Evaluating the Impact of Facebook Live on E-commerce: A Data-Driven Exploration of Seller Practices in Thailand

Ms. Sneha Ramesh Dharne
Msc in Data Analytics
National College of Ireland
Dublin, Ireland
x23195703@student.ncirl.ie

Abstract**—**This report analyses a Comma-Separated Values (CSV) panel of Facebook posts across different types (text, deferred, and live videos, images). These posts were taken out from the Facebook pages of ten Thai fashion and cosmetics retail sellers from March 2012 to June 2018. The data were collected through the Facebook API, and the data collected are anonymized to comply with the Facebook Platform Policy for Developers. In the scope of the current engagement metrics (shares, comments, and emoji), with emoji reactions being traditional "likes" in addition to the newly presented emoji reactions are "love", "wow", "haha", "sad", and "angry". The report provides the dataset that can be used as a tool for the engagement studies with the newest sales channel', Facebook Live with the following approaches: the comparative studies actions across other forms of writing such as text, deferred videos, and pictures and the seasonal trend analysis of posting performance and outlier posts.

Keywords— Social Media Marketing , component, Principal Component Analysis , Direct Selling, Customer Engagement.

I. Introduction

The use of live streaming has clearly incorporated a new turn in how the business communicates with its clients courtesy of the social media platforms. Of these platforms, the most potentially effective tool in the context of live communication with the audience for the sellers, and therefore a new potential for sales, is Facebook Live. Prior to the live stream, the customer engagement on Facebook was primarily based on the four post types, which include the status update, link, photo, and pre-recorded video. These studies again and again emphasized the fact that photos were the most effective type of content, as it received the maximum likes, comments, share, etc. Nevertheless, live streaming that now became a part of the strategy changes the overall picture dramatically providing an opportunity for various and probably much more engaging interaction.

Such urgency and purpose of the topic is heightened by the fact that many businesspeople especially those within the Thailand area adopt the Facebook Live streaming while marketing their products. It is therefore important to gain insight into the specifics of customer engagement in this regard so as to achieve better content tactics, enrich the customer relationship, and thus guarantee additional sales. As more organizations carry out experiments and gradually start tapping the prospects of live streaming, it is important to evaluate how as a medium it fares in engaging the audience against other formats of content.

The goal of this report is to understand usage and response correlated with Facebook Live selling in Thailand, with a further emphasis on the difference between selling through live broadcasts and conventional post types.

The research aims to answer the following questions:

1. How does engagement breakdown for the Facebook Live videos to the ordinary post types as; photos, and pre-recorded videos?

2. In the process of live streaming which patterns of user interaction can be distinguished and how do these patterns affect the overall efficiency of Facebook Live as the means of selling?

3. In what extent do various types of posts, such as live videos, influence the sentiment of users, in terms of emoji reactions?

The report is an attempt to shed light on the changing nature of customer interactions in the context of Facebook Live and offer practical tips for sellers who strive to enhance their social media performance.

II. Related Work

A. *Title: "Examination of Content Types and Social Media Engagement Indicators on Facebook: Case Analysis of 5-Star Hotels of Visegrad Group Countries"*

Analysis : The attached work qualitatively analyses the level of engagement on Facebook, and pays much attention to the visualization which also corroborates my observation. But, their use of exploratory statistics, especially the Kruskal-Wallis test is quite different from my utilization of scalable, parallel methods such as BFR and K-Means in MapReduce. It lets for a better dealing with the big amount of data and generating the more complicated nonlinear relations, which cannot be produced with the help of the traditional techniques.[1]

B. *Title: K-means clustering for the analysis of incomplete business data[2]*

Analysis: This work explores the application of K-means clustering combined with various imputation techniques to handle missing data in business datasets. While their focus is on improving clustering accuracy through data imputation, my approach leverages scalable parallel algorithms like BFR within a MapReduce framework to handle large-scale, complete datasets. This difference is crucial for real-time social media data analysis, where speed and scalability are key. The authors' use of "Big Data" K-means parallels my focus on optimizing clustering for large datasets, although my application specifically targets the dynamic engagement metrics of Facebook Live.

