FACEBOOK BUSINESS ACTIVITY TRENDS FOR DISASTER RECOVERY

Introduction To Data Science



INTRODUCTION TO DATA SCIENCE

1 ABSTRACT

Posts from Facebook business pages are essential tools for businesses to engage with their audience and share information. These posts serve multiple purposes, including promoting products or services, providing updates, sharing content, and interacting with customers. They often include text, images, videos, and links to external content. Facebook's algorithm determines the visibility of these posts to users based on factors like engagement, relevance, and timing. Business pages can use these posts to build brand awareness, drive website traffic, and foster a community of loyal customers. Monitoring and analyzing the performance of these posts is crucial for businesses to refine their social media strategies and enhance their online presence.

In this report let us see that how these posts are leveraged to monitor and analyze business activity, providing valuable insights into recovery and resilience after disruptive events like COVID-19 pandemic.

2 INTRODUCTION

The COVID-19 pandemic, Since its onset in late 2019, has unleashed unprecedented disruptions to global economies. This crisis has disproportionately affected small businesses, particularly in low-income countries, amplifying the urgent need for effective recovery strategies. Traditional survey methods and remote sensing offer limited insights into this complex issue. However, the emergence of data from mobile usage and web products, coupled with a novel methodology inspired by Eyre, De Luca, and Simini, presents a promising solution. This report explores the implementation of this methodology through Facebook Business Activity Trends (BAT) datasets, designed to provide real-time, global insights into business recovery. By examining practical applications, case studies, and limitations, this report aims to guide strategies for economic revival and resilience in the face of adversity and the important causes of the business downtime.

3 BASIC STATS

3.1 Understanding the data

To extract the csv files from the folder the following process was done:

- . Start with creating the dataframe to store the csv files.
- . Iterate through each csv file inside the folder, load the file into a dataframe and concatenate the dataframes into one.

```
path =r'/Users/namia/Downloads/Dataset 1 - Bussiness Activities/Part 1'
filenames = glob.glob(path + "/*.csv")
dfs = pd.DataFrame()
for df in filenames:
    df1 = pd.read_csv(df)
    dfs = pd.concat([dfs, df1])
```

. The process was repeated for the Dataset 1 - Bussiness Activities/Part 2 and then concatenated to form the dataframe df.

3.2 Missing Value Analysis and Anomaly Detection

The following code was executed to find the missing value percentage in the dataframe

```
round((df.isnull().sum()/len(df))*100,2)
```

· Detected the presence of missing values in the column; 'Country', 'gadm1 name', 'gadm2 name'.

- . Given the global scale of the COVID-19 crisis, state-level and county-level data, as indicated by columns 'gadm1 name' and 'gadm2 name', holds limited relevance and are therefore dropped.
- . The column 'gadm name' provides the expanded form of the country name, rendering the country column unnecessary.
- . Redundant columns, such as 'gadm0 name', have been removed to increase data efficiency.
- . The 'gadm id' column is unnecessary as we now have the country name, providing the required context.
- The column 'gadm level' is dispensable since the global nature of the COVID-19 crisis consistently renders it at level zero, indicating its impact on entire countries. The columns are dropped as shown below. df=df. drop(['gadm_id', 'gadm0_name', 'gadm1_name', 'gadm2_name', 'country', 'gadm_level'], axis=1)
- Filtered the dataframe for the presence of duplicated rows and found none.
 df[df. duplicated()==True]

3.3 Further Analysis

. Count the number of countries/regions being mentioned in the dataset.

country_count=df['gadm_name'].nunique()

Total number of countries in the data: 220

 Count the total number of dates available for each country/state (or equivalent geo-area), in terms of the type of business.

df. groupby(['gadm_name', 'business_vertical'])['dates']. count()

Total number of dates available for each country across each business sector: 1004

This signifies that data has been documented for every country within various business sectors over a period of 1004 days following the commencement of the crisis.

