**Figures included in the narrative PDF**

*Figure 1. Features of Amazon Dataset*

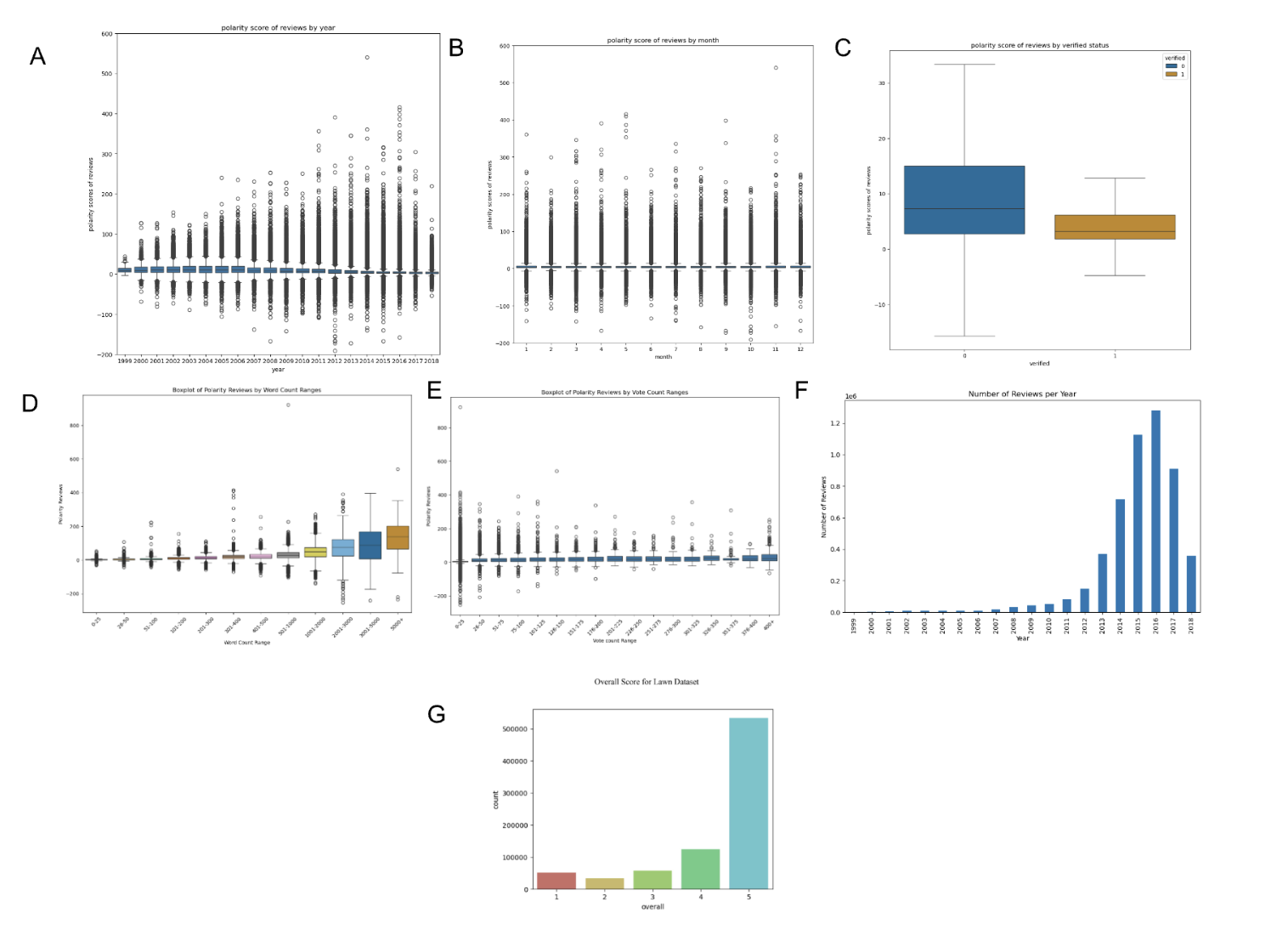
| **Feature** | **Description** | **Feature Type** | **Storage type** |
| --- | --- | --- | --- |
| overall | Rating of the product, discrete values from 1 to 5 | Quantitative- discrete | Double |
| vote | Number of helpful votes on the review | Quantitative- discrete | Int |
| verified | Whether of not the user giving the review has a verified purchase or not | Qualitative - dichotomous | Boolean |
| asin | ID of the product | Qualitative - nominal | Int |
| reviewerName | Username of the reviewer | Qualitative - nominal | String |
| reviewText | Main text portion of the review | Qualitative - nominal | String |
| summary | Summary text of the review | Qualitative - nominal | String |

