Sentiment Analysis

An existing sentiment analyser formulated for Social Media research was used to establish a baseline sentiment for all the available tweet data. The specific analyser (VADER - *Valence Aware Dictionary and sEntiment Reasoner*) is a lexicon and rule-based sentiment analysis tool. The tool was choosen as is specifically attuned to sentiments expressed in social media.

Configuration and code from multiple sources (https://github.com/cjhutto/vaderSentiment (https://github.com/cjhutto/vaderSentiment) and https://github.com/codingupastorm/vadersharp (https://github.com/codingupastorm/vadersharp)) were merged. A number of performance issue where addressed with the available implementations.

The analyser is capability generating scores from multiple perspectives. The **compound** score provides a single unidimensional measure of sentiment for a given sentence. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive).

It can be used standardized thresholds for classifying sentences as either positive, neutral, or negative. Typical threshold values (used with approach) are:

- positive sentiment: compound score >= 0.05
- neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
- negative sentiment: compound score <= -0.05

Specifically, the scoring process:

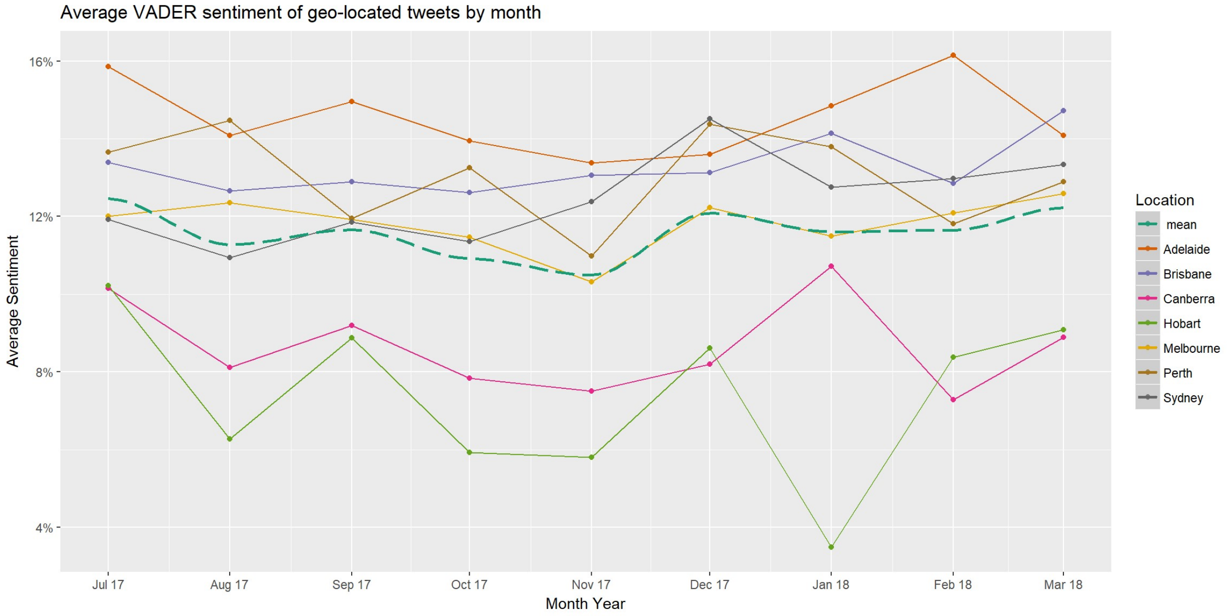
It covers all features we are meant to address:

- scores word for positive & negative associations (e.g. awesome +) (e.g. dickhead -)
- increased sentiment with capitalised words (e.g. I LOVE my dog, I HATE you)
- increased positive sentiment for !!!!!
- understands some double negatives (e.g. not bad)
- understand old smiles :)):< - it has prescribed weighting for each of these combinations.
- understands emojis ❤️💞😊💙 - emojis are converted to word equivalents before sentiment analysis is applied e.g. 😊 becomes *“smiling face with smiling eyes”*