C. *Title: Model based clustering of multinomial count data*

Analysis : The attached paper "Model based clustering of multinomial count data" by Panagiotis Papastamoulis explores clustering in multinomial datasets using a combination of Expectation-Maximization (EM) and Bayesian methods. The work is relevant to my approach, particularly in its emphasis on handling complex, heterogeneous datasets through scalable algorithms. However, while the paper applies sophisticated clustering techniques to multinomial data, my approach leverages parallel algorithms like K-Means and BFR within a MapReduce framework for real-time social media data analysis, emphasizing scalability and adaptability to dynamic environments. This contrast highlights the need for efficient, parallelizable methods in handling large-scale, real-time datasets.[3]

D. *Title: Influence of Facebook brand-page posts on online engagement[4]*

Analysis: The referenced study explores the effects of media and content types to the engagement of Facebook users; employs manual coding of Facebook posts and traditional quantitative approach in determining users' response (likes, comments, shares). As much as it is a good and effective method, this approach suffers from one major drawback; it is tedious and cannot handle huge volumes of data let alone real-time data. In contrast, my work uses first K-Means and then BFR parallel algorithms selected for the MapReduce work environment that enables processing large datasets in real time. This makes it possible to extract detailed engagement patterns and user interactions, which are essential in real-time platforms such as the Facebook Live as the stream of data produced is massive and very rapid.

E. *Title: Live streaming commerce from the sellers' perspective: implications for online relationship marketing*

Analysis: The most direct type that has been recently practiced is live streaming which has provided self-employed/ sellers high levels of consumer interactions. In the literature, most prior studies are focused on consumer motivation and shopping intention in live streaming environment, yet the insight from the sales personnel's viewpoint is not well researched. This particular work employs mixed with a method research design to carry out an evaluation of the Facebook data from live streaming sellers in order to assess the activity data and the interactive processes of the sales activities. This research categorizes four specific natures of sales and twelve of the strategies employed in reaching for the consumer and holding them. Such strategies are aligned to the relationship processes and outcomes, thus offering a framework for analysing relationship mechanisms in live streaming commerce.[5]

F. *Title: Facebook (A)Live?: Are Live Social Broadcasts Really Broadcasts?*

Analysis: The return to 'live' has gained new characteristics such as live-streaming of content, content broadcast to a global audience, and has evolved from professional content producers to 'live' streaming by individuals and using portable technologies: smart mobile devices. This shift, especially the move to Facebook Live helps in interaction of the broadcaster with the viewer as the process is live. Based on the Facebook Live data analysis of 3000 TB collected during a month, the authors define the peculiarities of the live broadcasting and propose the ways to minimize the load on the network. It also looks at the universal and the local view of live video, in terms of audience activity during live streams as compared to when watching a batch of videos on the social network. In an even more curious manner, most of the features associated with Facebook Live can be seen as differing with secondary and primary meaning orientations of 'live' and 'broadcast.'
[6]

G. *Title: The role of live streaming in building consumer trust and engagement with social commerce sellers[7]*

Analysis: The referenced study focuses on live streaming as a direct selling technique and analyses issues to do with perceived value, trust and customer response. As helpful as it is for explicating such dynamics, the article does not really get into how such ideas play out at the technological level, namely, how real-time analysis of stream-video data would be computationally handled. On the other hand, my approach uses parallel algorithm such as the K-Means and BFR that are suitable for large scale environments more enhanced in the MapReduce form of execution. This makes it possible to analyse large volumes of data in real-time and get insights that can be used to improve the strategies towards increasing consumer participation in social commerce.

III. METHODOLOGY

A. *Data Description and Data Dictionary*

The dataset used here for analysis includes engagement data concerning Facebook Live broadcasts and other posts, including photos and video clips. Data is collected from the Facebook with the special emphasize on the sellers from Thailand. Some of the likely attributes include post_type, num_reactions, num_comments, num_shares, num_likes and other emoji reactions such as num_loves, num_wows, among others since they indicate the audience interaction thus a way of understanding how different content types do in terms of engagement.
They used it to analyze the frequency of users' activity, the efficiency of live streaming as a sale promotion, and the

kind of posts that create the positive or negative attitude among them.

Below is a data dictionary describing the attributes:

Attribute	Data Type	Description
status_id	Integer	Unique identifier for each post
status_type	Categorical	Type of the post (e.g., live video, photo, pre-recorded video).
status_published	Categorical	Date and time when the post was created.
num_reactions	Integer	Total number of reactions received on the post
num_comments	Integer	Total number of comments received on the post.
num_shares	Binary	Total number of times the post was shared.
num_likes	Integer	Total number of "Like" reactions.
num_loves	Binary	Total number of "Love" reactions.
num_wows	Binary	Total number of "Wow" reactions.
num_hahas	Binary	Total number of "Haha" reactions.
num_sads	Binary	Total number of "Sad" reactions.
num_angrys	Binary	Total number of "Angry" reactions.