*Figure 2. Added Features of the Amazon Dataset*

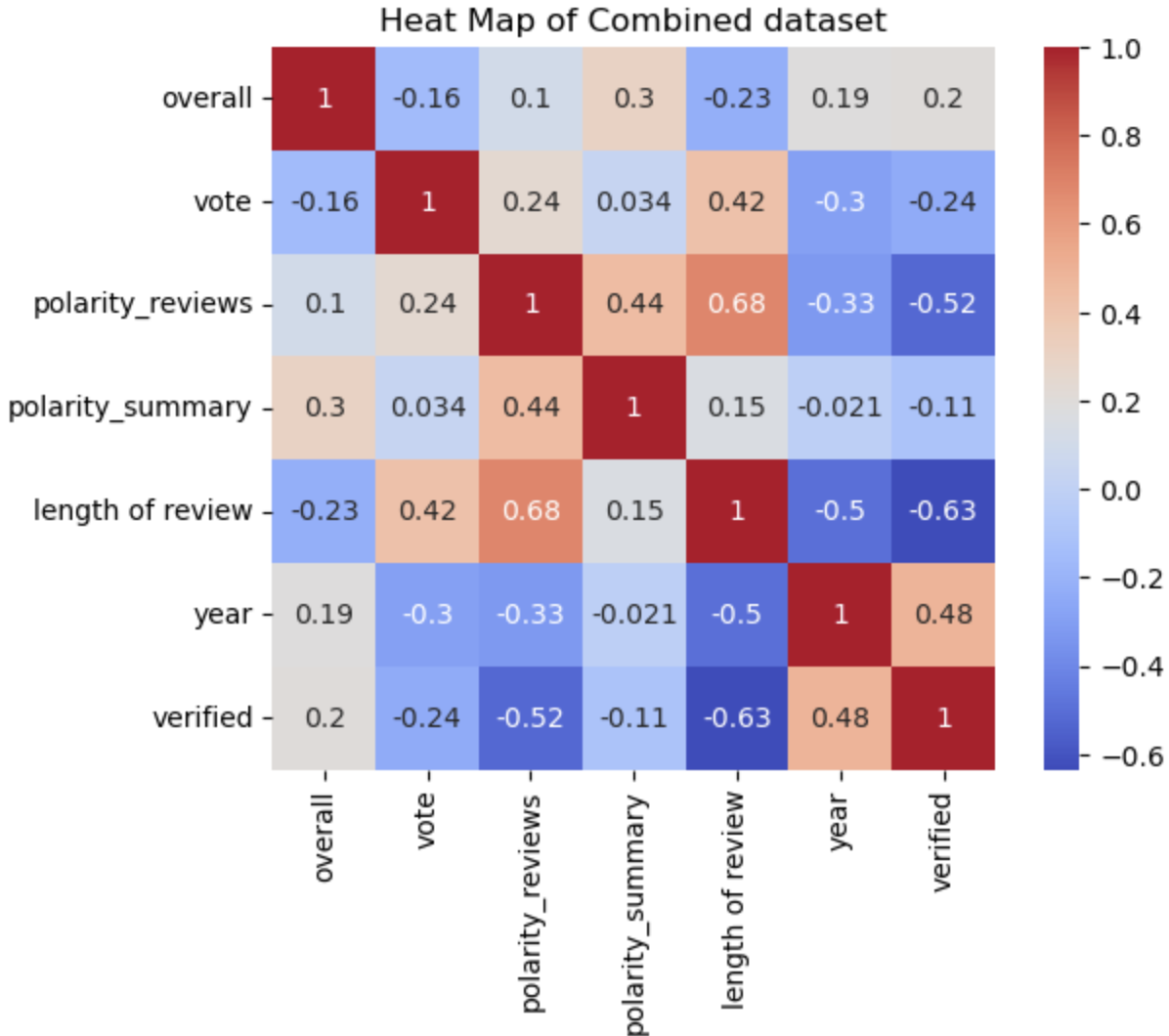
| **Feature** | **Description** | **Feature Type** | **Storage type** |
| --- | --- | --- | --- |
| polarity\_review | The VADER sentiment score of the reviewText | Quantitative - continuous | double |
| length\_of\_review | The number of characters in the reviewText | Quantitative- discrete | double |
| length\_of\_username | The number of characters in the reviewerName | Quantitative- discrete | double |
| count | The number of reviews for each asin | Quantitative - discrete | double |
| percentile | The percentile rank of values in the 'polarity\_reviews' column | Quantitative- continuous | double |