Explorative Statistics

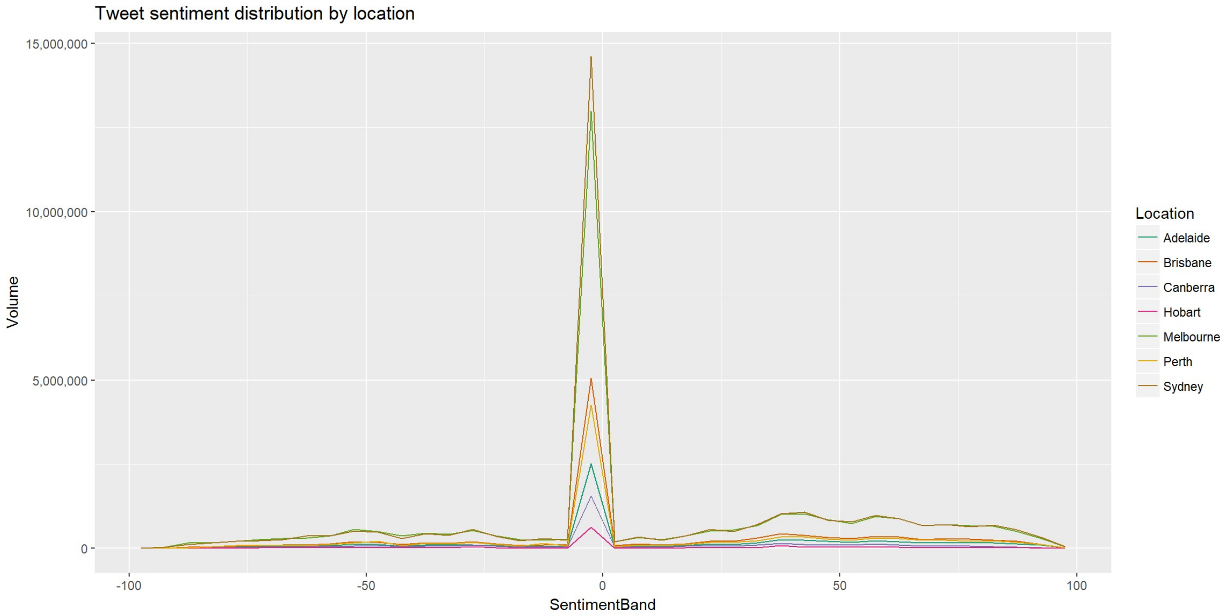
The following is the observed VADER sentiment over 87,028,224 distinct Geo-located tweets.

