**Title suggestions:**

1. **Open data public participation is mostly positive and inquisitive**
2. **Most public commentators on open data policy suggest improvements to the policy**

*Audience: users of open data*

A majority of commenters on draft open data policies recommend changes or ask questions related to the policy. 32% of total comments expressed enthusiasm whereas only 5% were dissatisfied.

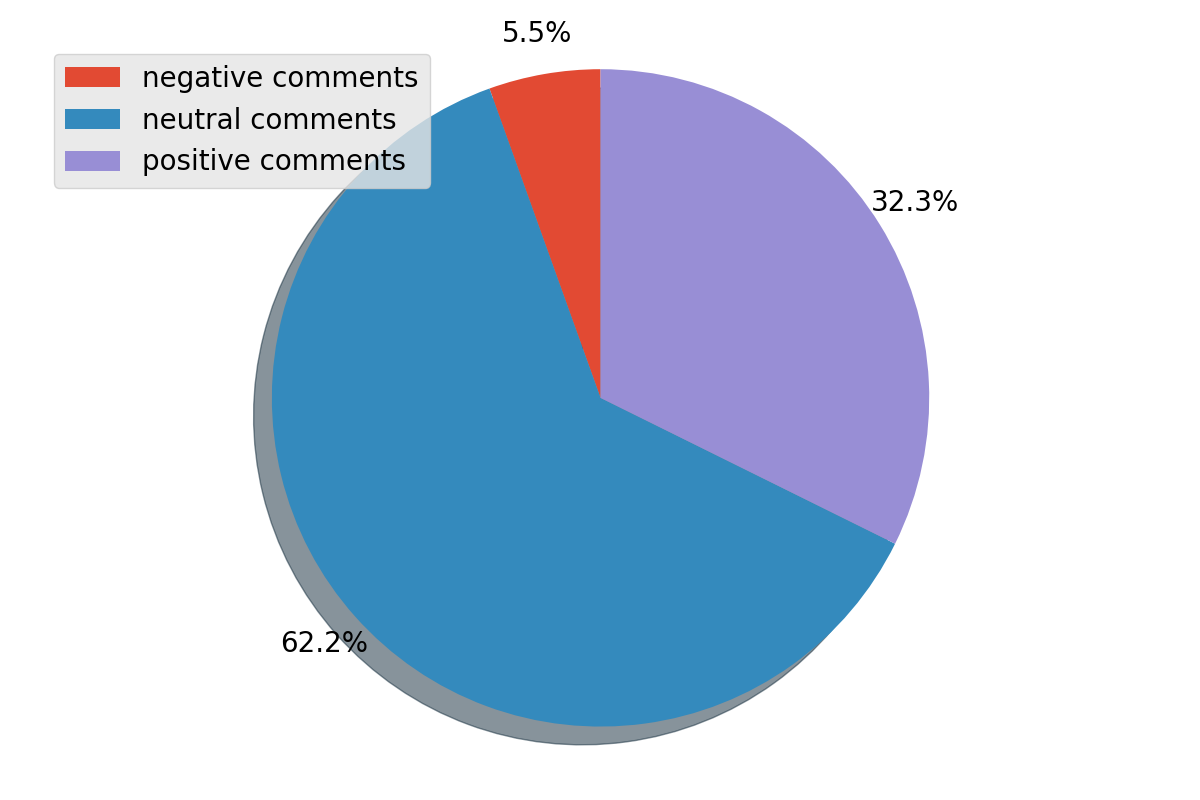
This blog post is the last one in the series of posts made to reveal insights from open data policy making. In this post, we will focus on the application of sentiment analysis, a tool to categorize a reaction as positive, neutral or negative, on comments made by users on draft open data policies.

***What is sentiment analysis?***

Sentiment analysis involves analyzing text to determine emotions. It takes a series of words and assigns a numerical score to it, representing how positive, neutral or negative it is. Then, based on some threshold, it is categorized as a positive, neutral or a negative sentiment.

While the field of sentiment analysis in computing has not fully matured [link], we were still excited to see the results a sentiment analysis tool can generate if we applied it to public participation on draft open data policies. Our goal is to identify topics that receive most positive, neutral or negative reactions.

