



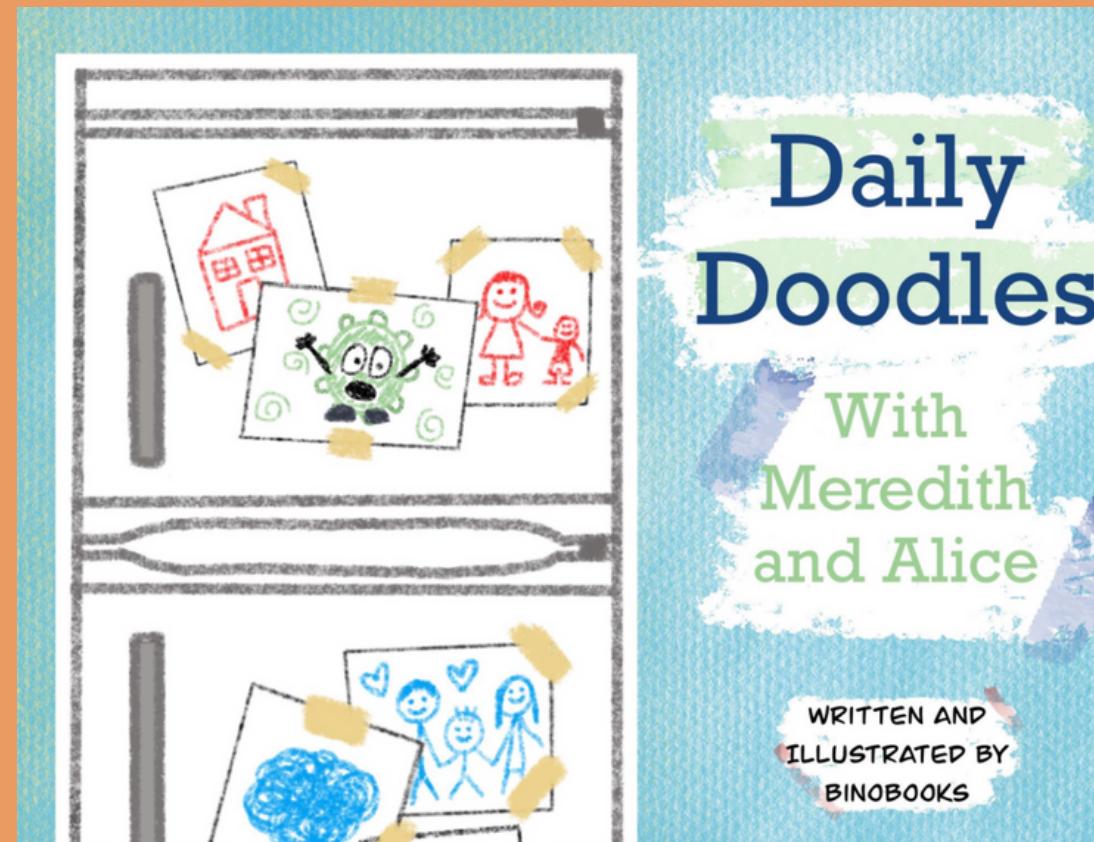
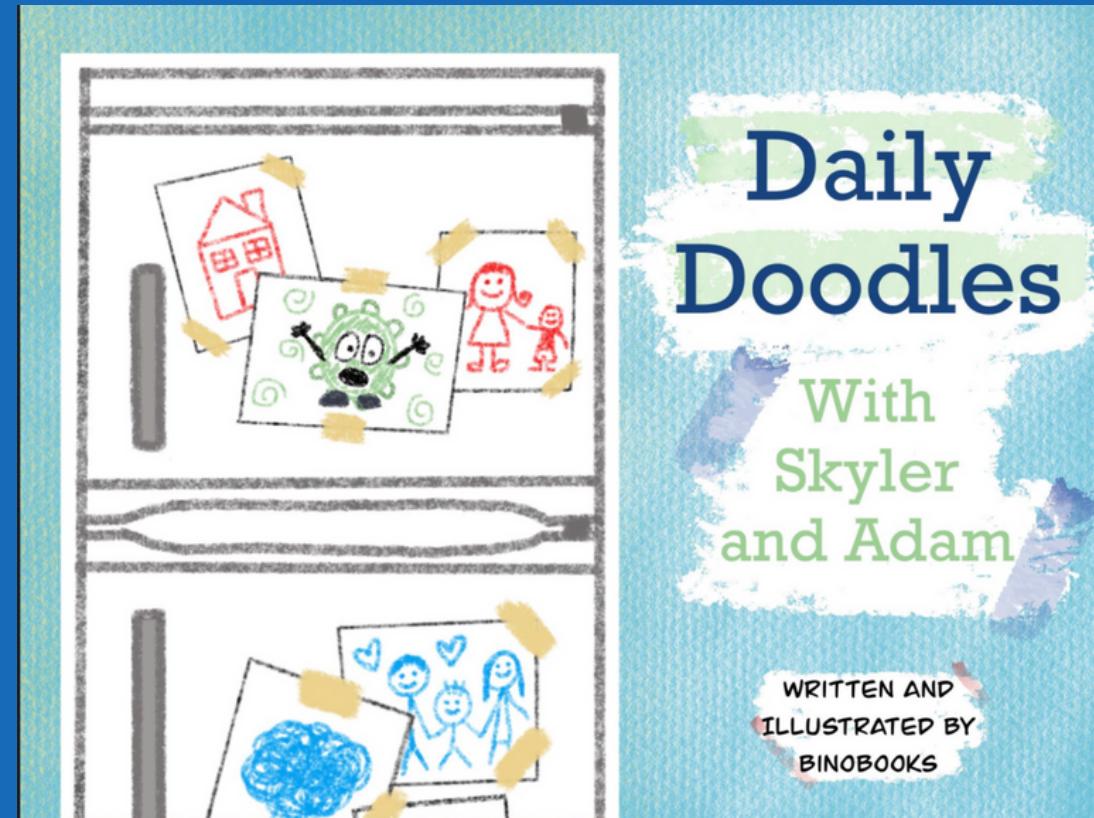
Using Data Science to Determine Popular Children's Books Elements

