Amazon Review Language and Rating Analysis

Group 12: Connor Huang, Zeyu Li, Sahil Mehra, Linxiao Xia

Objective of the Project

- Amazon has an extensive dataset of review data and metadata
- Is there a way to find trends based on the category of products with their ratings and reviews?
- Could this information be used to figure out what is important to customers based on their positive and negative reviews?
- Does pricing affect the perception of customers?

★☆☆☆☆ **Beds should look like beds**

By Adam on June 30, 2016

I ordered this when I was drunk because I thought it was a giant ice cream sandwich.

Methodology

- Restructure and Filter the Amazon Data
- Filter the data by category, rating, and pricings.
- Eliminate common stop words by using the NLTK Stopwords Library (includes words such as “the”, “and”, “then”, etc.)
- Find the most common words for categories, ratings, and categories with ratings.
- Map ratings, categories, and prices together and make inferences on their relations with each other and the words used in the reviews. Subsequently, finding any other indicator of how well products may be perceived by a more broad generalization.

Dataset

Example Review in Data:

```
{
    "overall": 5.0,
    "verified": true,
    "reviewTime": "02 9, 2015",
    "reviewerID": "A18TW0CAFRIC1P",
    "asin": "0972683275",
    "reviewerName": "Dorvan",
    "reviewText": "It's a great tv mount I love the fact that u can
pretty much put your screen in any position u want",
    "summary": "great tv mount",
    "unixReviewTime": 1423440000
}
```

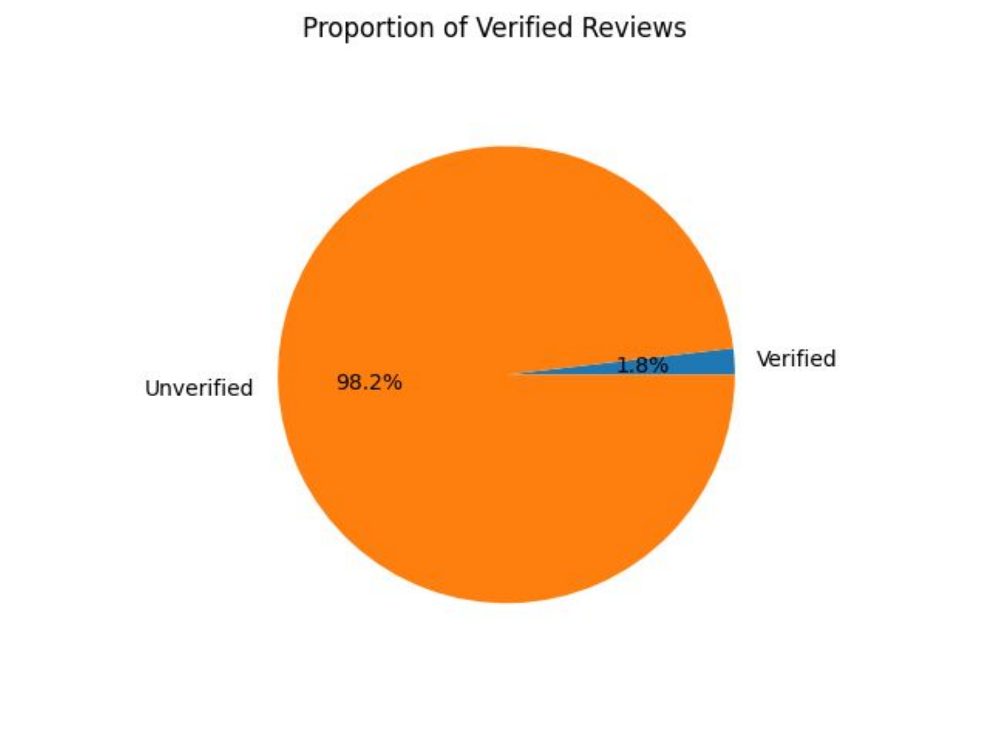
Example Product Metadata in Data:

```
{
    "asin": "0000031852",
    "title": "Girls Ballet Tutu Zebra Hot Pink",
    "feature": ["Botiquecutie Trade…],
    "description": "This tutu is great for dress up play …",
    “price": 3.17,
    "imageURL": "http://ecx.images-amazon.c…",
    "imageURLHighRes":
"http://ecx.images-amazon.com/images/I/51fAmVkTbyL.jpg",
    "also_buy": ["B00JHONN1S", "B002BZX8Z6",
"B00D2K1M3O", …],
    "also_viewed": ["B002BZX8Z6", "B00JHONN1S",
"B008F0SU0Y", …],
    "salesRank": {"Toys & Games": 211836},
    "brand": "Coxlures",
    "categories": [["Sports & Outdoors", "Other Sports",
"Dance"]]
}
```

