Matthew Sutherland, Jordan Spector, Kenny Wu

Digital City: A Data Analysis of the City of San Diego

IST 736 Final

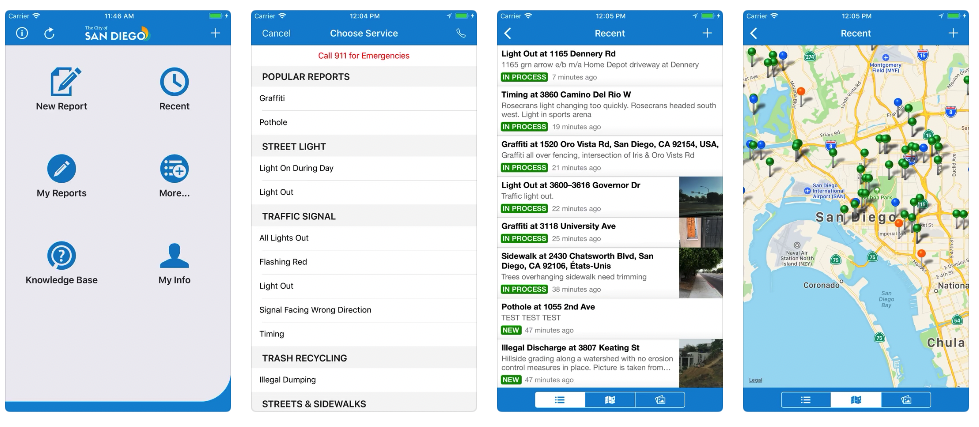
­

# Digital City: A Data Analysis of the City of San Diego

## Introduction:

Digital services are becoming pervasive across the United States in cities like San Diego. One of the services San Diego offers to its constituents is the Get It Done San Diego mobile app. The mobile app enables citizens and visitors to submit non-emergency type requests, commonly referred to as 311-requests.



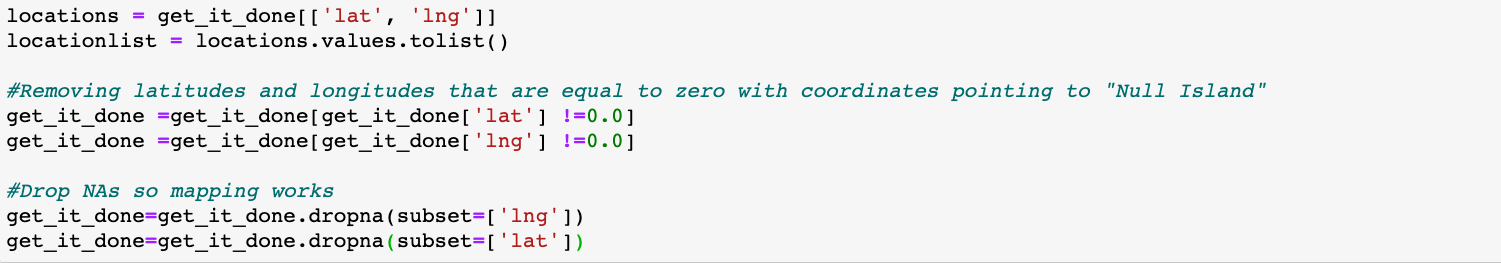
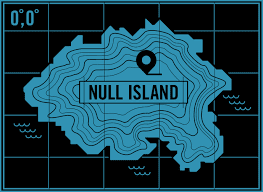


City service groups should aim to serve their constituents as best they can and cities like San Diego are at the forefront of promoting digital services to reach this objective. The data collected from the Get It Done San Diego app is publicly available and therefore readily available to analyze.

