

Data For Good: Business Activity

Trends For Disaster Recovery

Vedant Adawadkar - 730042744 (va296@exeter.ac.uk)
November 2023

Contents

1	Introduction	3
2	Basic Stats and Metrics	4
2.1	Count the number of countries/regions being mentioned in the data set	4
2.2	Count the total number of dates available for each country/state (or equivalent geo-area), in terms of the type of business. Describe missing values/duplicates (if there are any), and how you deal with anomalies (if there are any).	4
2.3	There are two metrics in the data: ‘activity_quantile’ and ‘activity_percentage’. Look at the metadata of the data set and write your understanding about these two metrics, e.g., how are they calculated (using equations), and what do they mean in the real world?	6
2.4	Choose 5 different countries (preferably from different continents), and plot time-series figures for your chosen type(s) of business. Comment on what you see.	8
2.4.1	Restaurant Activity Quantiles and Percentages	9
2.4.2	Retail Activity Quantiles and Percentages	12
2.4.3	Travel Activity Quantiles and Percentages	15
2.5	Are there any other patterns you could find from the data - e.g., weekdays vs. weekends? You could propose other discoveries. Comment on what you see.	18
2.5.1	Comparing weekdays and weekends	18
2.5.2	Understanding distribution of different business verticals using Boxplots of Australia, Germany, India, United Kingdom, United States	22
2.5.3	Other observations	25
2.6	Visualisation. Please choose two days and visualise different countries/states (depending on the gadm_level of the data set) with the business level (preferably using colormaps), in those two days.	26
3	Events	29
3.1	Identify changes. Identify the obvious changes in the business trends across different types of business. You could work in more countries if you want to!	29
3.1.1	Restaurant Business Vertical	30
3.1.2	Retail Business Vertical	33
3.1.3	Travel Business Vertical	36
3.2	What do you discover when using different metrics? Please make comments.	38

1 Introduction

This report provides a concise analysis of the Meta Data For Good dataset, formerly associated with Facebook, focusing on the impact of disasters such as floods, earthquakes, hurricanes, and the recent global pandemic on businesses of varying scales [5] [9]. Recognizing the significant repercussions of these disasters is crucial for devising strategies to mitigate their effects, benefiting both businesses and government agencies.

Traditional methods, like surveys, are often employed to study disaster effects, but they may lack consistency, efficacy, and timeliness due to their resource-intensive nature [7]. In the digital landscape, Eyre, De Luca, and Simini propose leveraging the digital footprint of businesses to identify post-disaster effects and mitigation strategies, especially in terms of economic recovery. The COVID-19 pandemic disrupted businesses globally, with small enterprises in low-income countries facing disproportionate challenges. Studies indicate heterogeneous impacts, with some U.S. businesses experiencing substantial losses while others reported net-zero or positive impacts, emphasizing the need to consider varied government policies affecting different sectors [7].

This report specifically delves into data collected during the COVID-19 pandemic, examining global business activity trends across sectors. Conducted by researchers at the University of Bristol, the analysis aims to provide insights that can inform global-scale strategies. Subsequent sections detail the data collection process, the metrics utilized, their calculation methods, and offer observations based on the provided dataset.

2 Basic Stats and Metrics

2.1 Count the number of countries/regions being mentioned in the data set

The dataset encompasses 12 columns, namely gadm_id, gadm_name, gadm_level, gadm0_name, gadm1_name, gadm2_name, country, activity_quantile, activity_percentage, crisis_ds, ds, and business_vertical, collected between March 1, 2020, and November 29, 2022.

The geographical scope is based on polygons from the Database of Global Administrative Area (GADM) at the country level, given the global nature of the COVID-19 pandemic. GADM comprises three polygon levels, with level 0 representing country-level data. The dataset exclusively focuses on GADM level 0 data, with 216 countries identified by their gadm0_id (3-character ISO alpha-3 country code). The associated gadm0_name provides the country name, and the country column contains the 2-character ISO alpha-2 country code.

The dataset introduces gadm_level, specifying the GADM level of each data point, while gadm1_name and gadm2_name are present but left blank, a topic to be explored in section 2.2.

Additional columns include activity_quantile and activity_percentage, explored in section 2.3, capturing quantitative and percentage aspects of activity. The crisis_ds column signifies the crisis start date (YYYY-MM-DD), with the baseline period being the 365 days preceding this date. The ds column represents activity dates in YYYY-MM-DD format, according to Pacific Time.

Finally, the dataset incorporates the business_vertical column, delineating business sectors internally defined within Facebook. Examples include Business and Utility Services, Grocery and Convenience Stores, Home Services, Lifestyle Services, Local Events, Manufacturing, Professional Services, Public Good, Restaurants, Retail, Travel, and 'All,' signifying all businesses in a specific country.

2.2 Count the total number of dates available for each country/state (or equivalent geo-area), in terms of the type of business. Describe missing values/duplicates (if there are any), and how you deal with anomalies (if there are any).

The dataset spans from March 1, 2020, to November 29, 2022, encompassing 1004 dates for each of the 216 countries and 12 business verticals. Notably, there are no duplicate values within the dataset.

While the dataset includes gadm1_name and gadm2_name columns for GADM levels 1 and 2, respectively, the global focus of the data results in the availability of only GADM level 0 data. Consequently, gadm1_name and gadm2_name columns, having no relevant values, are omitted from the dataset as they play no role in calculations or visualizations.

Additionally, it was observed that some rows lacked values in the country column. In these instances, null values in the country column are substituted with the corresponding gadm_id column value.

Below is the code snippet used for dealing with null columns and empty data point values in the data set:

```

import pandas as pd
from os import walk

# Set pandas display options
pd.options.display.max_columns = None
pd.options.display.max_rows = None

# Read files from dataset
dir_path = ''
filename_list = []
for (dirpath, dirnames, filenames) in walk('Datasets'):
    filename_list = filenames
    dir_path = dirpath

# Create an empty dataframe
df_dfg = pd.DataFrame(
    columns=['gadm_id', 'gadm_name', 'gadm_level',
             'gadm0_name', 'gadm1_name', 'gadm2_name', 'country',
             'business_vertical', 'activity_quantile',
             'activity_percentage', 'crisis_ds', 'ds'])
data_list = []

# Read the dataset from DFG and represent the data
for file_name in filename_list:
    df = pd.read_csv(dir_path + '//' + file_name)
    data_list.append(df)

# Populate the dataframe with all values from the csv file
df_dfg = pd.concat(data_list)

# Set index to the dataframe
df_dfg['ds'] = pd.to_datetime(df_dfg['ds'])
df_dfg['crisis_ds'] = pd.to_datetime(df_dfg['crisis_ds'])
df_dfg.columns = df_dfg.columns.str.strip()
df_dfg.set_index('ds', inplace=True, drop=False)

# Replace empty country data points with gadm_id data points
df_dfg['country'] = df_dfg['country'].fillna(df_dfg['gadm_id'])

# Drop all empty columns where none of the rows have any

```

```
# value in them for the respective columns
df_dfg.dropna(axis=1, how='all', inplace=True)
```

- 2.3 There are two metrics in the data: ‘activity_quantile’ and ‘activity_percentage’. Look at the metadata of the data set and write your understanding about these two metrics, e.g., how are they calculated (using equations), and what do they mean in the real world?**

The raw data collected for measuring business activity involves counting the number of Facebook posts made by each business from March 1, 2020, to November 29, 2022. This encompasses all content authored by Facebook business page administrators, including posts on the business page, in Facebook groups, on the Facebook feed, stories, and reels. Each post is accorded equal weight in the metrics. The COVID-19 crisis is assigned an immediate 365-day baseline period preceding the crisis event date, serving as a reference for "normal" business behavior or activity when there is no crisis [7].

The activity quantile is computed by comparing the business activity on a specific day with its baseline period. This comparison helps understand current performance relative to historical performance, providing insights into trends and anomalies. The calculation methodology, inspired by Eyre, De Luca, and Simini, treats each gadm record-business vertical combination as a cell, producing data at that cell level. Initially, an activity mid-level quantile value for a business is calculated, considering the daily post count and comparing it with the distribution of daily post counts during the baseline period. Subsequent transformations are applied to produce a cell-specific daily activity quantile, ranging from 0 to 1. Values close to 0.5 indicate normal business behavior, while extreme values nearing 0 or 1 for extended durations suggest anomalous behaviors. For privacy protection, cells with less than ten business pages are excluded from the dataset. The quantile data is generated for the 'all' business vertical [7].

The calculation steps for activity quantile follow the notations of Eyre, De Luca, and Simini's methodology, with minor differences. Here, the baseline period is fixed at 365 days for the COVID-19 pandemic, and a fixed-cohort method is assumed, where the number of businesses remains constant across days during a given crisis, without adjusting for linear trends.

The mid-quantile is determined using the Probability Integral Transformation (PIT) Metric under non-crisis circumstances, assuming a standard uniform distribution.

1. $x_i(t)$ = Daily post count for business page i and day t .

$P_i(x)$ = The empirical cumulative distribution function for the daily post count during the baseline period.

$P_i(x_i(t))$ = Proportion of days comparative to baseline period where for business page i has $x_i(t)$ fewer posts.

