UNIVERSITY OF TARTU

Institute of Computer Science

Computer Science Curriculum

Joosua Adler

Multidimensional analysis to understand public perception for a health initiative project

Bachelor's Thesis (9ECTS)

Supervisors: Rajesh Sharma, PhD

Ruth Hunter, PhD

Tartu 2019

Multidimensional analysis to understand public perception for a health initiative project

Abstract:

Speeding is one of the major causes in road accidents in city areas that end in severe injuries or death. In this respect, the United Kingdom government launched a health oriented transport initiative across several sites by proposing reduction of speed limit to 20 miles per hour to achieve fewer casualties and to lower traffic volumes, leading to an improvement in the perception of safety and a subsequent increase in people's physical activity. Various studies have been performed in the past to understand the impact of speed in accidents and road safety. Results have shown that lower speed limit in urban evironment makes traffic safer. In this work, social media data is analysed by taking Twitter as a use case to understand general public's perception to 20 miles per hour speed limit. The data consists of tweets which span from February 2015 to March 2017. The analysis was performed using opinion mining and social network analysis. The results of the analysis indicated that people showed positive reaction to the change and were ready to accept the new speed limit for the benefit of safety and health.

Keywords: health related policy,speed limit, opinion mining, social media analysis

CERCS:P170 Computer science, numerical analysis, systems, control

Mitmemõõtmeline analüüs mõistmaks avalikkuse arusaama tervise algatuse suhtes ühe projekti näitel

Lühikokkuvõte:

Tõsiste vigastustega või surmaga lõppevate liiklusõnnetuste üks suurimatest põhjustajatest linnades on kiiruse ületamine. Sellega seoses käivitas Ühendkuningriigi valitsus tervisele orienteeritud transpordialgatuse mitmetes kohtades, tehes ettepaneku vähendada kiirusepiiranguid 20 miilini tunnis, et vähendada ohvrite arvu ja liiklusmahtu. Koos sellega paraneks ohutustunne ja suureneks inimeste füüsiline aktiivsus. On läbi viidud mitmeid uuringuid, et mõista kiiruse mõju õnnetustes ja liiklusohutuses.. Tulemused on näidanud, et madalam kiiruspiirang linnades muudab liikluse ohutumaks. Bakalaurusetöös analüüsitakse sotsiaalmeedia andmeid, võttes aluseks Twitteri postitused, et mõista üldsuse taju kiirusepiirangu 20 miili tunnis suhtes. Andmed koosnevad Twitteri postitustest, mis on tehtud 2015.a veebruarist 2017.a märtsini. Analüüs viidi läbi hoiakute kaeve ja sotsiaalvõrgustiku analüüsi teel. Analüüsi tulemused näitasid, et inimesed reageerisid muudatusele positiivselt ja aktsepteerisid ohutuse ja tervise nimel uue kiirusepiirangu.

Võtmesõnad: tervisepoliitika, kiiruspiirang, hoiakute kaeve, sotsiaalmeedia analüüs

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of contents

1. Introduction

Driving faster than the obligatory speed limit is one of the major causes of accidents in downtowns, residential areas and school areas.Therefore it is important for drivers to follow the traffic rules. In this respect city councils have announced new lower speed limits to reduce accidents (King, 2018). In addition lower speed limit reduces air pollution (Group, 2013).

In previous works researchers have analysed the impact of speed limits using computer modelling or by gathering statistics over periods of time with different speed limits. This has been studied in many places around the world such as South Australia (J.E.WOOLLEY, 2002) and United States of America (Wen Hu, 2019) among others. These studies relied on gathering new data and comparing it to the older data.

The purpose of this thesis is to understand public response for speed limit policy change by studying the social media interactions of individuals with the speed limit policy handle on Twitter. These interactions have been studied using a multidimensional approach by incorporating descriptive analysis, sentiment analysis, social network analysis (Wasserman, 1994), topic analysis, text mining and geomap analysis.

In particular, the problem was studied by using 3422 posts of Twitter data which span across February 2015 to March 2017 to understand public's reaction and mindset with respect to health oriented initiative, launched by the U.K. government, to set new speed limit of 20 miles per hour in different city areas. The result of the analysis shows that these initiatives have been taken positively by the people in general.

The rest of this thesis is organized as follows. Section 2 discusses about the effects lowered of speed limits in urban areas, provides an overview what other places around the world have done to make streets safer and how different data analysing methods can be used on Twitter data. Section 3 describes the methodology used in this work. Section 4 presents the results of the analyzes done with the data. Conclusion is provided in Section 5.

2. Related work

In this section, the effects of speed limits in city areas will be described first (Section 2.1). Secondly, an overview is given of how other places have dealt with speeding (Section 2.2 and 2.3). Thirdly, it is discussed how different data analyze methods can be used on Twitter data to achieve different results (Section 2.4).

2.1 The Effects of lowered speed limits in urban areas

Nowadays as traffic volume is only increasing, it is getting harder to find affordable solutions to sustain the needs of todays transport system without sacrificing too much mobility or making huge investments into road infrastructure. New ways are sought by society. One option for improving safety and traffic flow in urban areas is lowering speed limits. This is cheap, but can affect travel times. It has been found out by a study conducted by OECD(Organisation fot Economic Co-operation and Development) and the ECMT(Eutopean Conference of Ministers of Transport) that speeding is the number one safety problem in most contries in the world (Jeffery Archer, 2008). Furthermore lowering speeds by 5 percent lowers fatalities by 20 percent. Another good side to lower speed besides safety is better traffic flow which decreases traffic jams, especially on roads that cannot handle more traffic. Additionally this helps with vehicle operating costs, energy efficiency, pollution and noise.

Determining the right speed limit on roads comes down to the design of the road. Speed must be suited for the road. Drivers tend to choose their speed according to the road they are on rather than the set speed limit. According to studies, speed limits are set combining three principles. Adapting the speed limit to actual speed levels, varying the speed limit to the standard and design of the road, choosing a speed limit that has the lowest cost to society (Jeffery Archer, 2008).

While travelling, a driver looks for good mobility and travel time with enough safety. Regrettably, majority of drivers are prone to choose faster travelling before safety on account of themselves and others. They choose their speed by checking for enforcement on the road, road design, how much traffic is nearby, what condition is their vehicle in, how good are their driving skills, and accident rates. Still, the main factor in choosing a speed is the posted speed limit (Jeffery Archer, 2008). Newly licenced drives are shown to think they are better than

they really are and therefore they drive faster than they should without noticing possible dangers. Speed enforcement is needed to keep unnecessary high speed at bay.

Second way to control speed is with road design. This includes enhancing streets with obstacles such as speed-bumps, chicanes, raised intersections and so on. However, as this may be a good way to calm down traffic, it is bad for emergency vehicles trying to get to places with the lowest amount of time possible. These obstacles also tend to divert traffic to streets in the vicinity causing overload of traffic.

Thirdly, it is possible to use operational measures to keep traffic at bay and lower the speed. This includes multiway stop, yield signs, traffic signals, turn prohibitions and one-way streets (Jeffery Archer, 2008). Used correctly by optimizing to posted speed limit, traffic signals can be efficient in keeping the traffic flowing without delays. Variable message signs are great for letting drivers know of upcoming issues on the road. Futhermore, a way to keep drivers from speeding, is installing mandatory speed limiting devices within the vehicle. These are already used on heavy vehicles. Lastly, another key factor of maintaining safer speeds is drivers attitudes. Understanding the dangers of driving with higher speeds has generally increased.

