

## CONTENTS

<b>Business Problem.....</b>	<b>1</b>
<b>Background/History .....</b>	<b>1</b>
<b>Data Explanation.....</b>	<b>1</b>
<b>Conclusion.....</b>	<b>5</b>
<b>Challenges .....</b>	<b>5</b>
<b>Future Uses/Additional Applications.....</b>	<b>5</b>
<b>Recommendations .....</b>	<b>6</b>
<b>Implementation Plan.....</b>	<b>6</b>
<b>Ethical Assessment .....</b>	<b>6</b>
<b>GitHub.....</b>	<b>6</b>
<b>References .....</b>	<b>7</b>
<b>Questions .....</b>	<b>7</b>

## **Business Problem**

The project addresses the challenge of efficiently managing the increasing posts and discussions within a rapidly growing community Facebook group. As the community expands from 400 to an anticipated 700 houses, the need for streamlined communication and extracting actionable insights from group posts becomes critical for moderators and administrators.

## **Background/History**

The Facebook group's significance as a communication hub as the primary communication channel amongst community members has intensified with the community's expansion. Utilizing NLP and LDA, this study analyzes 3,500 posts from March 2022 onwards (starting of the Facebook Group), aiming to extract actionable insights for group moderators and administrators.

## **Data Explanation**

The analysis utilized a dataset comprising approximately 3,500 Facebook group posts collected from March 2022 to the present, focusing on non-personal data, including post creation dates, comment counts, reaction counts, and textual content—a rigorous data preprocessing phase involved anonymizing user identities and preparing the textual data for NLP analysis.

## **Methods**

The approach included:

**Data Collection:** Data Collection: Automated collection of posts and metadata using Apify, focusing on non-personal data of our Community Facebook group.

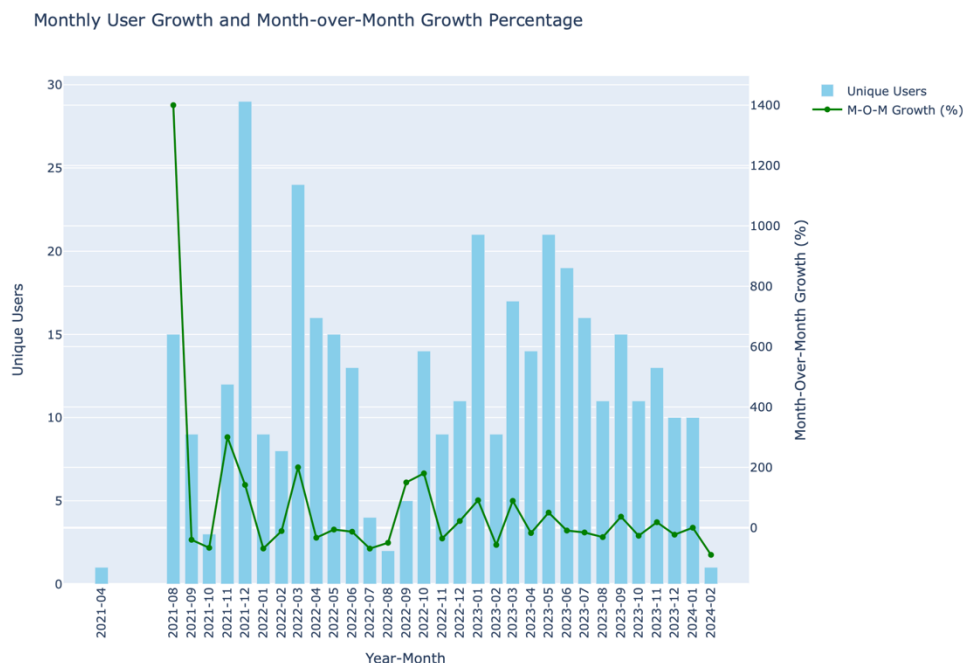
**Data Preprocessing:** Removal of potential personal identifiers and preparation of the dataset for analysis. For example, Facebook's assigned user id was removed/hidden in favor of random generated id.

**Analysis Techniques:** Application of NLP techniques, particularly LDA, to identify main discussion topics and analyze their evolution over time, offering insights into changing community interests.

