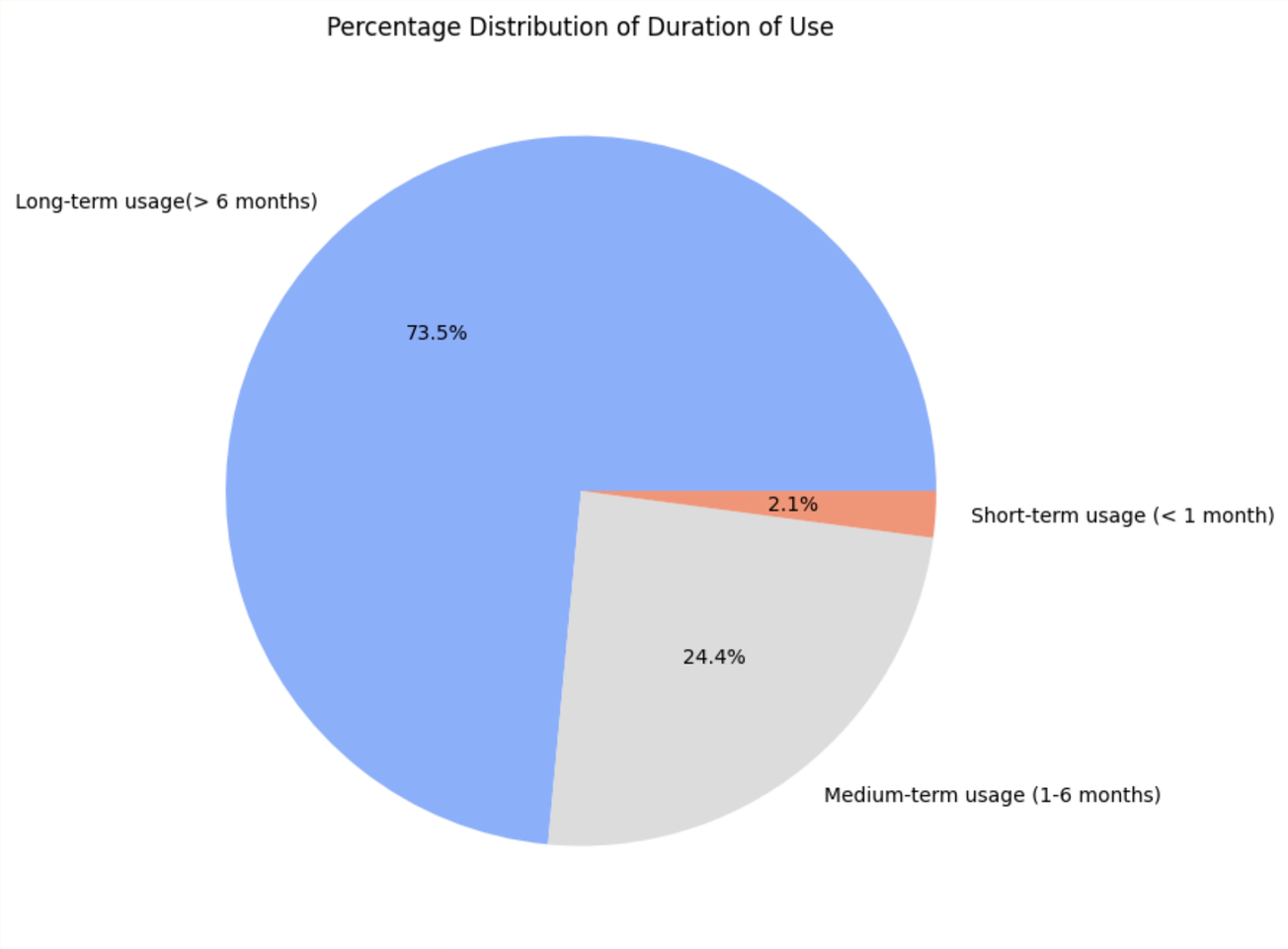


Top 10 Adjectives in Reviews

This reveals the presence of positive adjectives, representing nearly twice the frequency of neutral or negative terms, indicates overwhelmingly positive user sentiment toward melatonin products

