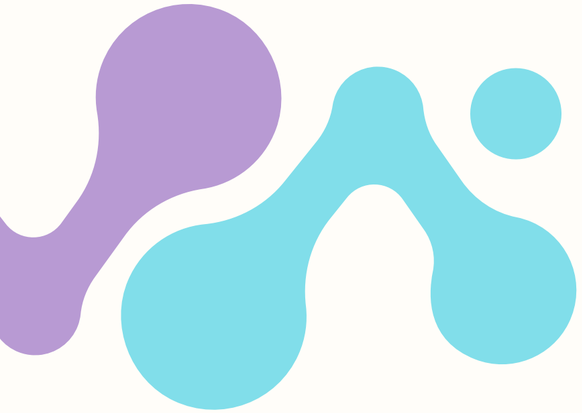
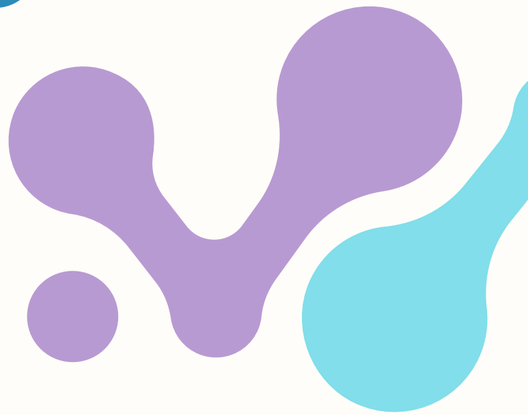


TEXT ANALYTICS WITH PYTHON ON MEDIUM ARTICLES

Nur Azmi Prasetyo



PYCON APAC 2024
Yogyakarta, Indonesia
October 25th-27th, 2024

HI, EVERYONE! MY NAME IS

Azmi

Business Intelligence at a Distributor Company

Thank you for having me. →

you can find it on
Poster list of Pycon
APAC 2024 Venue..

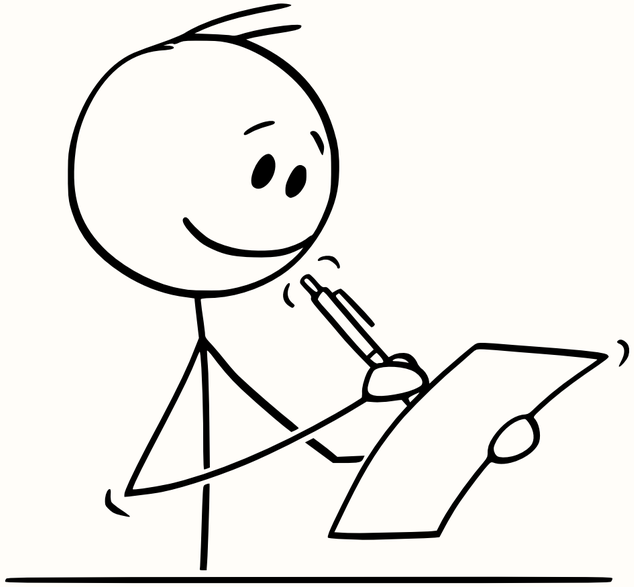
bitly

www.datawizards.id

find more at
www.exsight.id

PYCON APAC 2024
Yogyakarta, Indonesia
October 25th-27th, 2024

Story: The Beginning

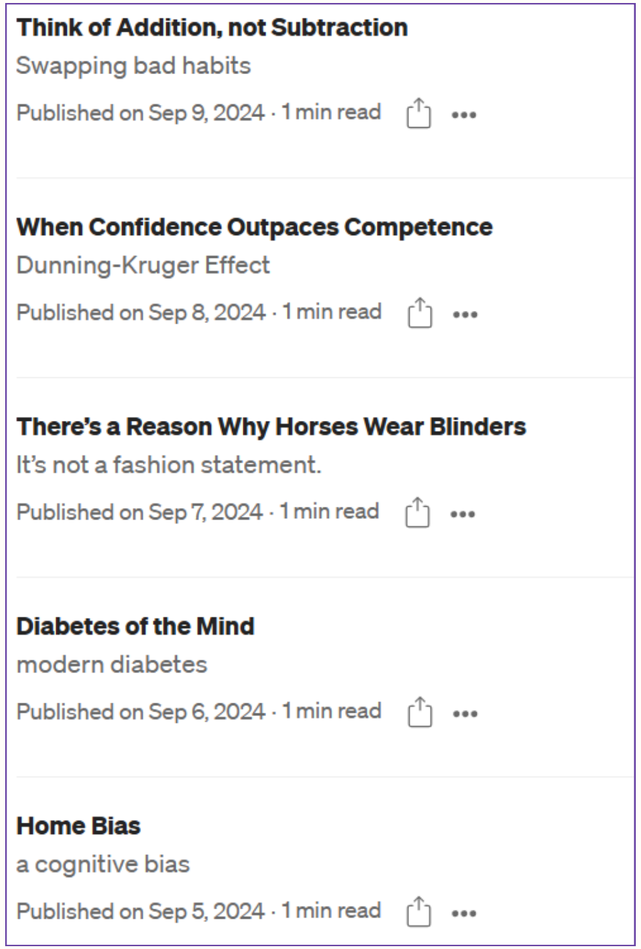


About six months ago, in **April 2024**, I made a commitment to write on Medium **every day**.

Each post is a reflection of **whatever topic** slides into my mind, from personal thoughts to professional insights.

PYCON APAC 2024
Yogyakarta, Indonesia
October 25th-27th, 2024

Story: The Beginning



As I kept up this daily writing habit, I became **curious** about what I had actually been writing over the past few months.

Today, I'm excited to **share** with you what I discovered through this journey using **text analytics** to dive deeper into my own work.

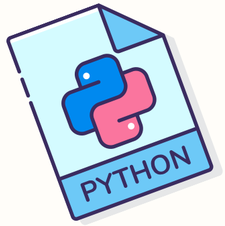
Methodology

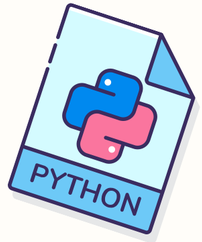
To begin the process, I used **Beautiful Soup**, a Python library for web scraping. This allowed me to extract the content of my Medium articles by navigating through the HTML structure of the site.