Location Analysis

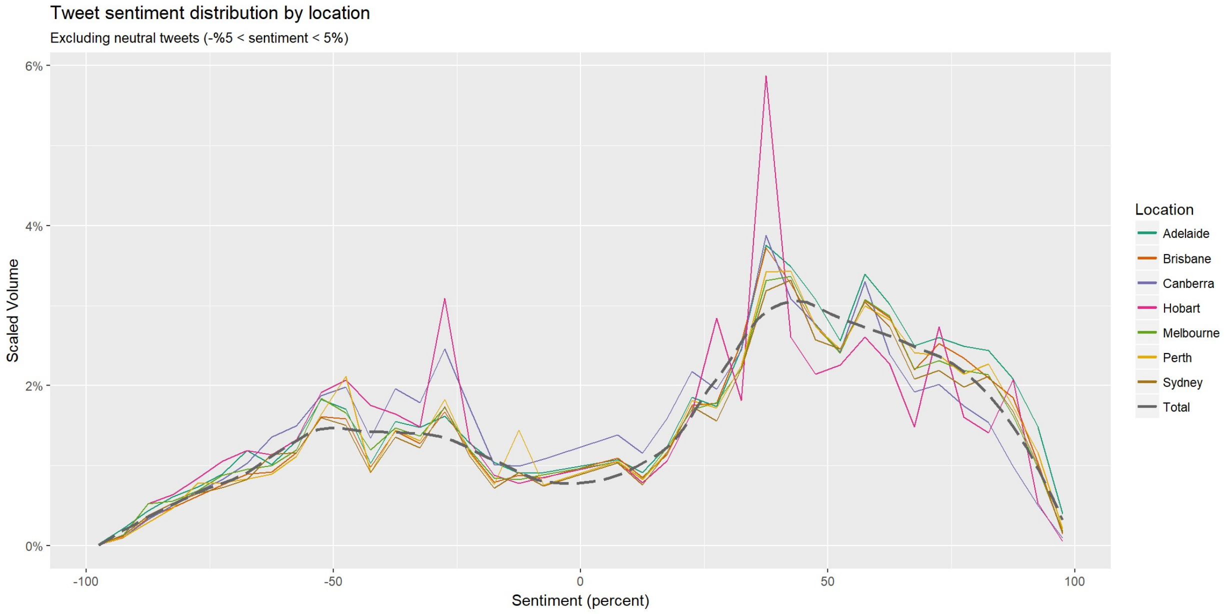


Distribution by Location

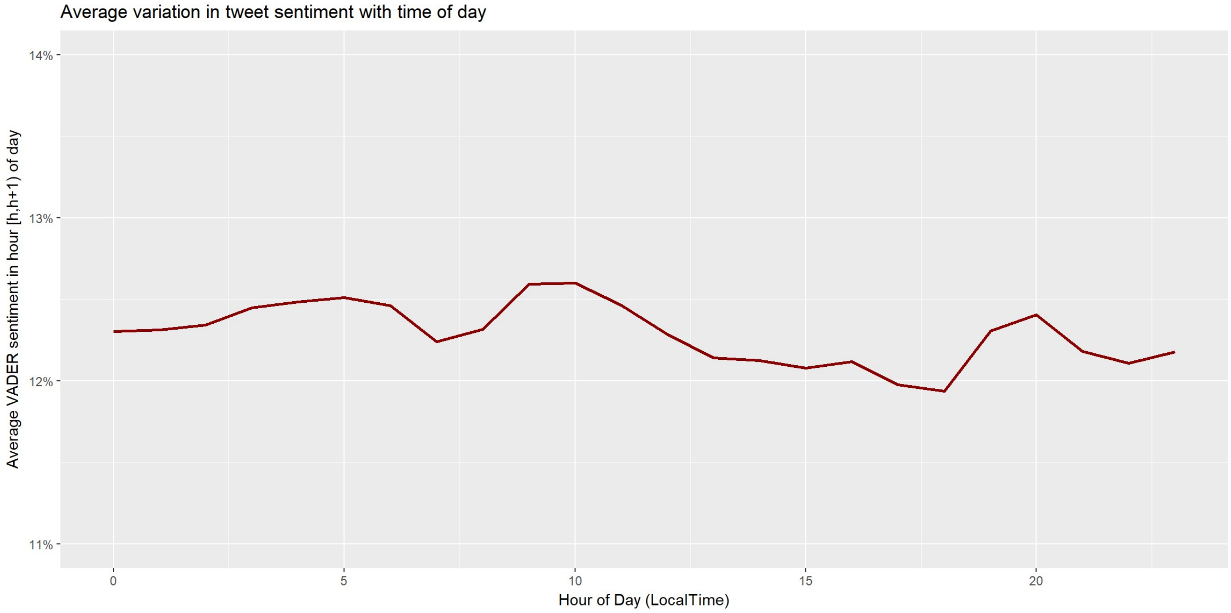
Sentiment analysis is heavily distorted by large volume of neutral tweets.