3.4 Metrics and their computation

Activity Quantile:

- . Definition: Activity quantile measures the level of activity relative to a baseline period, typically using a 7-day average of the aggregated probability integral transform metric.
- Calculation: It involves approximating quantiles (midquantiles) of daily activity compared to baseline activity for each Page. These quantiles are summed and then subjected to a series of transformations, including shifting, rescaling, and variance adjustment to align with a standard normal distribution. The adjusted sum is further probability transformed to yield a value between 0 and 1. A 7-day rolling average is then applied to smooth daily fluctuations.

$$q_{i}(t) = \frac{1}{2} \left[P_{i}(x_{i}(t)) + P_{i}(x_{i}(t) + 1) \right]$$

$$rP/T(t) = \sum_{i \in \mathcal{B}} q_{i}(t)$$

$$rN(t) = \frac{rP/T(t) - n(t)}{\sqrt{n(t)/12}}$$

$$rN(t) = \frac{rN(t)}{\sigma}$$

$$rU(t) = PN(rN(t))$$

Average over the last 7 days taken to obtain the smoothed daily fluctuations.

. Interpretation: This metric provides a quantile interpretation by comparing daily activity to the baseline distribution, with a value around 0.5 indicating normal activity. It treats all businesses equally and is less affected by prolific posters. This metric is recommended for robustness against outliers and numerical stability.

Activity Percentage:

- . Definition: Activity percentage is the 7-day rolling sum of total activity (e.g., total posts) expressed as a percentage of the average weekly baseline activity.
- Calculation: The weekly baseline average is computed as the mean of 7-day sums of total activity, with each sum starting on a Monday within the baseline period. For each day during the crisis, the 7-day rolling sum of total posts is divided by this weekly baseline average and multiplied by 100.
- Interpretation: This metric is straightforward to interpret, with a value near 100 indicating normal activity. However, it is sensitive to businesses that post frequently and can yield less stable results when post counts are low. It is recommended when interpretability is the primary concern.

3.5 Time Series

Let's examine the business activity trends over the period from March 2020 to November 2022. We will utilize the metric of activity quantile, measuring the activity level in comparison to a baseline period 365 days before the crisis. This analysis spans various countries and business verticals. A quantile value of 0.5 signifies normal business activity. Values greater than 0.5 indicate improved business activity during the COVID-19 period compared to before the crisis, while values less than 0.5 suggest that business activity before the crisis was higher than during the COVID-19 period.



Fig. 1. Exploring Trends in Activity Quantiles Over the Years: An In-Depth Analysis

The declaration of COVID-19 as a pandemic in March 2020 initiated widespread global lockdowns. Business activity exhibited notable declines in March 2020, January 2021, August 2021 and January 2022. The global lockdown, implemented to control the virus, had profound effects on various aspects of daily life, impacting food security, the global economy, education, tourism, and more. The reduced business activity during these months can be attributed to waves of the virus, prompting lockdowns and preventive measures, such as social distancing and business closures, to mitigate its spread which is evident from the fall of activity quantile from 0.54 to a value of 0.36.

3.6 Feature Engineering

I am conducting feature engineering on the 'ds' column (Date of the activity), extracting information such as year, month, quarter, weekday, and weekend. The code is applied to enhance the dataset, and these derived features will be utilized for subsequent analyses.

```
df['dates']=pd. to_datetime(df['ds']). dt. strftime('%b', %Y')
df['week_no']=pd. to_datetime(df['ds']). dt. strftime('%U')
df['year']=pd. to_datetime(df['ds']). dt. strftime('%Y')
df['month']=pd. to_datetime(df['ds']). dt. strftime('%b')
df['quarter']=pd. to_datetime(df['ds']). dt. quarter
df['weekday']=pd. to_datetime(df['ds']). dt. strftime('%a')
df['week_end']=df['weekday']. apply(lambda x:'Yes' if x in ['Sat','Sun'] else 'No')
```

