

ANALYSING MEDIUM AUTHORS

PROJECT LAB

Prepared by:
Sanyam Jain (214104020)

Guided By:
Dr. Amit Bhagat
Dr. Dheeraj Dixit



◆ **WHAT IS MEDIUM?**

Medium is a popular online platform that allows authors to publish articles on a diverse range of topics.

For data analysts, this is a rich source to explore the diverse data available on it to gain valuable insights.

**60-100M
Monthly
Active Users**

**>1.2M
New Stories
Every Month**

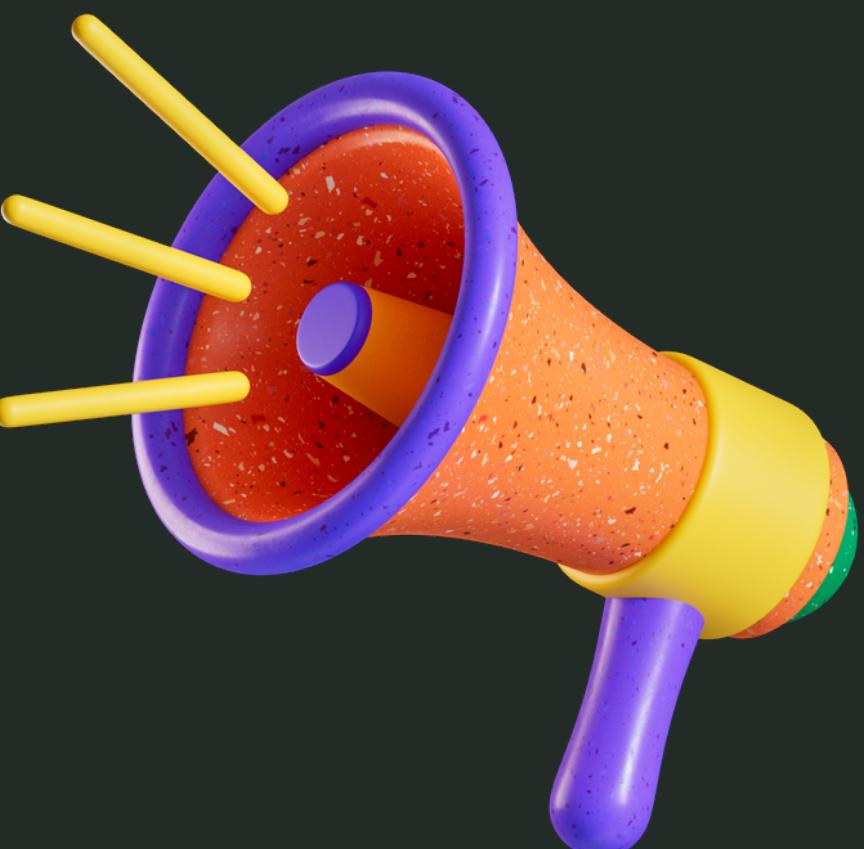




◆ INTRODUCTION

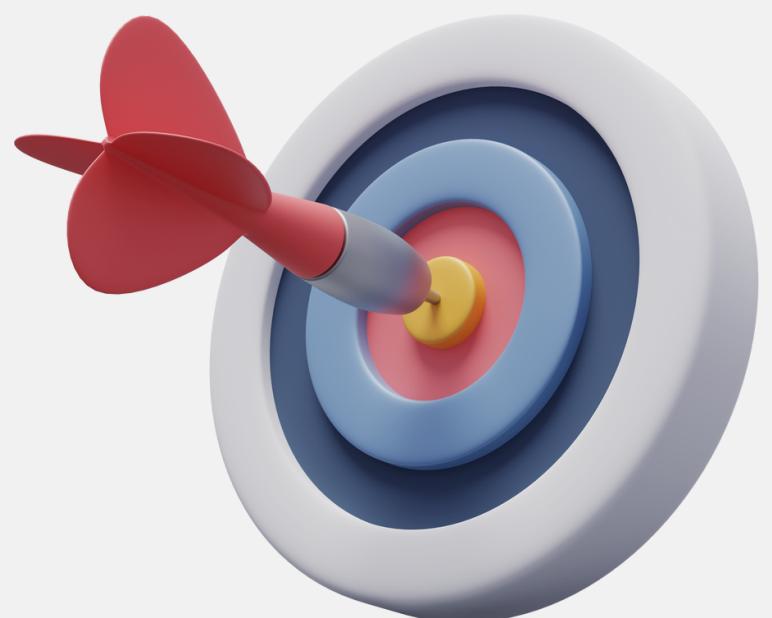
Conducting data analysis on Medium data can provide valuable insights into the various writer's ways of writing, their article content, and platform dynamics, which can inform decision-making and strategy development in various domains.

