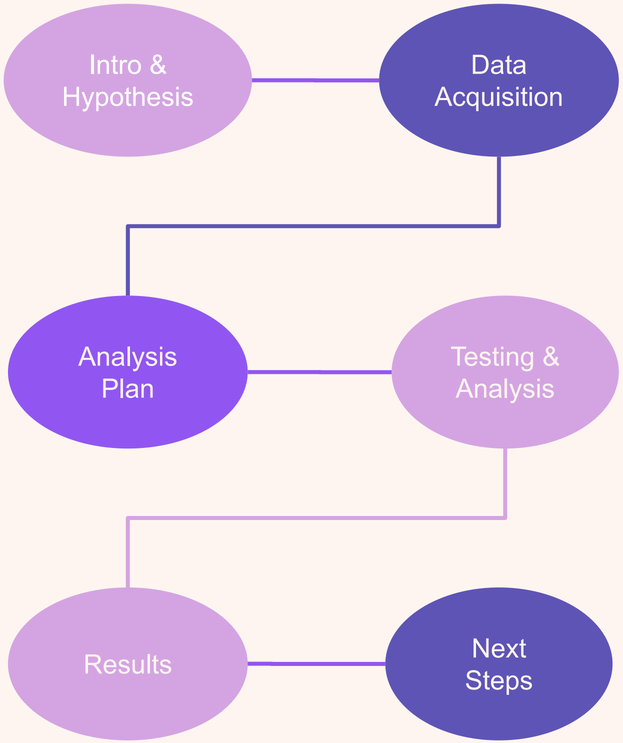
Investigating Online Review Readability as a Predictor of Helpfulness Ratings

DS 4002
9/29/24

Group 8: Grace Brasselle (Leader), Kristy Luk, & Isabel O'Connor

Presentation Outline



Motivation

Amazon is one of the largest e-commerce companies, and they are an important metric within today's online marketplace, signaling product quality to consumers and alerting sellers to possible areas of self-improvement.

According to the Pew Research Center [1], 40% of US adults check online reviews before purchasing a product for the first time.

Goal: Analyze whether the readability index, based on the Flesch Reading Ease scale, of Amazon food influences how helpful other users perceive the review.

Research Question

Do Amazon food reviews with higher reliability scores more likely to be rated as “helpful” by other internet users?

Modeling Approach

In order to test this hypothesis, we will generate Flesch-Kincaid reading ease scores for Amazon reviews within a public dataset. Subsequently, we will conduct a regression analysis mapping the correlation between Flesch Kincaid scores and helpfulness ratings. Flesch Kincaid scores will be generated using the “Readability” python package [2].

Flesch-Kincaid Scale

Reading Ease	Grade	Description of Style	Syllables per 100 Words	Av. Sentence Length
90–100	5	Very easy	123	8
80-90	6	Easy	131	11
70–80	7	Fairly easy	139	14
60–70	8–9	Standard	147	17
50–60	10–12	Fairly difficult	155	21
30–50	College	Difficult	167	25
0–30	College graduate	Very difficult	192	29

Data Acquisition/Explanation

Data Dictionary of Original Dataset

Column	Description	Potential Responses
Id	Row Id	3420
ProductId	10-character string with a random combination of letters and numbers used to identify a product	B005K4Q1VI
UserId	14-character string with a random combination of letters and numbers used to identify a user	A1ISS4F3IRVTS0
ProfileName	The username created by the user, can be any length and may contain letters, numbers, spaces, or other special characters	Morton Family
HelpfulnessNumerator	Integer that indicates the amount of users who marked a review was helpful	9
HelpfulnessDenominator	Integer that represents the amount of users who indicated the helpfulness of a review	11
Score	An integer between 1 and 5 symbolizing the user's rating of a product (like star ratings)	2
Time	A timestamp that indicates the date and time at which the review was posted	1322179200
Summary	A short synopsis of the review contents	"Why does this have sucralose?"
Text	The entire text of the review	"My whole family (7 people) tasted all 3 flavors and they are sickly sweet. Very disappointed that product has sucralose. I wouldn't buy it again."

