WASHINGTON
Traffic Safety
COMMISSION

Measuring Current Traffic Safety Culture via Social Media Mining

Report Prepared by:

Fei Dai, Ph.D.
Mohhammad Sujon

Department of Civil and Environmental Engineering
West Virginia University

WTSC Report Number 2019-AG-2856

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
10/18/2019	Final Research Project Report	Oct. 2018 - Sep. 2019

4. TITLE AND SUBTITLE
Measuring Current Traffic Safety Culture via Social Media Mining

5a. CONTRACT NUMBER
2019-AG-2856

5b. GRANT NUMBER
2019-AG-2856

5c. PROGRAM ELEMENT NUMBER
n/a

6. AUTHOR(S)
Fei Dai
Mohhammad Sujon

5d. PROJECT NUMBER
n/a

5e. TASK NUMBER
n/a

5f. WORK UNIT NUMBER
n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Civil and Environmental Engineering
West Virginia University

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Washington Traffic Safety Commission
Olympia, WA

10. SPONSOR/MONITOR'S ACRONYM(S)
WTSC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
2019-AG-2856

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release.

13. SUPPLEMENTARY NOTES
The contents of the manuscript are solely the responsibility of the authors and do not necessarily reflect the official views of the funding agency.

14. ABSTRACT
This project exploited Twitter analysis and the technique of big data mining to address current traffic safety culture in Washington State. Twitter data was collected considering its high acceptance by people for sharing opinions and its huge active users. Twitter has also provided open Application Program Interfaces (APIs) for access to its historical data that can make the analysis feasible. Sentiment analysis and topic modeling were applied on the collected Twitter data to analytically reveal the public's opinions and their tendencies towards traffic safety over the past four years.

15. SUBJECT TERMS
social media mining, traffic safety culture, public sentiment, public opinion, Twitter mining, Washington State, data mining, target zero, traffic safety

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Staci Hoff, Ph.D., WTSC Research Director
					19b. TELEPHONE NUMBER *(Include area code)*
None	None	None	None	63	(360) 725-9874

Measuring Current Traffic Safety Culture via Social Media Mining

Report Prepared by:

Fei Dai

Mohhammad Sujon

Department of Civil and Environmental Engineering

West Virginia University

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Traffic safety is one of the main public concerns in the United States. Each year, traffic-related injuries and fatalities cause massive damage to human lives and properties. According to National Highway Traffic Safety Administration (NHTSA), 37,133 people lost their lives in motor vehicle traffic crashes on U.S. roadways during 2017 (NHTSA 2017). Studies have revealed that most traffic crashes were caused by factors associated with drivers (Lum and Reagan 1995). Road stakeholders sometimes neglect the importance of traffic safety, and it causes hazardous crashes that result in devastation to themselves and to other peoples. In the state of Washington, traffic data showed that the rate of fatalities increased by 5.4% from 2016 to 2017 (WTSC 2018). Traffic safety remains a concern.

To make the road safer and to reduce the number of casualties, Washington, as the first state in the nation, set a vision called "Target Zero" in 2000. The goal of this vision is to reduce the number of traffic fatalities and serious injuries to zero by the year 2030 (WTSC 2013). It seeks to evolve all residents in Washington into responsible citizens on road, who will behave responsibly for the safety of their lives and others. However, achieving zero deaths and grievous injuries on road is extremely challenging. It requires creation and implementation of new strategies including improving the current traffic safety culture.

Traffic safety culture (TSC) is an emerging concept in the field of traffic safety (Hedlund 2007). It can be referred to as an assembly of underlying assumptions, beliefs, values and attitudes shared by members of a community, which interact with a community's structures and systems to influence road safety related behaviors (Edwards et al. 2014). Recently, TSC has been identified as a priority for investigation in the Washington State Strategic Highway Safety Plan 2016: Target Zero (WTSC 2016). The Target Zero has included a call to action in shaking the roots of our belief that "accidents happen as a price of mobility" and shaping a culture of safe driving in Washington. To enable such an action, a key link is establishing a better understanding of the current traffic safety culture in Washington.

However, because the application of safety culture to traffic safety is new, there is little consensus about the nature and content of traffic safety culture (Dwards et al. 2014). It is also unclear what are the shared values, beliefs, and attitudes that are the most influential on Washington drivers' behavior. A number of questions have yet to be answered. For instance, what do we think is normal for Washington drivers? Do we all see traffic safety as an important issue to most people in our communities? Do we all believe it is possible to prevent fatal and serious injury crashes? Do we all believe seatbelts are effective in saving lives? Do we all think most people obey the speed limit? Do we all have the attitude that police enforcement of traffic laws is beneficial? From a pragmatic perspective, there is a pressing need to measure our current traffic safety culture, without which identifying the factors that should be modified to bring about the TSC improvement will remain difficult.

Nowadays, social media platforms (e.g., Twitter, Facebook) play a vital role in the data collection of the public's opinions. Many more people favor social media platforms for building social relationships, disseminating and obtaining information, advertising products and businesses,

and sharing opinions and interests (Akram and Kumar 2017). Social media analysis has tremendous potential to understand public's opinion with respect to a variety of topics. Moreover, in comparison with traditional survey methods, which have limitations of (1) limited number of participants in comparison to population, (2) being time-consuming and expensive in collecting data from participants, and (3) biased feedback from personal experience that can be manipulated, capitalizing on social media possesses benefits of (1) ease of accessing massive amount of data for analysis, (2) obtaining natural data on human behaviors, and (3) convenience of conducting a longitudinal study. As a result, social media analysis is a good complementary to survey studies and this project encompasses the social media analysis strategy to represent, measure, model, and mine meaningful patterns that characterize the Washington TSC from social media data.

This project exploited the social media and the technique of big data mining to address the above identified need. The objective of this project was to apply Twitter analysis as an alternative source of data to measure the current traffic safety culture in the state of Washington. The analysis was done to understand the traffic safety culture and get a realistic picture of public attitude toward traffic safety in Washington. In this project, Twitter data was collected considering its high acceptance by people for sharing opinions and its huge active users. Twitter has also provided open Application Program Interfaces (APIs) for access to its historical data that can make the analysis feasible. Sentiment analysis and topic modeling were then applied on the collected Twitter data to analytically reveal the public's opinions and their tendencies towards traffic safety over the past four years.

2. LITERATURE REVIEW

The concept of safety culture was coined after the tragic accident of the Chernobyl disaster. That unfortunate event caused a huge loss of assets and human lives. After the accident, the aspects of managerial and human factors in ensuring safety were discussed in several occasions (Flin et al. 2000). The term "Safety Culture" was first introduced in the International Nuclear Safety Group (INSAG)'s report in 1986 (INSAG 1986). In the "Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident", the term "Safety Culture" was described as "The assembly of characteristics and attitudes in organizations and individuals which establishes that, as an overriding priority, nuclear plant safety topics receive the attention warranted by their significance". This description was referred to by the traffic safety community, who later developed a standard term "Traffic Safety Culture" in 2007. A well-accepted definition of traffic safety culture was defined as "the assembly of underlying assumptions, beliefs, values and attitudes shared by members of a community, which interact with a community's structures and systems to influence road safety-related behaviors" (Edwards et al. 2014).

Every culture consists of cognition, behaviors, and artifacts. Cognition guides and drives cultural-based behaviors where behavior relates to activities that produce information about the culture like custom and ceremonial behavior. Artifacts are the depiction of symbols, expressions, and tools of a culture which can include rules that govern cultural consent. To have a better understanding, traffic safety culture can be reviewed from a cognitive perspective. In the cognitive perspective, "traffic safety culture can be defined as the perceptions people have about what behaviors are normal in their peer group and their expectations for how that group reactions to violations to these behavioral norms. In terms for traffic safety, this definition applies to behaviors that either increase crash risk (e.g., speeding) or are protective (e.g., wearing seatbelts), as well as behaviors related to acceptance or rejection of traffic safety interventions" (Ward et al. 2010).

Traffic safety culture has drawn high attention in recent years. It has been selected as a preference for improvement in the Washington State Strategic Highway Safety Plan 2016: Target Zero (WTSC 2016). The Target Zero has included a call to action in shaking the roots of our belief that "accidents happen as a price of mobility" and shaping a culture of safe driving in Washington. To enable such an action, a key link is establishing a better understanding of the current traffic safety culture in Washington. Studies have pointed out the benefits of understanding the traffic safety culture (Ward et al. 2015) including, but are not limited to, facilitating explanation of high-risk driving behaviors (Conner et al. 2007), assisting in identification of high-risk-taking drivers (Li et al. 2014), supporting the decision-making of traffic safety policies and programs (Rakauskas et al. 2009), and ameliorating cultural shift for a vision of zero traffic fatalities (Ward et al. 2014; WTSC 2016).

However, because the application of safety culture to traffic safety is recent, there is a limited consensus about the structure and substance of traffic safety culture (Edwards et al. 2014). It is still ambiguous what are the shared values, beliefs, and attitudes that are the most prominent among Washington drivers and therefore direct their behaviors. Several questions have still to be dealt with. For example, what do we think is natural for Washington drivers? Do we all think

traffic safety as an important issue for most people in our communities? Do we all believe it is feasible to prevent fatal and serious injury crashes? Do we all believe seatbelts are adequate in saving lives? Do we all think most people obey the speed limit? Do we all have the attitude that police enforcement of traffic laws is beneficial? From a realistic viewpoint, there is a compelling demand to measure our present traffic safety culture, without which identifying the factors that should be adjusted to bring about the TSC improvement will remain problematic.

On the Internet, social media platforms generate an enormous volume of data with valuable information. These platforms allow millions of active users to convey their views and spread their opinions on a specific topic (El Alaoui et al. 2018). Today, many people appreciate social media platforms for developing social relations, disseminating and collecting information, promoting products and businesses, and sharing points of view and interests. Social media analysis has great potential to reveal the public's opinions in regard to a range of topics (Bian et al. 2016; Cobb et al. 2011; Paul and Dredze 2011; Tumasjan et al. 2010). As a popular venue among the existing platforms, Twitter has crossed the border of countries and brought its service practically to all over the world. Its universal presence stimulates people to share their viewpoints with millions of users. The nominal cost of surfing the web and cost-free account opening policy is appealing to people. The volume of generated data on Twitter is enormous. Twitter made the data accessible to researchers and businesses. Data availability, affordable cost, and universal geographic coverage have made Twitter to be an appealing source for information. Due to these reasons, twitter data mining has been applied in a diversity of transportation-related studies such as transportation incidents analysis (Mai and Hranac 2013), traffic detection on road (Panchal and Apare 2016), division between traffic congestions and traffic crashes (D'Andrea et al. 2015), traffic flow analysis for city event detection (Zhou et al. 2016), and traffic event detection (Albuquerque et al. 2016).

Among existing techniques to interpret social media data, sentiment analysis is one of the most popular. It involves the designation of posts generated in social media and detects the sentimental polarity of these posts and classifies them into different groups like positive, negative and further (Mäntylä et al. 2018). The quantity of groups depends on the special necessity of the analyzed problem. The methods associated with sentiment analysis can usually be categorized into two groups:

1. The lexicon-based method that calculates the polarity of a document from the linguistic orientation of words or phrases in the document (Tausczik and Pennebaker 2009).
2. Machine learning (ML) based method that requires developing models from the labeled training dataset to detect the orientation of a document (Pang and Lee 2008).

The lexicon-based method requires measuring orientation for a document from the linguistic orientation of words or phrases in the document (Turney 2002). It opens a text document and analysis the file by word by word. Every word in that text file matched with the word in the dictionary file denoted with a sentimental value (Tausczik and Pennebaker 2009). The dictionary utilized for matching the word can be generated manually or automatically (Taboada et al. 2011). After evaluating all the words in the sentence, the sentiment values are later sum up. If the sum is larger than zero, the sentence is classified as positive. If the sum is smaller than zero, the sentence is classified as negative. The sentence will be classified as neutral if the sum is zero.

The machine learning-based method is generally represented as a classification problem. A classifier is provided with a text and yields the comparable group, e.g. positive, negative, or neutral. The analysis is usually was done with supervised or unsupervised methods. In the supervised method, the model is trained with a labeled dataset where it learns to identify a text with the corresponding group. The feature extractor transforms the text input into a feature vector. Pairs of feature vectors and categories like positive, negative, or neutral are provided into the machine learning algorithm to develop a model. After developing the model, an unlabeled dataset is handed over to that model to analyze. The model predicts the sentiment polarity based on the features extracted from the text. The most accepted models applied for opinion mining are Naïve Bayesian classification, the maximal entropy principle, and the support vector machine. The unsupervised method functions without a training dataset. Unavailability of the trained dataset for a distinct topic rarely takes place and in that exceptional case, the unsupervised methods are implemented (El Alaoui et al. 2018).

For the lexicon-based method, labeled dataset and the practice of generating the model is not necessary. For the machine learning-based method, classifier trained in one domain cannot be suitable to other domains in most situations. In this research, the lexicon-based method has been selected to analyze the sentiments of tweets thanks to the above reasons and considering the widespread availability of the lexical resources. This research have utilized a computerized text analysis method named Linguistic Inquiry and Word Count (LIWC). LIWC is a straightforward text analysis program that generates a psychologically relevant category for a word. Experimental results from the LIWC-based analysis established its strength in detecting meaning in a broad range of empirical settings (Tausczik and Pennebaker 2010). LIWC is user-friendly, rapid and accurate (Crossley et al. 2017).

Besides sentiment analysis, topic modeling is another useful technique to uncover insights and meanings from text. In the realm of machine learning and natural language processing, topic modeling is a frequently applied text-mining tool for the exploration of concealed linguistic structures (i.e., topics) from a wide unregulated set of documents by analyzing the words within the texts (Blei et al. 2010). One of the most significant features of topic modeling is that it does not require any prior annotations or labeling of the documents. This is particularly suitable for processing of massive, large-scale text messages contained in social media data. In specific, this project has selected the Latent Dirichlet allocation (LDA) method (Blei et al. 2003). As a generative probabilistic model, it is capable of categorizing the set of tweets into latent topics. In the LDA model, each document (e.g., tweet set), regarded as a vector of word counts applying the bag-of-words method. It is a mixture of probabilities over the topics, where each topic is represented as a probability distribution over a set of words (i.e., the dictionary).

3. METHODOLOGY

The objective of this research project was to implement Twitter analysis as an alternative data source to measure the current traffic safety culture in the state of Washington. The overarching goal was to establish a cost-effective method to better understand Washingtonians' shared value, beliefs, and attitudes that are most influential on Washington drivers' behavior such that more effective safety strategies can be identified and developed. The specific research objectives of this project included:

1. Establish procedures to collect, clean, and pre-process pertinent social media data in Twitter.
2. Analyze the public's general attitudes toward traffic safety.
3. Analyze the public's specific attitudes toward the high-risk behaviors determined in the "Target Zero" plan.
4. Identify key culture determinants that drive these high-risk behaviors.

To accomplish these research objectives, this project has established and implemented the procedures as shown in Figure 1. Detailed explanation of these procedures are provided in the following sections.

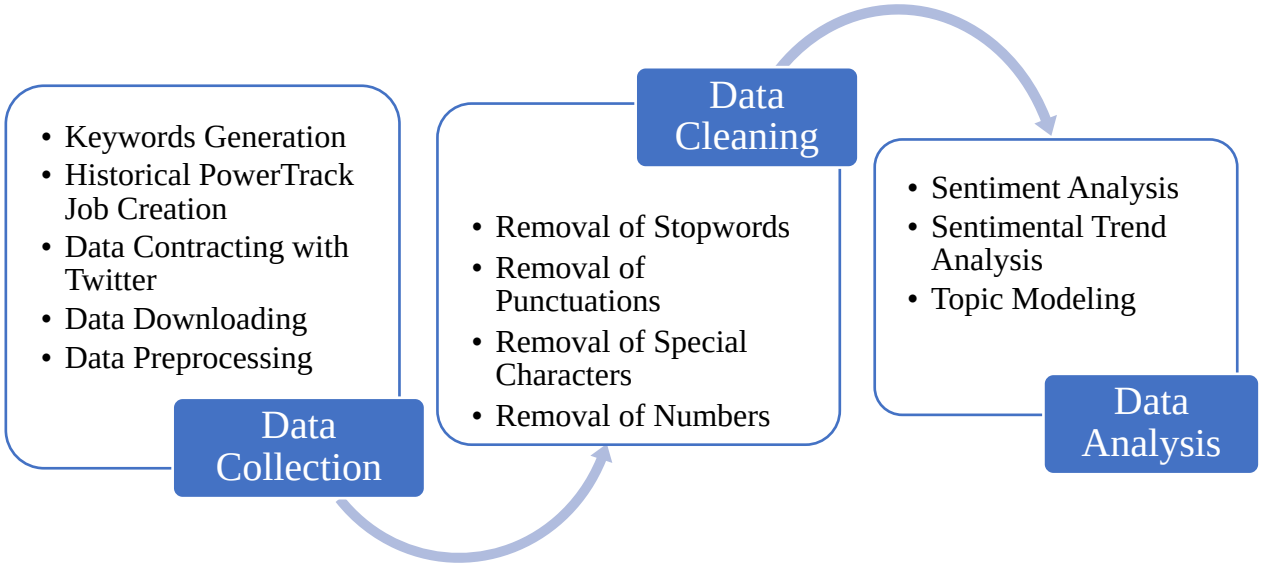


Figure 1. Data collection, cleaning and analysis process flowchart

3.1 Data Collection

To reflect the current traffic safety culture in Washington State, the following questions were raised as measures to guide the data collection and the subsequent analysis. These questions were formatted based on the chapter "Improving our Traffic Safety Culture" in the 2016 "Target Zero" plan and discussion with the experts in Washington Traffic Safety Commission (WTSC):

1. Do we all see traffic safety as an important issue to most people in our communities?
2. Do we all believe it is possible to prevent fatal and serious injury crashes?
3. Do we all have the attitude that police enforcement of traffic laws is beneficial?
4. What are the public's attitudes toward the high risk behaviors in terms of impairment involved, speeding involved, distraction involved, unrestrained vehicle occupants, young driver involved, and older driver involved?

The timeframe of tweets collected for this research project started in March 2015 and ended in February 2019. Only tweets that were relevant to the studied topics and generated in the state of Washington were collected. There are a number of attributes such as retweets and users' information associated with the collected tweets. As the data collection process plays an important role in the subsequent study planned in this project, a rigous steps were set forth and implemented inlucding keywords search and generation, Historical PowerTrack job creation for data extraction, contracting with Twitter for data transaction, data downloading, and data preprocessing.