Contrarily to general belief by drivers that driving faster reduces travel time, in the urban environment drivers cannot maintain cruise speed, but must stop often or slow down due to intersections, pedestrian crossings, rail crossings, reduced speed areas. Another obstacle is traffic jams. Lower speed limit however increases traffic flow which helps with these obstructions. Travel time is affected only minimally. This is a small price to pay for safer traffic.

2.2 Dealing with speed in the cities of the United States of America

2.2.1 New York City

In relation to a project called Vision Zero, which is a multi-national road traffic safety project that aims to achieve a road system with no fatalities or serious injuries involving traffic, New York City has come up with strategies to decrease speed. According to a crash summary of New York City (ITSMR, 2015), drivers who drive faster than the speed limit cause fatal accidents more than other reasons combined that end up in a fatality. Other reasons being a driver is drunk or using a cell phone while driving.

First strategy was lowering speed limit from 30 miles per hour to 25 miles per hour. In addition, city's staff made sure people are aware of the new speed limit by launching a wide spread education campaign including fliers, news, drive-time radio ads, social media ads and adding new speed limit information on municipal parking receipts (Ferrier, 2017). Awareness went up from 28% to 62%. Second strategy was automated speed enforcement which is installing fixed speed cameras that detect and capture images of vehicles travelling at dangerous speeds. This was eventually added to 140 school zones. New York has done well in lowering speeds as shown in this report (City, 2014-2016) 2 years after the changes.

2.2.2 Chicago

Like New York, speeding causes the most deaths in Chicago. Chicago placed emphasis on speed management and speed prevention to reduce speed related crashes 25% by 2020 (Ferrier, 2017). To discourage speeding and manage safe speeds, Chicago's Department of Transportation put their effort into designing safer streets. This method included narrowing two travel lanes and adding a protected bike line. Another solution used for disorganised conditions was to change design, add safety posts and new paint to define different areas and lanes for areas full of pedestrians and cyclists. Furthermore, on these roads, speed limit was lowered to 20 miles per hour. Equally to New York, Chicago uses automated speed enforcement in school zones and park zones otherwise known as Children Safety Zones. Cameras have so far helped keep speedings at bay.

2.2.3 Portland

Speeding has become a problem. To encounter this, Portland set up a High Crash Network to mark out intersections and streets where crashes happen most frequetly (Network, 2018). Main connection between these choices are high death rate and higher speed limits. Solution consists of many different plans of action.

To start off, Portland established appropriate speed limits to roads, which do not match the required conditions for current speed limit. Furthermore the idea was to focus more on pedestrians and cyclists safety. Speed limits and roads should consider all the road users rather than just motorists. Portland also started Alternate Speed Zone investigation, which allows the city to request a change in speed limits on local streets through Oregon Department of Transportation (Network, 2018).

Secondly, the city puts emphasis on designing safer streets. Safety design elements include protected bike lanes, raised curbs to create barriers between different road users (Network, 2018). The plan is to create a street, where drivers do not want to drive faster than the allowed speed. Authorities pick out the streets by three metrics (Network, 2018). These are: the number of fatal and injury causing crashes in a chosen four to five-year time frame, the collision rate, the total value of crashes, which assigns a monetary value to the severity of injuries. Lastly, Portland started using speed safety cameras to motivate people to care more about traffic regulations.

2.2.4 Harrisburg

Since the number of car crashes involving pedestrians, bicyclists and motorcyclists was high in last years in Harrisburg, authorities had to come up with strategies to improve the situation after evaluating the most dangerous places. One part was installing red light cameras to stop individuals from running red lights. However, as Harrisburg is not a first-class city, this action is denied by state law (Vendel, 2019). One very problematic intersection with high crash rate was redesigned with safer features. Project was sent for review to get federal funding. Another problem in downtown consisted of an intersection without marked crosswalk for pedestrians. There is hope that the newly made project will be included in the city's 2020 budget (Vendel, 2019).

In addition, an idea is to use police resources to put officers on locations with high crash rate. The point is to keep them visible thereby enforcing the law. What's more, illegally parked cars which degrade pedestrian safety would get higher penalties than before. Moreover, to avoid relying only on public complaints, city officials plan to walk around streets to find obstructions that make pedestrians walk the on road instead of staying on the sidewalk (Vendel, 2019). Getting rid of these obstacles would improve safety.

2.2.5 Seattle

Making an effort to minimize traffic injuries Seattle limited speed to 25 miles per hour in main streets and throughout downtown. In residential areas the new number is 20 miles per hour. Still, a question remains, whether Seattle's traffic unit capable to enforce the new limit. Furthermore there would be school-zone enforcement, elimination of not signalled crosswalks, adding speed bumps, making wider sidewalks and switching four lane roads to three lane roads (Lindblom, 2016).

2.3 Lowering speed limits in South Australia

After many years of debate, thanks to local communities for standing up, finally it was decided that urban limit 60 kilometres per hour was too high. In Australia, individual states and territories make decisions of setting speed limits. People demanded lower speed limits to have less dangerous traffic and safer environments. The new limit in urban areas is 50 kilometres per hour. This was applied to most jurisdictions. However, getting drivers to follow new limit presented another problem. Maintaining the right speed requires enforcement before people come to terms with the new limit as drivers with a speeding habit have hard time not to do it.

For research about the lower urban speed limit and the effects of it, first method was theoretical modelling. This involved comparing results from assessing traffic and environmental performance of test networks under different speed limits and traffic conditions (J.E.WOOLLEY, 2002). Results included that the only thing better apart from safer traffic with lower speeds was congestion index.

Second method used was real world modelling, which was a real model of a local road network, observing mornings from 8 am to 9 am. It incorporated the whole traffic. Data for input was collected previously by performing trips through the area. Road network consisted of different types of roads: residential streets, collector and arterial roads. Different scenarios were investigated by applying varying speeds on each road type. Results concluded that by lowering speed limit, increase in travel time increase is just minimal instead of high. The only thing missing from this test was not optimizing signalised intersections to changed speeds. For that reason, the results do not include some potential gains from it, if such action were to be done.

Prior work in other areas, such as Adelaide and Queensland, where speed limit was also reduced, had a six month trial. In Adelaide, during that time injury crashes had notable reduction, although crash frequency and crash types did not have a major impact. In Queensland, assessment of the test period showed substantial reduction in crash risk and the number of crashes (J.E.WOOLLEY, 2002). Thanks to this, communities support went up and the new speed limits were accepted. Also, drivers showed signs of adapting to lower speed.

In 1999, city of Unley, implemented a 10 units lower speed, 40 kilometres per hour, to all local streets. Data for evaluation was collected for 21 months. This was compared to historical data. One main objective was to reduce the 85th speed percentile. It is the speed below which 85% of all the speeds recorded lie (J.E.WOOLLEY, 2002). Bearing in mind the 85th speed percentile, streets with already low speed traffic before did not show any remarkable change. However, streets with a little faster traffic before this change showed small reduction and streets which sustained faster traffic before had remarkable reduction in 85th percentile and mean speeds. Lowering of speed limits on local streets also lowered traffic volumes. Telephone questionnaire surveys indicated that Unley's community support for 40 kilometres per hour speed limit was high. Those few people against it were only those that were caught speeding and got fined.

