Project 1

STQD6114 – Unstructured Data Analytics

P152419 – Hazim Fitri Bin Ahmad Faudzi

Part 1 – Task 3

INTRODUCTION

The Amazon Fine Food Reviews from Kaggle has been selected for this analysis. This data is a food product review from Amazon starting from October 1999 to October 2012. It has a total of 568,454 reviews which is contributed by 256,059 users reviewing 74,258 different products. Each review includes the user ID, rating score (1-5), helpfulness vote counts (upvotes/downvotes), review and a timestamp. Our goal is to analyse the sentiment score and emotion of users from Amazon. Since this data is too big and can't be uploaded into UKMFolio, I'll provide link to both google drive and Kaggle as source of data. The link to the Kaggle is https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews while the link to the google drive is https://drive.google.com/drive/folders/1BvJ4SHkWh6zjTtQxLqGorkaqQhuQfCC7?usp=sharing
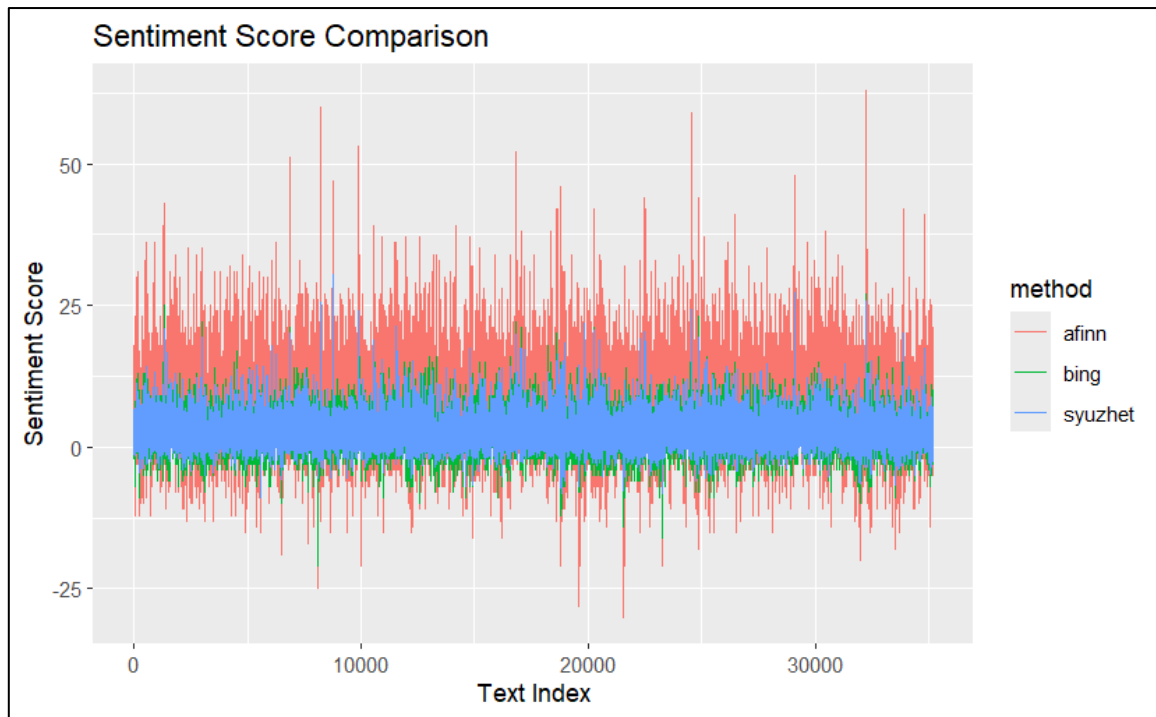
DISCUSSION