In this project, we tried to uncover interesting insights about Medium authors that can be helpful for content strategies and marketing trends. And that wasn't all! We also demonstrated how we employed sentiment analysis and summarization techniques to gain even deeper insights into the content of Medium articles.



AIM OF THIS PROJECT

To gain a deeper understanding of the characteristics of authors who are most active on the platform, including their demographic profiles, writing styles, topics of interest and major publication they are active into. This analysis aims to provide valuable insights that can inform content creation strategies, platform policies, and marketing approaches for authors, publishers, and Medium as a whole.

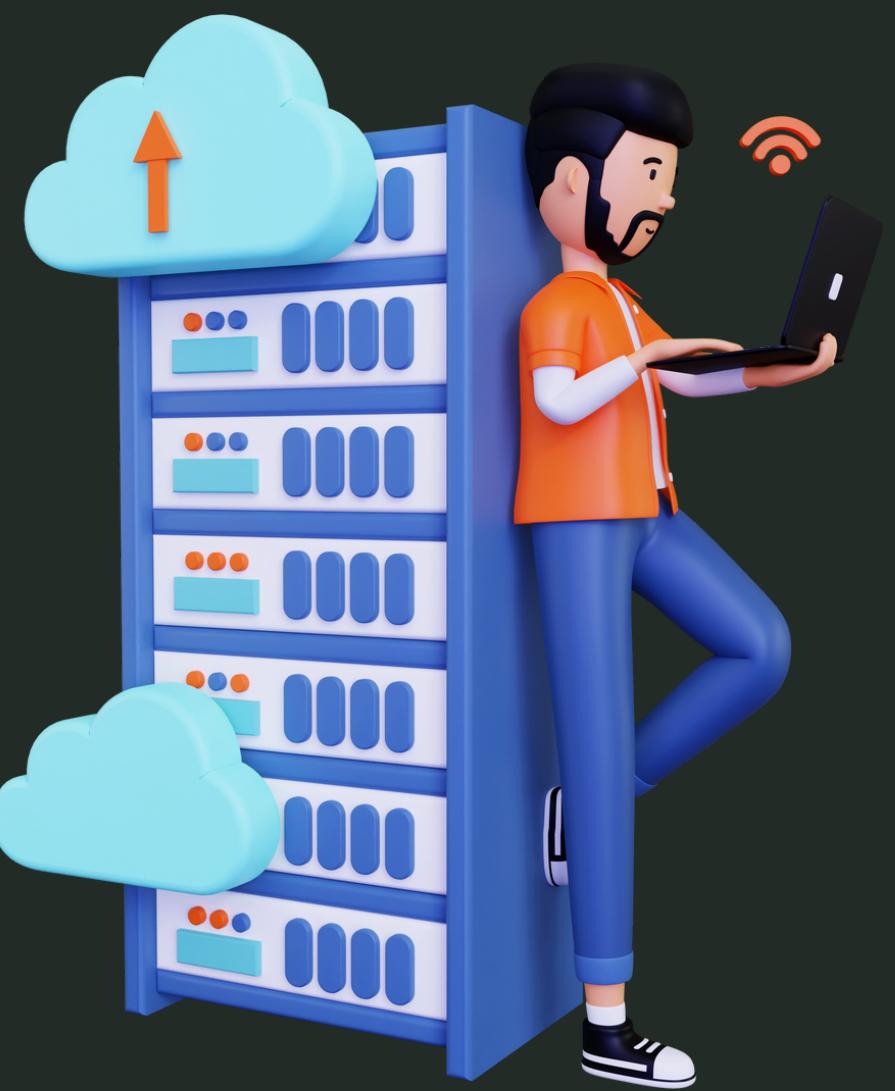


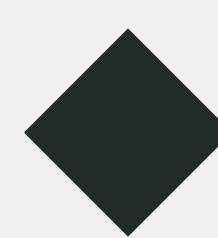
1. Finding top authors writing for a specific input topic. So that we can do our further analysis on the right sample data.
2. Finding user data including : bio, followers, following, social handle and more.
3. Finding each article's data published by the author.
4. Finding topics the author writes about.
5. Finding the top publication the author publishes in.
6. Finding popularity of articles, based on clap count and voter count.
And how it is related to corresponding topics and publications.
7. Operations on the content of article:
 - Doing sentimental analysis of the content.
 - Performing Summarization.