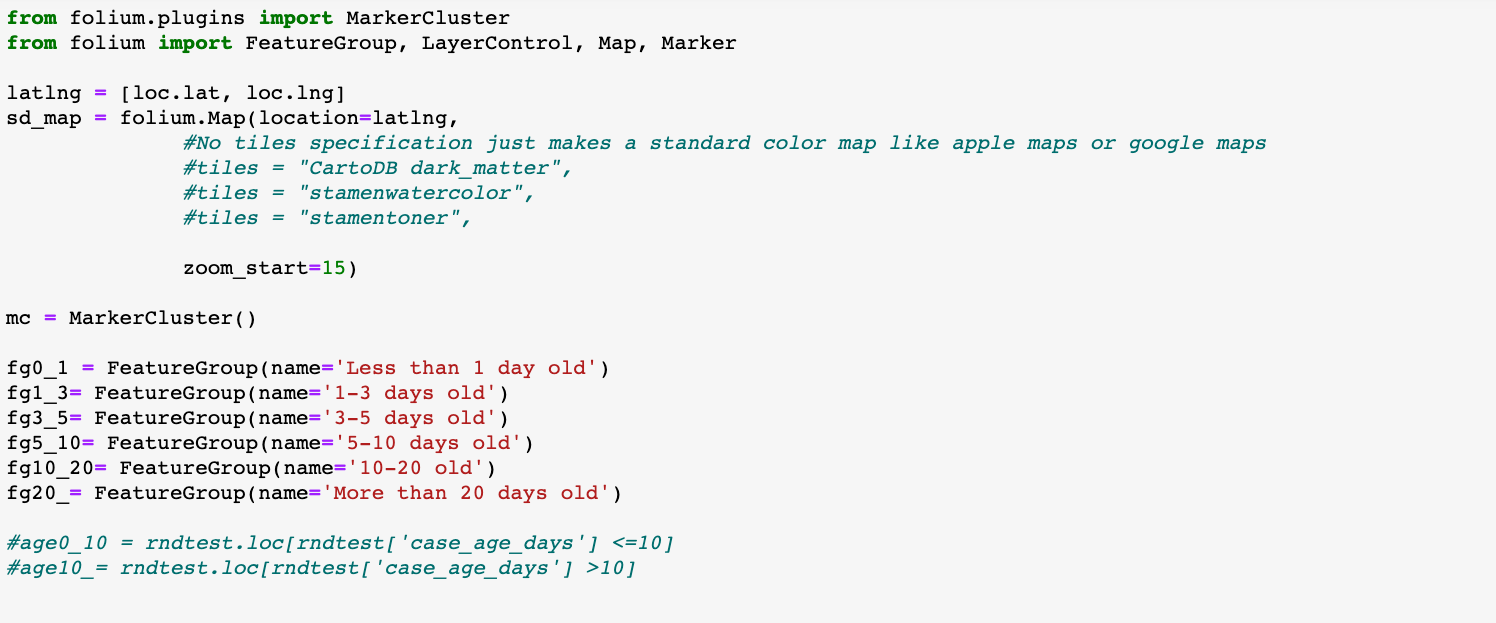
There are or course many other avenues to understand more about the needs of the people of San Diego. Looking at multiple different sources of San Diego related data creates a more complete picture of what is important to the city and its constituents and visitors. Twitter is another great resource to combine with the 311 data. Twitter similarly opens its data to the world (upon approval of a developer account) and is an incredible resource for topic modelling and sentiment analysis.

## Analysis/Models:

Latitude and longitude data for requests can be analyzed and then plotted on a map. It is important to specify that latitude and longitude does not equal zero. Requests that are missing data are plotted at what cartographers affectionately have labeled “null island”. Null island is located at 0,0 and located in the Gulf of Guinea off the west coast of Africa. Requests with no location data will be filtered out of the data set.



An open source base layer map is used to plot the location of the requests.