Review Insights

- **Usage Duration DistributionLong-term users (>6 months): 73.5%**
- **Medium-term users (1-6 months): 24.4%**
- **Short-term users (<1 month): 2.1%**
- **Word Cloud AnalysisPrimary terms: "sleep," "help," "fall," "work"**
- **Product experience: "taste," "good," "great"**
- **Usage patterns: "night," "take," "within"**

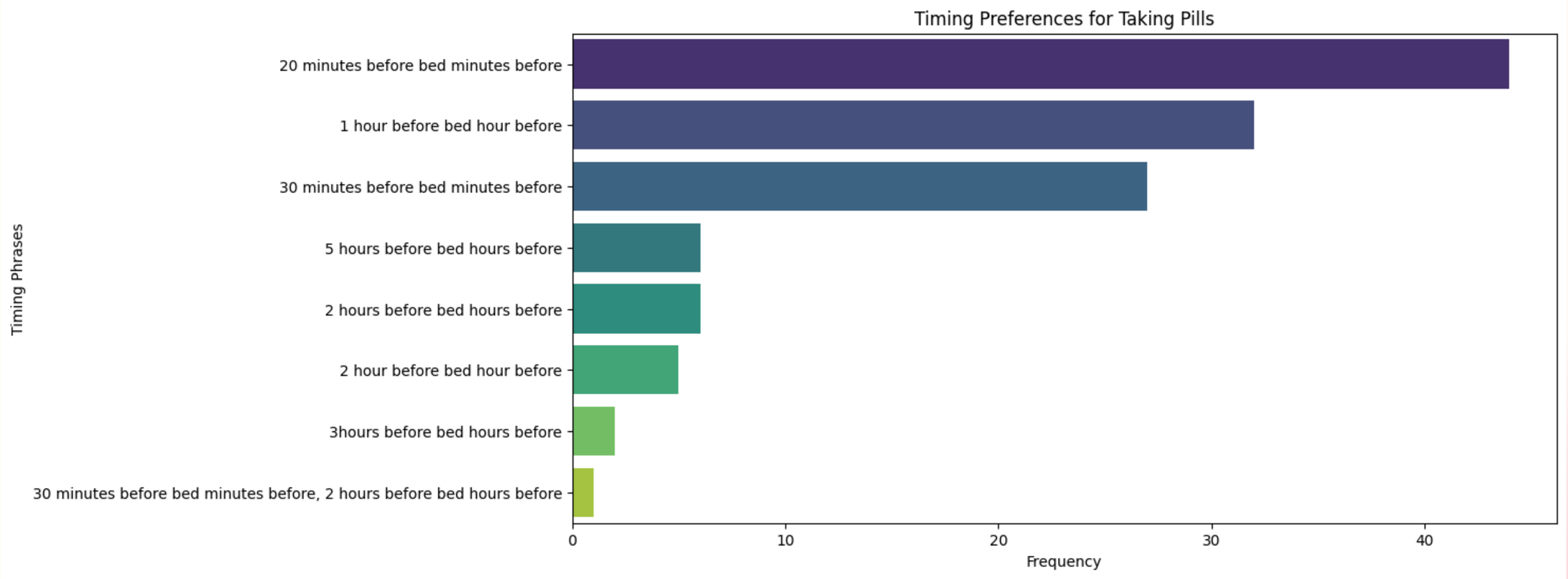


Duration of use

The data indicates strong product loyalty with most users continuing long-term use, suggesting high satisfaction and effectiveness of melatonin products.