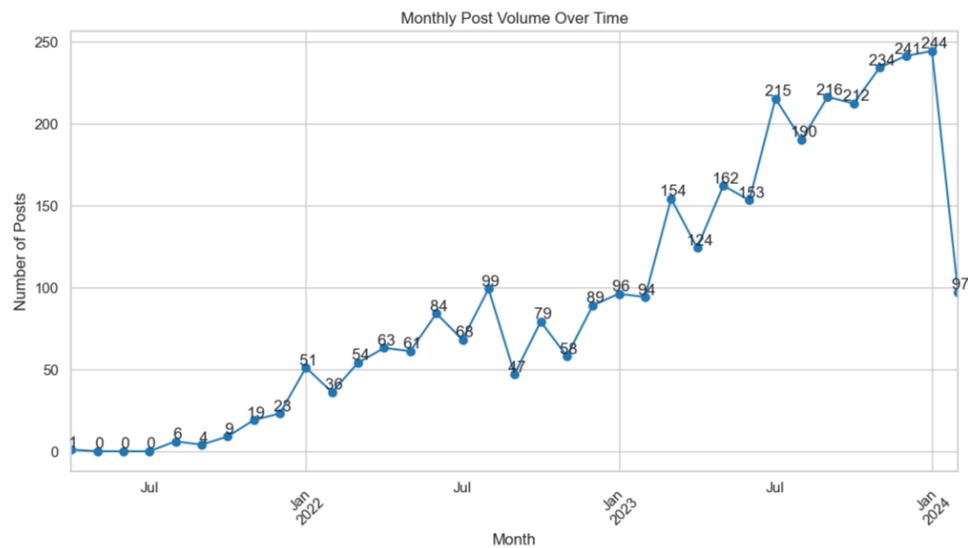
## Analysis

The comprehensive analysis unveiled various findings across several stages.

**Exploratory Data Analysis (EDA):** A remarkable growth in unique users was observed, from just one in April 2021 to a peak of 244 in January 2024, with a notable 90.9% growth in January 2023 alone. This signifies the group's expanding engagement. Interestingly, the top 10 users were responsible for about 23% of all posts, indicating a concentration of content creation among a few.



**Trend Analysis:** Monthly post volumes showed fluctuating engagement levels, with a significant surge to 215 posts in July 2023, and an upward trend continuing to January 2024, marking the highest post volume of 244. This indicates periods of heightened activity and engagement within the community.



**Topic Analysis:** Through Latent Dirichlet Allocation (LDA), we uncovered diverse topics of discussion, including:

- Package Delivery and Security: Discussions around "package", "delivered", "missing", and "mailbox" reflected significant concerns over package delivery issues, exacerbated by the rise of online shopping and the newness of the community, leading to frequent mis deliveries or lost packages.

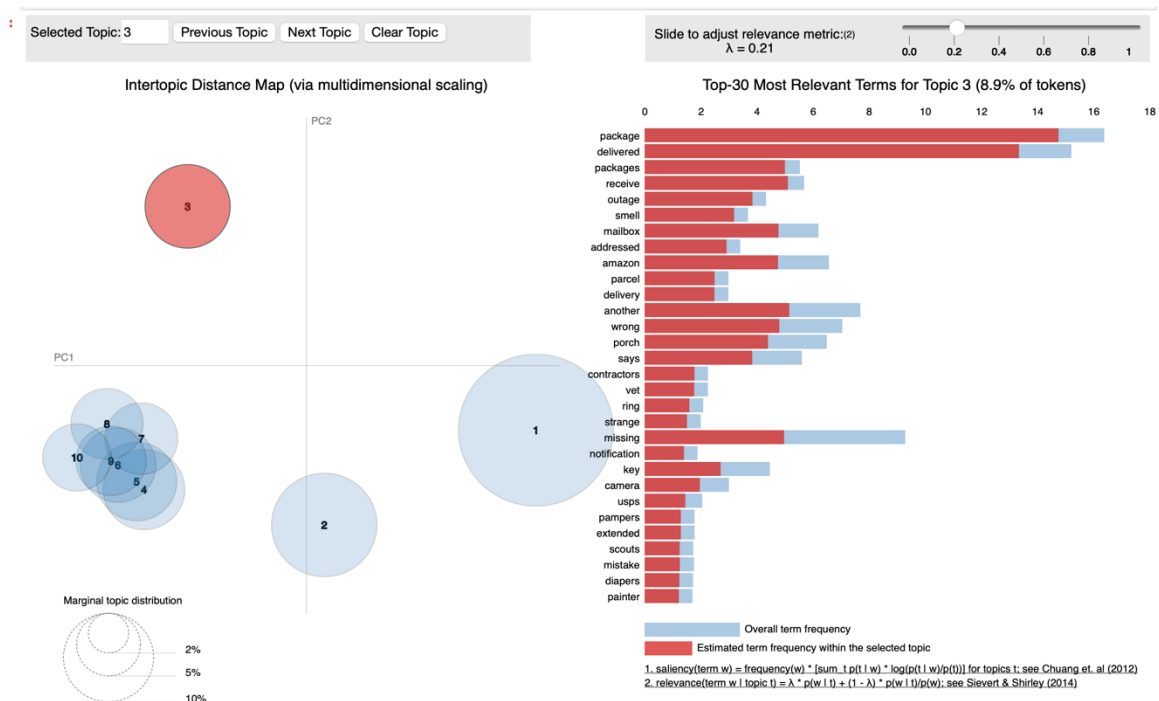
- Local Services and Community Engagement: Mentions of "cleaning", "business", and "notary" highlighted a focus on local services and businesses, suggesting conversations around service recommendations, experiences, or scheduling.