Jessica Dassanayake
CISC372



A customizable literacy platform

that helps parents explain complex topics to their children with ease while making all children feel represented in their readings.





Home

Choose your story

Customize

Payment

Child's name

Skyler

Location

Vaudreuil-Dorion

Hair length

Short

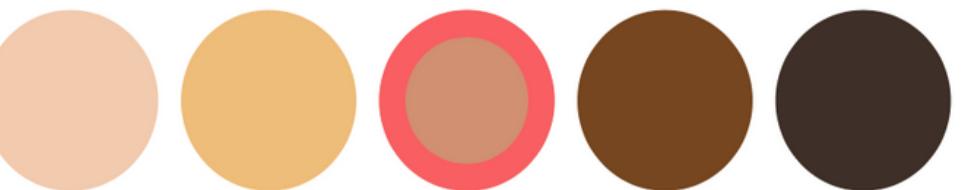


Hair texture

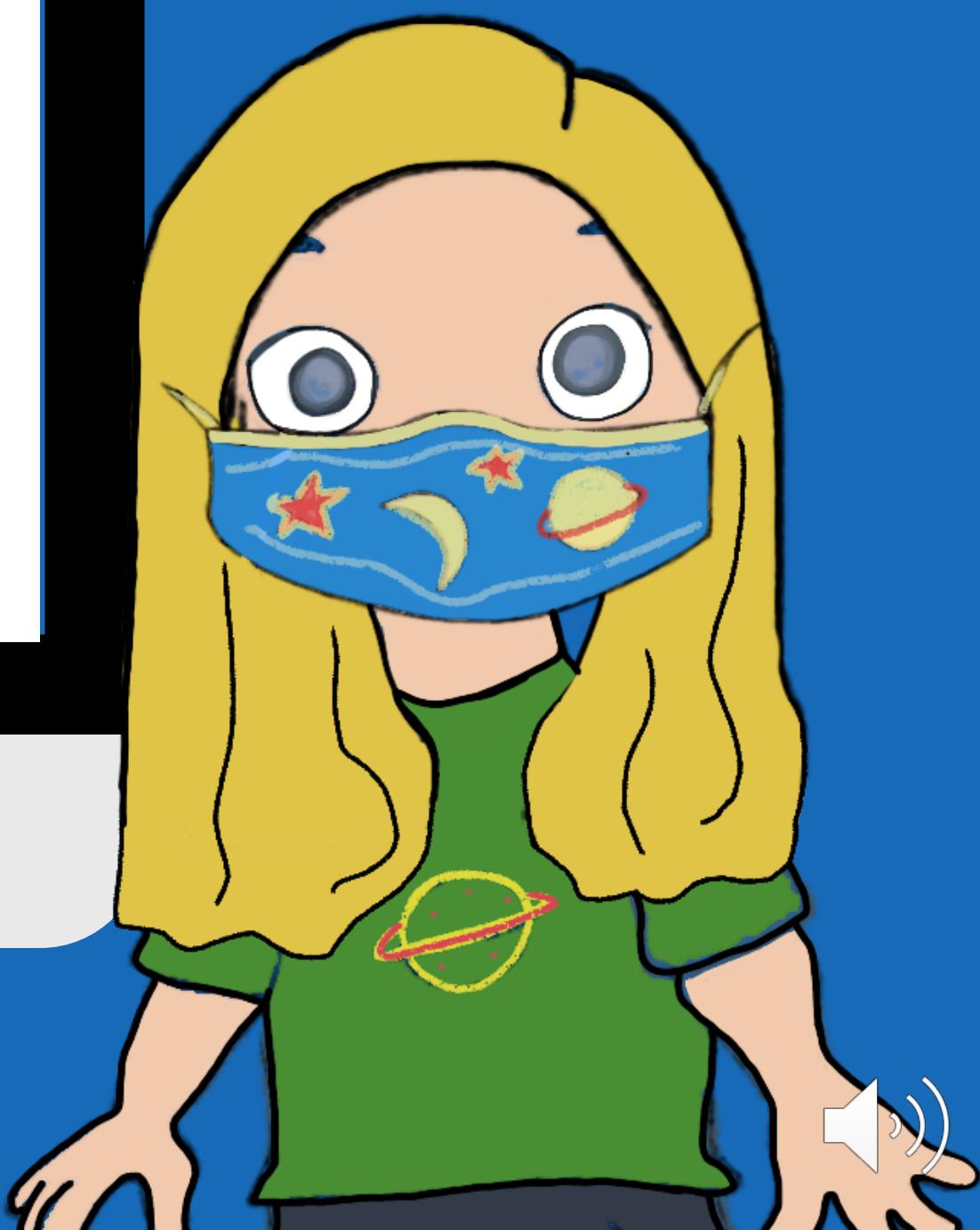
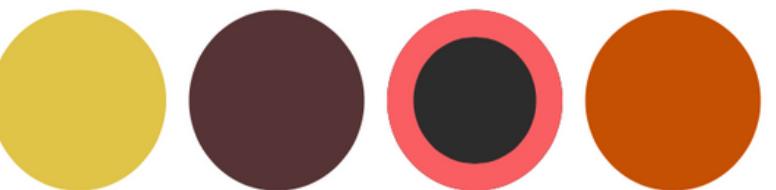
Curly