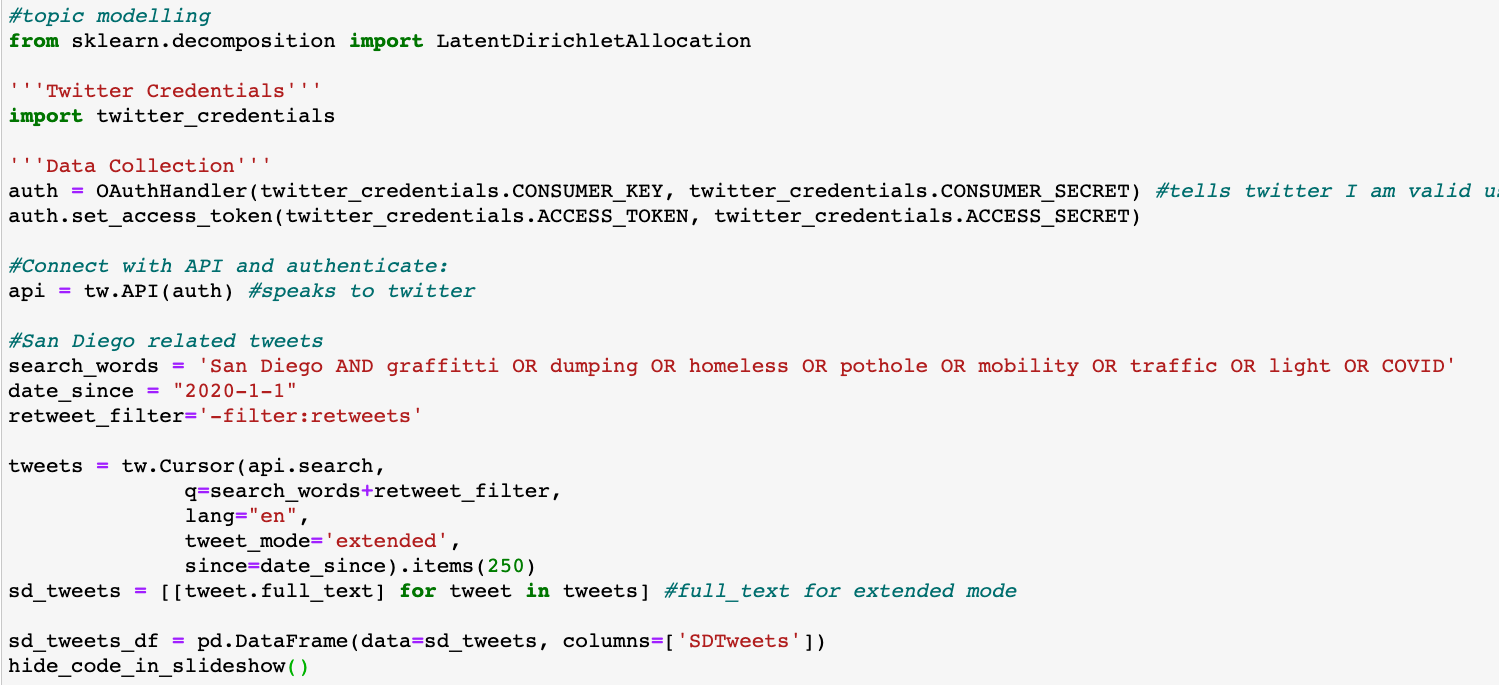


Toggleable layers are also added to filter the data based on request age. Request age is determined by how long a request has been open without having been closed / remediated. From a business use case perspective, the city does not want to have request fulfillment pending for months. The point of deploying a mobile application to intake requests is to respond to issues more quickly.

Visualizing the highest percentages of requests by type is important. Staffing the teams with the appropriate skills and experience can be forecasted with an accurate prediction of request types. Orchestration workflow processes for fixing a broken traffic light require a different field response team than an “Illegal Dumping” request workflow.

Predicting requests based on text data is also important in a similar way. The correct remediation team should be dispatched correctly the first time. Proper workflow orchestration saves the city money and is more efficient. It is useless and costly to deploy a garbage removal team accidently to deal with an encampment situation. A growing encampment might initially require social workers and police to respond. With traffic light outages, police field service technicians must be deployed. If instead the situationally useless graffiti removal team arrives on-site then there can be severe car accidents and even deaths. A broken traffic light request should kick off a remediation workflow to deploy a police officer to direct traffic and find the nearest located field service worker who has the skill set to fix the light. It would be even better to cross reference asset data to make sure that the field service technician is not only within close proximity to the issue but also confirm that they have the correct parts / assets to fix the traffic light.