Review Insights

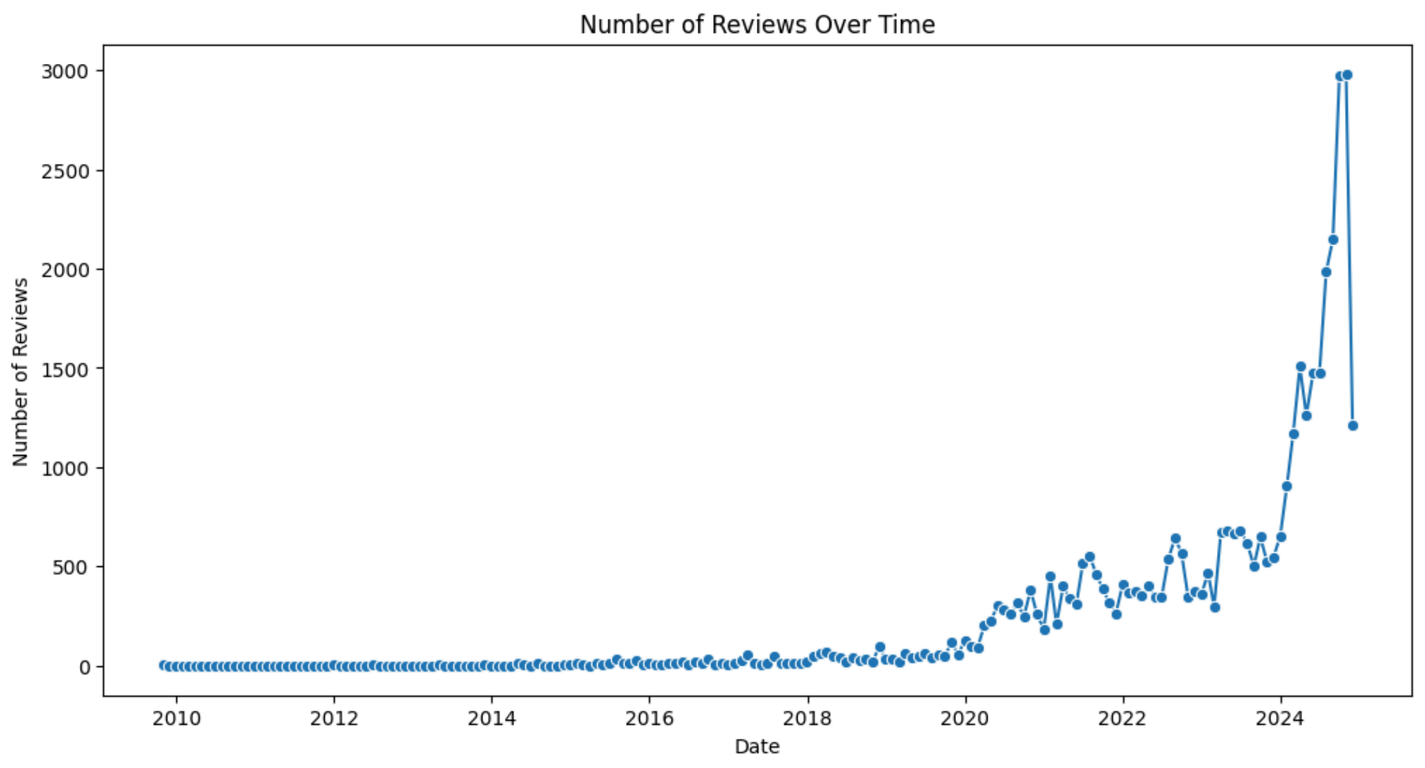
Different time taking pills



Timing Preferences Analysis

- Most popular timing: 20 minutes before bed
- Second most common: 1 hour before bed
- Third preference: 30 minutes before bed
- Less common timings: 2-5 hours before bed
- Minimal usage: 3 hours or longer before bed

Review Insights



- **Dramatic growth from minimal engagement (~50 reviews/month in 2010-2018) to explosive popularity (3,000 reviews peak in 2024)**
- **60x increase in review volume over 14 years**
- **Post-2020 exponential growth suggests significant market expansion**