Therefore, mid-quantile is calculated as,

$$q_i(t) = \frac{1}{2}[P_i(x_i(t)) + P_i(x) + 1]$$

2. This is followed by summation of individual page level mid quantile as,

$$r_{PIT}(t) = \sum_{i \in B} q_i(t)$$

. Here, B denotes a set of businesses in a cell.

The summation should result in values following uniform distribution.

3. Perform cell-wise transformation so that values follow standard normal distribution,

$$r_N(t) = \frac{r_{PIT}(t) - \frac{n(t)}{2}}{\sqrt{\frac{n(t)}{12}}}$$

Here, $n(t)$ is representative of number of businesses in an individual cell

4. Adjustments are then made for identifying correlation in posting behaviors,

$$\tilde{r}_N(t) = \frac{r_N(t)}{\sigma}$$

Here, σ is representative of standard deviation of $r_N(t)$ calculated within a baseline period for each cell.

5. Probability Integral Transformation metric is then applied followed by which average of last 7 days is taken which results in values between 0 and 1 as follows,

$$r_U(t) = P_N(\tilde{r}_N(t))$$

It's crucial to emphasize that the following equations are directly drawn from the research paper provided by the University of Bristol researchers, as cited in [7].

In contrast to the activity quantile, the activity percentage operates on a scale not limited to 0 to 1 but typically extends into the hundreds. Businesses with activity percentages close to 100 are considered to have normal activity levels. While this metric facilitates easier interpretation of activity levels, it is often less reliable due to its sensitivity to the number of posts made by businesses. This metric is especially prone to inaccuracies when dealing with businesses that have a lower number of posts. Consequently, it is generally preferred when interpretability is more critical.

To calculate the activity percentage, a 7-day rolling sum of a business's total posts is expressed as a percentage of the average weekly baseline. Each weekly baseline average is determined as the 7-day sum of total activity for each week, starting from Monday during the baseline period. The 7-day rolling sum of total posts is then divided by the weekly baseline average and multiplied by 100.

2.4 Choose 5 different countries (preferably from different continents), and plot time-series figures for your chosen type(s) of business. Comment on what you see.

To analyze and interpret various activity levels over time, the report focuses on five countries—Australia, Germany, India, the United Kingdom, and the United States. Additionally, three business verticals—Travel, Retail, and Restaurant—are selected to identify trends in activity levels for the years 2020, 2021, and 2022. These verticals are chosen due to potential correlations among them.

Both activity quantile and activity percentage metrics are employed to study the time-series data for the chosen countries and verticals. For each vertical, a time-series is plotted, featuring the five countries. The objective is to comprehend how the COVID-19 pandemic may have impacted the selected sectors throughout the specified years. The dual use of metrics provides a comprehensive perspective on business activity trends, allowing for a nuanced analysis of the data.

2.4.1 Restaurant Activity Quantiles and Percentages

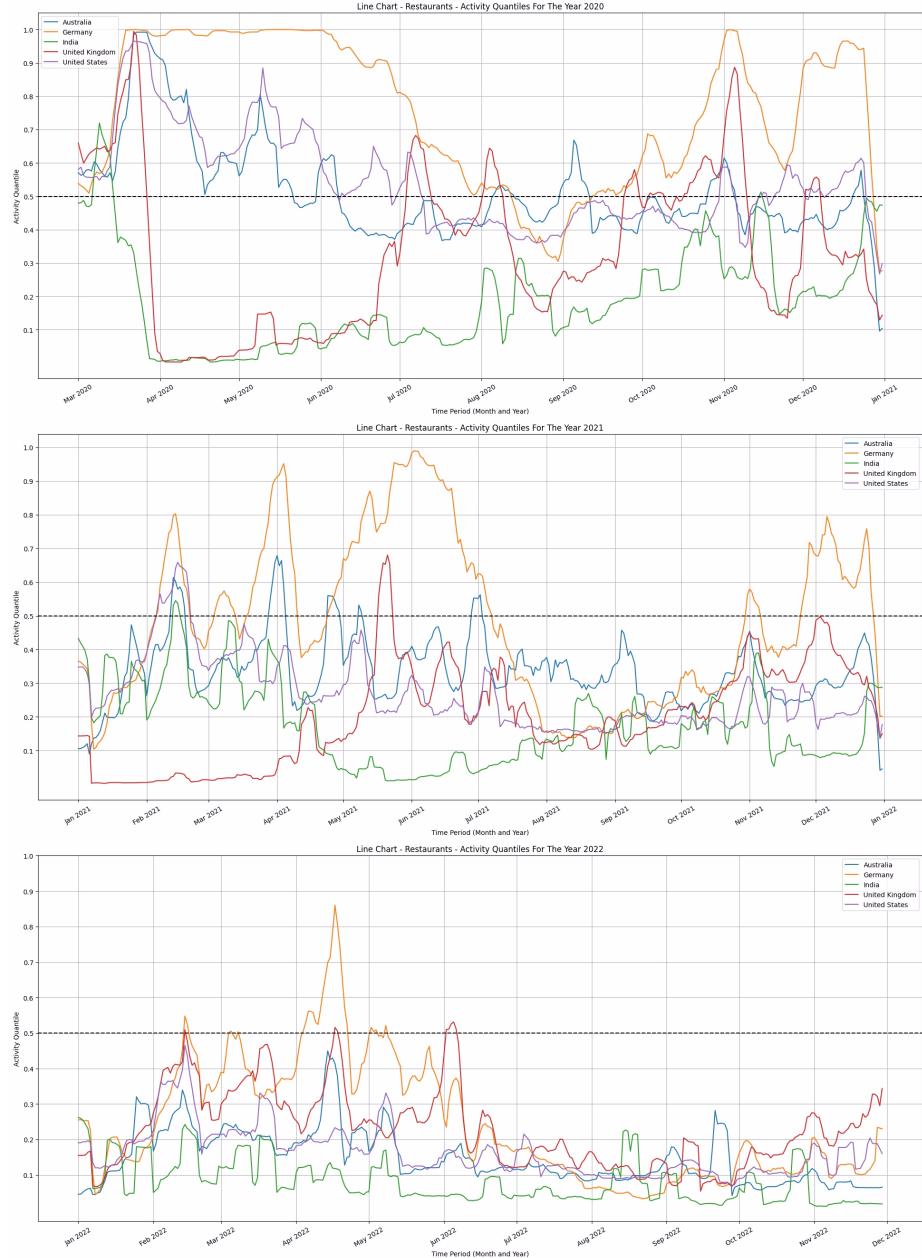


Figure 1: Restaurant Activity Quantiles

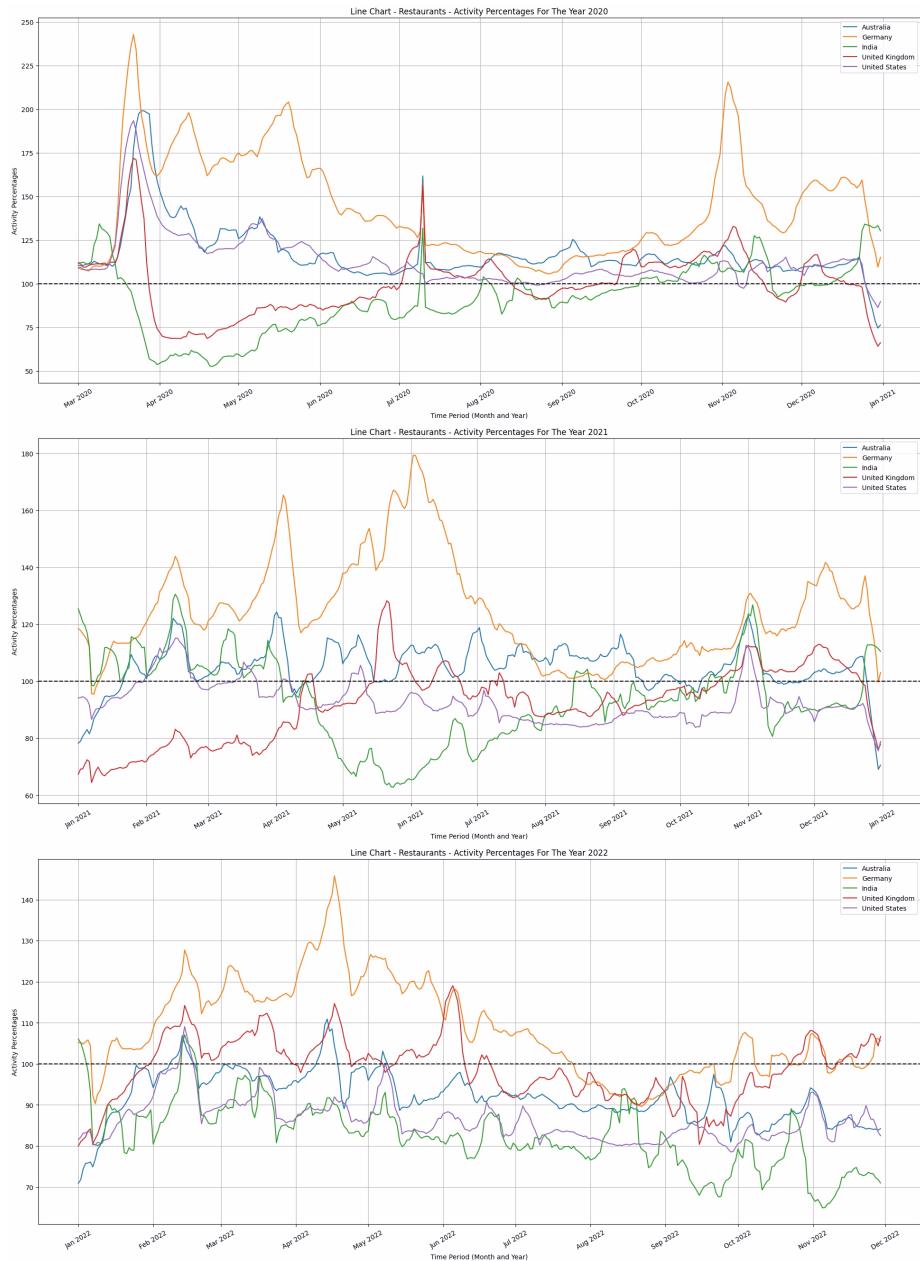


Figure 2: Restaurant Activity Percentages

The initial observation highlights a significant surge in restaurant business activity levels across various countries, occurring promptly after the implementation of government-mandated lockdowns aimed at curbing the spread of the COVID-19 pandemic. This surge reflects the immediate response of businesses as they navigated the challenges posed by the pandemic and the ensuing lockdown measures.

