

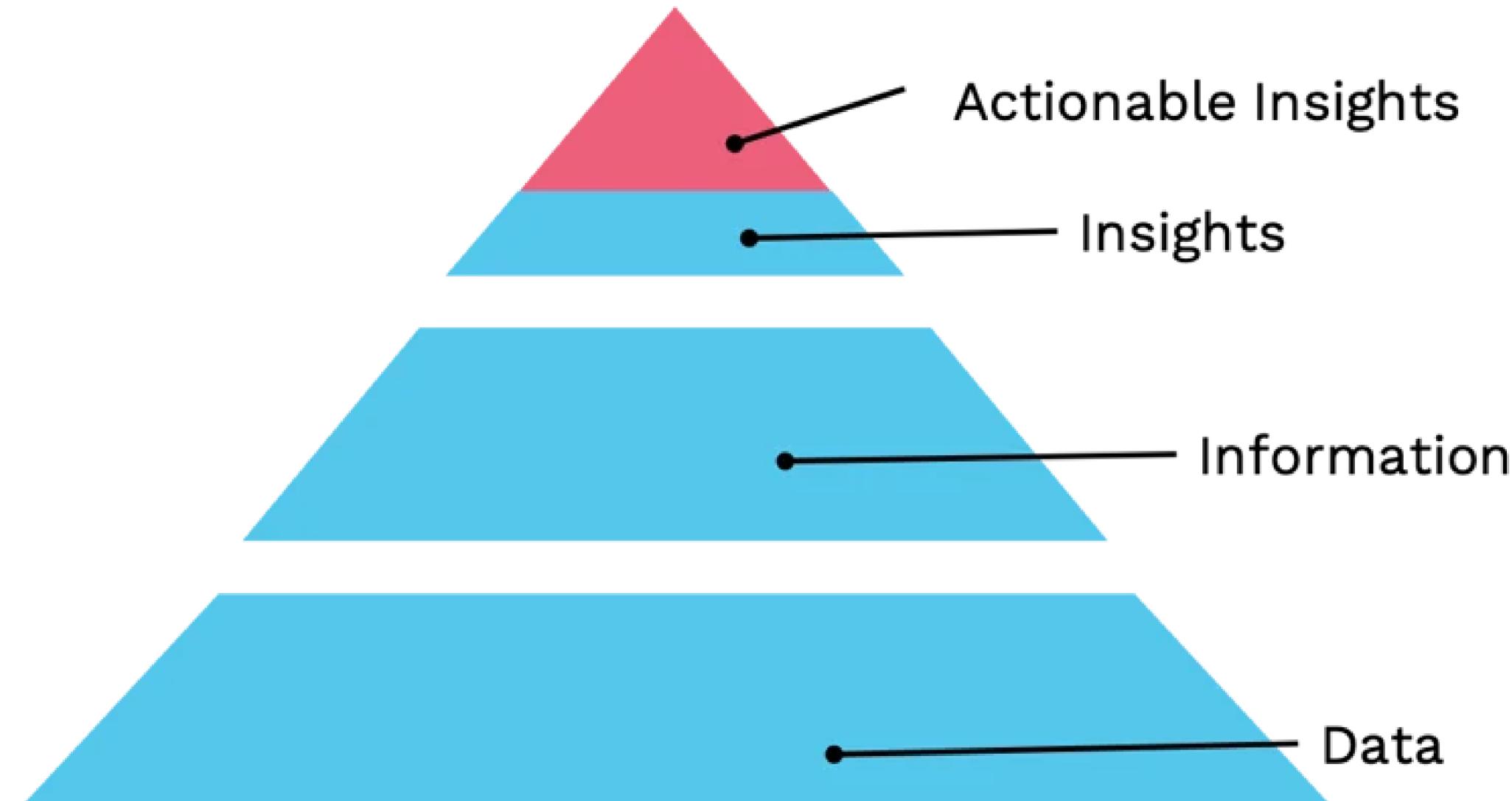
# **Identifying Consumer Preferences and Marketing Insights with BERTopic from E-commerce Reviews**

**YUFEI HUANG**

**Bocconi**

## INTRODUCTION

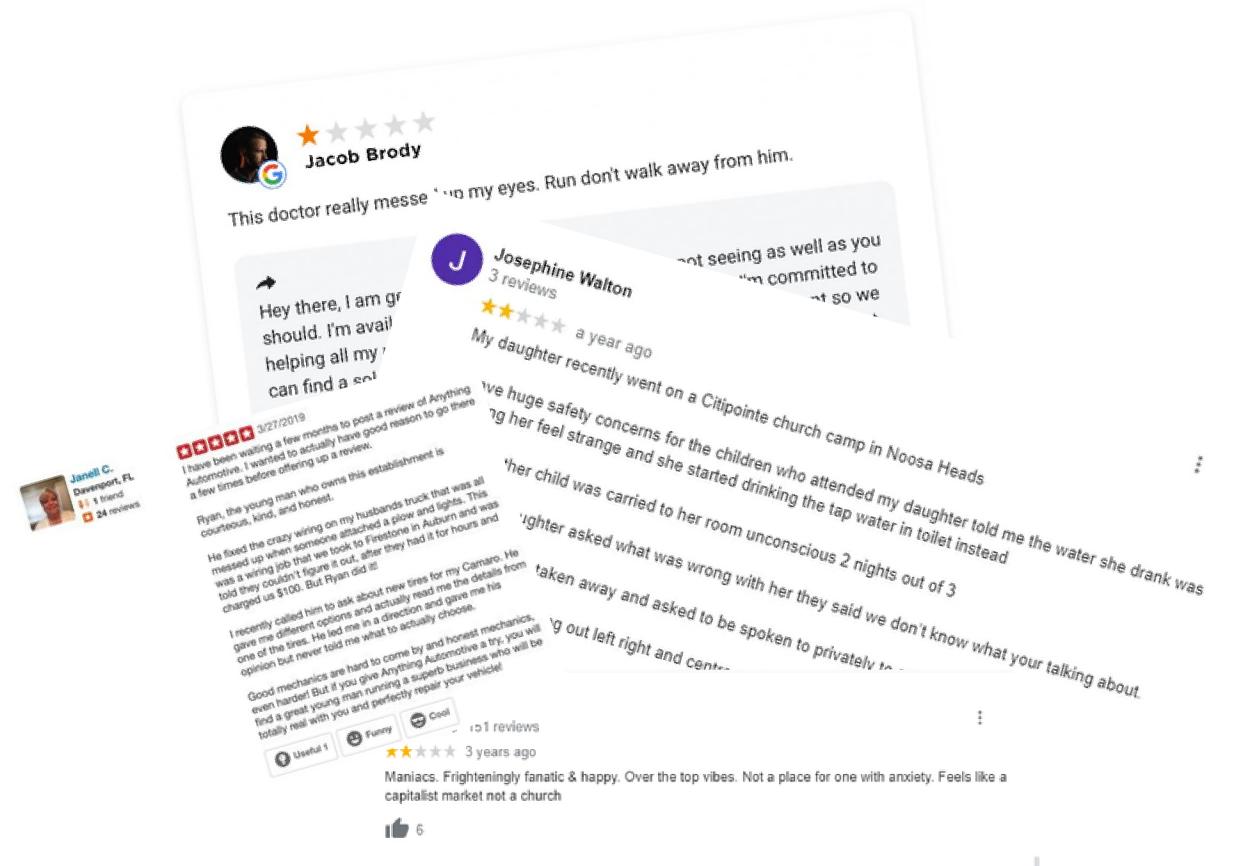
### Why Online Reviews



# INTRODUCTION

## The Power of Consumer Reviews

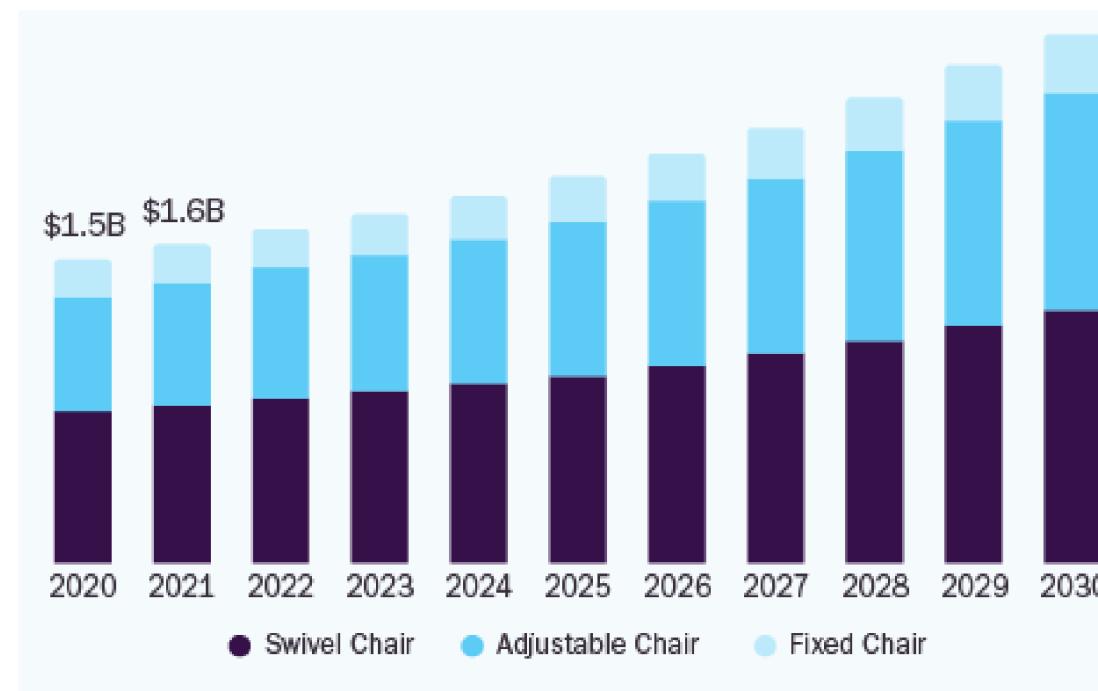
	Average levels of conformity per GRI guideline sector (1 if <25%, 2 if 26-50%, 3 if 51-75% and 4 if >75%)					
	Vision and Strategy	Profile	Governance Structure and Management Systems	Economic Performance Indicators	Environmental Performance Indicators	Social Performance Indicators
FTSE100 companies who reported use of the GRI (9)	3.1	3.0	3.1	2.9	3.2	2.8
Content analysis of the FTSE100 companies who reported use of the GRI (9)	4.0	4.0	3.8	3.3	3.1	3.4
Content analysis of other FTSE100 companies identified as conducting some form of sustainability reporting (68)	3.9	3.9	3.7	3.1	2.5	2.7



# INTRODUCTION

## Why Ergonomics Chair?

The argument for ergonomic seating is simple: People who are comfortable in their chairs are more productive. They can sit longer. They have less back, neck and leg pain. They're happier and healthier overall. But designing for humans is complicated. No two bodies are exactly alike. But that's where the biggest benefit of an ergonomic chair comes in: adjustability. Here are eight reasons why a well-designed, well-built ergonomic chair is worth the investment...



Compound annual growth rate (CAGR) of 7% from 2022 to 2031, culminating in a global market value of USD 1.6 billion in 2021.



