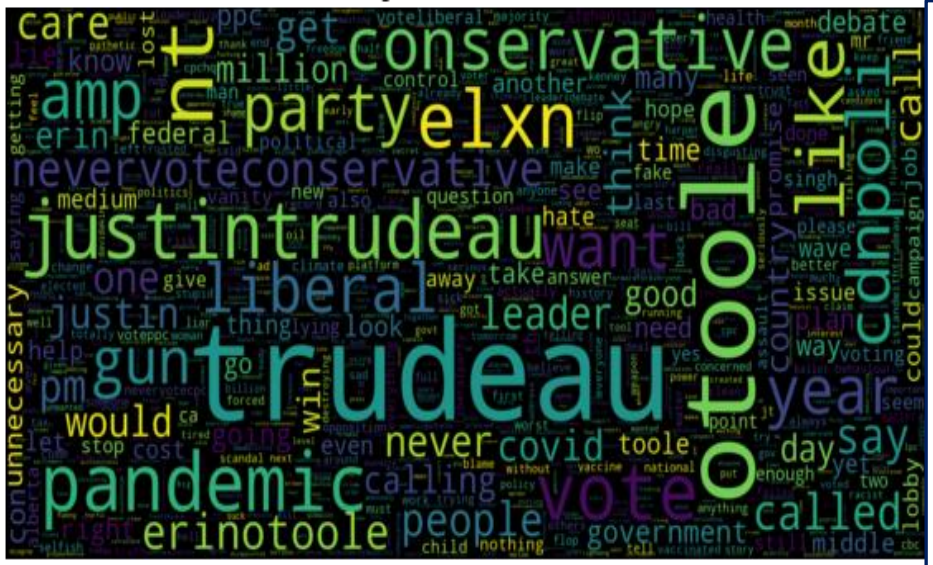
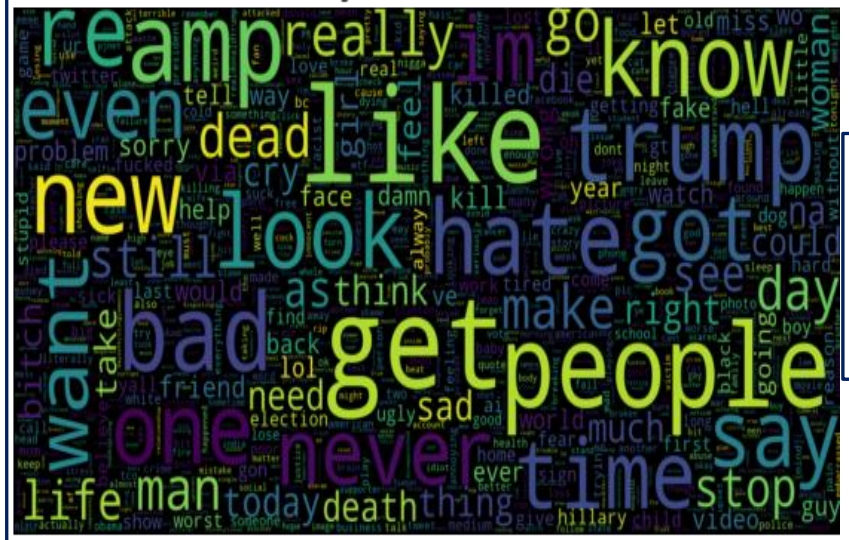
## Key finding from word clouds of elections and generic negative tweets

Word cloud for negative sentiments in elections data




-Two major politicians in negative tweets- Trudeau and Otoole
-Two major parties- Liberal and Conservative.
-Different hatred reasons -pandemic, promises broken, gun violence, lies, controlling attitude, wanting many answers.