After the initial lockdown in March 2020, both the United States and Australia experienced a significant drop in restaurant activity levels, hovering between 0.7 and 0.3, with a notable spike in May. In contrast, Germany and the UK exhibited unstable activity quantile levels, prompting the use of activity percentages for interpretation in 2020. In Germany, restaurant activity witnessed fluctuations throughout the year, with spikes surpassing 200% in March, April, May, and November. Similarly, the UK displayed varying values throughout the year, reaching a peak just above 150% and a low below 75%. Although there was a brief period in the UK where values approached 100%, resembling normal business activity, it can be considered an outlier due to the persisting strict restrictions in September 2020 [1]. Conversely, India experienced a steady increase in restaurant activity levels after the decline following the March 2020 lockdown, culminating in levels approaching near 0.5 by the end of 2020.

In 2021, activity levels remained relatively stable in Australia and the US. Germany witnessed significant fluctuations in February, April, and June 2021, with activity quantile levels nearing 1. Conversely, India and the UK saw a drop in activity levels to values below 0.1 in May 2021 and January 2021, respectively.

Moving into 2022, all countries, except Germany, exhibited stable activity levels below 0.5. Germany reached a value near 1 in April 2022, possibly attributed to the substantial relaxation of restrictions in the country [3] [4] [6] [10].

2.4.2 Retail Activity Quantiles and Percentages



Figure 3: Retail Activity Quantiles



Figure 4: Retail Activity Percentages

Similar to the restaurant business vertical, the retail sector also experienced the announcement of lockdown measures in March 2020. While activities in 2020 remained relatively stable for countries like Australia, Germany, the UK, and the US, with minimal fluctuations beyond the initial spike, India presented a volatile retail sector. In India, activity levels oscillated between below 0.1 and near 1 on the activity quantile scale. A notable spike occurred in July 2020, reaching nearly 250%, around 150% higher than the norm. This surge can be attributed to India's "unlock period" during July 2020. The United Kingdom also exhibited a similar spike in July 2020 in terms of activity percentage, reflecting some relaxation lockdown restrictions [3] [4] [6] [10].

In the subsequent year, India and Germany demonstrated fluctuating activity quantile levels, varying above and below 0.5 throughout the year. However, the activity percentage values for these countries consistently remained above 100%. In contrast, Australia, the UK, and the US displayed a more stable distribution of activity percentages throughout 2021 compared to India and Germany. While the activity quantile metric fluctuated similarly to India and Germany, the values tended to hover closer to 0.5 than those of the two aforementioned countries.

Moving into 2022, activity levels, when measured in percentages, returned to near-normal values for most countries as restrictions were gradually eased. However, activity quantile values for India and Germany consistently remained below 0.5, ranging between 0.2 and 0.4, which can still be considered close to 0.5. Both India and Germany experienced higher business fluctuations, with Germany exhibiting the highest throughout 2022, as reflected in the activity quantiles. Notably, India witnessed a sharp decrease in activity levels at the end of October 2022.

2.4.3 Travel Activity Quantiles and Percentages

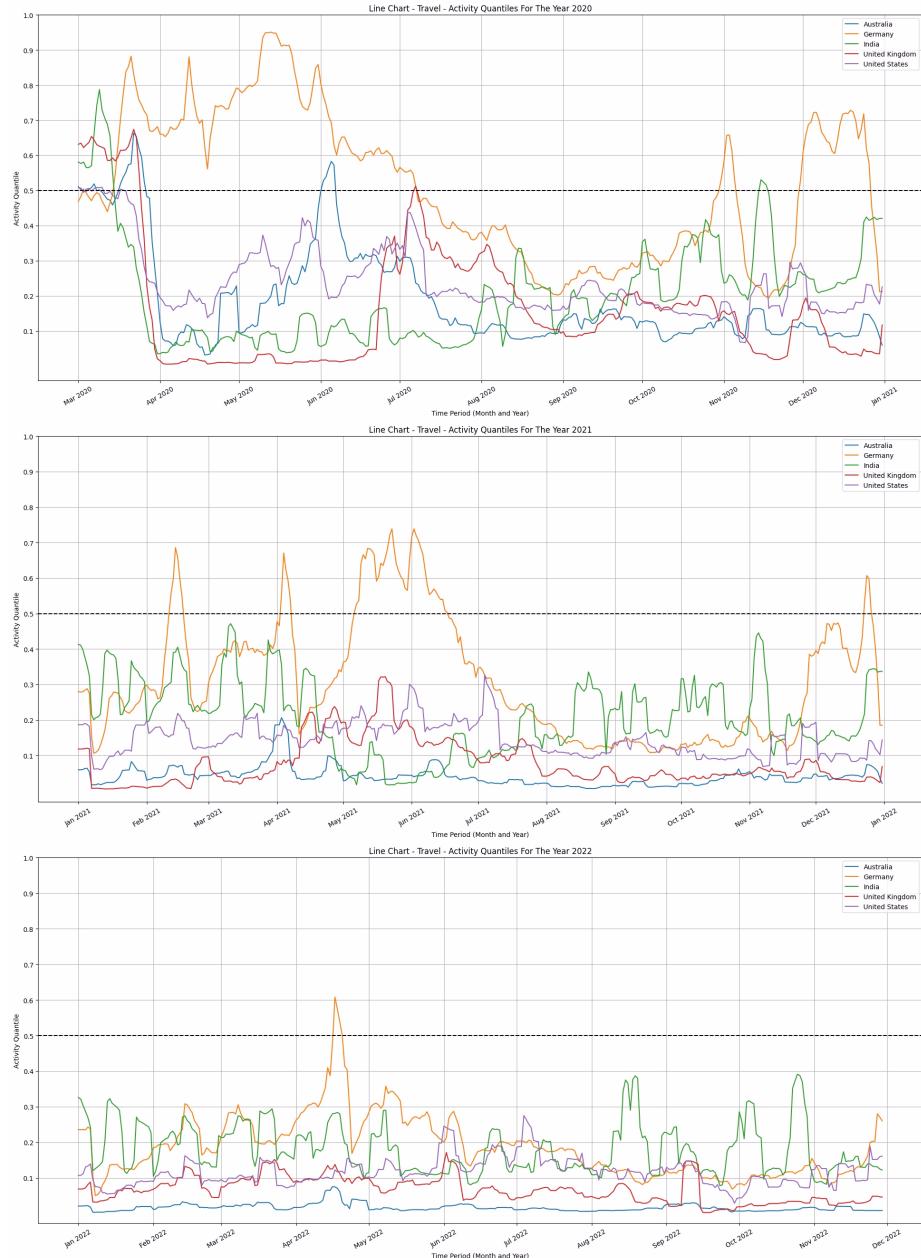


Figure 5: Travel Activity Quantiles



Figure 6: Travel Activity Percentages

In contrast to the restaurant and retail verticals, the travel sector did not experience a pronounced surge in activity for Australia, the US, and the UK in either metric during March 2020. However, India and Germany witnessed a sharp increase in activity during this period, with quantiles surpassing 0.9 and percentages exceeding 140. Another notable spike in activity percentages occurred for India and the UK in July 2020. Subsequently, India's activity levels remained above the 100 percentile mark, while Germany gradually tapered off to near the 100 percentile mark by year-end. The United Kingdom, after a steep decline in travel activity in April 2020, exhibited a sharp recovery in July 2020, followed by activity levels dropping below the 60% mark by the close of the year [3] [4] [6] [10].

In 2021, India exhibited highly unstable travel activity levels, fluctuating above and below the normal 100 percentile mark throughout the year. Germany, on the other hand, began 2021 with relatively normal activity levels but experienced fluctuations, including a sharp spike at the beginning of June 2021.

Moving into 2022, despite the easing of restrictions worldwide, there was only a marginal increase in activity levels, with quantile values for almost all countries remaining close to 0 and 0.3. Germany stood out with a spike in mid-April 2022, reaching an activity quantile of 0.6.

2.5 Are there any other patterns you could find from the data - e.g., weekdays vs. weekends? You could propose other discoveries. Comment on what you see.

