**A screenshot of a cell phone

Description automatically generated**

**Twitter Data:**

Using GetOldTweets3 (python library) import tweets between 2010 to 2019 with search keywords including #latraffic, #losangeles, #lapd to correlate LA road traffic collisions and their trend analysis from social media tweets. Here, trend in tweets are correlated directly to the number of accidents/collisions reported on Kaggle’s collision dataset.

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

After pre-processing tweets collection has 252,432 rows and 18 columns (including derived columns such as date, year, hour, minute, month/monthname, weekday etc.)

A screenshot of a cell phone

Description automatically generated

Dataframe details below

A close up of text on a white background

Description automatically generated

**Trend Analysis by Date/Time**

* **Tweets across Years 2010 through 2019**

Below chart shows the number of tweets mentioning latraffic spread across 2010 to 2019. Twitter API returned more data on 2018 compared to other years in scope.

A screenshot of a cell phone

Description automatically generated

* **Tweets across Years 2010 through 2019 by Month**

Below chart shows the number of tweets mentioning latraffic spread by month across 2010 to 2019. Twitter API returned more data on December 2018.Hence the spike in month of December.

A screenshot of a cell phone

Description automatically generated

* **2017- 2018 Tweets by month**

In alignment with data analysis on Kaggle’s collision data – twitter analysis restricted to 2017-2018 data as well. Here, trend in tweets are correlated directly to the number of accidents/collisions reported on Kaggle’s dataset. Based on that, below table summarizes overall spread of number of tweets (collisions) by month.

A screenshot of a cell phone

Description automatically generated

Let’s, zoom in on 2017 tweets as the data is consistent across whole year. To balance out the numbers, whenever there is a seasonal change – the number of tweets seems to increase.

A screenshot of a cell phone

Description automatically generated



* **2017- 2018 Tweets by Weekday**

Moving onto weekday analysis. Chart below shows, higher number of tweets on Friday followed by Thursday and least on Sundays. Suggesting, more collisions/accidents towards end of the week than beginning of week. This pattern is also in alignment with Kaggle’s collision report suggesting more accidents on Fridays and least on Sundays.

A picture containing person, sitting, umbrella

Description automatically generated

* **2017- 2018 Tweets by Hour of day**

Further into hourly analysis - Chart below shows, higher number of tweets between 15 to 20h. Suggesting, more collisions/accidents towards end of the day as more people returning from work. This pattern is also in alignment with Kaggle’s collision report suggesting more accidents during this hour range. With 10 AM being the least suggesting less to minimal traffic during this window.

In addition, towards end of the month, there seems like more accidents/collisions based on the day of the month analysis followed.

A close up of a map

Description automatically generated

* **Wordcloud**

Finally, a simple Wordcloud depicting the overall talk about the words(tokens) trended in tweets are correlated directly to the number of accidents/collisions reported on Kaggle’s dataset.

A screenshot of a cell phone

Description automatically generated

* **Most trending Users**

A screenshot of a cell phone

Description automatically generated