3.1.1 Traffic safety-related keywords generation

The data collection process started with generating a keyword list with traffic safety-related words. To retrieve data from Twitter for the subsequent analysis as specified earlier, the research team generated keywords and phrases from the following sources that were deemed to be relevant to this research project:

- https://www.wtscpartners.com [Washington Traffic Safety Culture Partner's website]
- http://wadrivetozero.com [Washington's Target Zero's official website]
- http://wtsc.wa.gov [Washington Traffic Safety Culture Official Website]
- https://www.facebook.com/WATargetZero/ [Official Facebook page for Target Zero]
- https://twitter.com/targetzero [Official Twitter page for Target Zero]
- "Washington State Strategic Highway Safety Plan 2016" by Washington Traffic Safety Commission
- "A framework for conceptualizing traffic safety culture" by Jason Edwards, James Freeman, David Soole, and Barry Watson
- "Measuring Minnesota's Traffi¬c Safety Culture" by Minnesota Department of Transportation
- "2011 Tennessee Traffic Safety Culture Survey" by UT Center for Transportation Research and The Center for Applied Research and Evaluation
- 2015 Traffic Safety Culture Index, 2016, AAA Foundation for Traffic Safety
- 2016 Traffic Safety Culture Index, 2017, AAA Foundation for Traffic Safety
- 2017 Traffic Safety Culture Index, 2018, AAA Foundation for Traffic Safety
- Twitter website relevant topics review
- Google website relevant topics research and review

The keywords were refined and checked several times to test their suitability for collection of tweets with respect to topics including the public's general attitudes toward the importance of traffic safety, the public's general beliefs on possibility of zero fatalities, and the public's specific attitudes toward to a diversity of determined unsafe driving behaviors (i.e., impairment involved,

speeding involved, distraction involved, unrestrained vehicle occupants, young driver involved, and older driver involved). Nine sets of keywords were obtained, which are provided in Appendix A.

3.1.2 Creating Historical PowerTrack job to collect tweets from Twitter

To collect relevant tweets from Twitter using the generated keywords, a Historical PowerTrack (HPT) job needed to be created based on the requirement of Twitter. The following explain the procedures associated with creating a Historical PowerTrack job in Twitter for data collection.

3.1.2.1 Twitter API

Twitter's API [https://developer.twitter.com/en/docs/tweets/search/overview] provides three types of API. To get the basic response from Twitter for our context, we have used the Standard API [https://developer.twitter.com/en/docs/tweets/search/guides/standard-operators]. It helped in creating the dictionary of keywords. However, due to the limitation of this API (time limit and the number of tweets), we had to use Premium API [https://developer.twitter.com/en/docs/tweets/search/guides/premium-operators] for our research work. This premium API provided us the tweets through a Historical PowerTrack job.

3.1.2.2 Historical PowerTrack

Historical PowerTrack (HPT) [https://developer.twitter.com/en/docs/tweets/batch-historical/overview] is an enterprise API that grants access to a historical archive of public Twitter data. It can get a tweet from the beginning of Twitter. The process of getting a batch of the historical tweets was done with the rule-based-filtering system, ready to deliver complete coverage of historical Twitter data.

Getting data through Historical PowerTrack begins through creating historical 'jobs' [https://developer.twitter.com/en/docs/tweets/rules-and-filtering/overview/premium-operators] which is a set of PowerTrack filtering rules [https://developer.twitter.com/en/docs/tweets/rules-and-filtering/overview/premium-operators] and a historical time frame for which customer would like to retrieve matching data from the Twitter archive. These jobs could be created and managed through the Historical PowerTrack API.

There are a few steps needed to complete a job in Historical PowerTrack. They are

1. The job provider must give a period and the PowerTrack rules to create a job in Historical PowerTrack API. The API then estimated the number of data volumes and the time needed to complete the job.
2. Then the provider must accept or reject the job. If the rejection is selected, then the job provider can edit the job to meet the volume estimation and budget. When the job is accepted, the job will be run to generate data files.
3. After the job is complete, the job provider must send a request to API to gain the list of URLs, which was used to download the data files.

4. By using this list, the job provider can download the files. Each file contains a 10-minute segment of the overall job and for 15 days, the data is available for download.

3.1.2.3 Our Historical PowerTrack Job

By following the procedure and example provided by Twitter, we created our job. The job was then used to collect tweets from Twitter. The part of the job is provided in JSON format.

The partial job is given below:

{

"from": "2015/3/1",

"to": "2019/2/28",

"rules": [

{

"value":"((road OR traffic OR driving OR highway) safety OR (traffic OR road OR highway OR vehicle OR car OR bus OR automobile OR truck OR SUV OR van OR taxi OR motorbike OR motorcycle OR motorcyclist OR pedestrian OR intersection OR lane departure) (accident OR accidents OR collision OR collisions OR crash OR crashes OR crashing OR crashed OR fatality OR fatalities OR injury OR injuries OR death OR deaths OR casualty OR casualties OR killed OR negligence) OR (aggressive OR aggressively OR risk OR risky OR unsafe OR untrained OR illegal OR negligent OR reckless) (driver OR driving) OR "road mishap" OR "deadly traffic" OR "traffic offender" OR "traffic offenders") profile_region:Washington",

"tag": "the importance of traffic safety"

},

.....

.....

{

"value":" ((drive OR driving OR driver OR drivers OR drove) ("dozed off" OR sleep OR asleep OR sleepy OR sleeping OR slept OR drowsiness OR drowsy OR "shift work" OR insomnia)) profile_region: Washington",

"tag": "drowsy driver involved high-risk behavior"

]

}

3.1.2.4 Result estimation

The job estimation returned with a result of 5.7 million tweets over 1,461-days spanned.

3.1.3 Data contracting with Twitter

After the job estimation by Twitter, the next phase was to establish a contract with Twitter to purchase the data officially. Twitter provided its historical tweets and related user's information through a Twitter Developer Agreement. This process was called data contracting, which included signing the Twitter Developer Agreement that had to be made between an individual (or an entity) and Twitter (Twitter, Inc. and Twitter International Company). This agreement governed an individual's access to and use of the Licensed Material. There are four steps needed to be followed in order for granting access to Twitter data.

3.1.3.1 Submission and gaining approval on Use Case

At the beginning of the process, an individual needed to submit a Use Case and should gain approval on this Use Case by Twitter's Public Policy team. Use Case described the reasons for, and possible applications of the data acquired from Twitter. Individuals needed to provide as many details and specifics as possible. This needed to be done precisely to avoid follow-up questions which can draw the process out. Detailed use case policy where restrictions on the use of licensed materials were provided at https://developer.twitter.com/en/developer-terms/agreement-and-policy.html. Detailed used case description was prepared and submitted. Approval was obtained.

3.1.3.2 Signing of Master License Agreement and Order Form

After the approval of the Use Case, Twitter provided the Master License Agreement and Order Form. These documents provided the conditions which the individual needed to agree before purchasing data from Twitter. These documents were reviewed and signed by the WVU procurement office. The Master License Agreement was fully executed. The entire process could take 2-3 weeks depending on how much data Twitter were providing and how quickly the individual could turn around Use Case/Signatures.

3.1.3.3 Payment

Twitter required the payment processed before providing the data access to the individual. The payment was made through the WVU procurement office.

3.1.4 Downloading data

After the payment was completed, Twitter granted WVU access to and use the tweet data returned from its archive using the earier created HPT job. Twitter sent a Data URL link with login credentials. This Data URL provided a list of download links via the Historical PowerTrack (HPT) API. As this list contained thousands of links, some form of download automation was needed to retrieve the data. The size of data returned by a job was huge, both in the number of activities and in the storage size of the output payload. The size of the resulting dataset is about 4 gigabytes. In order to ensure files that are quick to download, the API generated data files in a 10-minute time-series. That is, each file covered a ten-minute period of data. Therefore, the total number of 210,335 files needed to be downloaded. Depending on the data volumes associated with the job's filters, even these 10-minute files could contain many thousands of tweets. The data files generated

were hosted at Amazon's Simple Storage Service (S3) and were available for 15 days. These files were gzip-compressed JSON files and were based on the UTF-8-character set. All timestamps used in the job description, included in API responses, used in filenames, and in the returned tweet data were in UTC. The whole downloading process was time-consuming. The process was interrupted several times because of frequent connection timeout. So, the process was closely monitored. After downloading, the next step started with preprocessing these data. The data files were converted from JSON to CSV format to make the data applicable to sentiment analysis. Database software MongoDB and Studio 3T were used to complete the process.

3.1.4.1 Accessing download links

Once payment for the job was completed, Twitter sent an email containing a Data URL that contained the data files links that we needed to download. The Data URL is shown as below, and we used it in a later step.

> https://gnip-api.gnip.com/historical/powertrack/accounts/{ACCOUNT_NAME}/publishers/twitter/jobs/{JOB_UUID}/results.json

3.1.4.2 Downloading data

Considering that the HPT job contained thousands of files that need to be downloaded, unzipped, and combined, we needed to use an automated process. We have listed the example strategy below to help data files downloading automation. We followed this strategy because it has an interactive command prompt, and the ability to download and resume at where the last downloading was interrupted.

1. Replaced the ".json" extension of the Data URL/Job URL as in:
   https://gnip-api.gnip.com/historical/powertrack/accounts/{ACCOUNT_NAME}/publishers/twitter/jobs/{JOB_UUID}/results.json
   with a ".csv" extension as in:
   https://gnip-api.gnip.com/historical/powertrack/accounts/{ACCOUNT_NAME}/publishers/twitter/jobs/{JOB_UUID}/results.csv
   Here, the {ACCOUNT_NAME} is WestVirginiaUniversity; the {JOB_UUID} is the job ID provided by Twitter. This link was used to download the CSV data link files in the next step.
2. Ran the link in the browser to download CSV file "results.csv" containing the list of download links that includes data. Each line contained a file name and the corresponding Amazon S3 link. This file was used by later steps to download, unzip, and combine data files.
3. To automate the downloading process, the PTDataDownload tool was installed on the computer by:
   - Downloading PTDataDownload.zip [Link: https://github.com/gnip/support/releases].

- Unzipping the file to deploy the files and folder structure needed to download the data files.

4. Added the CSV file "results.csv" from Step 2 into the 'PTDataDownload/input' folder and it was assured that the 'PTDataDownload/download' folder was empty.
5. Installed Cygwin [Link: http://cygwin.com/setup-x86_64.exe]. In the installation process, Cygwin deployed a set of 'Base' packages, which included Cygwin itself along with common Unix-like utilities such as grep, gawk, gzip, sed, which and bash. A critical package for running most of our work commands was the 'curl' package. The 'curl' package was not deployed by default so it must be selected under the 'Web' set of packages with the most updated version. After making the selection, the installer started downloading the libraries and utilities and required dependencies.
6. After Cygwin with the cURL package installed, opened 'Cygwin Terminal' interface and navigated working directory to 'PTDataDownload'.
7. Typed the following into command line:
   $ ./run.sh
8. A reply was prompted by the following:
   Historical Download Options:
   d: Download files.
   D: Delete downloaded file.
   q: Quit/Exit.
   Enter selection:
   Entering 'd' and then 'return' initiated the download of 10-minute interval gzip files into the 'downloads' folder within 'PTDataDownload.' This took a while depending on the size of the job to download.
   Please be noted: when the downloading job was interrupted for any reason, deleted the most recently added gzip file from the 'downloads' folder and re-entered $ ./run.sh into the command line.
   The data downloading process started on 27th May and was completed on 2nd June. The whole process took almost a week.
9. Once all the files were downloaded, made working directory PTDataDownload and entered the following into command line:
   $ zcat -d -r downloads/ > filename.json
   This automatically unzipped and combined all files into a single file, which could be named whatever. This file showed up in the folder PTDataDownload once completed.

3.1.5 Preprocessing data

The format of the downloaded data was JSON. JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for machines to parse and generate. But the data amount is too big for every tweet, and for our sentiment analysis purpose we only need related tweets, matching tags, and time of tweeting. So, we have transferred the data from JSON to CSV (Comma-separated values). It is an Excel related format that can be easily used to process big data. To do so, first, we installed MongoDB, which is a document database that stores and converts the

format of data. We also installed Studio3T (Professional GUI and IDE for MongoDB) for visualization of our operation of data in MongoDB (Note: in MongoDB, one can only use commend to operate data). Next, we exported our desired part of the data in CSV format. The steps are provided below:

3.1.5.1 Installing and applying MongoDB

1. Went to MongoDB download page [https://www.mongodb.com/download-center] and downloaded MongoDB Community Server.
2. Ran the installation file and chose all default settings. After completion, created a folder named 'data' in the installed folder of MongoDB.
3. Ran the command window and changed the path to the bin folder of the installed MongoDB folder. Typed the command "mongod" to start the MongoDB server. Then, Ran another command window and typed the command "mongo" to start the mongo shell.
4. In the Advanced System Settings, added the path of the bin folder to environment variables. It helped automatic detection of commands such as "mongod" and "mongo". After that, we didn't need to change the path anymore.

3.1.5.2 Installing and applying Studio3T

To have a better GUI to work with MongoDB, we have installed and used Studio3T.

1. Downloaded Studio3T from the download page [https://studio3t.com/download/].
2. Installed with the default setup. After installation, ran the program.
3. Created a new database for our work.
4. In the Import section, imported the JSON file.
5. Executed to convert formats of data. Based on the size of the file, the execution time varied.
6. After processing, exported the data in the desired CSV format. We choose only selected columns based on our work need.
7. The data was also separated based on set time intervals for such preprocessing process. Once done, data cleaning and sentiment analysis were subsequently carried out.

3.2 Data Cleaning

After collecting the data from Twitter, it is very important to clean and process the raw data to remove noise and other unnecessary objects (Haddi et al. 2013). The raw data commonly possess many noises and analysis of such data can yield unintended results. There are also many unusable words that are not necessary for the analysis. Therefore, the collected data were cleaned before analysis. The data cleaning process was completed using the Python code.

3.2.1 Removal of stop-words

When data analysis needed to be data-driven at the word level, the commonly occurring words (stop-words) should be removed. Stop words are frequently used words (such as "the", "a", "an", "in") appear in the tweet for grammatical purpose and removing them increased accuracy of sentiment analysis.

3.2.2 Removal of punctuations

All the punctuation marks according to the priorities was dealt with. For example: ".", ",","?" are important punctuations that should be retained while others needed to be removed.

3.2.3 Removal of numbers

The numbers had no sentimental values and considered as noise elements in sentiment analysis. They were eliminated in the cleaning process.

3.2.4 Removal of expressions and symbols

Textual data (usually speech transcripts) contained human expressions like [laughing], [Crying], [Audience paused]. These expressions were non-relevant to the content of the speech and hence needed to be removed. Also, the symbols like $, * has no sentimental expression. They were also removed. After cleaning the data, all tweets were converted to lower case to remove heterogeneity from the data.

3.3 Data Analysis

3.3.1 Sentiment and sentimental trend analysis

Based on the literature review, this project applied LIWC for the sentiment analysis of the collected tweets. LIWC calculates the frequency of words in a category defined by sentiment polarity. It also acknowledges standard language categories like articles and prepositions, psychological processes such as emotion, cognitive, sensory, social, and denoted words relativity as time and space (Kahn et al. 2007). The process of applying LIWC was materialized by using SentiStrength. SentiStrength is an executable application that can automatically analyze the sentiment polarity of a sentence. This application is free of charge for academic research and the GUI is very user-friendly. The efficiency of this application to predict positive emotion is 60.6% and for negative emotion, it is 72.8% accurate on a scale of 1-5 (Thelwall et al. 2010). In this project, the sentiments over the aforementioned period for the three general and six high-risk behavioral topics were analyzed individually. To get a better understanding of traffic safety cultural shift, the trend of these sentiments was then analyzed. For that, every 6 months was taken as a period and a total of 8 periods of results were generated.

3.3.2 Topic modeling

After sentiment analysis, topic modeling was formed to generate latent topics from the Twitter data, in order to possibly inform effective strategies in improving the current traffic safety. To construct meaningful latent topics in our context, the same cleaned data was used. The whole process of applying Latent Dirichlet Allocation (LDA)-based topic modeling was materialized by LDAvis. LDAvis is an online-based method for interactive visualization of topics generated by applying LDA. This visualization method provides a comprehensive observation of the topics with consequential differences among them, while at the same time providing for a deep analysis of the terms that most highly correlated with each individual topic. LDAvis provides users the

flexibly of exploring the topic-term relationships through LDA model (Sievert and Shirley 2014). The LDAvis was applied through the python code. For each general and high-risk behavior topic, 5 topics were created. The name of each topic was given manually. The name was dependent on the keywords and the possible relations among them. The layout of LDAvis is illustrated in Figure 2. To see the keywords generated for each topic, the circle with denoted topic number had to be clicked. The area of the circle is based on the amount of keywords frequency. Topic 1 had the highest area and topic 5 had the lowest.

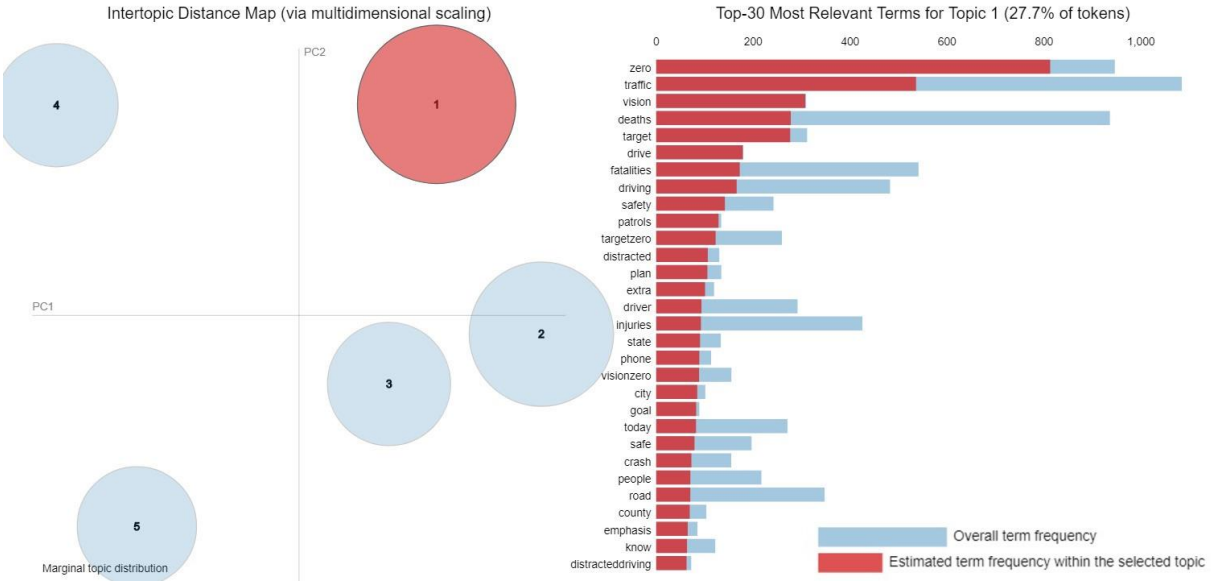


Figure 2. The layout of LDAvis

4. RESULTS

4.1 Sentiment and Sentimental Trend Analysis

4.1.1 Attitude towards importance of traffic safety

Figure 3 illustrated the results of sentiment analysis regarding the public's general attitudes towards the importance of traffic safety in Washington. The total tweets collected for this topic were around 5.5 million. The results of the analysis showed that in general, 55 percent of the residents agreed that traffic safety is important in their daily lives. It might be interpreted that these people paid attention to traffic rules and regulations and acted responsibly. By accepting the importance of traffic safety, they could save lives of themselves and other people. Based on the results, it can be said that the policies adopted by WTSC yielded a positive result. The traffic safety officials did their job efficiently by creating awareness among the public about the traffic safety. The analysis also showed that 30 percent of residents didn't take the traffic safety seriously and 15 percent of residents had a neutral reaction to the traffic safety. This result might be linked, to certain extent, with why the crashes, while reducing, still occurred. This calls for attention of the public and the officials, and it sets need for continued improvement of our traffic safety culture. New effective strategies for improving our traffic safety conditions are still desirable, which will lead to more positive feedback from people regarding the importance of traffic safety.