B. *Data Translation and Manipulation Rules*

Several steps were applied to prepare the data for clustering:

1. Handling Missing Values: Naive values were excluded from the observations where there were inputs missing in order to guarantee the validity in the process of clustering.
2. Combining Reactions: The various types of reactions which included, num_likes, num_loves, num_wows, and others were kept different so as to examine the sentiment of the users in detail.
3. Creating a Feature Vector: Such numeric attributes concerning engagement were handles ; num_reactions, num_comments, num_shares, etc., merged into one in PySpark as feature vector using VectorAssembler.
4. Filtering by Post Type: Again, only such post types were included in the analysis (live videos, photos, pre-recorded videos only and so on) that are potential to have the maximum impact on the community.

C. *Architecture and Application Workflow*

Databricks platform was used for the analysis with focused on its scalability and used of distributed computing environment. The workflow follows a scalable architecture designed to handle large datasets efficiently, making use of the following key components:

The workflow follows a scalable architecture designed to handle large datasets efficiently, making use of the following key components:

- Data Ingestion: Data was imported into the Databricks environment from the data sources in form of CSV files which are stored on the Databricks File System.
- Data Processing: Data processing was done using PySpark and here Map Reduce was used so as to divide the data processing jobs into different nodes. It provided for high throughput and management of what would amount to large data sets relative to the analysis being performed.
- Clustering Algorithms: K-Means and BFR (Bradley-Fayyad-Reina) algorithms were chosen because of the applicability in big data parallel computing. These algorithms are useful to manage the big datasets and have capabilities to explain the structure of the whole dataset.
- Visualization and Export: Matplotlib and Seaborn were used to visualize the results of the clustering analysis that was carried out. The final and processed data and graphs and tables were exported for reporting.

- Scalability Approach: Therefore, the selection of K-Means and BFR clustering as well as the MapReduce structureability guarantees that the analysis can fit large and constantly fluctuating social media data. Another factor resulting into this is the distributed architecture of Databricks that has aided scalability for real time processing and analysis.

D. Data Processing Activities

The data processing activities were carried out in the following order:

1. **Data Ingestion**: Data was read from a DBFS into a PySpark DataFrame.
2. **Data Cleaning**: Many authors have used deletion, that is, cases where one or more of the original variable values are missing are simply discarded, and no analysis is done on variables that contain missing values; therefore, the missing values were processed by deleting rows with missing data.
3. **Feature Engineering**: In order to cluster the records Dataframe the VectorAssembler was used to assemble the records data into feature vector selecting relevant numeric fields in the record.
4. **Clustering**: Cluster analysis was performed on the feature vectors using the K-Means as well as BFR algorithms to classify the posts in clusters that reflect the pattern of users' engagement and sentiment.
5. **Visualization**: After clustering, the positioning of the clusters was done through different forms of plots such as scatter plots, 3D scatter plots.
6. **Exporting Results**: The final clustered data and visualizations were exported out for more analysis and then for reporting .
7. **Justification for Methods**: The activities for data processing were thus a result of the vast data that was involved in social media analysis and the ability to extract knowledge for decision making purposes. The MapReduce framework was particularly useful in dividing the processing work, making sure that the analysis was to be as elastic as the data itself.

E. Ethical Considerations :

The following ethical factors were taken into account when gathering and processing the data:

1. Data privacy: No personally identifiable information (PII) was processed because the data used in the analysis was anonymized.
2. Consent: Facebook posts that are accessible to the public were the source of the data, as individuals had implicitly consented to share their information with the public.
3. Transparency: The techniques and algorithms employed in the analysis were selected to guarantee repeatable and transparent outcomes, enabling responsibility for the handling and interpretation of the data.

Because of these ethical considerations, the analysis was carried out in a responsible manner that respected user privacy and followed best practices for data processing.