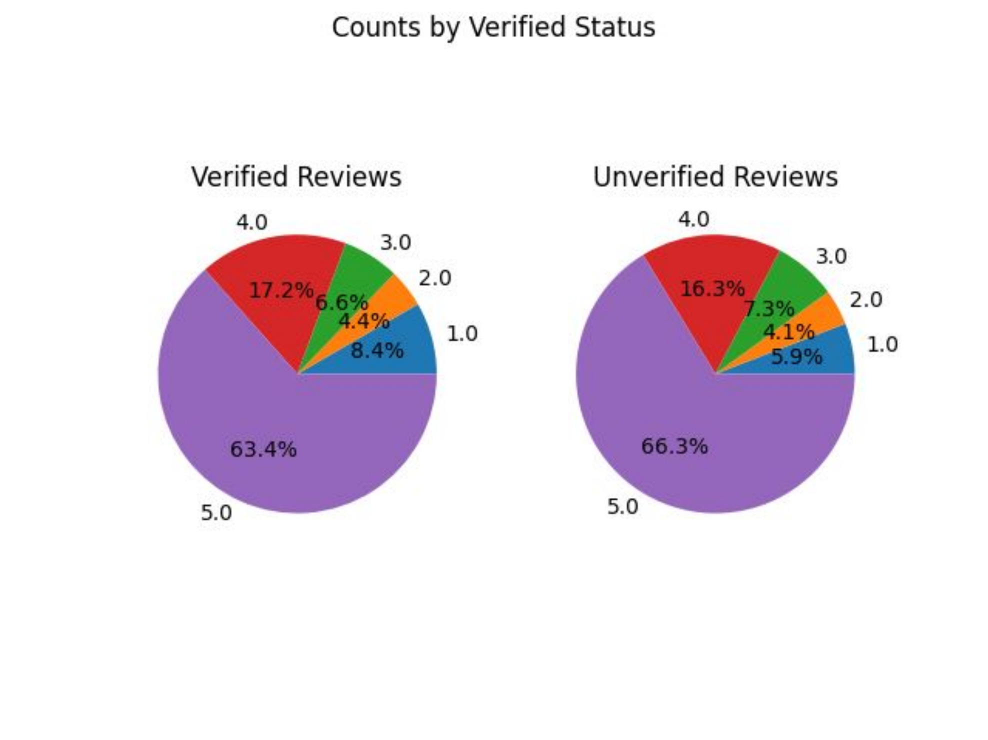
Filtered Dataset

Source Category	Product ID	Reviewer ID	Rating	Review Summary	Review Text	Has Image	Product Price	Is Verified
AMAZON_FASHION_5.json	B000K2PJ4K	ALJ66O1Y6SLHA	5.0	Five Stars	Great product and price!	False	$10.02 - $25.01	False

