Tell me a story...
A Children's Book Recommender

Elyse Renouf, MBA
Data Scientist

When Once Upon A Time…
There was a Need:

"We've read this book so much, we have it memorized!!

Help me find new children's books
similar to the stories my kids already love!"

A Hero Appeared: Tell Me A Story (TMAS)

- A content-based child book recommender system

- Recommends books based on similarities in the book descriptions
  using natural language processing (NLP) and co-similarity scores

In the beginning, there was Raw Data:

- **Source**
  - Kaggle dataset scraped from GoodReads.com
  - No category or age column available

- **Format(s)**:
  - **Then:** 11 .csv files with 100k books each and 13 columns

	name	authors	isbn	rating	publishyear	publishmonth	publishday	publisher	countsofreview	pagesnumber	description	is_preschooler
0	russell's secret	johanna hurwitz	0688175740	3	2001	6	11	harpercollins	8	32	have you ever heard the words "'you can s...	1
1	oye, hormiguita	phillip hoose	1582460892	4	2004	1	6	tricycle press	2	32	the spanish translation of bestseller hey, lit...	1

DATA Contents

- **Book Name**
- Author(s)
- **ISBN:** unique book identifier #
- **Rating:** Avg book rating (1-5)
- **Publish Year / Month / Day** (dates)
- **Publisher**

- Counts of Reviews
- **Pages Number**
- **Description:** a short synopsis
- **Is_Preschooler:**
  - 1 = yes (ages 3+ by >30 pages)
  - 0 = no (ages 0-2 by <=30 pages)

Ages and Pages

Publisher General Guidelines:

- Board Books (ages 0-2):
  - 10 or less pages

- Picture Books (2-3):
  - Pages in multiples of 8
  - From 16-32 pages in length
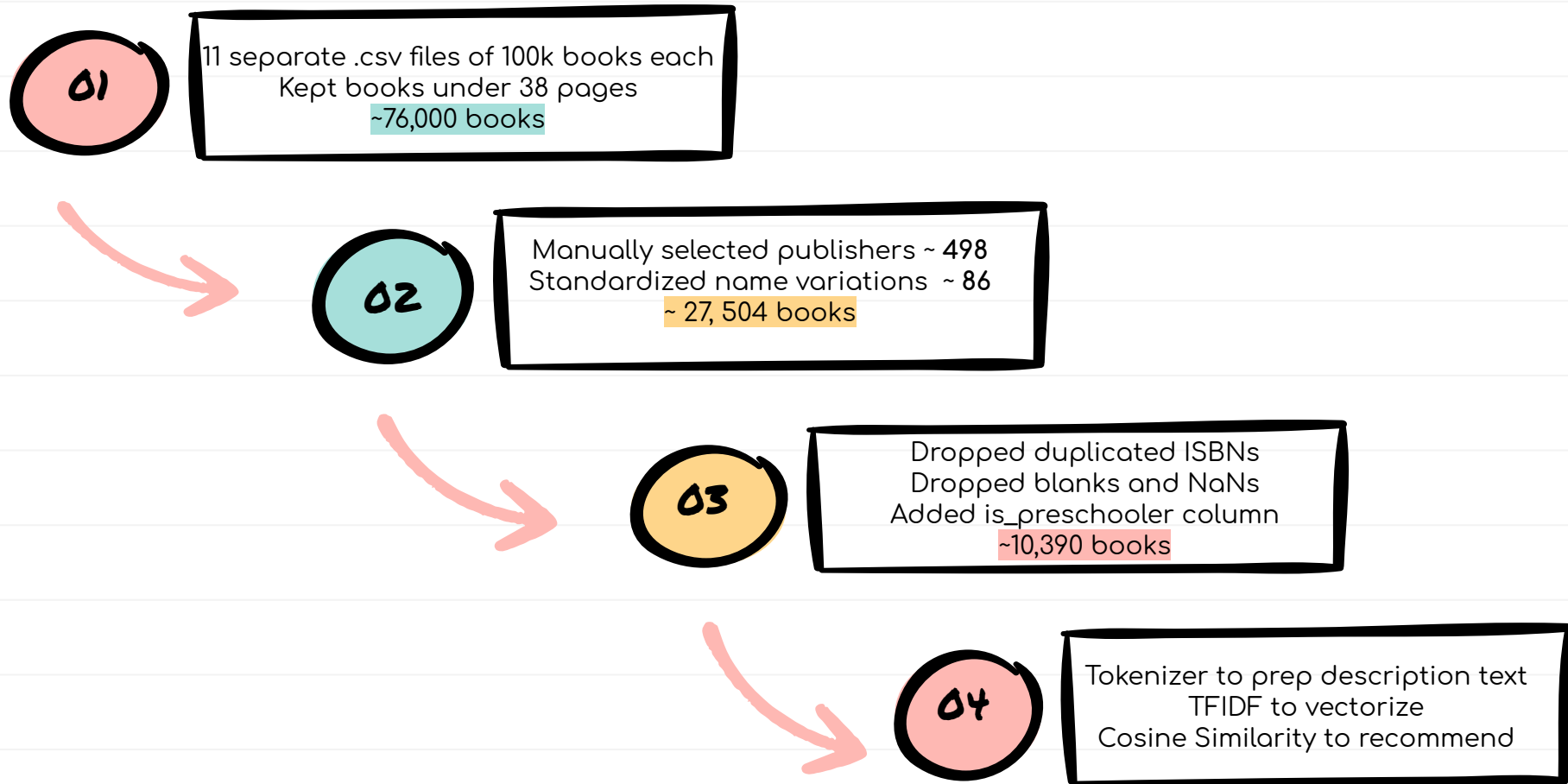
- Early Readers (4-6):
  - 32-48 pages