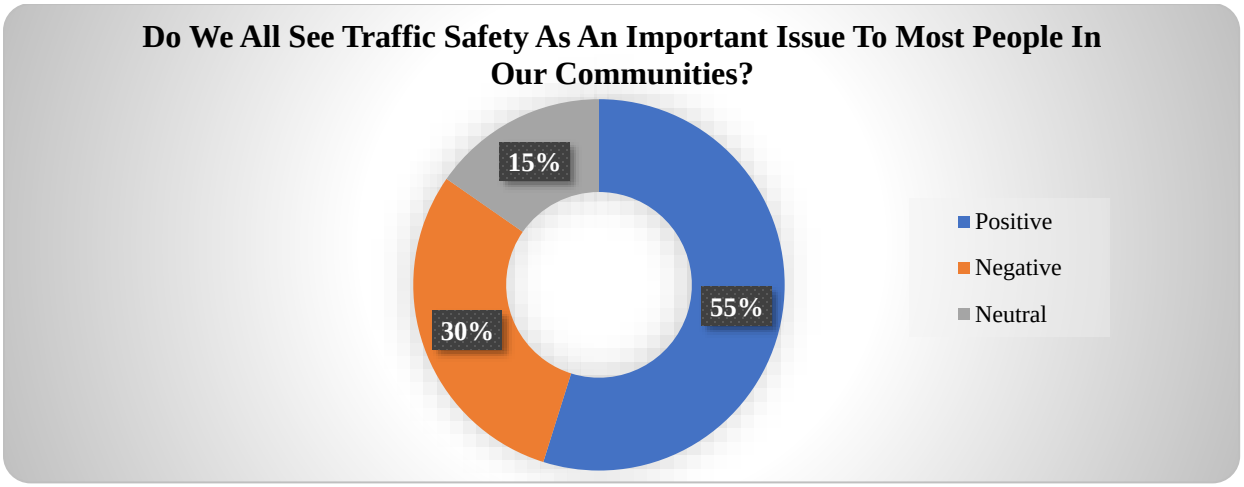
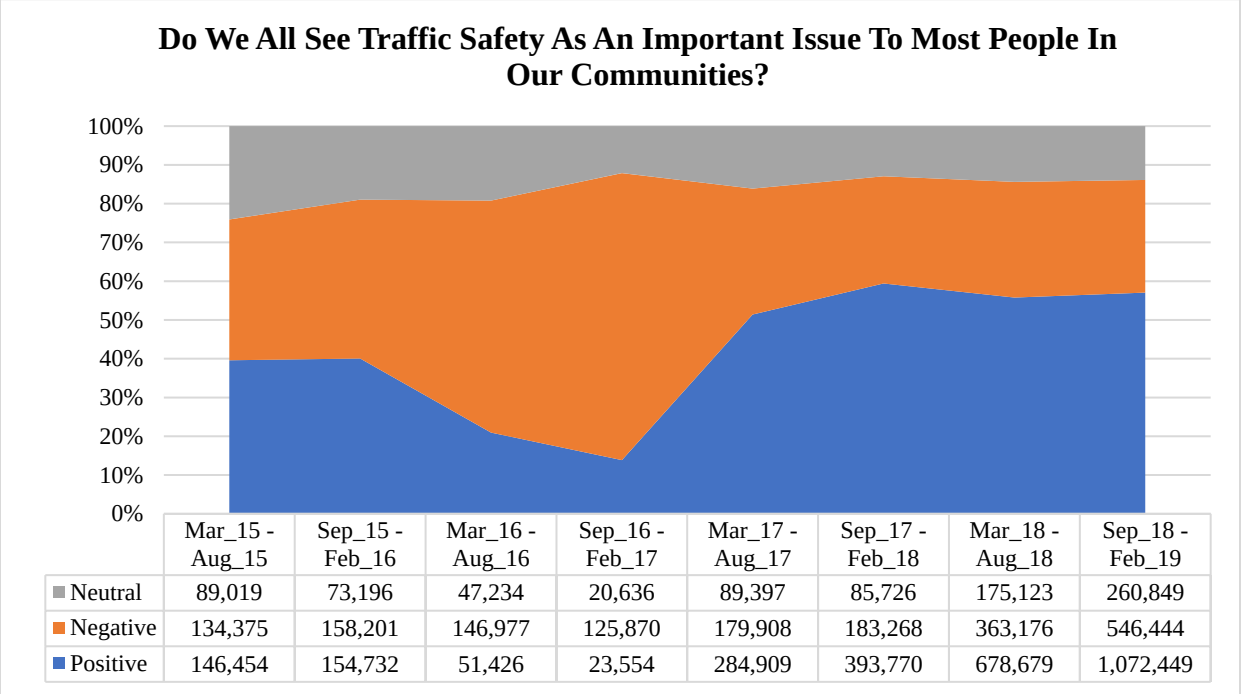


Figure 3. Sentiment for the question "Do we all see traffic safety as an important issue to most people in our communities?"

In Figure 4, the trend analysis of sentiments regarding the public's general attitudes toward the importance of traffic safety in Washington was illustrated. It showed that the perception of the people changed over time. Around 40 percent of people believed safety as an important issue from March 2015 to February 2016. The trend of this positive thinking dropped from March 2016 to February 2017. After this period, the trend of the positive thinking grew significantly between March 2017 and February 2018, reaching up to 60 percent. This trend indicated that people are becoming more concerned about traffic safety and they are acting responsibly on road. It implied

good behavioral changing trend had generated among the residents. This could represent a positive change in traffic safety culture. However, ever since, the trend had remained relatively unchanged from March 2018 to February 2019. On the other hand, the trends of the negative and neutral thinking, though narrowing, still accounted for a certain portion with the passing of time.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	89,019	73,196	47,234	20,636	89,397	85,726	175,123	260,849
Negative	134,375	158,201	146,977	125,870	179,908	183,268	363,176	546,444
Positive	146,454	154,732	51,426	23,554	284,909	393,770	678,679	1,072,449

Figure 4. Sentimental trend for the question "Do we all see traffic safety as an important issue to most people in our communities?"

4.1.2 Belief on possibility of preventing fatal and serious injury crashes

Figure 5 presented the results of sentiment analysis regarding the public's general beliefs on possibility of zero fatalities. The number of tweets collected for this topic were 10,827. The analysis revealed that most residents had a neutral and negative view for the possibility of preventing fatal and serious injury crashes. The reaction can be explained as most of them are not sure how they should react to this idea. They didn't seem to understand the belief that zero fatalities can be achievable by strictly following the traffic rules. The negative response could be the result of general thoughts that the amount of fatalities could be reduced, but could not reach the point zero. The reason could be most people believe that the strategies adopted by traffic safety agencies could make most people follow the traffic rules but still there are some people who would overlook the rules and thus could cause harm to other people. Those violated traffic rules could lead to serious crashes on streets and cause damage to the public. Also, some pedestrians usually crossed the road without paying attention to traffic signals. There are many factors that could influence people's beliefs. This result highlighted the challenge for achieving the "Target Zero" goal. Based on the analysis, only a few individuals believed on the possibility to obtain zero casualties on the streets. Such results require more consideration on how to change the public's belief on such a topic.

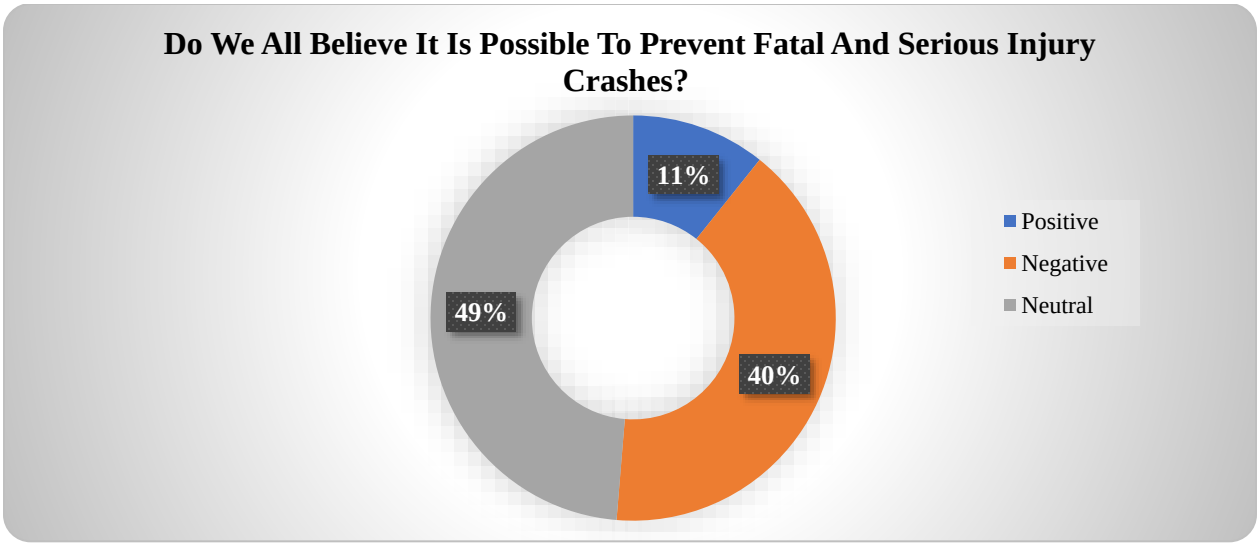
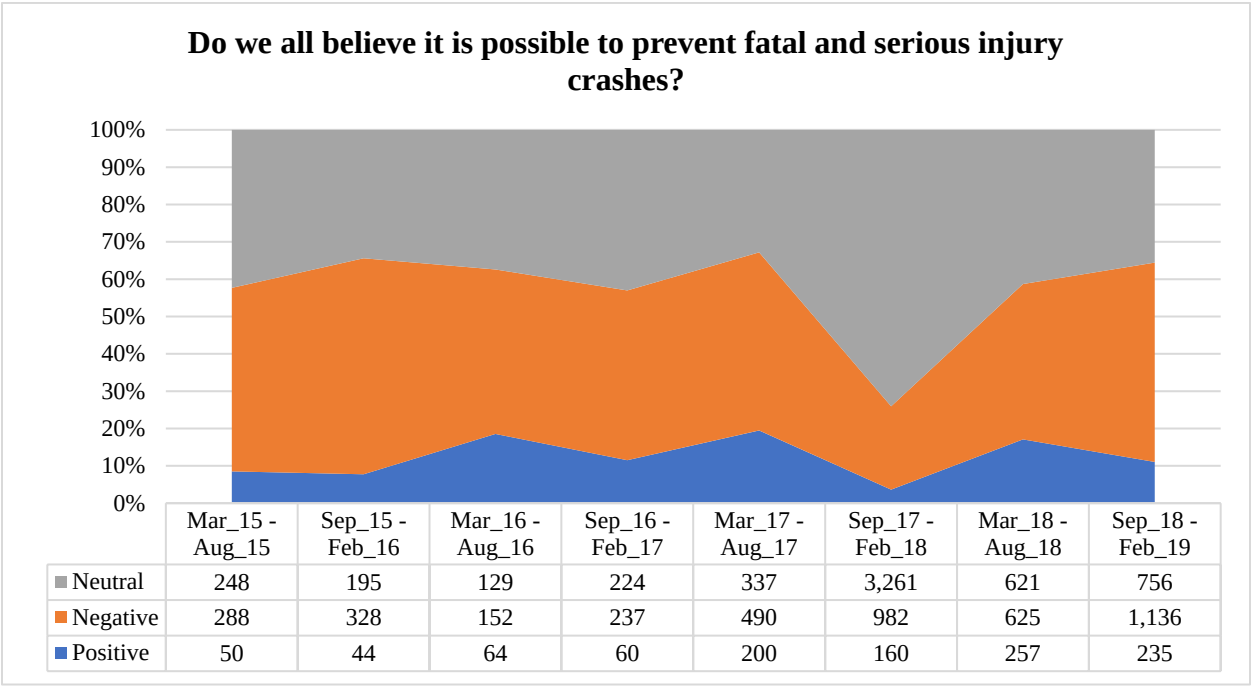


Figure 5. Sentiment for the question "Do we all believe it is possible to prevent fatal and serious injury crashes?"

Figure 6 presented the sentiment trend about the public's general beliefs on possibility of zero fatalities. The analysis revealed that the positive perceptions on possibly preventing fatal and serious injuries on road were constantly low throughout the period, while the amounts of negative and neutral percentions sentiments were always high. This trend implies cultural reflection of the public and might be explained as people didn't believe that traffic fatalities could be completely eliminated at this point and over past years.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	248	195	129	224	337	3,261	621	756
Negative	288	328	152	237	490	982	625	1,136
Positive	50	44	64	60	200	160	257	235

Figure 6. Sentiment Trend for the question "Do we all believe it is possible to prevent fatal and serious injury crashes?"

4.1.3 Attitude towards benefits of police enforcement of traffic laws

The graphical depiction of sentiment analysis on the topic "Do we all have the attitude that police enforcement of traffic laws is beneficial" was shown in Figure 7. The number of tweets collected for this topic was 23,997. The analysis showed that many people had a negative attitude towards the benefits of police enforcement of traffic laws. This might reflect emotional reactions of the public to the traffic law enforcement when they were just caught and got a ticket. It is well received that endeavors implemented by law enforcement agencies would, to a large extent, ensure safety on street; however, it could be futile if the public had insufficient consciousness of the traffic rules. Strategies for improvement might include constructive punishment and rewarding policies for following traffic rules. The points system could be a great tool for that. If any driver followed rules constantly for a period, they should be rewarded with positive points. The points could have some benefits like their vehicle insurance would be less costly. The example of constructive punishment could be that they pay more for every service that their car received. According to the analysis, the majority of the public is neutral about the benefits of traffic law enforcement. This calls for deep understanding and improvements on effects of promoting passive ways to make people better attentive of traffic laws. There would be needs for improvements of current strategies that make the communities more prudent and appreciative of traffic laws.

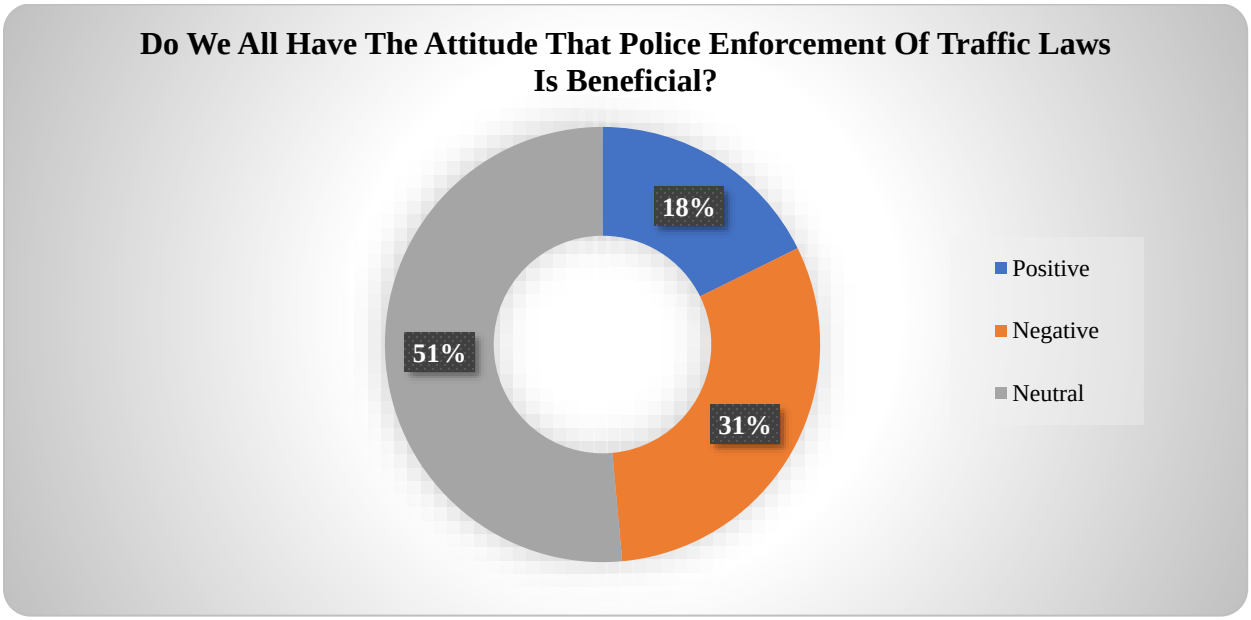
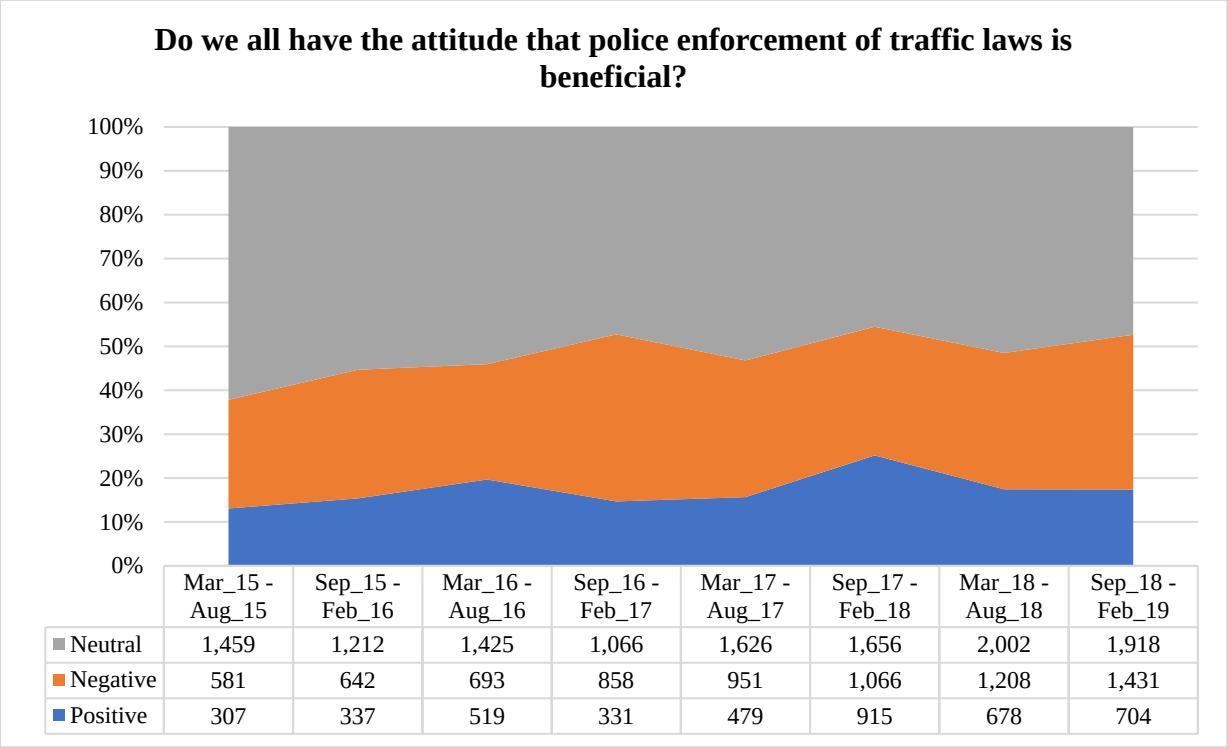


Figure 7. Sentiment for the question "Do we all have the attitude that police enforcement of traffic laws is beneficial?"

