MEGAPUTER NLP

MARKETING CONSULTING PROJECT

MS MARKETING
PURDUE UNIVERSITY

OUR TEAM

Bhawishya Juneja

Devanshi Parekh

Gautham Shankar Muthukumar

Mansi Swami

Shubhi Srivastava

Siqi Zhou

CONTENT

INTRODUCTION

Megaputer

Your Knowledge Partner™

UNDERSTANDING POLYANALYST

Product offering, content strategy, market positioning

RECOGNISING COMPETITORS

Blog strategy and customer pain points

OPEN-SOURCE SCRAPING

Unified data collection method

NLP

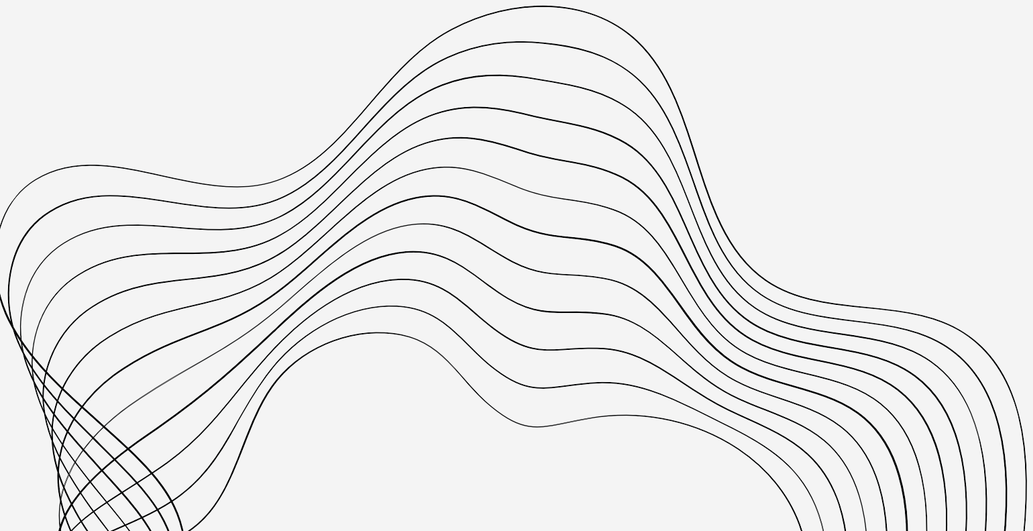
Large source data analysis method

BLOG STRATEGY

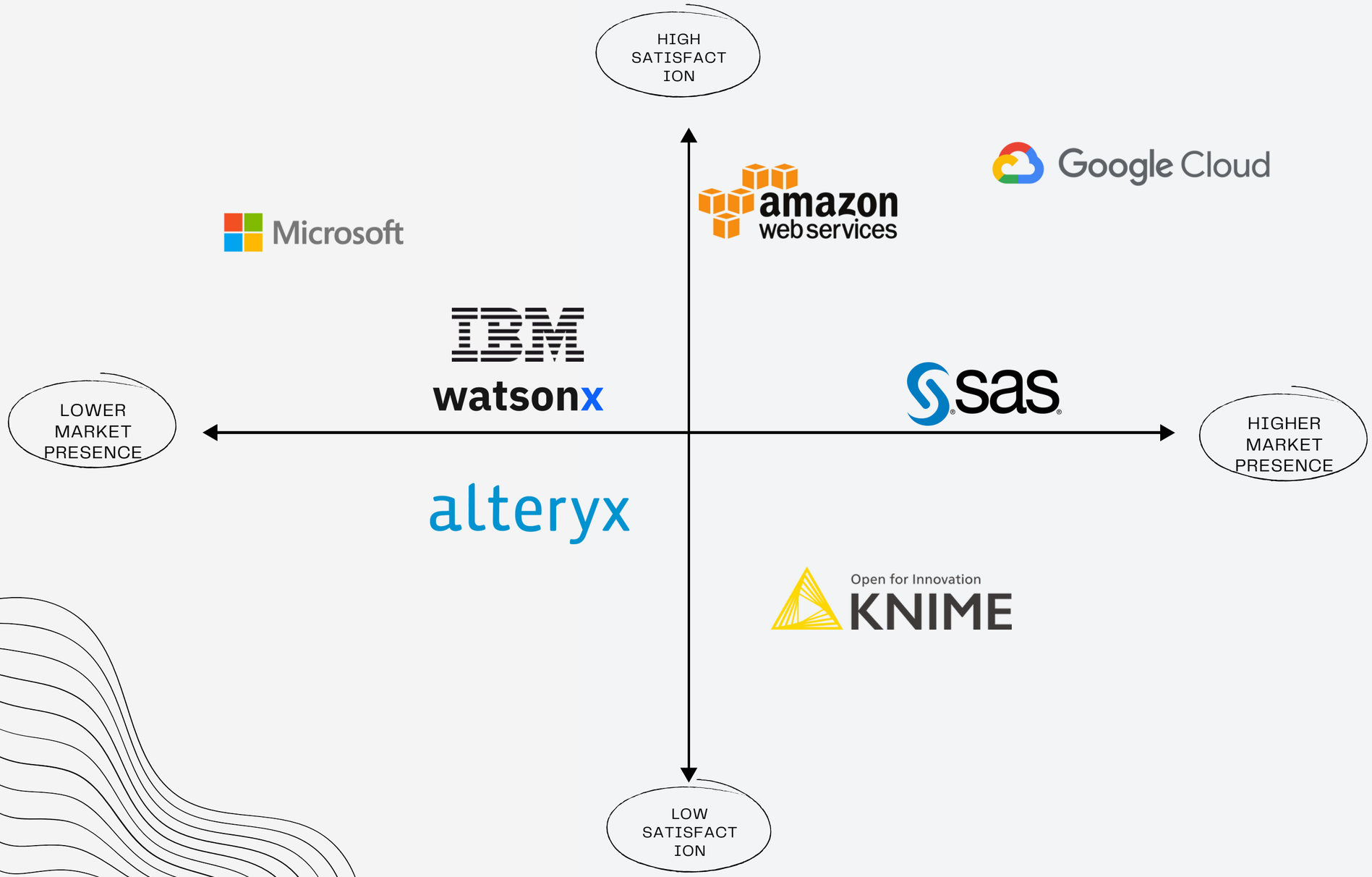
Redefine Megaputer's content strategy

CUSTOMER REVIEWS

Differentiating Megaputer's market positioning



COMPETITIVE ANALYSIS



Utilizing sources such as Gartner, G2, and Google Search, we identified the major company competitors and their market positioning.

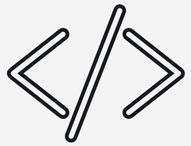
SOURCE: G2

METHODOLOGY

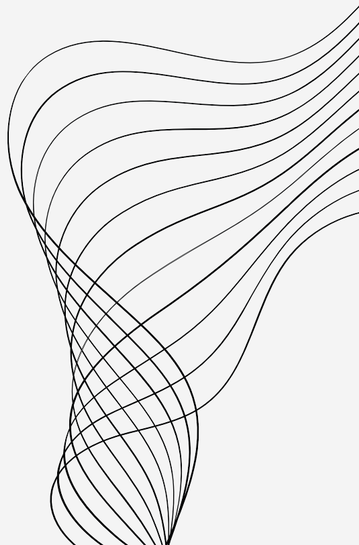
COMPREHENSIVE DATA SCRAPING STRATEGY

Accurately gather critical data from competitor websites and online sources for further in-depth competitive analysis and strategic planning.

Competitor websites



Software Review Websites



TOPIC MODELING TO DISCOVER INSIGHTS

Objective: To extract and identify popular themes and topics from a vast collection of blog text, providing insights into competitors' content strategies and focus areas.