# INTRODUCTION

## Research Gap/Challenges

	Average levels of conformity per GRI guideline sector (1 if <25%, 2 if 26-50%, 3 if 51-75% and 4 if >75%)					
	Vision and Strategy	Profile	Governance Structure and Management Systems	Economic Performance Indicators	Environmental Performance Indicators	Social Performance Indicators
FTSE100 companies who reported use of the GRI (9)	3.1	3.0	3.1	2.9	3.2	2.8
Content analysis of the FTSE100 companies who reported use of the GRI (9)	4.0	4.0	3.8	3.3	3.1	3.4
Content analysis of other FTSE100 companies identified as conducting some form of sustainability reporting (68)	3.9	3.9	3.7	3.1	2.5	2.7

01

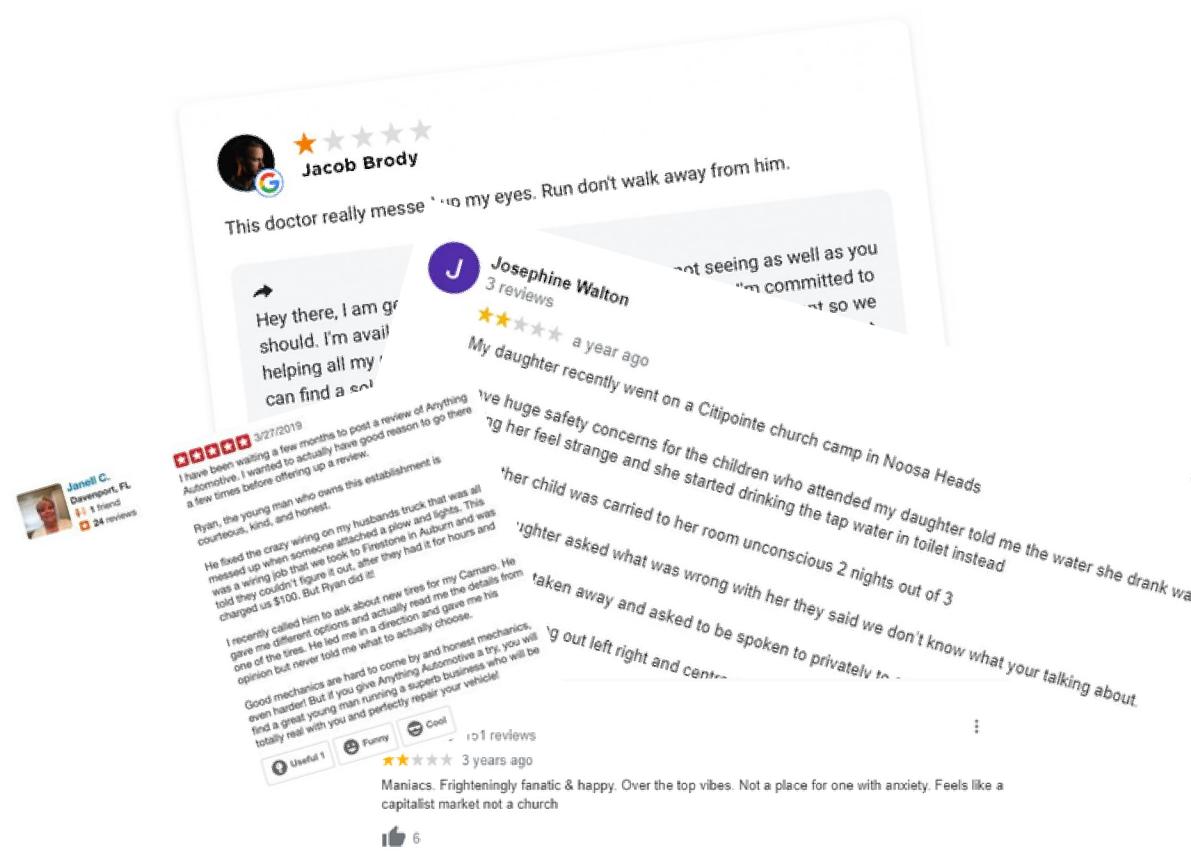
Traditional models and methods, which rely on **labeled data**

02

Latent Dirichlet Allocation (**LDA**) has been widely used in marketing to analyze consumer-related topics across different domains

03

Scholarly attention has predominantly been focused on empirical research which reveals a lag in the widespread integration of the term "**ergonomics**" **within marketing strategies**, particularly within the furniture industry.

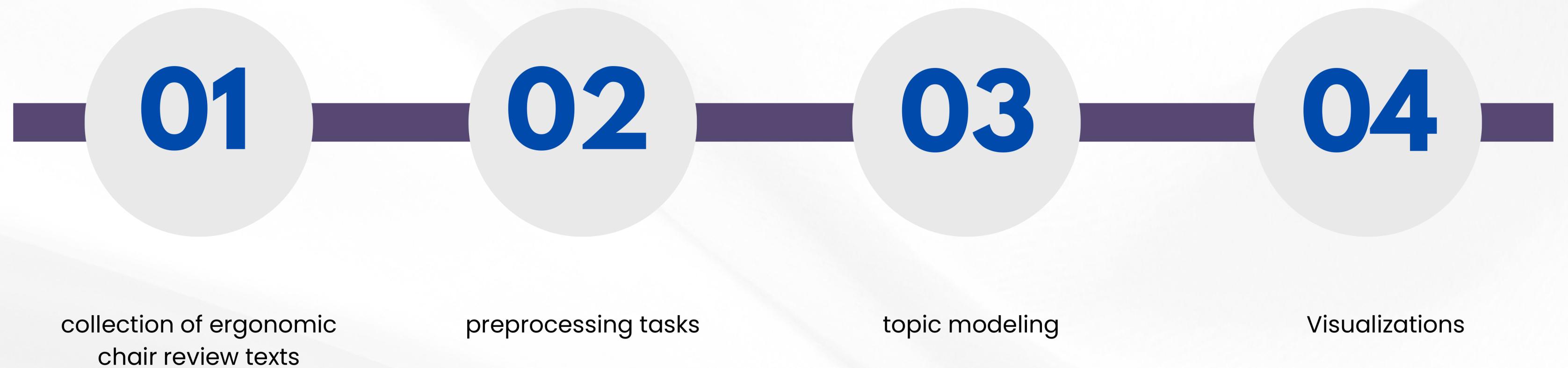


# INTRODUCTION

## Research Objectives

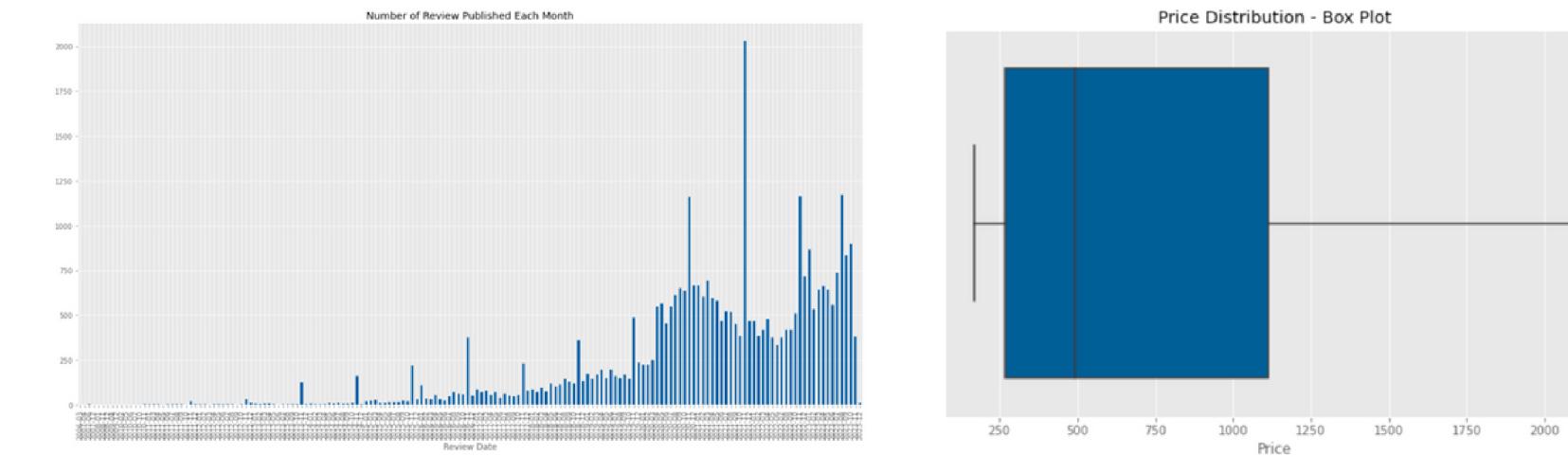
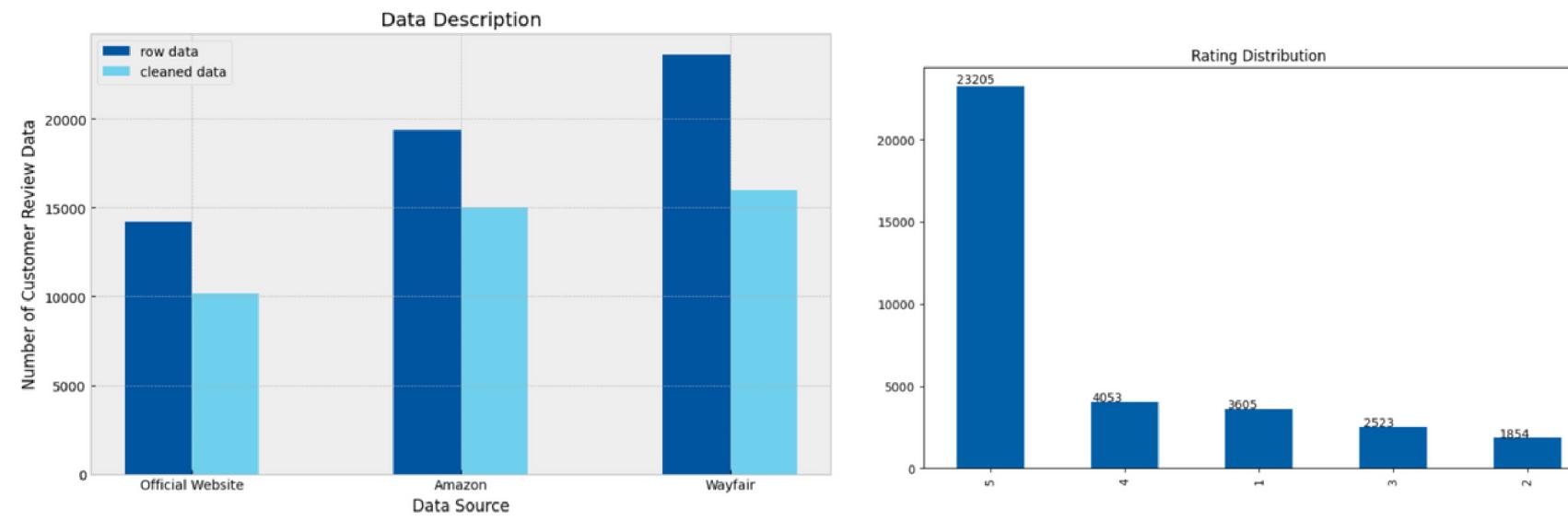
- 01** Create a robust **framework** that integrates the novel **NPFST** approach with established methods such as Word2Vec, HDBSCAN, MMR, and UMAP, enhancing the **BERTopic model**'s ability to extract and analyze marketing-related information from **review data**.
- 02** Introduce a methodological framework that can be adopted by other researchers and companies.
- 03** Showcase how topic modeling, particularly the Bertopic model, can be directly or indirectly employed to address a variety of marketing problems.
- 04** offer a detailed understanding of consumer behavior towards ergonomic chairs