Another reason for lowering speed limits is fuel consumption and air emissions. It is believed that lower speeds produce less emissions due to cars travelling slower. This might not be true for vehicles driving at a constant speed. Nevertheless higher speeds have more potential for higher emissions as they require more acceleration to get to those speeds. This is more relevant especially in cities, not on highways, due to intersections which cause even more deceleration and acceleration. Conducting a study needs to take into consideration the real driving conditions encountered on the road network. Driving at lower speed for a long period of time produces more emissions than higher speed, but in residential areas where there are short street sections, it is another situation, because for most of the time, cars do not stay at steady speed, they mostly accelerate or decelerate. These actions produce more emissions than just driving at a steady low speed. For testing, a vehicle was driven in real traffic, while its emissions were monitored. It was done in 40 kilometres per hour speed zone and 60 kilometres per hour speed zone with three different driving styles. A slow conservative driver, an average driver and an aggressive driver (J.E.WOOLLEY, 2002). All three accelerated, cruised and decelerated according to their type. Results confirmed that on short streets its better for the environment for cars to travel with lower speeds, however on longer streets where there is more cruise time, it is better to have higher speeds.

2.4 Analyzing Twitter data

A research project was done in the Analytics Research Weeks in Australian Department of Immigration and Citizenship(DIAC), with Twitter as a use case to analyse social media data

and study the feasibility of using them for better understanding and improving the business of this department (Zhao, 2013). For this, the official DIAC twitter account was used to get the data for analysing. Futhermore, followers of this account were examined with social network analysis with the objective of finding out how posts spread over Twitter network.

Before analysing, tweet texts were cleaned of misspelling, html links, stop words and words that were in the majority of the tweets. Next, tweets were analysed with text mining techniques to create a wordcloud of frequent words with R package *wordcloud* and *tm*. This gave base to creating dictionary of frequent terms. Results were shown in a Term Network graph. Following, tweets were looked through for topics with R package *topicmodels*. Number of tweets of each topic presented more density to an edge. Outcome was made to a stream graph to visualise temporal variation of topics (Zhao, 2013).

Social Network analysis consisted of analysing retweets and followers. The locations of followers were put on a map for geomap. Followers were categorized by their descriptions on their accounts using text mining. For evaluating the impact of each user, different details were collected like number of tweets, followers of the user, users they themselves followed and account creation date. The more followers a user had and less users he followed the more influential a user was considered. Another aspect was the activeness of a user. This produced a figure of most influential users. In addition an analysis was done by sorting out most retweeted tweets to see which ones were the most popular and how far had they spread.

The study concluded that it was a preliminary research. Tweets spread on Twitter network had been studied. Tweets, topics and their variations over time had been recognized with text mining. Used methodology is general and applicable for others to use on other Twitter accounts. There are many ways to expand this research, like adding more main accounts and data from other social media networks into analysis and doing comparison between them.

With the number of tourists increasing with each coming year in Perú and tourism being an important source of income in Perú, government tourism bureaus are working hard to keep these numbers growing (Luis Cajachahua, 2017). To further keep Perú appealing for tourists a research was made to understand tourists preferences and sentiments by analysing tweets from Twitter to find out which places tourists prefer and how they feel about them and how to improve these locations. For this, different analysis techniques were applied.

Text mining consisted of four steps: tweets collecting, pre-processing, tagging and classification, characterization. Tweet collecting was done by finding tweets that contained words associated with places in Perú. Pre-processing consisted of cleaning, fixing spelling mistakes, terms filtering and ortohgraphical review. Tagging and classification was done by understanding the dataset and analysing the comments with different R packages. Characterization was used to identify the types of tweets. Moving on, sentiment analysis was done with the Libary Syuzhet in R. This tool is known to have good performance in scoring sentiments or computing polarity. Through it eight sentiments were used: anger, fear, sadness, disgust, surprise, anticipation, trust and joy. The higher the score the more popular sentiment is. Correspondence analysis was done with qualitative statistical dimension reduction technique that is useful for representing original frequencies in a bi-dimensional space after performing a matrix decomposition of the analysed crosstab (Luis Cajachahua, 2017).

Filtering tweets helped to identify most active users. Cleaning process lead to finding the most frequent words to create a list and a wordcloud on them. Libary syuzhet was used to create an overview of positive, negative or neutral tweets. This was compared between regions to find out the more positive and more negative places. This was also applied to the most popular places. In addition words with the biggest appearance were found from positive and negative tweets to find specific words related to those places. Link analysis was done between terms using a metric strength for building graphs of relationships between most mentioned words and other words. Furthermore, libary syuzhet of R was used to classify emotions that were mentioned before in tweets. Correspondence analysis was done by piecing all previous results together to construct correspondence plots by emotions and place or concept and place. In addition a geomap of all mentioned places was formed.

Conclusion consisted of finding out percentage of active users, most commented place in Perú, most valued tourist site, notices of negative influence incidents, different opinions of Perú and peak time of Twitter activity.

3. Methodology

This section describes the work process of various approaches that have been used for analyzing the data.

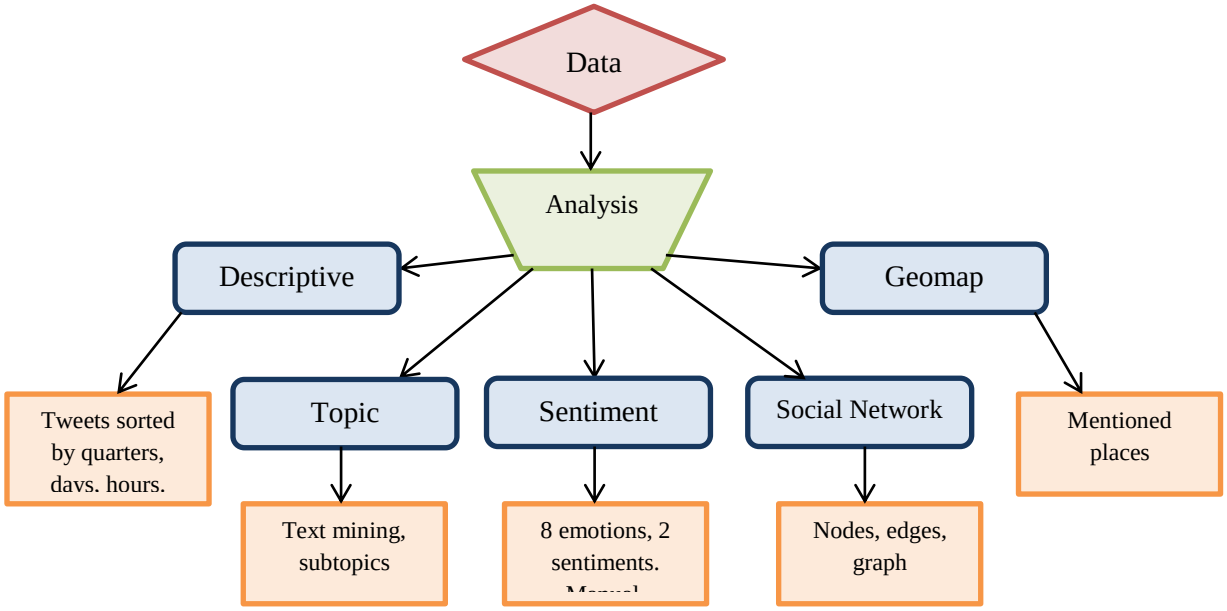

Figure 1. Flowchart of workflow

3.1 Descriptive analysis

For descriptive analysis the number of tweets was sorted by quarters, days of the week, hours of day in R (Laboratories, 1995) using Rstudio (RStudio, Online).