- Scikit-learn and Gensim
- Latent Dirichlet Allocation (LDA)

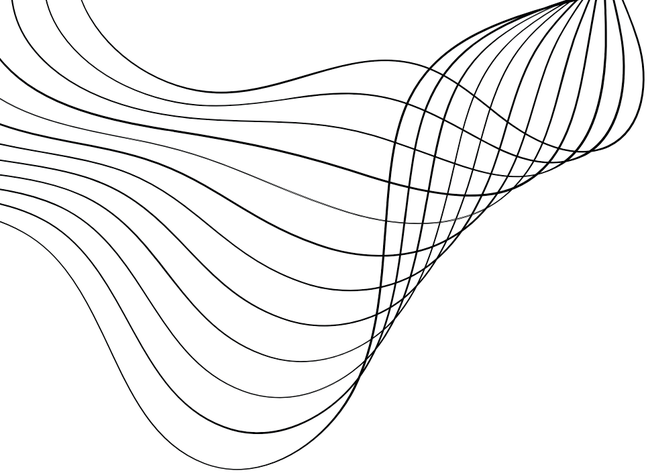
Output: identified 15 distinct topics for each competitor

TOOLS UTILIZED IN OUR METHODOLOGY

- JSON Script: Extracting URLs from various competitor blog pages.
- Scikit-learn: A machine learning library, utilized for text processing and topic modeling.
- NLTK (Natural Language Toolkit): Used for natural language processing tasks such as tokenization and stop word removal.
- Gensim: Elevates topic modeling by statistically crafting semantic document representations, unlocking deeper thematic insights.
- BeautifulSoup: Utilized for parsing HTML and XML documents, facilitating data extraction from web pages.
- G2 Scraping Extension: Gathering user reviews and insights.
- Gen AI: Interpreting results from the topic model.

PROCESS & CHALLENGES

FOR COMPETITOR BLOG ANALYSIS



1
COMPETITOR IDENTIFICATION

2
COMPETITOR BLOG URL SCRAPING

3
BLOG TEXT SCRAPING

4
TOPIC MODELLING AND INTERPRETATION

5
DATA VISUALIZATION

COMPETITOR IDENTIFICATION

PolyAnalyst was chosen as the focus; competitors identified via G2 and Gartner, confirmed by client: Alteryx, KNIME, AWS Comprehend, IBM WatsonX, SAS Viya, Google Cloud NLP, Microsoft Azure NLP.

