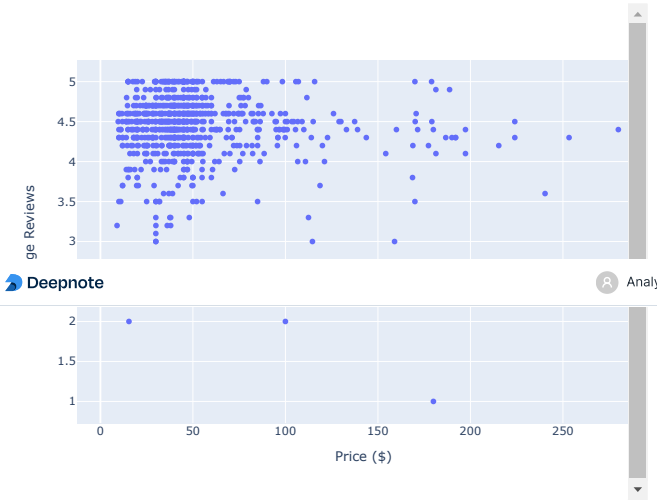
Reading Dataset

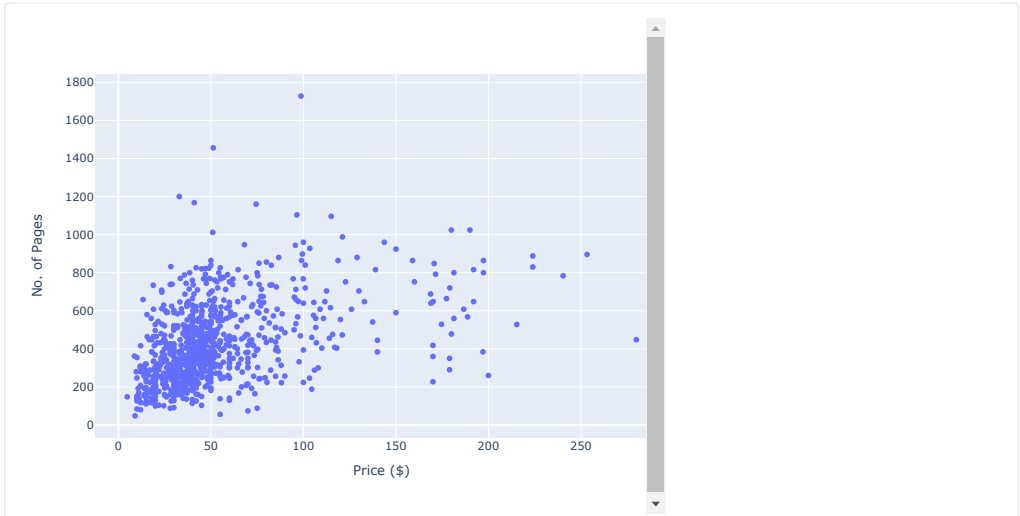
```
df = pd.read_csv("books.csv")
df.head()
```

EDA on Data Science Books

💵 Price vs. reviews



💵 Price vs. pages



📚 Best Python books

```
python_books = df[df['title'].str.contains('Python')]
best_python_books = python_books.nlargest(7, ['n_reviews', 'avg_reviews'])
best_python_books
```

	title object	author object	price float64	pages float64	avg_reviews float...	n_reviews int64	star5 float64	star4 float64
104	Python Crash Course 2nd...	[Eric Matthes]	21.49	544.0	4.7	7425	0.81	0.1
368	Python: - The Bible- 3...	[Maurice J. Thompson]	27.97	375.0	4.3	4033	0.64	0.1
819	Python: For Beginners: A Cras...	[Timothy C. Needham]	17.97	135.0	4.3	3034	0.66	0.1
827	Automate the Boring Stuff with...	[Al Sweigart]	26.49	592.0	4.7	2538	0.82	0.1
320	Python for Everybody:...	[Dr. Charles Russell Severance,Sue...	9.99	247.0	4.6	2467	0.76	0.1
218	Python for Data Analysis: Data...	[William McKinney]	53.99	547.0	4.6	1631	0.76	0.1
428	Deep Learning with Python	[Francois Chollet]	32.49	384.0	4.6	1349	0.76	0.1