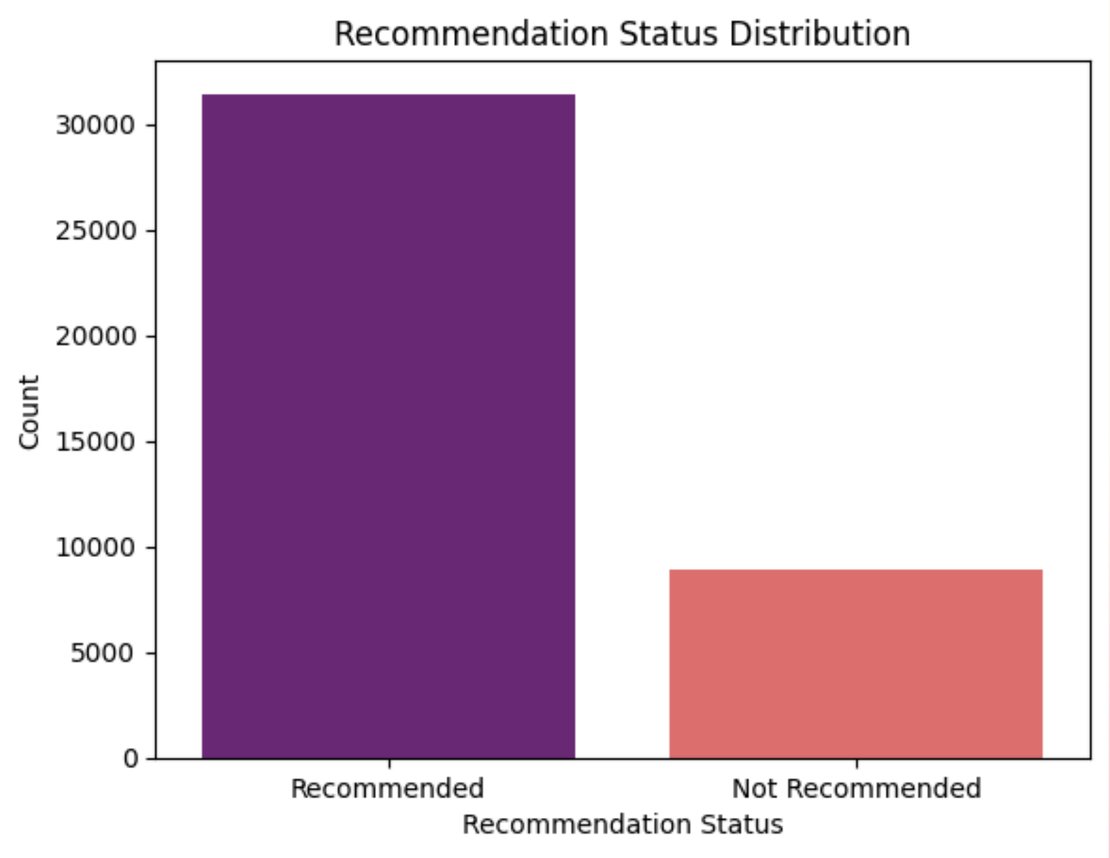
Number of Reviews Trend Over Time

The visualization reveals a clear evolution in review patterns, with recent years showing dramatically increased user engagement compared to the product's early history. This suggests growing popularity and market presence of melatonin products over time.

Review Insights

The visualization reveals:
Strong positive
- **Recommendations**: 30,000 users
- **Not recommended**: 8,000 users
- Approximately 80% recommendation rate

Key Insights
- Helpful votes from other users
- Review scores (weighted 10x)
- Clear majority of users recommend melatonin products
- Threshold set at 15 points distinguishes recommendations