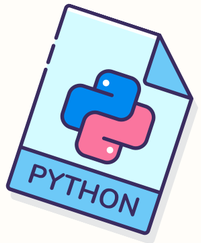
By automating the extraction process, I was able to gather all my articles efficiently, creating a dataset of my writing for further analysis

Once I had the text, I turned to **text mining** to uncover patterns. Using techniques like word **frequency analysis** and **n-grams**, I identified the **most common words and phrases** I tend to use. This helped me see **recurring themes** and **ideas** across my posts, offering a clearer picture of what I've been writing about **unconsciously.**

Finally, I applied **text analytics** to dive deeper. I used sentiment analysis to gauge the emotions behind my writing, and topic modeling to categorize articles by themes. These methods allowed me to explore not just what I wrote, but also how I wrote it, revealing insights into tone, sentiment, and underlying trends in my content.







Data Collection

	title	content	date_published
0	Know Thyself, Know Thy Enemy	azmi_ord\nFollow\n--\nListen\nShare\nThis is a...	2024-04-15 19:41:50.859000+07:00
1	The Most Insecure Person in the Room	azmi_ord\nFollow\n--\nListen\nShare\nI previou...	2024-04-17 19:27:07.213000+07:00
2	Are You Here Now?	azmi_ord\nFollow\n--\nListen\nShare\nWhen you ...	2024-04-18 19:08:24.342000+07:00
3	Certainty Means Inaction	azmi_ord\nFollow\n--\nListen\nShare\nWhen we g...	2024-04-19 16:48:07.793000+07:00
4	The Entrepreneur, The Manager, and The Technician	azmi_ord\nFollow\n--\nShare\nSaat gw nulis ini...	2024-04-20 19:41:23.876000+07:00

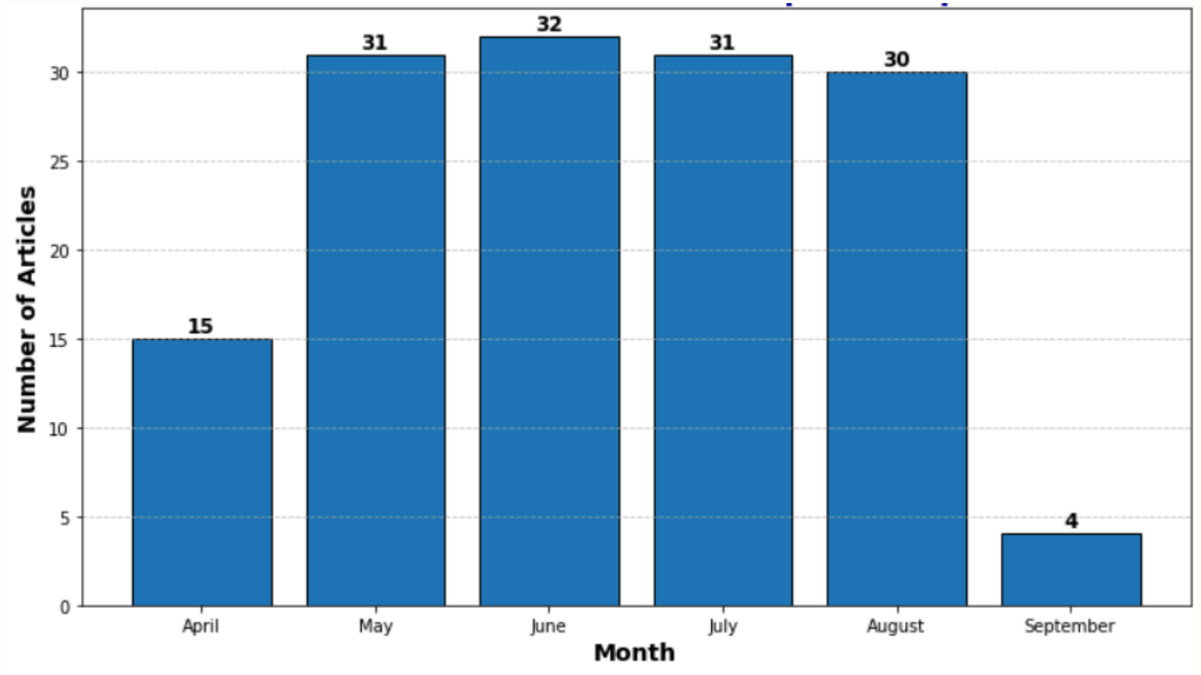
my medium account: azmi_ord

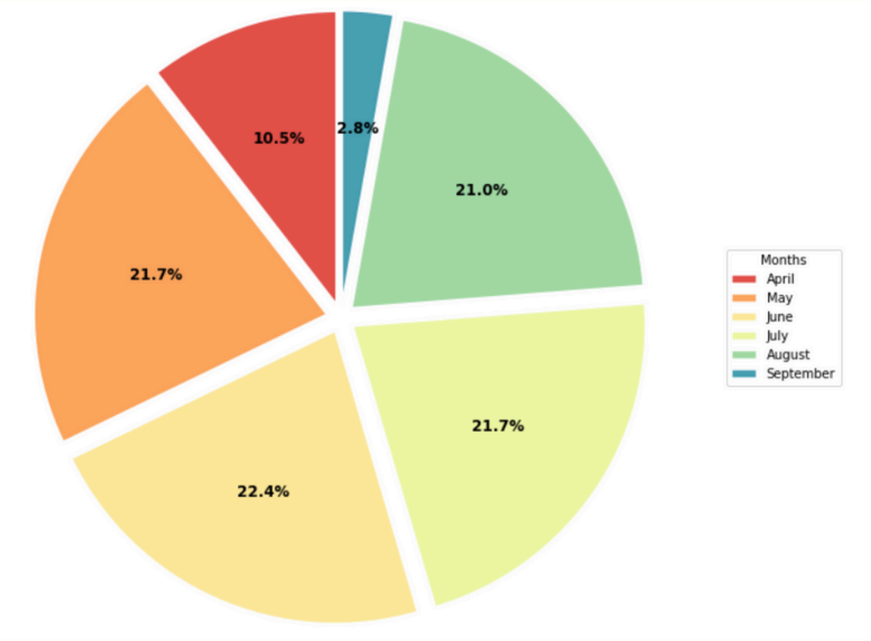
To begin the analysis, I **scraped** my **Medium articles** using **Beautiful Soup**, collecting the raw text from each post.