My decision:

- Board & shorter picture books (ages 0-2):
  - 30 or less pages

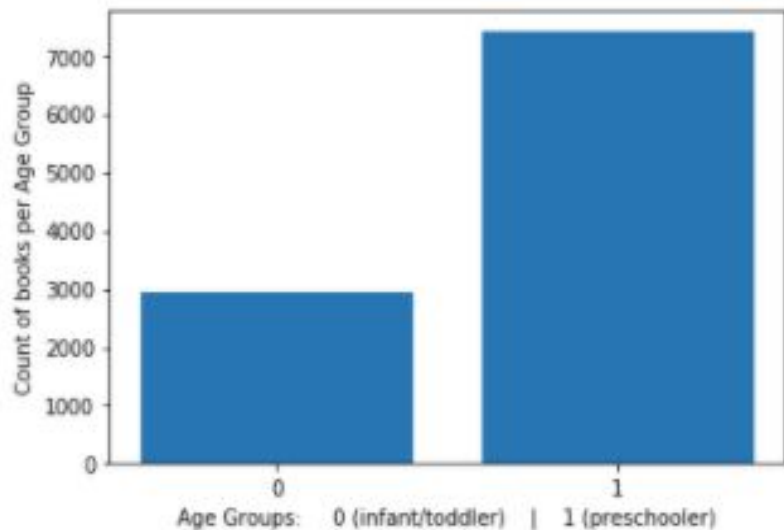- Longer picture books & early readers (3+)
  - 30-38 pages

NOTE:
Books over 38 pages were dropped to reduce the higher probability of adult material

Getting To Happily Ever After

01
11 separate .csv files of 100k books each
Kept books under 38 pages
~76,000 books