# METHODOLOGY



# DATA COLLECTION

## Data retrieval



- March 2006 to December 2023.
- **51,225 reviews → 35,240 reviews**
- Price: 4 equal categories, each constituting 25% of the total products(250-400, 400-550, 550-1000, over 1000, USD)
- 65.8% : 5-star rating
- 10% : 1-star rating

index	Review Date	country	Rate	Price	Review
26701	2021-11-30	United States	4	735.0	Great for back supportMy back feels better but the seat isn't that comfy after 2 months of use.
33449	2022-03-29	United States	5	259.0	I'd give it 6-Stars if I could.I've used it everyday for over a year. I'm sitting in it now as I write this review. It is still comfortable and performing as advertised. Best thing I have bought off of the Zon for quite a while. I'd give it 6-satrs if I could
3499	2021-06-22	United States	3	229.99	So far, okay So far this chair isn't exactly what I was expecting. I love the adjustable armrests, but for me the lumbar support is too low (even adjusted on the highest setting) and pretty hard / uncomfortable on my back. Going to order an extra lumbar support pillow and hopefully that'll make it more comfortable. Overall not bad, but not the best.
4324	2021-12-03	United States	4	189.99	Great chair except for the armrests Ergonomically and seat-comfort wise I really like this chair. What I don't like is that the armrests are cupped/concave and not adjustable (they adjust vertically but the horizontal angle is fixed). If I'm sitting with my arms resting on the armrests perpendicular to my body it's ok, but when I bend my arms in to a typing position the armrests are uncomfortable. If the armrest were horizontally adjustable this chair would be nearly perfect.
7757	2023-04-09	United States	5	319.99	Love this chair! This chair is so comfortable! I love it. I was afraid it would stain easily and so far I've had no issues. I did buy rollerblade wheels for it as I feel the wheels on it didn't work great on the carpet in my office. I'm 5'4" and it's perfect for me.
8362	2023-06-29	United States	5	249.5	Outstanding Office Chair! Chair arrived quickly & was easy to assemble. Comfort is perfect for a tall person (6'4").Customer Service (Cayla) has been great in providing timely responses to product inquiries.Highly recommend!

# NATURE LANGUAGE PREPROCESSING

## Removal of Stop Words

index	Review Date	Review	Remove Stopword	Unigram Word	Review word count
26701	2021-11-30	Great for back supportMy back feels better but the seat isn't that comfy after 2 months of use.	back supportmy back feel better comfy 2 month use .	['back', 'supportmy', 'back', 'feel', 'comfy', 'month']	18
33449	2022-03-29	I'd give it 6-Stars if I could.I've used it everyday for over a year. I'm sitting in it now as I write this review. It is still comfortable and performing as advertised. Best thing I have bought off of the Zon for quite a while. I'd give it 6-satrs if I could	'd give 6-stars could.i used everyday year . 'm sitting write . still comfortable performing advertised . best bought zon quite .'d give 6-satrs could	['give', '6-stars', 'could', 'i', 'use', 'everyday', 'year', 'sit', 'write', 'still', 'comfortable', 'perform', 'advertised', 'best', 'quite', 'give', '6-satrs', 'could']	52
3499	2021-06-22	So far, okay So far this chair isn't exactly what I was expecting. I love the adjustable armrests, but for me the lumbar support is too low (even adjusted on the highest setting) and pretty hard / uncomfortable on my back. Going to order an extra lumbar support pillow and hopefully that'll make it more comfortable. Overall not bad, but not the best.	far , okay far exactly expecting . love adjustable armrest , lumbar support low ( even adjusted highest setting ) pretty hard / uncomfortable back . going extra lumbar support pillow hopefully 'll make comfortable . overall bad . best .	['okay', 'exactly', 'expect', 'love', 'adjustable', 'armrest', 'lumbar', 'support', 'even', 'adjust', 'highest', 'set', 'pretty', 'hard', 'uncomfortable', 'back', 'go', 'extra', 'lumbar', 'support', 'pillow', 'hopefully', 'make', 'comfortable', 'overall', 'best']	63
4324	2021-12-03	Great chair except for the armrests Ergonomically and seat-comfort wise I really like this chair. What I don't like is that the armrests are cupped/concave and not adjustable (they adjust vertically but the horizontal angle is fixed). If I'm sitting with my arms resting on the armrests perpendicular to my body it's ok, but when I bend my arms in to a typing position the armrests are uncomfortable. If the armrest were horizontally adjustable this chair would be nearly perfect.	except armrest ergonomically seat-comfort wise really . armrest cupped/concave adjustable ( adjust vertically horizontal angle fixed ) . 'm sitting arm resting armrest perpendicular body ok , bend arm typing position armrest uncomfortable . armrest horizontally adjustable nearly perfect .	['except', 'armrest', 'ergonomically', 'seat-comfort', 'wise', 'really', 'armrest', 'cupped/concave', 'adjustable', 'adjust', 'vertically', 'horizontal', 'angle', 'fix', 'sit', 'arm', 'rest', 'armrest', 'perpendicular', 'body', 'bend', 'arm', 'type', 'position', 'armrest', 'uncomfortable', 'armrest', 'horizontally', 'adjustable', 'nearly', 'perfect']	80
7757	2023-04-09	Love this chair! This chair is so comfortable! I love it. I was afraid it would stain easily and so far I've had no issues. I did buy rollerblade wheels for it as I feel the wheels on it didn't work great on the carpet in my office. I'm 5'4" and it's perfect for me.	love ! comfortable I love . afraid stain easily far ' issue , rollerblade wheel feel wheel ' work carpet office . ' 5 ' 4 " . perfect .	['love', 'comfortable', 'love', 'afraid', 'stain', 'easily', 'issue', 'rollerblade', 'wheel', 'feel', 'wheel', 'work', 'carpet', 'office', 'perfect']	55
8362	2023-06-29	Outstanding Office Chair! Chair arrived quickly & was easy to assemble. Comfort is perfect for a tall person (6'4"). Customer Service (Cayla) has been great in providing timely responses to product inquiries.Highly recommend!	outstanding office i arrived quickly & easy assemble . comfort perfect tall person ( 6 ' 4 " ) . customer service ( cayla ) providing timely response inquiries.highly recommend !	['outstanding', 'office', 'arrive', 'quickly', 'easy', 'assemble', 'comfort', 'perfect', 'tall', 'person', 'customer', 'service', 'cayla', 'provide', 'timely', 'response', 'inquiries', 'highly', 'recommend']	32

## unigram bag of words

- Remove punctuations
- Remove words smaller than 3 letters
- Remove stopwords
- Remove punctuation marks

## Part-of-Speech Tagging

index	Review Date	Review	After POS Tag
26701	2021-11-30	Great for back supportMy back feels better but the seat isn't that comfy after 2 months of use.	['supportmy', 'feel', 'comfy', 'month']
33449	2022-03-29	I'd give it 6-Stars if I could.I've used it everyday for over a year. I'm sitting in it now as I write this review. It is still comfortable and performing as advertised. Best thing I have bought off of the Zon for quite a while. I'd give it 6-satrs if I could	['6-stars', 'could', 'i', 'use', 'everyday', 'year', 'sit', 'comfortable', 'perform', 'give', '6-satrs']
3499	2021-06-22	So far, okay So far this chair isn't exactly what I was expecting. I love the adjustable armrests, but for me the lumbar support is too low (even adjusted on the highest setting) and pretty hard / uncomfortable on my back. Going to order an extra lumbar support pillow and hopefully that'll make it more comfortable. Overall not bad, but not the best.	['adjustable', 'armrest', 'lumbar', 'support', 'hard', 'uncomfortable', 'extra', 'lumbar', 'support', 'comfortable', 'overall']
4324	2021-12-03	Great chair except for the armrests Ergonomically and seat-comfort wise I really like this chair. What I don't like is that the armrests are cupped/concave and not adjustable (they adjust vertically but the horizontal angle is fixed). If I'm sitting with my arms resting on the armrests perpendicular to my body it's ok, but when I bend my arms in to a typing position the armrests are uncomfortable. If the armrest were horizontally adjustable this chair would be nearly perfect.	['seat-comfort', 'wise', 'armrest', 'cupped/concave', 'adjustable', 'adjust', 'horizontal', 'angle', 'fix', 'sit', 'arm', 'rest', 'perpendicular', 'body', 'arm', 'type', 'position', 'uncomfortable', 'adjustable', 'perfect']
7757	2023-04-09	Love this chair! This chair is so comfortable! I love it. I was afraid it would stain easily and so far I've had no issues. I did buy rollerblade wheels for it as I feel the wheels on it didn't work great on the carpet in my office. I'm 5'4" and it's perfect for me.	['love', 'comfortable', 'love', 'afraid', 'rollerblade', 'wheel', 'wheel', 'work', 'carpet', 'office', 'perfect']
8362	2023-06-29	Outstanding Office Chair! Chair arrived quickly & was easy to assemble. Comfort is perfect for a tall person (6'4"). Customer Service (Cayla) has been great in providing timely responses to product inquiries.Highly recommend!	['outstanding', 'office', 'arrive', 'quickly', 'easy', 'assemble', 'comfort', 'tall', 'person', 'service', 'cayla', 'response', 'inquiries', 'highly', 'recommend']

- nouns (both singular and plural, denoted as "NN" and "NNS," respectively)
- adjectives (represented as "JJ") are retained.