2.5.1 Comparing weekdays and weekends

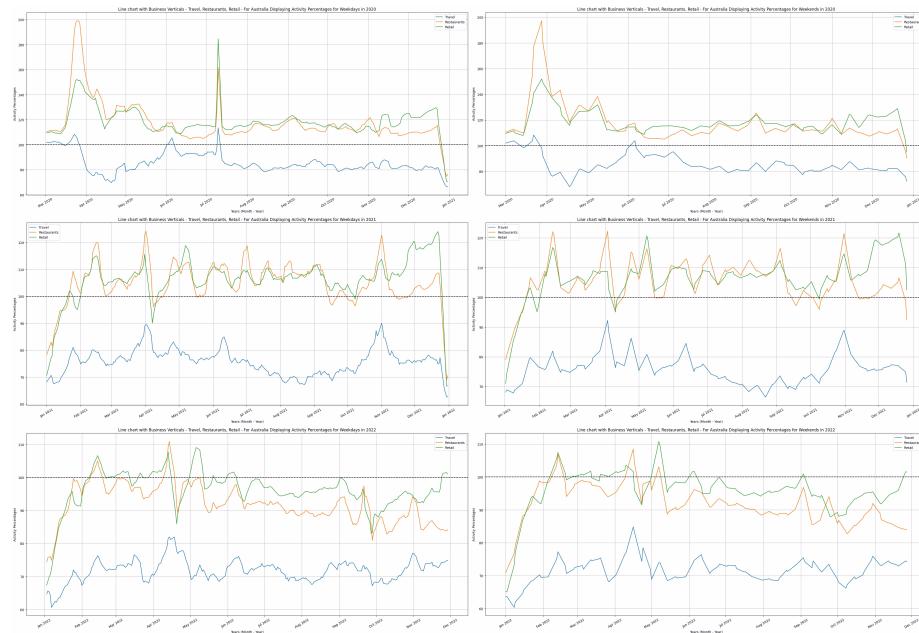


Figure 7: Australia Line chart comparing weekdays and weekends

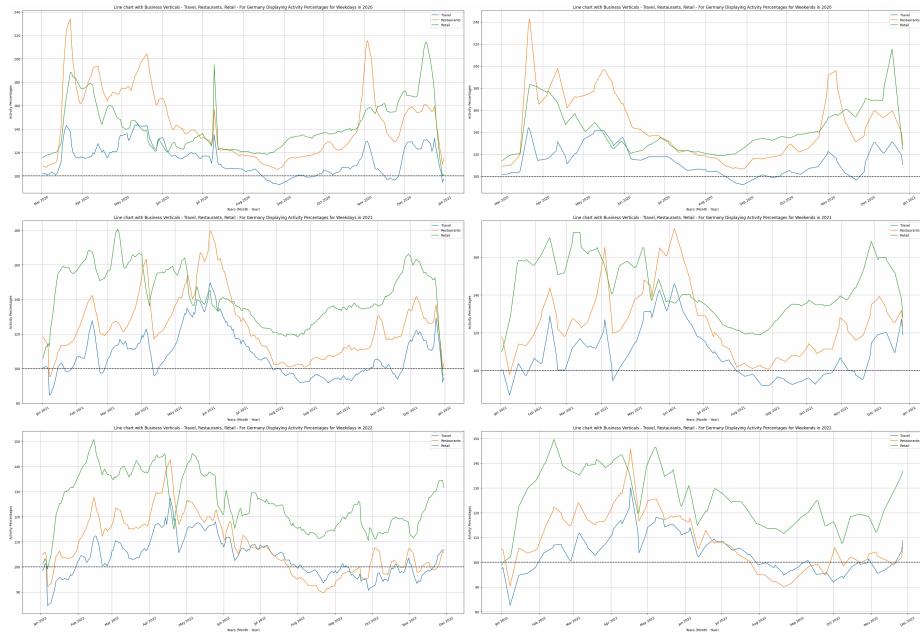


Figure 8: Germany Line chart comparing weekdays and weekends

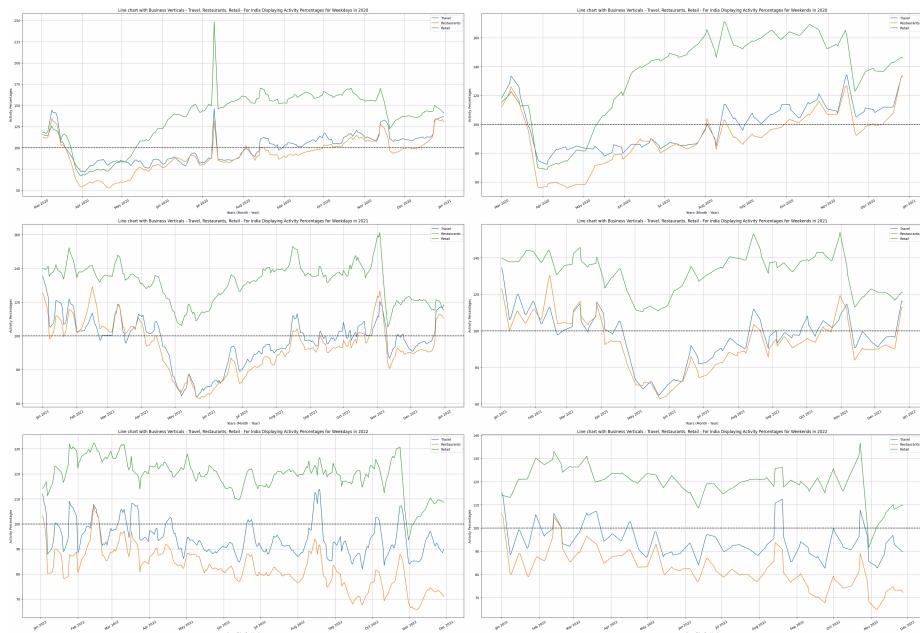


Figure 9: India Line chart comparing weekdays and weekends

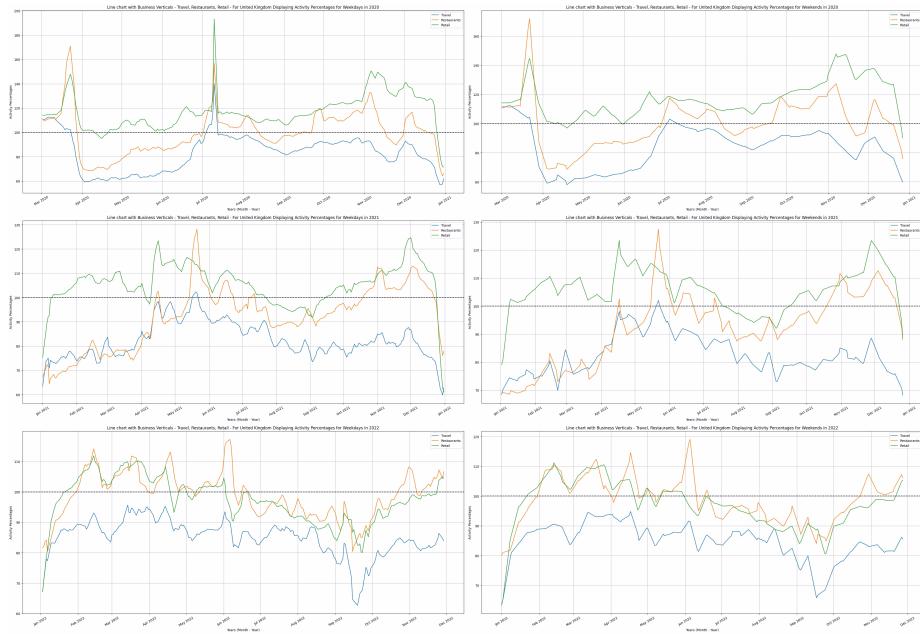


Figure 10: United Kingdom Line chart comparing weekdays and weekends

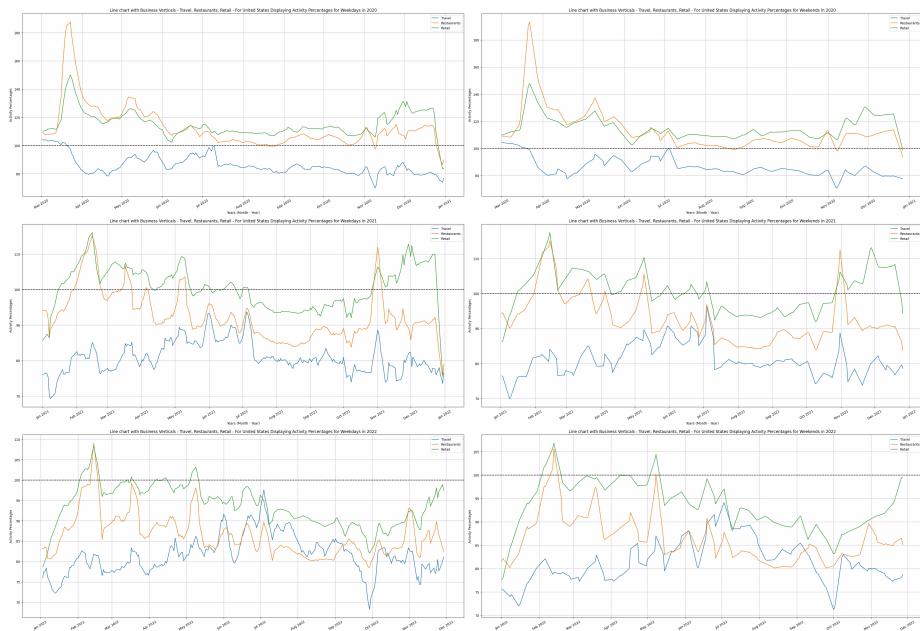


Figure 11: United States Line chart comparing weekdays and weekends

The provided figures illustrate the activity percentages for Restaurant, Retail, and Travel business verticals in Australia, Germany, India, the United Kingdom, and the United States. Upon careful observation, it becomes evident that significant fluctuations in activity levels concerning weekdays and weekends are not prevalent in the years 2021 and 2022. The most notable variations are apparent in the initial year for each country. This observation aligns with the fact that during the first year of the pandemic and ensuing lockdowns, countries were navigating diverse strategies to mitigate the spread of the disease while responding to the rapidly evolving global situation [3] [4] [6] [10].