This allowed me to create a **dataset** of my writing over the past six months.

Here's a quick glimpse of the articles **I've written**, now transformed into a structured format, **ready for further exploration** through text mining and analytics.

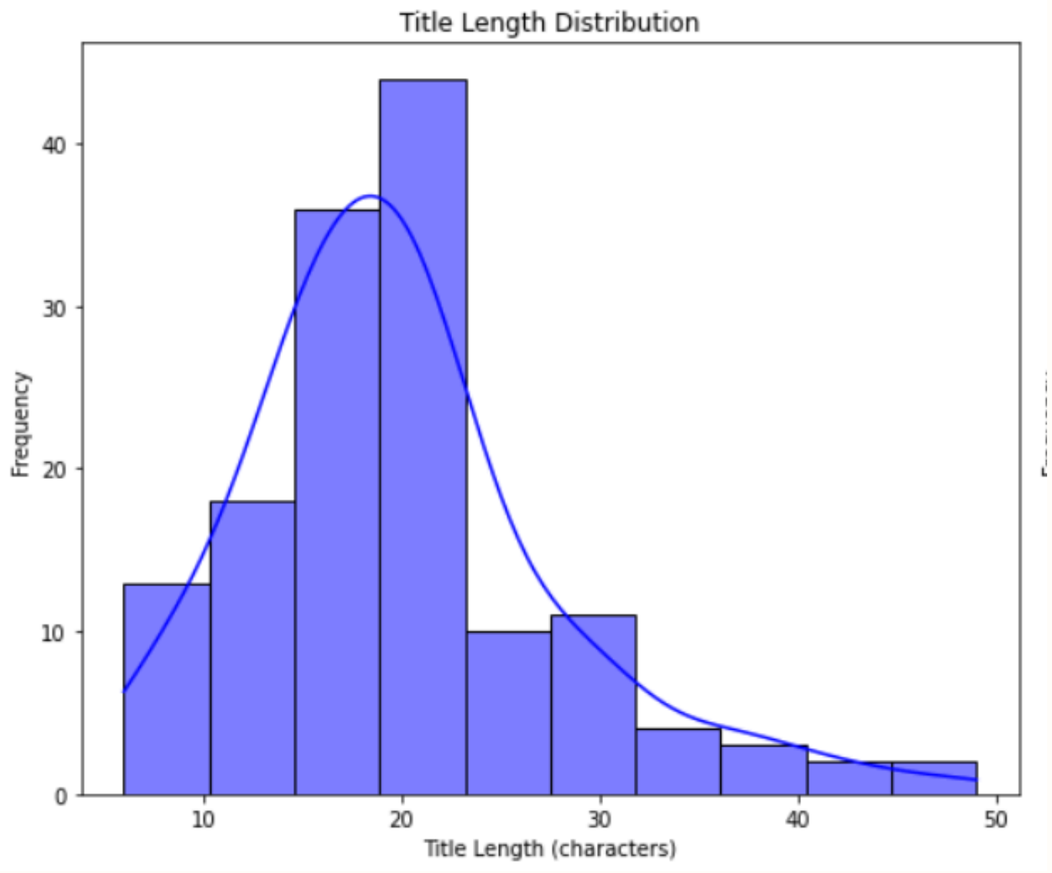
What my writing reveals





This chart illustrates my daily publishing journey on Medium from **mid-April** to **early September 2024**. After beginning with 15 articles in a partial April, I maintained a robust daily writing schedule, publishing 30-32 articles each month during the full months of **May** through **August**.