7

MAJOR COMPETITORS

SCRAPING AND ANALYSIS

Competitor Blog URL’s

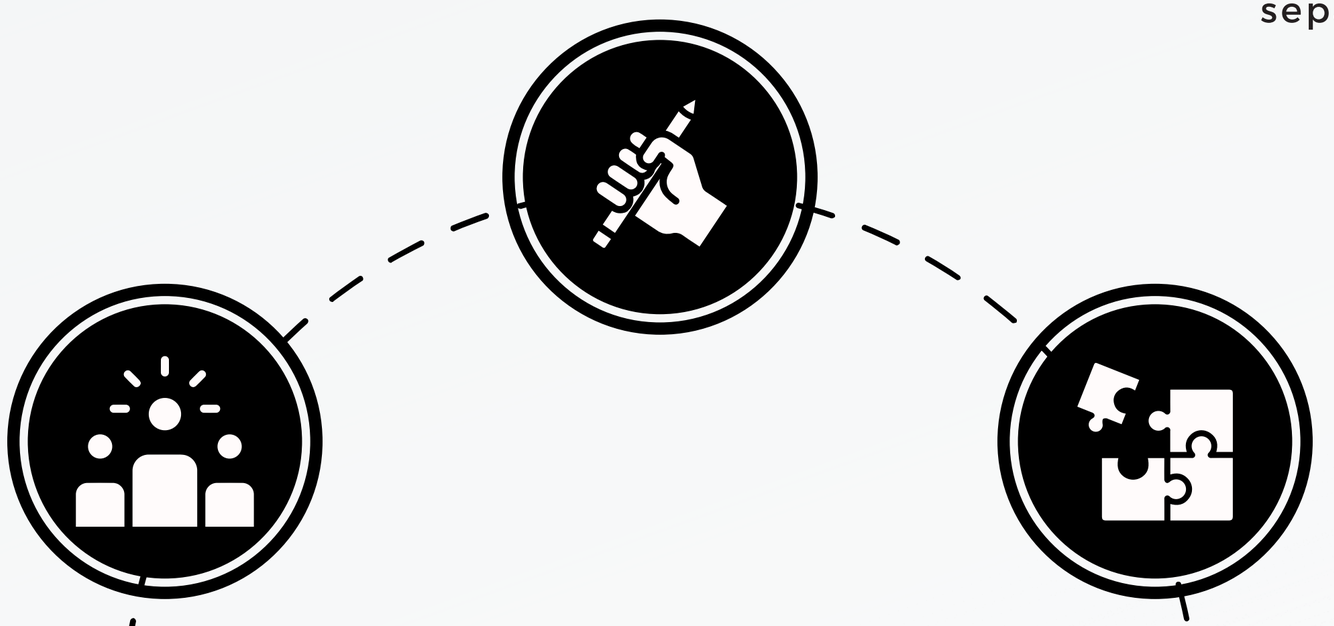
A JavaScript efficiently extracts competitor blog URLs, generating a JSON file with the first 100 entries from each blog

Blog text

BeautifulSoup library is used to scrape text from URLs, extract relevant data, and save it in ".txt" format for text analysis

Text Analysis

NLTK library used for text cleaning, removing stop words; keyword frequency analysis conducted, generating separate CSVs for top 100 keywords



TOPIC MODELING & VISUALISATION

Topic modeling with Scikit-learn and Gensim libraries uncovered 15 thematic clusters from blog text files, with Gensim yielding better results for Microsoft, KNIME, & Alteryx, and Scikit-Learn for SAS VIYA, AWS, IBM Watson, & Google.