3.2 Topic analysis

Before text mining, a corpus was made from tweet texts. R package 'tm' (Ingo Feinerer, 2018) was used for cleaning the corpus by removing punctuation, number, stopwords, html links and some words that appeared in most of the tweets. As a next step, term document matrix was created from this clean dataset. R package Wordcloud2 (R-project, 2018) was used to create a visualisation of most frequently used words. Subtopics were found by manually reading through the tweets.

3.3 Sentiment analysis

Firstly, tweets were divided into groups in RStudio (RStudio, Online), groups being all tweets, tweets each year, weekday tweets from Monday to Friday, Saturday and Sunday for weekends. R package Syuzhet (Jockers, 2017) was used to give sentiment scores over tweets to create barplots. Sentiment analysis was done over 8 different emotions(*disgust, anger, sadness, fear, anticipation, suprise, joy, trust)* and 2 sentiments(*positive, negative)*. Also tweets were rated manually with 5 different categories: -2 for very negative, -1 for slightly negative, 0 for neutral, 1 for slightly positive and 2 for very positive. Categories were chosen by reading through a tweet and evaluating its rating depending on content. However some conversations were one sided, so a precise rating could not be determined. Those mostly received rating 0.

3.4 Social Network analysis

For social network analysis the content of tweets was filtered in Python (Python, Online) for nodes and edges to formulate a graph using Gephi (Gephi, Online) software. Layout used was OpenOrd. Each Twitter handle is considered as a node. Edges from the main node which is the main account went to all other mentioned accounts. Other edges were determined as follows: if usernames were mentioned in the same tweet, it meant that they were connected and they would all get an edge between them. Weight was determined by how many times source user and target user were mentioned throughout the data.

3.5 Geomap analysis

For this analysis, all the cities were filtered out from the tweets to understand which cities were mentioned during the whole interaction dataset.

4. Data analysis

Dataset consists of 3422 tweets from years 2015 to 2017. After removing empty tweets, 3413 were left for analysing. This section shows the results of the analyses done with the data.

4.1 Descriptive analysis

Here is an overview of the data analysed by number of tweets.

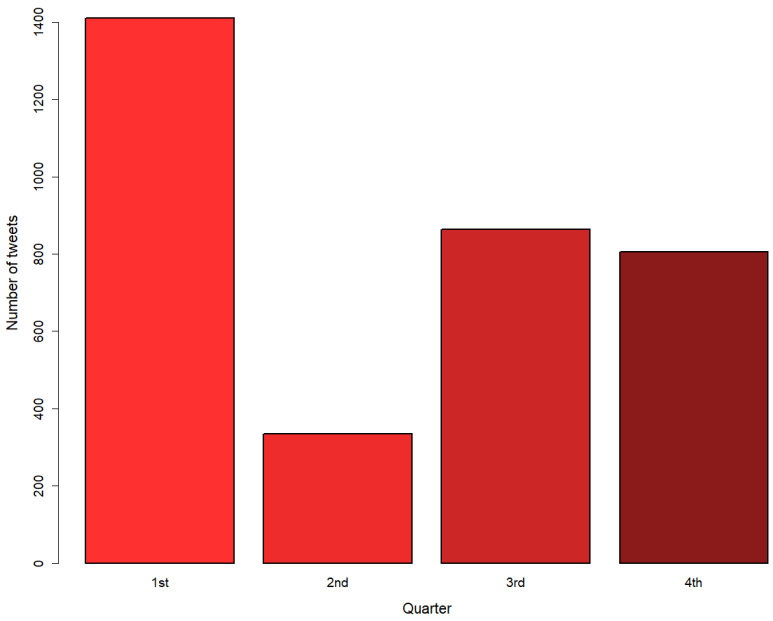

Figure 2. Total number of tweets divided over quarters

Figure 2 shows total number of tweets by quarters. First quarter consists of January, February and March. Second quarter consists of April, May and June. Third quarter consists of July, August and Septermber. Fourth quarter consists of October, November and December. As seen from figure 2, most tweets were done in the 1st quarter. However, considering that in the year 2017 there was only data of the first quarter, roughly 600 tweets, the number of tweets done in first, third and fourth quarters are of quite equal size. Quarter with the lowest numer of tweets during these years is the 2nd quarter. Hence, during second quarter of a year, the activity in twitter was at its lowest. Throughout other three quarters, the volume of activity stayed the same.

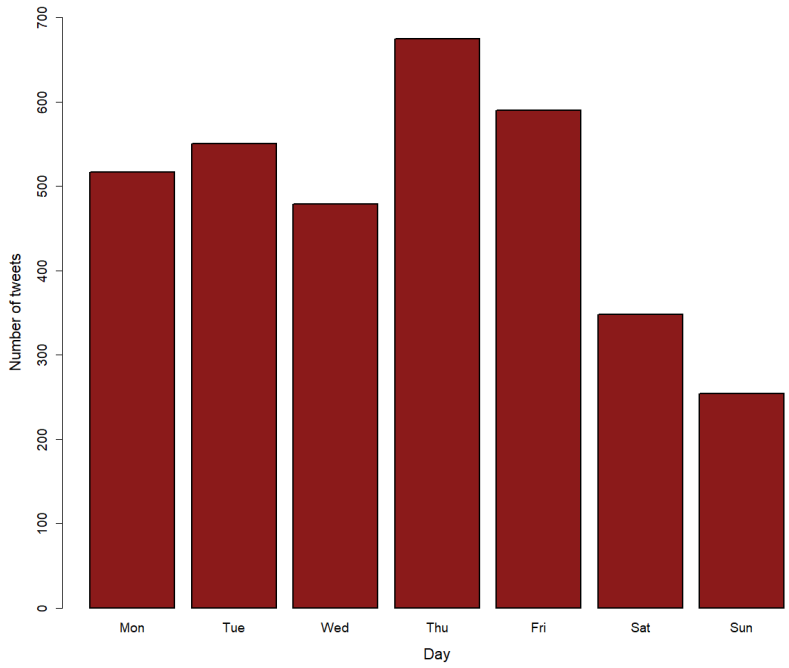

Figure 3. Total number of tweets divided over days

Figure 3 shows an overview of number of tweets posted by day. Days are from Monday to Sunday. Figure 3 shows that most tweeting is done from Monday to Friday. It makes sense to discuss the new speed limit as more people use roads to travel between work and home As seen on figure 3, most tweets were done on Thursday. This could be because it is close to the end of workweek. New events have developed in the start of the week, Thursday is better suited to wrap things up and as it is before Friday, people are more productive. Number of tweets on Friday is a bit lower, as people want to get to the weekend already and attend their free time, but at the same time finish their week properly. Start of the week from Monday to Wednesday shows results with no significant difference. However, Saturday is last but one in the lowest number of tweets before Sunday. This proves that people tend to relax more on weekends and deal less with social media. Sunday having the lowest number of tweets is no suprise, as Sunday is considered to be meant for resting by most people.