◆ DATA COLLECTION

The first step in the process was to collect data from Medium using the Medium API. Their python package allowed us to access a wide range of information about articles, including author profiles, top contributing writers, article titles, publication dates, and more. The retrieved data was saved in a Pandas DataFrame for further analysis.





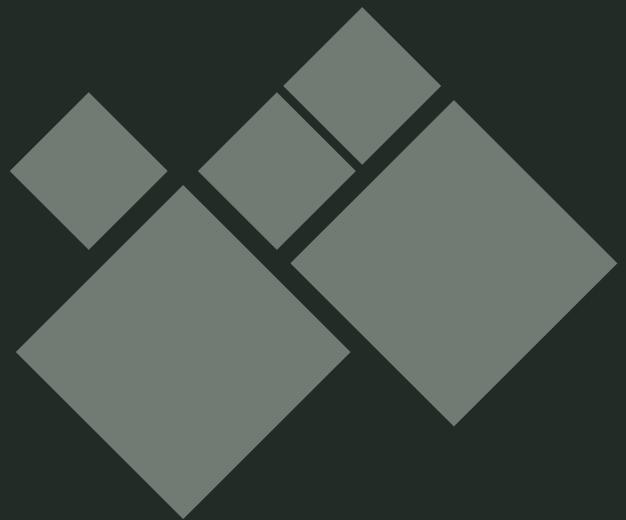
WHO TO ANALYZE? THE TOP WRITERS OF MEDIUM...

It was a hectic task to decide on whose profile to perform the analysis. Whose profile could serve as an excellent test subject for our project?

In the end, we selected the profile of '**frank-andrade**', who was the winner of the top writers list under the topic of '**data-science**'.

Top writers of "data-science" on Medium

The PyCoach --- [frank-andrade](#)
Giorgos Myrianthous --- [gmyrianthous](#)
Sanjay Priyadarshi --- [priyadarshisanjay](#)
Jim Clyde Monge --- [jimclydemonge](#)
Molly Ruby --- [molly.ruby](#)
Bex T. --- [ibexorigin](#)
B. Chen --- [bindichen](#)
Youssef Hosni --- [youssefraafat57](#)
Lars Nielsen --- [pythoslabs](#)
Thuwarakesh Murallie --- [thuwarakesh](#)



Username: frank-andrade
Fullname: The PyCoach
Bio: 8M+ Views on Medium || Early Bird Discount: Make money by writing about AI, programming, data science or tech ↗ <http://bit.ly/3zfbgiX>
Profile Image: https://miro.medium.com/1*veEX4-CiLz5jqUjwWfQo_Q.jpeg
Top Writer In: ['artificial-intelligence', 'technology', 'science', 'entrepreneurship', 'business']
User Written Articles: 249
Followers: 47887
Following: 7

From the number of articles written by it can be concluded that this person is quite active and has written a lot about this field. He is successful in terms of followers - 47887!



◆ FETCHING DATA OF EACH ARTICLE WRITTEN BY USER

These are the fields of our dataset

```
Index(['id', 'tags', 'claps', 'last_modified_at', 'published_at', 'url',  
       'image_url', 'is_series', 'lang', 'publication_id', 'word_count',  
       'is_locked', 'title', 'reading_time', 'responses_count', 'voters',  
       'topics', 'author', 'subtitle'],  
       dtype='object')
```

Let's do some cleaning!