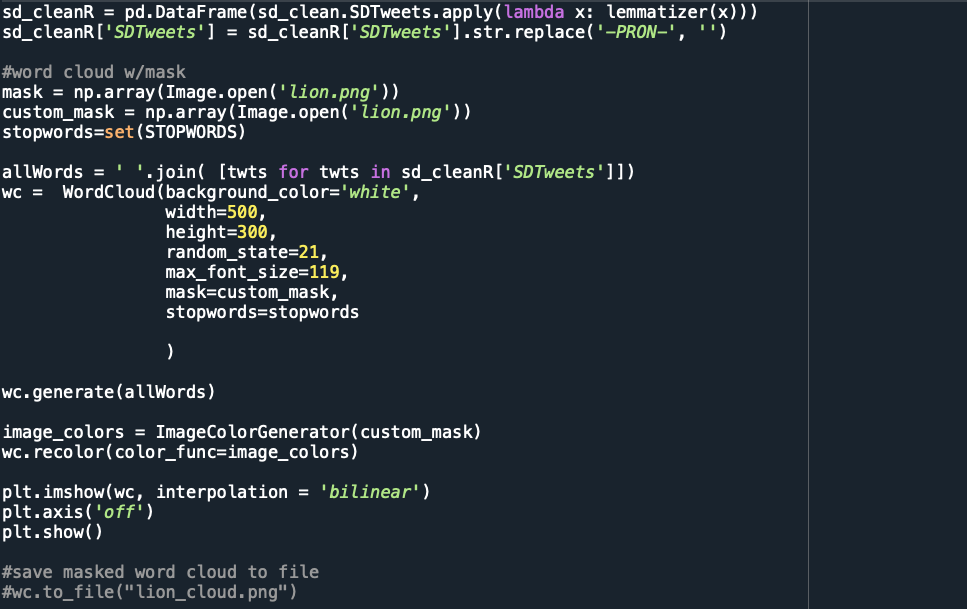
Data from Twitter can also be integrated into streamlining the city of San Diego’s 311 business service. Using the Twitter API data with city services related terms can be imported and then analyzed.



The data from twitter is not in a clean form and has to be cleansed before determining the statistical representation. A function is applied to the dataframe that uses regular expressions and word lemmetization. Word lemmatization is similar to stemming (finding the root word) except it brings more context to the words. Lemmatization links words with similar meaning to one word. This is performed by defining a lemmatization function and then applying to a now “cleaner” data frame.



To visually understand the top words used in San Diego tweets, word cloud are useful tools.

