Identify the major complaint topics

-Establish this through the generation of tags with 50 topics

-show this using a topic pyramid, including the sentiment associated with each topic

-a Heat map displaying the relative frequencies of each tag within the data set

SHEET 1: TopicPyramid

We generate a population pyramid for the 50 different tags (or topics) (we have to use two measures (avg neg and pos sentiment)

We show how each topic ranges in sentiment showing excel having the largest positive sentiment on average amongst all categories

-We filter by year allowing the user to see how each of the topics change over time

SHEET 2: Heat Map

We generate a heat map showing the density of each topic within the set, we can filter this by creation year as well

SHEET 3: Tree Map

We generate a tree map to best illistrate the frequency of the topic based on years. Showing most significant years in the top left hand corner.

We can filter this plot by months within the creation date ranges to gain more insite into the monthly trends

Answers:

Since our answers are not linked to specific questions within our sets, as we took random samples of each, we begin to identify Positive contributors to answers by first examining the user accounts that contributed most often to answering questions.

SHEET 4: TOP10

Top 10 is a scatter plot overlaid on a bar chart representing the top ten user accounts by number of answers in the set. We Identify the positivity through the pink circles which represent the average positive sentiment aggregated by the user’s individual answer sentiment scores.