# NATURE LANGUAGE PREPROCESSING

## Feature extraction

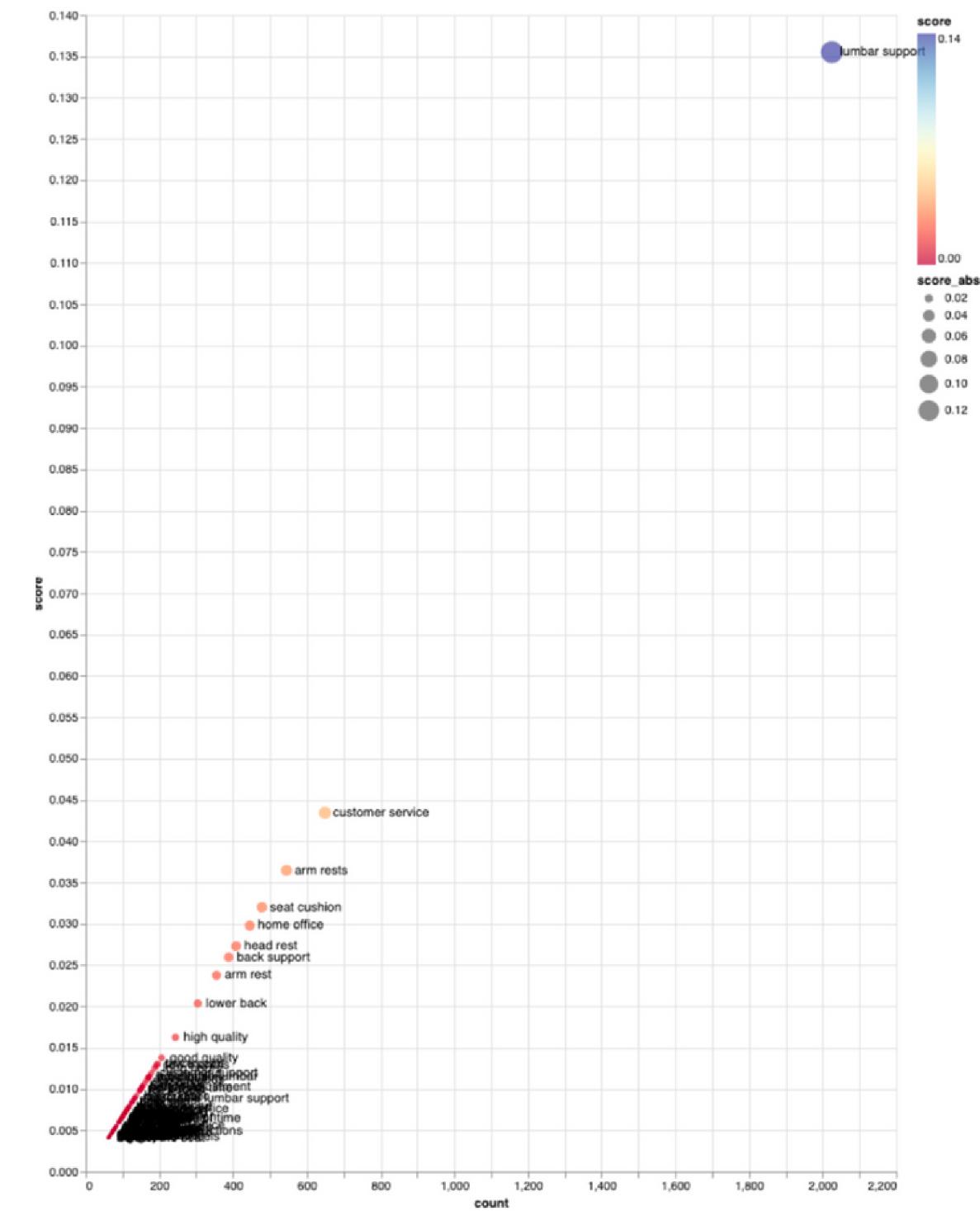
Figurine of a dog, already in the early dynastic period, egyptians deposited faience figurines of wild animals in temple precincts. These figurines were reintroduced in the twelfth dynasty, but as a component of burial equipment and with new species added to the repertoire. The controlled representation of desert animals may have assured the Egyptians of eternal safety, though they also likely had symbolic meanings. The dog may represent a personal wish for companionship.

NPFST

figurine of a dog, [○ ○ ○]  
early dynamics, [○ ○ ○]  
faience figurines, [○ ○ ○]  
Wild animals, [○ ○ ○]  
Burial equipment [○ ○ ○]  
... ...

- Novel Phrase Finding and Segmentation Tool (NPFST)
- It transforms text into fixed-length vectors by counting how many times each word appears in a document. It ignores the order of words, grammar, and even punctuation, treating every word as equally important in a "**bag**."

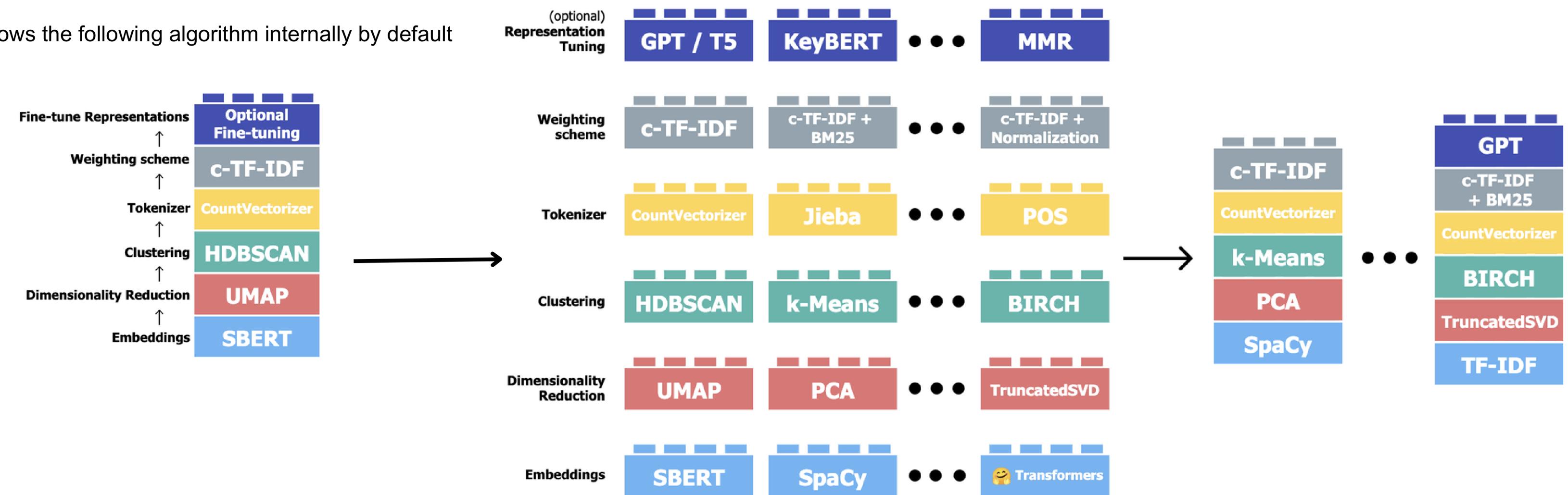
### Why Noun Phrases?



# MODEL SETUP

## Bertopic Overview

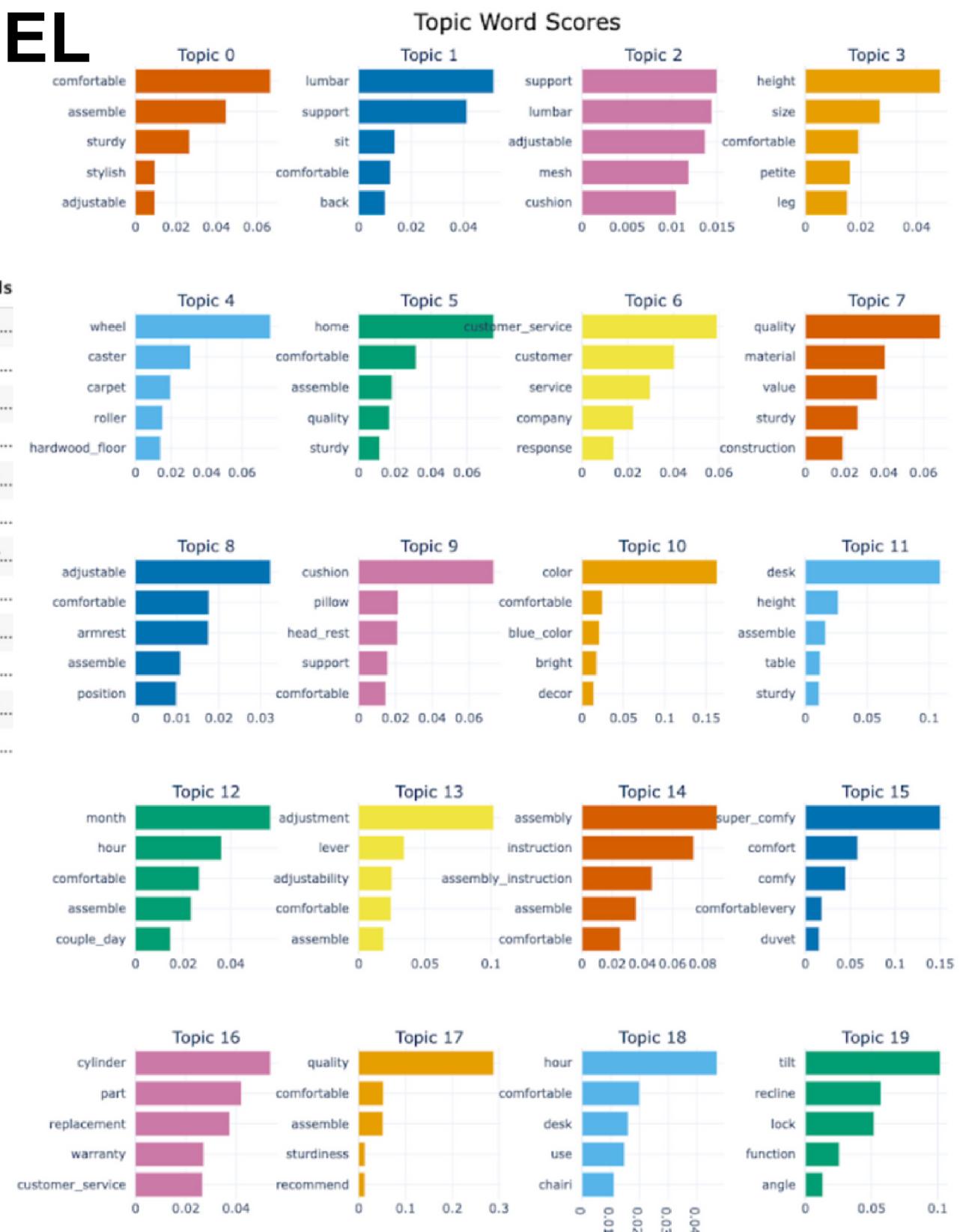
BERTopic follows the following algorithm internally by default