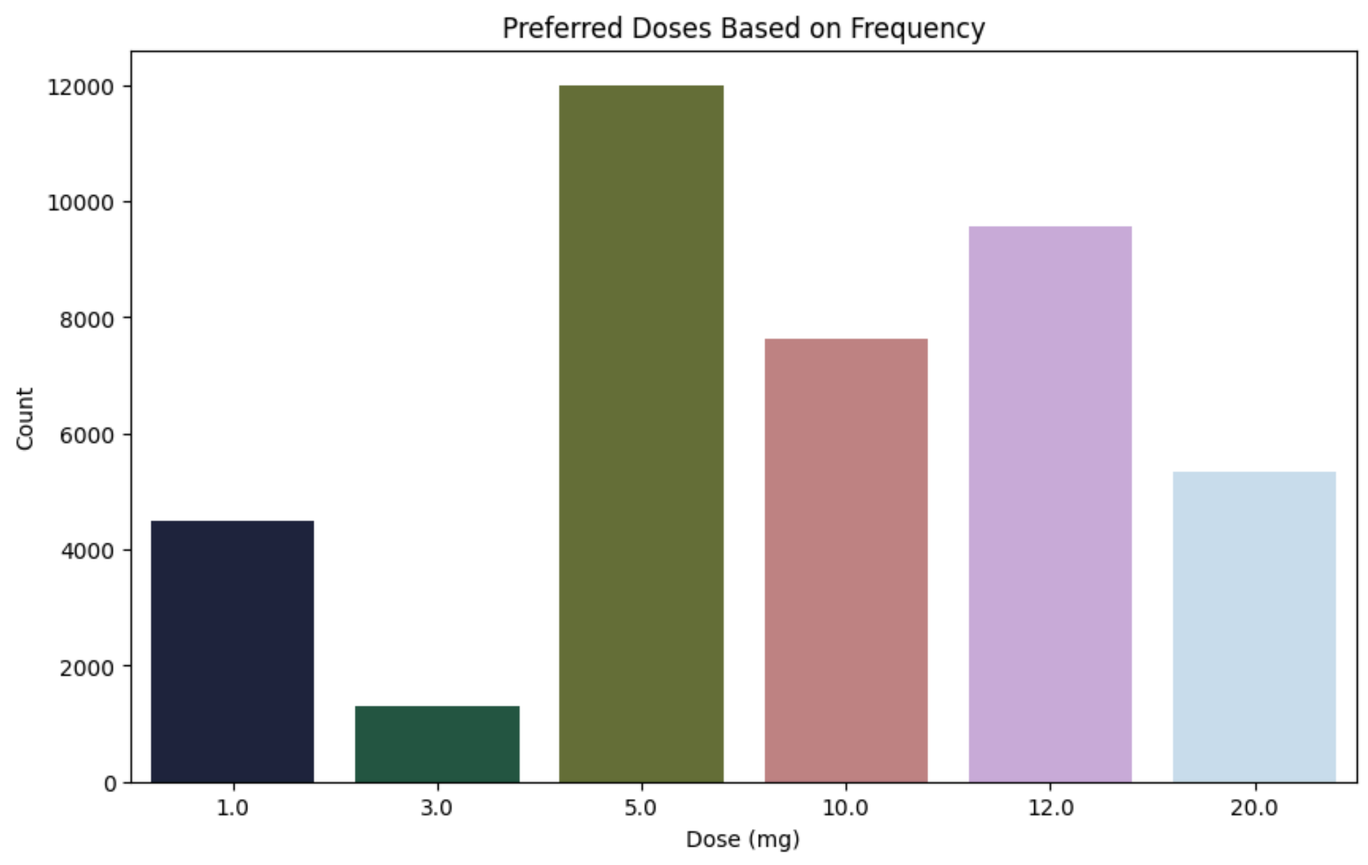


Melatonin Recommendation

- Highly recommended: 30,000 users
- Not recommended: 8,000 users
- Clear 4:1 positive recommendation ratio

Review Insights

- **Clear preference for 5mg dosage (12,000 users) over all other options**
- **Higher dosages (10-12mg) account for significant market share (~17,000 users combined)**
- **4:1 ratio between most popular (5mg) and least popular doses (1-3mg)**



Preferred Dosage Count

The data indicates that consumers strongly prefer moderate dosages around 5mg, suggesting this may be the optimal balance between effectiveness and safety. Higher doses remain popular but aren't the primary choice, while lower doses see comparatively less adoption among users.

Contributions

Sunny

- Loading the dataset.
- Data processing
- Extract MG details from the product title.
- Concatenate the data frames.
- Analyze and Visualize various variables (univariate, bivariate, multivariate if possible)
- Perform NLP Data cleaning
- Check the Frequency of words like top 10 or something.
- Form N Grams
- Form Word Clouds
- Perform Sentiment Analysis
- Check Sentiment of each doses.
- Distribution of the ratings for each dose.
- Average rating over doses.

Shushil

- How users take melatonin: thirty minutes before bed, ten minutes before bed, etc.
- Users' reactions to taking melatonin: Hot words extraction and analysis, such as 2-gram, or
- 3-gram etc.
- The intensity of user emotions: statistics and analytical adjectives
- Sentiment analysis: positive, neutral, negative analysis, Polarity and Subjectivity, etc.
- Check the sentiment over the time.
- Number of reviews trend over time
- What doses People Prefer?

Thank you