3.7 Weekdays vs Weekends

The plot displays the average activity quantile across business vertical for different countries spanning three years, differentiating between weekdays and weekends. It is evident that the median remains constant, indicating that weekdays and weekends do not significantly influence business activity during the COVID-19 period.

```
sns. boxplot(x=df['week_end'], y=df['activity_quantile'])
sns. barplot(x=df['week end'], y=df['activity quantile'], hue=df['year'])
```

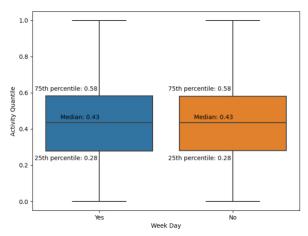


Fig. 2(a). Weekend Vs Weekdays

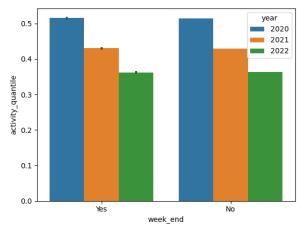
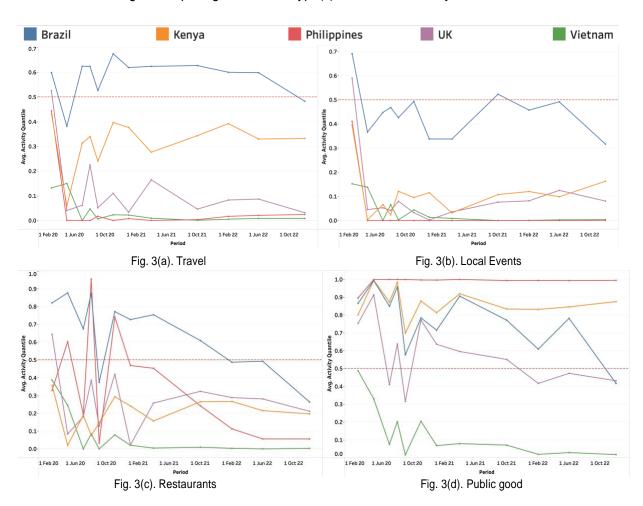


Fig. 2(b). Weekends and Weekdays over a span of 3 years

Although the day of the week is not affecting the business, COVID-19 has badly affected the business lowering the activity quantile to 0.43

3.8 Time-series figures depicting the chosen type(s) of business activity in five diverse countries



3.8.1 **Travel**.

- (1) Philippines and Vietnam Downturn: Both Philippines and Vietnam, as major tourist hubs near China, experienced a significant downturn during COVID-19 due to mobility restrictions and virus control measures.
- (2) London's Travel Industry Decline: London, a renowned global tourism hub celebrated for historical sites and cultural events, faced a decline in the travel industry amid the pandemic.
- (3) Kenya's Rebound: Kenya is rebounding from the COVID-19 downturn, benefiting from local travelers capitalizing on reduced prices. However, international visitor numbers are yet to reach pre-pandemic levels.

- (4) Brazil's Tourism Renewal: Brazil revitalized its tourism industry by emphasizing short-haul domestic travel, implementing health protocols, achieving increased vaccination rates, and making infrastructure investments.
- (5) Positive Trends in Brazil: Positive numbers in international arrivals and tourism spending contribute to Brazil's successful restoration of its tourism industry to pre-pandemic levels.