The trend analysis for the topic of the public's general attitudes towards police enforcement of traffic laws was showed in Figure 8. Over time, the positive attitudes were increasing. This could imply that people gradually appreciated the benefits that police enforcement of traffic laws had brought to them and the society. On the other hand, based on the figure, the negative attitudes were increasing. This reflected the challenges on increasing more people to be appreciative of the beneficial aspects of following traffic rules.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	1,459	1,212	1,425	1,066	1,626	1,656	2,002	1,918
Negative	581	642	693	858	951	1,066	1,208	1,431
Positive	307	337	519	331	479	915	678	704

Figure 8. Sentiment Trend for the question "Do we all have the attitude that police enforcement of traffic laws is beneficial?"

4.1.4 Six high-risk behaviors

The sentiment analysis was also used to undertand people's specific attitudes towards the six high-risk behaviors (impairment, speeding, distraction, unrestrained vehicle occupants, young and old driver involved). The result was presented in the following sections.

4.1.4.1 Impairment involved high-risk behavior

Among the six high-risk behaviors, impairment because of alcohol and drug is one of the most important. The number of tweets collected for this topic was 374,599. The graphical depiction of the sentiment analysis was provided in Figure 9. The results revealed that most people thought this high-risk behavior as a serious offense (i.e., negative). Only very few people accepted this practice (i.e., positive). Impaired drivers have been creating risks on the road. This riskly behavior was detrimental to the drivers because they could become prone to potential serious events. Drunk and drug-impaired drivers were frequently repeated offenders. Devices such as ignition interlocks were generally applied to hold them from driving. There were impairment detection tools like breath testing devices that have been applied to detect alcohol impairment. Legalization of drug could create a concern to traffic safety. Driving under influence (DUI) could result in harsh punishment, and the license could be revoked for a long time or life time. Also, refusing to take a breath test could have punishments. These activities showed that traffic safety agencies took this issue seriously, and law enforcements that attempted to prevent such risk behavior are appreciated by the public.

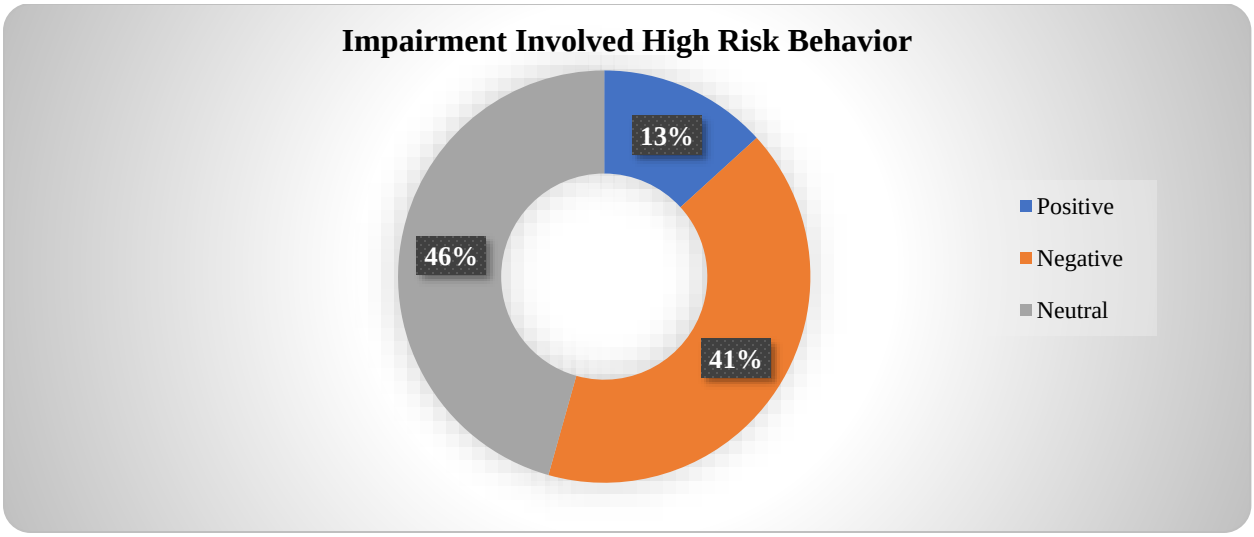
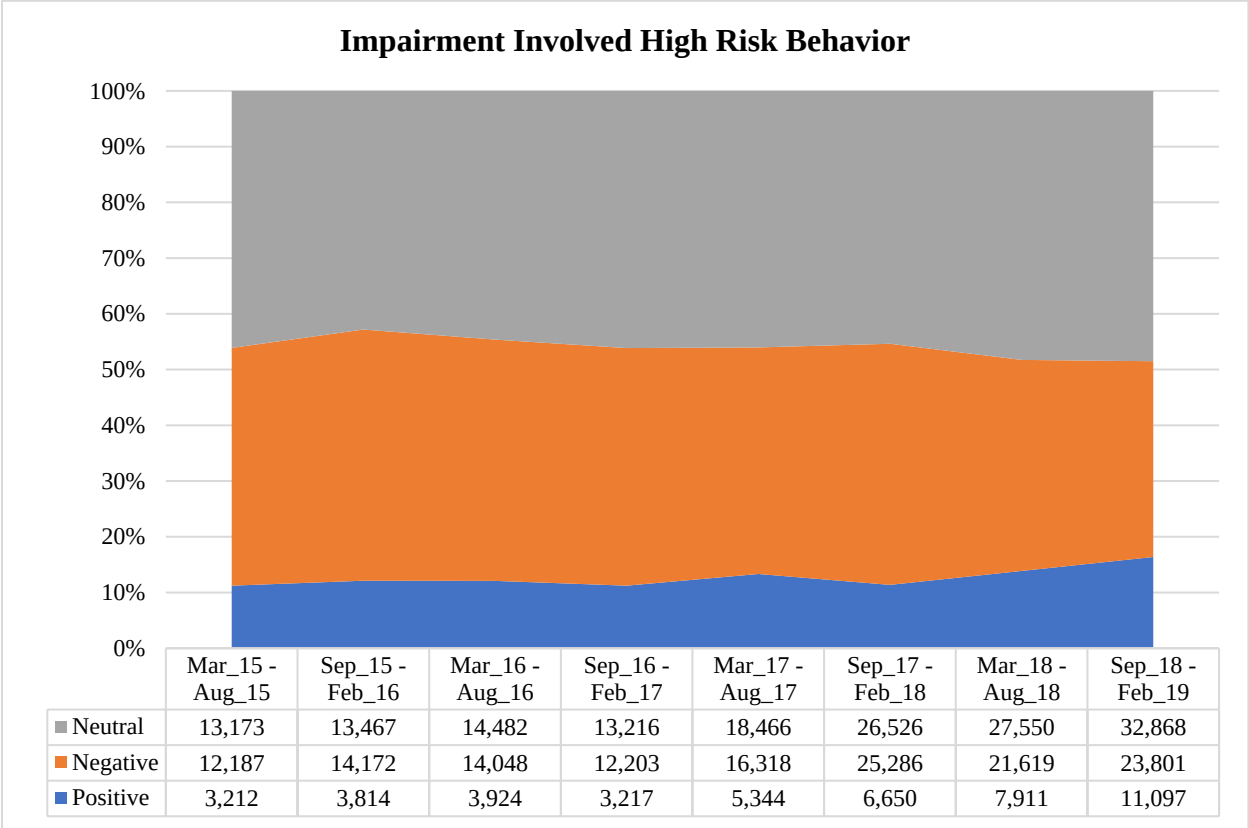


Figure 9. Sentiment for "Impairment involved high-risk behavior"

The trend analysis on the public's specific attitutes towards impairment involved high-risk behavior was displayed in Figure 10. It indicated that the public's attitudes towards this high-risk behavior were consistent over the time. People invariably took the impairement involved behavior as dangerous and were likely not to tolerate the impaired drivers on road.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	13,173	13,467	14,482	13,216	18,466	26,526	27,550	32,868
Negative	12,187	14,172	14,048	12,203	16,318	25,286	21,619	23,801
Positive	3,212	3,814	3,924	3,217	5,344	6,650	7,911	11,097

Figure 10. Sentiment trend for "Impairment involved high-risk behavior"

4.1.4.2 Speeding involved high-risk behavior

Speeding on the highway resulting in serious crashes is a common scenario. It is also one of the six high-risk behaviors. The graphical result of the sentiment analysis was shown in Figure 11, evaluated based on 139,566 tweets collected for this topic. Speeding is described as exceeding the set speed limit and running extremely fast. It is also called racing on the street or aggressive driving. Speeding had serious implication and can cause many crashes. It created danger for the driver and for the other people and property. In the analysis, it seems that most people had a neutral attitude towards this behavior. This is likely due to the reason that many people think that speeding is quite common in the street. This also reflects the fact that while the traffic law enforcement took measures, still many people didn't pay much attention to the speed limit on the street. Only very few people considered this behavior as high risk to traffic safety. More measures would be needed to improve the situation.

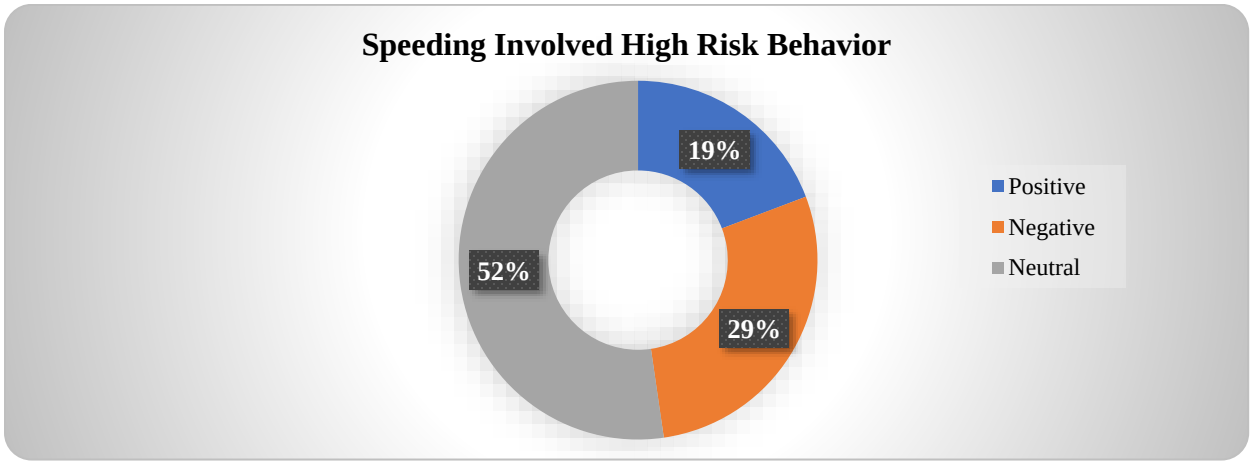
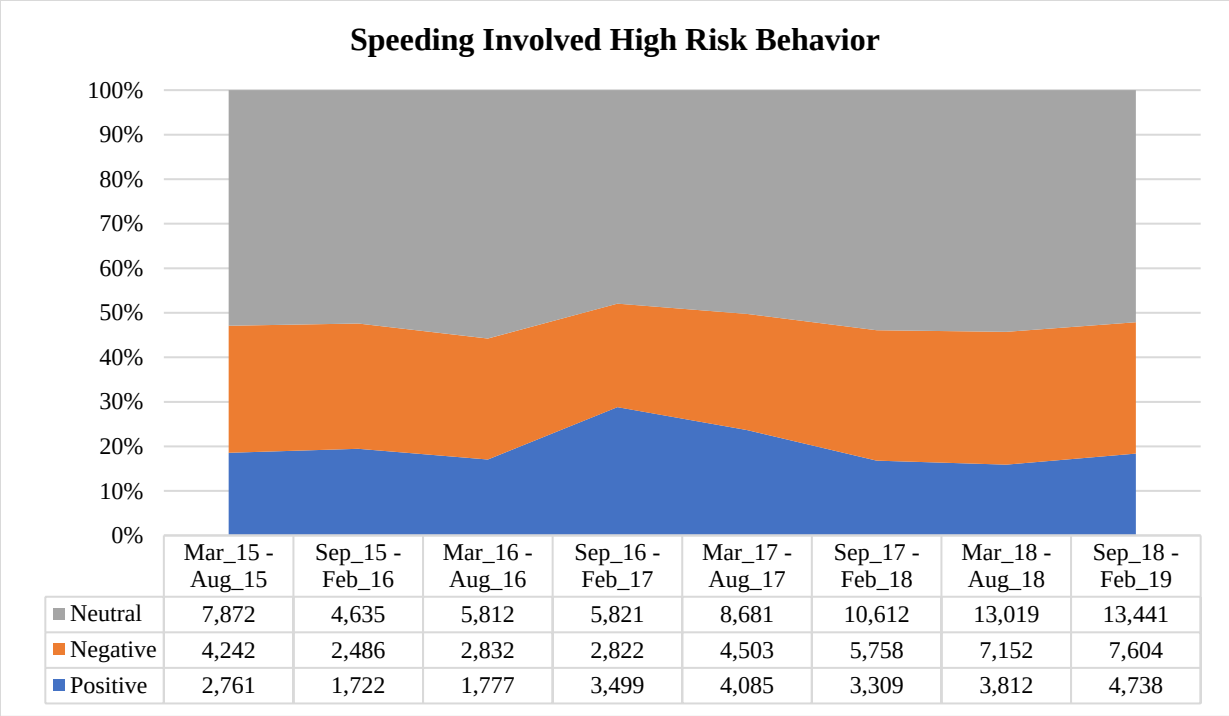


Figure 11. Sentiment for "Speeding involved high-risk behavior"

Figure 12 depicted the trend analysis regarding the public's attitudes towards the speeding involved high-risk behavior. It showed that most people had a neutral and negative sentiment about this high-risk behavior over time. It could be explained as many people are likely to overlook this high-risk behavior and took it as granted. Throughout the period of analysis, people had the similar level of attitude toward speeding. Though many crashes occurred because of this, people were not as concerned as about other risk behaviors and frequently drove more than the instructed speed limit. Law enforcement agencies might need to pay more attention to this issue. New policies might need to be developed. For example, with the help of technology, it is easy to detect speed via Google Maps or similar apps. The Blackbox installation in the vehicle is also a new technology to keep the track of a vehicle speed. By implementing these technologies, the speeding would be monitored and better controlled.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	7,872	4,635	5,812	5,821	8,681	10,612	13,019	13,441
Negative	4,242	2,486	2,832	2,822	4,503	5,758	7,152	7,604
Positive	2,761	1,722	1,777	3,499	4,085	3,309	3,812	4,738

Figure 12. Sentiment trend for "Speeding involved high-risk behavior"

4.1.4.3 Distraction involved high-risk behavior

Another significant high-risk behavior is distraction during driving. The number of tweets collected for this topic was 147,160. The graphical representation of the sentiment analysis was provided in Figure 13. This analysis showed that a certain portion of people did not think using cell phones or other distractive devices were harmful, regardless of the fact that exercises like texting and twitting during driving have caused many fatalities as a result of crashes. There should be more measures to get attention and awareness about severity of this high-risk behavior.

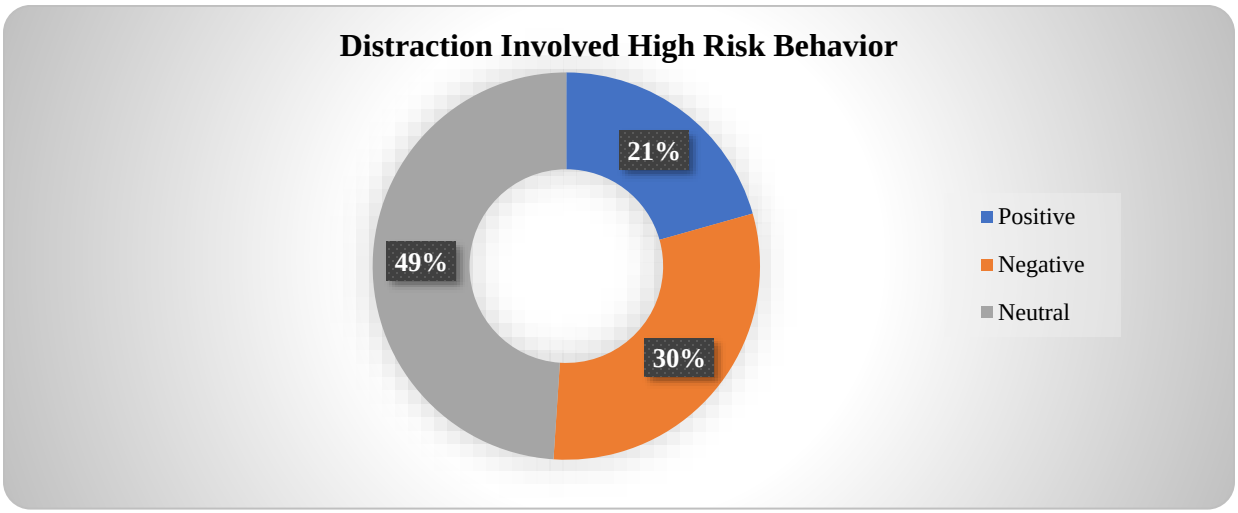
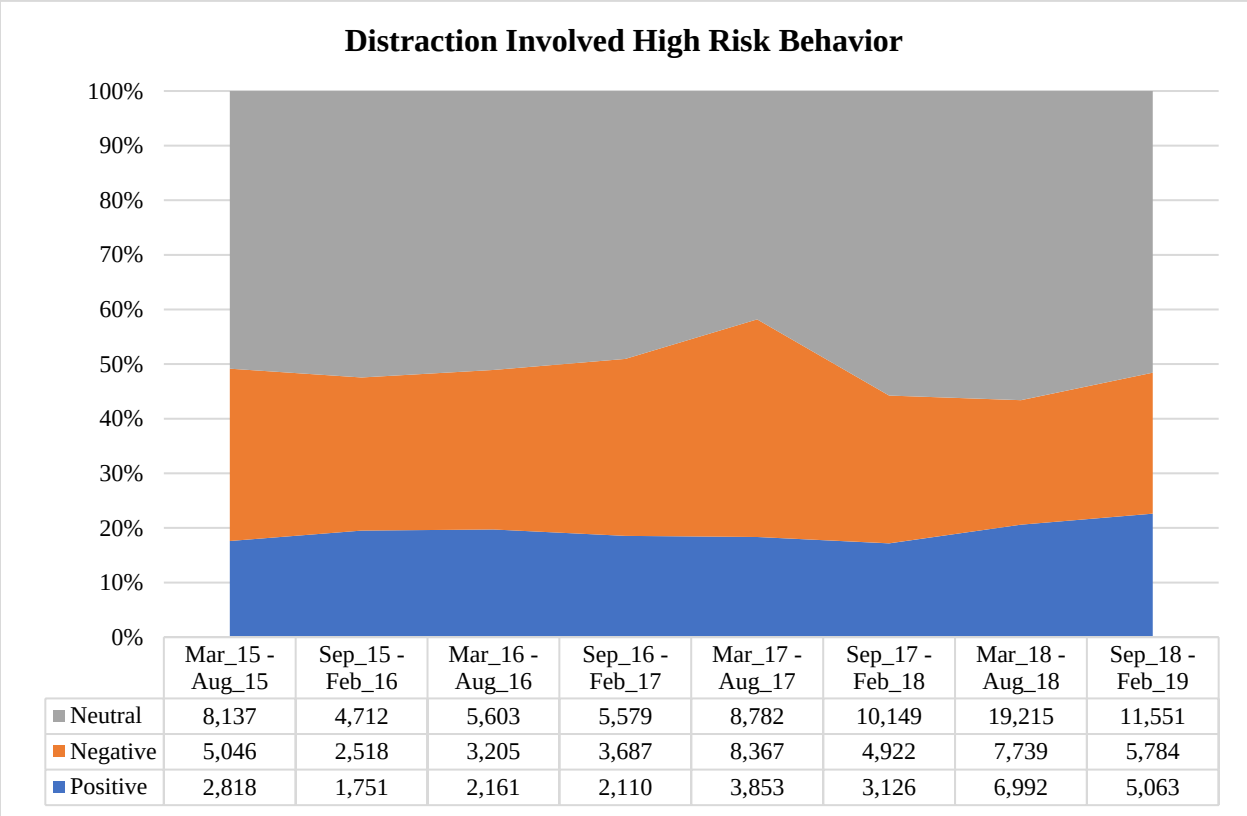


Figure 13. Sentiment for "Distraction involved high-risk behavior"

The trend analysis for the topic was illustrated in Figure 14. The result showed that people's attitudes towards the distraction involved high risk behavior was consistent over time. They did not believe that distraction is a serious issue. For example, many still think that using a phone is not a big risk while driving. The trend highlights the need for new methods to reduce and/or stop distraction involved driving on the road.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	8,137	4,712	5,603	5,579	8,782	10,149	19,215	11,551
Negative	5,046	2,518	3,205	3,687	8,367	4,922	7,739	5,784
Positive	2,818	1,751	2,161	2,110	3,853	3,126	6,992	5,063

Figure 14. Sentiment trend for "Distraction involved high-risk behavior"

4.1.4.4 Unrestrained vehicle occupants involved high-risk behavior

Figure 15 showed the sentiment analysis result for the public's attitudes towards the unrestrained vehicle occupants involved high-risk behavior. The number of tweets collected for this topic were 3,981. The result revealed that most people had a neutral and negative opinion about not using a seatbelt. The seatbelt is an occupant protection tool, and it is extremely important for reducing the damage of any severe crash. Using seatbelts is mandatory for the passenger and driver sitting in the front seat. The seatbelts can minimize the effect of any fatal crash. But still, there is room for improvements. In the result of the analysis, it appears that a few people still overlook the benefit of using the seatbelt. This demonstrates the need for traffic safety commissions to continue taking actions reminding and increasing awareness among the public about the usefulness of wearing a seatbelt. People need to wear seatbelts willingly and as a habit.