For instance, in Australia, the figure indicates a substantial spike in Restaurant and Retail activity in July 2020, a pattern not mirrored in the corresponding weekend graph for that month. Similarly, Germany's Retail business vertical displays analogous spikes in the weekday graphs for July 2020, which are notably absent in the weekend graphs. India and the United Kingdom exhibit a comparable surge in July 2020, with the Retail business vertical experiencing a notable upswing in activity levels. This behavior in the weekday graphs can be attributed to the relaxation of restrictions in specific regions of Australia, Germany, India, and the UK. Interestingly, the United States does not display such a spike in its activity levels during the mentioned period.

2.5.2 Understanding distribution of different business verticals using Boxplots of Australia, Germany, India, United Kingdom, United States

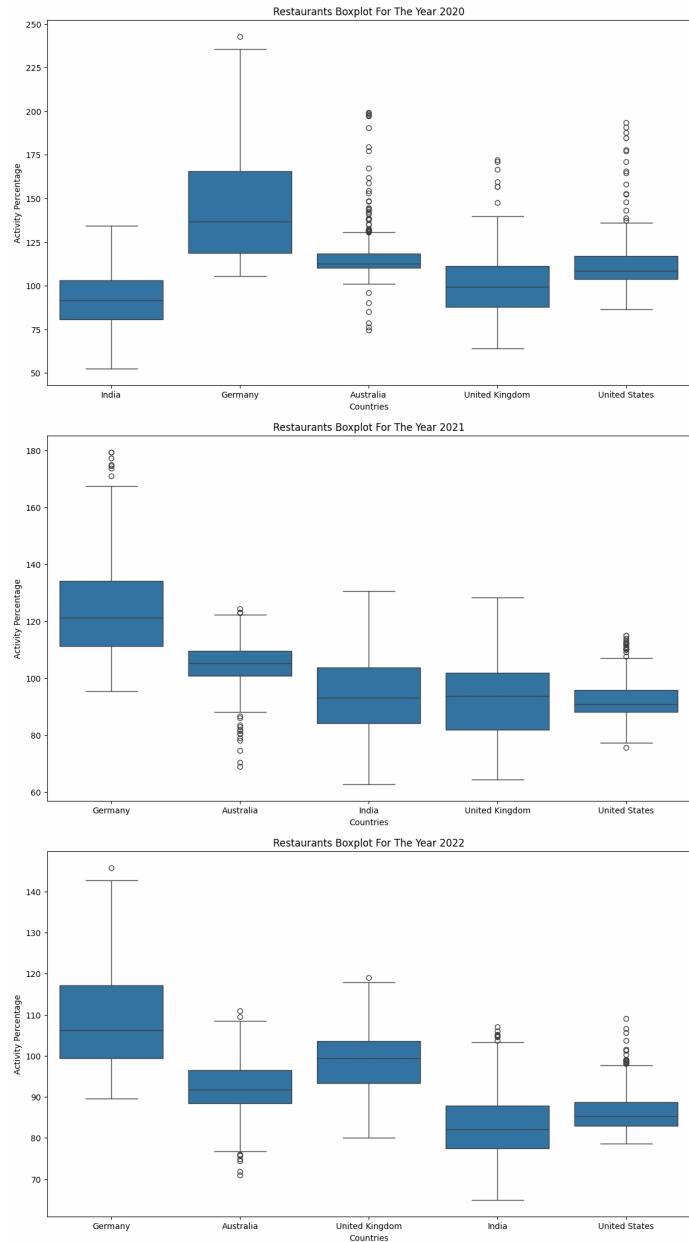


Figure 12: Boxplot of restaurant vertical featuring Australia, Germany, India, United Kingdom and United States with respect to activity percentage

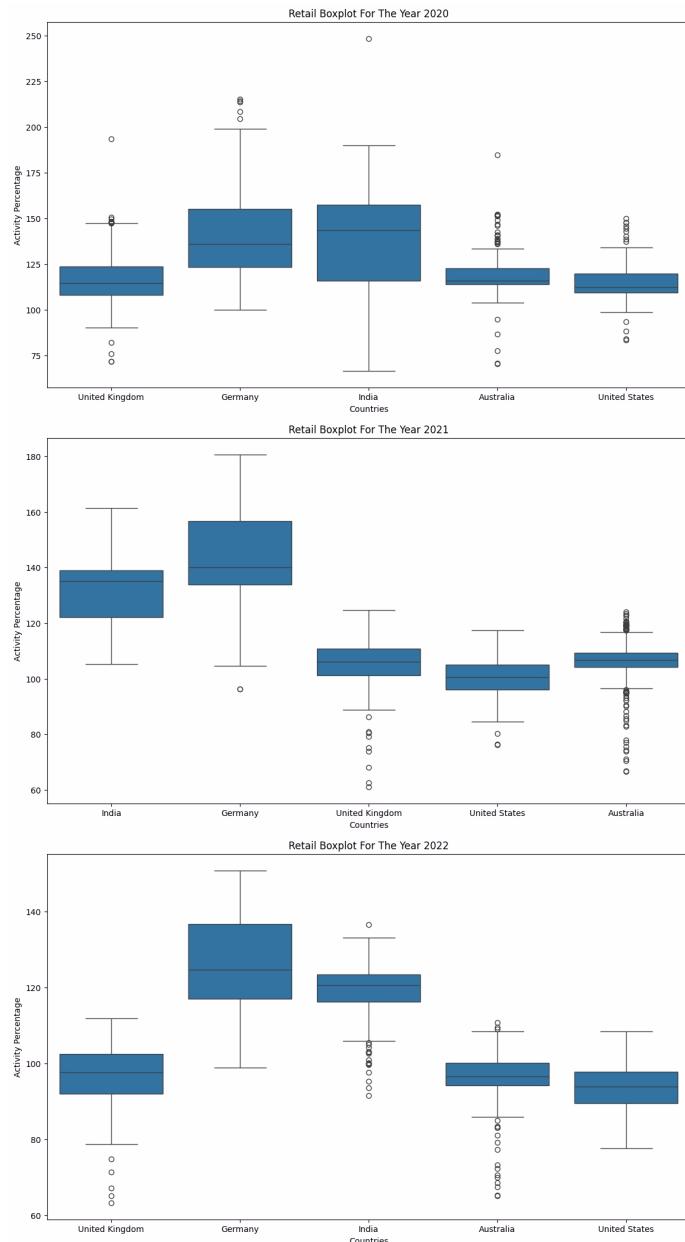


Figure 13: Boxplot of retail vertical featuring Australia, Germany, India, United Kingdom and United States with respect to activity percentage

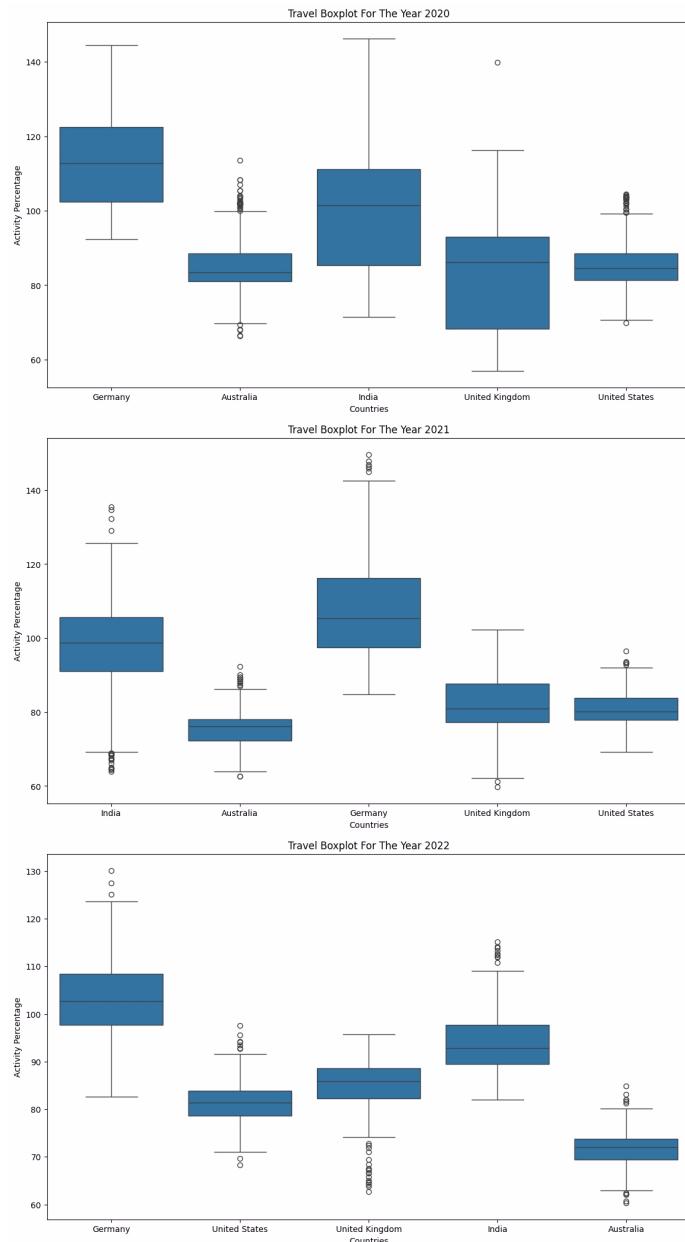


Figure 14: Boxplot of travel vertical featuring Autralia, Germany, India, United Kingdom and United States with respect to activity percentage

Boxplots are usually used for providing a quick summary distribution the data which highlights the spread and the median of that data.