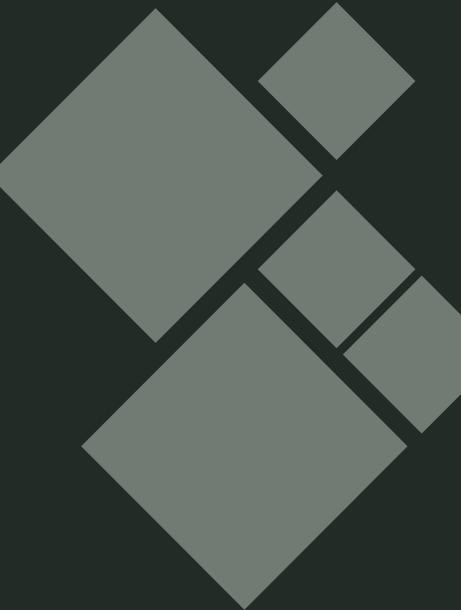


◆ DATA CLEANING

```
articles_df = articles_df.drop(['image_url',
                                'last_modified_at',
                                'url',
                                'author',
                                'is_locked',
                                'reading_time',
                                'is_series',
                                ], axis=1)
```

Removing
unnecessary
fields

```
Index(['id', 'tags', 'claps', 'published_at', 'lang', 'publication_id',
       'word_count', 'title', 'responses_count', 'voters', 'topics',
       'subtitle'],
      dtype='object')
```