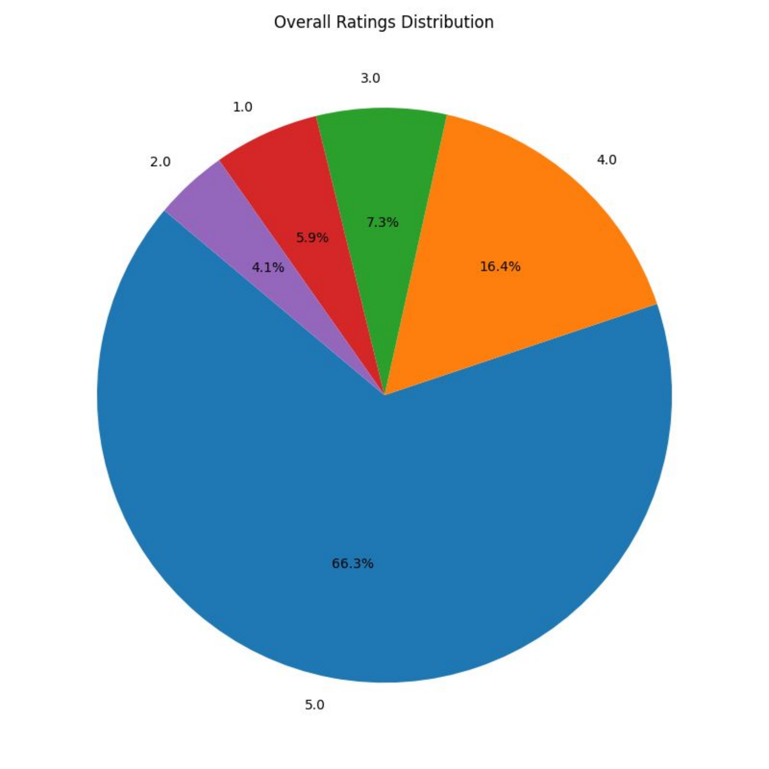
Proportion of Verified and Unverified Ratings



