

## Code ▼

SGSF Executive Director Name: Loretta Williams Gurnell

Date: 6 Dec 2022

Code

The dataset refers to 18 SUPERGirls of SGSF Foundation. It includes data collected from a survey filled out by those Girls. Every SUPERGirl provided us with her first Name, last Name, home Address, email, cell phone number, date of birth, race or ethnicity, age,current grade,the name of the school that she is currently attending, her preferred mentor, her professional interest or expertise, if she has or has not reliable transportation in case she needs to meet her mentor, if she had a mentoring role before, her language proficiency, her classification, her major/minor or certifications. also, she answered many questions.

Looking ahead, of interest for the SUPERGirls SHINE Foundation will be to apply a sentiment analysis based on the SUPERGirls expectations from the mentorship program.

Here is a glimpse of what we will be working with.

Code

For this project, we are not going to dive deeper into cleaning the data, our main focus will be on the text columns only and particularly on the question: "What are your expectations in this mentoring program?"

In the first step, we will extract this text column from the dataset. Then we will start cleaning it by removing the undesirable words, especially mentioned in the question: "mentor", "Mentor", "mentoring", "Mentoring", "Program", "program".

After that, we will extract the words out of the sentences, and get only the words other than the undesired words, also we will get rid of the stop words using the "tidytext package" and those words that are less than 3 words such as "I", and finally we will get the counts of the words

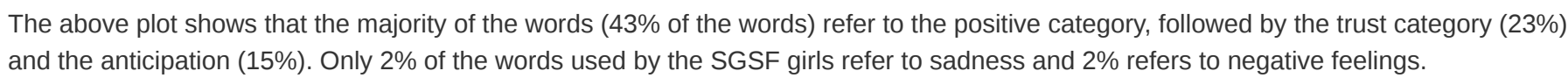
Code

Now, we will assign every word into NRC lexicon categories: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust by inner joining the words that we got previously with the NRC sentiments dataset provided by the tidytext package.

Code

Lets visualize our SUPERGirls sentiments.

Code



Lets look at the words for each category.

Code



Code