Figure 1 Sentiment Score Comparison

The line-chart comparison reveals that the three lexicons, while broadly agreeing on which reviews are more or less positive, differ greatly in scale and sensitivity. AFINN (red) produces the largest swings in score, reflecting its use of integer word scores (–5 to +5) summed across a review. Bing (green) stays clustered near zero because it simply counts positive vs negative words. Any review with roughly equal positive and negative terms yields a near-zero score. Syuzhet (blue), which uses the NRC lexicon, falls between these methods and it also effectively counts positive minus negative emotion words but does so on a smaller scale than AFINN. In effect, AFINN's seems "more dramatic" while Bing's binary seems almost flat baseline. This behaviour matches prior observations: AFINN has relatively few words but assigns them weighted values, while Bing's simple binary labels result in less gradation.
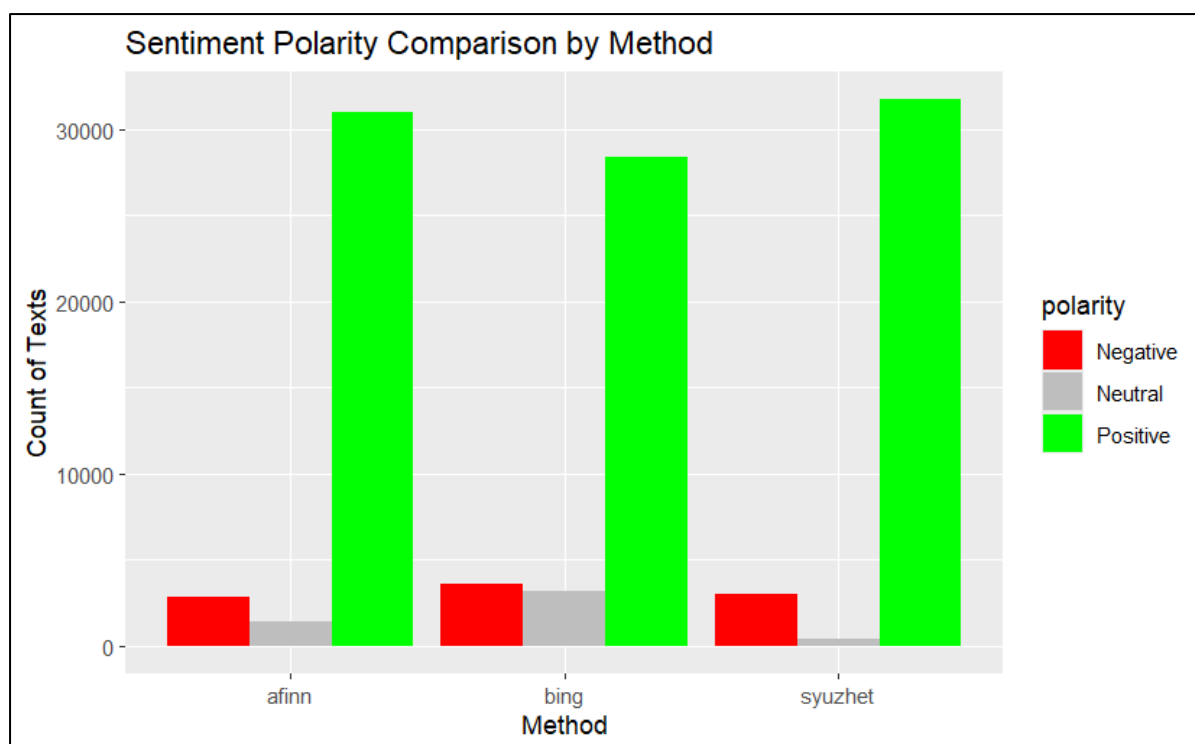
Figure 2 Sentiment Polarity Comparison