The first box-plot (Figure 12) illustrates the restaurant vertical for Australia, Germany, India, the United Kingdom, and the United States. Across all years, Germany consistently exhibits the highest median activity percentage, while the United States consistently shows the lowest median. This indicates that activity levels varied significantly in Germany, with the least fluctuations observed in the United States over the three years. Additionally, Australia tends to have the highest number of outliers, suggesting both low and high activity level fluctuations, while India has the fewest outliers, indicating less variability. The Inter-Quartile Range (IQR) for Germany is notably wider throughout 2020, 2021, and 2022, indicating greater variation in activity levels compared to other countries.

The second box-plot (Figure 13) represents the retail vertical for the same countries. India consistently has the highest median activity percentage, while the United States has the lowest. Similar to the restaurant vertical, India exhibits more significant fluctuations, and the United States displays the least. Australia consistently has the most outliers, indicating varied activity levels, while Germany has the fewest outliers, suggesting more stability. The IQR for Germany is generally wider, indicating greater variation in activity levels over the three years.

The third box-plot (Figure 14) portrays the travel vertical for the selected countries. Germany consistently has the highest median activity percentage, while Australia has the lowest. Similar to the other verticals, Germany experiences more pronounced fluctuations, and Australia shows the least. India and the United Kingdom consistently have the most outliers, indicating varied activity levels, while Germany consistently has the fewest, suggesting more stability. The IQR for India and the United Kingdom is highest in 2020, while Germany's IQR is widest in 2021 and 2022, demonstrating greater variation in activity levels for those years.

2.5.3 Other observations

Other observations reveal a correlation between travel and restaurant business verticals in India, where their activity levels appear to follow a similar pattern.

2.6 Visualisation. Please choose two days and visualise different countries/states (depending on the gadm_level of the data set) with the business level (preferably using colormaps), in those two days.

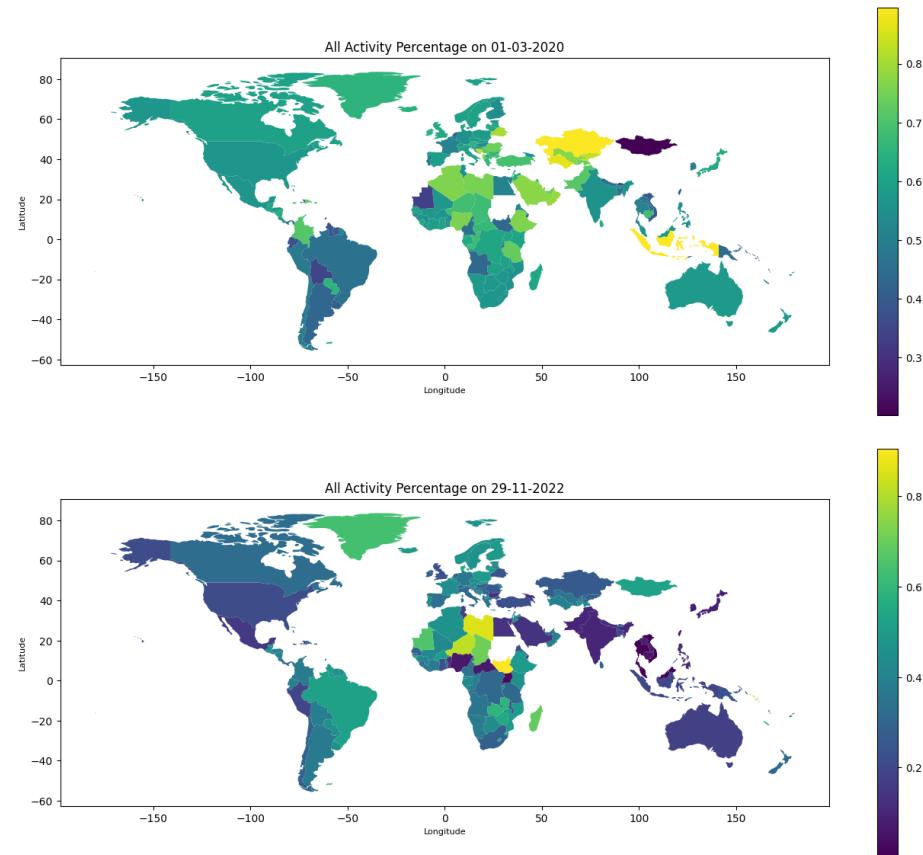


Figure 15: Worldmap Heatmap for 'All' business vertical using Activity Quantiles for March 1, 2020 and November 29, 2022

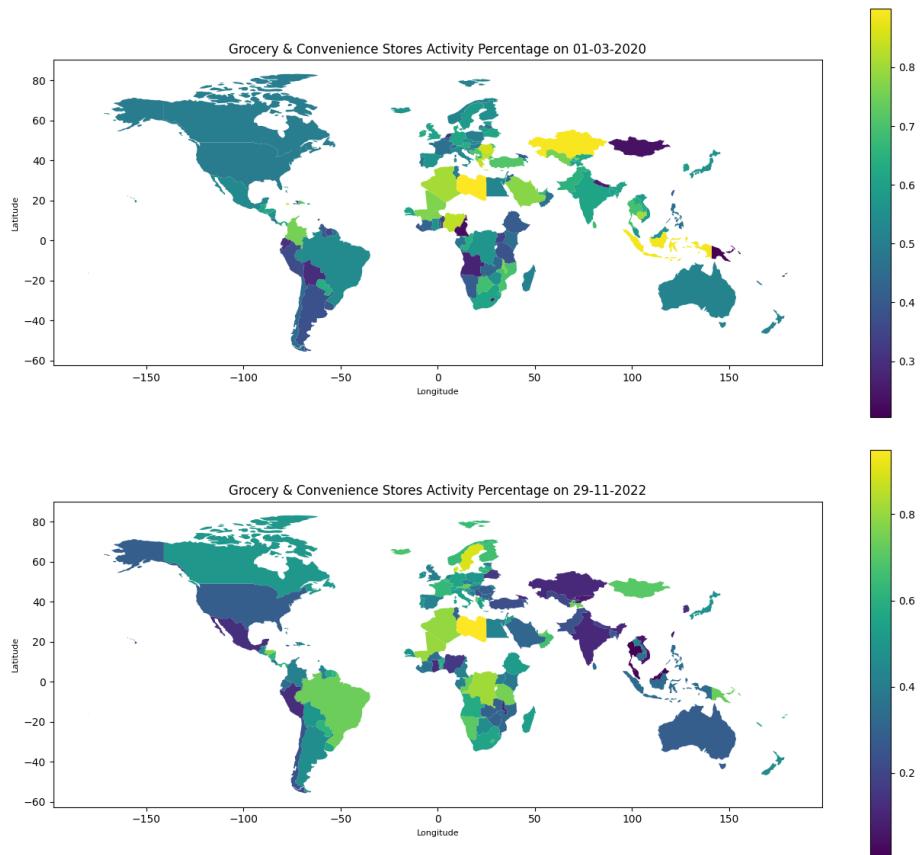


Figure 16: Worldmap Heatmap for 'Grocery and Convenience Store Services' business vertical using Activity Quantiles for March 1, 2020 and November 29, 2022