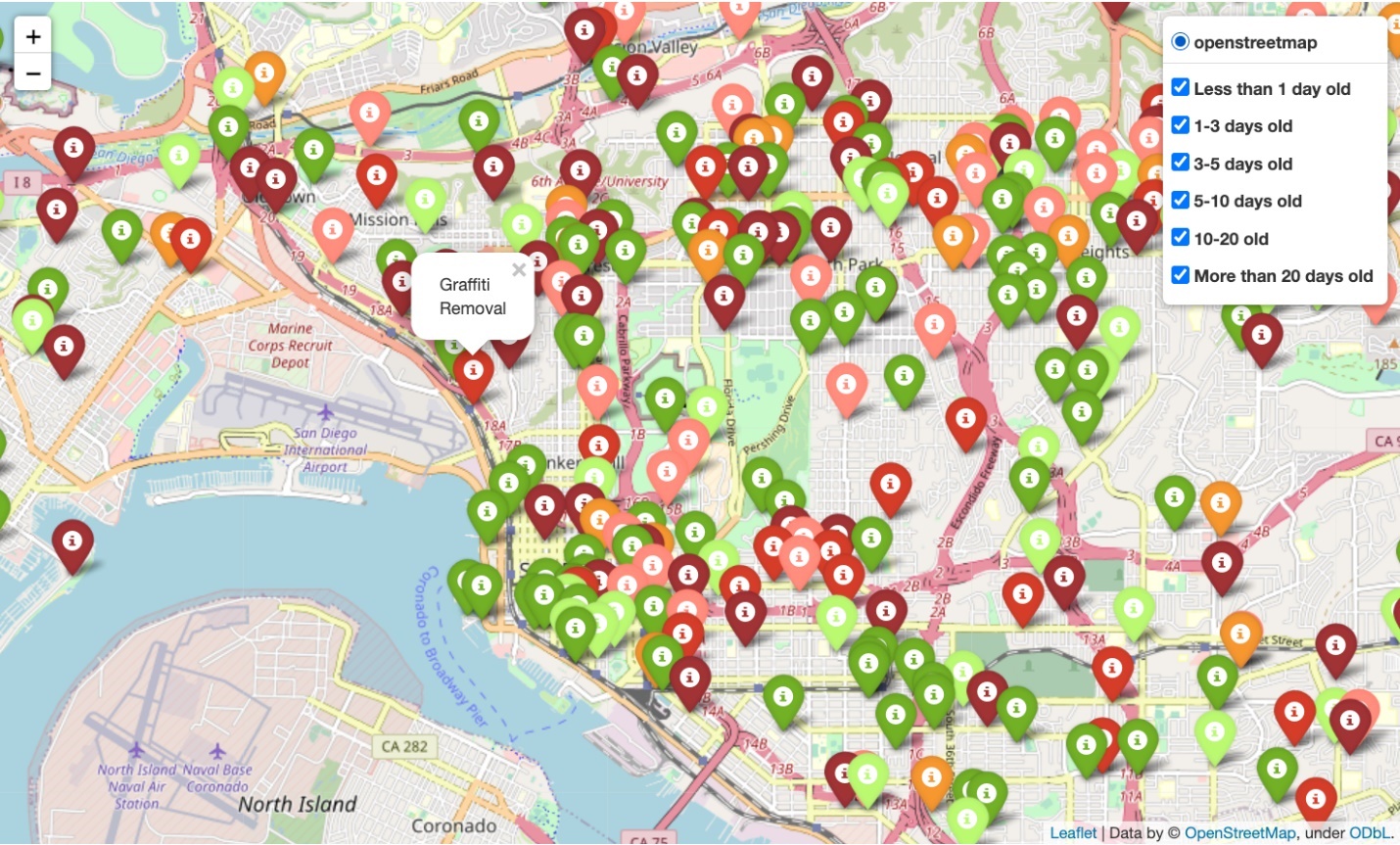


Two functions are defined to get the subjectivity and polarity of the twitter test using TextBlob. The two functions are then applied to the data frame to create two new columns for subjectivity and polarity. A third function is also applied that calculates sentiment score and its corresponding column is added to the data frame and populated with the results.