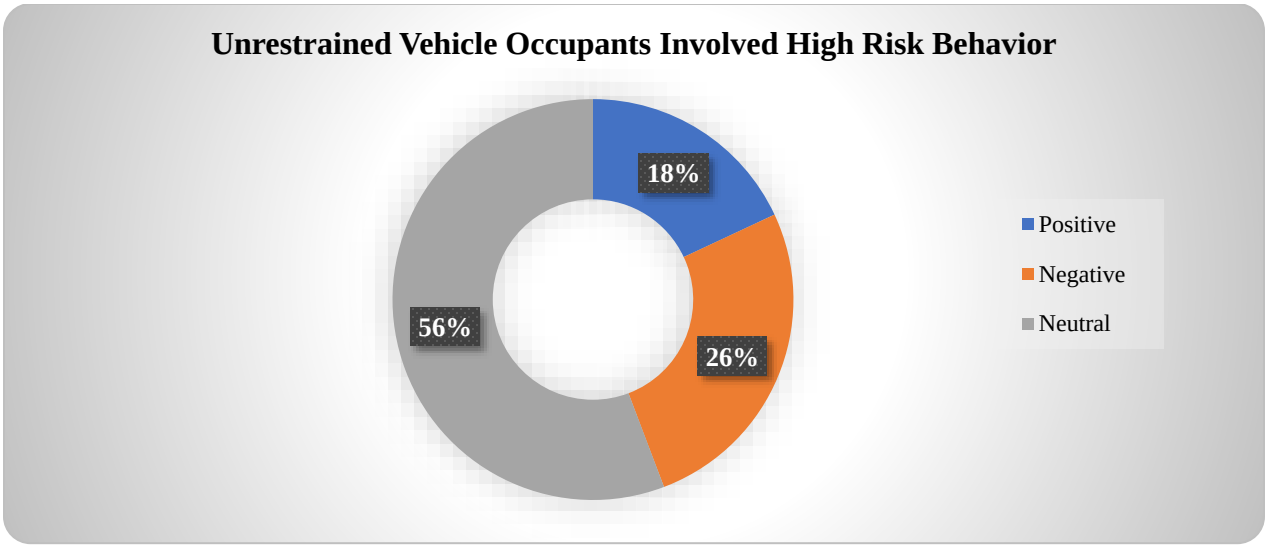
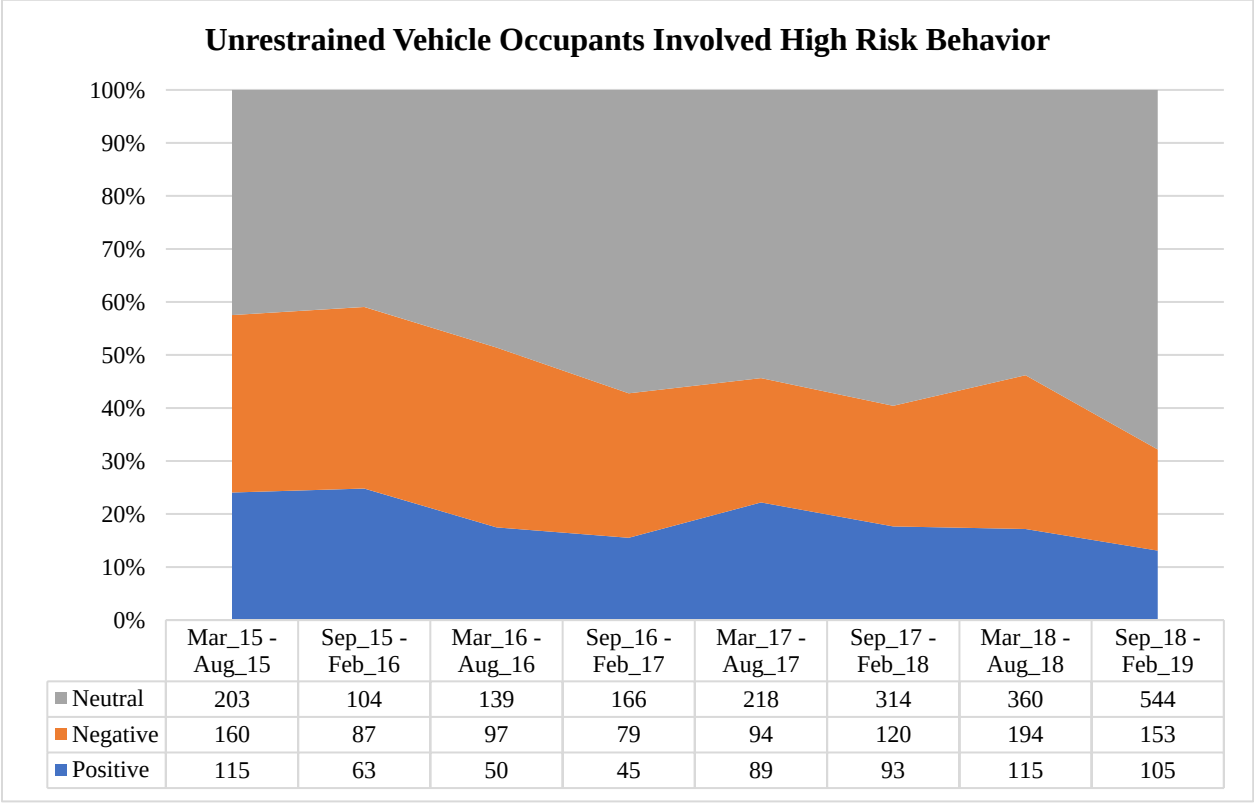


Figure 15. Sentiment for "Unrestrained vehicle occupants involved high-risk behavior"

The trend analysis was illustrated in Figure 16, according to which, people were getting less positive about not using seatbelts and child safety seats as time progressed. While this is a positive sign, it still required to improve the situation that the number of people who were negative about not using seatbelts and child safety seats are reducing.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	203	104	139	166	218	314	360	544
Negative	160	87	97	79	94	120	194	153
Positive	115	63	50	45	89	93	115	105

Figure 16. Sentiment trend for "Unrestrained vehicle occupants involved high-risk behavior"

4.1.4.5 Young driver involved high-risk behavior

The result of sentiment analysis was illustrated in Figure 17. For this topic, 42,976 tweets were collected. From the result, it seems that most of the public possessed neutral attitudes towards teenage driving. Around 21 precent considered that the young drivers were not harmful. This could reflect the public's impression about the phenomenon of teenage driving. However, the teen drivers were inexperienced. They could care little about the consequences and wanted to show their superiority by speeding. As a result, they could be prone to be victims of a road crash. Their risk-taking behaviors were common for the age they belong to. Their possible behaviors were speeding, consuming alcohols and drugs, and not wearing seatbelts. They were more vulnerable. Strategies like graduated driver licensing (GDL) reduced the risk and many states had already adopted this policy. They could learn about traffic rules if driving learning was intense and well managed. They could have positive points if they completed the learning phase without making any crash. The parents could also bring a great change if they could teach the process properly. Several practices could increase the influence of the parent to the teen driver. The teen drivers were new to driving and proper care could turn them into a good driver. The parents, school, and society together could improve the situation and hence increase the traffic safety.

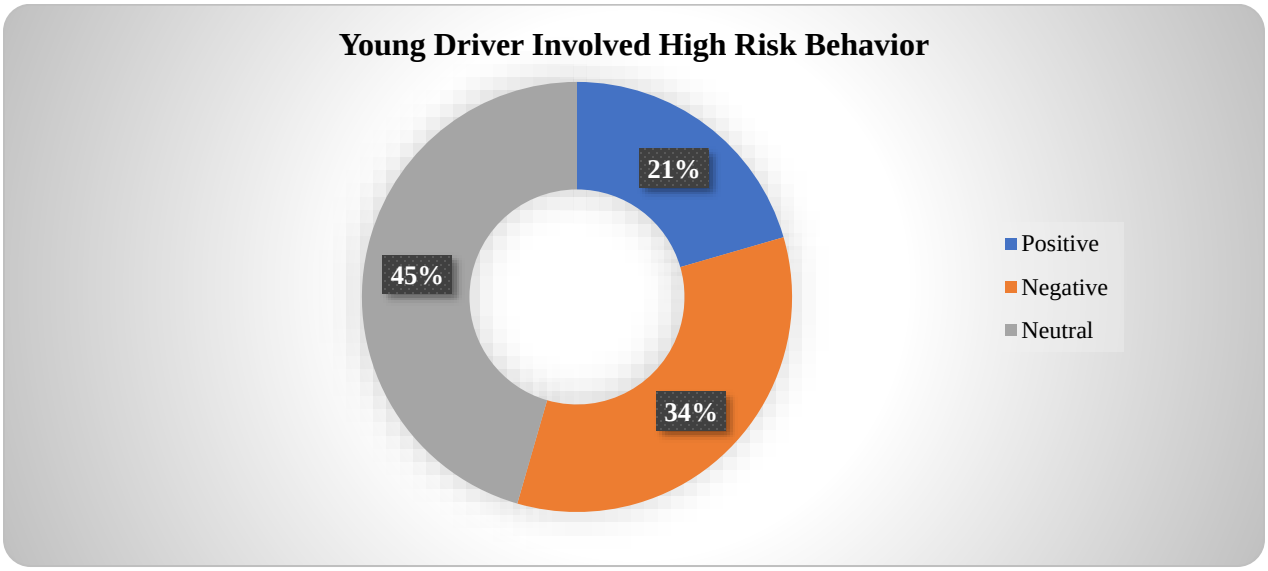
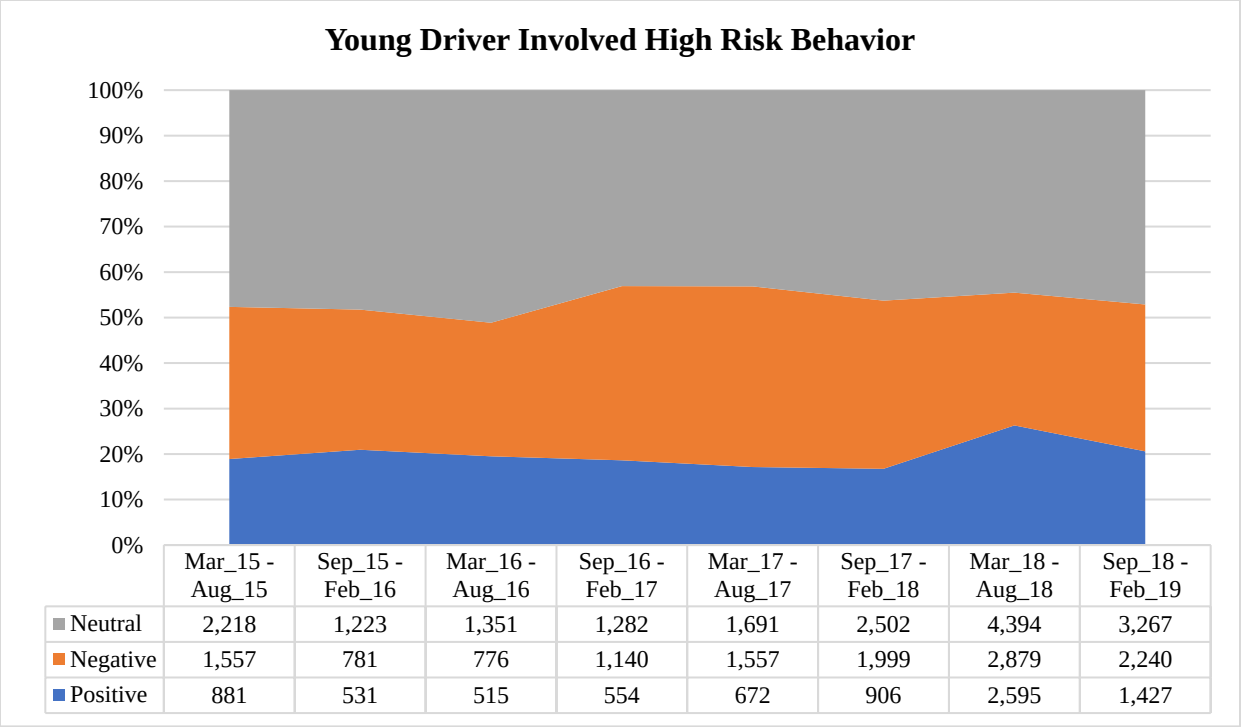


Figure 17. Sentiment for "Young Driver involved high-risk behavior"

The trend analysis for this topic was illustrated in Figure 18. The trend showed that the opinions of people regarding teen drivers were quite consistent over time.



	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	2,218	1,223	1,351	1,282	1,691	2,502	4,394	3,267
Negative	1,557	781	776	1,140	1,557	1,999	2,879	2,240
Positive	881	531	515	554	672	906	2,595	1,427

Figure 18. Sentiment trend for "Young Driver involved high-risk behavior"

4.1.4.6 Older driver involved high-risk behavior

The graphical representation of the sentiment analysis was provided in Figure 19. The number of tweets collected for this topic were 89,250. The precent of negative attitudes is about twice of the precent of positive attitudes, indicating that more people had concerns about older people driving. Older people usually have impairments with vision, cognition and motor function, which may hinder their ability in proper driving. Public bus services could be an option for older people to go side for their daily activities.

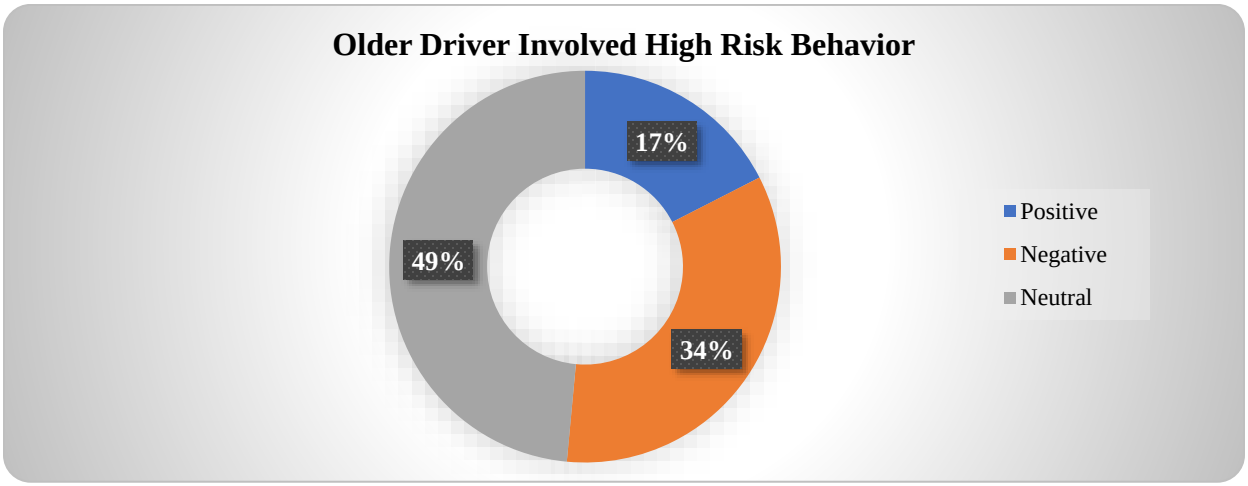
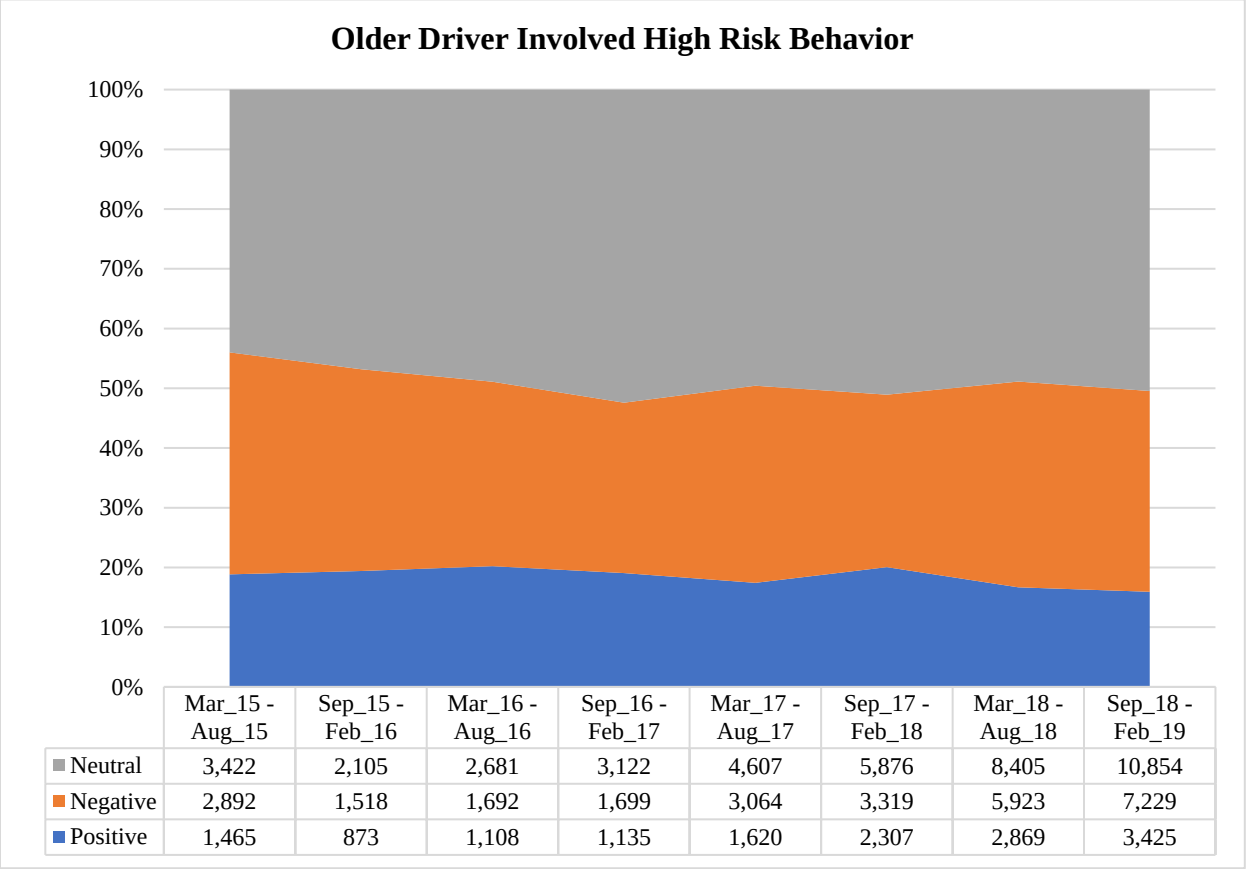


Figure 19. Sentiment for "Older driver involved high-risk behavior"

The trend analysis was showed in Figure 20. It was consistent over time. Most people had a neutral and negative attitude toward the risk of the older driver. There was no major change in the sentiment distribution.