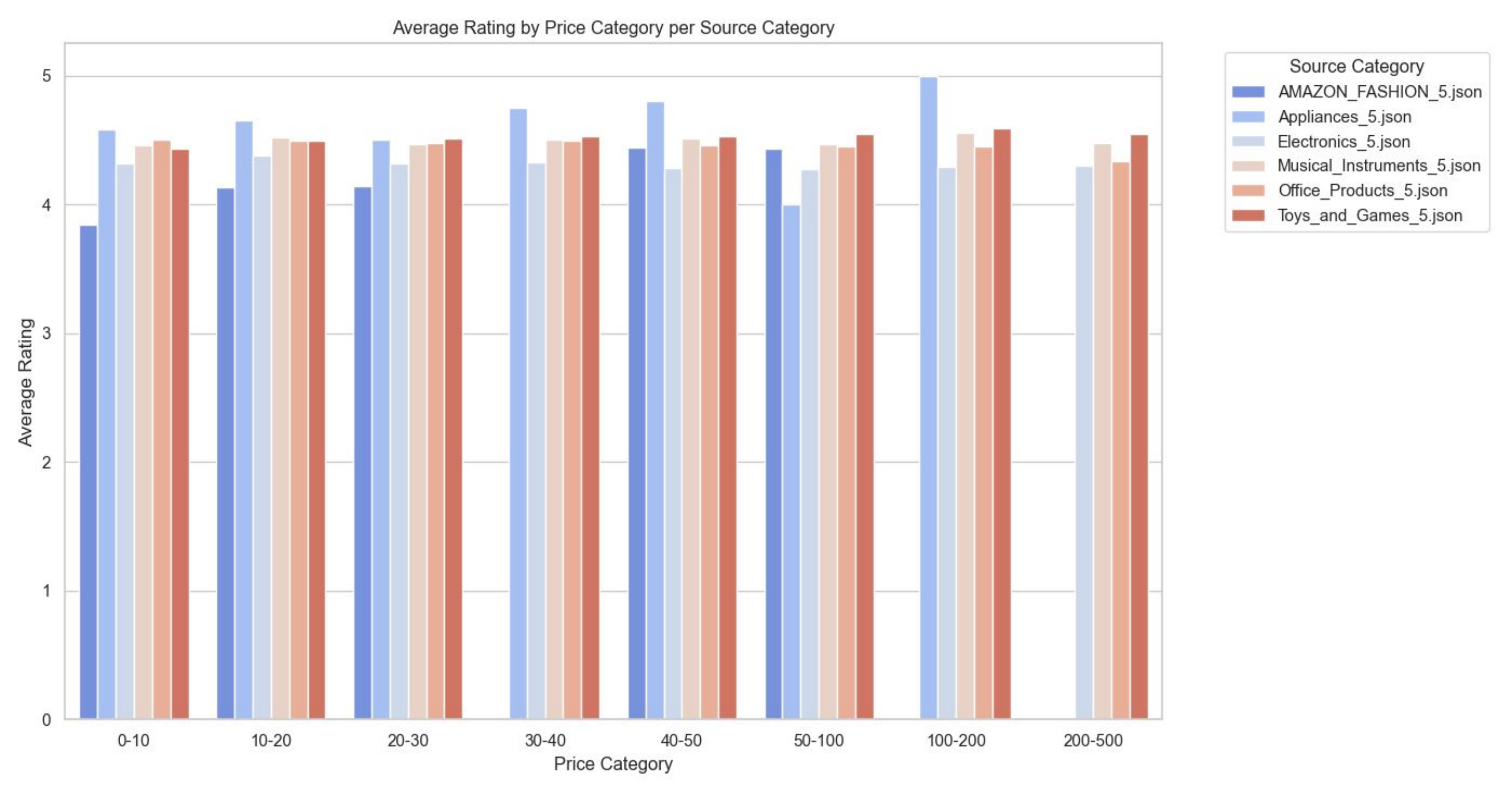
Verified to Unverified Ratings



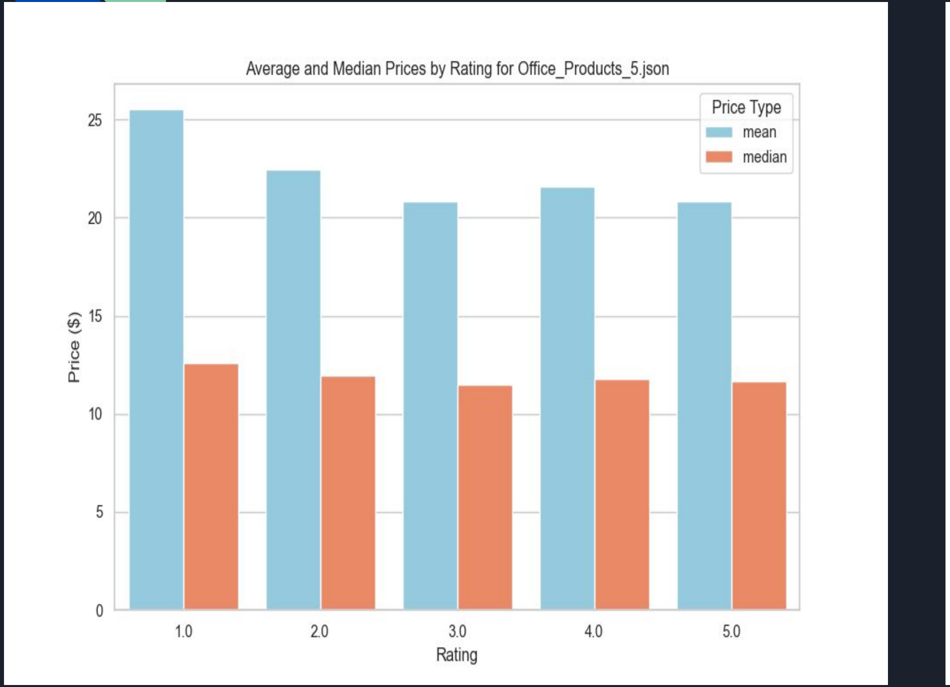
Overall Ratings Distribution

