Data pre-processing

JSON reviews file from Yelp

Mapped Boston restaurant IDs to Yelp Business IDs

Joined inspection dataset to reviews using key range. All reviews written between inspection data and previous inspection date. Reviews with rating of 4 or 5 labeled as positive reviews, reviews with rating of 3 or less labeled as negative reviews. Removed rows that had no reviews and what appear to be “follow up” reviews. These “follow up” reviews appear to happen approximately one week after the original review and often have the same exact number of violations. It’s not clear if this is some quirk of the data recording process or true results, but would be best modeling by other techniques.