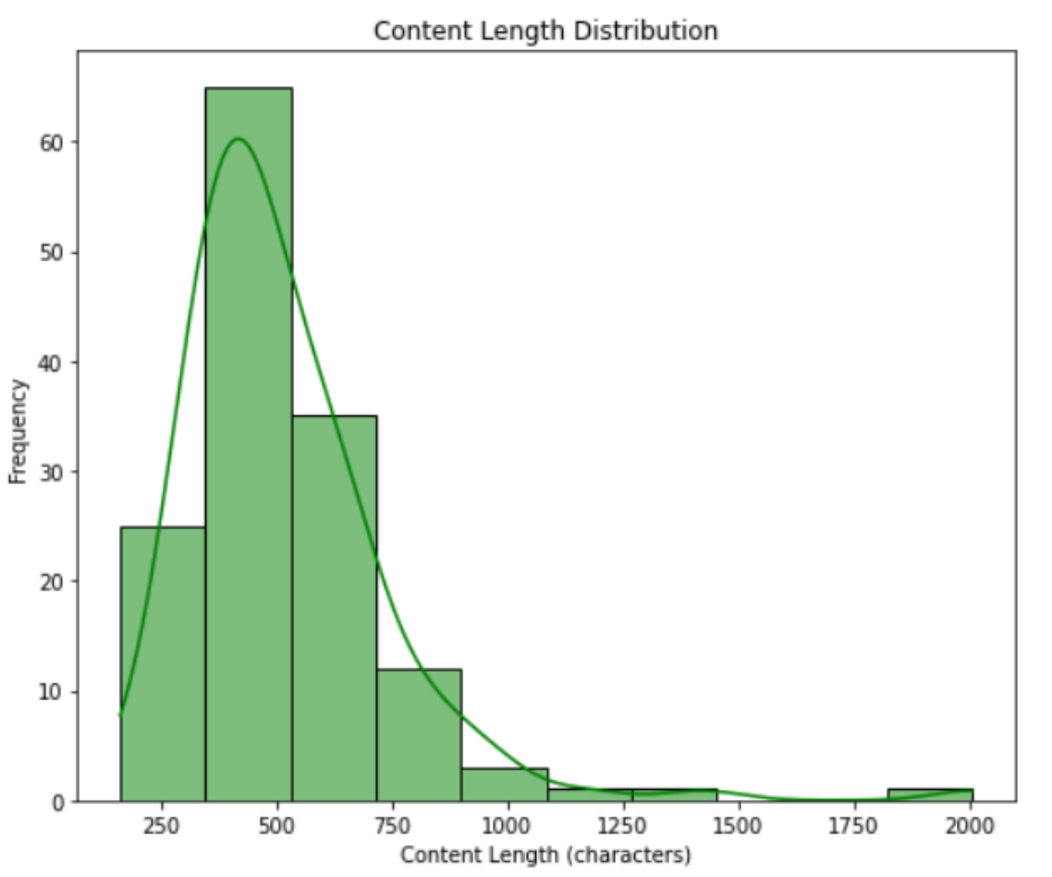
What my writing reveals



The majority of article titles have a length between **15 to 25 characters**, with the peak frequency around **20 characters.**

The distribution appears to be **right-skewed**, meaning there are some titles with longer lengths (up to 50 characters), but they are **less frequent**.

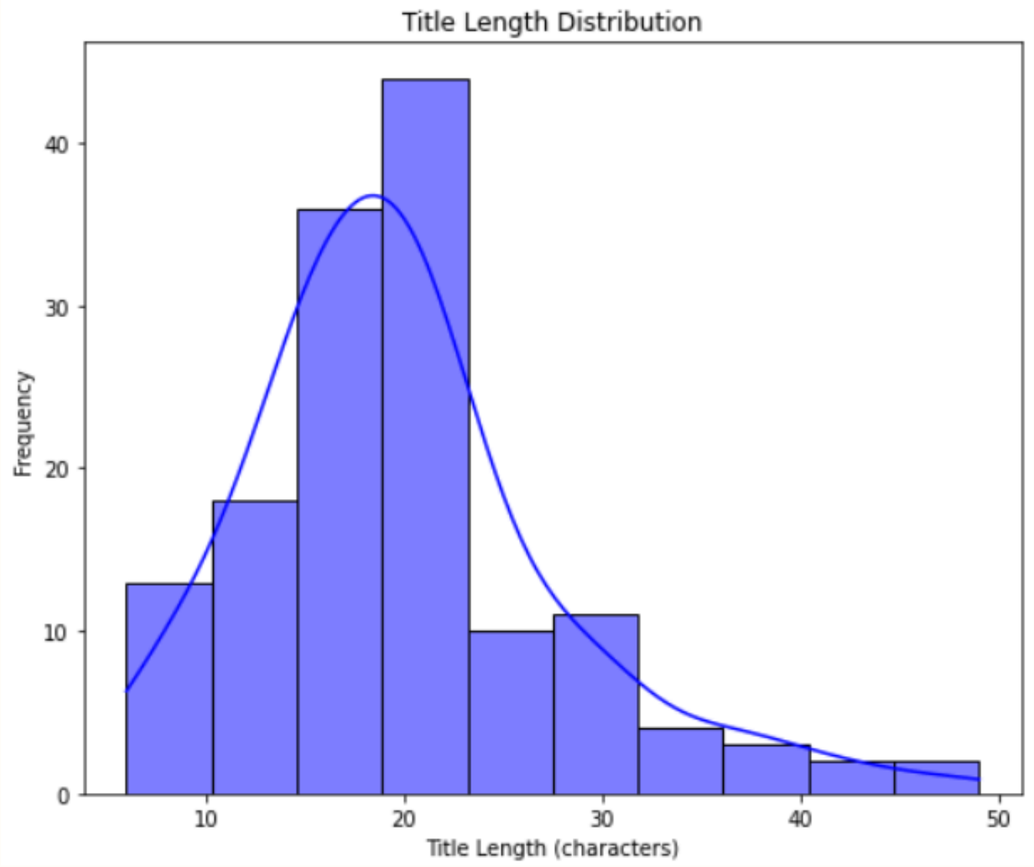
What my writing reveals

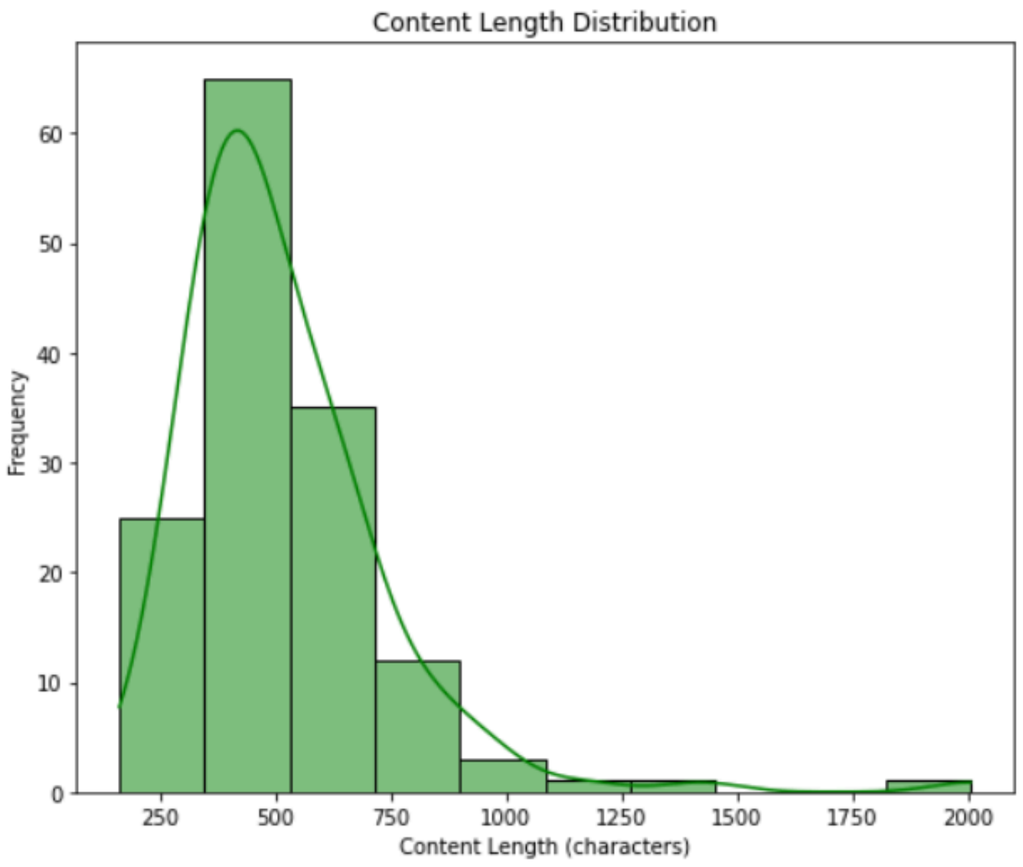


Most articles have a content length between **250** and **750 characters**, with the peak around **500 characters**.