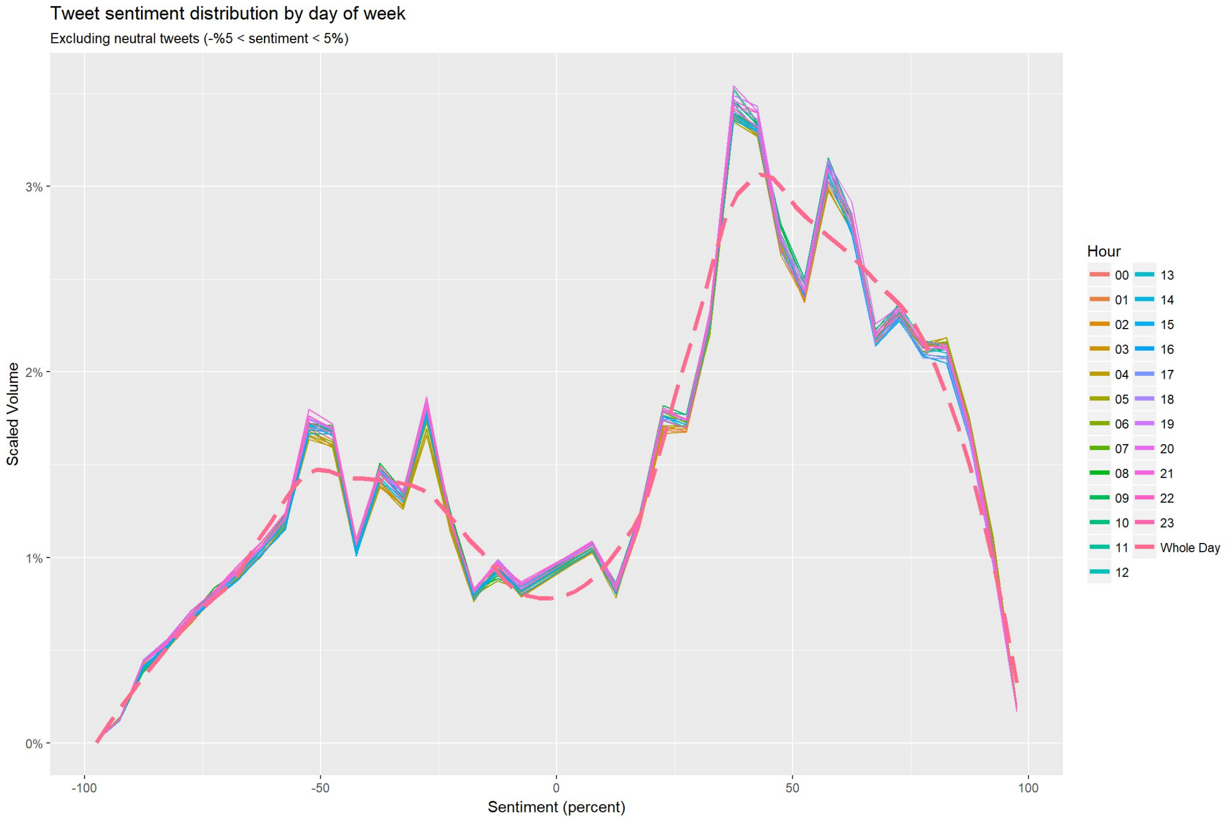
Scaling by total number of tweets per location and excluding central neutral tweets, it’s clear there is a skew in the distribution.



