

### **The research question we've decided on**

Our main research question is *"How could a book with so much bad attention, like Fifty Shades of grey, become a bestseller?"*.

Sub questions:

- Is the overall opinion of the book indeed negative or is this just an assumption made?
- Why did people give negative reviews just after the book was published?
- What is the general motivation of the negative/positive reviews?
- What is the motivation for people to start reading this book?

### **A short description of the steps we've planned**

In finding an answer to the sub question if the overall opinion of the book is indeed negative, we first of all want to look at the proportion of positive/negative votes using unix scripts.

Secondly, by looking at the usefulness of the reviews we want to check the findings of the question above? With unix commands we want to find out which ratings got the most votes for useful reviews. Besides that, we want to know whether we can speak of an instant negative hype or not. For this we want to use unix commands and look at the ratings just after publishment. With close reading we would also like to find out why the reviews just after publishment of the book were positive/negative. this could confirm the results of the unix commands. With the answers to these questions we will have a general idea about the overall opinion of the book.

To find out what the general motivation of people is to either like or dislike the book, we will first of all divide all the reviews in two parts. One part with all the ratings lower than 3, a second part with all the ratings higher than 3. By counting the words in both groups of reviews, some frequently reoccurring words could be identified per group, visualized in a wordcloud. By close reading on those frequently used words and identifying their contexts, we can see what people are actually saying.

For the question why people actually started reading this book, we will look at the contexts of motivational sentences like *"i started reading this book..."* or *"... decided to read ..."*. By close reading these motivations we will hopefully get insight in the general motivation.

We want to visualize using several programs, amongst which Tableau and online visualization applications. For example make a wordcloud online.

### **A short discussion of issues you've encountered and dealt with**

- We found out that there were reviews about random products in the dataset, which were written before the book was published. These results could falsify our findings so we are now thinking of a good way to filter these reviews out. This could make our results more plausible.
- There were some problems finding the right searchterm when searching for reasons the reviewers read the book. Since a lot of terms were too general to get useful results some close reading was needed in order to find the right terms.
- 

### **What you've learned about answering your research question so far**

First we wanted to find out whether the overall opinion about the book was really negative or whether this was just an assumption most people made. With the results we've got right now we could conclude that most people rate the book a 1.0 and therefore we could assume that the overall opinion about the book is really negative. Besides that, we found out that in the first months after publication of the book already more negative reviews were given than positive reviews. So we could also speak of an instant negative hype about the book. The positive reviews just after release are mostly written by loyal fans who read the fan fiction version of the book. With close reading we still have to find out why most reviews just after release were negative.

One other thing we've already learned is that the most given reason for reading the book is the popularity. Most reviewers say that they read the book because they wanted to see what the hype was all about.

Also two basic visualizations can be seen in the github repository.