Visualization of Twitter Trends in 12-15 August 2020 Riot in Netherland with Microsoft Power BI

M. Serkan Kilic[1]

Introduction

This visualization and document prepared for the assignment week-4 of "Microsoft Azure Data Engineer Trainee" at ITPH Academy[2].

In 12 August 2020 night in Schilderswijk, Den Haag some young men group threw rocks and fireworks to police, damaged bus stops and set at least one car on fire. Then the riot spread to Kanaleneiland, Utrecht[3]. According to youth workers and experts, months of boredom, the heatwave and the ongoing corona crisis had to wait for problems[4]. Dozens of young men have been arrested.

This situation can also be deduced from the images reflected on the Twitter, such that stones were thrown into the police vehicle, a fire was lit on the road in Den Haag, a bus stop and shop were damaged.

Question of The Study

Did Twitter has a role in the August 2020 Riot in Netherland?

The Framework of The Study

Twitter was founded in 2006 and has 330 million monthly active user. Roughly 42% of Twitter users are on the platform daily. 71% of Twitter users say they use the network to get their news. It is the very popular social media platform.

Twitter has #hashtag (hashtag) feature, users can create a personal agenda about a topic which they consider important and initiate a public conversation about any topic they identify[5]. Users can easily follow the Twitter agenda with Trend Topics (TT), which are determined by the most used words and topics on Twitter.

In this study, it is preferred to collect Twitter data under 2 popular hashtags (#schilderswijk and #Kanaleneiland) which is related to Riots which mentioned detailly above.

Limitation of The Study

Because of the aim of this study is using MS Power BI to visualization of tabular data, so the study was focused to visualization. Normally, in this studies, texts (tweets for this study)

---

[1] https://www.linkedin.com/in/mehmet-serkan-kilic/

[2] https://www.itph-academy.nl/

[3] https://www.dutchnews.nl/news/2020/08/disturbances-spread-to-more-cities-as-bored-youngsters-clash-with-police/

[4] https://www.tellerreport.com/news/2020-08-15-schilderswijk-in-the-hague--the-scene-of-riots--who-is-causing-the-unrest-and-why-.Bkeidq9HfP.html

[5] Ceng, E. p. 673, A Political Analysis of the Use of Twitter as a Perception Management Tool) Erciyes Publication, 5 (4), 663-689.

should be analyzed with NLP (Natural Language Processing) techniques. So, the unused of NLP techniques is the biggest limitation of this study.

Data Collection Phase

In this study, the data was collected from Twitter via Twitter API with Python. Firstly, a Twitter Developer account was created. Then, tweepy (which is Twitter famous open source library) was to connect the Twitter and raw data was obtained. After then, the raw data was manipulated (organized) with pandas, another famous open source Python library. Finally, the data saved to single file in CSV format.

As a result; under the #schilderswijk hashtag 17596 rows and under the #Kanaleneiland hashtag 7797 rows and 34 columns data collected which belongs to date between 12.08.2020-26.10.2020.

The full python code was published in GitHub[6] as public. Because of this study focused to visualization with MS Power BI, the details of data collection phase don't mentioned here.

Data Visualization Phase

Power BI is a business analytics service by Microsoft in 2011. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.[7]

After saved the twitter data under two hashtag in 2 CSV file, this files were imported to Power BI with Power Query Editor.

Normally the tweet_id should be unique but after viewing data with "Column distribution" it is understood that happened some mistake while data collected phase. In spite of some tweets have different content their tweet_ids are same. If there wasn't any problem, it can be easily eliminate the duplicated tweet by "Remove Duplicates" function. Because some tweets both have #schilderswijk and #Kanaleneiland hashtag. So, to solve the problem; new column "Hashtag_Name" created with "Conditional Column" function both for two table by containing the hashtag name or not.

Then two file merged with "Append Queries As New" function with the table name of "HashtagBoth". Firstly the duplicated rows removed with the help of "Hashtag_Name" column. After this operation the rows number of HashtagBoth table decreased from 25393 (17596 rows from #schilderswijk and 7797 rows from #Kanaleneiland) to 24272.

Then creating a new column, "tweet_language_long" with "Append Queries As New" function for full name of language names from language codes, like "nl" to "Dutch".

Because of "tweet_create_date" column consists of date and time together, 2 new columns were created and the date and time data were separated with DAX function as below.

```
tweet_create_date_only = FORMAT('HashtagBoth'[tweet_create_date],"DD-MM-YYYY")
```

For only hour data;

---

[6] https://github.com/msklc
[7] https://powerbi.microsoft.com/en-us/

```
tweet_create_time = FORMAT('HashtagBoth'[tweet_create_date], "hh:00")
```

It is also needed total user and total tweets number. So two measures were created as;

```
Total User = DISTINCTCOUNT('HashtagBoth'[username])
```

and

```
tweet_sum = COUNT('HashtagBoth'[tweet_id])
```

While visualization the data totally 8 different charts were used. For visualization of the tweets texts as word cloud, WordCloud 2.0.0 was imported from Power BI AppSource by "Get More Visuals" option.

- 2 Cards
- 2 Multi-Row Card
- 1 Stacked Bar Chart
- 1 Area Chart
- 2 Table
- 1 Pie Chart
- 1 Donut Chart
- 1 WordCloud

Result of The Study

After the visualization the Twitter data, it is easily seen that;

- Twitter was used highly during the 12-18 August 2020 which is parallel to the riots. If the usage was continued, it could be mentioned about trying to affect people for violence. But with this data, it could not be claim.
- 24272 tweets were shared by 6479 user. %99.79 of accounts was created before the 12 August 2020. The oldest account was also created in 2006. Only 31 account was created after 12 August 2020. So it could not be claim that lots of accounts created newly for manipulate the people.
- 23764 tweets were send from normal platform (phone, tablet and PC). Only %2 of all tweets were send from 3.party applications. So, it could not be claimed that tweets were send automatically (scheduled) from one center.
- Tweet shared hours are also not remarkable / abnormal. It is seen that the shared hours were parallel to normal Twitter usage.
- 18 different language used for 24272 tweets. But the %96.70 of tweets shared in Dutch. So there isn't any suspicious situation exists with the used language.
- 6479 Twitter accounts have 1514 followers in average. Some of the accounts have more than 500k followers. With this network these 24272 tweets seen by nearly 3.7M user. In spite of this huge network not much people attended the violence.

As a result, **it could not be claimed that Twitter played an important role to support the riot for spreading between the people.**