This distribution is also **right-skewed**, with some articles having much longer content (up to **2000 characters**), though these longer articles are relatively rare.

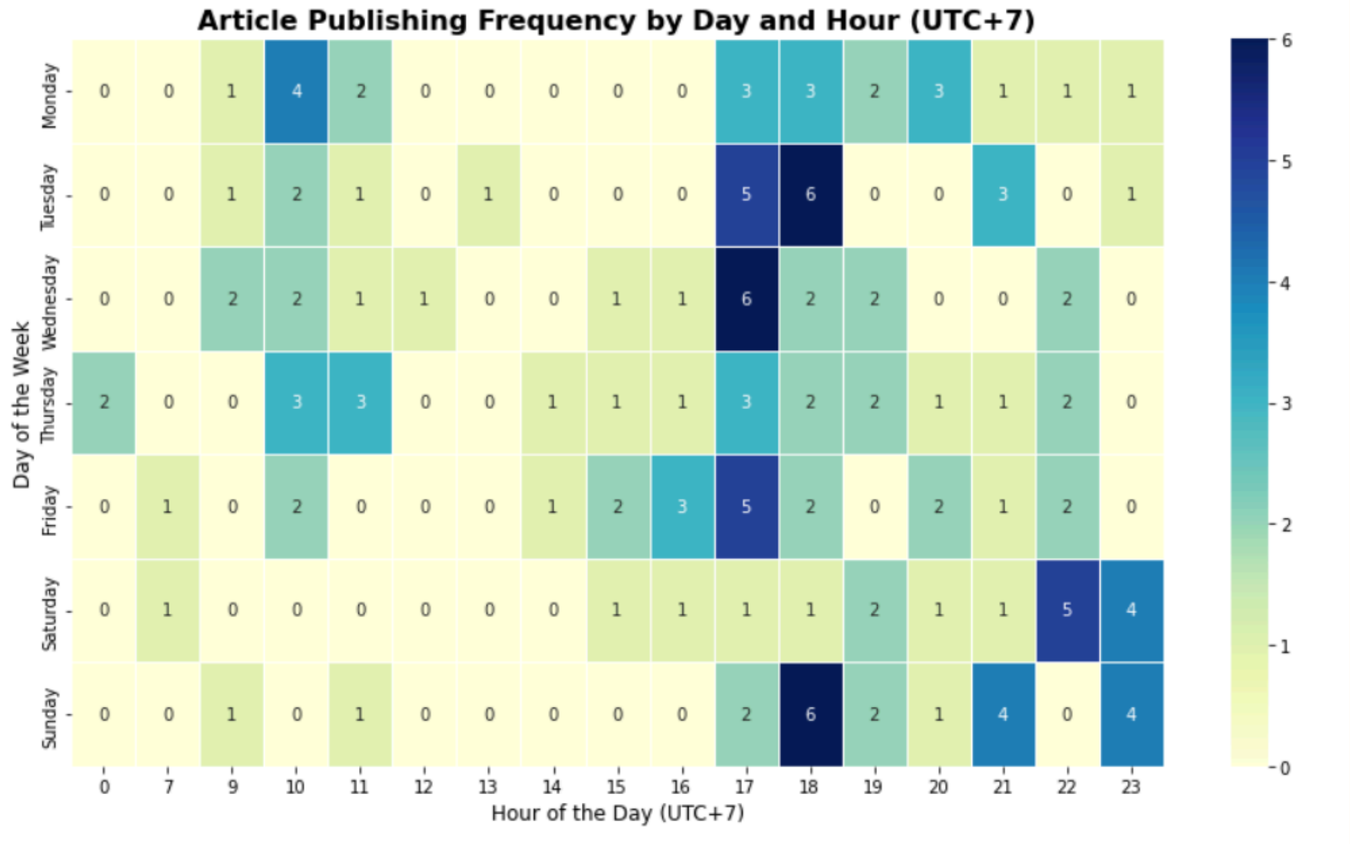
What my writing reveals





Both the title and content lengths follow a **right-skewed distribution**, where the majority are **short** to **medium** in **length**, with fewer long articles or titles. Titles are generally **shorter** and fall within a **narrower range**, while **content length varies more widely**.

What my writing reveals



Article publishing appears to be more **frequent** in the **late afternoon** and **evening** (from 16:00 to 21:00) **across most days.**

The **morning hours** (7:00 to 10:00) generally have **very few or no articles** being **published**.