- The "Amazon Fine Food Reviews" text dataset is sourced from Kaggle, containing over 500,000 food reviews which were documented on the Amazon website between October 1999 and October 2012 [3]
- After loading the dataset, we identified the primary variables of interest pertaining to each review, including:
  - "ProductId": string of unique characters identifying each product
  - "HelpfulnessNumerator": total number of users who marked review as helpful
  - "HelpfulnessDenominator": total number of users who left a helpfulness rating
  - "Score": average products rating by users, between 1-
  - "Text": product review text
- We also added a column for the word count of the review, as well as the FRE values generated using the Readability package in Python [4]

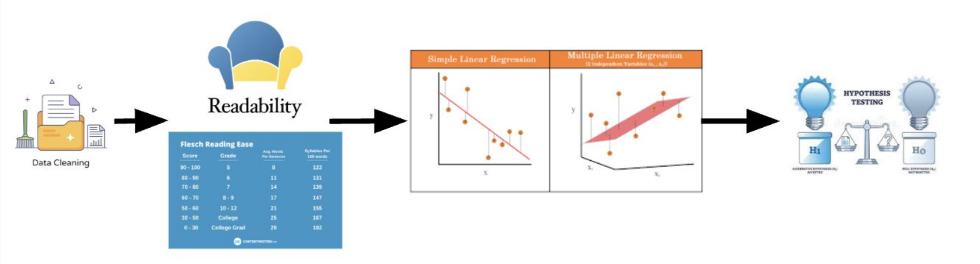
Analysis Plan

After reviewing the raw data set more and conducting an EDA, we noticed the text data we chose was very diverse and cleaning the data was very difficult.

Hypothesis: Reviews receive more helpful ratings when the readability score of the review is higher

Phase 1: Use Readability Python package to compute readability FRE scores. Perform simple linear regression of the Flesch Reading Ease Score on the HelpfulnessNumerator for reviews. Calculate and analyze the R^2 value to determine the strength of the relationship.

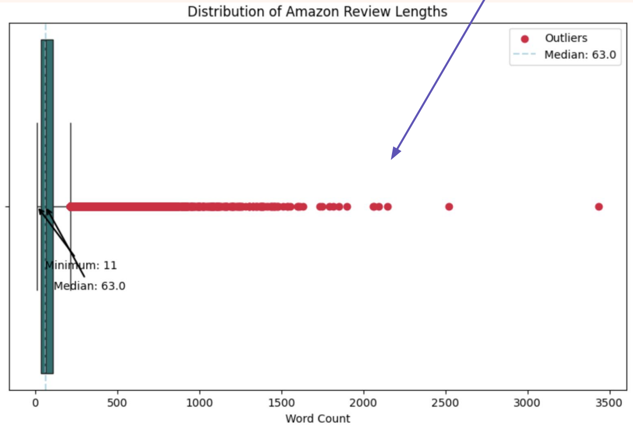
Phase 2: Conduct multilinear regression to assess the combined effects of review length and readability index on perceived helpfulness. The R^2 value, again, will be calculated in order to examine the relationship between the variables.



Tricky Analysis Decisions

the outliers
were skewing
our data

- Our data consisted of a lot of outliers
  - Some reviews contained a word count of over 3,000, which was much larger than the median of 32 words; we decided to retain these **outliers** within the dataset
- Some of the Flesch-Kincaid scores generated were negative
  - Through closer inspection, we discovered that many reviews with negative readability scores contained **symbols** or **“irregular”** sentence structures; we decided to retain reviews with negative readability values within the dataset
- Helpfulness ratio vs. helpfulness numerator
  - A helpfulness ratio column was first created to account for the amount of total helpfulness indication
    - Thought about normalizing the data
  - We ultimately decided to use the **helpfulness numerator** for our analysis due to simplicity



removed reviews like
this

Example: “The mouth says, "How do I love thee, let me count the ways..."**
**If you like apple products a must have item. The only drawback, shipping cost. These are very heavy.”

Bias and Uncertainty Validation

During the cleaning process, we found that a lot of reviews had hyperlinks/breaks, html code, or other odd text that was significantly skewing the FRE scores.

We decided to modify these rows in our dataset, however, this could have introduced bias into the study because we only ultimately modified these reviews because they were were affecting one of our variables of interest, the FRE scores.