3.8.2 Local Events.

- (1) Cultural Impact in Brazil: The closure of cultural spaces and event suspension in March 2020 due to social distancing measures significantly affected Brazil's cultural sector, which annually contributes R\$170 billion to the economy and employs five million people.
- (2) Government Aid Lei Aldir Blanc: Congress responded to the crisis by approving the Law for Cultural Emergency (Lei Aldir Blanc), allocating R\$3 billion from the Federal Cultural Fund. This provided three months of emergency aid and up to six months of tax exemption for the cultural industry and creative businesses.
- (3) Innovative Cultural Responses: Creative initiatives, including livestreamed concerts, online festivals, and virtual productions, emerged during the pandemic. Trends like "Lives Solidarias" in live streaming garnered millions in donations. Brazilian artists showcased resilience and adaptability, dominating global online concerts.
- (4) Comparison with Other Countries: In contrast to Brazil's approach, other countries implemented strict lockdown measures, resulting in widespread event cancellations. The events industry, covering live performances, conferences, and sports events, faced significant disruptions.
- 3.8.3 **Restaurant:** Despite variations in preparedness and development status, all countries experienced a significant downturn in their restaurant businesses.
- 3.8.4 **Public good:** In 2021, decreased government pandemic aid in the Philippines resulted in a drop in recipients from 80% in 2020 to 50%. Approximately 70% of households faced reduced food consumption due to financial constraints and increasing prices. To manage, impoverished households postponed payments, utilized credit, borrowed, and sold assets. Community pantries, started by private individuals and groups, proliferated. Interestingly, in 2021, more households received support from church-related and private organizations than from the government, marking the second year of the pandemic (World Bank 2021).

3.9 Business Levels Across Countries on Two Selected Days

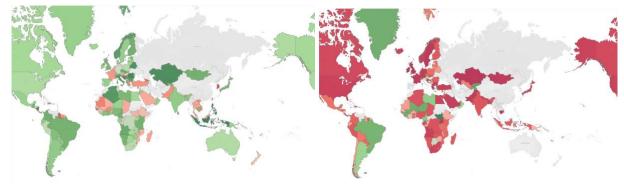


Fig. 4(a). Onset of COVID-19 pandemic

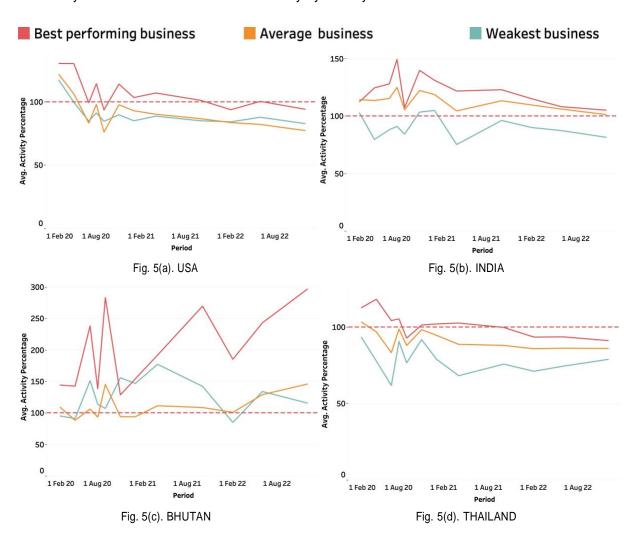
Fig. 4(b). Conclusion of the COVID-19 pandemic

The maps provide a detailed view on the change in activity quantile and the impact of COVID-19 on various countries across diverse business verticals. The maps are plotted for the dates 5th March 2020 and 8th January 2022, marking the commencement and near conclusion of the COVID-19 pandemic. Observing this visualization reveals that South America emerges as the least affected continent, with Brazil experiencing the least impact, followed by Greenland. In contrast, many Asian countries, particularly those in close proximity to China, such as **Pakistan** and **Vietnam**, appear to be significantly affected, especially when considering the mentioned dates.

4 EVENTS

This section will offer a comprehensive analysis of the impact of the COVID-19 crisis, examining its effects continent-wise, country-wise, and business-wise. Stringency index data collected from Blavatnik School of Government, University of Oxford for further analysis on government policies and Lockdown

Analysis of business verticals on a country-by-country basis



The top-performing sectors comprise grocery and convenience stores, manufacturing, public goods, and retail. On the other hand, businesses in the average-performing category encompass business and utility services, home services, and lifestyle services. The worst-performing sectors include travel, local events, and restaurants.