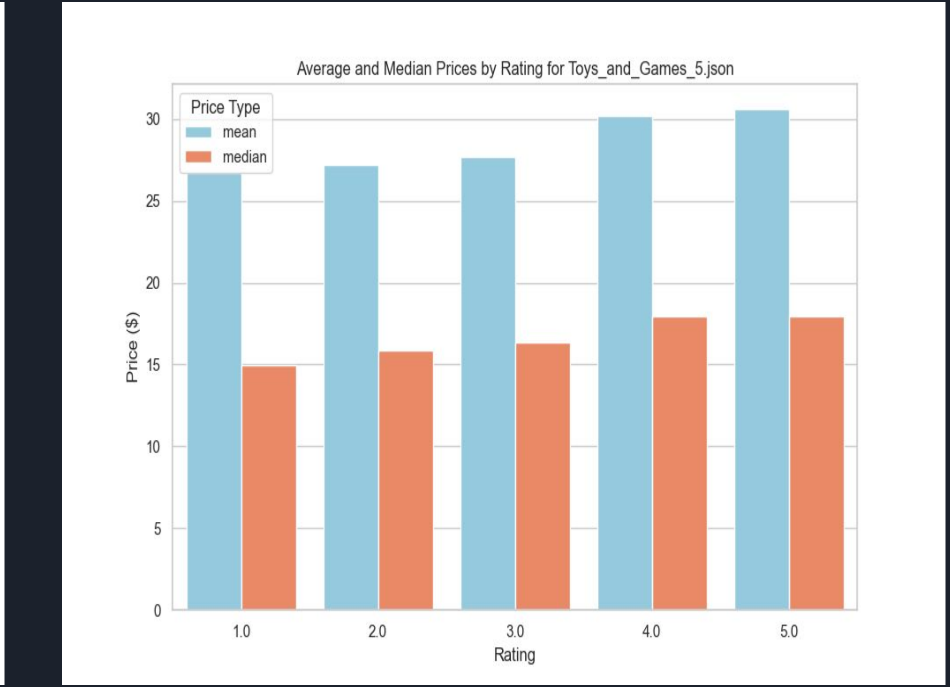


Ratings Distribution per Source Category

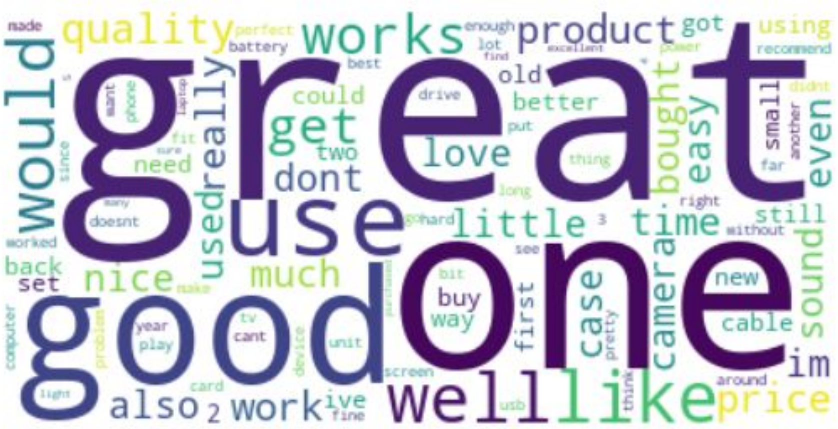


Average and Median Prices by Rating for each category

