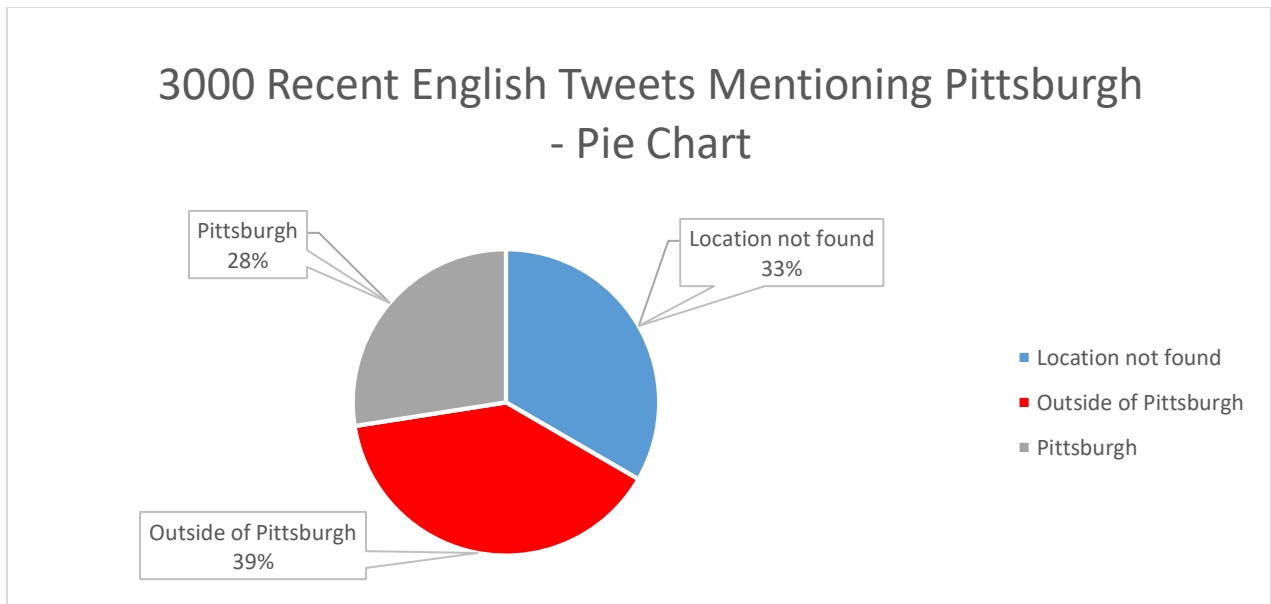
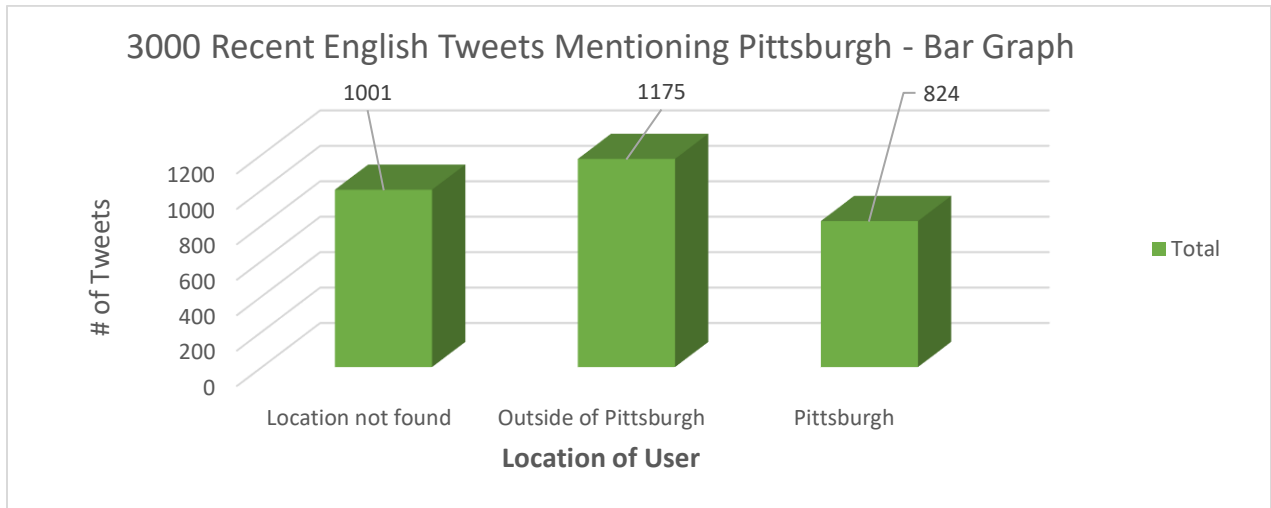


