UNDERSTANDING PUBLIC FEEDBACK FROM SOCIAL MEDIA AND USING IT TO IMPROVE CITY INFRASTRUCTURE

INTRODUCTION

It is difficult to get day to day public feedback on various government activities to improve city infrastructure or amenities. It is even harder to understand public behavior and impact. But, through social media we may slightly get to understand the public opinions. By understanding the public sentiments, opinions and behavior we may we be able to give valuable insights to governments, and companies.

It may also answer questions on current state of infrastructure of the city and how the development of new infrastructure can bring in more investments.

We can understand behavior of people in various cities and predict the behavior in another.

But here we will only focus on the understanding part of our project.

PROJECT OVERVIEW

I have extracted the data from social media in which the word Atlanta is appearing. By extracting the data for the given time frame. I can answer questions like how frequently people are talking about the City and in what languages people are talking about the city.

In the Data Analysis section I have shown the graphs and let the numbers speak the truth.

TOOLS

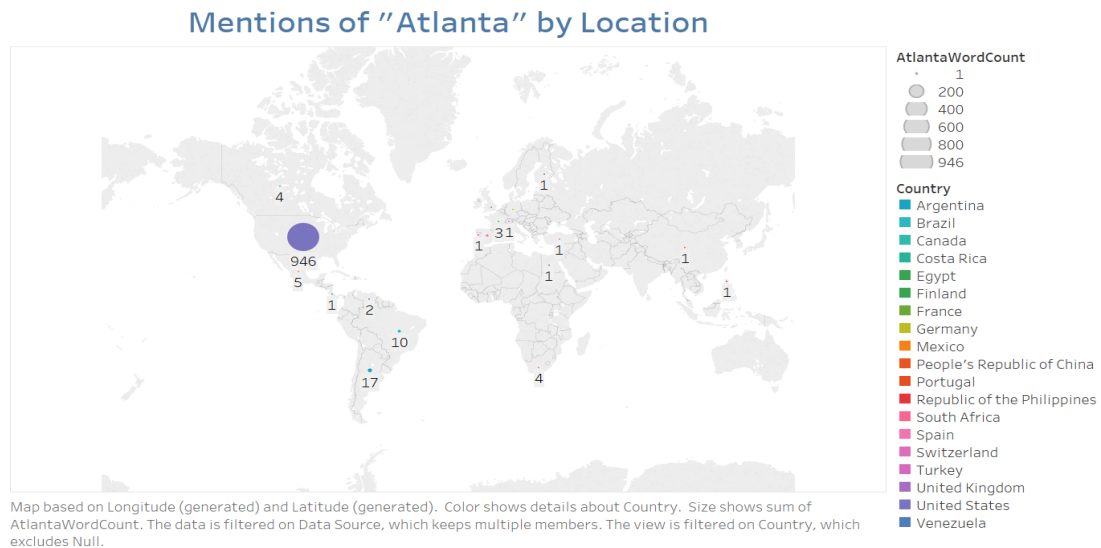Data Extraction on python and Analysis and Visualization on Tableau.

DATA

The Data has been extracted from various sites but mainly through sites like Reddit and Twitter. Below are the python packages used are-
1) Tweepy for twitter
2) Praw for Reddit

I have integrated the official API of twitter and Reddit by integrating with 3rd party python for packages. The data includes the tweets from Twitter and News Feed from Reddit. For my analysis I have extracted the Data for the city of Atlanta for the month of January, 2018.
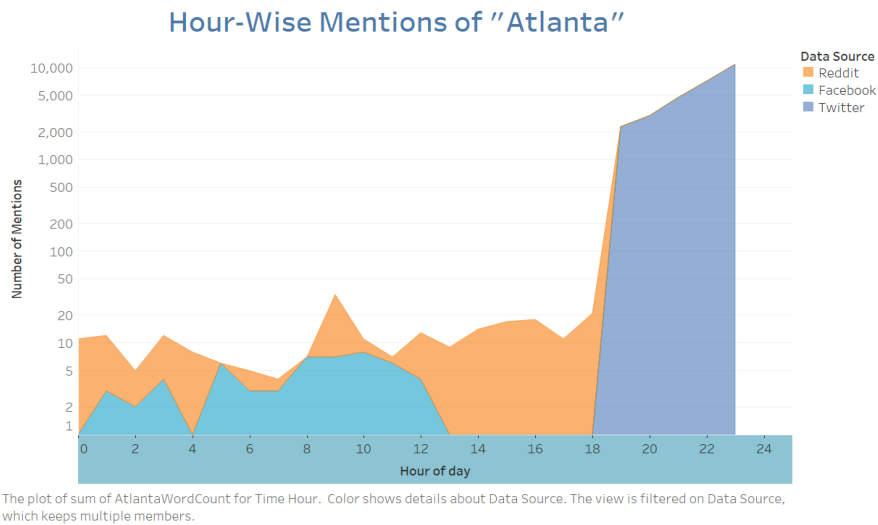
THE DATA EXPLORATION

Atlanta Popularity:

Mentions of "Atlanta" by Location

Map based on Longitude (generated) and Latitude (generated). Color shows details about Country. Size shows sum of AtlantaWordCount. The data is filtered on Data Source, which keeps multiple members. The view is filtered on Country, which excludes Null.

This graph shows an overview Atlanta word count in Data across the world.

In which hour of the day during which people about the Atlanta most

Hour-Wise Mentions of "Atlanta"

The plot of sum of AtlantaWordCount for Time Hour. Color shows details about Data Source. The view is filtered on Data Source, which keeps multiple members.

It can be seen that people mostly talk about Atlanta in the late evening or at night. Either before bed they talk about things happening in Atlanta as they discuss with family members.

Languages people spoke:

Mentions of "Atlanta" by Language

Language(group)
- English
- Other Languages
- spanish
- tagalog

1,463   280

25,760
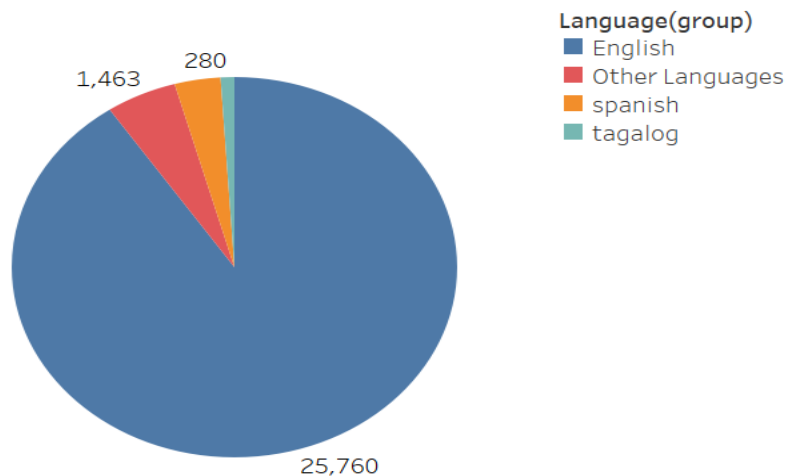
Language(group) (color).

Most of the mentions are in English language, but some are in Spanish and Tagalog (Philippines Language). The above chart shows can give us an insight of the people nationality or inheritance which gives us an indicator of their culture as well.

Conclusion:

In the above we can see the potential of this data and insights that it can give us on various parameter. Using this abundant unstructured data, we can get various analysis and might be able to predict the public feedback on certain occasions using classification algorithms. The public sentiment analysis can provide us information which can further be used to extract opinions.