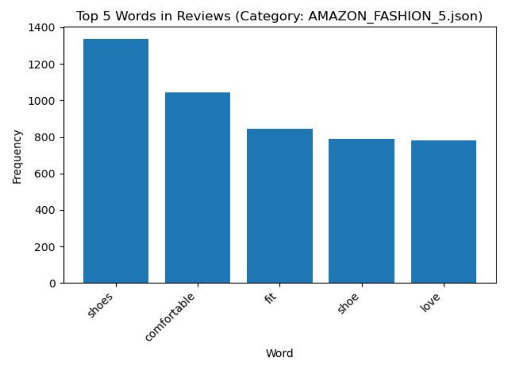


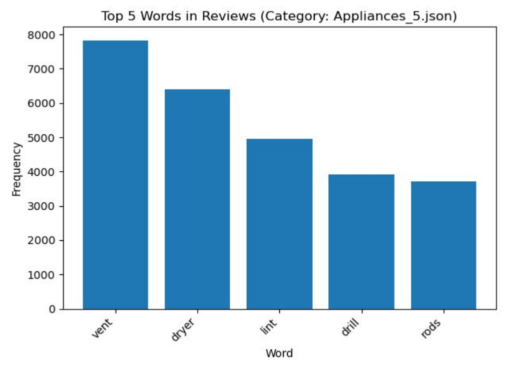


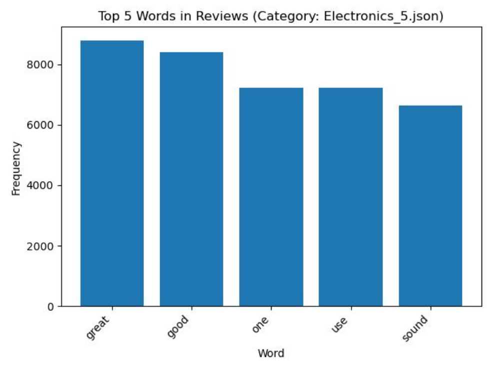
Overall Word Frequencies (word cloud)