📚 Best ML books

```
# Selecting books with title containing 'Machine Learning'
ml_books = df[df['title'].str.contains('Machine Learning')]

# From ML Books, selecting top 7 based on no of reviews and avg reviews(using no of reviews to make data less biased)
best_ml_books = ml_books.nlargest(7, ['n_reviews', 'avg_reviews'])
best_ml_books
```

	title object	author object	price float64	pages float64	avg_reviews float...	n_reviews int64	star5 float64	star4 float64
400	Deep Learning (Adaptive...	nan	54.25	800.0	4.3	1862	0.73	
200	The Hundred-Page Machine Learning...	[Andriy Burkov]	31.99	160.0	4.6	816	0.81	

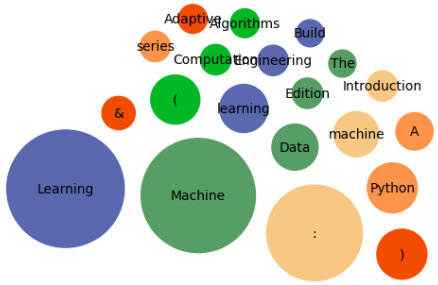
571	Pattern Recognition and...	[Christopher M. Bishop]	76.1	738.0	4.6	663	0.76
215	Mathematics for Machine Learning	nan	46.54	398.0	4.7	580	0.8
559	Introduction to Machine Learning...	nan	45.0	398.0	4.5	565	0.76
567	Advances in Financial Machin...	[Marcos Lopez de Prado]	40.49	400.0	4.5	514	0.76
608	Reinforcement Learning second...	nan	66.44	552.0	4.6	460	0.82

Clustering book titles

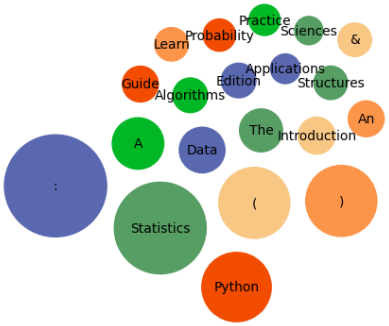
Using K-Means to form clusters

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

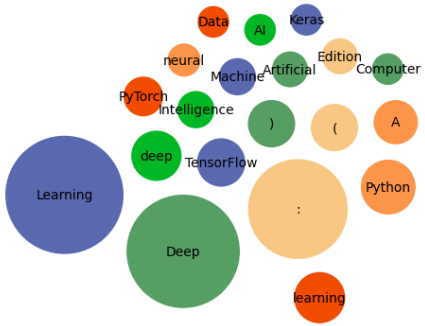
Frequently occuring words in cluster 0



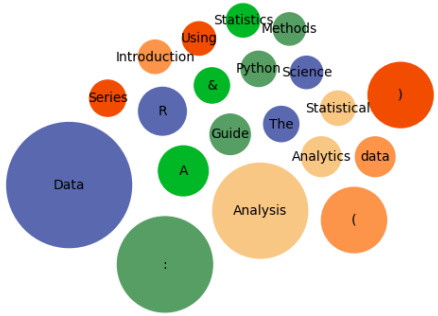
Frequently occuring words in cluster 1



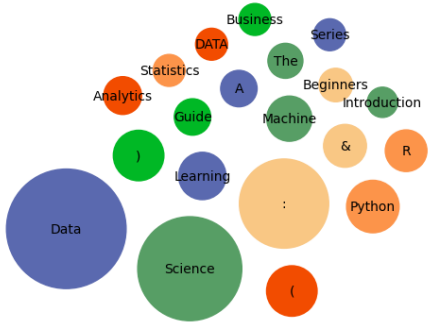
Frequently occuring words in cluster 2



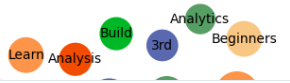
Frequently occuring words in cluster 3



Frequently occuring words in cluster 4



Frequently occuring words in cluster 5



Scrape Amazon Book Reviews

Two types of summarization of text:

1. Abstractive - Paraphrases the given text. Uses AI or machine learning.
2. Extractive - Takes out important sentences from paragraph as they are. Uses bert model. There may be other models available.