Child's Skin tone



Child's Hair colour





Educational Content



Storybook Content

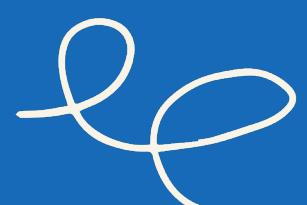
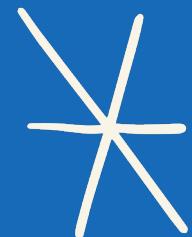
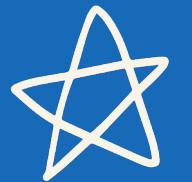




**What are
common
elements among
successful
children's books?**



“ CHILDREN'S BOOKS DATASETS



**Project
Gutenberg**

**Goodreads
(Kaggle)**

**Highly-rated
books (Kaggle)**



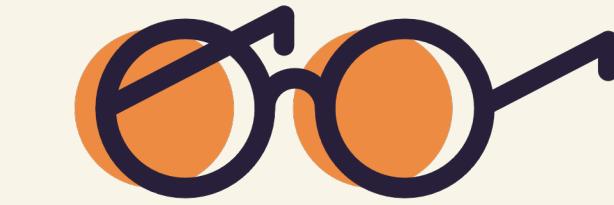
ATTRIBUTES OF DATASET



Book Name

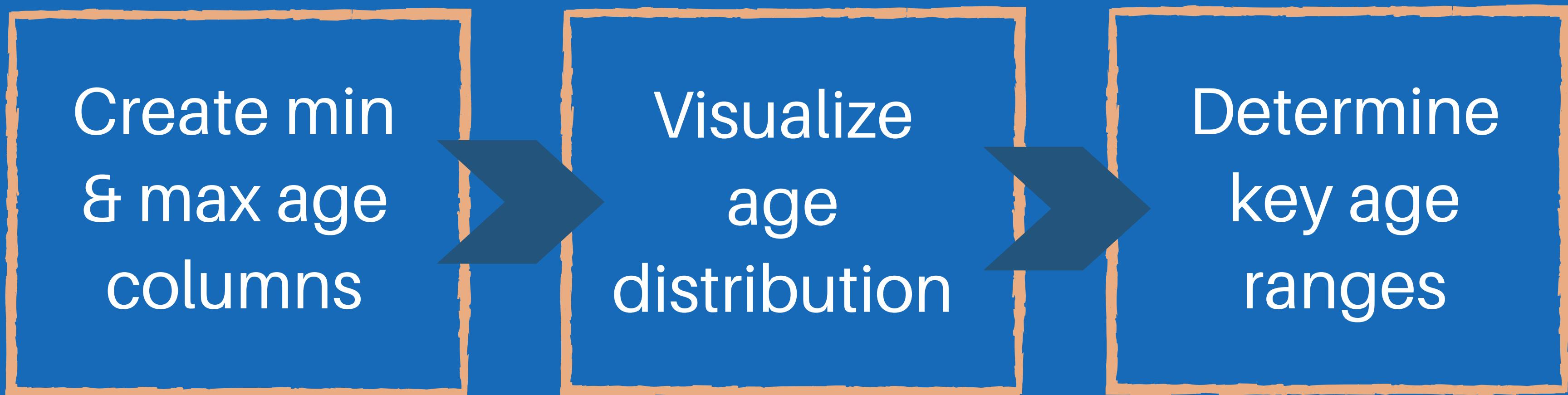


Book Description



Age Range

PROPOSED METHODOLOGY

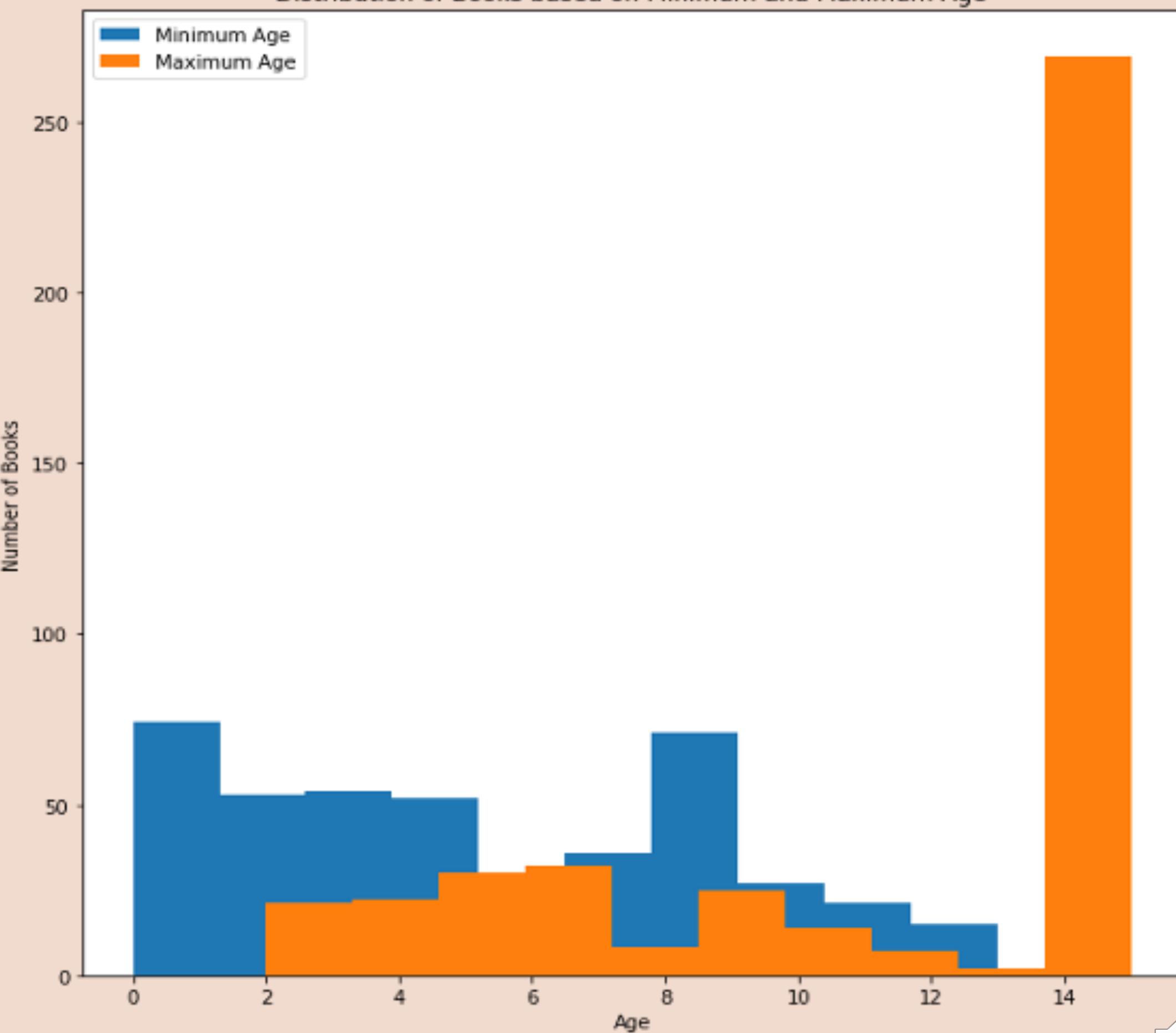


names cats desc max_age

min_age

0	55	55	55	55
1	19	19	19	19
2	53	53	53	53
3	54	54	54	54
4	30	30	30	30
5	22	22	22	22
6	27	27	27	27
7	36	36	36	36
8	31	31	31	31
9	40	40	40	40
10	27	27	27	27
11	21	21	21	21
12	12	12	12	12
13	3	3	3	3