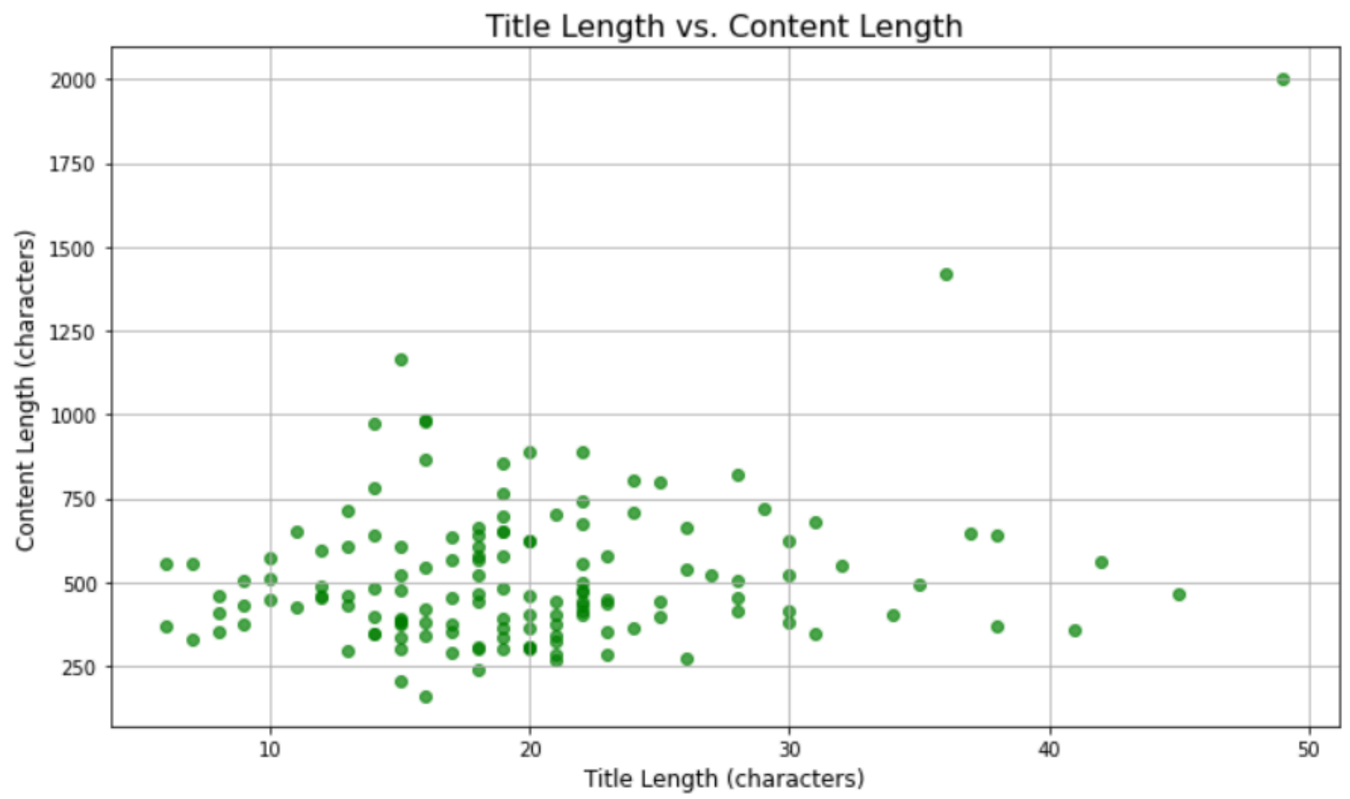
Most articles tend to be published during the **late afternoon to evening hours**, with notable peaks on **Tuesday, Wednesday, and Sunday.**

What my writing reveals

There **doesn't appear** to be a **strong correlation between title and content length.**

Most titles and content lengths are **concentrated** within **certain ranges**, but **content length** varies widely even for **similar title lengths**.

This suggests that the **author** (me) may not use **longer titles** for **longer articles**, and **vice versa**.



What my writing reveals

- **People-centric content**: The largest and most prominent word is "people," suggesting that many articles focus on human experiences, relationships, or societal issues.

- **Self-improvement and personal growth**: Words like "better," "think," "know," "change," "work," and "get" indicate a strong focus on personal development, learning, and self-improvement.

What my writing reveals

- **Emotional and psychological topics:** Words such as "feel," "fear," "like," and "need" suggest that many articles deal with emotions, mental states, and psychological concepts.

- **Life and lifestyle**: "Life" is prominently featured, indicating that many articles discuss life experiences, lifestyle choices, or life philosophy.

What my writing reveals

- **Decision-making and choices:** Words like "choices," "make," and "use" imply content related to decision-making processes and taking action.

- **Time-related concepts**: "Time" and "often" are visible, suggesting discussions about time management or temporal aspects of various topics.

What my writing reveals

- **Cognitive processes:** "Think" and "see" are prominent, indicating articles that encourage critical thinking or new perspectives.

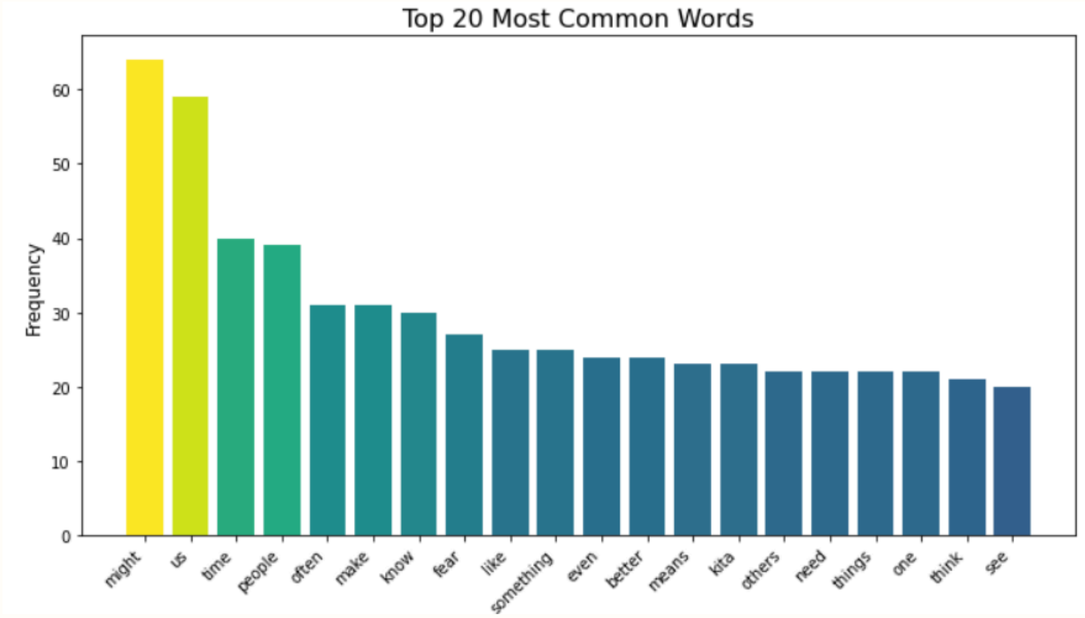
- **Social interactions:** Words like "someone," "others," and "real" might relate to social relationships and authentic connections.