# RESULTS AND DISCUSSION OF TRAINED MODEL

## Topic Representation

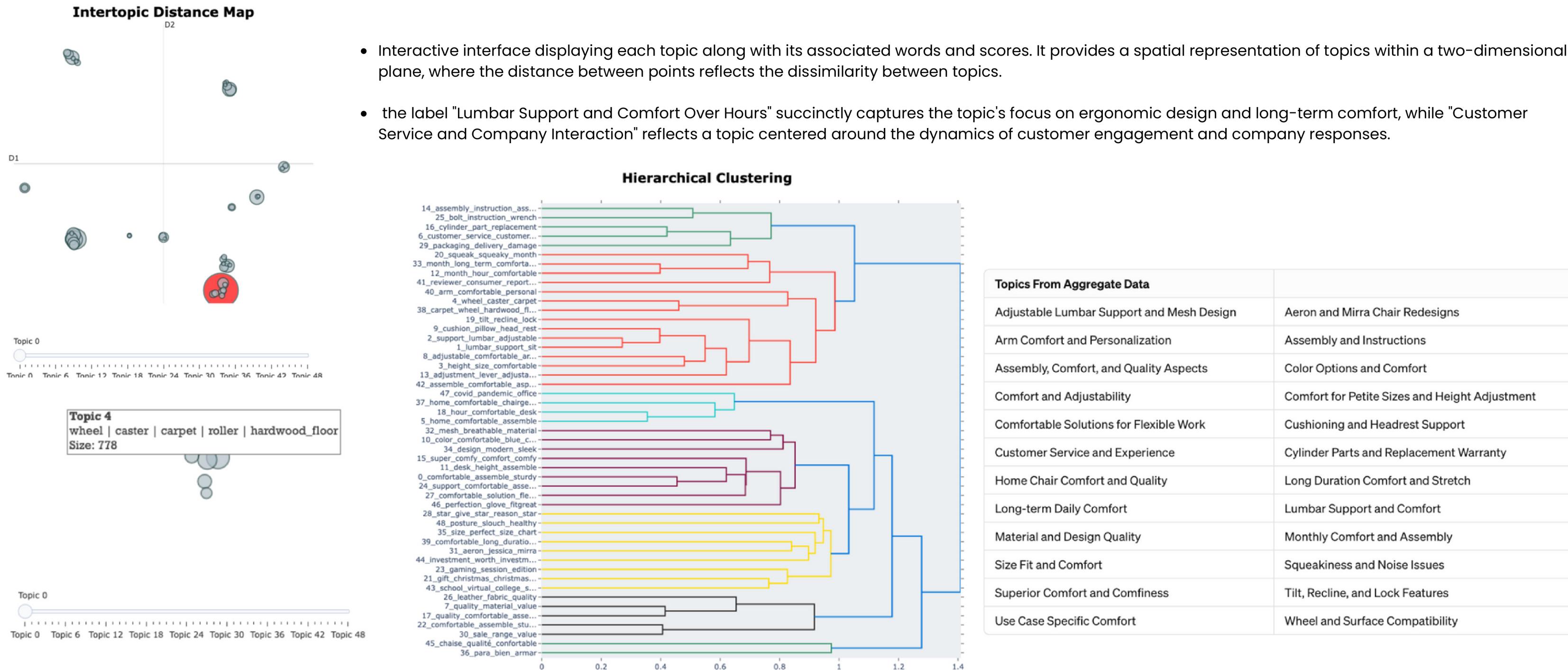
Topic	Name	Representation	Representative_Docs	Top_n_words
-1	-1_support_comfortable_hour_lumbar	['support', 'comfortable', 'hour', 'lumbar', ...]	["[uncomfortable", "lumbar", "support", "hop"...]	support - comfortable - hour - lumbar - desk - ...
0	0_comfortable_assemble_sturdy_stylish	['comfortable', 'assemble', 'sturdy', 'stylish...]	["[easy", "assemble", "comfortable]", "[lov...]	comfortable - assemble - sturdy - stylish - ad...
1	1_lumbar_support_sit_comfortable	['lumbar', 'support', 'sit', 'comfortable', 'b...]	["[lumbar", "support", "lumbar", "support", "...]	lumbar - support - sit - comfortable - back - ...
2	2_support_lumbar_adjustable_mesh	['support', 'lumbar', 'adjustable', 'mesh', 'c...]	["[embody", "investment", "pro", "disappear",...]	support - lumbar - adjustable - mesh - cushion...
3	3_height_size_comfortable_petite	['height', 'size', 'comfortable', 'petite', 'l...]	["[comfortable", "tall", "person", "love"], ...]	height - size - comfortable - petite - leg - f...
4	4_wheel_caster_carpet_roller	['wheel', 'caster', 'carpet', 'roller', 'hardw...]	["[price", "assemble", "perfect", "short", "...]	wheel - caster - carpet - roller - hardwood_ fl...
5	5_home_comfortable_assemble_quality	['home', 'comfortable', 'assemble', 'quality',...]	["[home", "office]", "[home", "office]", "...]	home - comfortable - assemble - quality - stur...
6	6_customer_service_customer_service_company	['customer_service', 'customer', 'service', 'c...]	["[customer_service", "easy", "assemble]", "...]	customer_service - customer - service - compan...
7	7_quality_material_value_sturdy	['quality', 'material', 'value', 'sturdy', 'co...]	["[quality", "price]", "[quality", "price"]...]	quality - material - value - sturdy - construc...
8	8_adjustable_comfortable_armrest_assemble	['adjustable', 'comfortable', 'armrest', 'asse...]	["[rest]", "[comfortable", "love", "rest", ...]	adjustable - comfortable - armrest - assemble ...
9	9_cushion_pillow_head_rest_support	['cushion', 'pillow', 'head_rest', 'support', ...]	["[cushion]", "[cushion]", "[cushion", 'c...	cushion - pillow - head_rest - support - comfo...
10	10_color_comfortable_blue_color_bright	['color', 'comfortable', 'blue_color', 'bright...]	["[nice", "color", "comfortable", "happy"], ...]	color - comfortable - blue_color - bright - de...



- **bar charts out of the c-TF-IDF scores for each topic representation.**
- The descending order of word scores within each topic unveils the relative weight and significance of individual terms, providing a gradation of relevance that further clarifies the thematic structure. Terms with lower scores, while still relevant, may not define the topic's core as strongly.

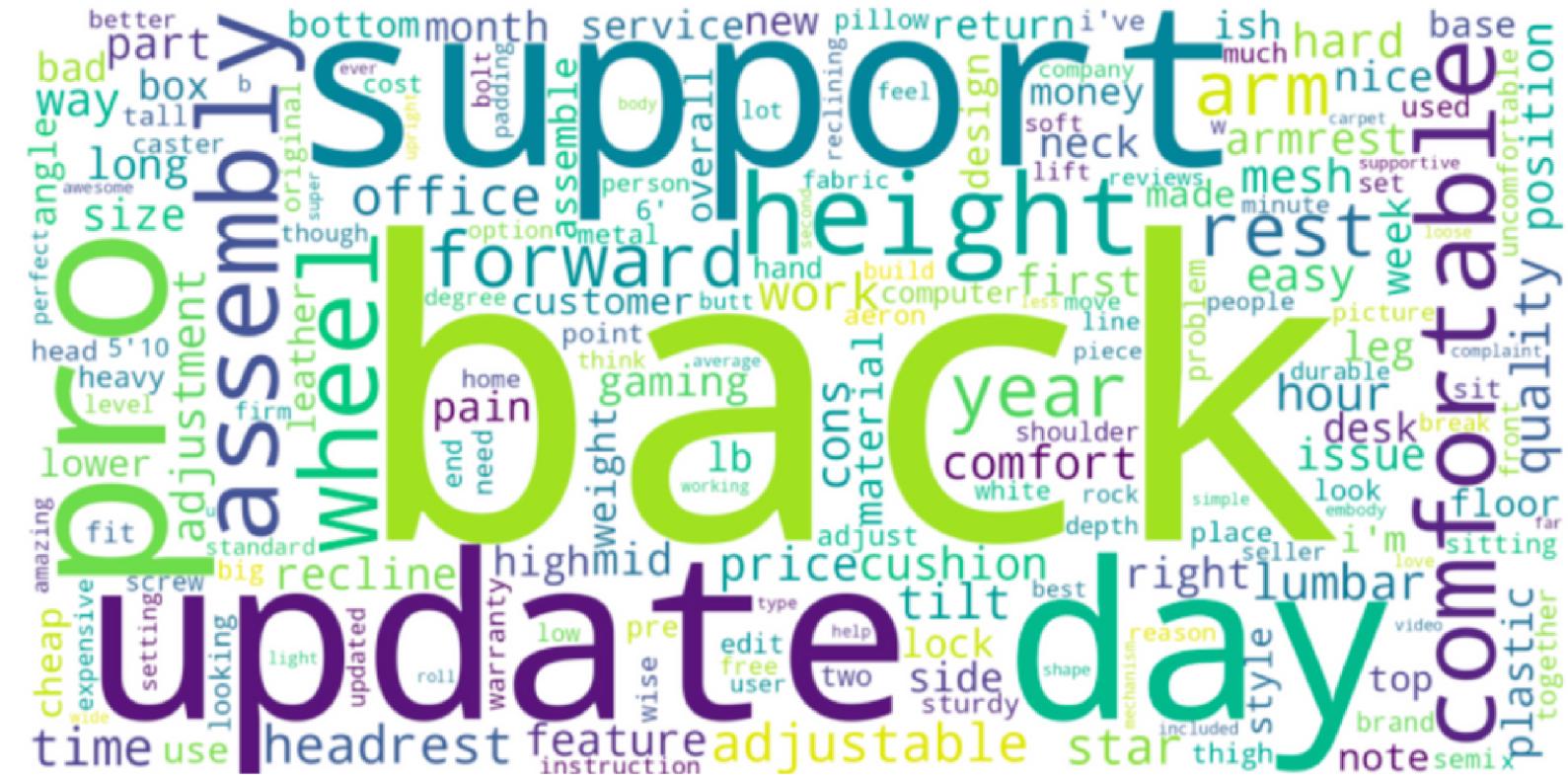
# RESULTS AND DISCUSSION OF TRAINED MODEL

## Topic Representation



# SUPPLEMENTARY ANALYSIS

# Word/Phrase Analysis



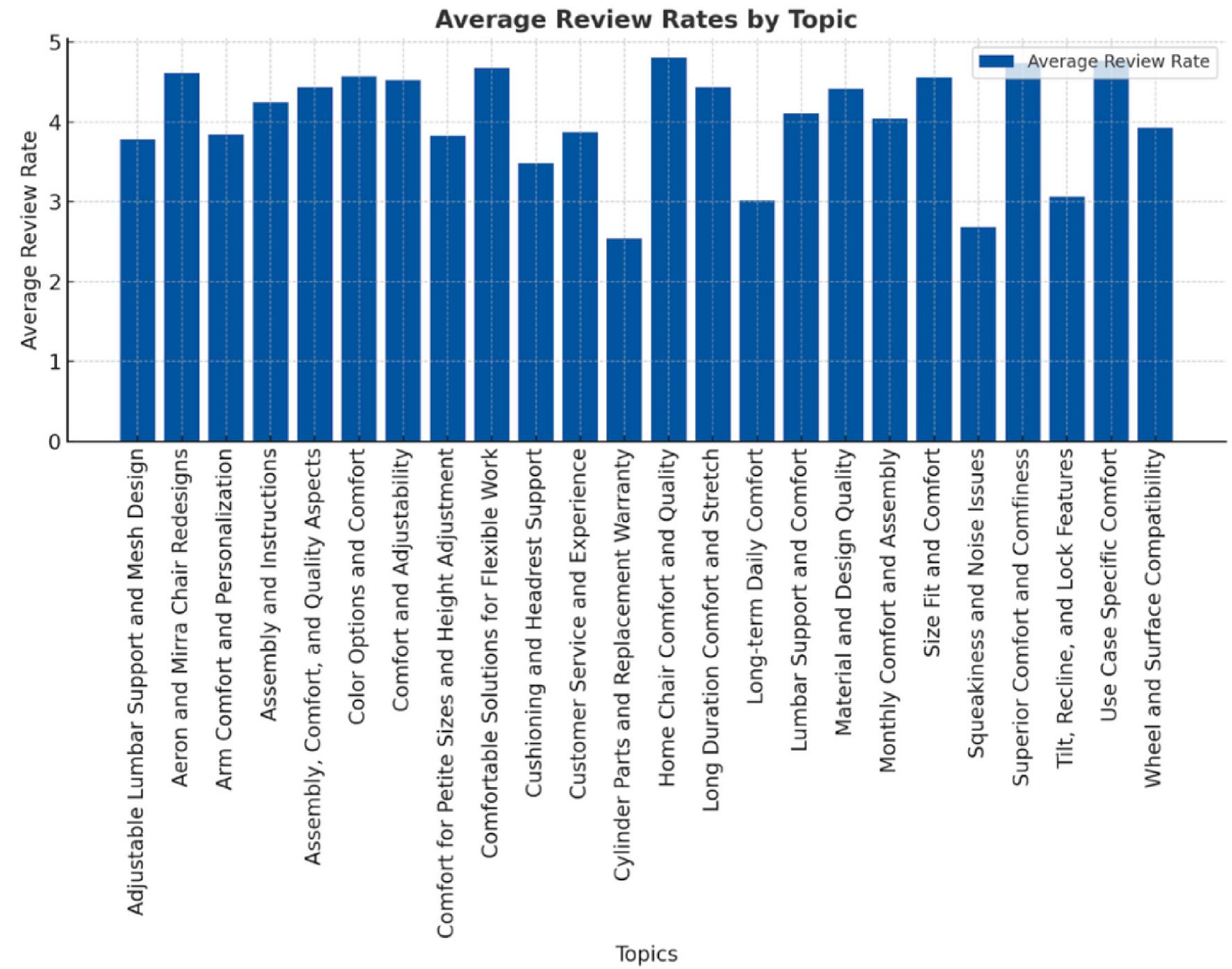
- underscore the critical importance of ergonomic features in customer satisfaction
  - customers highly value straightforward assembly processes.
  - marketing communications should emphasize the brand's commitment to superior customer service.
  - "head rest," "long\_period" of use, and "worth\_money." Creating content that addresses these topics, from blog posts to buying guides, can improve search engine optimization (SEO) and engagement by aligning with customer interests.
  - "office chair," 'tall\_people,' should be incorporated into website content, product descriptions, and blog posts to improve search engine visibility and attract organic traffic.