Distribution of Books based on Minimum and Maximum Age

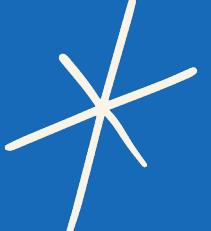
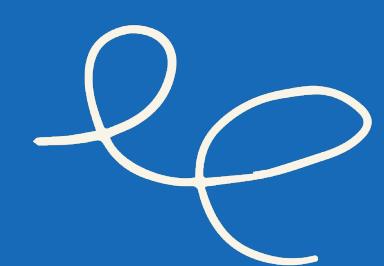
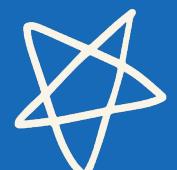


“ TEXT_TO_DICT() FUNCTION ”

Tokenize
and remove
stopwords

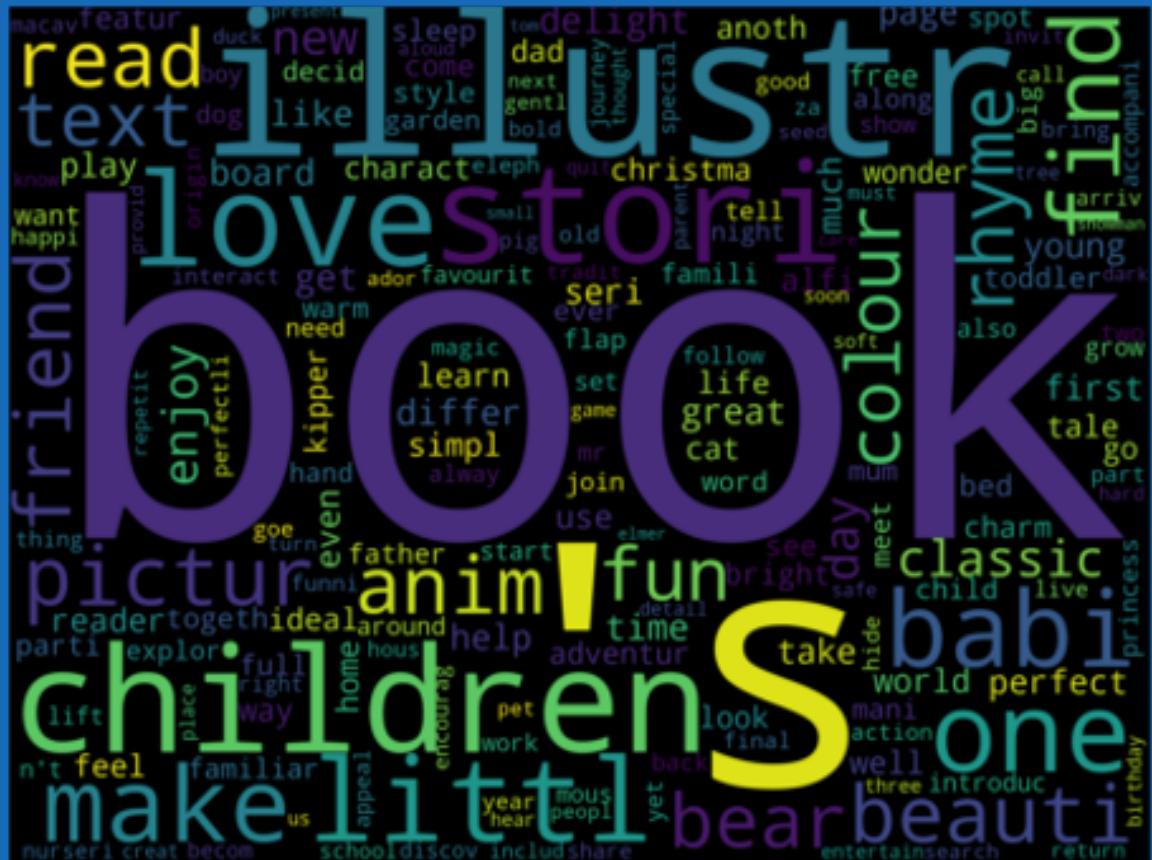
Stem the
words with
PortStemmer

Bag-of-
words for
dictionary

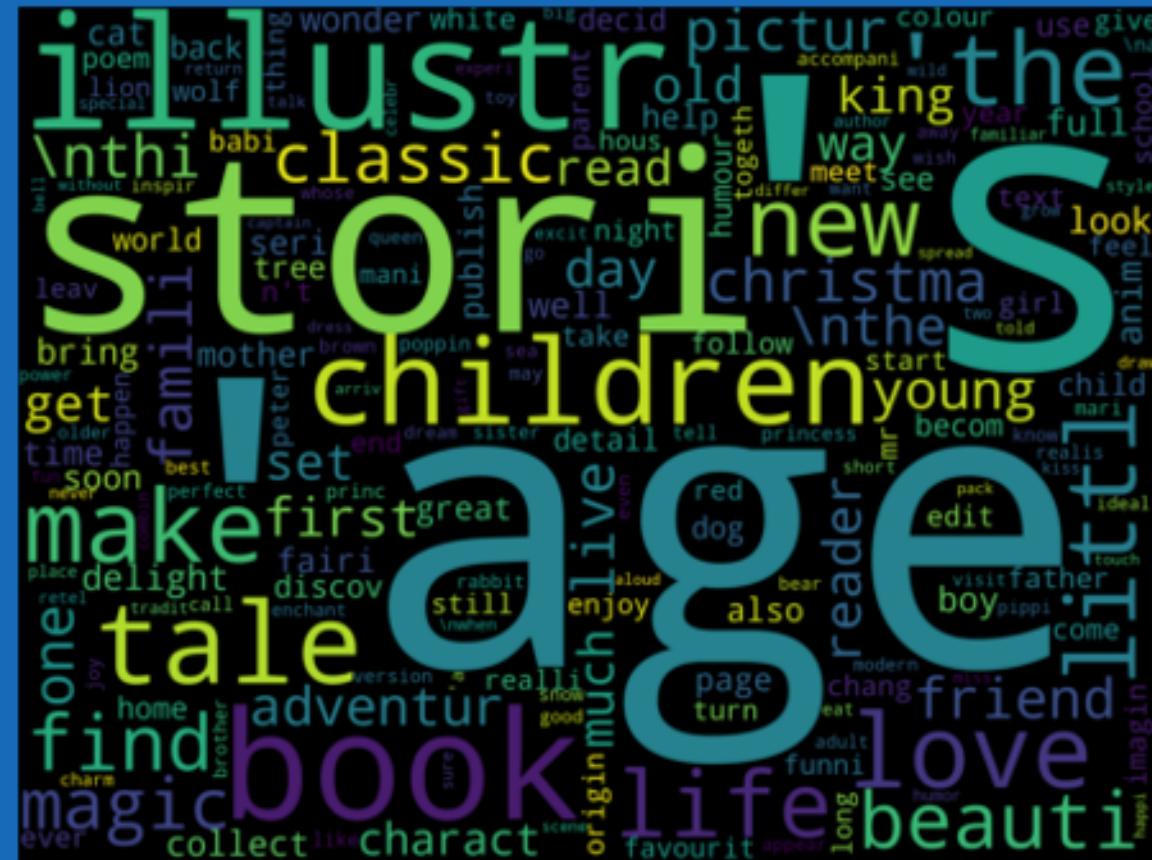


GENERATED WORD CLOUDS

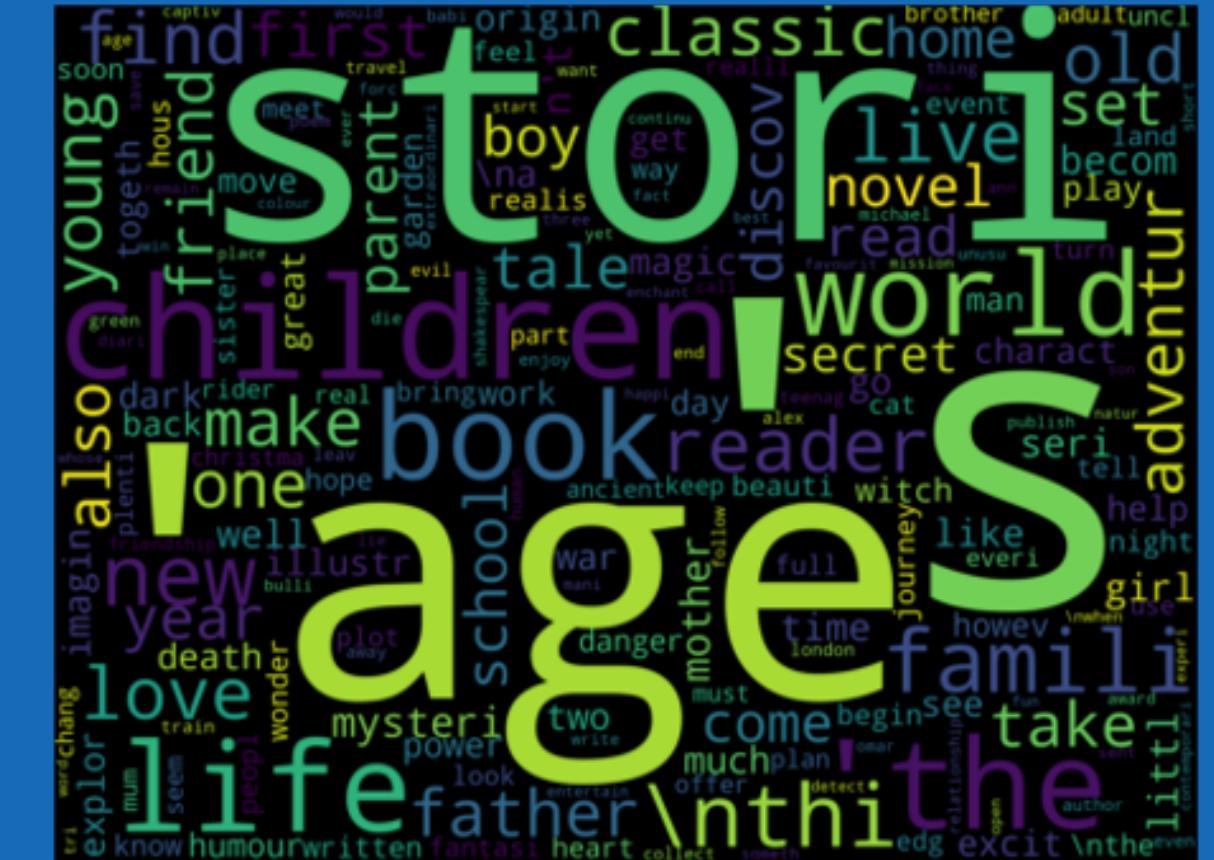