What my writing reveals

- **Practical advice:** Terms such as "things," "way," "use," and "means" suggest that many articles offer practical tips or methodologies.

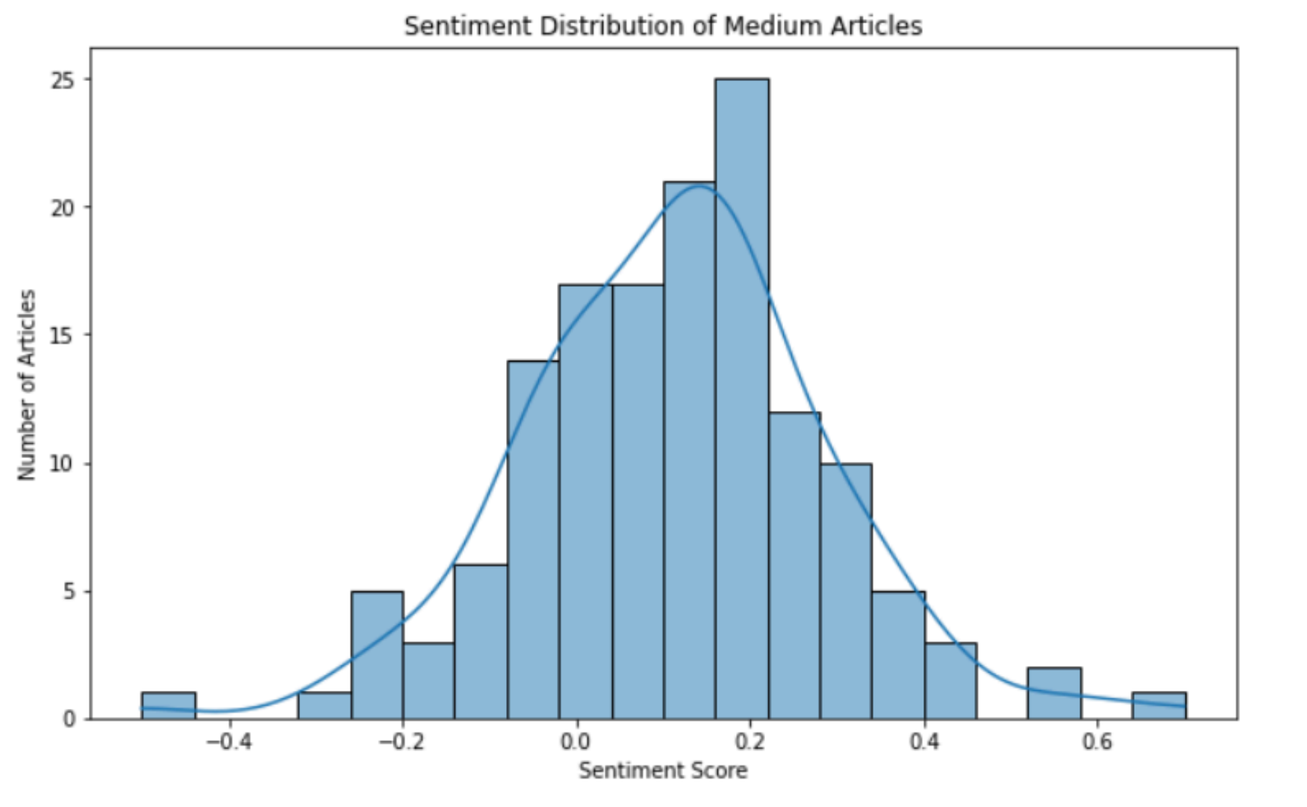
- **Abstract concepts:** Words like "world" and "something" hint at discussions of broader, more abstract ideas or global issues.

What my writing reveals



Overall, this word cloud suggests that the Medium articles analyzed tend to focus on **personal growth, human experiences, emotional well-being, and practical life advice.**