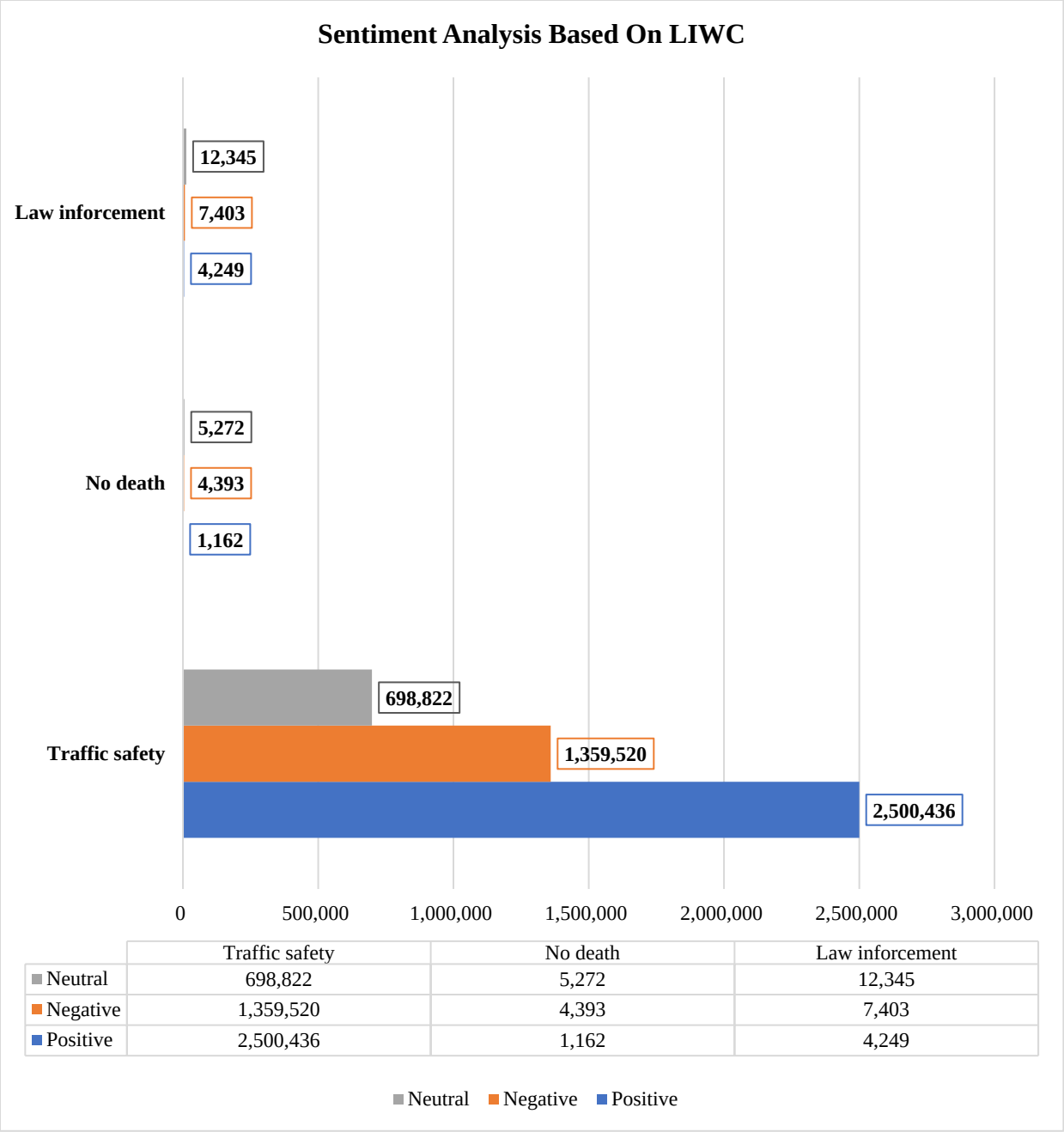


	Mar_15 - Aug_15	Sep_15 - Feb_16	Mar_16 - Aug_16	Sep_16 - Feb_17	Mar_17 - Aug_17	Sep_17 - Feb_18	Mar_18 - Aug_18	Sep_18 - Feb_19
Neutral	3,422	2,105	2,681	3,122	4,607	5,876	8,405	10,854
Negative	2,892	1,518	1,692	1,699	3,064	3,319	5,923	7,229
Positive	1,465	873	1,108	1,135	1,620	2,307	2,869	3,425

Figure 20. Sentiment trend for "Older driver involved high-risk behavior"

4.1.5 Comparison of results of attitudes to traffic safety, zero death and law enforcement

The result of the comparison of the public's attitudes to traffic safety importance, zero death possibilities, and benefits of law enforcement was illustrated in Figure 21. The result showed that most of the tweets were generated for the topic of the importance of traffic safety. The reason for that is, most of relevant keywords that were applied for the twitter data collection were related to traffic safety, which means most people on Twitter are concerned about the importance of traffic safety rather than other topics such as law enforcement's benefit or possibility of zero deaths.



	Traffic safety	No death	Law inforcement
Neutral	698,822	5,272	12,345
Negative	1,359,520	4,393	7,403
Positive	2,500,436	1,162	4,249

Figure 21. Sentiment analysis based on LIWC

4.1.6 Comparison of results of attitudes to six high-risk behaviors

For the attitudes to six high-risk behaviors on road, the data showed that people are mostly concerned about impairment-related high-risk behaviors from the number of tweets collected on this topic. The DUI always seemed to be a serious action from the public's perspective. Speeding and distractive driving are also two major high-risk behaviors that people often talked about on Twitter. The comprision results were presented in Figure 22.

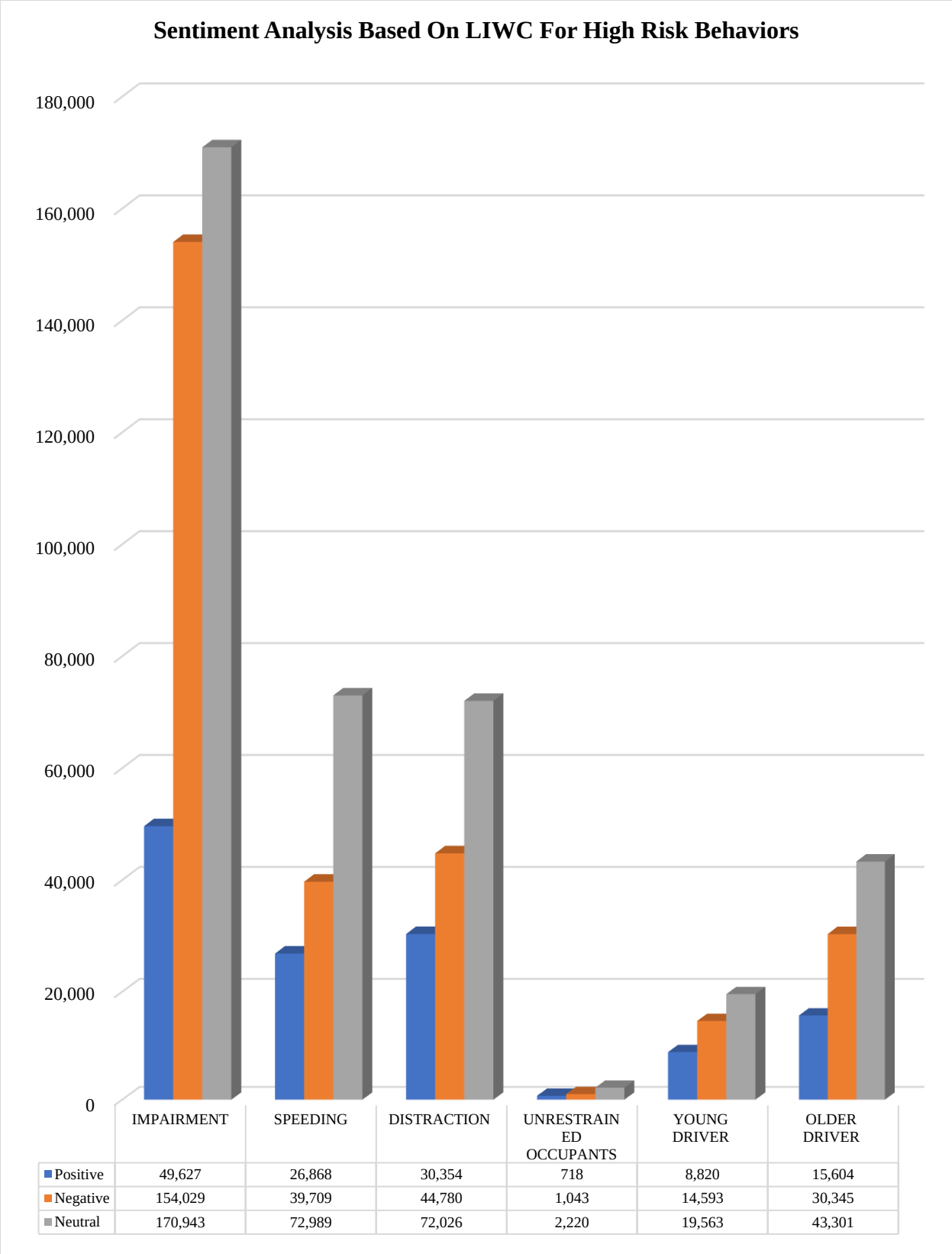


Figure 22. Sentiment analysis based on LIWC for six high-risk behaviors

4.2 Topic Modeling

The LDA technique has been utilized for the topic modeling in this project. For three traffic safety and six high-risk behavioral topics, five topics were generated with the preprocessed and cleaned data for each through applyinhg Python code. These topics were generated based on the interrelation among the keywords. Table 1 illustrated the topic modeling results. Potential latent ideas based on these topics and relationship of the keywords were also described in the implication column.

Table 1. Topic modeling with associated keywords

Q1. Do we all see traffic safety as an important issue for most people in our communities?		
Topic	**Keywords**	**Implication**
Safe driving	Light, driving, safe, driver, drive, vehicle, reported, accident, cars, drivers, stop, lane, police, dead, collision.	Traffic light, policing and traffic lane facilitate safely driving to reduce accidents and deaths on road.
Pedestrians and bike safety	Safety, crossing, road, safe, pedestrians, lanes, safely, safer, drive, killed, lane, bike, drivers, cyclists, mishap, traffic, pedestrian, street.	Crossing and specific lane are provided to ensure the safety of cyclists and pedestrians and to prevent a mishap.
wearing helmet	Helmet, bike, motorcycle, headset, ears, video, wear, riding, bicycle, ride, safety, wearing, helmets, lights.	Wearing helmet reduces and stops any injuries in head and provide safety for bike and motorcycle.
Traffic accident and road blockade	Lane, collision, crash, accident, lanes, right, blocking, traffic, left, blocked, near, just, cleared, update, open, vehicle, closed.	Traffic accidents block the traffic and actions are taken to clear the closed road and update the information in the system.
Traffic safety and law violation	Helmet, bike, bicycle, traffic, reported, beat, moving, violation, motorcycle, camera, action, super, waterproof, road, block.	Traffic safety equipment should be used to prevent traffic law violation and cameras help to detect the violation.
Q2. Do we all believe it is possible to prevent fatal and serious injury crashes?		
Topic	**Keywords**	**Implication**

Target zero	Zero, traffic, vision, deaths, target, drive, fatalities, driving, safety, patrols, targetzero, distracted, plan, driver, injuries, state, visionzero.	The Target Zero program aimed to make fatalities amount to zero by providing traffic safety and continues patrolling on the road.
Traffic crashes	Accident, cars, fatalities, injured, train, traffic, injuries, zero, derailed, roadway, Amtrak, struck, motorists, truck.	Traffic accidents and crashes result in injuries and fatalities on the roadway.
Effect of traffic accidents	Accident, traffic, injuries, road, targetzero, driving, crash, troopers, people, involved, update, drivers, good, deaths, tonight, crashes.	People's lives are involved in traffic crashes and people get afraid if accidents happen on road.
Deaths on road	Deaths, year, commercial, free, tonight, driving, safe, driver, uber, drunk, stop, lyft, just, accidents, traffic, police, death, road, scene.	Each year lot of people lost their lives on road and uber and other ride-sharing vehicles are parts safe driving.
DUI and accident	Accidents, arrests, tonight, rips, duis, jesus, traffic, accident, deaths, report, increase, road, just, driving, fatalities, like, says, driven.	DUI involvement can result in arresting and it is one of the main reasons for traffic accidents.

Q3. Do we all have the attitude that police enforcement of traffic laws is beneficial?

Topic	**Keywords**	**Implication**
Laws enforcement	Traffic, laws, enforcement, state, Washington, drivers, patrol, rules, people, follow, obey, cars, driving, cyclists, police, road, bike.	The Washington law enforcements are active to make people following traffic rules and reduce no of incidents.
Patrolling and cameras	State, Washington, patrol, cameras, police, trooper, stop, troopers, ignition, sign, crash, sobriety, light, checkpoint, interlock, killed.	Washington police patrolling and cameras are active in detects the sign of possible crash and ensure sobriety in the street.
Ticketing	Speed, cameras, ticket, camera, click, speeding, school, tickets, enforcement, parking, fine, driver, traffic, city, fines, running, license.	Speed camera and ticketing are an effective method of law enforcement to detect speeding and other issues in the road.

Ignition interlock to prevent the crash	Speed, cameras, camera, state, Washington, patrol, ignition, work, interlock, drivers, time, super, crash, county, taking, memo, date.	The ignition interlock is an effective system for preventing a crash.
clickitorticket	Camera, speed, drive, clickitorticket, sobriety, high, seat, pulled, checkpoint, belt, good, sober, video, speeding, campaign police	Clickitorticket is an effective campaign of police to proper using of seatbelts and checkpoint are used to detect speeding.
Impairment involved high risk behavior		
Topic	**Keywords**	**Implication**
Drunk drivers and braking laws	Driving, drunk, drivers, drinking, drug, people, impaired, marijuana, high, free, alcohol, laws, lead, tonight, getting, good, abuse, driver.	Alcohol consumption, using drug make people high and they are prone to breaking traffic laws.
DUI and after effect	Driver, arrested, police, drunk, dui, crash, injury, accident, suspected, charged, killed, alcohol, officer, arrest, lawyer, checkpoint, stop.	DUI is serious offense and police are on the street to arrest people with DUI by setting up checkpoints.
Imprisonment for DUI	Drive, drunk, drink, driver, like, court, just, beer, judge, home, want, family, good, jail, does, need, drinking, make, victim, ride, better, tonight, http, threw, fucking, wine, time, weekend, people, going.	DUI bring a serious charge against the driver and jail time can seriously impact the family of the driver.
Drunk driving kills people	Crash, drunk, woman, arrested, killed, driving, county, charged, http, driver, state, years, Washington, arrest, child, away, police, wife, fatal, year, time, charges, guilty, right, takes, husband, gets, trump.	Driving under the influence of alcohol can result in arresting and people get killed by these drivers.
Drugs hurt people	Driving, influence, beat, reported, http, block, fight, girl, cannabis, young, heart, weed, failed, breaking, marijuana, charged, woman, drive, year, aurora, courage, need, high,	Drugs have a serious impact on the driver and most of the time they fail to react properly on the street.

	black, trafficked, time, country, house, city, cops.	
Speeding involved high risk behavior		
Topic	**Keywords**	**Implication**
Speeding gets ticket	fast, speeding, drive, http, just, food, driver, train, driving, does, like, speed, crazy, night, tickets, police, faster, changes, really, furious, truck, bullet, long, ahead, life, drove, need, pull, high, tried.	Fast and speedy driving will pull the driver out of the road and results in getting a ticket.
Speed limit	speed, limit, drive, driving, speeding, drivers, time, people, driver, today, late, like, slow, fast, lane, just, poor, high, skills, dare, management, attention, care, going, focus, health, secret, advice, sessions, limits.	The driver should follow the speed limit.
Fast driving and pedestrians	driving, fast, speeding, speed, drivers, going, like, drive, driver, cars, people, think, pedestrians, self, road, student, make, courteous, need, getting, boat, actually, uber, police, just, conditions, trump, lane, traffic, america.	Fast driving cars create risk for pedestrians.
Pull over by police	speeding, ticket, just, http, pulled, like, know, driver, tickets, said, officer, driving, stop, police, time, getting, sure, wasn, today, fast, gets, woman, hard, think, change, didn, school, shit, going, dont.	Police pull over speeding vehicles.
Breaking laws	speeding, people, court, driver, ticket, news, drive, home, year, having, fast, stopped, saying, laws, live, night, freedom, school, caught, great, breaking, traffic, morning, light, state, going, white, dinner, happen, kept.	Speeding results breaking of laws and can drag the driver to court.
Distraction involved high-risk behavior		

Topic	Keywords	Implication
Distraction speeding getting ticket	Driving, distracted, http, speeding, drive, phone, texting, drivers, just, state, distracteddriving, message, washington, text, roadtodayticket, driver, know, marketing, need, video, police, crash, black, time, like, effect, life, pledge.	Distracted driver fails to follow the speed limit and get a ticket. It also causes a crash.
Using mobile while driving	drive, driving, fast, mobile, phone, text, calling, people, http, like, know, just, think, speed, friends, okay, mother, clear, good, texting, need, Muslim, terror, message, likes, typing, function, hesitation, microphone, vshphoob.	Driver using a mobile phone while driving cause distraction and they think it is ok. It is a wrong perception and cause of several crashes.
Distraction cause speeding	speeding, phone, driving, drive, just, people, message, driver, right, speed, calling, drove, ticket, away, make, ride, police, like, said, drivers, stop, white, road, told, tickets, alabama, texting, single, school.	Driver using the phone on driving cause speeding and failed to follow the speed limit.
Uber and lyft drivers	driver, phone, driving, drivers, uber, speed, just, http, cell, calling, drive, mobile, texting, using, truck, limit, drove, stop, lyft, cars, talking, high, traffic, need, phones, left, time, pedestrians, street, license.	Ridesharing vehicle drivers always engaged with the phone for their pickup of the customer and get distracted.
Drink driving and texting	drive, text, driving, texting, safe, drivers, people, drink, things, home, like, distracted, coming, says, reminder, mobile, dont, fucking, message, calling, date, distracteddriving, years, vote, girl, kids, just, friends, used, wait.	Texting and drinking drivers get highly distracted.
Unrestrained vehicle occupants involved high-risk behavior		
Topic	**Keywords**	**Implication**