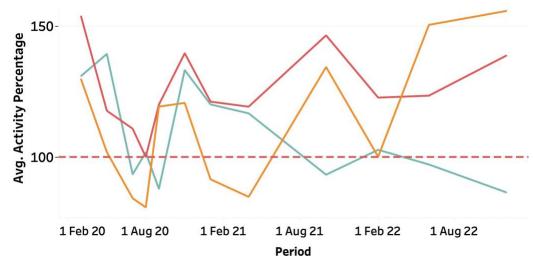


Fig. 6. GREENLAND

Top-Performing Sectors:

- . Grocery and Convenience Stores: These businesses experienced increased demand as people stockpiled essentials and preferred shopping at local stores to avoid crowded spaces.
- Manufacturing: Certain manufacturing sectors, particularly those involved in producing essential goods like medical supplies and packaged foods, remained resilient during the pandemic.
- Public Goods: Industries involved in producing public goods, such as healthcare and pharmaceuticals, were in high demand, given the global focus on health and safety.
- . Retail: Some segments of the retail sector, especially those with a strong online presence, saw increased sales as consumers shifted to online shopping.

Average-Performing Sectors:

- . Business and Utility Services: While some businesses in this category may have faced challenges due to disrupted operations, others that provided essential services likely fared better.
- Home Services and Lifestyle Services: The impact on these sectors varied. Home services related to
 essential maintenance may have continued, while non-essential lifestyle services faced disruptions due to
 lockdowns and social distancing measures.

Worst-Performing Sectors:

- . Travel: Travel-related industries, including airlines, hotels, and tourism, were severely impacted by lockdowns, travel restrictions, and a decrease in consumer confidence to travel.
- . Local Events: With restrictions on gatherings and events, businesses involved in organizing local events faced cancellations and a significant downturn in revenue.
- . Restaurants: The restaurant industry suffered due to lockdowns, capacity restrictions, and changes in consumer behavior, with many people opting for home-cooked meals or takeout instead of dining out.

4.2 Unveiling Insights: Exploring Varied Metrics

Australia

The line plot below depicts the retail business trend from March 2020 to November 2022 for specific countries. It is evident that the activity quantile demonstrates a distinct deviation from 0.5, highlighting anomalous business activity. In contrast, the activity percentage does not exhibit the proper deviation, Since it gets affected by outliers

Bahrain

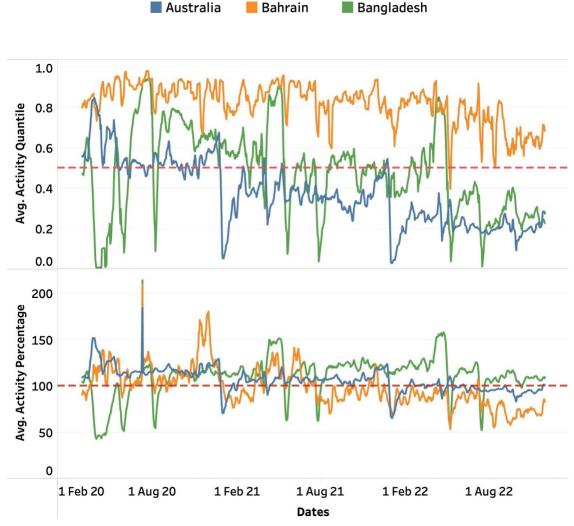


Fig. 7. Activty Quantile Vs Activity Percentage

Activity Quantile:

. Robustness and Numerical Stability: The Activity Quantile metric is recommended for its robustness against outliers and numerical stability. This implies that it is less susceptible to extreme values or anomalies, providing a reliable measure of daily activity levels.