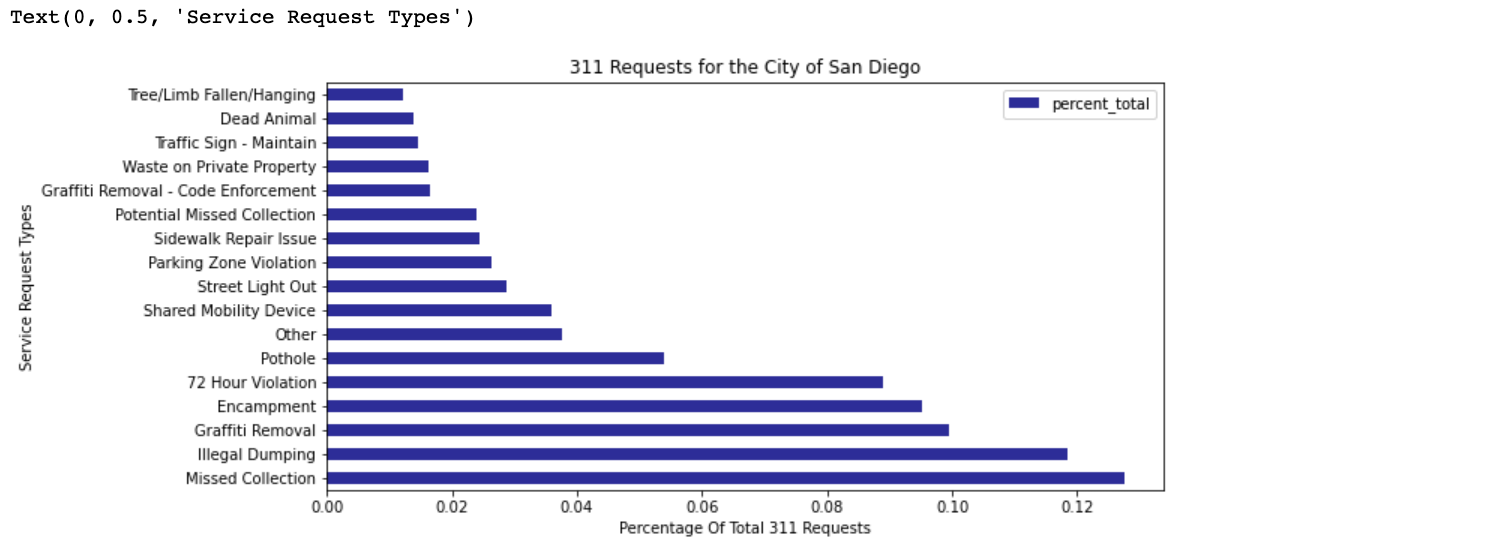


## Results:

The map displays all the most recent 500 311 requests with icons. The icons when clicked display the name of the request type. In the example below, one of the more common requests is highlighted, “Graffiti Removal”. The filter in the top right of the map enables the viewer to quickly filter between requests based on how long the request has been open. The icons are also intuitively color coordinated with the green icons representing new cases and the dark red icons representing requests that have been open and unresolved for over 20 days.

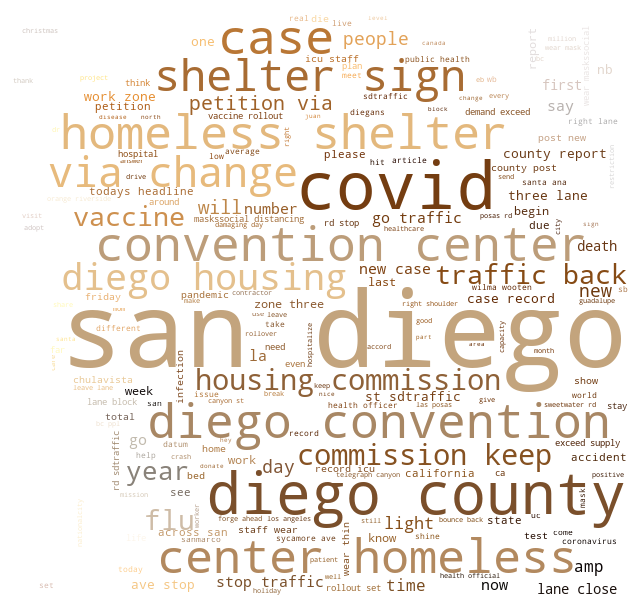


When compared with the other types of requests that were collected via the Get It Done San Diego app, Graffiti removal represents the third highest percentage of requests. The highest percentage of requests is represented as “Missed Collection”. It will be important for the city of San Diego to focus on these high volume requests from a business efficiency and continual improvement perspective. Also, the requests indicated in the map view as “more than 20 days old” is an attribute that should be targeted for improvement.