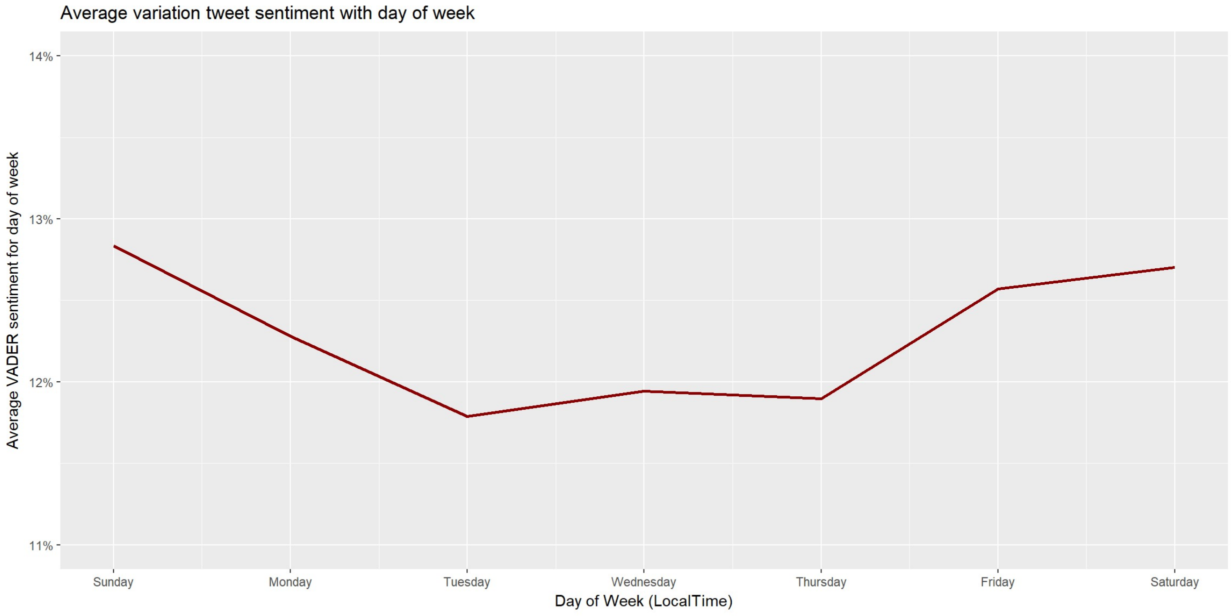
Time of Day Analysis



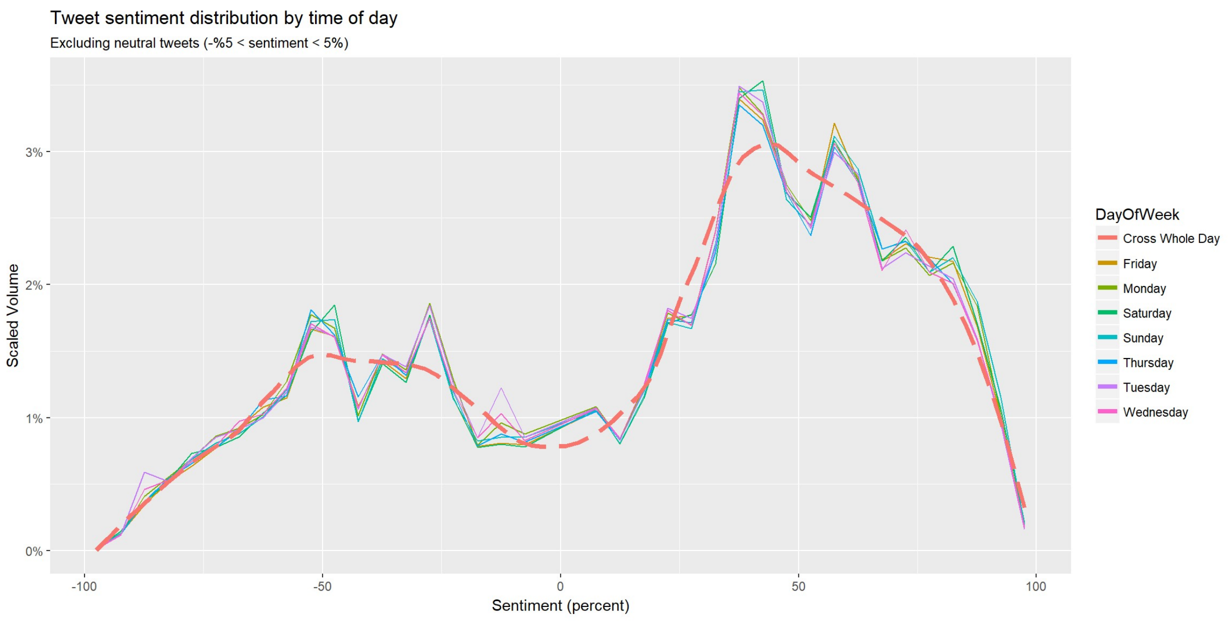
Scaling by total number of tweets per location and excluding central neutral tweets