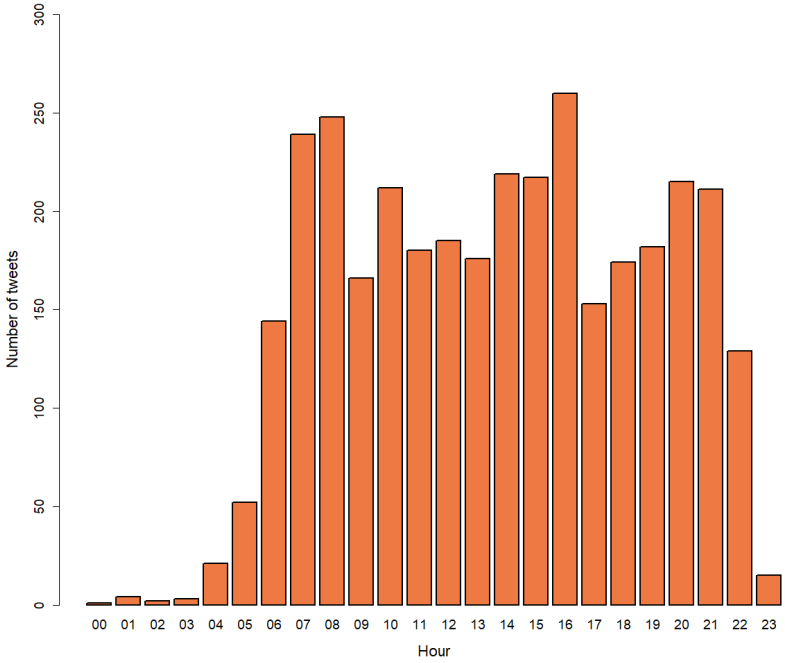

Figure 4. Total number of tweets divided over hours

Figure 4 shows number of tweets posted by hours. Hours are from 00:00 to 23:00. Peaktime for tweets seems to be in the morning from 7 to 9 and after lunch from 16 to 17. This might refer to the start and the end of the workday for people working in day shifts. As seen, most of the tweets are being tweeted in the daytime from 6 in the morning till 23 in the evening, when people are doing their daily routine. During nighttime, from 23 till 6 in the morning, there is none to very little amount of tweeting as people are mostly sleeping and getting ready for the next day.

4.2 **Topic analysis**

Figure 5. Wordcloud

Figure 5 gives an overview of most frequent words with a Wordcloud. As seen most topics point to *speed, streets, safety, roads, people*. Adjectives that point out most are *good* and *great*. This shows that people are focused on the new speed rules and the safety they provide, they are mostly feeling good or great about it. As expected the most frequent word is *speed* followed by *road*. It can also be observed that conversation mainly revolved around *community, city* and how this policy is *better* in connection with *safety*, *children* and *residential* areas.

Upon reading through all the posted tweets, apart from the main theme in the discussion being implementation of 20 miles per hour speed limit, there were other topics as well. Firstly, different speed zones for 20 miles per hour limit. People wanted to get this limit into different zones. Mainly for safety near their homes, they put emphasis on getting this limit to residential areas. Furthermore, there was desire to have 20 miles per hour speed limit wider zones around schools. Authorities only wanted this limit around school, so children would be safe near school, but this did not satisfy the needs of parents as this area was too small. This would not help parents to let their kids go to school on their own either by walking or cycling because of the dangerous traffic. However a wider 20 miles per hour speed limit zone around schools which reaches residential areas would prove beneficial as traffic is slower and less dangerous. In addition people mentioned city centres as a place where to have a lower speed limit. This would make city centres a lot safer.

Secondly, a larger topic discussed was health. Health is connected to lower speed limit. With speed being a factor in safety, if people do not feel safe on the streets, they tend to spend less time outside doing physical activity, like walking or cycling. Through implementing this new speed limit and making streets safer, where people can safely get from one point to another they can choose to leave their car behind and be more active. This would also reduce volumes of vehicles. Important part of this is the health of children, as it is very important to be physically active from an early age on. If the way to school is safe enough for parents to let their kids walk to school, this already improves the situation. Another reason related to health in setting the 20 miles per hour speed limit is pollution and emissions of the cars. As these areas, city centre, residential strees, streets around school, tend to have short streets and lots of intersections, it requires a lot of acceleration and deceleration for a car to travel. These activities produce most emissions during driving. Setting a lower speed limit reduces the time for these processes. Therefore, car emissions go down and air becomes cleaner. In addition, a car going 20 miles per hour can stop in the distance within 40 feet, a car doing 30 miles per

hour is still doing 24 miles per hour at this distance and will not stop for another 35 feet. Hence, lower speed limit makes cars travel slower and stop faster before having to inflict fatal injuries on pedestrians or cyclists.

Thirdly, people discussed about different behaviours of driving. Specifically, how would this new speed limit effect drivers. In general there are three different types of drivers: slow conservative, average and aggressive. Slow conservative driver being the one who follows all the rules, drives a bit below the speed limit and takes his time. Average driver being the one who drives at a mild pace, who rarely may break some rules as a result of inattention. Aggressive driver being the one who wants to get to his destination as quick as he can, trying to overtake other cars, decelerate and accelerate as fast as possible, being a danger to himself and other road users. Setting the new 20 miles per hour speed limit does not change much for slow conservative and average drivers as they would just have to get used to the new limit and drive slower in those areas than before. However, dealing with aggressive drivers, they might start speeding more because of the lower speed limit. To encounter this, people demanded more enforcement of the new limit, so traffic would eventually calmer and safe.

4.3 Sentiment analysis

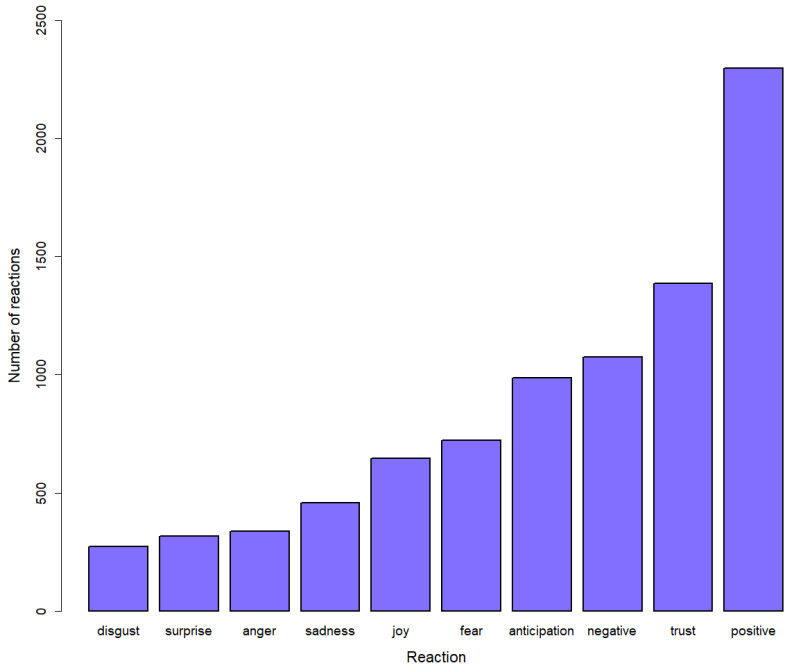

Figure 6. Reactions over all tweets

Figure 6 shows emotions and sentiments divided over all tweets. *Positive* is ahead by a huge lead, followed by *trust*. This is great as it shows that overall people are truly fond of this project. In the 4th place is *anticipation*, meaning people are looking forward to getting streets safer and healthier. However, *negative* is in the 3rd place with a slightly higher score than *anticipation* and this indicates that some people are against changes. Other bad feelings like *disgust*, *anger*, *sadness* and *fear* achieved very low scores. To summarize, people like this project and support it.