In order to mitigate some of this bias, we conducted linear regression models before and after scaling the data using arcsinh transformation to provide insight of how the regression differs with the raw data and scaled data [5].

Results and Conclusions

Phase 1:

- Very weak negative correlation between readability scores and the number of people who rated the review as helpful
- Very low predictive power, where only 3.7% of the variation in the outcome can be explained by the IVs in the model

Phase 2:

- Very high SSE value indicating that our model may not be the best for our data
- 4.04% of the variation in the outcome can be explained by review length and perceived helpfulness

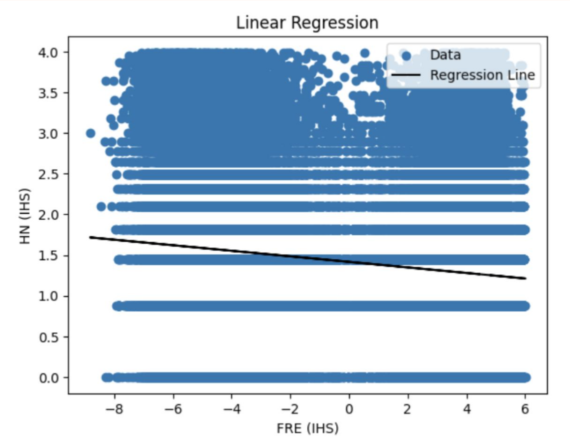
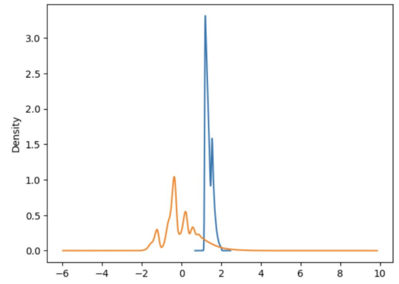


Table of Values from SLR

	b0	b1	rsq
0	1.459383	-0.041568	0.03744

Density Plot of Predicted Values (Blue) & Residuals (Orange)



MLR Stats table

	b	sse	rsq
0	[1.7886524254278842, -0.006794718957230515, -0...	156040.522711	0.040384

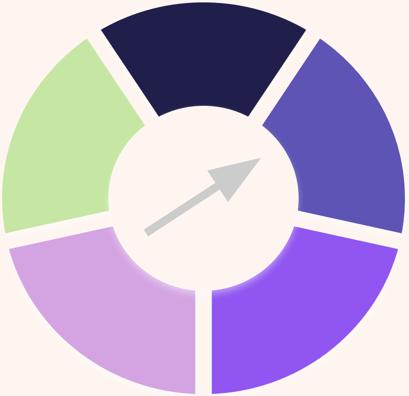
Next Steps

Improvements

- Further clean the dataset
- Research possibly a better readability index tool
- Add more meaningful variables/transformations

New Lines of Exploration

- Explore correlation between review score and readability index
- Analyze whether negative, neutral, or positive reviews tend to have lower/higher readability
- Examine readability on other reviews like clothes or electronics



New Questions

- What are common words that appear in one star vs. 5 star reviews?
- How can the readability index scores vary depending on product? Do niche products have higher FRE scores?
- How does sentiment play a role in readability scores?

References and Acknowledgements

[1] A. Smith, "Online Shopping and E-Commerce," Pew Research Center, https://www.pewresearch.org/internet/2016/12/19/online-reviews/ (accessed Sep. 13, 2024).

[2] "Readability," PyPI, https://pypi.org/project/readability/ (accessed Sep. 13, 2024).

[3] S. N. A. Project, "Amazon Fine Food Reviews," Kaggle, https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews (accessed Sep. 13, 2024).

[4] "Flesch reading ease and the Flesch Kincaid grade level," Readable, https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/ (accessed Sep. 13, 2024).

[5] "Asinh Demo," matplotlib, https://matplotlib.org/stable/gallery/scales/asinh_demo.html/ (accessed Sep. 27, 2024).

https://github.com/gkbrasselle/DS4002Project1

Thank you to Professor Alonzi and our TA Layla for their assistance on this project!

Thank you!

* beware of fake reviews *