Seatbelt increase survival possibility	seatbelt, wearing, seat, http, child, pulled, trip, lean, moon, wear, forh, online, gujfh, like, month, duty, nicole, survive, orleans, crash, info, cutter, heavy, cables, automobile, jumper, hammer, fireextinguisher, emergency.	Wearing seatbelts increase survival chance and decrease crash fatalities.
Child restraint seat	seatbelt, wearing, driving, wear, unrestrained, things, drive, year, child, http, people, seat, speed, ride, children, remember, news, fast, looks, month, roller, coaster, appear, buckleup, fake, desperate, regai, know, limit.	The vehicle pulled driver due to unrestrained child seat.
Seatbelt and safety	seatbelt, wear, wearing, mlkzo, just, fasten, like, seat, time, says, know, going, didn, right, safety, accident, http, passenger, wasn, need, good, doesn, life, getting, died, think, riding, buckleupamerica, save, superman.	Wearing seatbelt increase safety and reduce death by accident.
Seatbelt and getting ticket	seatbelt, wearing, wear, driver, today, truck, just, said, ticket, lots, pick, tread, colorado, stickers, like, life, crash, break, woman, lost, does, best, shows, friend, people, passenger, save, mother, unrestrained, kids.	Not wearing seatbelts will result getting a ticket.
Seatbelt is security	seatbelt, emergency, travels, security, http, zbqzunbq, info, adds, great, morning, road, unrestrained, yesterday, greater, change, windshield, extinguisher, emergencykit, auto, wreck, fireextinguisher ditch, flipped, slid, classmates, climate, said, child, problem, remember.	The seatbelt is an emergency for traveling security.
Young driver involved high risk behavior		
Topic	**Keywords**	**Implication**
Teen driver distraction and fatalities	teen, driving, driver, young, http, drivers, black, safety, safe, person, white, teens, girl, work, parents, trump,	Teen driver often get distracted and cause serious crash. It can result death of people.

	road, drive, said, help, crashes, distracted, crash, uber, week, death, dude, father, teendriver, people.	
Young driver, family and policing	young, drive, driver, teen, killed, driving, parents, nearly, daughter, family, uber, drivers, police, home, year, kids, shooting, kindergarten, http, radio, charged, shot, drunk, seen, like, drove, crash, murder, says, traffic.	Young driver can cause serious harm to the safety of family and frequently get charged by police.
Drinking and getting killed	teen, drive, young, drivers, driver, just, drunk, killed, year, trust, junior, drinking, driving, friend, time, jail, want, teenage, going, drove, weekend, crash, cool, number, seattle, americans, prison, accidents, night.	Teen driver frequently gets drunk and jailed for that.
Future of driver	young, driving, driver, people, just, stop, woman, better, world, future, africa, wait, black, talented, girl, friends, create, spend, drove, good, drivers, http, doing, going, video, goes, teenage, lady, thinking, dumb.	Reckless driving in young age can destroy the future life of the driver.
Drug and young people	young, people, driving, drive, drivers, deaths, teenage, change, marijuana, scary, like, stats, attributed, annually, excessive, drove, state, great, women, money, family, american, make, vote, native, saved, brothers, tour, colorado, today.	Young people with drug addiction is a scary thing for the safety on the road.
Older driver involved high risk behavior		
Topic	**Keywords**	**Implication**
Older driver and checking driving	drive, year, driving, shot, shooting, family, people, older, killed, hard, kids, http, really, just, drove, school, check, like, work, trump, want, truck, near, cars, fuck, girl, jazmine, story, fucking, time.	The older driver should get frequently checking for driving to understand their ability to driving.

Senior driver and drinking	year, driving, driver, drive, senior, just, stop, http, drivers, like, older, drove, good, school, police, father, watch, home, past, drunk, child, away, says, need, time, goes, speed, imagine, america, ride.	Senior people with an alcohol problem can harm people on road.
Senior people like driving	drive, year, years, driver, http, like, driving, today, drivers, drove, just, people, male, know, arrested, white, older, woman, said, women, good, killed, american, girl, time, music, want, cars, sexually, assaulted.	Senior people get pleasure while driving on the road.
Older drivers and night accidents	year, driver, driving, drive, older, drove, police, just, seen, missing, street, help, baby, drivers, white, woman, night, crash, years, breaking, shit, http, senior, killed, away, gonna, seattle, people, tesla, held.	Older drivers have low vision and not good in night driving. It causes crash and death.
Senior driver, crash and killing	driver, driving, year, thing, killing, uber, wish, crashed, open, infant, month, headline, duty, years, talking, school, stupid ,home, girl, http, wait, trash, drivers, killed, door, kinda, relationships, kids, license, romantic.	Senior drivers are responsible for crashes and killing on street.

5. DISCUSSION AND CONCLUSION

In this project, the social media mining techniques of sentiment analysis and topic modeling were applied to reflect the traffic safety culture in the state of Washington. As a representative example of this culture, tweet like “Take traffic deaths seriously - is one of the most important things that can be addressed in local politics” showed the public conceive the necessary awareness and regard safety as an important topic of their ways of life and others. Likewise, “If I’m in the car, everyone is wearing a seatbelt. Not up for debate” demonstrated that people are trying to be sincere about wearing a seatbelt to increase the safety and reduce the level of damages in case of unfortunate event. Moreover, the engagement of the law enforcement department can have a better idea from the tweet “Man charged for DUI. Investigations ongoing into an allegation of a linked traffic accident”. It showed their dedication to ensure safe roads and make the belief that the traffic law system is beneficial for every individual in the community. In addition, tweeting from traffic safety commission “Target Zero adds a guide for safety: Parent engagement reduces the risk for teens in cars” can help to make parents more cautious about the safety of their child that will finally benefit to reduce the number of traffic casualties and injuries to zero. Their endeavors in social media may lead to positive cultural changes in traffic safety culture. For instance, the Click It or Ticket program in North Carolina produced in rising seat belt usage from 63% to 80% and lowered rates of highway deaths and the success story constructed a model for other state and nationwide programs (Williams et al. 1996). This activity helped to boost up the seatbelt usage from 83% up to 95% (Salzberg and Moffat 2004).

Social media activities can build consciousness among the public about the constructive impact of following traffic laws and support to make the stakeholders more concerned about the safety of each other. People cannot be forced to reach a goal unless they realize the route of reaching that goal. Information about traffic safety should be available in social media in a constructive and lucrative way. People can understand traffic safety and relevant programs or campaigns from social media activities. If people are well informed about the policies of the traffic safety commission, they can act as a volunteer and can help to assure the success of that policy. From social media, people can learn about the aim of ‘Target Zero’ and can understand the phases associated with it. If they well acknowledge about the aim and process, they can relay the safety messages by retweeting and sharing. Their message can help to make other people aware of the traffic safety goal and probably in the future they will also relay the message to more people. In this way, they can prevent any unfortunate event and make their family and friends safe. It can also help to lessen the damage and can decrease the severity of any tragic incidents.

The most interesting part of the result was the positive response about importance of traffic safety. Most people in the Washington have accepted that traffic safety as an important issue. This positive attitude towards traffic safety will make the pathway smoother to reach the goal of ‘Target Zero’. The responsible stakeholders of the community can influence the whole community. Their act will setup a standard that can certainly help to bring the positive change in any culture. Also, the response of the people for the impaired driving can kindle the hope of reducing the number of causalities to zero. This approach can establish a healthy atmosphere for traffic. The long-term

effect of a good traffic safety culture can support wiping out the economic and time loss on the street. After all, these safety actions will contribute to societies and people.

The data collected from the Twitter was huge. There are different perspectives that can be generated in future work. Topic modeling gave some vague ideas about commonly discussed topic related to traffic safety. In future work, more analysis will sort out more interesting insights about traffic safety culture from this enormous data.

Limitations

There are limitations in this research. The first is that Twitter filtered information regarding user demographics when releasing data, because of its privacy protection policy to keep the user information confidential. Information such as age and gender that could have generated more insights was not available. The analysis of these insights might introduce innovative policies to improve the current status of traffic safety culture by targeting each specific group of people. The analysis in this research was only conducted on an overall situation of the traffic safety culture for the residents of Washington.

The second limitation is that the tweet data was non-geotagged. While this research intended to collect both geotagged and non-geotagged tweets in Washington, the geotagged tweets can only be collected from GPS-enabled devices that provide the exact location information of the user. Also, the users have full control of whether their tweets are geotagged or not, and the geotagging option in Twitter is off by default. So the majority of the data were not geotagged. If the majority of the data were geotagged, they could generate more interesting insights about the cluster of people based on location. The analysis could have differentiated people in several groups such as urban, suburban, and rural areas. The analysis could have also helped to understand the specific traffic safety issues in specific areas. Such analysis was not done in this research due to lack of geotagged data.

Lastly, this study applied the technique of SentiStrength to estimate the strength of positive and negative sentiments in the short texts. SentiStrength was based on the algorithm that rely on emotions, showing social support or as part of online arguments. Generated sentiment values based on this algorithm, tweets with an equal negative and positive values were regarded as neutral sentiments. A large volume of tweets were regarded as neutral sentiments in this research. This might be because many tweets posted were mainly about describing a fact, making an announcement, carrying out a campaign that did not carry explicit emotions or the tweets happened to have the same positive and negative strengths. Examples of such tweets include: “Teens are truly at risk while driving distracted. That’s why we’re seriously addressing the issue”; “Good Morning! Just a friendly reminder: use caution when traveling today as roads could be icy from last night’s cold temps”; “The driver of the Lumina, which hit the Jeep, will be cited for running a red light, officer Hank Paulson said. That driver also not injured”; “Apparently, she needs a new bike helmet”; “Those days when you get every red light while driving”; “I just blatantly blew a red light. I’m turning into a Florida driver. Oh boy”; “It takes a special kind of stupid to ride a motorcycle up US 59 without a helmet”. Nevertheless, the tweets that resulted in neutral sentiments are still valuable since they are relevant to traffic safety and to large extent reflect

attention and focus of the public, which could be further explored through more comprehensive investigations.

REFERENCES

Akram, W., and Kumar, R. (2017). "A study on positive and negative effects of social media on society." *International Journal of Computer Sciences and Engineering*, 5, 347-354.

Albuquerque, F. C., Casanova, M. A., Lopes, H., Redlich, L. R., de Macedo, J. A. F., Lemos, M., de Carvalho, M. T. M., and Renso, C. (2016). "A methodology for traffic-related Twitter messages interpretation." *Computers in Industry*, 78, 57-69.

Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., Guo, Y., Prosperi, M., Salloum, R., and Modave, F. (2016). "Mining Twitter to assess the public perception of the "Internet of Things"." *PloS one*, 11(7), e0158450.

Cobb, N. K., Graham, A. L., Byron, M. J., and Abrams, D. B. (2011). "Online social networks and smoking cessation: a scientific research agenda." *Journal of medical Internet research*, 13(4), e119.

Conner, M., Lawton, R., Parker, D., Chorlton, K., Manstead, A. S., and Stradling, S. (2007). "Application of the theory of planned behaviour to the prediction of objectively assessed breaking of posted speed limits." *British Journal of Psychology*, 98(3), 429-453.

Crossley, S. A., Kyle, K., and McNamara, D. S. (2017). "Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis." *Behavior research methods*, 49(3), 803-821.

D'Andrea, E., Ducange, P., Lazzerini, B., and Marcelloni, F. (2015). "Real-time detection of traffic from twitter stream analysis." *IEEE transactions on intelligent transportation systems*, 16(4), 2269-2283.

Edwards, J., Freeman, J., Soole, D., and Watson, B. (2014). "A framework for conceptualising traffic safety culture." *Transportation research part F: traffic psychology and behaviour*, 26, 293-302.

El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., and Kobi, A. (2018). "A novel adaptable approach for sentiment analysis on big social data." *Journal of Big Data*, 5(1), 12.

Flin, R., Mearns, K., O'Connor, P., and Bryden, R. (2000). "Measuring safety climate: identifying the common features." *Safety science*, 34(1-3), 177-192.

Hedlund, J. (2007). "Improving Traffic Safety Culture in the United States-The Journey Forward: Summary and Synthesis."

INSAG, I. N. S. A. G. (1986). *Summary Report on the Post-accident Review Meeting on the Chernobyl Accident*, INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.

Kahn, J. H., Tobin, R. M., Massey, A. E., and Anderson, J. A. (2007). "Measuring emotional expression with the Linguistic Inquiry and Word Count." *The American journal of psychology*, 263-286.

Li, W., Gkritza, K., and Albrecht, C. (2014). "The culture of distracted driving: Evidence from a public opinion survey in Iowa." *Transportation research part F: traffic psychology and behaviour*, 26, 337-347.

Lum, H., and Reagan, J. A. (1995). "Interactive highway safety design model: accident predictive module." *Public Roads*, 58(3).

Mai, E., and Hranac, R. (2013). "Twitter interactions as a data source for transportation incidents."

Mäntylä, M. V., Graziotin, D., and Kuutila, M. (2018). "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers." *Computer Science Review*, 27, 16-32.

NHTSA, N. H. T. S. A. (2017). "Fatal motor vehicle crashes: overview." *Washington, DC: US Department of Transportation*.

Panchal, S. G., and Apare, R. S. (2016). "A Survey on Traffic Detection from Twitter Tweets Analysis." *International Journal of Science and Research (IJSR)*, 5(12), 832-835.

Pang, B., and Lee, L. (2008). "Opinion mining and sentiment analysis." *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.

Paul, M. J., and Dredze, M. "You are what you tweet: Analyzing twitter for public health." *Proc., Fifth International AAAI Conference on Weblogs and Social Media*.

Rakauskas, M. E., Ward, N. J., and Gerberich, S. G. (2009). "Identification of differences between rural and urban safety cultures." *Accident Analysis & Prevention*, 41(5), 931-937.

Salzberg, P. M., and Moffat, J. M. (2004). "Ninety five percent: an evaluation of law, policy, and programs to promote seat belt use in Washington state." *Journal of Safety Research*, 35(2), 215-222.

Sievert, C., and Shirley, K. "LDAvis: A method for visualizing and interpreting topics." *Proc., Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 63-70.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). "Lexicon-based methods for sentiment analysis." *Computational linguistics*, 37(2), 267-307.

Tausczik, Y. R., and Pennebaker, J. W. (2009). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Language and Social Psychology,Volume: 29 issue: 1*, 24-54.

Tausczik, Y. R., and Pennebaker, J. W. (2010). "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of language and social psychology*, 29(1), 24-54.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). "Sentiment strength detection in short informal text." *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. "Predicting elections with twitter: What 140 characters reveal about political sentiment." *Proc., Fourth international AAAI conference on weblogs and social media*.

Turney, P. D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." *Proc., Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, 417-424.

Ward, N., Otto, J., Swinford, S., and Borkowski, J. (2015). "Measuring Minnesota's traffic safety culture." *Western Transportation Institute, Montana State University, Bozeman*.

Ward, N. J., Linkenbach, J., Keller, S. N., and Otto, J. (2010). "White paper on traffic safety culture." *White Paper*, 2.

Ward, N. J., Otto, J., and Linkenbach, J. (2014). "A primer for traffic safety culture." *ITE Journal*, 84(5), 41-47.

Williams, A. F., Reinfurt, D., and Wells, J. K. (1996). "Increasing seat belt use in North Carolina." *Journal of Safety Research*, 27(1), 33-41.

WTSC, W. T. S. C. (2013). "Washington State Strategic Highway Safety Plan 2013." Washington State Department of Transportation.

WTSC, W. T. S. C. (2016). "Washington State Strategic Highway Safety Plan 2016: Target Zero."

WTSC, W. T. S. C. (2018). "Washington State 2018 Traffic Safety Annual Report."

Zhou, Y., De, S., and Moessner, K. (2016). "Real world city event extraction from Twitter data streams." *Procedia computer science*, 98, 443-448.

APPENDIX A

THREE SETS OF KEYWORDS USED IN PHASE 1 SENTIMENT ANALYSIS

Keywords for Q (1): Do we all see traffic safety as an important issue to most people in our communities?