*Figure 3. Selected EDA Visualizations*

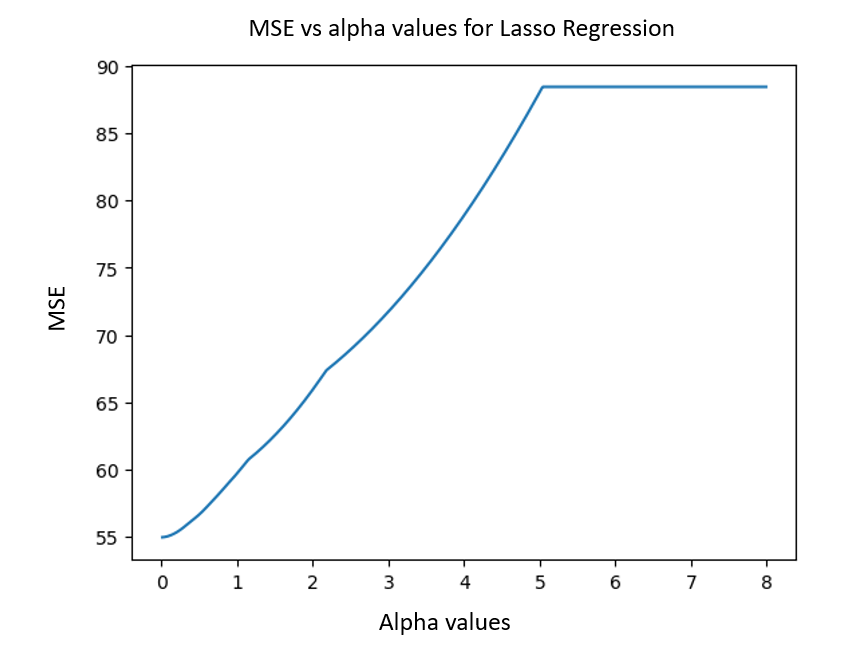
1. *Boxplot - Polarity Score of Review by Year*
2. *Boxplot - Polarity Score of Review by Month*
3. *Boxplot - Polarity Score of Review by Verified Status*
4. *Boxplot - Polarity Score of Review by Word Count Range*
5. *Boxplot - Polarity Score of Review by Vote Count Range*
6. *Barplot - Number of Reviews by Year*
7. *Barplot - Counts of Overall Score*



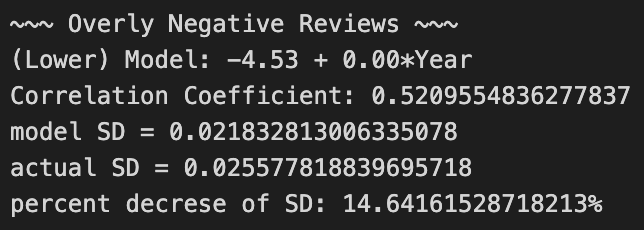
*Figure 4. Heat Map of Features for Model 2*