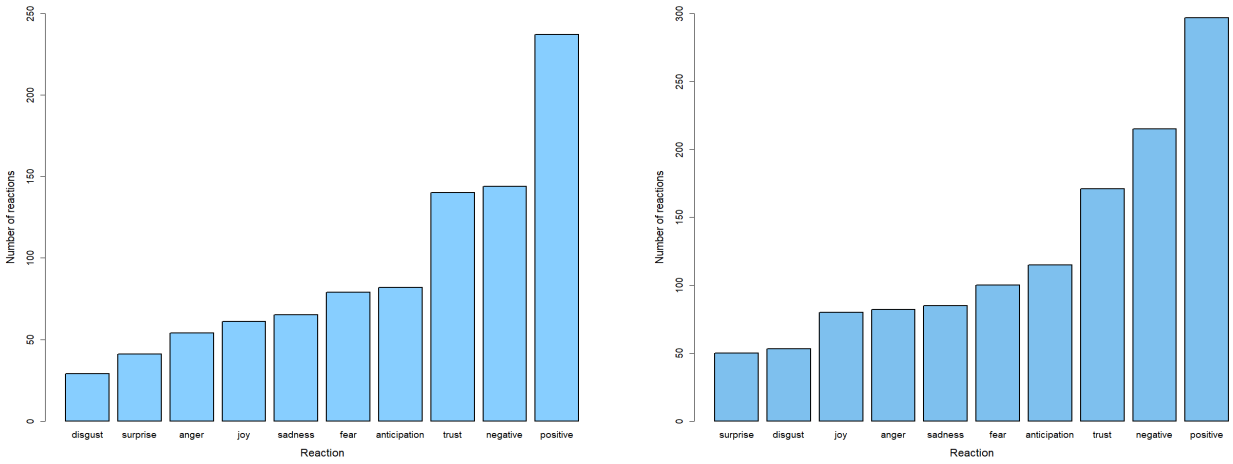

Figure 7. Reactions over year 2015 on the left and over year 2016 on the right

Figure 7 shows emotions and sentiments during year 2015 on the left and emotions and sentiments during year 2016 on the right. Mostly the scores are same. Few differences are at the smaller numbers of reactions. *Suprise* has switched places with *disgust* and *anger* with *joy*. A more visible change can be seen in *negative*. Probably as this project started in 2015, people did not know so much about it and as a year moved on, few more negative people showed up and pointed out why they did not like something.

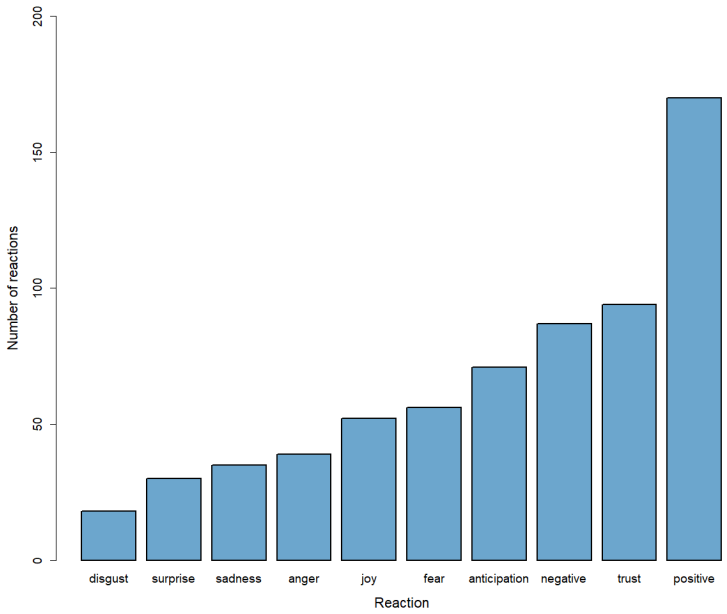

Figure 8. Reactions over year 2017

Figure 8 shows emotions and sentiments in year 2017. Compared to previous years, also here *positive* wins by a mile. A minor change can be noticed as *sadness* has fallen from 5th place to 8th. Despite that, the general outcome is mostly the same throughout all the three years.

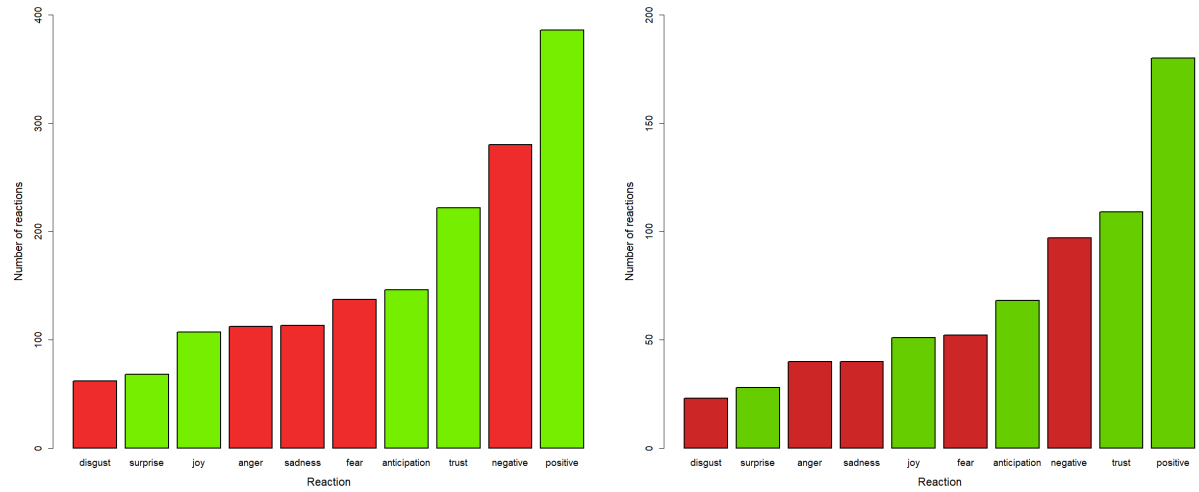

Figure 9. Reactions over weekdays on the left and over weekends on the right

Figure 9 shows distribution of feelings on weekdays on the left and on weekends on the right. Bad feelings are marked with red colour and good feelings with green colour. Main difference is with *negative*, on weekdays it has higher score and it is at 2nd place while over weekends it has third spot and lower score. This is logical as people are more relaxed on weekends and not so stressed from work bearing in mind that most people work from Monday to Friday. Another feeling switching places is *joy*, being 8th on weekdays, and 6th on weekends. In other aspects, like the order of feelings and number of feelings, results stay the same.