1.	ROAD	SAFETY
2.	ROAD	SAFE
3.	ROADWAY	SAFER
4.	ROADWAYS	SAFELY
5.	TRAFFIC	
6.	PEDESTRIAN	
7.	PEDESTRIANS	
8.	DRIVE	
9.	DRIVES	
10.	DRIVER	
11.	DRIVERS	
12.	DRIVING	
13.	HIGHWAY	
14.	LANE	
15.	LANES	
16.	CROSSING	
17.	PEDESTRIAN	RED LIGHT
18.	PEDESTRIANS	GREEN LIGHT
19.	DRIVE	YELLOW LIGHT
20.	DRIVES	
21.	DRIVER	
22.	DRIVERS	
23.	DRIVING	
24.	TRAFFIC	ACCIDENT
25.	ROAD	ACCIDENTS
26.	ROADS	COLLISION
27.	LANE	COLLISIONS
28.	LANES	CRASH
29.	HIGHWAY	CRASHES
30.	VEHICLE	CRASHING
31.	VEHICLES	CRASHED
32.	CAR	FATALITY
33.	CARS	FATALITIES
34.	BUS	INJURY
35.	AUTOMOBILE	INJURIES

36.	AUTOMOBILES	DEAD
37.	TRUCK	DEADS
38.	SUV	DEATH
39.	VAN	DEATHS
40.	TAXI	CASUALTY
41.	CAB	CASUALTIES
42.	MOTORBIKE	KILL
43.	MOTORCYCLE	KILLED
44.	MOTORCYCLIST	NEGLIGENCE
45.	MOTORCYCLISTS	NEGLIGENT
46.	PEDESTRIAN	
47.	PEDESTRIANS	
48.	INTERSECTION	
49.	LANE DEPARTURE	
50.	AGGRESSIVE	DRIVER
51.	AGGRESSIVELY	DRIVERS
52.	RISK	DRIVING
53.	RISKY	
54.	UNSAFE	
55.	UNTRAINED	
56.	ILLEGAL	
57.	NEGLIGENT	
58.	NEGLIGENCE	
59.	RECKLESS	
60.	ROAD	MISHAP
61.	STREET	
62.	HIGHWAY	
63.	BICYCLE	HELMET
64.	BIKE	
65.	MOTORCYCLE	
66.	DEADLY TRAFFIC	
67.	TRAFFIC	OFFENSE
68.		OFFENSES
69.		OFFENDER
70.		OFFENDERS
71.		VIOLATION
72.		VIOLATOR
73.		VIOLATORS
74.	#ROADSAFETY	
75.	#ROADSAFE	
76.	#ROADSAFER	

77.	#ROADSAFELY
78.	#ROADSSAFETY
79.	#ROADSSAFE
80.	#ROADSSAFER
81.	#ROADWAYSAFETY
82.	#TRAFFICSAFETY
83.	#TRAFFICSAFE
84.	#PEDESTRIANSAFETY
85.	#PEDESTRIANSAFE
86.	#PEDESTRIANSSAFETY
87.	#DRIVESAFETY
88.	#DRIVESAFE
89.	#DRIVESAFER
90.	#DRIVESAFELY
91.	#DRIVESSAFE
92.	#DRIVESSAFELY
93.	#DRIVERSAFETY
94.	#DRIVERSAFE
95.	#DRIVERSAFER
96.	#DRIVERSAFELY
97.	#DRIVERSSAFETY
98.	#DRIVERSSAFE
99.	#DRIVINGSAFETY
100.	#DRIVINGSAFE
101.	#DRIVINGSAFER
102.	#DRIVINGSAFELY
103.	#HIGHWAYSAFETY
104.	#CROSSINGSAFETY
105.	#CROSSINGSAFELY
106.	#TRAFFICACCIDENT
107.	#ROADACCIDENT
108.	#HIGHWAYACCIDENT
109.	#VEHICLEACCIDENT
110.	#VEHICLESACCIDENT
111.	#CARACCIDENT
112.	#CARSACCIDENT
113.	#BUSACCIDENT
114.	#AUTOMOBILEACCIDENT
115.	#TRUCKACCIDENT
116.	#SUVACCIDENT
117.	#VANACCIDENT

118.	#TAXIACCIDENT
119.	#CABACCIDENT
120.	#MOTORBIKEACCIDENT
121.	#MOTORCYCLEACCIDENT
122.	#PEDESTRIANACCIDENT
123.	#INTERSECTIONACCIDENT
124.	#TRAFFICACCIDENTS
125.	#ROADACCIDENTS
126.	#ROADSACCIDENTS
127.	#HIGHWAYACCIDENTS
128.	#VEHICLEACCIDENTS
129.	#CARACCIDENTS
130.	#CARSACCIDENTS
131.	#BUSACCIDENTS
132.	#AUTOMOBILEACCIDENTS
133.	#TRUCKACCIDENTS
134.	#TAXIACCIDENTS
135.	#MOTORBIKEACCIDENTS
136.	#MOTORCYCLEACCIDENTS
137.	#MOTORCYCLISTSACCIDENTS
138.	#PEDESTRIANACCIDENTS
139.	#INTERSECTIONACCIDENTS
140.	#TRAFFICCOLLISION
141.	#ROADCOLLISION
142.	#HIGHWAYCOLLISION
143.	#VEHICLECOLLISION
144.	#CARCOLLISION
145.	#BUSCOLLISION
146.	#AUTOMOBILECOLLISION
147.	#TRUCKCOLLISION
148.	#TAXICOLLISION
149.	#MOTORCYCLECOLLISION
150.	#PEDESTRIANCOLLISION
151.	#INTERSECTIONCOLLISION
152.	#TRAFFICCOLLISIONS
153.	#ROADCOLLISIONS
154.	#ROADSCOLLISIONS
155.	#CARCOLLISIONS
156.	#TRUCKCOLLISIONS
157.	#PEDESTRIANCOLLISIONS
158.	#TRAFFICCRASH

159.	#ROADCRASH
160.	#HIGHWAYCRASH
161.	#VEHICLECRASH
162.	#VEHICLESCRASH
163.	#CARCRASH
164.	#CARSCRASH
165.	#BUSCRASH
166.	#AUTOMOBILECRASH
167.	#TRUCKCRASH
168.	#SUVCRASH
169.	#VANCRASH
170.	#TAXICRASH
171.	#MOTORBIKECRASH
172.	#MOTORCYCLECRASH
173.	#PEDESTRIANCRASH
174.	#TRAFFICCRASHES
175.	#ROADCRASHES
176.	#HIGHWAYCRASHES
177.	#VEHICLECRASHES
178.	#CARCRASHES
179.	#BUSCRASHES
180.	#AUTOMOBILECRASHES
181.	#TRUCKCRASHES
182.	#VANCRASHES
183.	#TAXICRASHES
184.	#MOTORBIKECRASHES
185.	#MOTORCYCLECRASHES
186.	#CARCRASHING
187.	#CARSCRASHING
188.	#CARCRASHED
189.	#TRUCKCRASHED
190.	#TRAFFICFATALITY
191.	#ROADFATALITY
192.	#CARFATALITY
193.	#PEDESTRIANFATALITY
194.	#TRAFFICFATALITIES
195.	#ROADFATALITIES
196.	#HIGHWAYFATALITIES
197.	#CARFATALITIES
198.	#BUSFATALITIES
199.	#MOTORCYCLEFATALITIES

200.	#PEDESTRIANFATALITIES
201.	#TRAFFICINJURY
202.	#ROADINJURY
203.	#CARINJURY
204.	#BUSINJURY
205.	#MOTORBIKEINJURY
206.	#MOTORCYCLEINJURY
207.	#PEDESTRIANINJURY
208.	#TRAFFICINJURIES
209.	#ROADINJURIES
210.	#MOTORCYCLEINJURIES
211.	#PEDESTRIANINJURIES
212.	#TRAFFICDEATH
213.	#ROADDEATH
214.	#HIGHWAYDEATH
215.	#PEDESTRIANDEATH
216.	#TRAFFICDEATHS
217.	#ROADDEATHS
218.	#HIGHWAYDEATHS
219.	#PEDESTRIANDEATHS
220.	#ROADCASUALTY
221.	#PEDESTRIANCASUALTY
222.	#TRAFFICCASUALTIES
223.	#ROADSKILL
224.	#CARKILL
225.	#CARSKILL
226.	#CARKILLED
227.	#MOTORCYCLISTKILLED
228.	#PEDESTRIANKILLED
229.	#PEDESTRIANSKILLED
230.	#AGGRESSIVEDRIVER
231.	#RISKYDRIVER
232.	#UNSAFEDRIVER
233.	#UNTRAINEDDRIVER
234.	#ILLEGALDRIVER
235.	#NEGLIGENTDRIVER
236.	#RECKLESSDRIVER
237.	#AGGRESSIVEDRIVERS
238.	#RISKYDRIVERS
239.	#UNSAFEDRIVERS
240.	#UNTRAINEDDRIVERS

241.	#ILLEGALDRIVERS
242.	#NEGLIGENTDRIVERS
243.	#RECKLESSDRIVERS
244.	#AGGRESSIVEDRIVING
245.	#AGGRESSIVELYDRIVING
246.	#RISKYDRIVING
247.	#UNSAFEDRIVING
248.	#ILLEGALDRIVING
249.	#NEGLIGENTDRIVING
250.	#RECKLESSDRIVING
251.	#ROADMISHAP
252.	#HIGHWAYMISHAP
253.	#DEADLYTRAFFIC
254.	#TRAFFICOFFENSE
255.	#TRAFFICOFFENSES
256.	#TRAFFICOFFENDER
257.	#TRAFFICOFFENDERS
258.	#TRAFFICVIOLATION
259.	#TRAFFICVIOLATOR
260.	#TRAFFICVIOLATORS
261.	#SAFETYROAD
262.	#SAFEROAD
263.	#SAFERROAD
264.	#SAFELYROAD
265.	#SAFETYROADS
266.	#SAFEROADS
267.	#SAFERROADS
268.	#SAFETYTRAFFIC
269.	#SAFETRAFFIC
270.	#SAFEPEDESTRIAN
271.	#SAFEPEDESTRIANS
272.	#SAFETYDRIVE
273.	#SAFEDRIVE
274.	#SAFERDRIVE
275.	#SAFELYDRIVE
276.	#SAFEDRIVES
277.	#SAFETYDRIVER
278.	#SAFEDRIVER
279.	#SAFERDRIVER
280.	#SAFETYDRIVERS
281.	#SAFEDRIVERS

282.	#SAFETYDRIVING
283.	#SAFEDRIVING
284.	#SAFERDRIVING
285.	#SAFELYDRIVING
286.	#SAFEHIGHWAY
287.	#ACCIDENTTRAFFIC
288.	#ACCIDENTCAR
289.	#ACCIDENTCARS
290.	#ACCIDENTBUS
291.	#DEATHTRAFFIC
292.	#DRIVINGUNSAFE
293.	#DRIVINGILLEGAL
294.	#DRIVINGRECKLESS

Keywords for Q (2): Do we all believe it is possible to prevent fatal and serious injury crashes?

1.	TARGET ZERO	
2.	ZERO DEATH	TRAFFIC
3.	ZERO DEATHS	ROAD
4.	ZERO SERIOUS INJURY	ROADS
5.	ZERO SERIOUS INJURIES	TRANSPORTATION
6.	ZERO FATALITY	CAR
7.	ZERO FATALITIES	CARS
8.	ZERO FATAL ACCIDENT	DRIVER
9.	ZERO FATAL ACCIDENTS	DRIVERS
10.	0 DEATH	DRIVING
11.	0 DEATHS	HIGHWAY
12.	0 FATALITY	ROADWAY
13.	0 FATALITIES	
14.	0 SERIOUS INJURY	
15.	0 SERIOUS INJURIES	
16.	0 FATAL ACCIDENT	
17.	0 FATAL ACCIDENTS	
18.	NO DEATH	
19.	NO DEATHS	
20.	NO FATALITY	
21.	NO FATALITIES	
22.	NO SERIOUS INJURY	
23.	NO SERIOUS INJURIES	
24.	NO FATAL ACCIDENT	

25.	NO FATAL ACCIDENTS	
26.	0	TRAFFIC DEATH
27.	ZERO	TRAFFIC DEATHS
28.	NO	TRAFFIC FATALITIES
29.		TRAFFIC FATALITY
30.		TRAFFIC INJURIES
31.		TRAFFIC INJURY
32.		TRAFFIC SERIOUS INJURY
33.		TRAFFIC SERIOUS INJURIES
34.		TRAFFIC ACCIDENT
35.		TRAFFIC ACCIDENTS
36.		ROAD DEATH
37.		ROAD DEATHS
38.		ROAD FATALITIES
39.		ROAD FATALITY
40.		ROAD INJURIES
41.		ROAD INJURY
42.		ROAD SERIOUS INJURY
43.		ROAD SERIOUS INJURIES
44.		ROAD ACCIDENT
45.		ROAD ACCIDENTS
46.		CAR ACCIDENT
47.		CAR ACCIDENTS
48.	#TARGETZERO	

Keywords for Q (3): Do we all have the attitude that police enforcement of traffic laws is beneficial?

1.	HVE PATROL
2.	HVE PATROLS
3.	HVE ENFORCEMENT
4.	CLICK IT OR TICKET
5.	DRIVE SOBER OR GET PULLED OVER
6.	TRAFFIC ENFORCEMENT
7.	HIGH VISIBILITY ENFORCEMENT
8.	SOBRIETY CHECKPOINT
9.	SOBRIETY CHECKPOINTS
10.	DUI TICKET
11.	DUI TICKETS
12.	SUSPENDED DRIVER LICENSE
13.	SPEED CAMERA
14.	SPEED CAMERAS

15.	SPEEDING CAMERA
16.	SPEEDING CAMERAS
17.	SPEED MONITOR
18.	SPEED MONITORS
19.	WASHINGTON STATE PATROL
20.	IGNITION INTERLOCK
21.	IGNITION INTERLOCKS
22.	24/7 SOBRIETY PROGRAMS
23.	24/7 SOBRIETY PROGRAM
24.	TRAFFIC LAW
25.	TRAFFIC LAWS
26.	TRAFFIC RULE
27.	TRAFFIC RULES
28.	TRAFFIC REGULATION
29.	TRAFFIC REGULATIONS
30.	#CLICKITORTICKET
31.	#DRIVESOBERORGETPULLEDOVER
32.	#TRAFFICENFORCEMENT
33.	#SOBRIETYCHECKPOINT
34.	#SOBRIETYCHECKPOINTS
35.	#DUITICKET
36.	#SPEEDCAMERA
37.	#SPEEDINGCAMERA
38.	#SPEEDINGCAMERAS
39.	#SPEEDMONITOR
40.	#WASHINGTONSTATEPATROL
41.	#IGNITIONINTERLOCK
42.	#IGNITIONINTERLOCKS
43.	#TRAFFICLAW
44.	#TRAFFICLAWS
45.	#TRAFFICRULE
46.	#TRAFFICRULES
47.	#TRAFFICREGULATION
48.	#TRAFFICREGULATIONS

SIX SETS OF KEYWORDS FOR PUBLIC'S ATTITUDES TOWARDS THE HIGH-RISK BEHAVIORS IN PHASE 2 SENTIMENT ANALYSIS

Impairment Involved

1.	DRIVING UNDER INFLUENCE

2.	DUI	
3.	BLOOD ALCOHOL CONCENTRATION	
4.	BAC	
5.	IMPAIRMENT	DRIVER
6.	IMPAIRED	DRIVERS
7.	ALCOHOL	DRIVING
8.	COCAINE	BICYCLIST
9.	DRUG	BICYCLISTS
10.	MARIJUANA	MOTORCYCLIST
11.	OPIOID	MOTORCYCLISTS
12.	DRINKING	DRIVE
13.	DRUNK	DRIVES
14.	WINE	DROVE
15.	BEER	
16.	HEROINE	
17.	WEED	
18.	CANNABIS	
19.	#DRIVINGUNDERINFLUENCE	
20.	#DUI	
21.	#BLOODALCOHOLCONCENTRATION	
22.	#BAC	
23.	#IMPAIREDDRIVER	
24.	#DRUGDRIVER	
25.	#DRUNKDRIVER	
26.	#WINEDRIVER	
27.	#BEERDRIVER	
28.	#IMPAIREDDRIVERS	
29.	#COCAINEDRIVERS	
30.	#DRUGDRIVERS	
31.	#MARIJUANADRIVERS	
32.	#DRUNKDRIVERS	
33.	#BEERDRIVERS	
34.	#IMPAIREDDRIVING	
35.	#DRUGDRIVING	
36.	#MARIJUANADRIVING	
37.	#DRINKINGDRIVING	
38.	#DRUNKDRIVING	
39.	#BEERDRIVING	
40.	#WEEDDRIVING	
41.	#CANNABISDRIVING	
42.	#DRUNKBICYCLIST	

43.	#DRUGDRIVE
44.	#DRINKINGDRIVE
45.	#DRUNKDRIVE
46.	#DRUNKDRIVES
47.	#DRIVERDRUNK
48.	#DRIVINGIMPAIRED
49.	#DRIVINGDRUNK
50.	#DRIVEDRUNK
51.	#DRIVESDRUNK

Speeding Involved

1.	SPEEDING	
2.	FAST	DRIVE
3.	SPEED	DRIVER
4.	SPED	DRIVERS
5.		DRIVING
6.		DROVE
7.	#SPEEDING	
8.	#FASTDRIVE	
9.	#SPEEDDRIVE	
10.	#FASTDRIVER	
11.	#SPEEDDRIVER	
12.	#FASTDRIVERS	
13.	#SPEEDDRIVERS	
14.	#FASTDRIVING	
15.	#SPEEDDRIVING	
16.	#DRIVEFAST	
17.	#DRIVESPEED	
18.	#DRIVERSSPEED	
19.	#DRIVINGFAST	
20.	#DRIVINGSPEED	

Distraction Involved

1.	DISTRACTED	DRIVE
2.	DISTRACTION	DRIVER
3.	PHONE	DRIVERS
4.	MOBILE	DRIVING
5.	TEXT	DROVE
6.	TEXTING	PEDESTRIAN

7.	MESSAGING	PEDESTRIANS
8.	MESSAGE	
9.	CELLPHONE	
10.	CALLING	
11.	CARELESS	
12.	#DISTRACTEDDRIVE	
13.	#CARELESSDRIVE	
14.	#DISTRACTEDDRIVER	
15.	#PHONEDRIVER	
16.	#MOBILEDRIVER	
17.	#TEXTDRIVER	
18.	#TEXTINGDRIVER	
19.	#MESSAGEDRIVER	
20.	#CELLPHONEDRIVER	
21.	#CARELESSDRIVER	
22.	#DISTRACTEDDRIVERS	
23.	#TEXTINGDRIVERS	
24.	#CELLPHONEDRIVERS	
25.	#CARELESSDRIVERS	
26.	#DISTRACTEDDRIVING	
27.	#DISTRACTIONDRIVING	
28.	#PHONEDRIVING	
29.	#MOBILEDRIVING	
30.	#TEXTDRIVING	
31.	#TEXTINGDRIVING	
32.	#CELLPHONEDRIVING	
33.	#CARELESSDRIVING	
34.	#DISTRACTEDPEDESTRIAN	
35.	#DISTRACTEDPEDESTRIANS	
36.	#CARELESSPEDESTRIANS	
37.	#DRIVEDISTRACTED	
38.	#DRIVERDISTRACTED	
39.	#DRIVERCELLPHONE	
40.	#DRIVINGDISTRACTED	
41.	#DRIVINGDISTRACTION	

Unrestrained Vehicle Occupants

1.	SEATBELT OFF		
2.	SEATBELT	NOT	CLOSED
3.		HAVEN'T	PROPER

4.		HASN'T	WEAR
5.		HADN'T	WEARING
6.		DON'T	FASTEN
7.		DOESN'T	BUCKLE UP
8.		DIDN'T	BUCKLED UP
9.		ISN'T	WORN
10.		AREN'T	FASTENING
11.		WASN'T	FASTENED
12.		WEREN'T	
13.	CHILD		UNSAFE SEAT
14.	CHILDREN		FRONT SEAT
15.			UNRESTRAINED
16.			NON-SAFETY SEAT
17.			UNSAFETY SEAT
18.	UNRESTRAINED		DRIVER
19.			DRIVERS
20.			OCCUPANT
21.			OCCUPANTS
22.			PASSENGER
23.			PASSENGERS
24.	#SEATBELTOFF		
25.	#SEATBELTON		
26.	#NOSEATBELT		
27.	#SEATBELT		
28.	#NOSEATBELT		

Young Driver Involved

1.	YOUNG	DRIVE
2.	JUNIOR	DRIVING
3.	TEEN	DRIVER
4.	TEENAGE	DRIVERS
5.		DROVE
6.	#YOUNGDRIVE	
7.	#JUNIORDRIVE	
8.	#TEENDRIVE	
9.	#TEENAGEDRIVE	
10.	#YOUNGDRIVING	
11.	#JUNIORDRIVING	
12.	#TEENDRIVING	
13.	#TEENAGEDRIVING	

14.	#YOUNGDRIVER
15.	#JUNIORDRIVER
16.	#TEENDRIVER
17.	#TEENAGEDRIVER
18.	#YOUNGDRIVERS
19.	#JUNIORDRIVERS
20.	#TEENDRIVERS
21.	#TEENAGEDRIVERS
22.	#DRIVINGTEEN
23.	#DRIVERTEEN

Older Driver Involved

1.	OLDER	DRIVE
2.	SENIOR	DRIVING
3.	OLD	DRIVER
4.	ELDER	DRIVERS
5.		DROVE
6.	#SENIORDRIVING	
7.	#ELDERDRIVING	
8.	#OLDERDRIVER	
9.	#SENIORDRIVER	
10.	#OLDDRIVER	
11.	#ELDERDRIVER	
12.	#OLDERDRIVERS	
13.	#SENIORDRIVERS	
14.	#OLDDRIVERS	
15.	#ELDERDRIVERS	