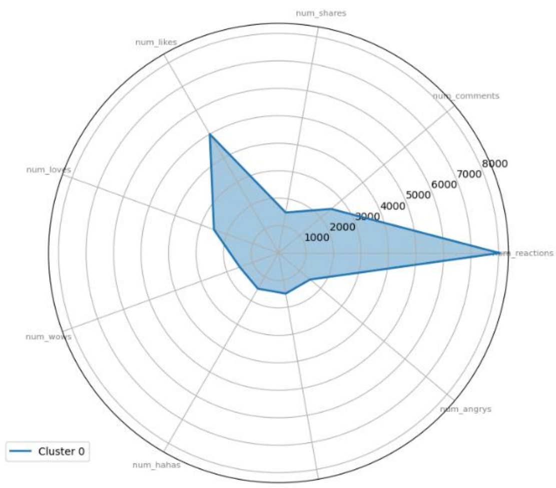
IV. Results

The main goals were to comprehend how various content ki nds affected user engagement, spot trends in user behavior d uring live streaming, and assess how users felt about various post kinds.

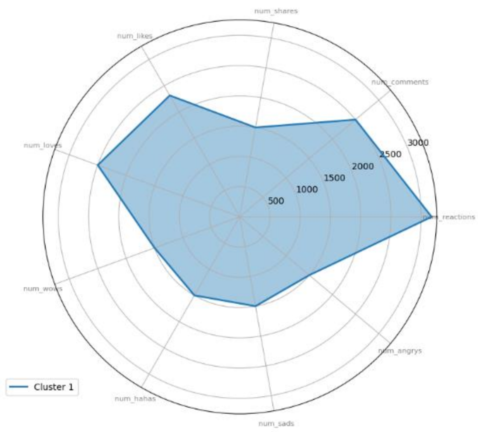
A. Clustering Results

1) **K-Means Cluster Centers**: Using engagement numbers as a guide, the K-Means algorithm created unique clusters that distinguished between different categories of material. One cluster, for instance, has significant numbers of shares and comments, indicating high levels of involvement and usually being connected to Facebook Live videos. A different cluster had a large number of reactions but a low number of comments, which is probably indicative of attention-grabbing images that don't spark discussion.

Cluster 0 Center



Cluster 1 Center



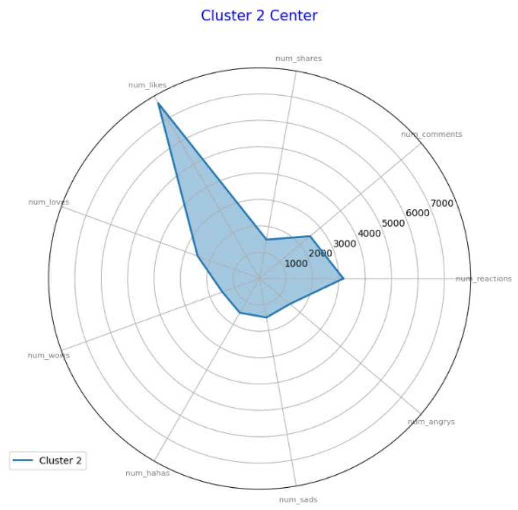


Fig. 1 K-Means Cluster Centers

2) **BFR Cluster Centers**: Similar clusters were generated by the BFR algorithm, supporting the K-Means results. A subset of posts with strong negative sentiment (num_angrys and num_sads) was also identified using the BFR clusters.