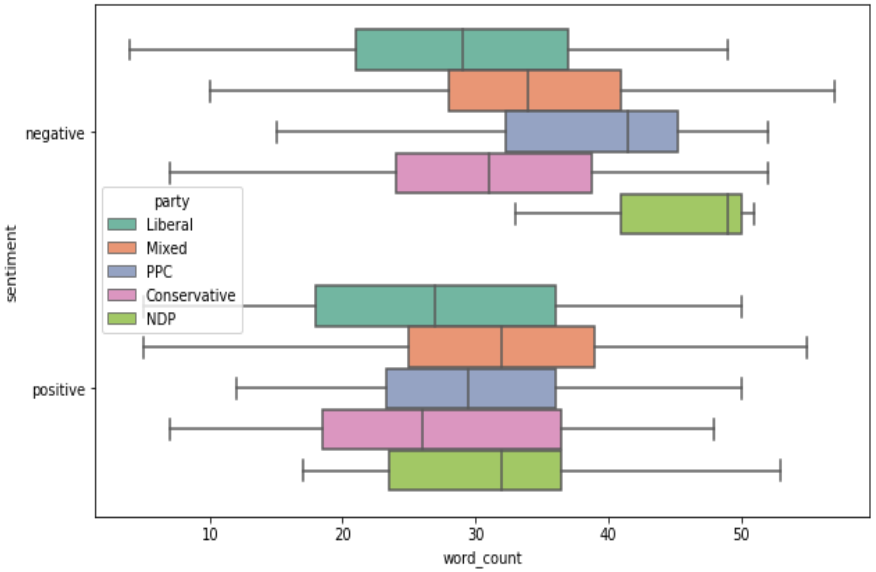
Word cloud for negative sentiments in sentiments data




Dead, bad, never, stop, hate, killed are some of the key words which help model to learn negative data

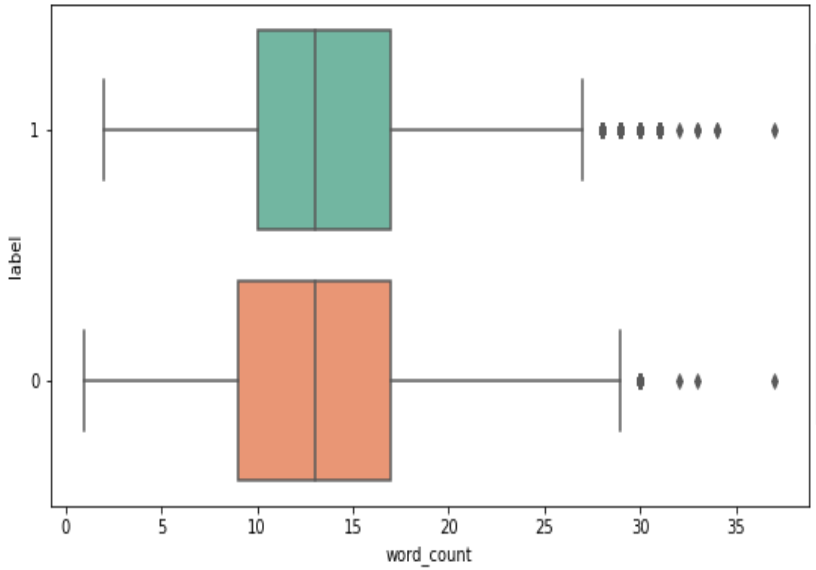
## Key finding from boxplots of length of elections and generic tweets

Word count of positive and negative tweets for each party in elections data




-Boxplot of word count
-In general, negative tweets are longer than positive tweets.
-There is considerable difference between length of negative and positive tweets for PPC and NDP. The difference is the least for Liberal party.