TOPIC MODELING

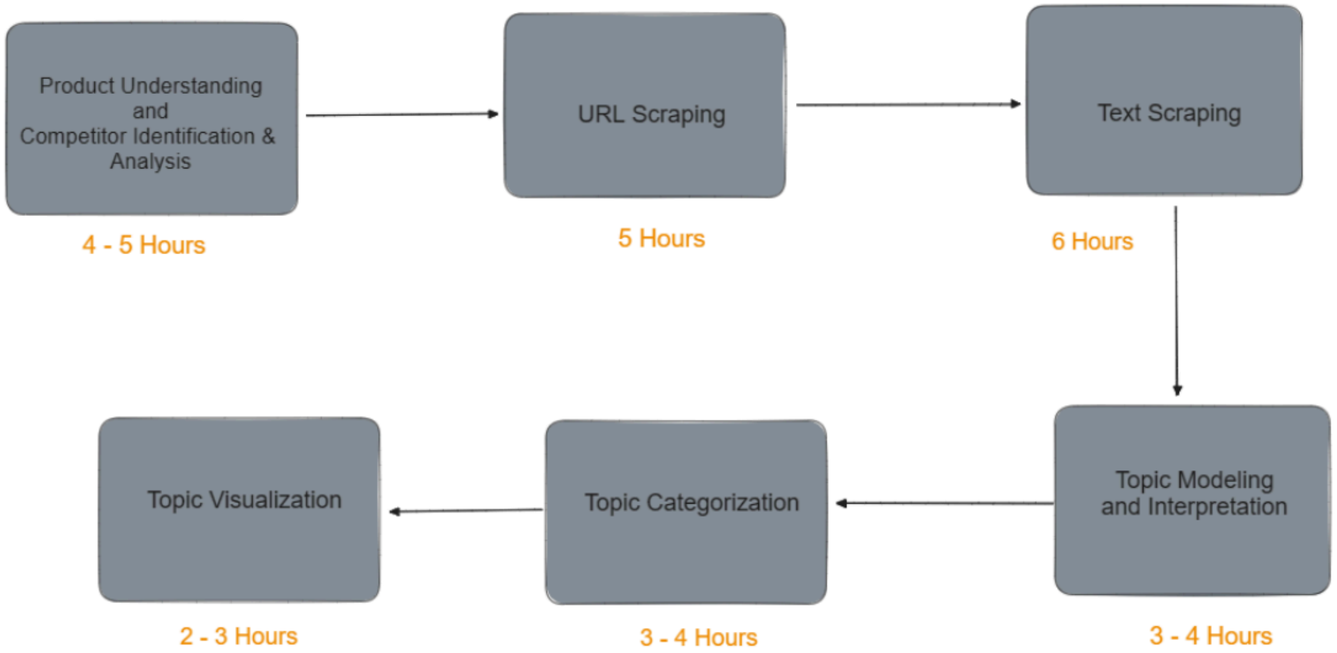
Gen AI tools interpreted LDA-generated topics; manual categorization was conducted based on blog content for industry use cases, informative blogs, testimonials, etc.

INTERPRETATION

Tableau and Excel used for two-way visualization: competitor blogs categorized under overarching themes and competitor-specific analysis for comprehensive insight into blog content strategy.

VISUALIZATION

FINAL TIMELINE (EXCLUDING ROADBLOCKS)



Total Time: 23 - 27 Hours

PROCESS & CHALLENGES

FOR CONSUMER PAIN POINTS

01

Competitor Pain Points Analysis

02

Identifying sources for data collection

03

Feasibility Check

04

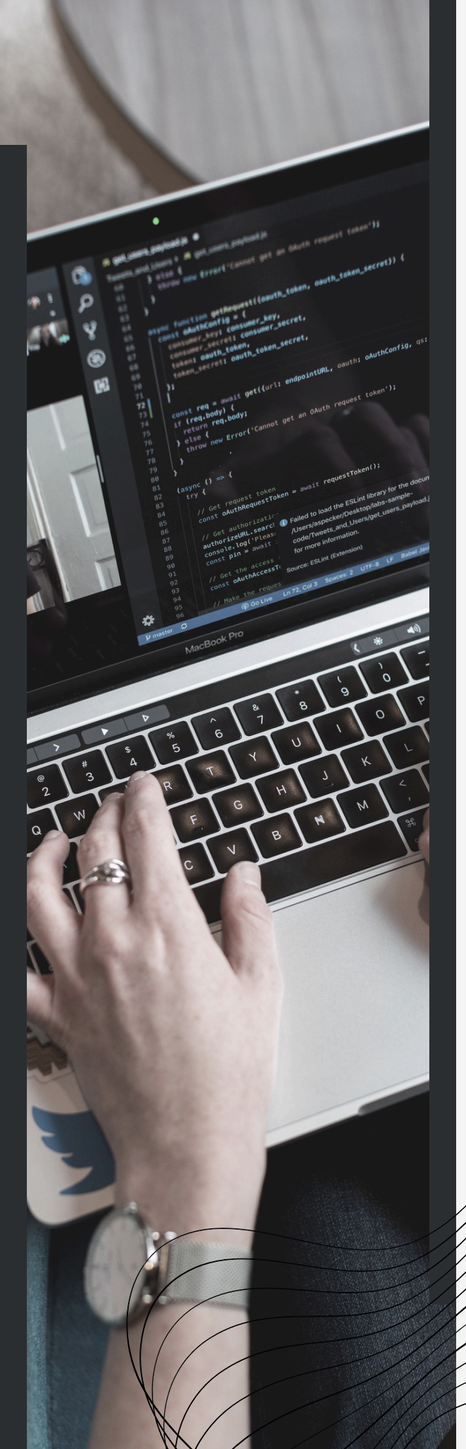
Data Scraping, Modeling and Interpretation

IDENTIFYING SOURCES FOR DATA COLLECTION

Social
Twitter/ Youtube/ Reddit

Third party analysts
G2/ Gartner

Developer communities



FEASIBILITY CHECK

REDDIT

Availability of only paid APIs

GARTNER

Scraping through API didn't work due to firewall security

TWITTER

Initial issues in obtaining the free API key/ authorization issues

USER COMMUNITIES

Lack of relevant information, mostly focused on FAQs

YOUTUBE

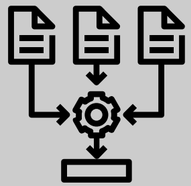
Lack of review videos

G2

Scraping possible only through browser extension

Scraping:

G2 via browser extension



Modeling:

Python libraries (NLTK + Gensim or NLTK+ Scikit)

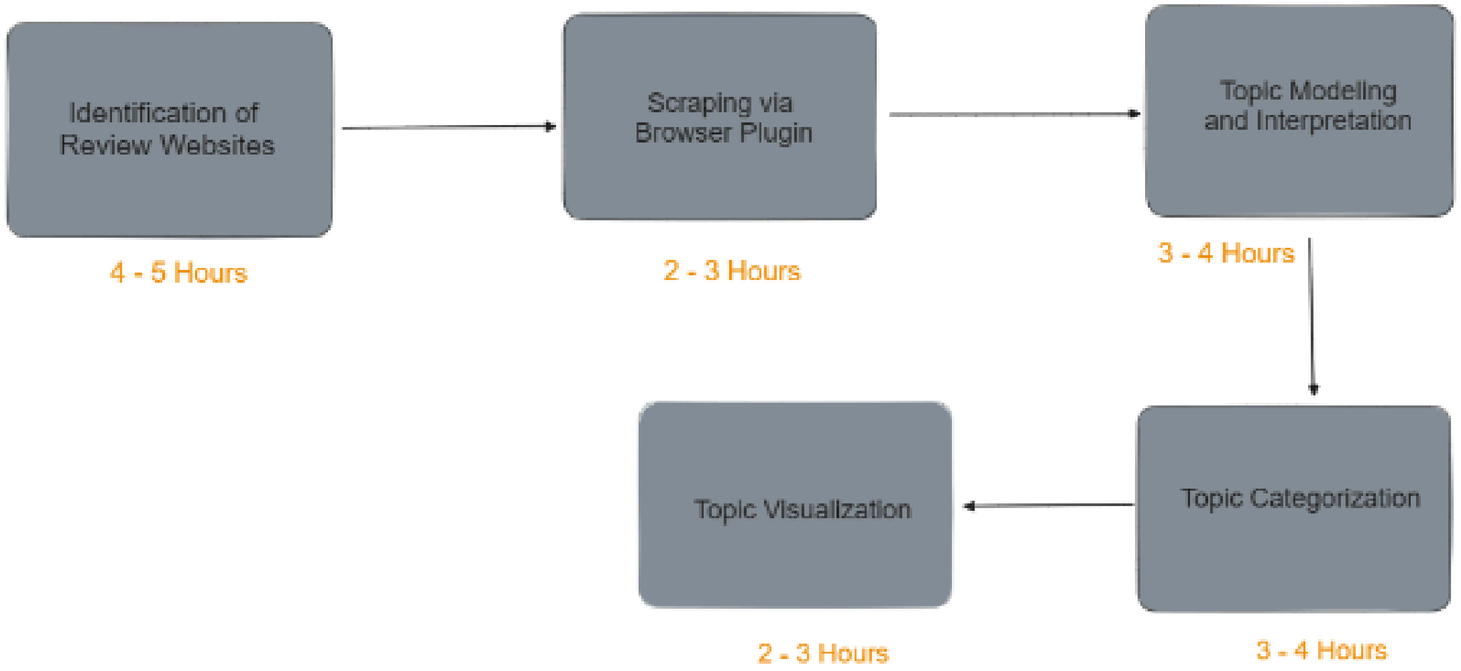
Interpretation:

Gen AI tools and manual categorization

DATA SCRAPING, MODELING AND INTERPRETATION

FINAL TIMELINE

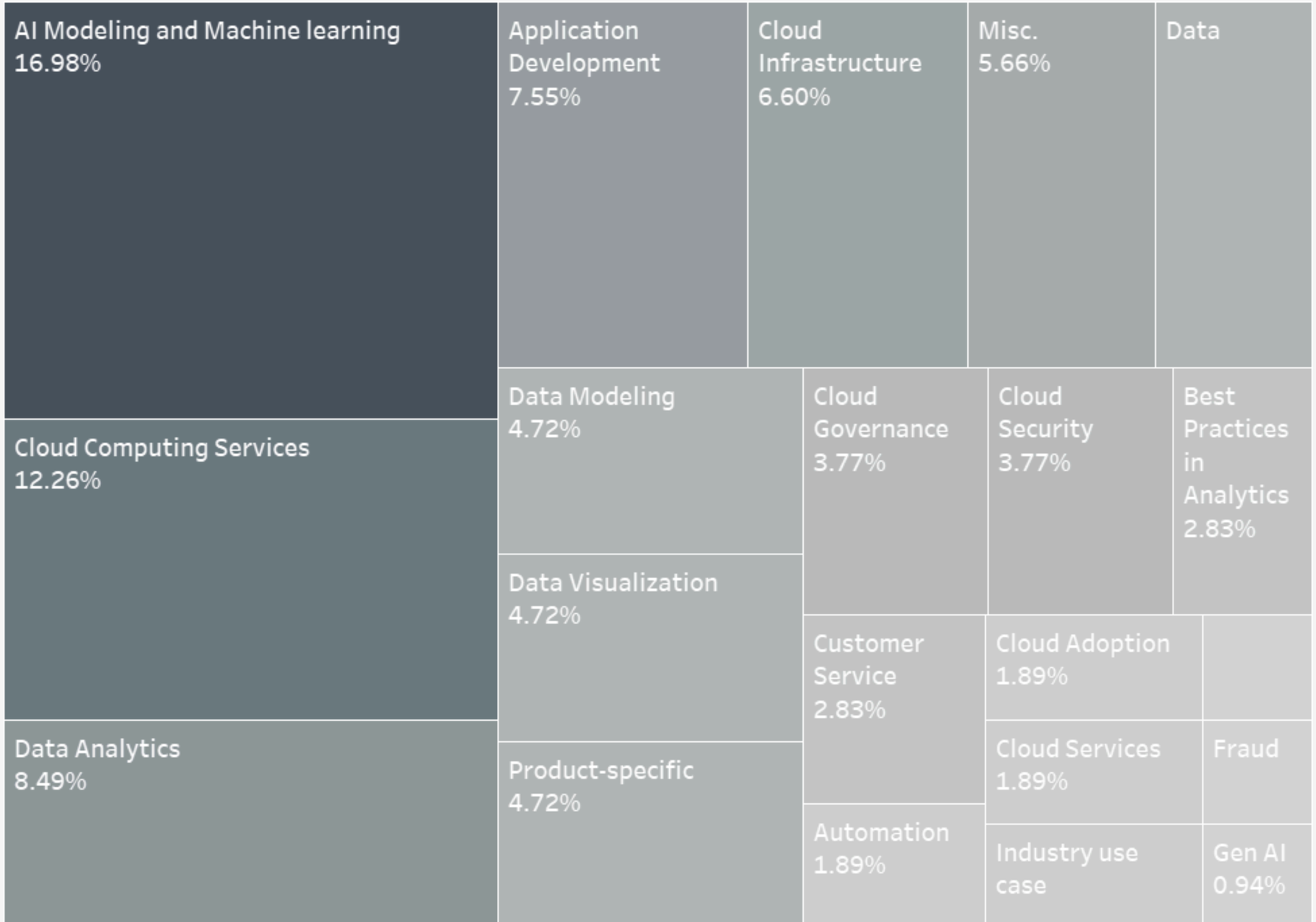
(EXCLUDING ROADBLOCKS)