# Social Data Report

Joseph Bender

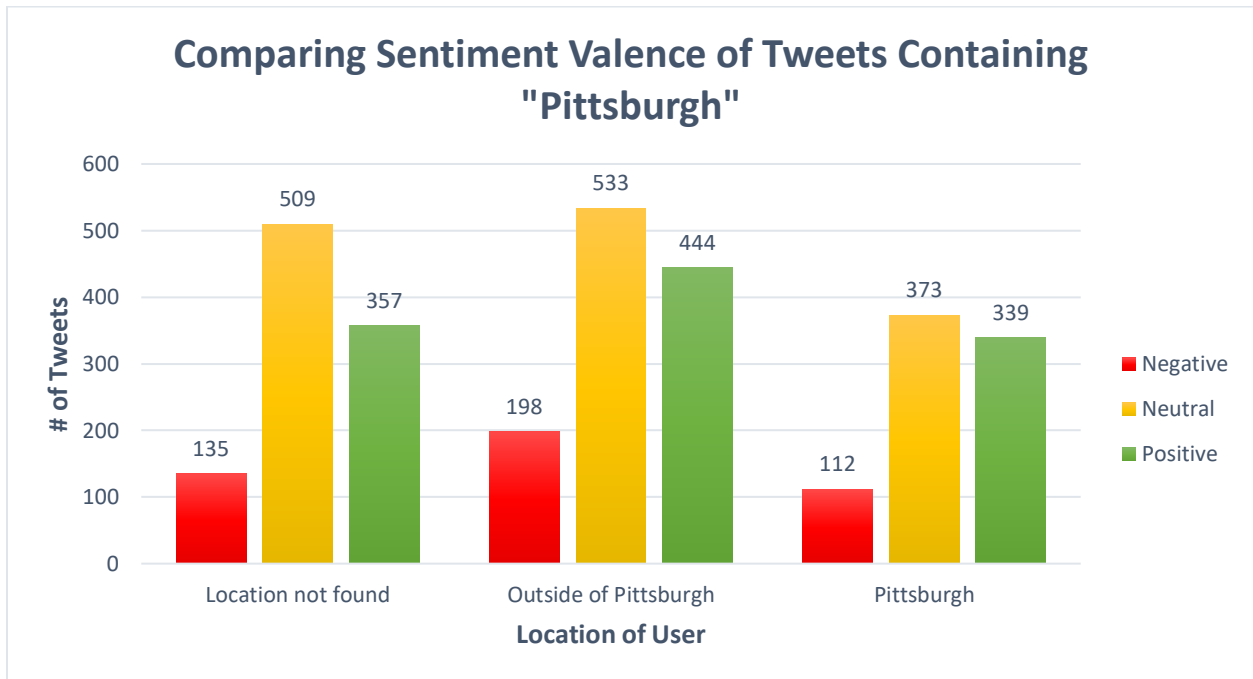
## Are tweets mentioning Pittsburgh written by people from Pittsburgh?



Tweets mentioning “Pittsburgh” can come from Pittsburgh, elsewhere, or unknown locations. It seems there is a fairly equal distribution between the three. There were slightly more coming from areas outside of Pittsburgh at 39%, and slightly less coming from inside Pittsburgh at 28%. However, the data is probably skewed because 33% of the tweets had no location attached and could have come from either inside or outside Pittsburgh. Although, if we use the known proportions as an indicator we can probably assume that a majority of the unknown tweet locations are from outside Pittsburgh. This would result in more significant findings that a

majority of the tweets about Pittsburgh come from outside the city. This is probably due to Pittsburgh having successful sports teams that are referenced by social media users all over the country.

***Does the valence of tweets mentioning Pittsburgh differ whether they are written by someone from Pittsburgh or not?***



Locations	Negative	Neutral	Positive	Totals
Location not found	135	509	357	1001
Outside of Pittsburgh	198	533	444	1175
Pittsburgh	112	373	339	824
<b>Grand Total</b>	<b>445</b>	<b>1415</b>	<b>1140</b>	<b>3000</b>

To normalize the sentiments, I divided the number of each sentiment by the total of tweets only from that specific location:

- **Unknown Location**
  - Negative ( $135/1001 = 13.4\%$ )
  - Neutral ( $509/1001 = 50.8\%$ )
  - Positive ( $357/1001 = 35.6\%$ )
- **Outside Pittsburgh**
  - Negative ( $198/1175 = 16.8\%$ )
  - Neutral ( $533/1175 = 45.2\%$ )
  - Positive ( $444/1175 = 37.7\%$ )
- **Pittsburgh**
  - Negative ( $112/824 = 13.5\%$ )

- Neutral ( $373/824 = 45.2\%$ )
- Positive ( $339/824 = 41.1\%$ )

Using these new proportional values we can compare the percentage of each sentiment for different locations:

- **Positive Tweets**
  - Unknown Location: 35.6%
  - Outside Pittsburgh: 37.7%
  - Pittsburgh: **41.1%**
- **Neutral Tweets**
  - Unknown Location: 50.8%
  - Outside Pittsburgh: 45.2%
  - Pittsburgh: 45.2%
- **Negative Tweets**
  - Unknown Location: 13.4%
  - Outside Pittsburgh: **16.8%**
  - Pittsburgh: 13.5%

This gives us a bit more information. As expected, the location with the highest proportion of positive tweets is from inside Pittsburgh. It is not surprising that citizens would enjoy the place that they call home. Also, the location with the highest proportion of negative tweets are locations outside of Pittsburgh. Since Pittsburgh does well in sports, it is probably fans of other teams being upset and tweeting negative things. In conclusion, residents of Pittsburgh are more likely to tweet positive things about their own city than outsiders.