02
Manually selected publishers ~ **498**
Standardized name variations ~ **86**
~ 27, 504 books

03
Dropped duplicated ISBNs
Dropped blanks and NaNs
Added is_preschooler column
~10,390 books

04
Tokenizer to prep description text
TFIDF to vectorize
Cosine Similarity to recommend
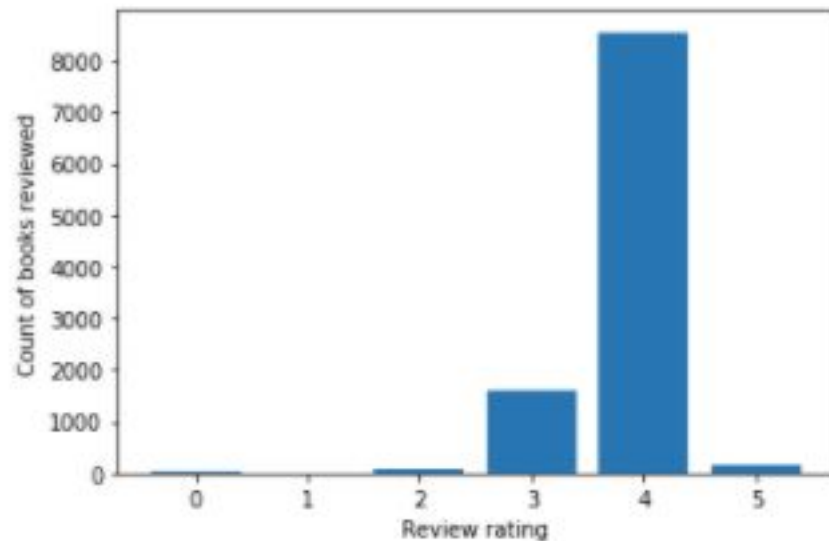
THE COLES NOTES ON THIS DATASET:

~ 30 % infant/toddler books (0-2) = 2961

~ 70% preschooler books (3+) = 7429

NOTE: +60% of books had exactly 32 pages

+85% of books in this dataset rated at a 4

~ rating not a major differentiating factor

What is a 'Good' similarity score?

```python
#testing out similarities between two very different kids books
book_1 = get_book_by_name('the velveteen rabbit', tfidf_scores, keys)
book_2 = get_book_by_name('a big city abc', tfidf_scores, keys)

print("Similarity:", cosine_similarity(book_1, book_2))
```

Similarity: [[0.03897897]]

```python
#testing out similarities between two very similar kids books
book_1 = get_book_by_name('a charlie brown christmas', tfidf_scores, keys)
book_2 = get_book_by_name('a charlie brown valentine', tfidf_scores, keys)

print("Similarity:", cosine_similarity(book_1, book_2))
```

Similarity: [[0.23813794]]

TMAS Version one:

Tell Me a Story...

📖A Content-Based Children's Book Recommender 📖

Select one of your child's favourite books from the drop down list below and TMAS will recommend some similar titles for you to try out.

| animals don't wear pajamas: a book about sleeping | ▾ |

📗That's a good one! 📗

Show me Similar Books

Titles Recommended for you:

	name	similarity
3420	cool bopper's choppers	0.2522
4913	the mole sisters and p…	0.2440
4498	the jazzy alphabet	0.2267
7275	henry's baby	0.2216
4736	mole and the baby bird	0.2137
5927	bravemole	0.2006
1326	baba yaga and the wise…	0.1816
1402	four little old men: a…	0.1780

Want to see the the entire dataframe behind TMAS?

☐ Give me the whole library!

```
similar_books = content_recommender("the bone keeper", True, tfidf_scores)
similar_books.head(10)
```

	name	similarity
3037	bones and the dog gone mystery (bones mysterie...	0.316017
3004	bones and the cupcake mystery (bones mysteries...	0.251207
6959	the bone talker	0.250863
5536	the bone man: a native american modoc tale	0.246213
6585	mister bones: dinosaur hunter	0.240057
904	bones and the birthday mystery (bones mysterie...	0.202886
5948	the teeny tiny woman: an old english ghost tale	0.199999
1670	benny	0.188995
7067	the teeny tiny woman	0.184541
2404	digging up dinosaurs	0.183855

TMAS Version Two:

- Filtered by is_preschooler age group
  - True = 1 (3+ years old) > 30pgs
  - False = 0 (0-2 years old) <30pgs

- Future:
  - Dataset with real ages / categories
  - Parts of speech categorization
  - Filter by topic or select keyword from word cloud

And They all Read Happily Ever After....

Let's connect:
elyse.renouf@gmail.com
https://www.linkedin.com/in/elyserenouf/

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

thank you

References / resources

Here are links to some of the outside resources that I've used throughout this project

1. Datasets gathered from:
   https://www.kaggle.com/bahramjannesarr/goodreads-book-datasets-10m
2. Children's Book page range averages and ages estimated based on info from:
   https://www.jennybowman.com/what-genre-is-my-childrens-book/