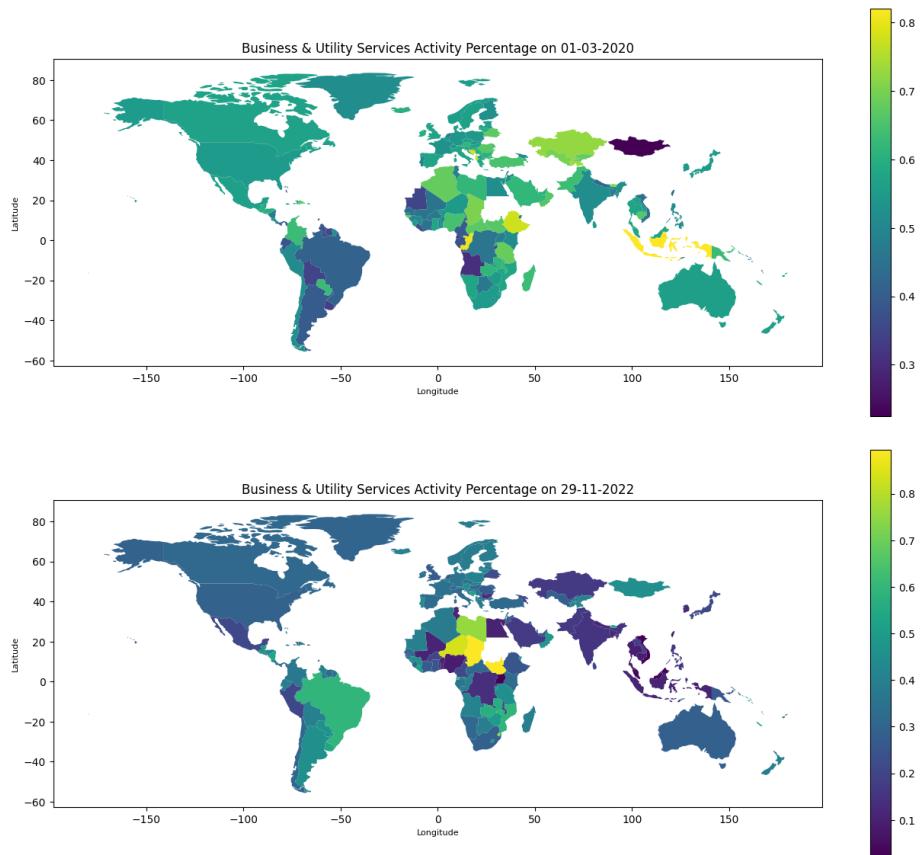


Figure 17: Worldmap Heatmap for 'Business and Utility Services' business vertical using Activity Quantiles for March 1, 2020 and November 29, 2022

3 Events

3.1 Identify changes. Identify the obvious changes in the business trends across different types of business. You could work in more countries if you want to!

This section is divided by business verticals with each figure divided into three graphs - small countries (Iceland and Luxembourg), medium countries (Australia and South Korea), large countries (India and United States). countries are divided into sizes by the population and the varying income of the countries respectively. The report uses 2 countries from each 'size' to have a better representation of the metrics, moreover, the report uses activity quantiles and activity percentages to understand how the two metrics can be used to understand and interpret the activity levels.

3.1.1 Restaurant Business Vertical

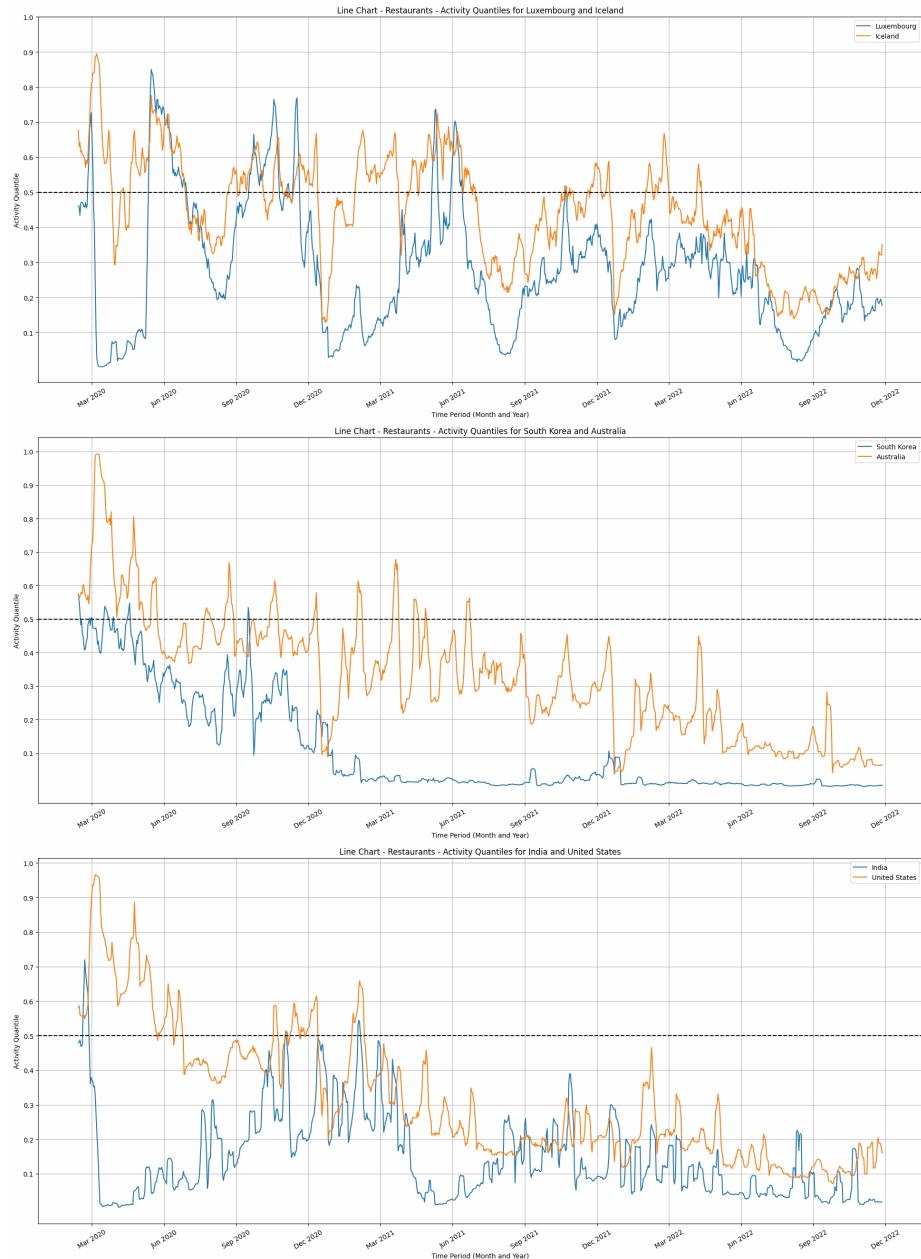


Figure 18: Activity Quantiles for Restaurant Vertical for the selected countries



Figure 19: Activity Percentages for Restaurant Vertical for the selected countries

For the restaurant sector, it is evident that relatively larger countries than Iceland and Luxembourg performed better during the pandemic, especially concerning the restaurant vertical and its associated industries. Activity levels, expressed as a percentage, appear to approach normal acceptable levels after the initial measures implemented by the governments of Australia, South Korea, India, and the United States. Although the values may dip below normal levels, implying stagnation in the vertical after mid-2020, these countries exhibit a less reactive stance compared to Iceland or Luxembourg, despite having larger populations and varying income levels.

India and the United States stand out as more stable than other nations, consistently maintaining values close to 100% throughout the pandemic period. This suggests either a high utilization of the restaurant vertical by the population or a less restrictive approach in handling the industries encapsulated in the restaurant sector when implementing measures.

To gain a clearer perspective on Iceland and Luxembourg, activity quantiles prove insightful, revealing multiple instances where quantile levels significantly drop below normal activity levels. This pattern indicates a fluctuating and reactive nature within the restaurant sector of smaller countries.

