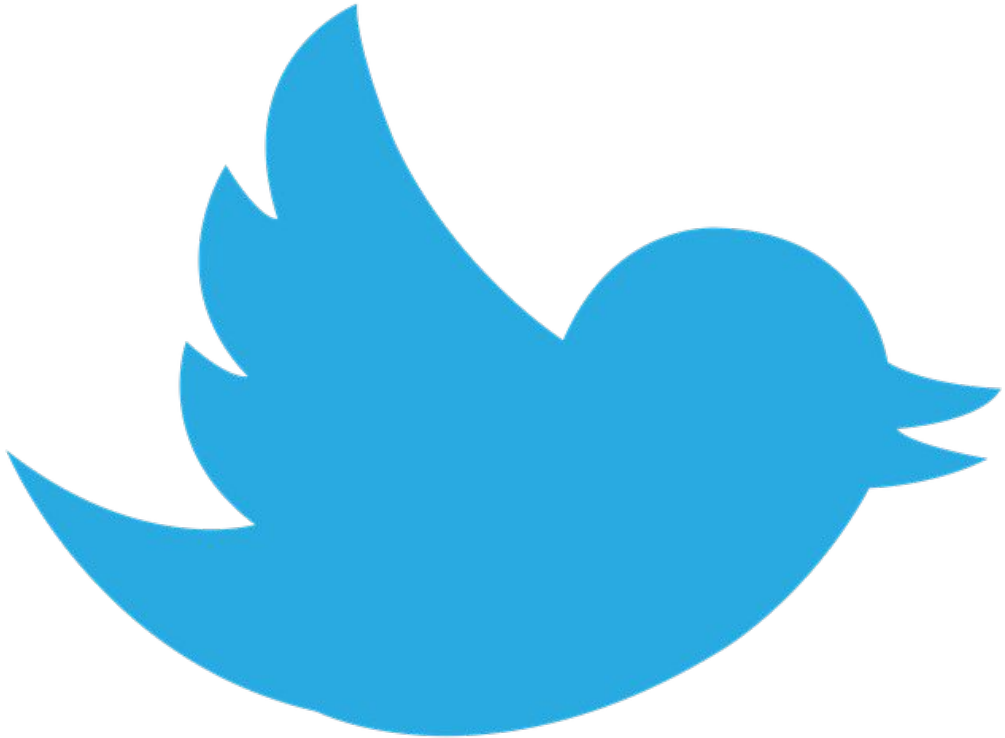
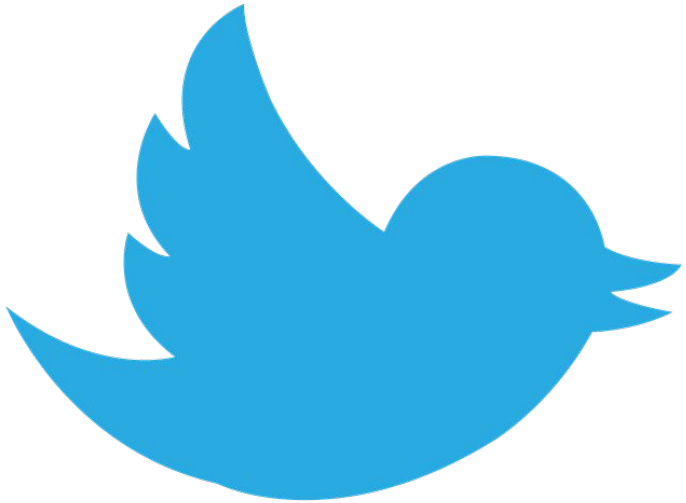
Social Media Sentiments