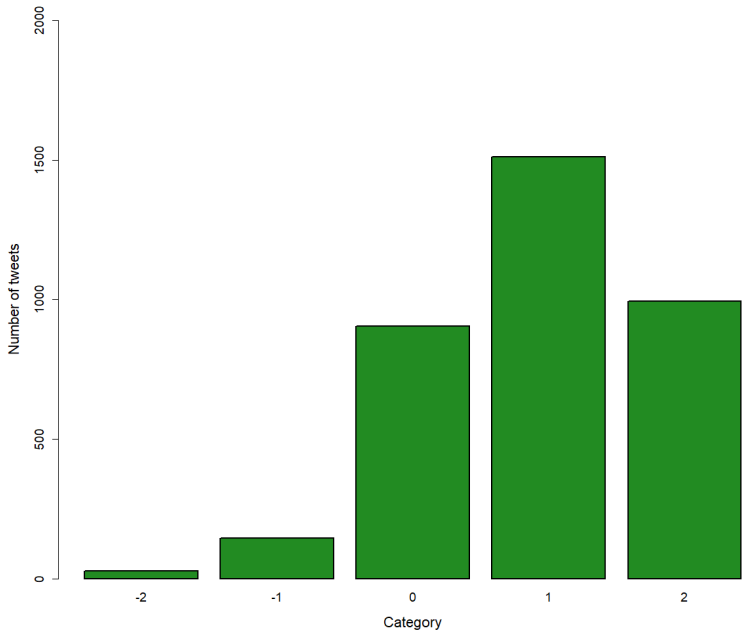

Figure 10. Manually categorized tweets

Figure 10 shows analysis of tweets done manually. As seen from figure 10, almost all the tweets received neutral or positive grade. Negative side has very little effect compared to others. Moreover, positively rated tweets have a big lead. Second in line are very positively rated tweets. This gives an outcome that people were well minded about this project.

4.4 Social network analysis

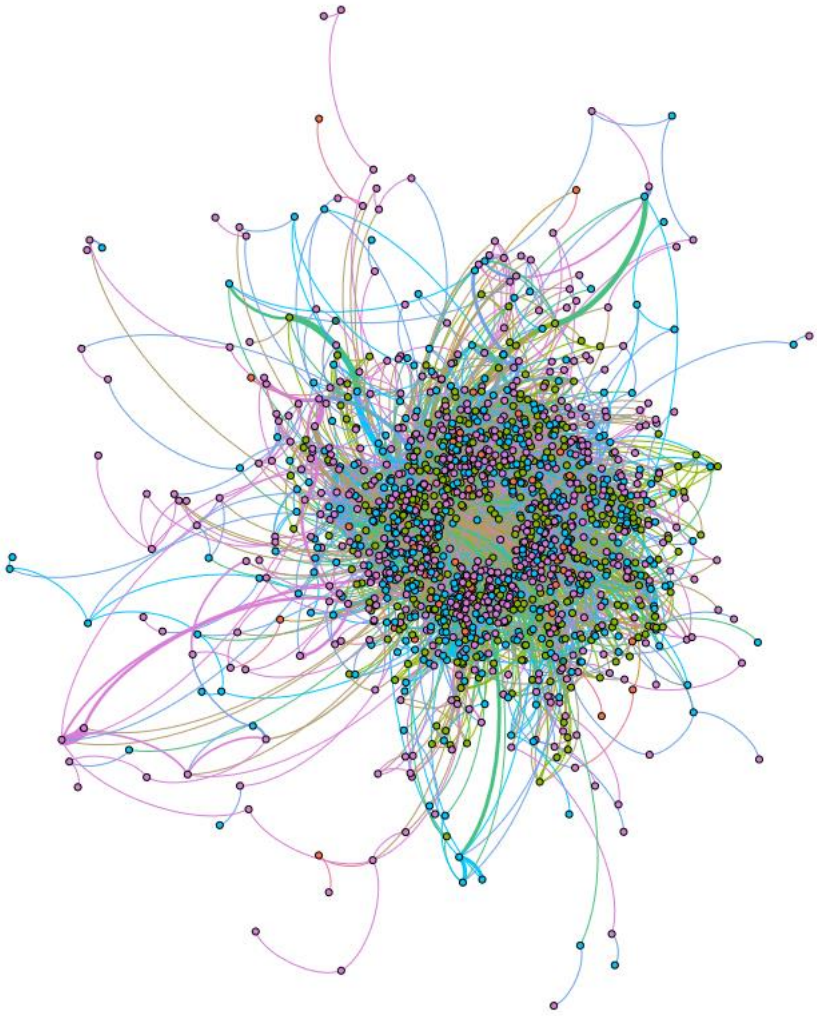

Figure 11. Visual representation of Twitter network as an undirected weighted graph

Figure 11 shows the results of network formation by using Gephi. The colours stand for ratings of nodes. -2 is dark green, -1 is orange, 0 is neon blue, 1 is purple and 2 is light green. In total there are 1541 twitter handles representing the nodes in the graph and 3044 edges between them. Only ~3% of nodes have a negative rating. Approximately ~26% of the nodes have neutral rating. Approximately ~44% of the nodes have a good rating and ~27% have an excellent rating. Results show that majority of the people involved in tweeting about topic had positive opinion and were happy towards the new speed limit and its benefits. Some of the people chose to stay neutral and a minor part of them were against the changes. Average

degree was small, which indicates that people mainly interacted with 20splenty handler and less among each other.

4.5 Geomap analysis

Figure 13. Map of mentioned places

Figure 12. List of mentioned places

Wimbledon Village	Birmingham	Castly Cary
Norwich	Dublin	Ansford
Dunbartonshire	Sefton	Halewood
Wandsworth	Grangetown	Lewisham
Torquay	Hastings	Acton
Whitekirk	Frieth	Croydon
Shifnal	Hopeman	Dunstabe
St Briavels	Bristol	Faversham
Stithians	Oxford	Bruton
Lanners	Leicester	Dundee
Windelsham	Merseyside	Edmonton
Edinburgh	Somerset	Victoria
London	Lancashire	Lambeth
Barrhill	St John	Southwark
Denbury	Ridingmill	Atkins
Sonora Fields	Calgary	Otford
Glasgow	Gabalfa	Wiltshire
Manchester	Sheffield	Devon
Otley	Westminister	Havenstreet
Calderdale	Isle of Wight	York
Kent	Tunbridge Wells	Sittingbourne
Tunbridgewells	Trafford	
Liverpool	Tower Hamlets	

Figure 12 shows the list of mentioned places and figure 13 shows mentioned places on the map of U.K. As expected, it is visible on the map that places mentioned are located all over the UK and not just only in one part of the country. Mentioned places are reaching from South to North. They are next to the shore as well as inland. In total there were 67 cities mentioned and 68% of the cities mentioned were from England and 20 % from Scotland were mentioned. This shows that people across the U.K. are conversing each other about this policy.