Day of Week Analysis

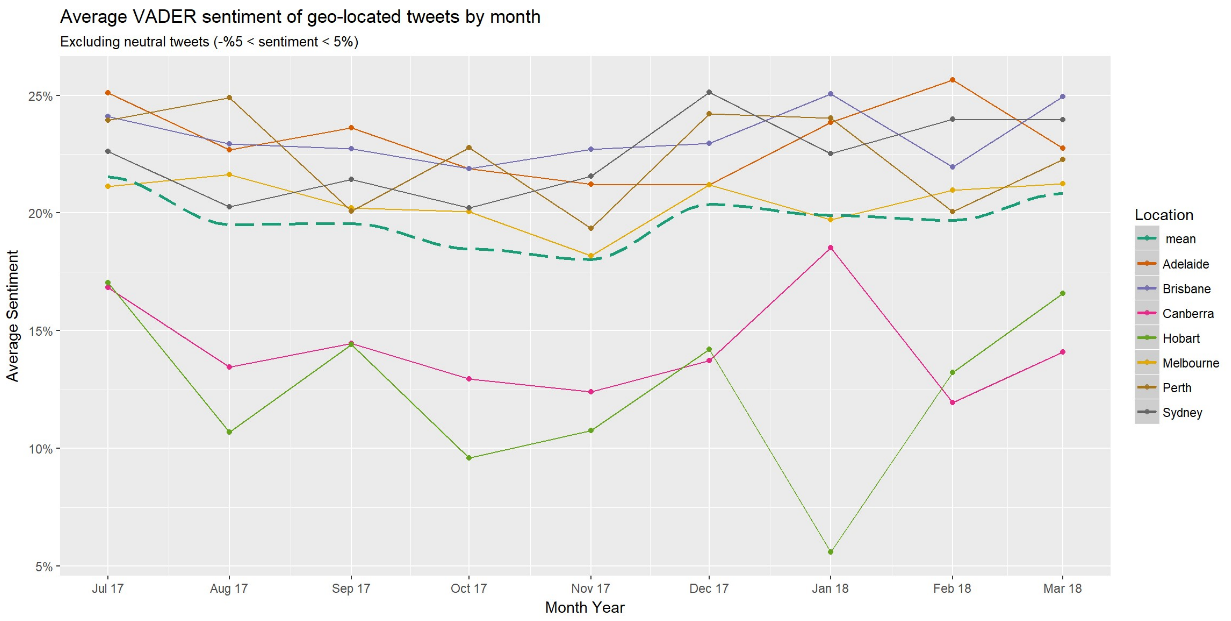


Scaling by total number of tweets per location and excluding central neutral tweets


Tweet sentiment distribution by time of day

Excluding neutral tweets (-%5 < sentiment < 5%)

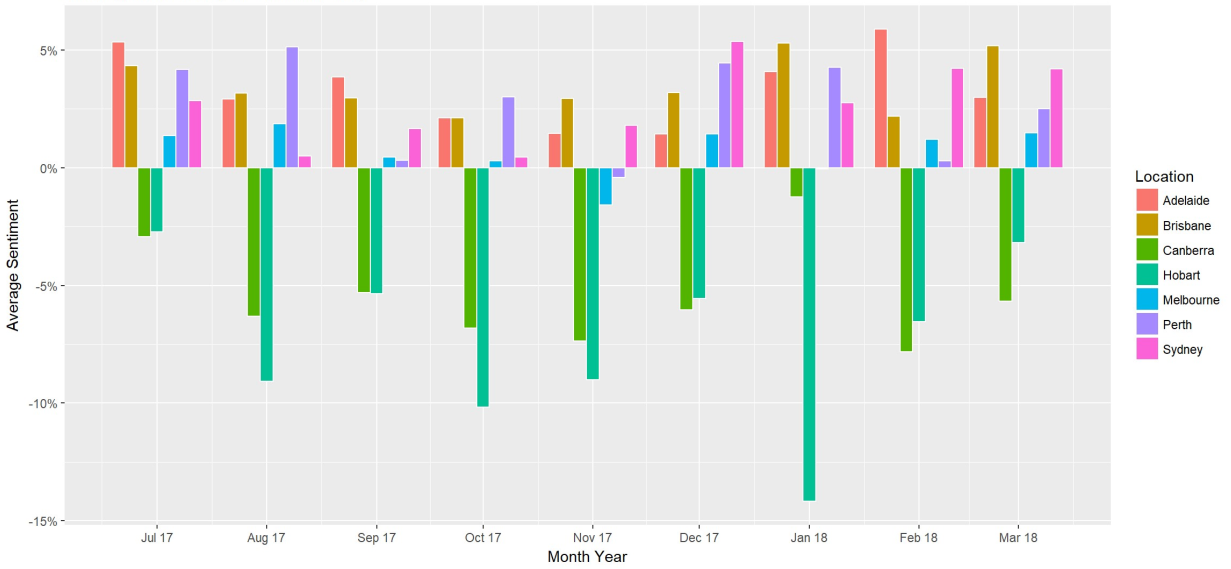
Location Analysis With Neutral Core Excluded