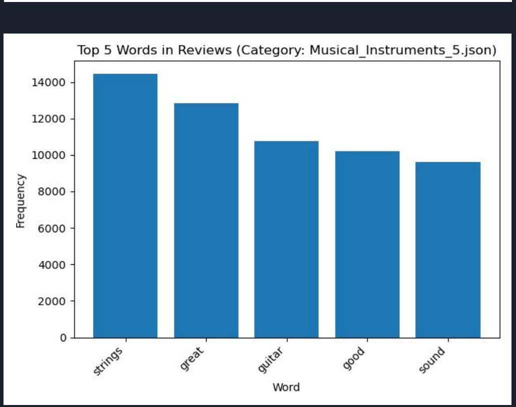


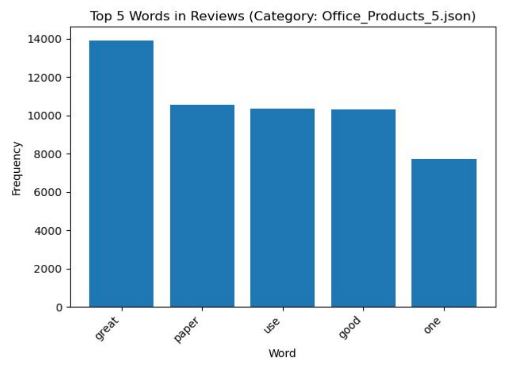
Most frequent words per category

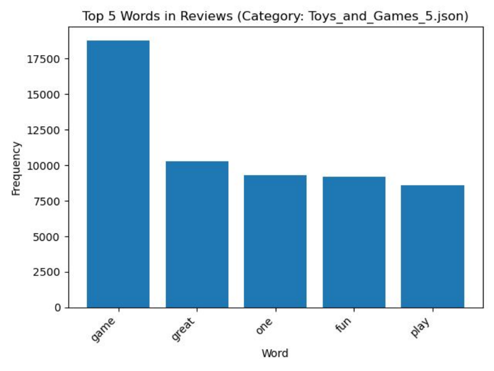




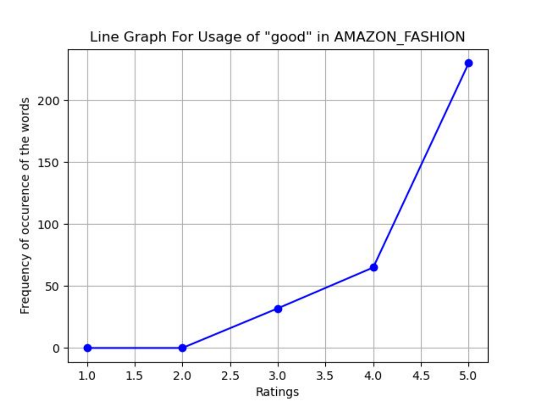


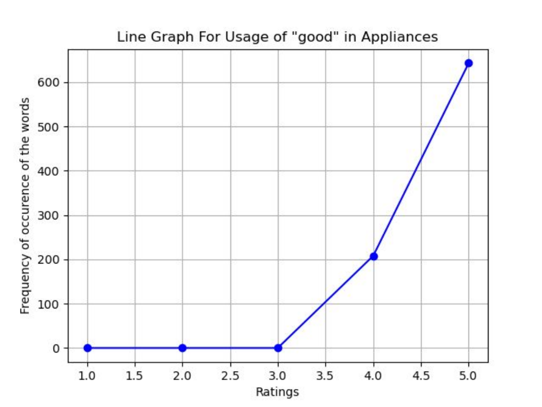


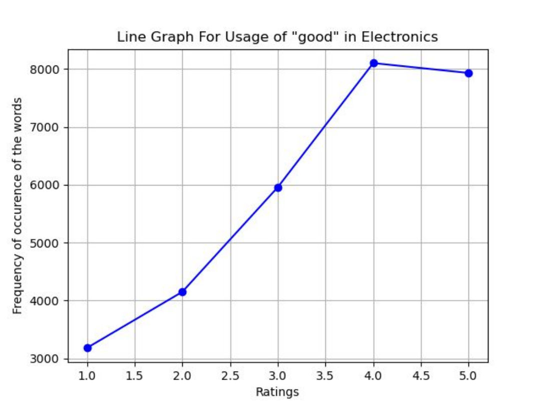


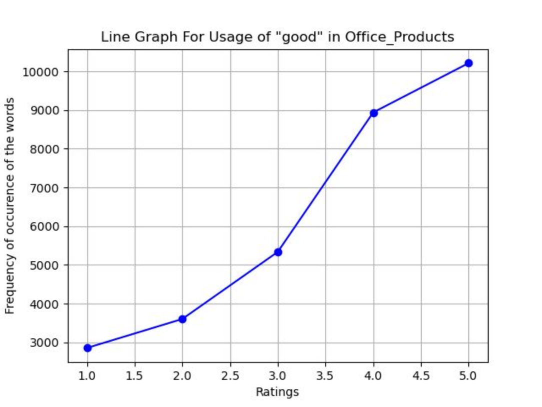


Usage of the word “good” and its synonyms