- . Equal Treatment of Businesses: This metric treats all businesses equally, irrespective of their posting frequency. This characteristic is advantageous when aiming for a balanced representation of different entities, avoiding biases introduced by prolific posters.
- Quantile Interpretation: The quantile interpretation provided by this metric, with a value around 0.5 indicating normal activity, allows for a nuanced understanding of how daily activity compares to the baseline distribution.

Activity Percentage:

- Interpretability Focus: The Activity Percentage metric is particularly useful when interpretability is the primary concern. It offers a straightforward interpretation, where a value near 100 signifies normal activity.
- . Sensitivity to Posting Frequency: This metric is sensitive to businesses that post frequently. Hence, if entities are prolific posters, the results may be influenced by their activity, leading to less stable outcomes when post counts are low.

4.3 Government Policies - Stringency Index Metrics

The Stringency Index Metrics encompass nine indicators such as school closures, workplace closures, public event cancellations, and international travel controls. Calculating the Stringency Index involves determining the mean score of these metrics, with values ranging from 0 to 100. A higher score indicates a more stringent government response, with 100 representing the strictest measures. Subnational variations are depicted based on the strictest sub-region. The index is calculated for three categories: vaccinated, non-vaccinated, and a national average weighted by vaccination share. It gauges the strictness of government policies but does not assess their appropriateness or effectiveness.

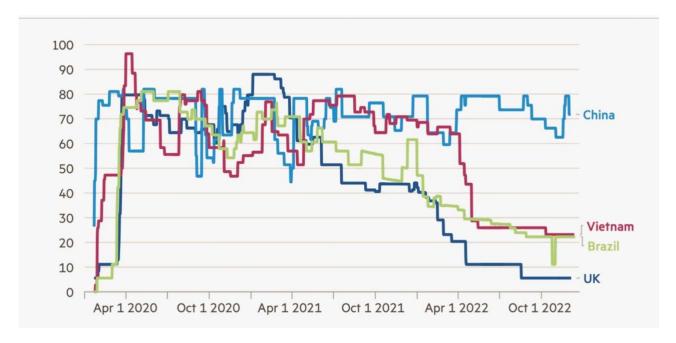


Fig. 8. Stringency Index

4.4 Impact of Global Lockdown on Tourism, Gender Relations and Education

Impact of Global Lockdown on Tourism

- . International tourism declined by 22% in Q1 of 2020, with a projected 60%-80% decline by the end of 2020.
- . Q1 2020 saw 67 million fewer international tourists, resulting in a loss of \$80 billion in exports. About 100–120 million jobs directly tied to tourism are at risk, with an estimated loss of \$910 billion to \$1.2 trillion in export revenues.
- · Countries considered COVID-19 hotspots lose millions in revenue due to canceled travel plans.

Impact of Global Lockdown on Gender Relations and Domestic Violence/Abuse:

- . Lockdown alters work-family and gender relations, affecting men and women differently. Increased domestic labor and care burden on women due to closed schools and nurseries.
- Women globally are responsible for 75% of unpaid care and domestic work, highlighting gender gaps. Gender gaps extend to work hours, job opportunities, income, and social standing.
- . Psychological impact includes anxiety, depression, distress, sleep disorders, and post-traumatic stress disorders. Causes of psychological distress include duration of lockdown, fear of infections, frustration and boredom, inadequate supplies, and inadequate information. Elevated levels of anxiety and depression during pandemics, leading to confusion, anger, anxiety, and post-traumatic symptoms.
- Helplessness, sadness, frustration, and loneliness recorded, exacerbated by self-isolation/quarantine, lack of lifestyle options, misinformation, and economic woes.

Global School Closures and impact on Education:

- . Implemented by the majority of governments worldwide as part of COVID-19 lockdown measures. Affected 143 countries, impacting 67.6% (1,184,126,508) of enrolled learners globally across different education levels.
- . Justification rooted in children's lower immunity and higher transmission tendencies of infectious diseases.
- . Global lockdown significantly affected education, leading to widespread school closures and a shift to online learning.
- Effects on young people included anxiety about education and disruptions in motivation and concentration. Researchers experienced varied impacts, with disruptions for some and increased productivity for others. Transition to online learning posed challenges, especially in regions with limited internet access and resources.