	NPFST	Frequency
2	lumbar support	2025
3	office chair	1591
4	great chair	903
5	customer service	649
6	herman miller	546
7	arm rests	545
8	comfortable chair	511
9	seat cushion	478
10	good chair	460
11	home office	445
12	head rest	408
13	office chairs	390
14	back support	388
15	arm rest	355
16	lower back	304
17	back of the chair	286
18	new chair	280
19	best chair	262
20	nice chair	257
21	desk chair	254
22	high quality	243
23	ergonomic chair	241
24	other chairs	236
25	old chair	225
26	good quality	206
27	price point	195
28	few months	192
29	long hours	187
30	quality chair	186
31	customer support	178

# SUPPLEMENTARY ANALYSIS

## Decoding Customer Satisfaction

- Manufacturers might consider a series of enhancements to improve product satisfaction.
- an upgrade in the design and warranty of the cylinder
- Offering an extended warranty for these components could reassure customers about the product's durability and quality.



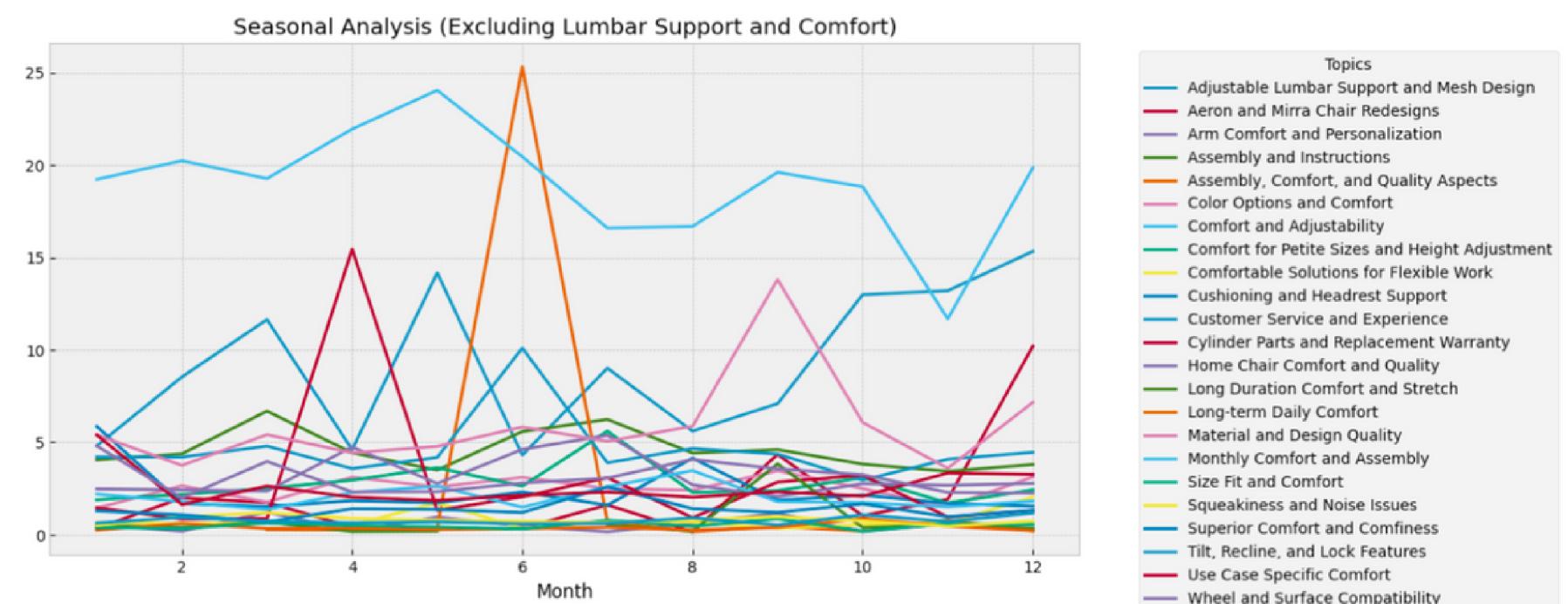
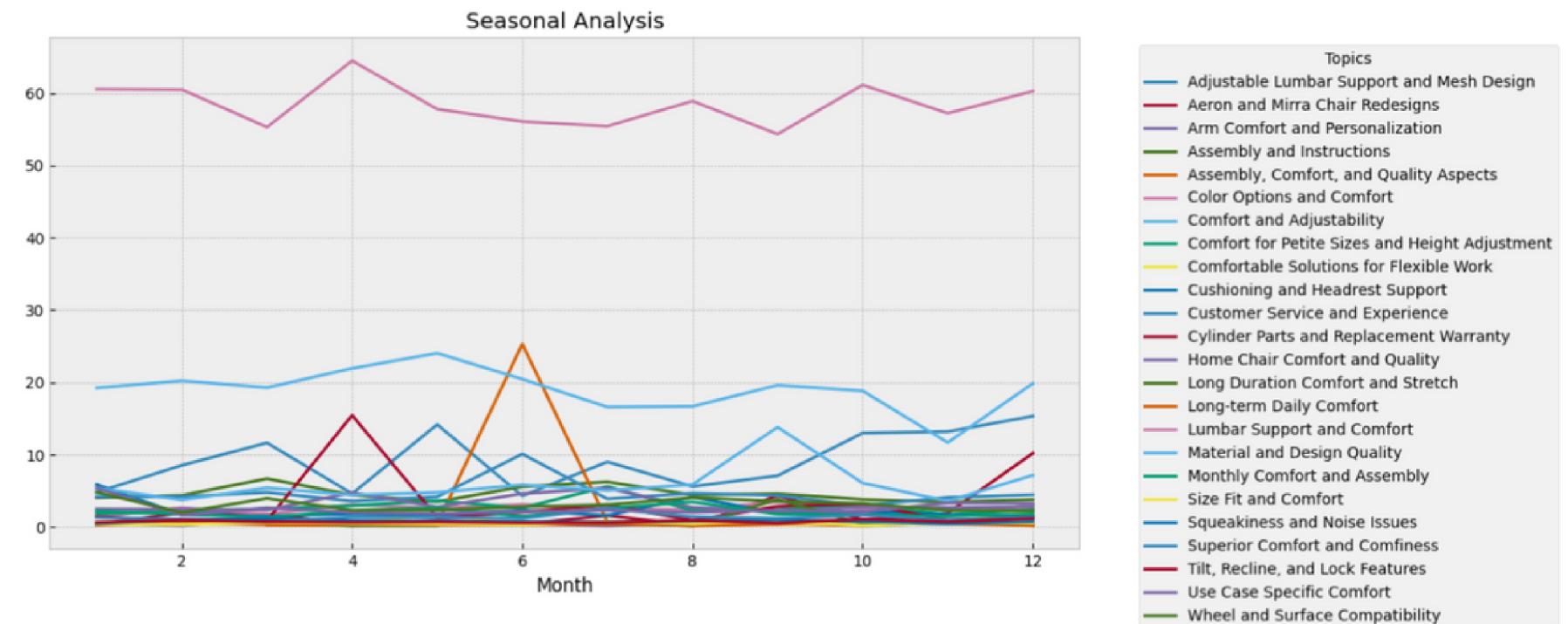
# SUPPLEMENTARY ANALYSIS

## Seasonal Analysis

### Investigate repeating patterns over time

- aggregated reviews by month (the average frequency of topics appearing in the same month across different years.)
- the cumulative topics represent 100% of the reviews for each month.

orange line representing the topic "Assembly, Comfort, and Quality Aspects" --a lack of diversity in the sample sources. However, on a general trend, September, October, and March are periods with high volumes of customer comments.



# SUPPLEMENTARY ANALYSIS

## Trending Recognition

**moving average of different topics over 6 months (from June 2023 to December 2023)**

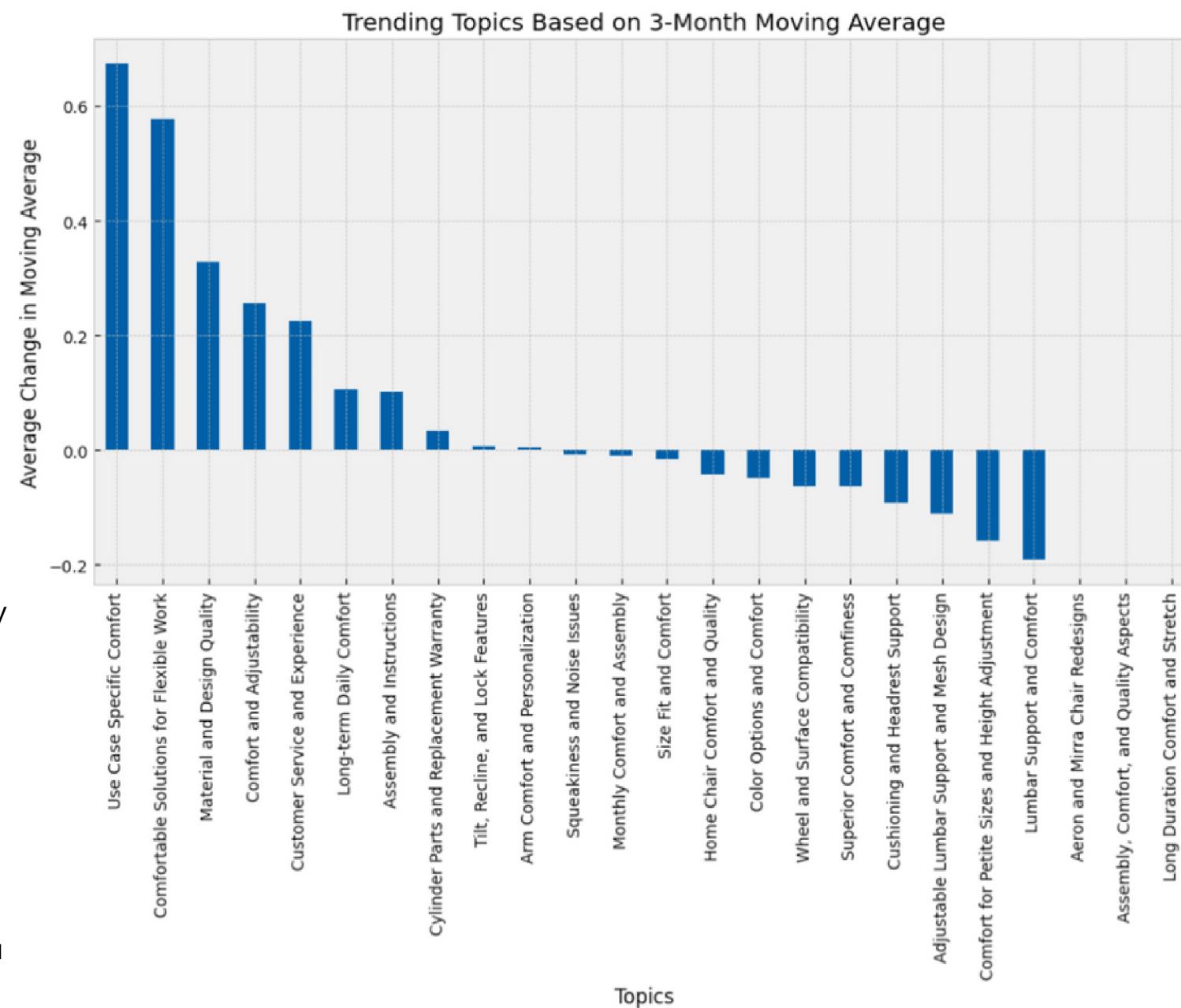
- the difference between each period's moving average and its predecessor with the latest 6 periods
- the average of these differences for each topic.