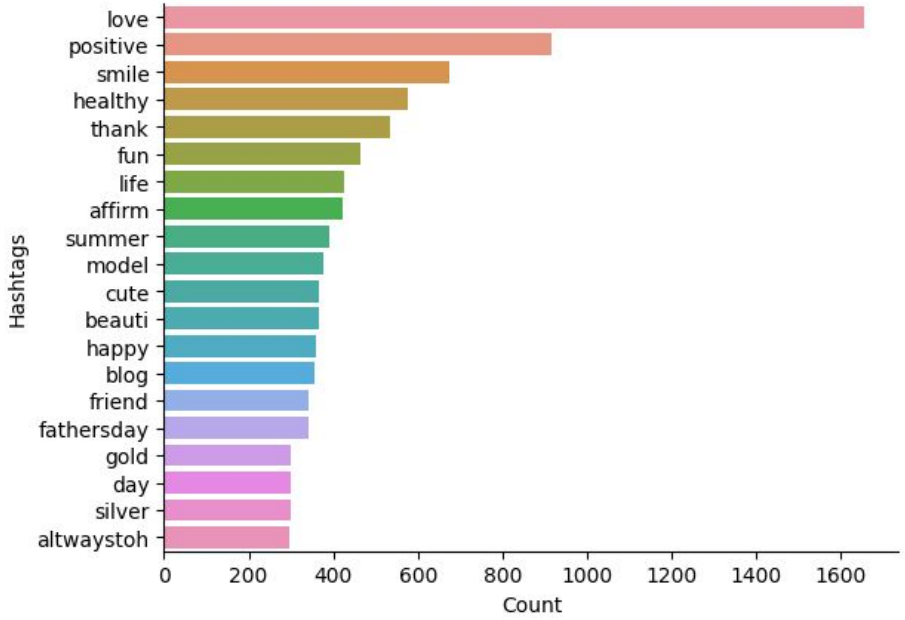




Twitter often is blamed for furmenting ill feelings on its user base. I wanted to know: at any given moment, what is the likelihood for a tweet to be either racist or misogynistic rather than positive and uplifting.

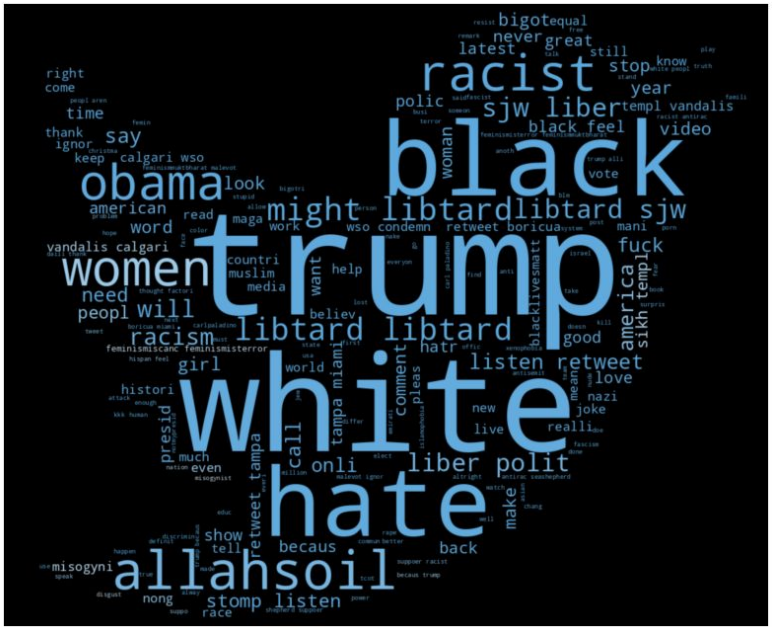
The dataset comprises nearly 32,000 tweets, and what ensues is an in-depth analysis of the sentiment frequencies.

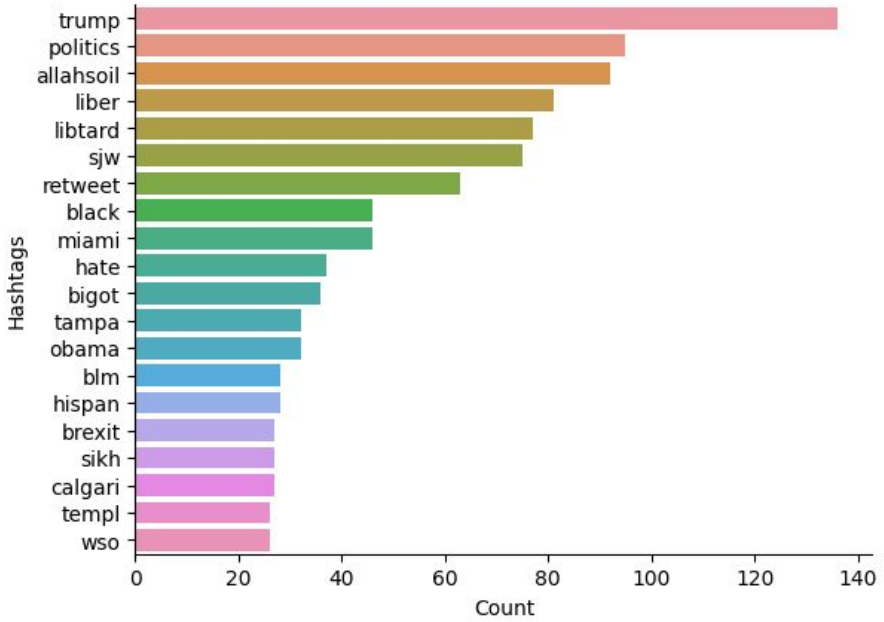
Positive Frequency



It's not unexpected that *love* would top the list as the most commonly used keyword to convey positivity. However, it is remarkable that the word *positive* takes a distant second place.