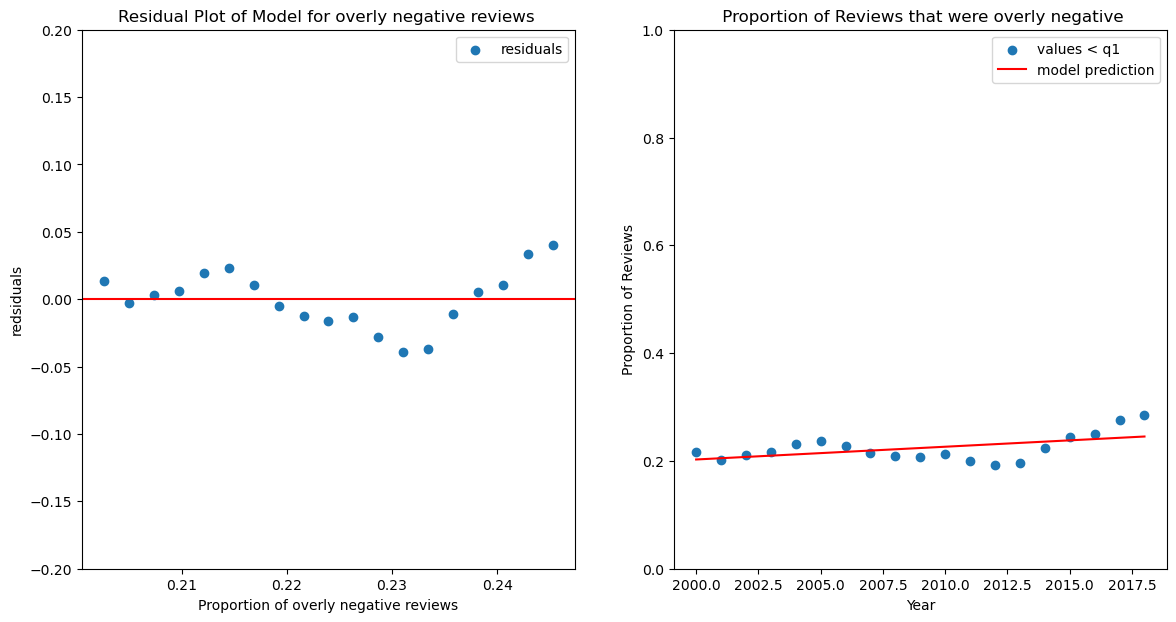


*Figure 5. Mean Squared Error for different alpha values*

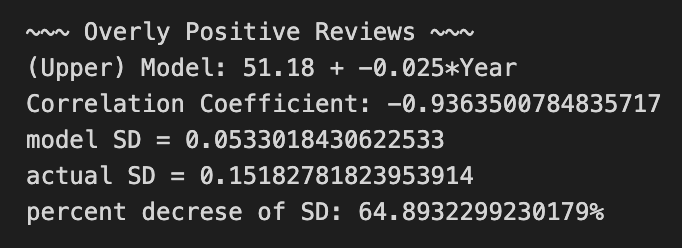


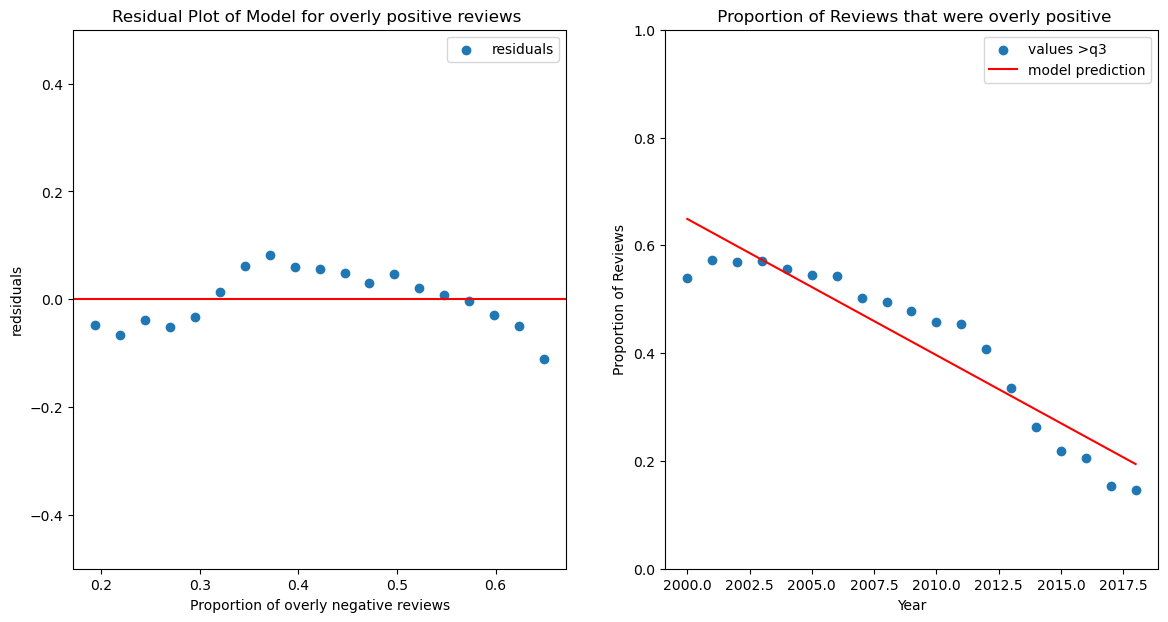
*Figure 6. Model 1 Assessment statistics of overly negative reviews*