Average VADER sentiment of geo-located tweets by month

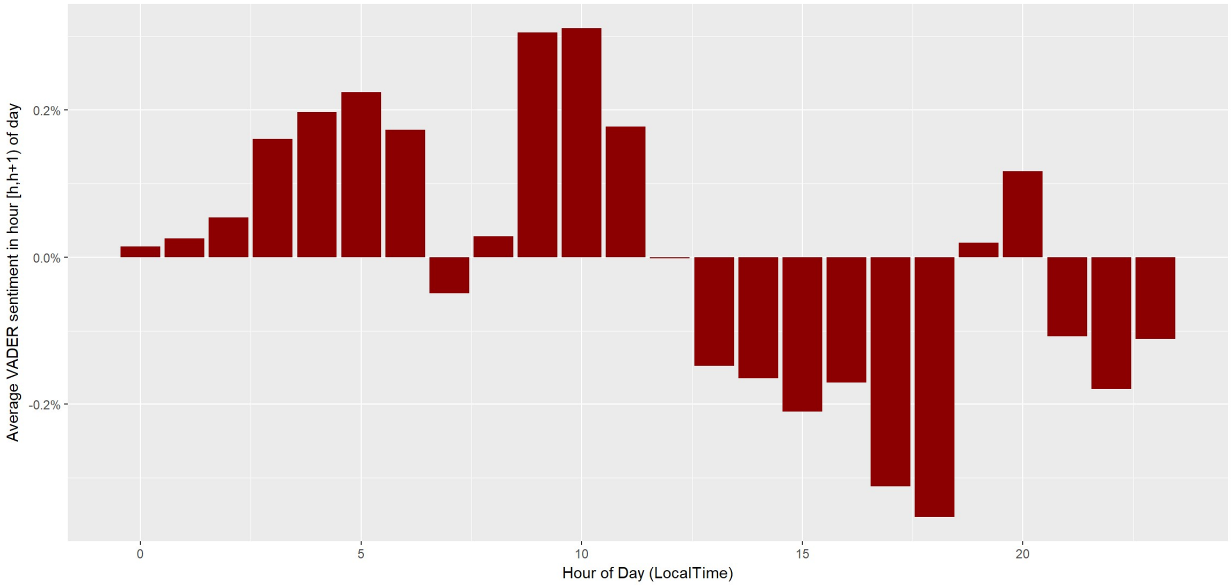
Excluding neutral tweets (-%5 < sentiment < 5%)

Baselined average VADER sentiment of geo-located tweets by month

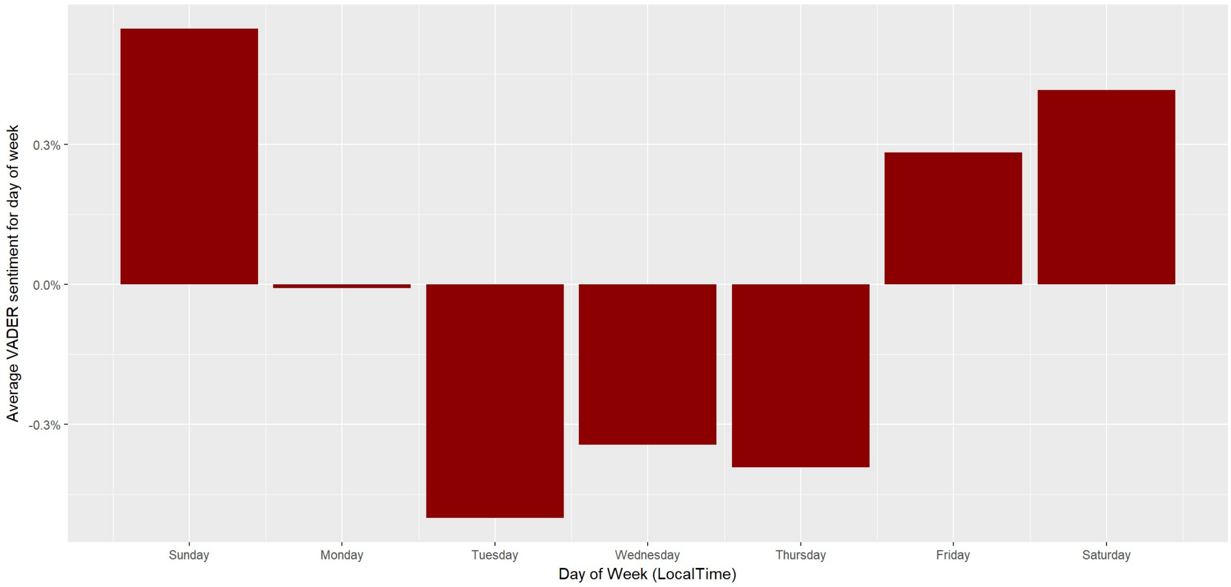
Excluding neutral tweets (-%5 < sentiment < 5%)



Baselined average variation in tweet sentiment with time of day



Baselined average variation tweet sentiment with day of week


Average VADER sentiment for day of week
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of Week (LocalTime)