	User ID	Top Terms
0	625297	admins, business, cibolo, community, days, good, members, one, welcome, would
1	715425	cibolo, community, dog, go, like, one, park, trash, work, year
2	305043	com, dog, duck, everyone, https, looking, pool, scoop, today, would
3	34901	community, estate, get, group, help, home, kings, one, real, realtor
4	262535	fixed, garage, get, home, house, need, thanks, time, update, want
5	386398	15, custom, details, dm, hey, interested, need, pricing, selling, size
6	205205	big, get, hills, like, looking, neighbors, package, parkway, side, tall
7	914447	black, clean, going, hoa, one, pool, someone, thanks, want, water
8	190359	com, community, get, group, like, looks, neighbors, page, see, us
9	62215	2022, area, city, community, ems, gold, isd, school, star, transit

Table showing what are the topics that the top 10 posters/users posting about

- Infrastructure and Education: Topics such as "power", "EMS", "school bus", and "tricity" pointed to discussions about local infrastructure and school-related logistics.

Other topics included around Home Maintenance and Local Issues, Community Rules and Maintenance, Amenities and Services Recommendations etc.



## **Conclusion**

This project demonstrates the utility of advanced analytics in enhancing social media-based community engagement. Discerning communication patterns within the Facebook group offers a blueprint for improved community support and interaction by providing insight on how the Facebook group is being used.

## **Assumptions and Limitations**

The study assumes data completeness and the efficacy of LDA in capturing predominant discussion topics, primarily focusing on textual data, which may not fully represent the community's engagement spectrum.

## **Challenges**

Navigating data privacy, accurately capturing community discussions with LDA, and ensuring model relevance presented challenges. Technical limitations of data collection tools and the nuanced interpretation of unstructured social media content required careful consideration and validation.

## **Future Uses/Additional Applications**

The approach outlined can extend to behavioral modeling to predict engagement trends and personalize content delivery, ensuring relevance. Additionally, monitoring community health becomes feasible, identifying areas requiring intervention. The methodology's adaptability to other platforms promises a comprehensive community engagement strategy, enhancing interactions across the digital landscape.

## **Recommendations**

To enhance community engagement, curating content reflecting identified interests, such as package security, is crucial. Introducing participation incentives could diversify content creation, which is currently dominated by a few. Updating community guidelines to support healthy discussion around these interests will further enrich community interactions.

## **Implementation Plan**

Given the active participation of members, the group presents an excellent opportunity to raise awareness about common issues such as missing packages. Furthermore, the creation of dedicated channels within the Facebook group, focusing on specific topics such as buy/sell activities and recommendations for local businesses, appears to be a highly effective strategy for fostering proper community engagement.

## **Ethical Assessment**

Maintaining data privacy and anonymity is crucial. Efforts to mitigate bias and ensure diverse representation in data analysis are paramount, with transparent data usage and explicit member consent being ethical foundations.

## **GitHub**

[https://github.com/rozank/dsc680\\_applied\\_datascience/tree/main/Project-3](https://github.com/rozank/dsc680_applied_datascience/tree/main/Project-3)

## References

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the Social Sciences: Opportunities, Challenges, Ethical Considerations, and Practical Guidelines.

[https://www.researchgate.net/publication/281621604\\_Facebook\\_as\\_a\\_Research\\_Tool\\_for\\_the\\_Social\\_Sciences](https://www.researchgate.net/publication/281621604_Facebook_as_a_Research_Tool_for_the_Social_Sciences)

Jiang, J., Wilson, C., Wang, X., Sha, W., Huang, P., Dai, Y., & Zhao, B. Y. (2013). Understanding latent interactions in online social networks.

<https://people.cs.uchicago.edu/~ravenben/publications/pdf/latent-tweb13.pdf>

## Questions

Here is the list of 10 possible questions that the audience might ask.

1. What led to the need for analyzing Facebook group posts?
2. How did the significance of the Facebook group evolve with the community's expansion?
3. What data was analyzed in this study?
4. What methods were employed for data analysis?
5. What were the key findings from the exploratory data analysis?
6. How did community engagement trends change over time according to the trend analysis?
7. What topics were identified as areas of concern or interest within the community?
8. How can the insights from this analysis improve community support and interaction?
9. What future applications does the project's approach have?
10. What ethical considerations were made during the study?