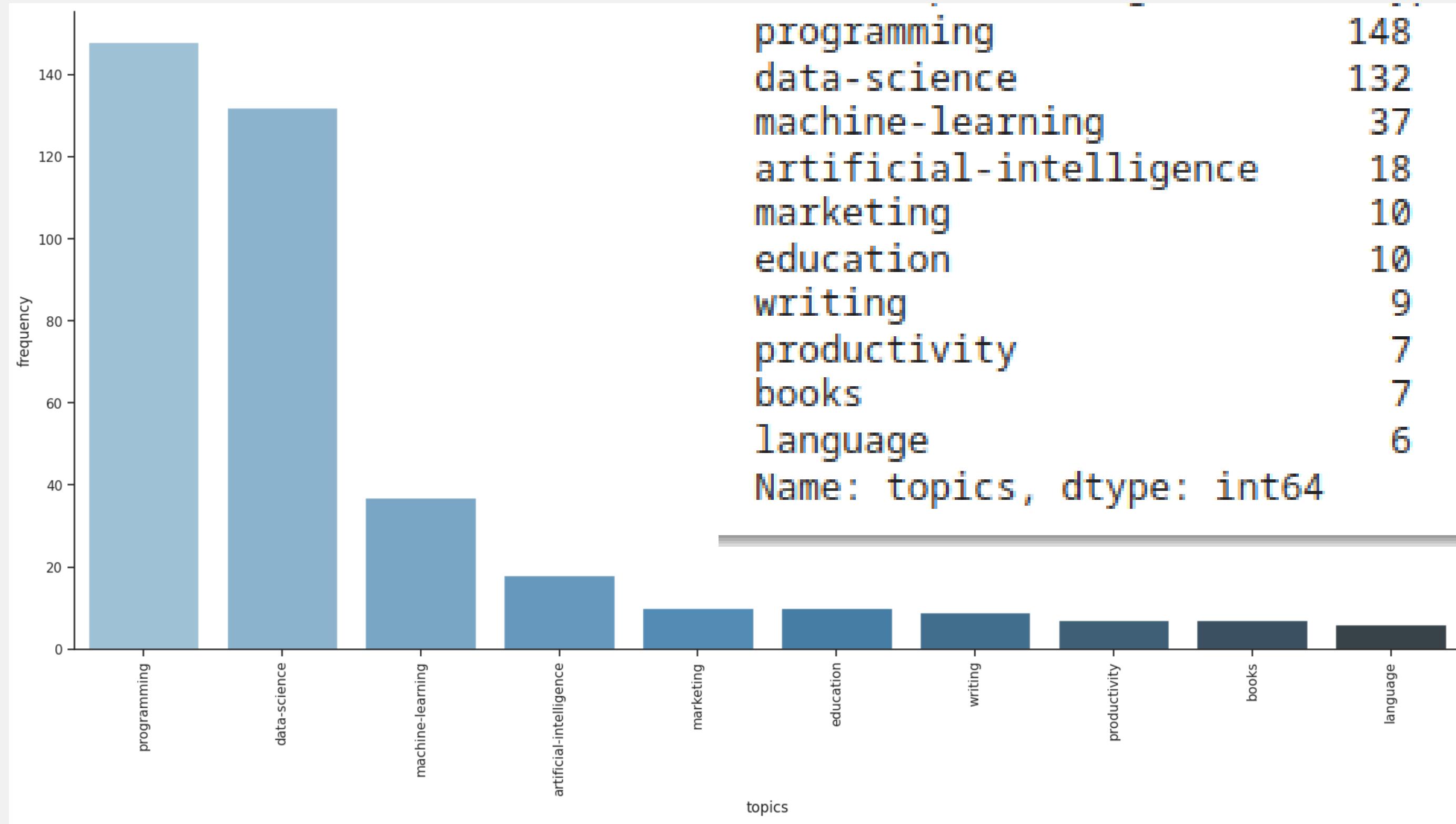


◆ DATA CLEANING

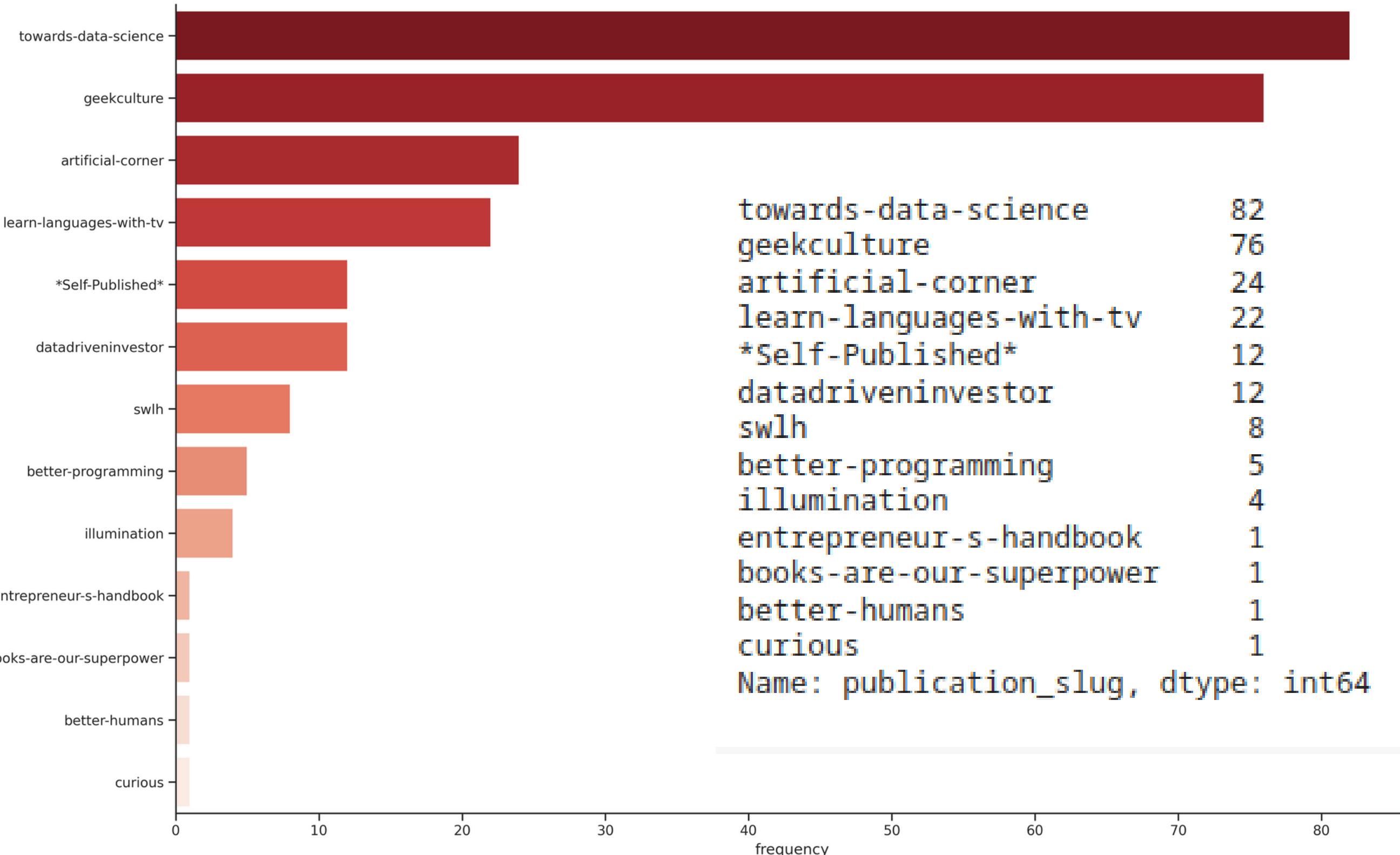
Here's how our final data looks like

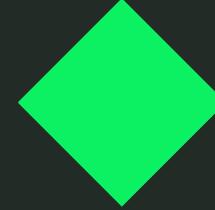
	id	tags	claps	published_at	lang	publication_id	word_count	title	responses_count	voters	topics	subtitle
0	8ac0ffb14af9	[artificial-intelligence, chatgpt, technology, ...]	114	2023-04-14 09:01:08	en	*Self-Published*	310	Share Your AI Knowledge With The World: Write ...	4	20	artificial-intelligence	Calling all AI and tech enthusiasts.
1	c876fbe7915e	[artificial-intelligence, technology, chatgpt, ...]	784	2023-04-12 15:39:52	en	76436a11a2b0	1272	You're Using Midjourney Wrong! Here's How to C...	12	101	design	Generate amazing images by learning how to cre...
1	c876fbe7915e	[artificial-intelligence, technology, chatgpt, ...]	784	2023-04-12 15:39:52	en	76436a11a2b0	1272	You're Using Midjourney Wrong! Here's How to C...	12	101	photography	Generate amazing images by learning how to cre...
2	bda045eed47f	[technology, chatgpt, python, data-science, ar...	742	2023-04-11 10:30:02	en	76436a11a2b0	868	The ChatGPT Skill That Pays Up to \$335,000 a Year	13	143	artificial-intelligence	AI is creating amazing new jobs.
2	bda045eed47f	[technology, chatgpt, python, data-science, ar...	742	2023-04-11 10:30:02	en	76436a11a2b0	868	The ChatGPT Skill That Pays Up to \$335,000 a Year	13	143	programming	AI is creating amazing new jobs.

◆ TOP TOPICS THE AUTHOR HAS MOST WROTE ABOUT (TOP 10) :