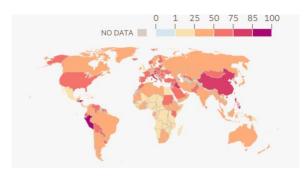


Fig. 9(a). Lockdown during onset of COVID-19 pandemic

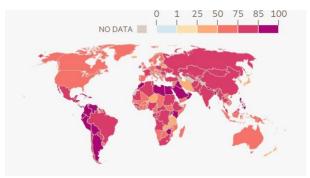


Fig. 9(b). Lockdown during COVID-19 pandemic

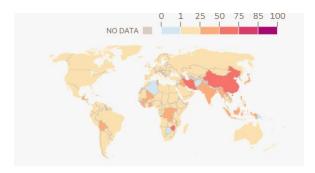


Fig. 10. Lockdown during conclusion of the COVID-19 pandemic

The map illustrates the worldwide extent of Covid-19 lockdown stringency, showcasing an inversely proportional relationship with business activity, Asian country being affected most due to its close proximity with China

Transformations in Sectors Amidst Social Distancing Measures:

- . After over 100 days of social distancing, the cultural sector is reinventing itself for survival.
- . Online cultural consumption and creative production are surging.
- . Music leads innovation with live concerts, while theatre companies produce web performances.
- . Drive-in cinemas, virtual short-film festivals, e-books, smart video games, and art auctions gain popularity.
- . TV audiences rise with re-runs and small format productions.
- . Emergence of web channels, podcasts, live streaming, film series, conscious donations, hybrid collaborative productions, crowd funding, and virtual public as alternatives.
- · International travel restrictions led to a focus on domestic tourism as mobility was limited

5 REFLECTION

5.1 Strengths and Weaknesses of BAT as data source

Strengths of the Business Activity Trends dataset:

- Baseline Period Selection: The dataset's choice of a fixed 365-day baseline period before the COVID-19 crisis is a robust approach. It ensures a well-balanced distribution of daily business activity levels, a critical factor for meaningful analysis. In contrast, the prior approach used by EDS, which divided data into 365-day blocks, was prone to comparing daily activities to future days within the same block, potentially leading to misleading results.
- . Removal of Long-Term Trends: Similar to EDS, the dataset effectively eliminates long-term trends and biases that favor highly active businesses. However, due to its real-time nature, it follows a fixed cohort approach where the number of active businesses remains constant during the crisis period. This method is suitable for shorter-term analyses, such as post-disaster business recovery, simplifying the process.
- . Activity Quantile Metric: The dataset introduces the "Activity Quantile" metric, a valuable tool for identifying anomalous business activities following a crisis. It treats all businesses equally, regardless of their posting rates, using a 7-day rolling average. This approach minimizes the impact of outliers and ensures numerical stability, making it a recommended choice.
- . Downtime Detection: The dataset includes a robust methodology for detecting downtime, measuring periods when aggregated activity significantly decreases. It considers variations in posting behavior

- during the week and employs the elbow method to establish threshold values effectively. This approach accurately identifies periods of reduced activity and estimates downtime length.
- . In conclusion, the Business Activity Trends dataset excels in baseline period selection, data transformation, metric robustness, and downtime detection. These strengths make it a valuable resource for understanding business recovery after disruptive events, particularly in the context of the COVID-19 pandemic..