Total Time: 14 - 19 Hours

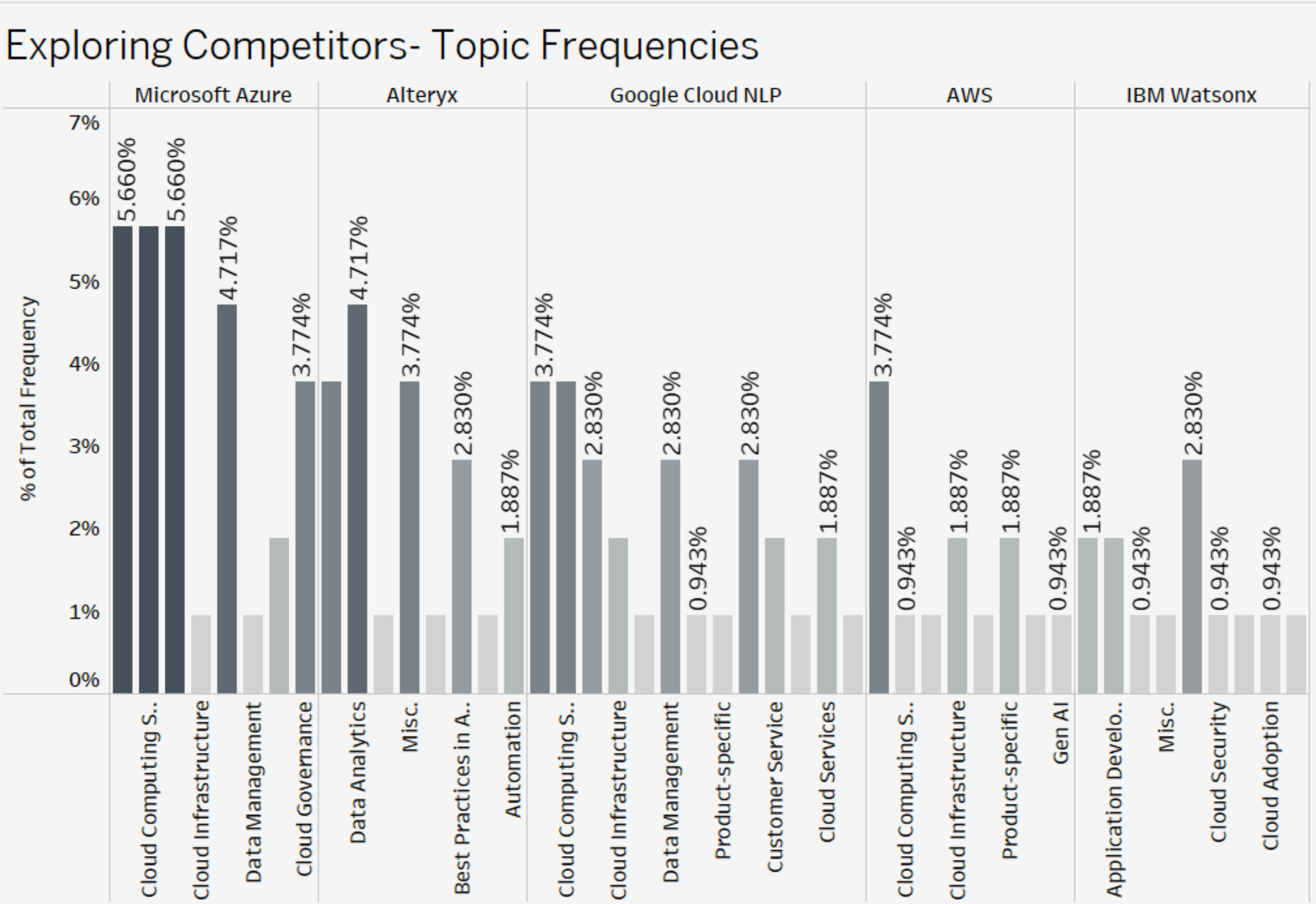
DATA VISUALIZATION

BLOG RECOMMENDATION : TOPIC FREQUENCIES



CONSUMER PAIN POINTS : TOPIC FREQUENCIES

FOR BLOG RECOMMENDATION



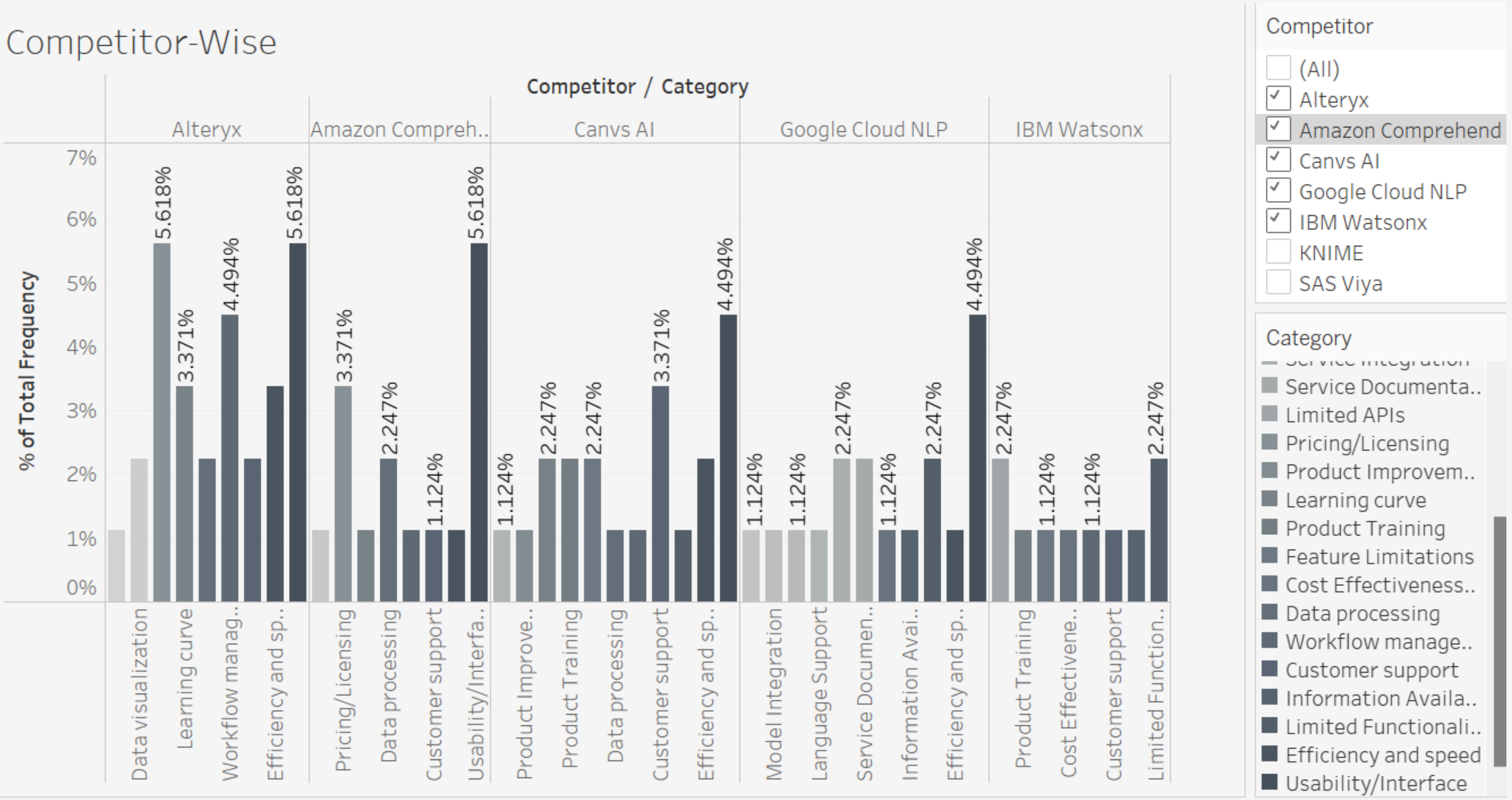
Action (Category)