Summary of 10 reviews of a book concatenated

I was by no means “a beginner” when I picked up this book. My complaint is against how the author chose to interject his explanation of the code. My feeling is that he's taken away part of the book for those like me who don’t need (and can’t stand) highly-commented code. It just gave me a headache trying to parse code from comments. When it’s all “black and white,” it just blurs together. My feeling is that Ruby was the most represented of the three. All in all worth a read for the visualizations of the structures and algorithms. This book is the opposite, easy to read, by that I mean, concepts are explained in simple terms with walk through examples, along with visual examples, makes it easy to grasp, thus I now understand it, recursion now demystified, this book speaks my language, highly recommend this book to anyone, whether you are starting out or looking for a primer. Even though most code is written in python, pearl, and JavaScript the code is readable to the point that you can comprehend the examples and can replicate them using your preferred programming language. By the way, I think step by step details were so good.

Original 10 Reviews concatenated

I was by no means “a beginner” when I picked up this book. But it's also been 20 years since I learned anything DS&A. I picked this book over others mostly on a whim, but being geared towards beginners, I was at least assured that the subject material would fly over my head.So don’t get me wrong, this is a fantastic, educational book. My complaint is against how the author chose to interject his explanation of the code. My feeling is that he’s taken away part of the book for those like me who don’t need (and can’t stand) highly-commented code. Heck, I’m okay if there were just a few, well-worded, well-placed comments that gave reasons for unclear aspects of the code. But at about a quarter through, I stopped trying to read too much of the code. It just gave me a headache trying to parse code from comments. When it’s all “black and white,” it just blurs together.Then, once you’re done trying to find the code amongst the comments, realize it's maybe 6 lines of code, he says, “This code is not trivial, so let's break it down.” Dude. You already did that. In the middle of the code. Using comments. And also. My man. Six lines is very nearly the *exact* meaning of trivial. Non-trivial is when you need to create classes and use *actual* programming techniques of architecture and design. Not writing 6 lines of code that you didn’t even write tests for because you’re not the first person to write it that way, in that language.So if you can get past the frustration involved with that aspect, the actual DS&A part is amazingly good and easy to understand. I’ve recommended the book on more than one occasion; I only mention the code comments when I think it’ll benefit the person to whom I’m recommending the book.I would certainly enjoy a second, less beginner-focused version that maybe expands on the concepts, or even repeats some with more succinct code/comments. As far as introducing the data structures this book takes a great visual approach. Algorithms are often easiest to understand in animated form but the step by step of this book is quite strong.The major weakness is by jumping between JS, Python, and Ruby you never fully develop any of them. Although the programmatic style is mostly translatable between these three languages it’s still typically better to do one thing fully rather than a few things partially. My feeling is that Ruby was the most represented of the three.All in all worth a read for the visualizations of the structures and algorithms. Got this book on a recommendation to help with the class that I’ve been struggling in have not finished it yet, but so far it’s helping me understand data structures much better I have read 7-9 books on this subject, majority are difficult to understand thus hard for the concepts to stick inside my brain.This book is the opposite, easy to read, by that I mean, concepts are explained in simple terms with walk through examples, along with visual examples, makes it easy to grasp, thus I now understand it, recursion now demystified, this book speaks my language, highly recommend this book to anyone, whether you are starting out or looking for a primer. I can’t say enough good things about this book. The author writes in everyday language so these concepts are so much more easily understood than any other resources I've found. This book is perfect for someone like me who does not have a strong mathematics or computer science background. I started a job as a dev and am going back to learn some CS fundamentals, and this is absolutely the best place you can start. If you're looking for a DSA book that is easy to understand and worth your time, THIS IS IT! If you're coming from a non-CS background, this is the book you want. I rarely think a book deserves 5 stars, but the author did an amazing job.In my journey to becoming a self taught software engineer, I found this book only after 3 years of working professionally, and I STILL gained a alot if value out of it, namely filling some gaps that may have still been present in my makeshift education.That's said, even if you are brand new, this is still the book you want, and I wish I found it years ago! I only got this book because others suggested it. Even though most code is written in python, pearl, and JavaScript the code is readable to the point that you can comprehend the examples and can replicate them using your preferred programming language. For example, I use mostly Kotlin and Java and as long as you understand basic programming syntax, then it’s

Source of Dataset: https://www.kaggle.com/datasets/die9origephit/amazon-data-science-books