Minimum age of 0-3 years old

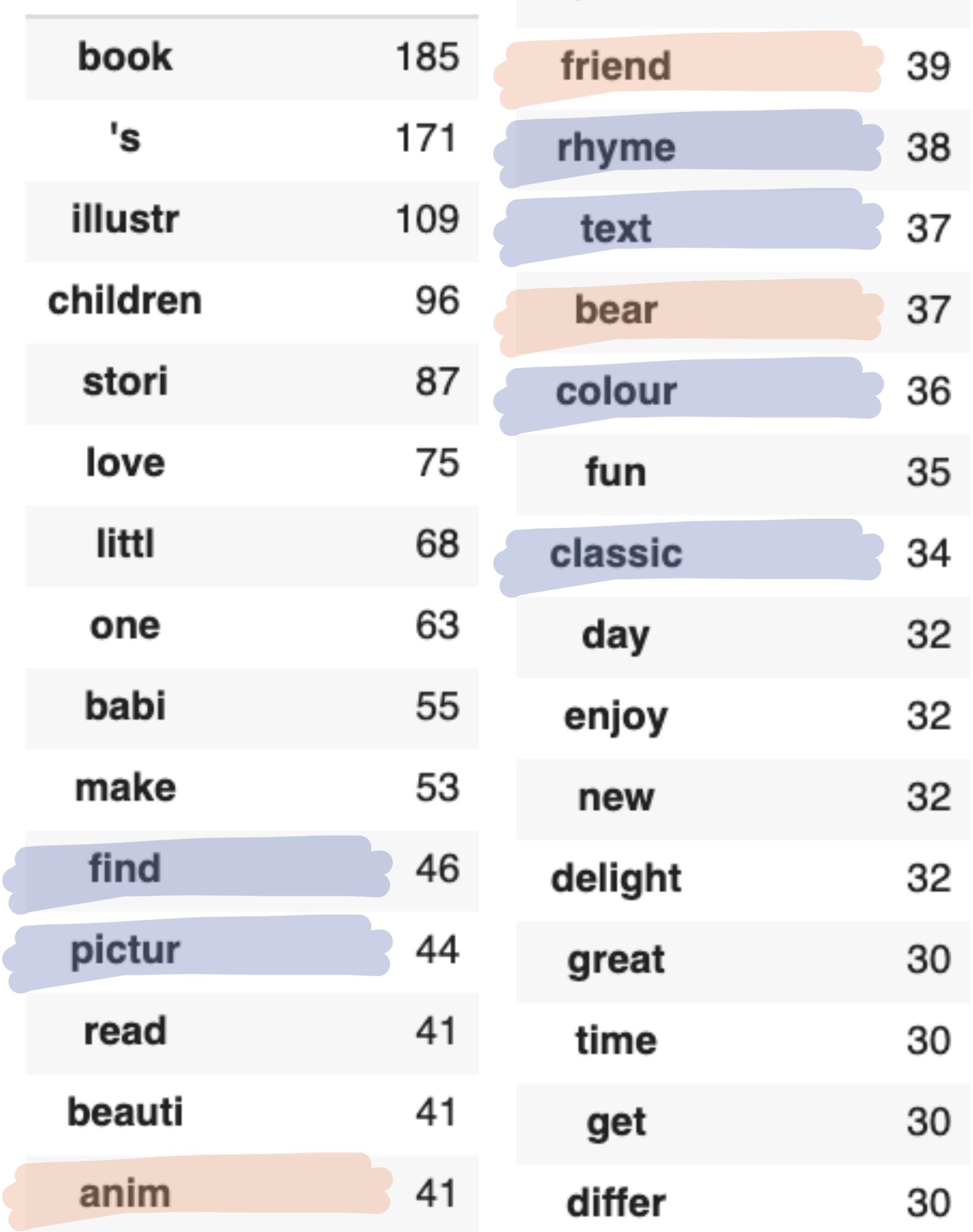


Minimum age of 4-7 years old

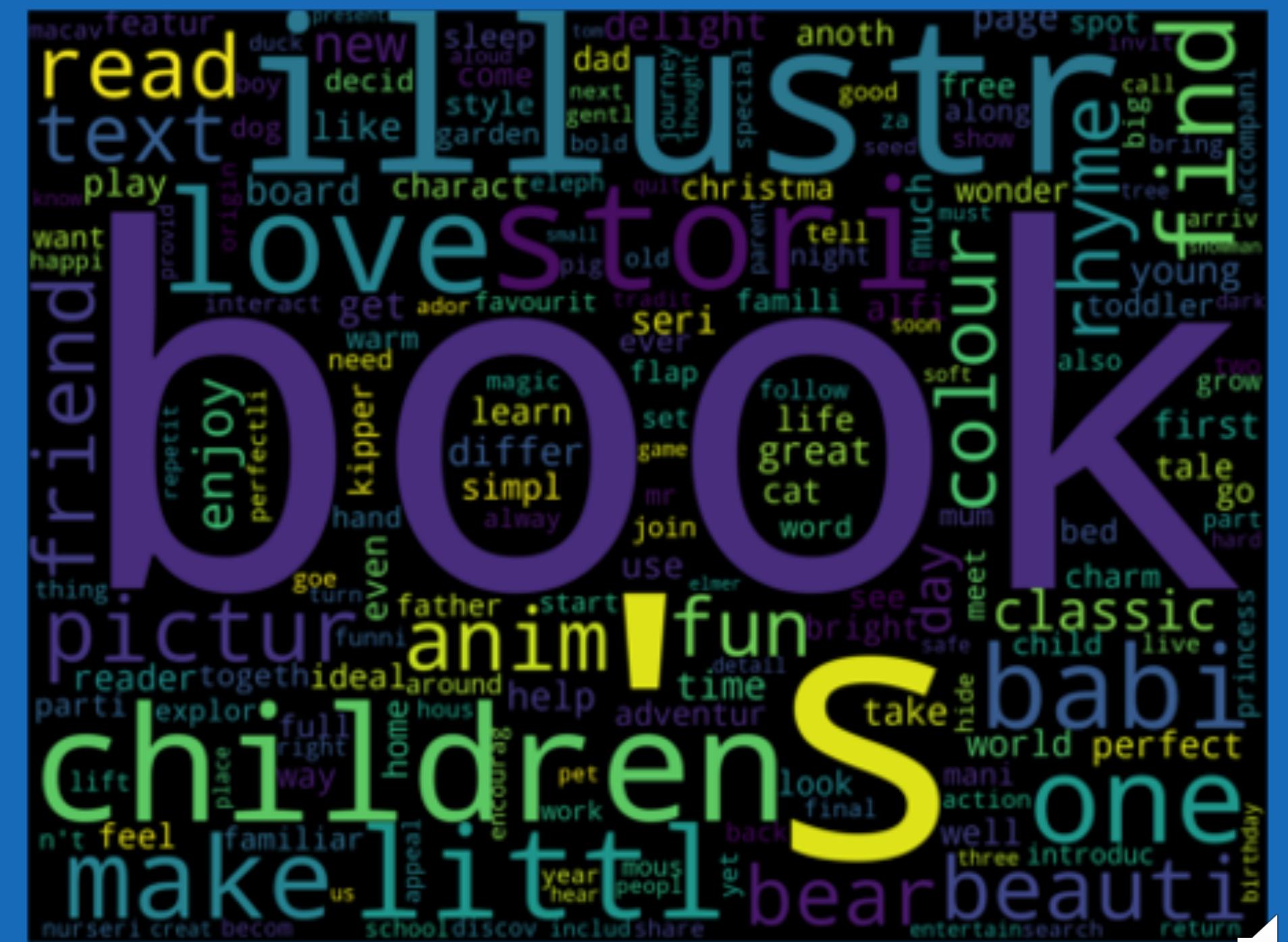


**Minimum age of
8+ years old**





RESULTS FOR THE 0-3 YEAR AGE RANGE



- Identified common book elements for different age groups
- limited analysis due to lack of a quality dataset
- Used tokens, stemming, bag-of-words, and word clouds
- Model identified age-appropriate story structures





Thank you.