When we categorize each review's overall sentiment (positive/negative/neutral), the methods shows distinction. All three mark the vast majority of reviews as positive (green bars), reflecting the generally favourable tone in the Amazon food data. However, Bing identifies noticeably more negative and neutral reviews than the other methods. Bing's lexicon contains many negative terms, so any review with even a few critical words often tips the balance to "negative." AFINN and Syuzhet, by contrast, found slightly more positives. Conversely, Bing produces a larger neutral category than AFINN or Syuzhet. In our analysis, Syuzhet produced very small number of neutrals. In sum, Bing tends to split off more negatives/ neutrals, whereas AFINN and Syuzhet usually label a larger share of texts as positive. These differences arise directly from the lexicons: Bing is strictly binary, so neutral cases occur only with perfect word-balance, while AFINN's numeric sum can more easily be nonzero, and NRC/Syuzhet also rarely yields exact ties.
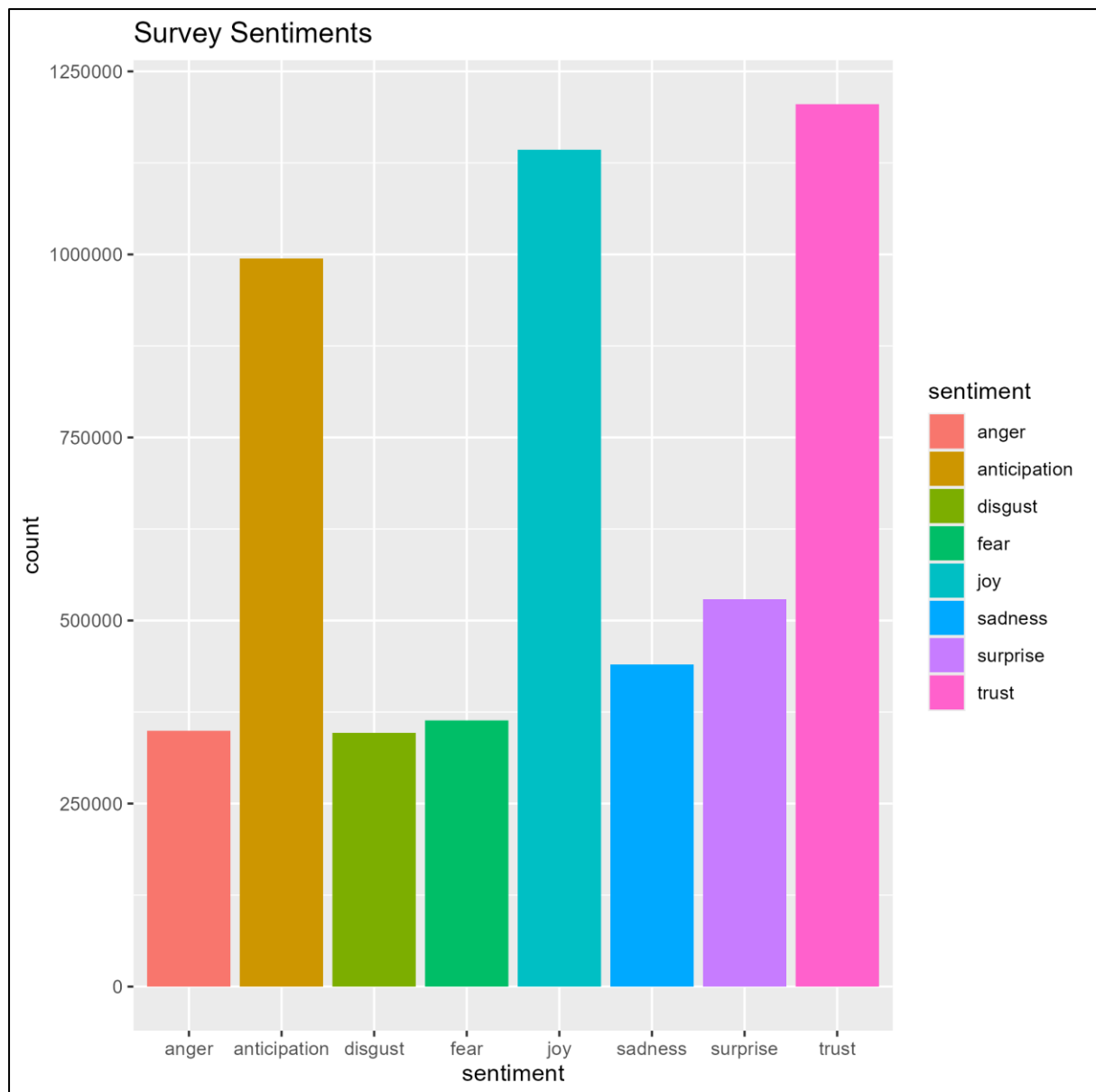
Figure 3 Emotion Scores

The NRC-based emotion breakdown via Syuzhet offers a richer view of customer feelings. Here trust (pink) and joy (cyan) dominate, far exceeding negative emotions. Anticipation (orange) and surprise (purple) are also prominent, while anger, fear, and disgust (red, green, olive) are comparatively rare. This pattern strongly suggests that reviews mostly convey positive sentiment as shown by customers express confidence and satisfaction (trust/joy) much more than frustration. Research confirms that joy and surprise in reviews correlate with higher satisfaction, whereas disgust and fear correlate with lower satisfaction. Our results shows that the high joy/surprise counts and low disgust/fear counts imply a generally favourable customer experience. For example, customers often describe delicious meals or good service which induce joy and report few truly upsetting issues, hence little disgust or anger. The prevalence

of anticipation indicates excitement or positive expectation, further reflecting engaged customers. In summary, the emotion chart reveals that positive-affect words (trust, joy, anticipation, surprise) greatly outnumber negative-affect words, underscoring that the overall tone of these reviews is upbeat and reassuring.

CONCLUSION

In conclusion, the comparative analysis of the Amazon Fine Food Reviews dataset through multiple lexicons demonstrates the value of a multi-faceted approach. By combining count-based polarity measures, intensity scores and nuanced emotion categories, it becomes possible to derive actionable insights. Companies may thereby focus efforts on improving products that inspire fear or disgust, while celebrating and leveraging feedback that fosters trust and joy. Such balanced analysis enhances decision making and ultimately supports better customer satisfaction and loyalty.