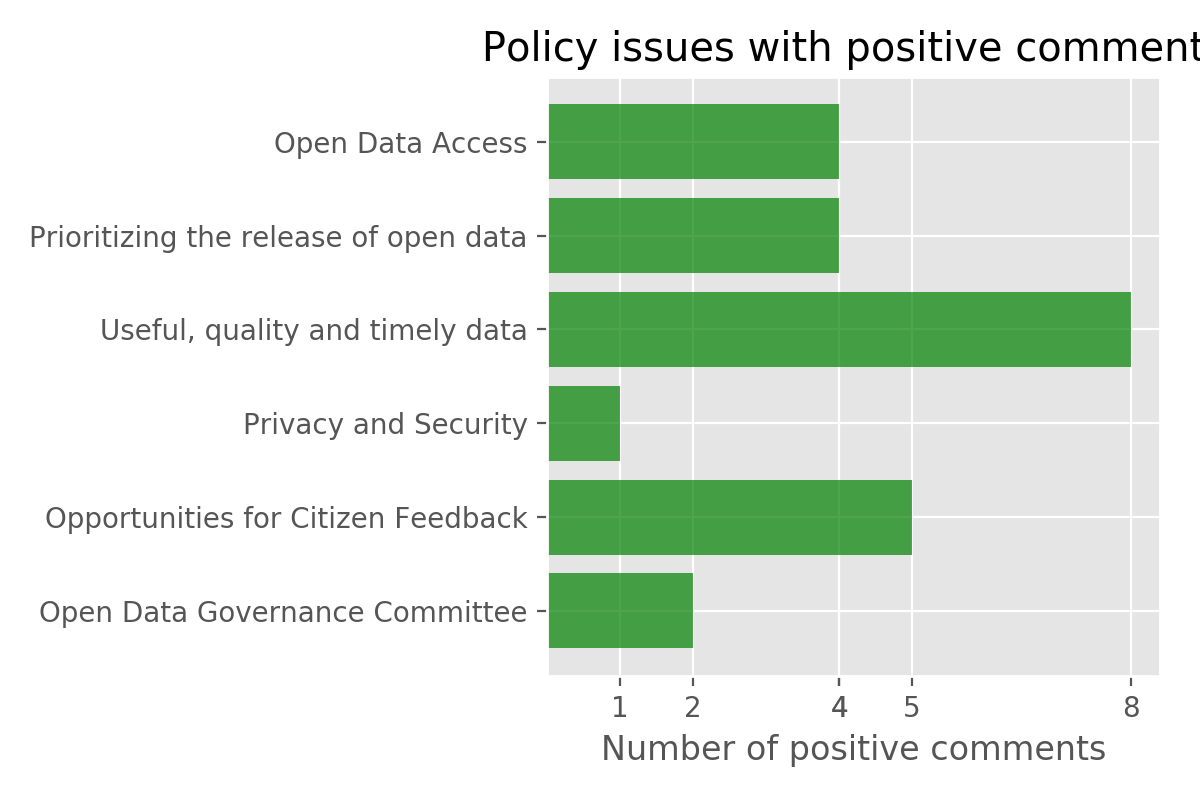
***Most users gave suggestions for improvements and asked questions***



Out of 164 comments, a majority of them were categorized as neutral. Neutral comments were either questions to understand policy better or suggestions for improvements to different policy sections. In general, people were curious to learn more about how confidential data will be handled, how public-records requests will be processed and whether they will be made available in a timely manner, and the potential for community organizations and academic institutions to host their data on open data portals.

Many people also made suggestions. Aaron Krolikowski, a researcher, on the “Privacy and Security” section of Buffalo’s open data policy, said:



A hefty number of users (32%) also expressed joy and enthusiasm. For example, Jurnell Cochren, a software developer, was delighted to see that the City of Nashville provided a Data Catalog, also known as metadata, for all its datasets so that users can understand and explore the datasets easily. A majority of positive comments cherished city’s capacity to make datasets interactive and friendly, keep open data policy a living document with feedback loops and provide data literacy to residents in order to access data.

There were a few comments that were classified as negative. A simple survey of those comments reveal that some of them were actually suggestions for improvements and did not really expressed disapproval of a certain policy section.

An example of a less favorable comment was about Glendale’s updated policy on getting citizen feedback. The comment emphasized the need to catch up with updated policy and stated that the city’s old policy was “awful” and “discouraged citizen involvement”.

***Is sentiment analysis always good?***

The major limitation posed by sentiment analysis for open data policy comments is that it hasn’t been trained on open data policy comments dataset. Given that the total number of comments (164) is insufficient to train a machine learning algorithm, we weren’t able to yield an accurate classification. Still, our model helped us make some sense of which policy issues generate what type of emotional response.