

DAMG6105 Instructor: Dr. Handan Liu

Project Code Review 10%

Project Presentation 5%

Grading: total 15 points.

Submission format:

Please finish the following questions in ONE Jupyter file with running results for each question in order. The detail policy can be found in the course syllabus.

Submission via Canvas. You have 2 chances to submit it.

Submission follows the submission policy. An extra day is given for late submission, but with a penalty. Penalties for late submissions can be found in the course syllabus.

Review and Grade:

Project Presentation – In class – counts for 5% in the final grade.

Code Review - TAs will conduct code review with you and grade your code. This part counts 10% in the final grade.

Data Background:

The market for melatonin products is very chaotic, with products of various doses abounding. The aim of this project is to analyze the distribution and user response of various doses of the product based on this data.

The datasets are scraped from Amazon and Walmart.

Requirements:

Two main parts: EDA and NLP

Suggested Analysis Content:

EDA and Visualization:

- Different brands
- Different doses: 1mg, 3mg, 5mg, 10mg, 15 mg, 20mg, 25mg etc.
- Different forms: Gummy or tablet
- Different flavors: strawberry, citrus, etc.
- After NLP, different rating, and with different doses
- And so on.

NLP from Review: Analysis of different doses from reviews:

- Different time taking pills, e.g. 10 minutes before go bed or after feel sleepy; I had no problem falling asleep within an hour of taking these tablets..... ..

- How long after taking the pills do you feel sleepy:
- Duration of use: e.g. Tasty and effective; been using for over a year
- Will you recommend it to others?
- And so on

Suggested Steps:

1. Load the dataset.
2. Check shapes.
3. Extract MG details from the product title.
4. Concatenate the data frames.
5. Analyze and Visualize various variables (univariate, bivariate, multivariate if possible)
6. Perform NLP Data cleaning (not only limited to)
 - a. Stemming
 - b. Lemmatization
 - c. Removing Stop words
 - d. Punctuations
 - e. Common words like Package etc.
7. Check the Frequency of words like top 10 or something.
8. Form N Grams
9. Form Word Clouds
10. Perform Sentiment Analysis
11. Check Sentiment of each doses.
12. Distribution of the ratings for each dose.
13. Average rating over doses.
14. How users take melatonin: thirty minutes before bed, ten minutes before bed, etc.
15. Users' reactions to taking melatonin: Hot words extraction and analysis, such as 2-gram, or 3-gram etc.
16. The intensity of user emotions: statistics and analytical adjectives
17. Sentiment analysis: positive, neutral, negative analysis, Polarity and Subjectivity, etc.
18. Check the sentiment over the time.
19. Number of reviews trend over time
20. What doses People Prefer?

Refer to Nov 15th class Notebook and Nov 22nd class Notebook.

Note: You're encouraged to develop additional questions and visualizations based on your creativity and the insights you uncover from the dataset.

Attention: Don't copy the solutions from chatGPT or other online resources, otherwise the submission will be treated as plagiarism.

All the best!!