Negative Frequency

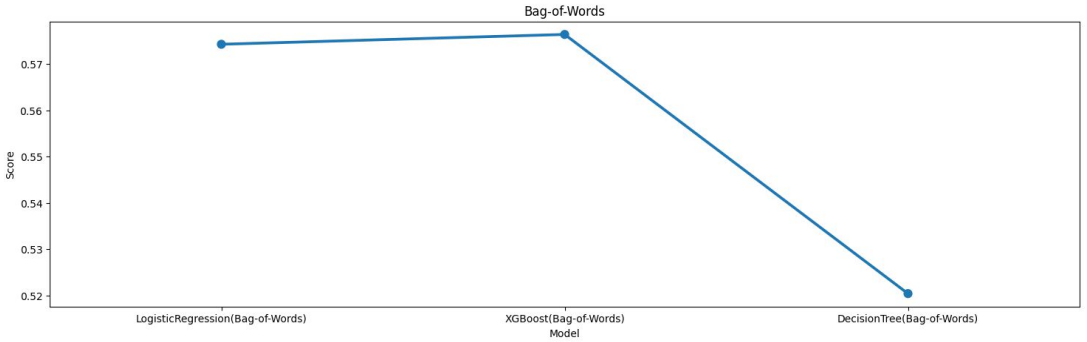


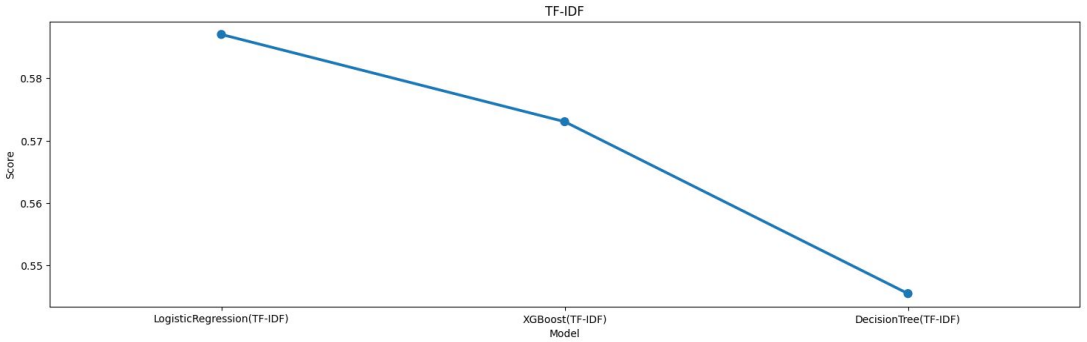


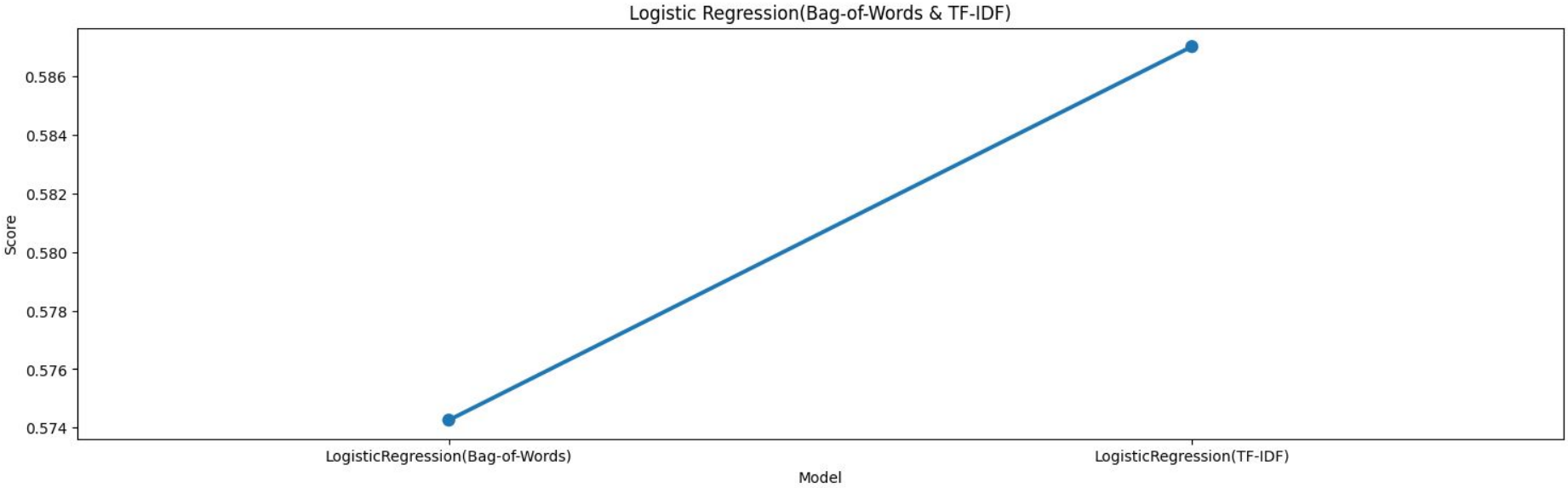
Political keywords abound! With the current heightened political tension, it's unsurprising to find representation of racial tensions between *black* and *white*, along with the most popular political figure of our era.

