

Milestone 3: Questions Answered

Felipe Rodriguez

Bellevue University

DSC 680 Applied Data Science

Professor Amirfarrokh Iranitalab

May 31, 2024

Questions:

1. Positive contains most of the values, is this skewing the model?

While positive is a big factor, it is not skewing the whole model. We will want to include more negative and neutral reviews so that the model can more accurately predict those two elements.

2. Are there any missing values or outliers in the comments?

There were a few missing comments, but they did not impact the model.

3. What is the key takeaway with this project?

Reviews are generally positive in the Boston area. The model also performed well in predicting the sentiment of the reviews. As mentioned, there is some room for improvement for the other two elements, but it is not affecting greatly.

4. What are some of the most common words in the reviews?

Some words that came out a lot were 'great location', 'highly recommend', 'great place', and 'great host'.

5. Would containing more features improve the performance in the model?

In this project, adding more elements wouldn't impact the model since we are only analyzing the comments and the sentiment. If we want to analyze different relationships between sentiment and review, it will allow for that exploration.

6. Can your team aid in model deployment?

This project will involve more web development, our team can aid in this but will need a dedicated deployment team as well.

7. What resources are needed for model deployment?

Primary knowledge on web development and model maintenance.

8. Is there any additional validation work needed for this?

No, during this project we analyze F-1 scores, R-squared and ROC to ensure model performance was acceptable.

9. Can you add more data to the model?

Yes, we can always use more reviews and process those to create a more robust data set that could potentially improve the model.

10. What needs to be done to maintain this model?

This model doesn't need much maintenance since sentiment has been predicted and determined based on the reviews provided. However, if there is a dataset that includes more reviews or different reviews, another model can be created based on that dataset.