PUBLICATIONS THE AUTHOR HAS PUBLISHED MOSTLY

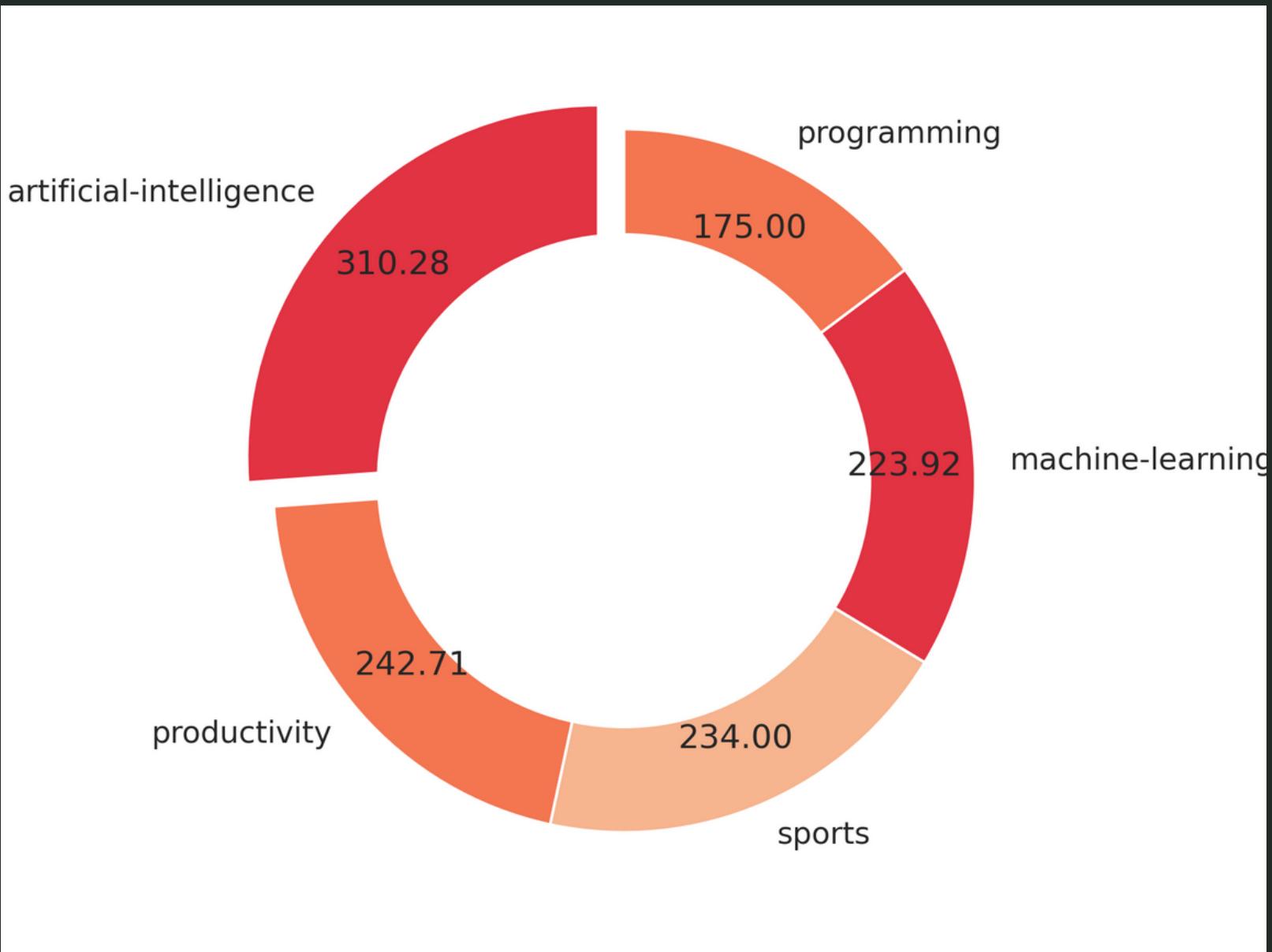




POPULARITY OF ARTICLES

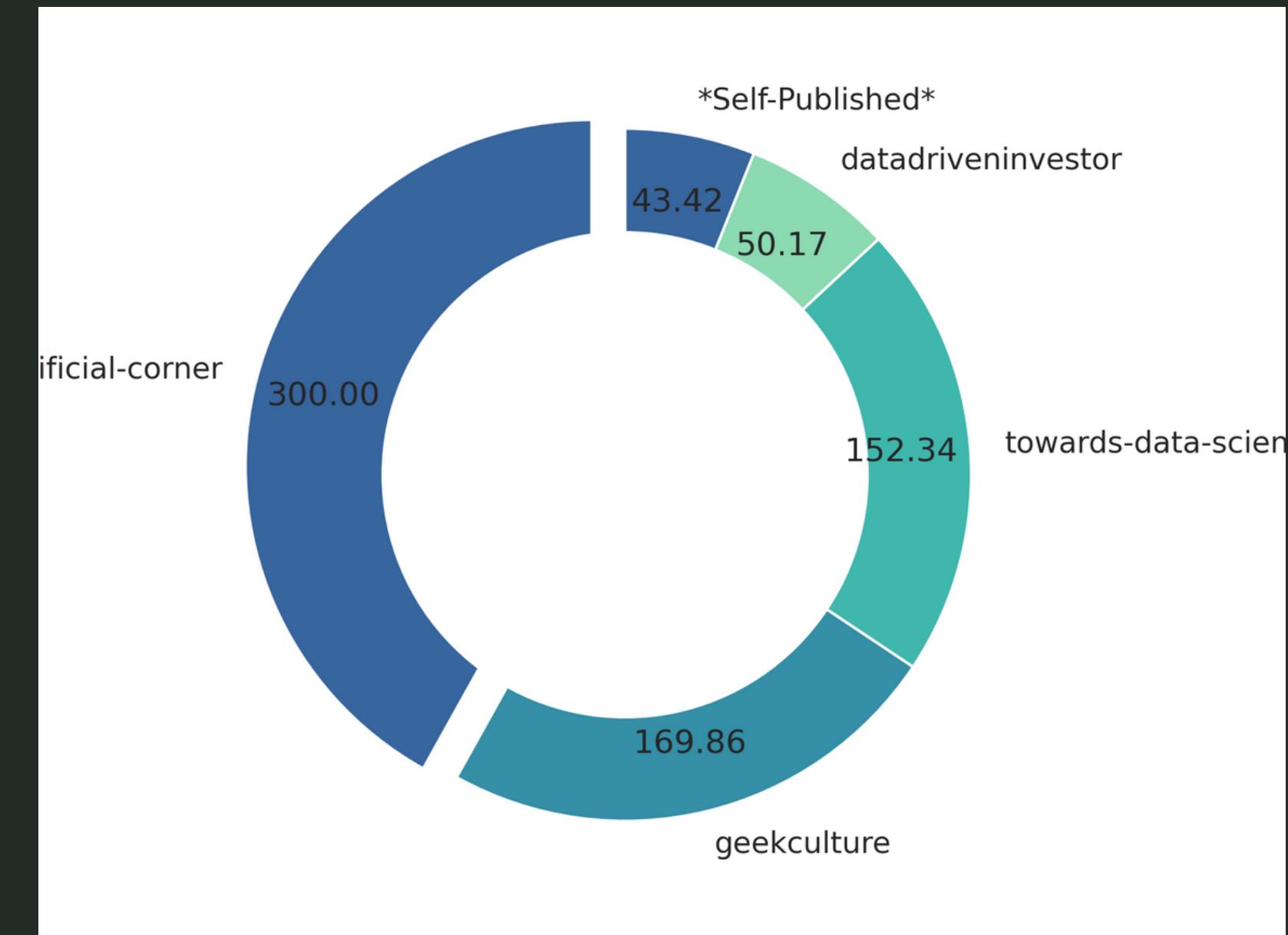
Based upon the clap count and voter count we get a rough estimation of how well an article has received attention from its readers. And we have shown how this is related to the topic of the article and the publication in which it was published.