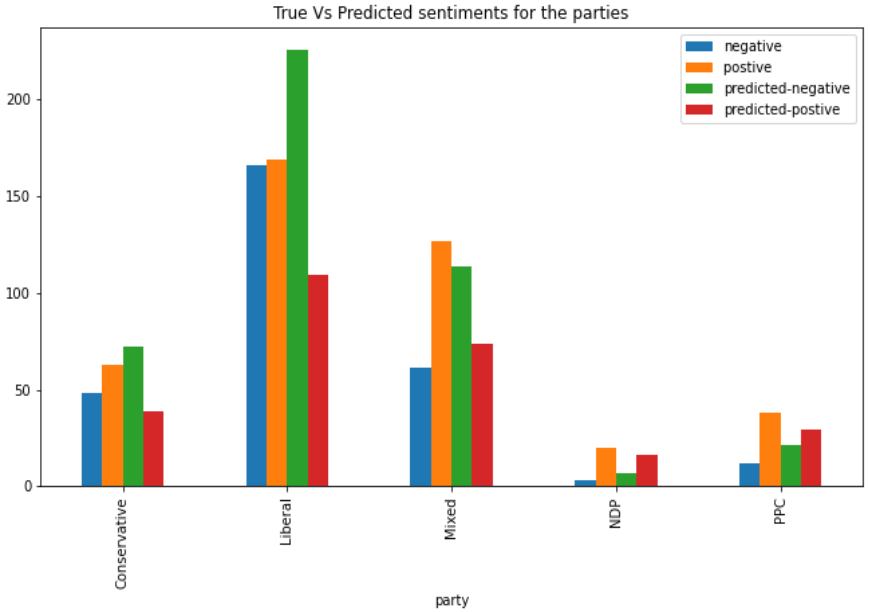
Word count of positive and negative tweets in generic tweets data

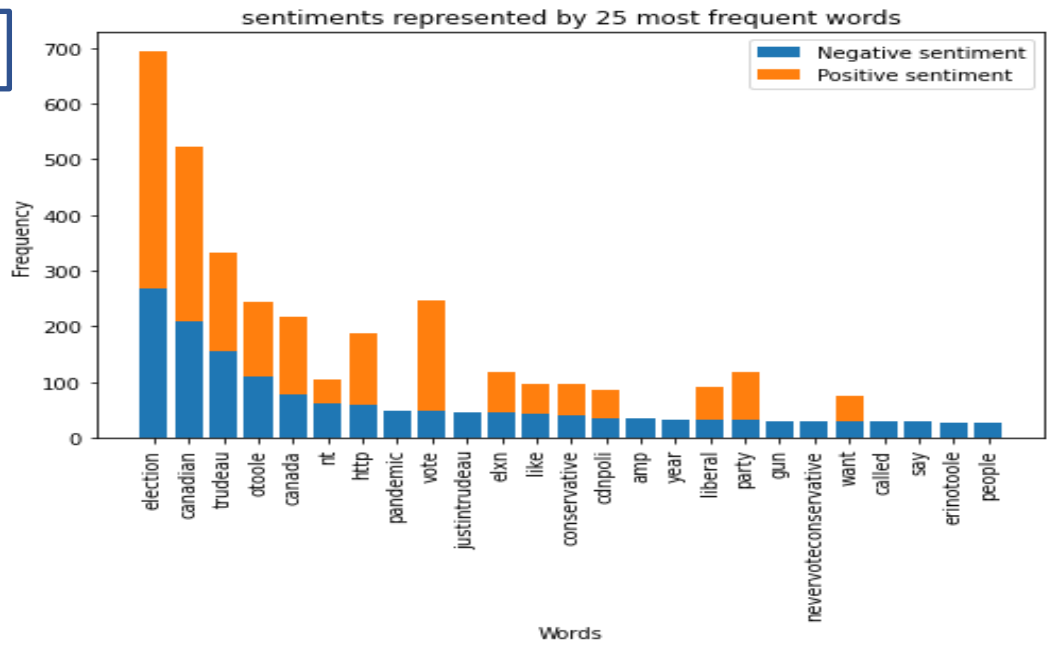



-For generic tweets, longer tweets are more common when sentiment is negative.
-Fewer positive tweets are longer than 25 words.

# Key finding from Sentiment classification

**-Best model: Logistic regression**
**-Model accuracy: 72%**
**-Negative predictions are much higher than actual negatives.**
**Positive predictions are lower than actual positives.**
To increase accuracy: Better feature engineering needed!!




**Many words with high frequency and tfidf scores represent more than one class reducing model accuracy. Soln: better feature engineering techniques like n-grams or using word2vec model**

# Key finding from Negative Reason classification

**-Imbalanced target classes**
**-Logistic model accuracy: 55%**
To increase accuracy: More samples reqd from under-represented classes for balanced targets!!