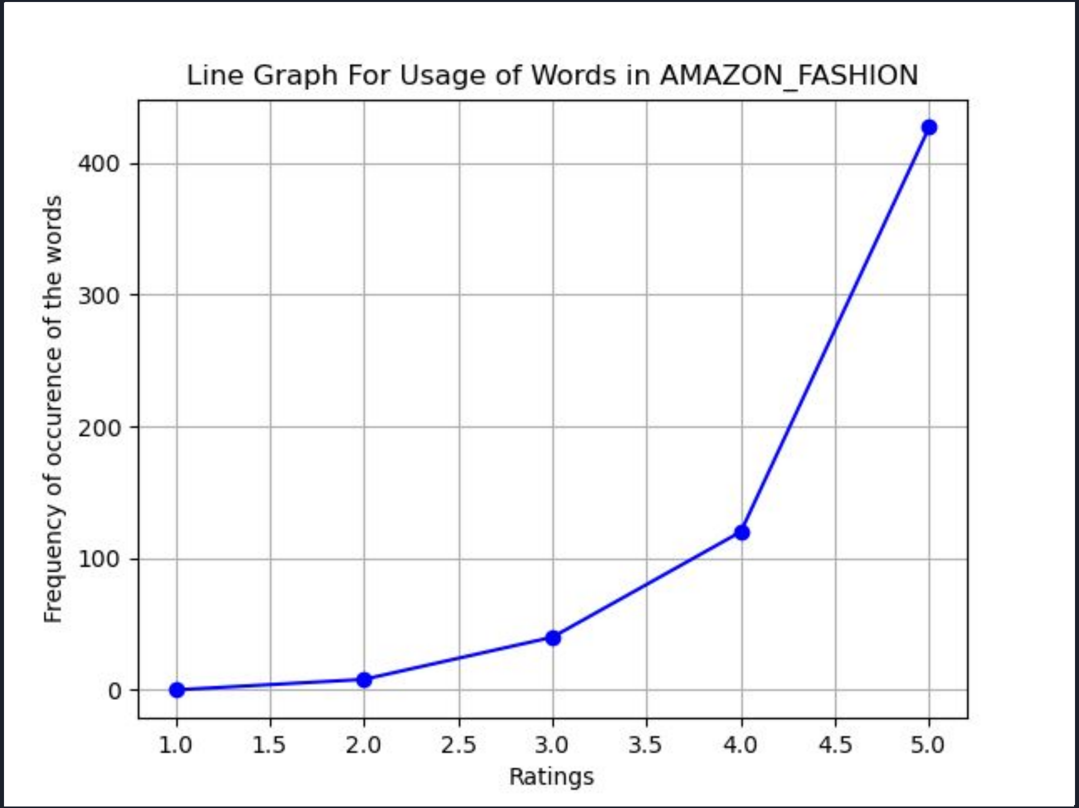




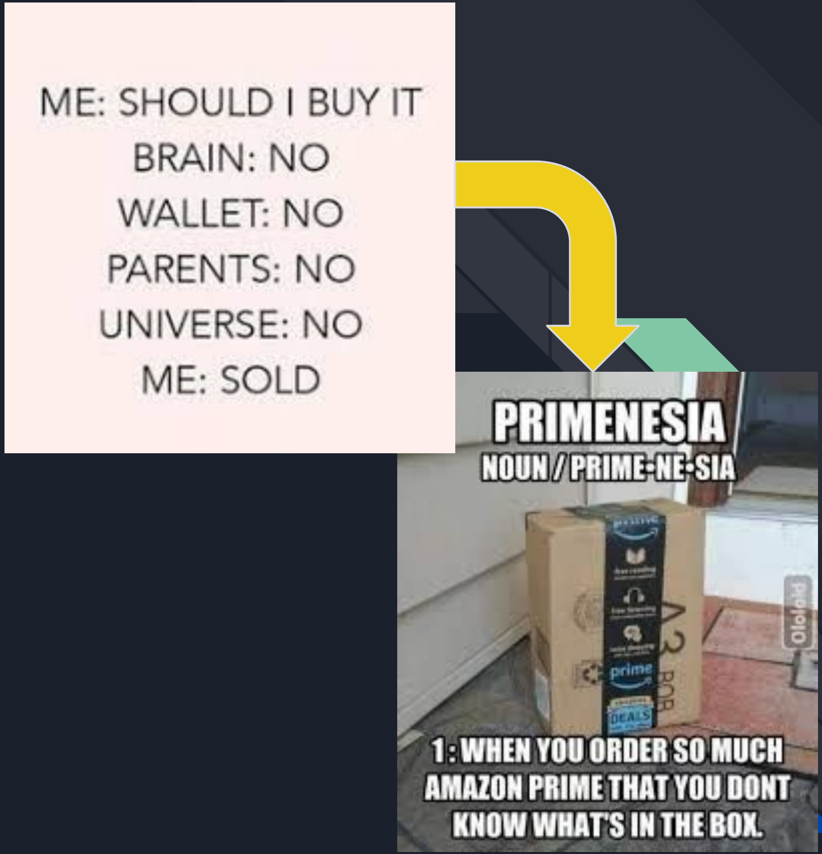




Usage of the word ‘comfortable’ in reviews in fashion category



Fin.



ME: SHOULD I BUY IT
BRAIN: NO
WALLET: NO
PARENTS: NO
UNIVERSE: NO
ME: SOLD

PRIMENESIA
NOUN / PRIME-NE-SIA
1 : WHEN YOU ORDER SO MUCH AMAZON PRIME THAT YOU DONT KNOW WHAT'S IN THE BOX.
prime
Ololold