- [x] (All)
- [x] AI Modeling and ...
- [x] Application Deve...
- [x] Automation
- [x] Best Practices in ...
- [x] Cloud Adoption
- [x] Cloud Computing...
- [x] Cloud Governance
- [x] Cloud Infrastruct...
- [x] Cloud Security
- [x] Cloud Services
- [x] Customer Service
- [x] Data Analytics
- [x] Data Management
- [x] Data Modeling
- [x] Data Visualization

Company Name

- [x] (All)
- [x] Alteryx
- [x] AWS
- [x] Google Cloud NLP
- [x] IBM Watsonx
- [x] Microsoft Azure

FOR CONSUMER PAIN POINTS



RECOMMENDATIONS

BLOG ANALYSIS

BLOG STRATEGY

Targeting Users of PolyAnalyst

Objectives:
Educate on PolyAnalyst features
Enhance customer support
Foster community and collaboration

Content Focus:
Product showcases, unique features
How-to guides and tutorials for PolyAnalyst usage

Targeting Decision Makers and Executives

Objectives:
Lead generation & customer
New business expansion
Brand awareness

Content Focus:
PolyAnalyst's business value
Industry insights
Customer success stories

THEMES AND TOPICS

Current Topics Strengths of Megaputer

Megaputer's Data Analysis Expertise

Tool-Centric Approach with PolyAnalyst

Themes from the results of Topic Modeling

Product/tool specific

Data analytics & management

AI modeling and machine learning

RECOMMENDATIONS

Industry-Specific Content: Integrate topics like "AI in Healthcare" and "AI in Finance."

Thought Leadership: Add "Future Trends in AI" and "Advancements in NLP."

Trust Building & Community : Incorporate "Ethical Considerations in AI" , "Navigating Data Privacy Laws", “Trouble Shooting”, “Feature Updates”.

PLAN OF ACTION

01

Create and distribute 3-4 blogs per week

02

Initiate content generation i.e, whitepapers & case studies

03

Distribute across social media and email.

04

Boost visibility, SEO, and lead gen

RECOMMENDATIONS

COMPETITOR PAINPOINTS

After a manual sweep, we saw similar key pain points to our topic modeling results..

INTERFACE/ USABILITY

What do you dislike about KNIME Software?

The user interface can be improved; it's not very beginner friendly. It lags sometimes and becomes very slow when working with large files. It takes up a lot of memory just to start the software.

What do you dislike about Amazon Comprehend?

One thing that I find slightly challenging is the pricing model. Amazon Comprehend charges based on the number of units processed, and it can get quite expensive when dealing with large volumes of text data.

For smaller businesses or startups on a tight budget, it might pose some limitations in terms of affordability. I wish there were more flexible pricing options or plans tailored for different business sizes.

PRICING/ LICENSING

CUSTOMER SUPPORT

What do you dislike about Alteryx?

The structure of their customer support and ticket limitation is rather strange and limits the amount of tickets that can be opened. It would also be great if the community would be given more transparency about the product roadmap.

What do you dislike about Canvs AI?

Sometimes it doesn't capture the nuances in the language and the coding ultimately isn't that helpful for our purposes, which is identifying common themes.

DATA PROCESSING

LEARNING CURVE

What do you dislike about KNIME Software?

Knime requires a lot of configuration for even the most simple excel tasks that are pretty challenging for new users to understand and work with. Apart from this, everything works well.

What do you dislike about SAS Viya?

While there are some resources on the SAS website, there is not as much open source support (ie Google results) for Viya as there is for some other visual analytic software, although this is likely to improve over time.

INFORMATION AVAILABILITY

3 PHASED MARKET REPOSITIONING

Differentiating Megaputer based on End-User Review

PRIMARY FOCUS

Usability & Interface 19.17%
Pricing or Licensing 7.50%

SECONDARY FOCUS

Efficiency & Speed 9.17%
Customer Support 8.33%

OTHERS

Data Processing 5.83%
Information Availability 4.17%
Feature Limitations 3.33%

LEAST FOCUS

Although recognized in our analysis, customers did not have many complaints related to actual product, NLP & API offerings.

THANK YOU

MR. JAYESH PANIGRAHI AND PROF. JINSUH LEE