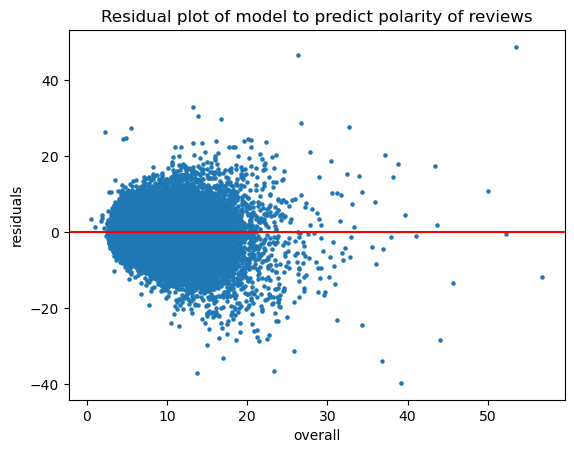
**

*Figure 7. Graphs showing residuals of Model 1 (left) and how the proportion of overly negative reviews changes over time (right)*

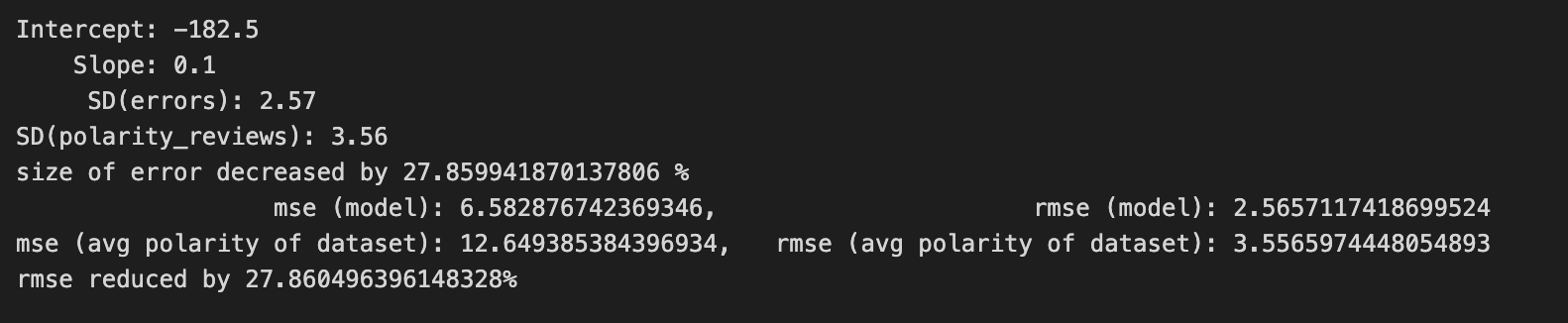
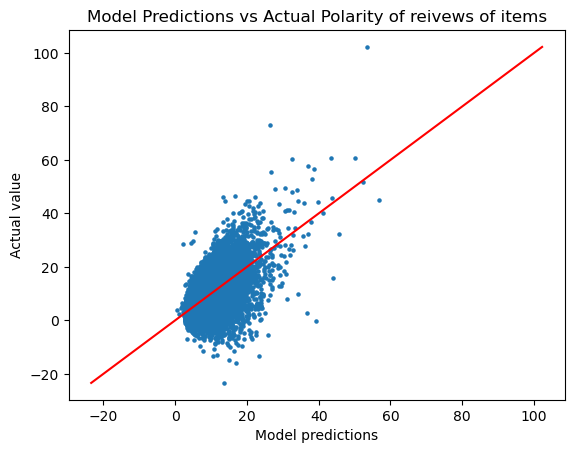
*Figure 8. Model 1 Assessment statistics of overly negative reviews*



*Figure 9. Graphs showing residuals of Model 1 for overly positive reviews (left) and how the proportion of overly positive reviews changes over time (right)*

*Figure 10. Residual plot for Model 2 on the full dataset*

*Figure 11. Results of Model 2: Multiple Linear Model on combined data. (A) SD and RMSE calculations of model and baseline. (B) Plot of actual vs predicted values of polarity scores*

1. *(B)*

*Figure 12. Residual plot for Model 3 on combined data*