Weaknesses of BAT as a data source:

- . Incomplete Economic Toll: While it effectively measures business downtime and recovery duration, it does not provide a comprehensive assessment of the full economic impact of a crisis. It focuses on activity levels but does not quantify the economic losses or gains incurred by businesses.
- . Limited Real-World Economic Insights: The dataset may not be well-suited for analyzing real-world economic changes, as it primarily focuses on changes in Facebook activity. Economic shifts related to sales, revenue, or overall business health may not be accurately reflected.
- Regional and Sector Variations: Business Facebook usage can significantly differ between regions and sectors. The dataset may not adequately capture these variations, potentially leading to biased or incomplete insights into specific economic niches.
- Fixed Cohort Approach: The use of a fixed cohort approach to account for inactive businesses raises questions about the cause of inactivity. It does not distinguish between businesses that become inactive due to normal business churn and those directly impacted by the COVID-19 crisis, potentially introducing ambiguity into the analysis.
- . Interpretation Sensitivity: While the "Activity Quantile" metric is straightforward to interpret, it can be sensitive to businesses that post frequently and may yield less stable results when dealing with businesses that have low post counts. It is recommended when interpretability is the primary concern, but it may not be the most suitable choice for all analytical scenarios.

5.2 Enhancing BAT Dataset Analysis with Additional Data Sources:

- . Economic Indicators: Include GDP, unemployment, consumer spending, and industrial data for a holistic view of economic impacts during disruptions.
- Geospatial Insights: Utilize location-based data, demographics, and infrastructure details to understand regional variations in business adaptability.
- . Industry-Specific Metrics: Sector-specific data like retail sales and manufacturing output complements BAT data, offering nuanced sector-specific insights.
- . Consumer Behavior Analysis: Study consumer actions, online shopping trends, and sentiment shifts to correlate with business activities.
- . Government Policy Information: Explore stimulus measures, lockdown data, and vaccination efforts to assess their impact on business recovery.
- . Supply Chain Analytics: Access data on supply chain disruptions, logistics, and inventory levels to explain business activity fluctuations.
- Financial Health Indicators: Incorporate financial data like credit scores and business finances to identify vulnerable or resilient businesses.

6 CONCLUSION

- In conclusion, the analysis conducted through the Facebook Business Activity Trends (BAT) dataset has provided valuable insights into the dynamic landscape of business activities during the unprecedented disruptions caused by the COVID-19 pandemic. The robust methodology, including the innovative metrics like Activity Quantile and Activity Percentage, has allowed for a nuanced understanding of how businesses across different sectors and regions have fared in the face of adversity.
- The report highlights the strengths of the BAT dataset, including its careful baseline period selection, metric robustness, and effective downtime detection. However, it also acknowledges certain limitations, such as the dataset's focus on Facebook activity, which may not fully capture the broader economic impact.
- To enhance future analyses, the report suggests incorporating additional data sources, such as economic indicators, geospatial insights, industry-specific metrics, consumer behavior analysis, government policy information, and supply chain analytics. This multi-dimensional approach could provide a more comprehensive understanding of the economic landscape during disruptive events.
- Moreover, the report emphasizes the significant role of a country's GDP in determining the frequency of downtime and its resilience. Understanding the economic context through GDP data can further refine the analysis and contribute to more targeted strategies for recovery and resilience.
- Overall, the findings presented in this report contribute to guiding strategies for economic revival and resilience. By leveraging the BAT dataset alongside complementary data sources and considering GDP variations, businesses and policymakers can make informed decisions to navigate challenges and foster recovery in the post-pandemic era.

7 BIBILOGRAPHY

- (1) https://dataforgood.facebook.com/dfg/resources/business-activity-trends-methodology-paper
- (2) https://ourworldindata.org/covid-stringency-index
- (3) https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker
- (4) https://data.harvestportal.org/dataset/oxford-govt-stringency/
- (5) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10015542/
- (6) https://pec.ac.uk/blog/brazilian-culture-and-creativity-during-covid-19-solidarity-and-lives
- (7) https://www.travelagentcentral.com/destinations/norway-announces-new-nordic-wellness-journey-through-scandinavia
- (8) https://ig.ft.com/coronavirus-lockdowns/