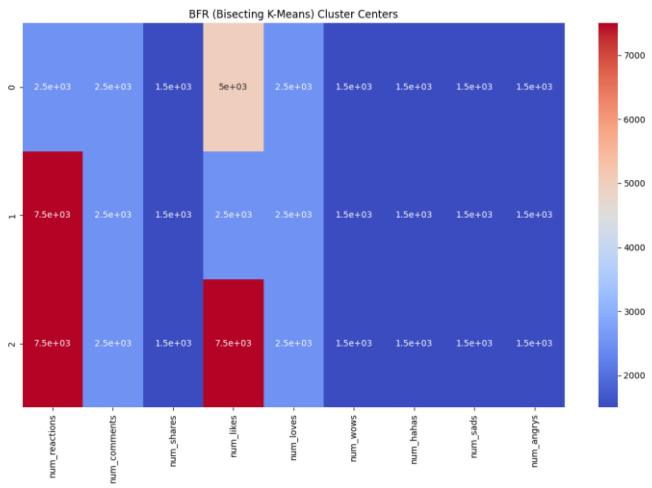


Fig. 2 BFR Cluster centres

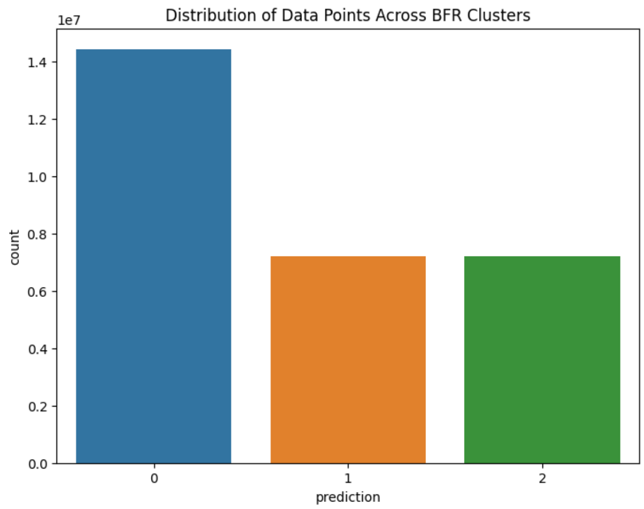


Fig. 3 Distribution of Data points across BFR Clusters

B. Research Question Analysis :

Research Question 1 : In general, Facebook Live videos garner more interaction (num_comments, num_shares), however some photographs also attain similar levels of engagement, according to the clustering data. This result highlights the fact that although live videos are useful, still images can occasionally outperform them, which helps to partially address the research topic.

Research question 2: The unique cluster that exhibits strong interaction metrics during live streaming sessions points to active user participation in the form of shares and comments as the primary driver of user engagement. This pattern increases Facebook Live's efficacy as a sales tool by promoting real-time engagement, which is essential for generating sales.

Research Question 3: The BFR algorithm found a cluster linked to negative sentiment, indicating that whereas live streams frequently elicit good feedback, they can also elicit negative feedback. According to this research, content strategy needs to be carefully controlled in order to reduce negative user interaction and preserve user sentiment.

Conclusion and Future Work

The results of the investigation challenged preconceived notions about the usefulness of content, showing that while Facebook Live broadcasts significantly increase user engagement, some photographs also garner similar levels of interaction.

The identification of negative sentiment clusters emphasizes the necessity for improved content strategies. Deeper sentiment analysis is however constrained by the study's dependence on grouping.

In order to more precisely forecast engagement trends, future research may investigate the integration of sentiment analysis with machine learning models. Furthermore, broadening the dataset to encompass a wider range of post kinds and extended time periods would yield a more comprehensive comprehension of content performance and user behavior over an extended period.

.

References

[1] L. Nastisin, R. Fedorko, B. Gavurova, and R. Bacik, “Examination of Content Types and Social Media Engagement Indicators on Facebook: Case Analysis of a 5-Star Hotels of Visegrad Group Countries,” *Mark. Manag. Innov.*, vol. 15, no. 1, pp. 112–119, 2024, doi: 10.21272/mmi.2024.1-09.

[2] “K-means clustering for the analysis of incomplete business data,” *Acad. J. Comput. Inf. Sci.*, vol. 4, no. 6, 2021, doi: 10.25236/AJCIS.2021.040606.

[3] P. Papastamoulis, “Model based clustering of multinomial count data,” *Adv. Data Anal. Classif.*, Jul. 2023, doi: 10.1007/s11634-023-00547-5.

[4] P. Luarn, Y.-F. Lin, and Y.-P. Chiu, “Influence of Facebook brand-page posts on online engagement,” *Online Inf. Rev.*, vol. 39, no. 4, pp. 505–519, Jan. 2015, doi: 10.1108/OIR-01-2015-0029.

[5] A. Wongkitrungrueng, N. Dehouche, and N. Assarut, “Live streaming commerce from the sellers’ perspective: implications for

online relationship marketing," *J. Mark. Manag.*, vol. 36, no. 5–6, pp. 488–518, Mar. 2020, doi: 10.1080/0267257X.2020.1748895.
[6] A. Raman, G. Tyson, and N. Sastry, "Facebook (A)Live?: Are Live Social Broadcasts Really *Broad* casts?," *Proc. 2018 World Wide Web Conf. World Wide Web - WWW 18*, pp. 1491–1500, 2018, doi: 10.1145/3178876.3186061.
[7] A. Wongkitrungrueng and N. Assarut, "The role of live streaming in building consumer trust and engagement with social commerce sellers," *J. Bus. Res.*, vol. 117, pp. 543–556, Sep. 2020, doi: 10.1016/j.jbusres.2018.08.032.
[1]