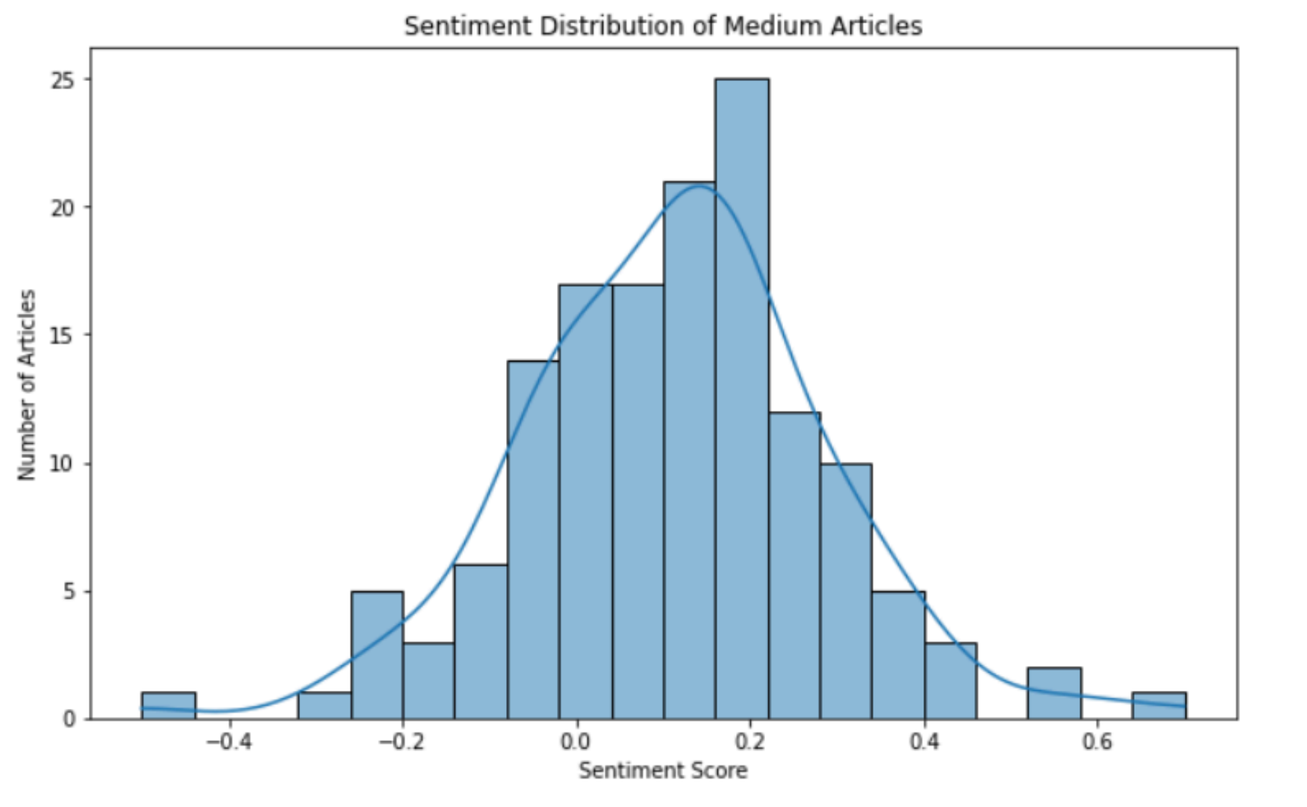
What my writing reveals



- The majority of my articles have a sentiment score **between 0.0 to 0.4**, with the peak **frequency around 0.2**.

- The distribution appears to be **slightly right-skewed**, meaning there are some articles with more positive sentiment scores (up to **0.6**), but they are **less frequent**.

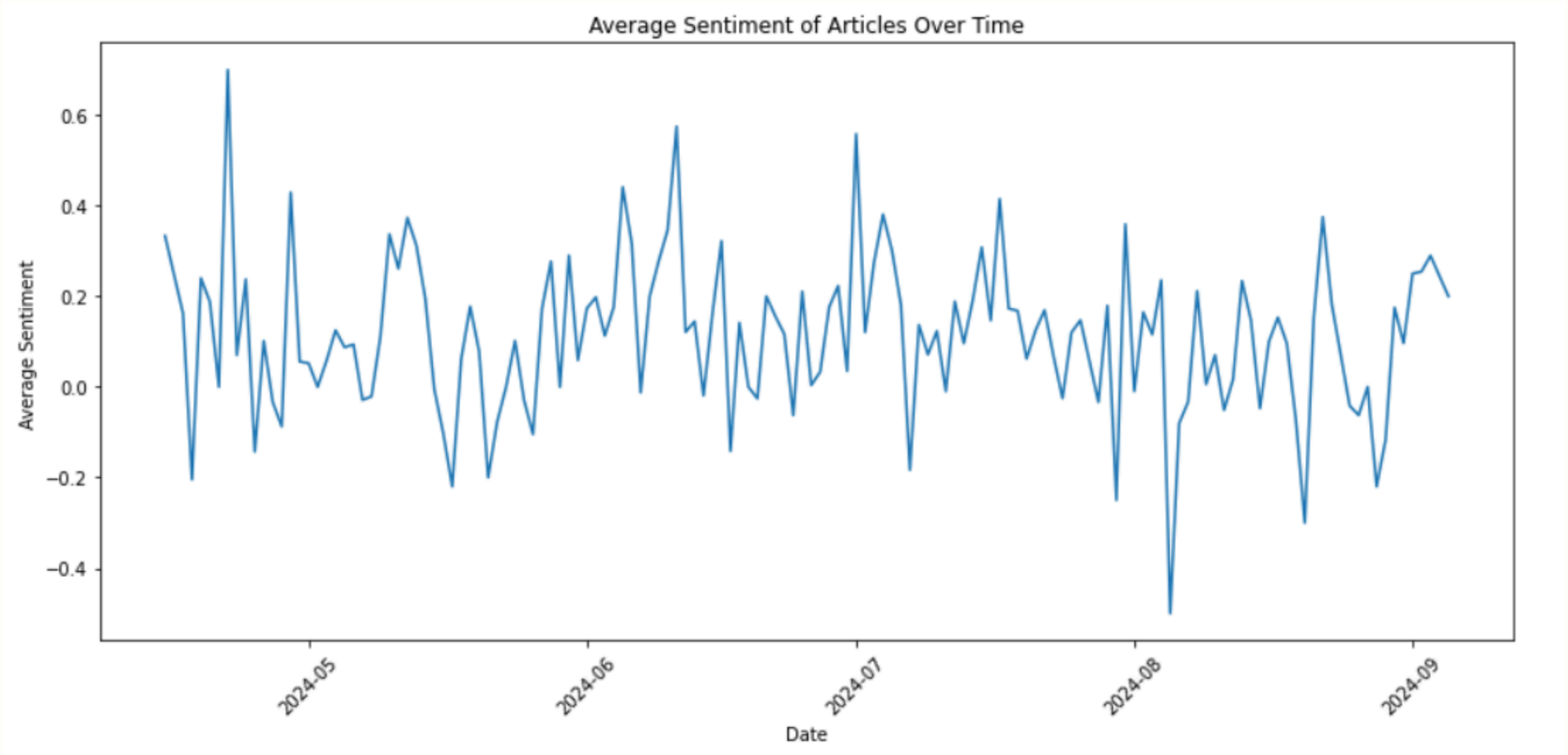
What my writing reveals



- Looking at the negative side, only a small number of your articles have sentiment scores below -0.2, suggesting that **I rarely write strongly negative content.**

- The distribution shows that around **25 articles** hit the **peak frequency at the 0.2 sentiment mark**, indicating your tendency toward **mildly positive writing.**

What my writing reveals



Despite these fluctuations, the sentiment consistently returns to a **mildly positive** baseline around **0.2**, indicating a **balanced** and **professional** writing style.

THANK YOU


NUR AZMI PRASETYO
Data Professional
Founder of Exsight Analytics
Founder of Data Wizards Community
+62-8956-0891-2030
NURAZMIPRASETYO@GMAIL.COM
NUR AZMI PRASETYO
JAKARTA, INDONESIA