Topics that have accumulated the most attention from readers

Publications that performed in favor of the author the most



Here the values show the average number of voters per article in that topic/ publication

◆ SENTIMENTAL ANALYSIS

The majority of the articles exhibit a highly positive sentiment, with scores approaching 1. This outcome was anticipated, as the subject matter pertains to learning data, which typically involves a positive and explanatory approach.

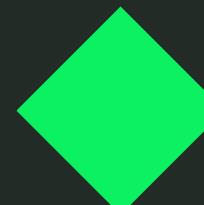
Following are the results of the sentimental analysis we got by using NLTK library:

	title	sentiment_score
0	Share Your AI Knowledge With The World: Write ...	0.9846
1	You're Using Midjourney Wrong! Here's How to C...	0.9990
2	The ChatGPT Skill That Pays Up to \$335,000 a Year	0.9985
3	The Chatbot Competition: A Hands-On Comparison...	0.9990
4	All AI Tools And ChatGPT Prompts in 1 Article	0.9988
...
244	Learn a Foreign Language with Friends	0.9997
245	Learn the language hacking method, jump on Hel...	0.9999
246	The Simpsons is the Best TV show to Increase Y...	0.9993
247	6 Months Without an iPhone—This Is How My Li...	0.9976
248	The Best Movies to Learn a Foreign Language	0.9997

◆ SUMMARIZATION

We utilized txt.ai's summarization model to effectively condense lengthy articles into concise summaries of less than 50 words, which were then stored for future use. This approach aimed to improve efficiency, provide a high-level overview, prioritize relevant information, enhance interpretability, and facilitate data reduction during the analysis process.





CONCLUSION

- We utilized the Medium API's Python package to import author's data and analyzed it using pandas dataframe, which proved to be effective and efficient.
- The retrieved data provided valuable insights into the author's performance and engagement metrics, allowing us to gain a deeper understanding of their content's reach and impact.
- The analysis revealed the author's popular topics, audience engagement patterns, and content performance over time, informing content strategy and identifying growth opportunities.
- Visualizations in appropriate charts and graphs aided in effectively communicating the findings to stakeholders, facilitating easier understanding of performance visually.

◆ RESOURCES

Code for this Project:

https://colab.research.google.com/drive/1YIArScpRodVWvWgOo_xtioxjRmvLtEte?usp=sharing

Medium API Documentation:

<https://docs.mediumapi.com/>

Medium API Python Package:

<https://pypi.org/project/medium-api/>