F1 Score

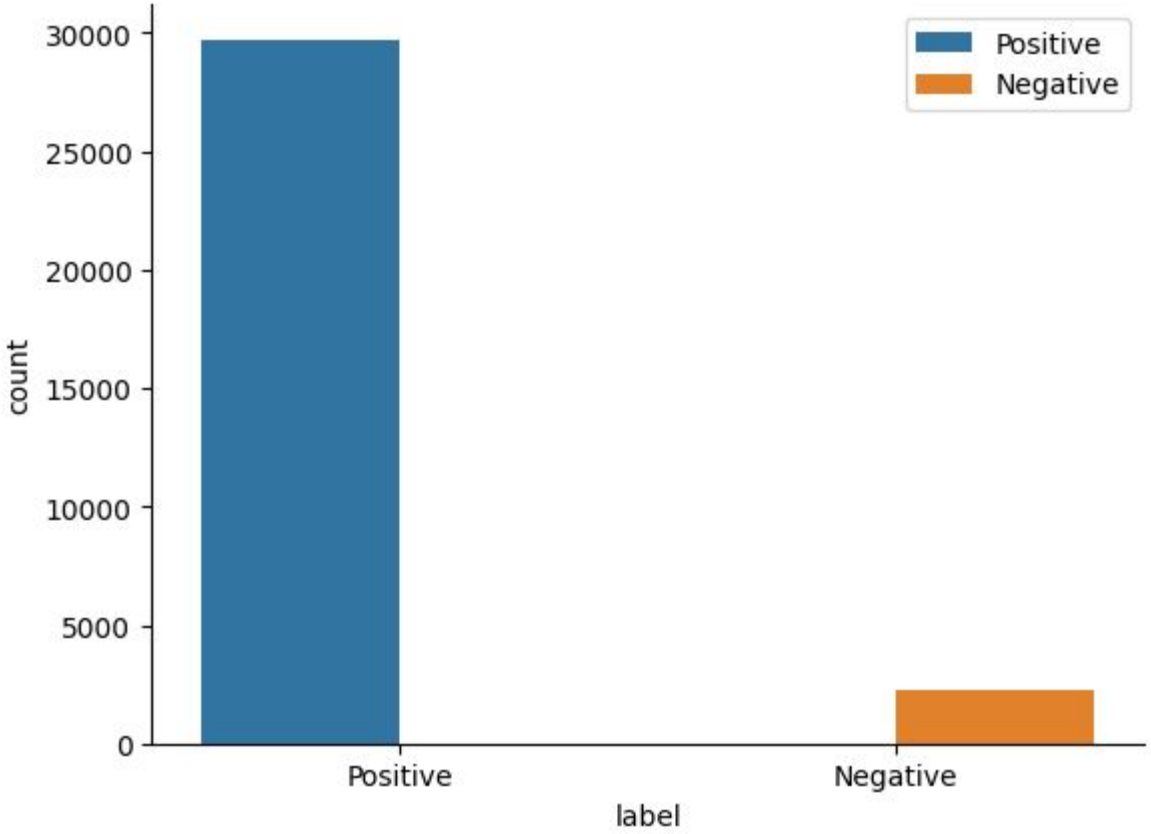
Training algorithms LogisticRegression, XGBoost, and DecisionTree on feature text data Bag-of-Words and TF-IDF. The closest number to 1 indicates the highest level of accuracy.







The LogisticRegression model trained on TF-IDF data is the most accurate and will be employed as a predictor for the test data



It is **1225.602%** more likely for a tweet to be positive without sexist or racist content than negative with sexist or racist content.