5. Conclusion

In order to make streets safer for vulnerable road users and to increase people's physical activity in the cities of the United Kingdom, a new lower speed limit was implemented in downtowns, residential areas and school areas.

The objective of this thesis was to analyse people's opinions on the new speed limit policy by analysing their interactions on social media with the speed limit policy handle on Twitter. The analysis used a multidimensional approach consisting of descriptive analysis, topic analysis, sentiment analysis, social network analysis and geomap analysis.

The data given was collected from Twitter. After removing empty tweets from the total amount, 3413 tweets were left for studying. The first analysis done was descriptive. Tweets were analysed by the number of tweets done in quarters, days of the week and hours of a day. Results showed that the lowest amount of activity was in the 2nd quarter. The activity in other quarters was equal. Mostly tweets were posted during daytime of the workweek from Monday to Friday.

Second analysis was topic analysis. First part was text mining, which gave an overview of most frequently used words. Most frequent words were *speed* and *road*. It was also seen that the conversation mostly revolved around *community, city* and how this policy is *better* in connection with *safety,children and residential areas*. Second part was manually analysing tweets for subtopics. These subtopics were speed zones, health and different behaviours of driving.

Third analysis was sentiment analysis. Results showed *positivity* being in front with a huge lead followed by *trust*. Third place was somehow taken by *negativity*. This indicates that some people were against changes. Closely behind was *anticipation*. Other emotions did not have significant scores. Manual analysis showed very low scores on negative side. Positive side won with flying colours. To summarize, mostly people were fond of these changes and happy about the new limits, were looking forward to safer streets, but there were some who opposed the new speed limit policy.

Fourth analysis was social network analysis. Results show positive side again as the winner as only ~3% of the involved people were negative. Final analysis was geomap analysis that

showed most of the cities mentioned in tweets were located in England, followed by cities in Scotland.

In conclusion, people's response in social media to the new speed limit policy was mainly positive, and people were happy about the changes.

For future perspectives same analysis could be used over larger data amounts to get more differentiated results. Another option is to add some other types of analysis to get more different aspects of the opinions. Furthermore, if data would be collected from other countries on same policies it would be possible to compare results between different countries.

6. References

City, N. Y. (2014-2016). *AUTOMATED SPEED ENFORCEMENT PROGRAM REPORT.* New York City: City of New York.

Ferrier, K. (7. December 2017. a.). *When Lowering Speed Matters Most – Successful Strategies from NYC and Chicago.* Kasutamise kuupäev: 17. April 2019. a., allikas https://visionzeronetwork.org/when-lowering-speed-matters/

Gephi. (Online). *Gephi - The Open Graph Viz Platform*. Kasutamise kuupäev: 6. May 2019. a., allikas https://gephi.org/

Group, T. a. (April 2013. a.). *An evaluation of the estimated impacts on vehicle emissions of a 20mph speed restriction in central London.* Kasutamise kuupäev: 9. May 2019. a., allikas https://www.cityoflondon.gov.uk/business/environmental-health/environmental-protection/air-quality/Documents/speed-restriction-air-quality-report-2013-for-web.pdf

Ingo Feinerer, K. H. (21. December 2018. a.). *Package 'tm'*. Kasutamise kuupäev: 6. May 2019. a., allikas https://cran.r-project.org/web/packages/tm/tm.pdf

ITSMR. (September 2015. a.). *Summary of New York City Motor Vehicle Crashes.* New York City: New York State Department of Motor Vehicles.

J.E.WOOLLEY, R. C. (2002). *IMPACTS OF LOWER SPEED LIMITS IN SOUTH AUSTRALIA - ScienceDirect.* Kasutamise kuupäev: 20. April 2019. a., allikas https://www.sciencedirect.com/science/article/pii/S0386111214600388#bib13

Jeffery Archer, N. F. (2008). *The Impact of Lowered Speed Limits in Urban Areas.* Kasutamise kuupäev: 21. April 2019. a., allikas http://casr.adelaide.edu.au/rsr/RSR2007/ArcherJ.pdf

Jockers, M. (13. December 2017. a.). *Introduction to the Syuzhet Package*. Kasutamise kuupäev: 6. May 2019. a., allikas https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html

King, R. (12. February 2018. a.). *BRISTOL'S 20MPH LIMITS HAVE LED TO VALUABLE REDUCTIONS IN SPEED AND CASUALTIES, AND BENEFIT ACTIVE TRAVEL.* Kasutamise kuupäev: 9. May 2019. a., allikas http://www.20splenty.org/bristol_20mph_analysis

Laboratories, B. (1995). *What is R?* Kasutamise kuupäev: 6. May 2019. a., allikas https://www.r-project.org/about.html

Lindblom, M. (16. September 2016. a.). *Seattle may drop speed limits to 25 mph on arterials, 20 mph in neighborhoods .* Kasutamise kuupäev: 18. April 2019. a., allikas https://www.seattletimes.com/seattle-news/transportation/city-could-reduce-downtown-residential-speed-limits/

Luis Cajachahua, I. B. (September 2017. a.). *Sentiments and Opinions From Twitter About Peruvian Touristic.* Kasutamise kuupäev: 8. May 2019. a., allikas http://ceur-ws.org/Vol-2029/paper17.pdf

Network, V. Z. (14. March 2018. a.). *Taming Speed for Safety.* Kasutamise kuupäev: 18. April 2019. a., allikas https://visionzeronetwork.org/wp-content/uploads/2018/03/0314_CaseStudy_Speed_FINAL.pdf

Python. (Online). *Python*. Kasutamise kuupäev: 8. May 2019. a., allikas https://www.python.org/

R-project. (3. 1 2018. a.). *Wordcloud2 introduction*. Kasutamise kuupäev: 8. May 2019. a., allikas https://cran.r-project.org/web/packages/wordcloud2/vignettes/wordcloud.html

RStudio. (Online). *RStudio*. Kasutamise kuupäev: 8. May 2019. a., allikas https://www.rstudio.com/products/rstudio/

Vendel, C. (20. March 2019. a.). *How to make streets safer for pedestrians? To start, red-light cameras and a police 'strike team'.* Kasutamise kuupäev: 18. April 2019. a., allikas https://www.pennlive.com/news/2019/03/how-will-harrisburg-make-its-streets-safer-how-about-red-light-cameras-and-a-police-strike-team.html

Wasserman, S. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.

Wen Hu, J. B. (January 2019. a.). *Lowering the speed limit from 30 to 25 mph in Boston: effects on vehicle speeds.* Kasutamise kuupäev: 9. May 2019. a., allikas https://www.iihs.org/bibliography/topic/2168

Zhao, Y. (2013). *Analysing Twitter Data with Text Mining and Social Network.* Kasutamise kuupäev: 7. May 2019. a., allikas https://www.researchgate.net/profile/Yanchang_Zhao/publication/264384706_Analysing_Twitter_Data_with_Text_Mining_and_Social_Network_Analysis/links/54917a160cf214269f296bba.pdf

Non-exclusive licence to reproduce thesis and make thesis public

I, Joosua Adler,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Multidimensional analysis to understand public perception for a health initiative project,

supervised by Rajesh Sharma and Ruth Hunter.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Joosua Adler

10/05/2019