'1,2 rise, influenced by the growing trend of remote work and home offices. A potential driver for this sentiment could be the COVID-19 pandemic, which forced many individuals to adapt their living spaces into functional work areas, thus prioritizing comfort and flexibility in product design.

3 rise, build quality and aesthetic appeal of products. In a market saturated with options, it is plausible that consumers are becoming more discerning, preferring products that offer durability and a pleasing design.

This shift could be driven by a consumer base that is increasingly willing to invest in higher-quality goods that promise longevity and a timeless appeal.

Conversely, topics such as 'Lumbar Support and Comfort' and 'Comfort for Petite Sizes and Height Adjustment' are trending **negatively**. This downward trend could suggest either a market **saturation** for these features or a possible indication of consumers finding the current offerings inadequate. It might also reflect a shift in marketing focus away from these features, leading to a decline in consumer discussions or sentiment.



# FUTURE SCOPE AND CONCLUSION

## Conclusions

### METHODOLOGY

Finetune Bertopic model for the task of topic recognition and analysis with natural language processing techniques

### INTUITION

This model completely analyses reviews from ergonomics chair industry.

### CONTRIBUTION

Introduced a novel framework designed to extract marketing-related insights from review data.

### EXTENSION

The final model is in a usable state for individual use, anthology development, or implementation into a more complex piece of software

## FUTURE SCOPE AND CONCLUSION

### Future work suggestions

- 01** Focus on developing metrics for evaluating the effectiveness of these methodologies, alongside comparisons across different methods and embedding models.
- 02** Explore strategies for managing imbalanced datasets, enhancing data quality for underrepresented categories.
- 03** Expand upon this foundation by investigating alternative data sources, enhancing text preprocessing methods, and applying the framework to various industries.