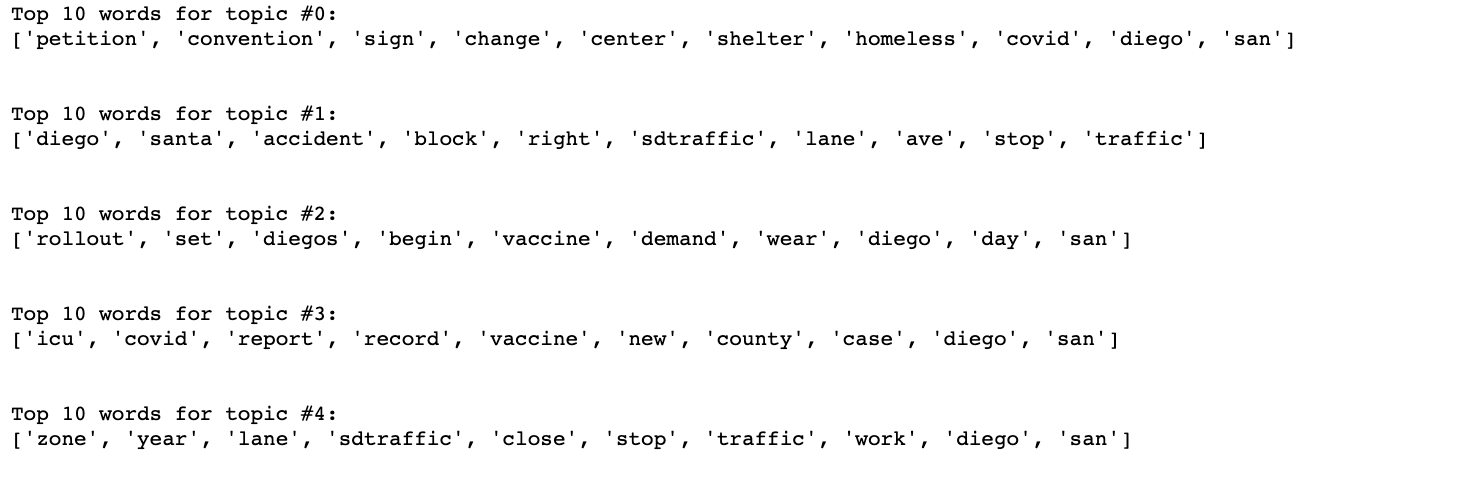


An image of a lion (San Diego Zoo them) is used as a “mask” to deliver the word cloud output. The word cloud displays top words used in the tweets represented by size. The “mask” picks up the shape of the lion and pulls in the colors from the lion image.





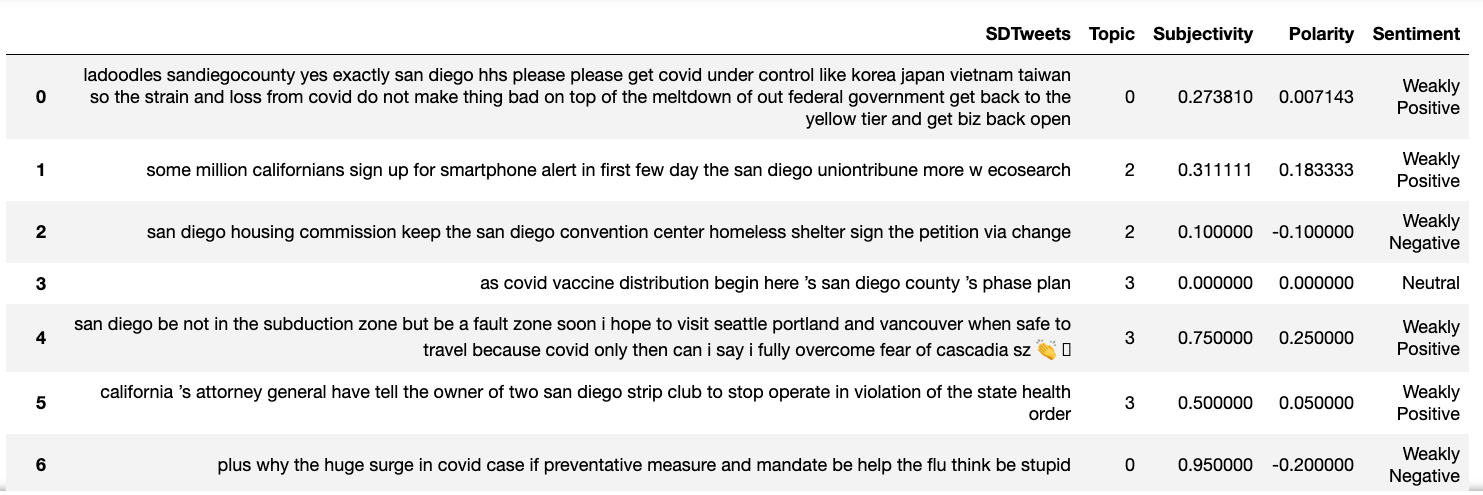
**5 topics generated from Twitter Data**



Out of the five topics the Latent Dirichlet Allocation model yielded, topic #4 seemed the most relevant to a non-emergency service request, while the others appeared as noise. Words like “sdtraffic,” “close,” and “stop,” indicate that there may have been a malfunctioning traffic light or another issue that might require city services.

Sentiment analysis of Tweets

With the three functions applied to the cleaned twitter data the sentiment of the tweets can now be observed.



Based on the nature of service requests, they generally tend to carry a negative sentiment label. Therefore, filtering tweets by negative sentiment and relevant topics allow for a targeted prediction of service requests.

## Conclusion:

For big metropolitan areas like San Diego, dealing with numerous requests for maintenance services can become an overwhelming task and a strain on city resources. By creating proactive measures for getting ahead of issues before they evolve into bigger ones, city workers can handle them with more efficiency. Social media channels like the Get it Done app and Twitter are great sources of information that can offer immense help in keeping a city functioning. As urban populations continue to grow, so too will the number of non-emergency service requests. Hopefully, these tools will enable San Diego and their counterparts to build up a strong infrastructure for managing them in the future.