# REFERENCE

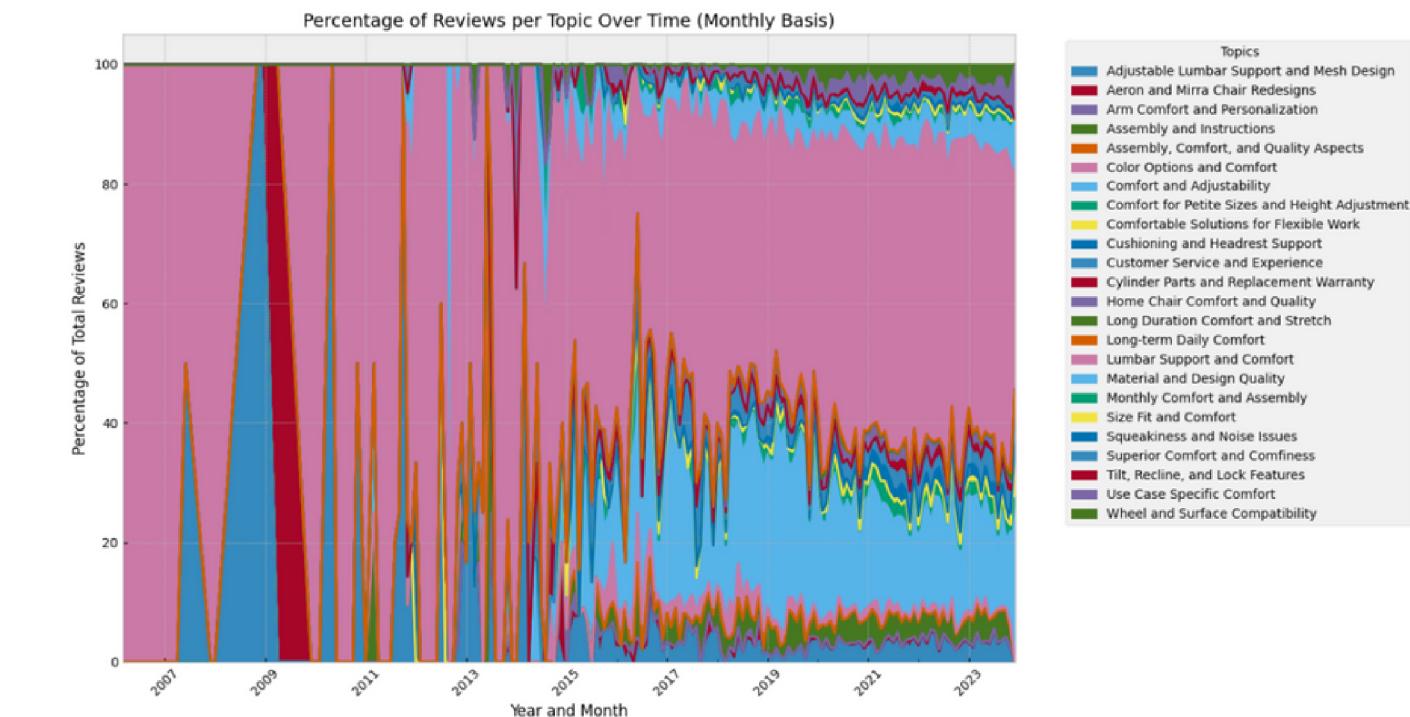
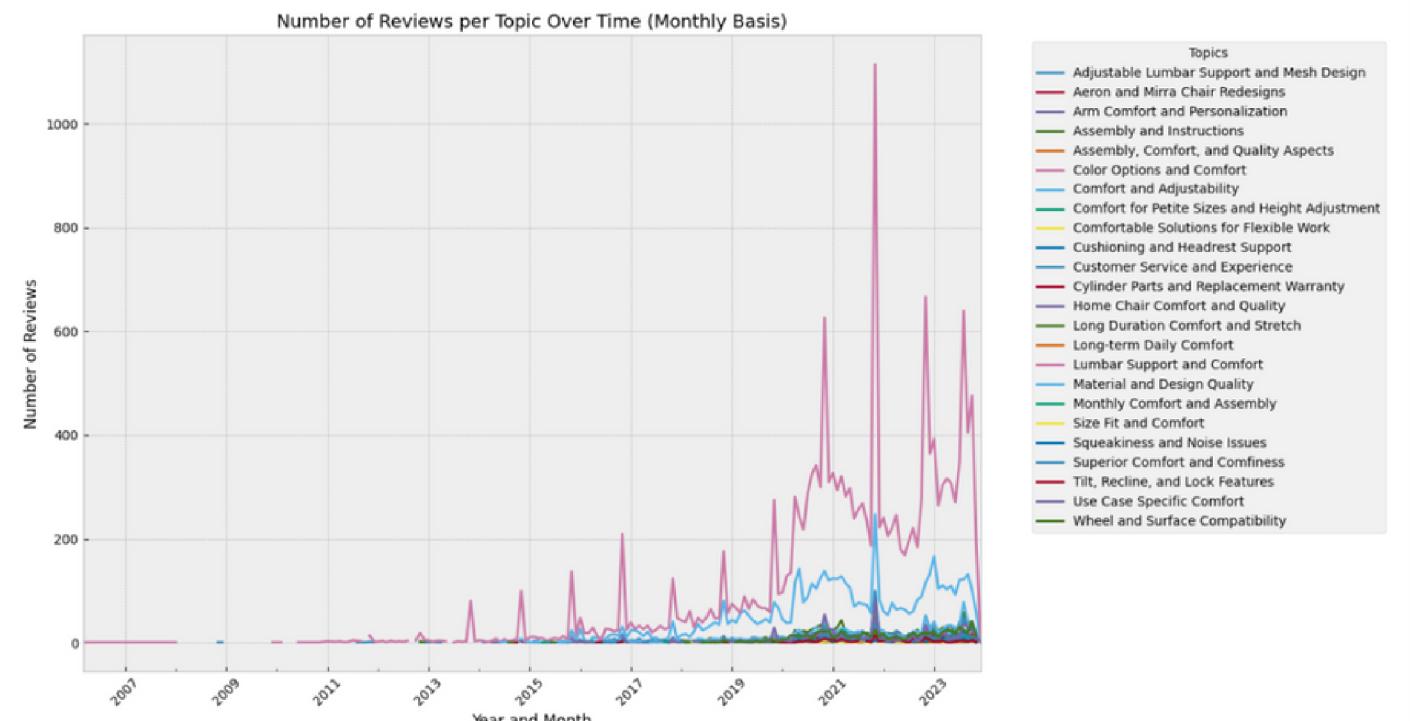
- Wang, Z., Chen, J., Chen, J., et al. (2023). "Identifying interdisciplinary topics and their evolution based on BERTopic". *Scientometrics*
- Yazıcı, G., Ozansoy Çadırcı, T. (2023). "Creating meaningful insights from customer reviews: a methodological comparison of topic modeling algorithms and their use in marketing research". *Journal of Marketing Analysis*
- Mosteller, J.R., Mathwick, C. (2016). "Online reviewer engagement: A typology based on reviewer motivation". *Journal of Service Research*
- An, Y., Oh, H., Lee, J. (2023). "Marketing Insights from Reviews Using Topic Modeling with BERTopic and Deep Clustering Network". *Applied Sciences*, 13(16), 9443
- Alvargonzález, D. (2011). "Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences". *International Studies in the Philosophy of Science*, 25(4), 387–403
- Callon, M., Courtial, J.P., Turner, W.A., Bauin, S. (1983). "From translations to problematic networks: An introduction to co-word analysis". *Social Science Information*, 22(2), 191–235
- Chen, B., Tsutsui, S., Ding, Y., Ma, F. (2017). "Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval". *Journal of Informetrics*, 11(4), 1175–1189
- Dong, K., Xu, H., Luo, R., Wei, L., Fang, S. (2018). "An integrated method for interdisciplinary topic identification and prediction: A case study on information science and library science". *Scientometrics*, 115, 849–868
- Grootendorst, M. (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". *arXiv preprint arXiv:2203.05794*
- Hall, D., Jurafsky, D., Manning, C.D. (2008). "Studying the history of ideas using topic models". In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 363–371
- Jiang, L., Zhang, T., Huang, T. (2022). "Empirical research of hot topic recognition and its evolution path method for scientific and technological literature". *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 26(3), 299–308
- Leydesdorff, L., Hellsten, I. (2006). "Measuring the meaning of words in contexts: An automated analysis of controversies about 'monarch butterflies', 'frankenfoods', and 'stem cells'". *Scientometrics*, 67(2), 231–258
- Leydesdorff, L., Ismael, R. (2011). "Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations". *Journal of Informetrics*, 5(1), 87–100
- Leydesdorff, L., Wagner, C.S., Bornmann, L. (2019). "Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient". *Journal of Informetrics*, 13(1), 255–269
- Li, M. (2017). "An exploration to visualize the emerging trends of technology foresight based on an improved technique of co-word analysis and relevant literature data of WOS". *Technology Analysis & Strategic Management*, 29(6), 655–671
- Li, J. (2014). "The concept and measurement of interdisciplinarity". *Documentation, Information & Knowledge*, 3, 87–93
- Aggarwal, C.C., Hinneburg, A., Keim, D.A. (2001). "On the surprising behavior of distance metrics in high dimensional space". *Proceedings of the International Conference on Database Theory*, pp. 420–434. Springer
- Allaoui, M., Kherfi, M.L., Cheriet, A. (2020). "Considerably improving clustering algorithms using UMAP dimensionality reduction technique: A comparative study". *International Conference on Image and Signal Processing*, pp. 317–325. Springer
- Angelov, D. (2020). "Top2vec: Distributed representations of topics". *arXiv preprint arXiv:2008.09470*
- Baturo, A., Dasandi, N., Mikhaylov, S.J. (2017). "Understanding state preferences with text as data: Introducing the UN General Debate Corpus". *\*Research & Politics\**, 4(2), 2053168017712821
- Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U. (1999). "When is "nearest neighbor" meaningful?". *International Conference on Database Theory*, pp. 217–235. Springer
- Bianchi, F., Terragni, S., Hovy, D. (2020a). "Pre-training is a hot topic: Contextualized document embeddings improve topic coherence". *arXiv preprint arXiv:2004.03974*
- Bianchi, F., Terragni, S., Hovy, D., Nozza, D., Fersini, E. (2020b). "Cross-lingual contextualized topic models with zero-shot learning". *arXiv preprint arXiv:2004.07737*
- Blei, D.M., Lafferty, J.D. (2006). "Dynamic topic models". In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 113–120
- Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). "Latent Dirichlet allocation". *Journal of Machine Learning Research*, 3, 993–1022
- Bouma, G. (2009). "Normalized (pointwise) mutual information in collocation extraction". *Proceedings of GSCL*, 30, 31–40
- Anoop, V.S., Ashraf, S. (2017). "A topic modeling guided approach for semantic knowledge discovery in e-commerce". *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4), 40–47
- Kyeong, J., Jiyo, K., Lee, J.W., Lee, Y., Lee, S.B. (2021). "Text Mining Analysis of Consumer Perception of Food Distribution Platforms: Focusing on Topic Modeling". *Journal of Foodservice Management*, 24, 71–100
- Bumjun, L., Heekyung, N. (2020). "Food tourism market segmentation approach using topic modeling analysis: Focusing on benefits sought". *Korean Journal of Hospitality and Tourism*, 29, 187–204
- Cho, M.-K., Lee, B.-J. (2021). "Comparison of service quality of full-service carriers in Korea using topic modeling: Based on reviews from TripAdvisor". *Journal of Hospitality and Tourism Studies*, 23, 152–165
- An, Y., Oh, H., Lee, J. (2023). "Marketing Insights from Reviews Using Topic Modeling with BERTopic and Deep Clustering Network". *Applied Sciences*, 13(16), 9443
- Lloyd, S. (1982). "Least squares quantization in PCM". *IEEE Transactions on Information Theory*, 28(2), 129–137
- Campello, R.J.G.B., Moulavi, D., Sander, J. (2013). "Density-Based Clustering Based on Hierarchical Density Estimates". *Advances in Knowledge Discovery and Data Mining*. PAKDD 2013. Lecture Notes in Computer Science, vol 7819. Springer, Berlin, Heidelberg
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise". *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Vol. 96, 226–231
- Xie, J., Girshick, R., Farhadi, A. (2016). "Unsupervised deep embedding for clustering analysis". *International Conference on Machine Learning*, pp. 478–487
- Yang, B., Fu, X., Sidiroopoulos, N.D., Hong, M. (2017). "Towards k-means-friendly spaces: Simultaneous deep learning and clustering". *International Conference on Machine Learning*, pp. 3861–3870
- Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). "Latent Dirichlet allocation". *Journal of Machine Learning Research*, 3(Jan), 993–1022
- Grootendorst, M. (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". *arXiv preprint arXiv:2203.05794*
- Reimers, N., Gurevych, I. (2019). "SentenceBERT: Sentence embeddings using siamese BERT networks". *arXiv preprint arXiv:1908.10084*
- Blei, D.M., Lafferty, J.D. (2006). "Dynamic topic models". *Proceedings of the 23rd International Conference on Machine Learning*, pp. 113–120
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W. (2021). "Informer: Beyond efficient transformer for long sequence time-series forecasting". In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 11106–11115
- Wu, H., Xu, J., Wang, J., Long, M. (2021). "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting". *Advances in Neural Information Processing Systems*
- Caron, M., Müller, O. (2020). "Hardening Soft Information: A Transformer-Based Approach to Forecasting Stock Return Volatility". *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, pp. 4383–4391
- Anoop, V.S., Ashraf, S. (2017). "A topic modeling guided approach for semantic knowledge discovery in e-commerce". *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4), 40–47
- Li, W., Yin, J., Chen, H. (2018). "Supervised Topic Modeling Using Hierarchical Dirichlet Process-Based Inverse Regression: Experiments on E-Commerce Applications". *IEEE Transactions on Knowledge and Data Engineering*, 30(6), 1192–1205
- Chen, R., Xu, W. (2017). "The determinants of online customer ratings: a combined domain ontology and topic text analytics approach". *Electronic Commerce Research*, 17, 31–50
- Dong, L.Y., Ji, S.J., Zhang, C.J., Zhang, Q., Chiu, D.W., Qiu, L.Q., Li, D. (2018). "An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews". *Expert Systems with Applications*, 114, 210–223
- Hajek, P., Hikkerova, L., Sahut, J.M. (2023). "Fake review detection in e-Commerce platforms using aspect-based sentiment analysis". *Journal of Business Research*, 167, 114143
- Ye, X., Lian, Z., She, B., Kudva, S. (2020). "Spatial and big data analytics of E-market transactions in China". *GeoJournal*, 85, 329–341
- Hong, W., Zheng, C., Wu, L., Pu, X. (2019). "Analyzing the relationship between consumer satisfaction and fresh e-commerce logistics service using text mining techniques". *Sustainability*, 11(13), 3570
- Wang, Y.D., Jiang, B.T., Ye, X.Y. (2016). "A method for studying the development pattern of urban commercial service facilities based on customer reviews from social media". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 577–578
- Kyeong, J., Jiyo, K., Lee, J.W., Lee, Y., Lee, S.B. (2021). "Text Mining Analysis of Consumer Perception of Food Distribution Platforms: Focusing on Topic Modeling". *Journal of Foodservice Management*, 24, 71–100
- Bumjun, L., Heekyung, N. (2020). "Food tourism market segmentation approach using topic modeling analysis: Focusing on benefits sought". *Korean Journal of Hospitality and Tourism*, 29, 187–204
- Soyeon, L., Yeongok, K. (2022). "Analysis of Apartment Interior Trend Using Topic Modeling: Focusing on 'Today's House' Review Data". *Proceedings of the KMIS 2022: 14th International Conference on Knowledge Management and Information Systems*, Valletta, Malta, pp. 141–149
- Cho, M.-K., Lee, B.-J. (2021). "Comparison of service quality of full-service carriers in Korea using topic modeling: Based on reviews from TripAdvisor". *Journal of Hospitality and Tourism Studies*, 23, 152–165
- Ekambaram, V., Manglik, K., Mukherjee, S., Sajja, S.S.K., Dwivedi, S., Raykar, V. (2020). "Attention based multi-modal new product sales time-series forecasting". *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3110–3118
- Kudo, T. (2006). "MeCab: Yet another part-of-speech and morphological analyzer". <https://sourceforge.net/projects/mecab/>
- Sentence Transformers. "all-mpnet-base-v2." Hugging Face, n.d., <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>.
- Lee, J. (2020). "KcBERT: Korean Comments BERT". *Proceedings of the 32nd Annual Conference on Human and Cognitive Language Technology*, pp. 437–440
- Derrick, T.R., Bates, B.T., Dufek, J.S. (1994). "Evaluation of time-series data sets using the Pearson product-moment correlation coefficient". *Medicine and Science in Sports and Exercise*, 26(7), 919–928
- Bouma, G. (2009). "Normalized (pointwise) mutual information in collocation extraction". *Proceedings of GSCL*, 30, 31–40
- Dieng, A.B., Ruiz, F.J.R., Blei, D.M. (2020). "Topic modeling in embedding spaces". *Transactions of the Association for Computational Linguistics*, 8, 439–453

# APPENDIX

## Monthly Topic Frequency

The x-axis represents time, spanning several years, while the y-axis represents the number of reviews. This chart can provide insights into trends over time, highlighting when certain features of ergonomic chairs became hot topics among customers and when they declined in discussion.

stacked area chart that traces the evolution of customer feedback on different product-related topics over time. Similar to the first plot but this time pivoting the “percentage” values for plotting to visualize the percentage of total monthly reviews that each topic represents over time.



# APPENDIX

## Embedding Model

This study used a hugging face open source sentence transformer pretrained model, particularly the "all-mpnet-base-v2" version, which represents a significant advancement in generating sentence embeddings.

The encoding text strings here used NPFST noun phrases. NPFST focuses on the most relevant aspects of the reviews for topic analysis, and reduces the noise in the data by omitting less informative parts of the text, such as filler words or off-topic comments.

## Reduce dimension/Cluster embedding

The tailored selection of UMAP and HDBSCAN parameters for analyzing ergonomics chair reviews demonstrates a strategic approach to capturing both detailed and broad trends in customer feedback.

On the clustering side, HDBSCAN's parameters are chosen to complement UMAP's data preparation by focusing on significant patterns through `min\_cluster\_size` and `metric='euclidean'`, ensuring that the clustering is both meaningful and interpretable.

## Representation Model

The study set a MMR diversity value of 0.3, the goal is to ensure the topics or summaries generated cover different aspects of the ergonomics chair reviews, from comfort and design to durability and customer service, providing a holistic view of customer sentiment.

## Model Fitting

```
topic_model_5 = BERTopic(  
    embedding_model=SentenceTransformer("sentence-transformers/all-mpnet-base-v2")  
  
    umap_model=UMAP(n_neighbors=15,  
                     n_components=5,  
                     min_dist=0.0,  
                     metric='cosine',  
                     random_state=42)  
  
    hdbscan_model=HDBSCAN(min_cluster_size=50,  
                           metric='euclidean',  
                           cluster_selection_method='eom',  
                           prediction_data=True)  
  
    representation_model=MaximalMarginalRelevance(diversity=0.3)  
    top_n_words=50,  
    verbose=True,  
    nr_topics=50  
)  
  
embeddings = sentence_model.encode(text, show_progress_bar=True)  
  
topics, probs = topic_model_5.fit_transform(bigram, embeddings) # fit of topic model  
topic_df = topic_model_5.get_document_info(bigram)  
topic_df.Topic.value_counts()
```