3.1.2 Retail Business Vertical

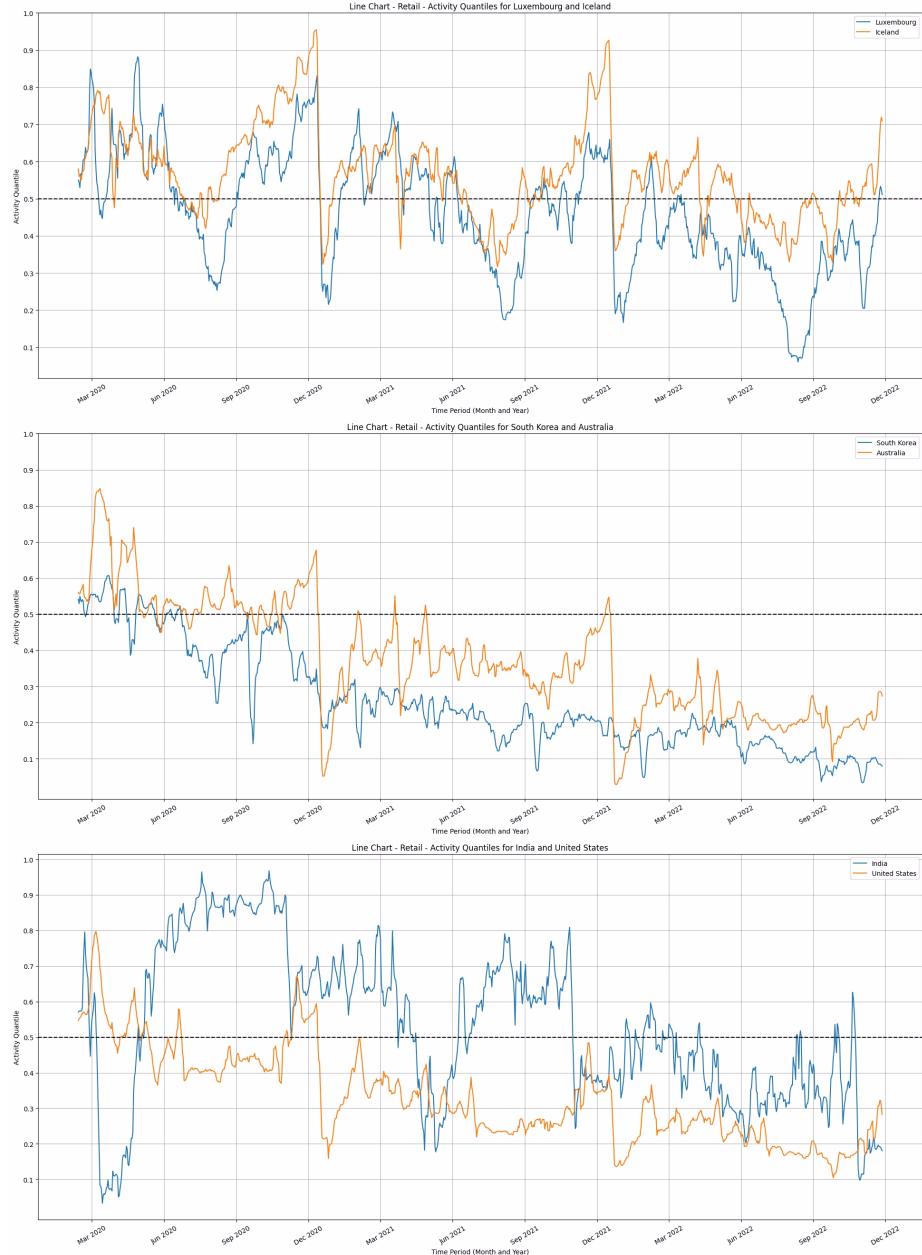


Figure 20: Activity Quantiles for Retail Vertical for the selected countries



Figure 21: Activity Percentages for Retail Vertical for the selected countries

In the context of the retail business vertical, the primary metric for representation is the activity percentage. While all six countries experience initial spikes, each at varying levels, South Korea and Australia emerge as having the most consistently stable vertical throughout the pandemic years, maintaining values close to 100%. In contrast, Iceland and Luxembourg exhibit a less favorable scenario for the retail sector compared to the larger countries of Australia, South Korea, India, and the United States.

The graphical representation illustrates a continuous ebb and flow of activity levels throughout 2020, 2021, and 2022 for Iceland and Luxembourg. This dynamic pattern suggests that the retail sector in these countries is more reactive to the diverse policies and measures implemented, rather than stabilizing itself to proactively adapt to them.

3.1.3 Travel Business Vertical



Figure 22: Activity Quantiles for Travel Vertical for the selected countries

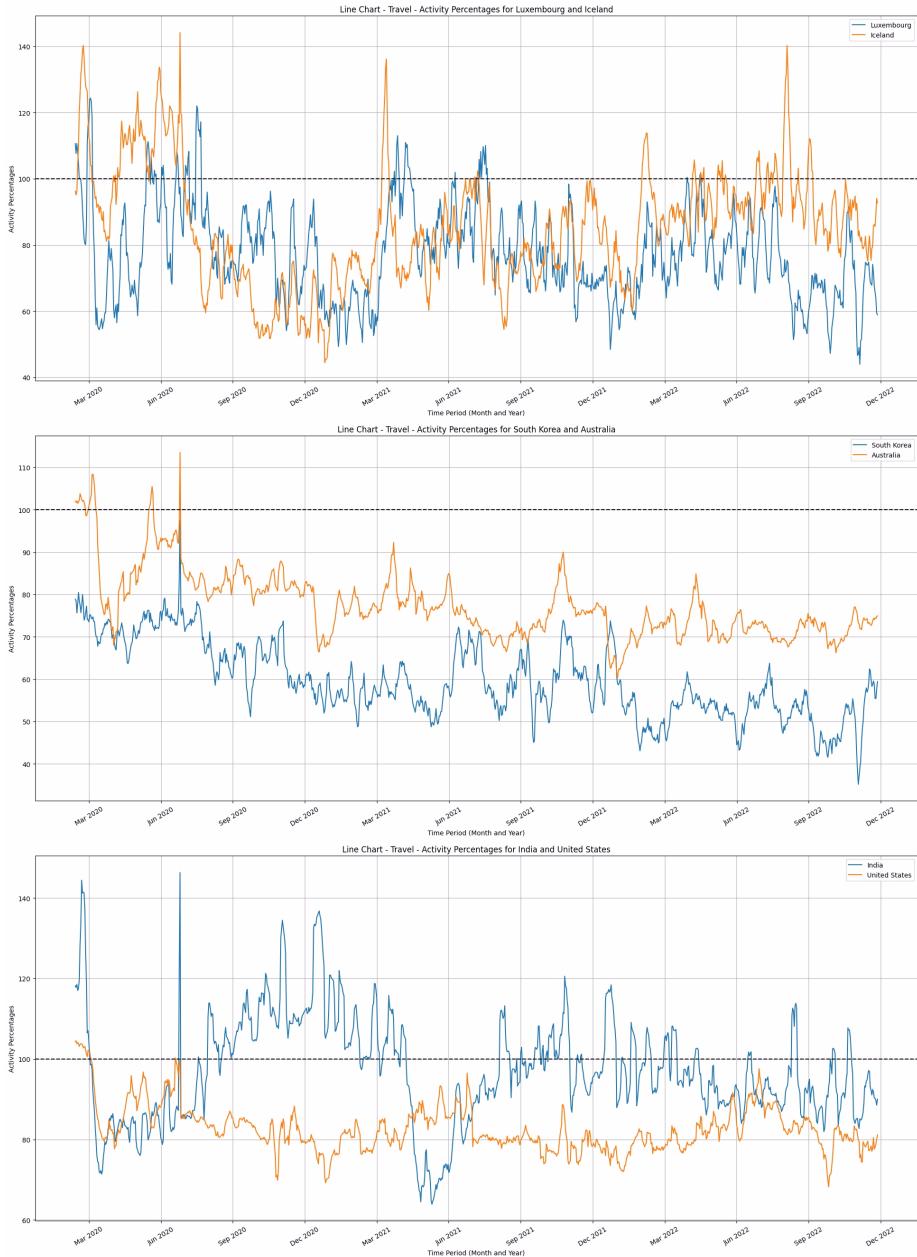


Figure 23: Activity Percentages for Travel Vertical for the selected countries

The travel vertical emerges as the most adversely affected sector across all countries, with South Korea and Australia appearing to bear the brunt of the impact. This suggests that these countries possibly implemented stricter restrictions during the pandemic, particularly due to the known airborne transmission of the disease. Interestingly, the travel vertical in both smaller and larger countries appears to be more accessible, as evidenced by activity levels that, while not reaching the pre-pandemic normal of 0.5, consistently fluctuate near that value over the three-year period.

A similar trend is observable in the activity percentage graph, where Australia and South Korea depict minimal activity in the travel vertical. In contrast, Iceland, Luxembourg, India, and the United States show a continuous rise and fall in activity levels. This pattern indicates a dynamic and fluctuating state within the travel sector, potentially reflecting the varying impact of restrictions and measures across these countries.

3.2 What do you discover when using different metrics? Please make comments.

The activity percentage metric proves valuable for interpreting activity levels, yet its reliability is susceptible to the volume of posts by businesses. Despite this, it remains a preferred choice for interpretation. In contrast, activity quantiles offer a clearer insight into the impact on business activity levels. Capped between 0 and 1, these quantiles compare daily business activity to baseline periods, facilitating an understanding of performance relative to historical benchmarks.

Throughout the report, instances arise where the two metrics may conflict. For example, the activity percentage for a specific country and business vertical may exhibit varying ranges in different years, while the activity quantile consistently remains within a range of 0 to 1. This potential discrepancy poses challenges in analysis, particularly without prior knowledge or thorough research on a specific country.

3.3 Try to understand the causes of these changes - please look for relevant policies and other events that help you understand these changes. Or reversely, how the business trends have enabled/disabled some policies to some extent. Use online information wisely! Comment on what you discover.

Analyzing the causes of these changes involves examining relevant policies and events that influenced fluctuations. Notably, a significant change across studied countries occurred with the initial implementation of lockdown rules in response to the virus spread. This initial spike is evident in March 2020 across various business verticals [1] [8] [4].

Subsequently, fluctuations occurred, particularly in the first half of 2020, as businesses adapted to government measures. For instance, in July 2020,

certain countries, including India, began easing lockdown procedures, only to later impose stricter rules, leading to another spike [6] [4] [1].

In India, sectors forecasted to perform well experienced uncertainty, especially as the country shifted between relaxed and strict lockdowns. This uncertainty impacted online food delivery services significantly, leading to a steep decline in business activity. Conversely, the travel sector, hit hard globally, saw innovations such as contactless boarding due to the pandemic [2].

During the pandemic, grocery and convenience stores experienced higher activity levels, particularly at the start of lockdowns, when panic buying surged. The travel sector, especially international travel, bore the brunt of restrictions, experiencing a severe decline.

The analysis suggests a return to normal activity levels for verticals like restaurants, retail, and travel by the end of 2022. Larger countries, such as the United States, India, and the UK, lifted lockdowns and normalized business operations by December 2022.

4 Reflection

The activity percentage metric proves valuable for interpreting activity levels, yet its reliability is influenced by the volume of posts from businesses. Despite this limitation, it remains a preferred choice for interpretation. In contrast, activity quantiles offer a clearer insight into the impact on business activity levels. Capped between 0 and 1, these quantiles compare daily business activity to baseline periods, providing an understanding of performance relative to historical benchmarks. However, activity quantiles may pose readability challenges, requiring readers to possess prior knowledge of the topic or data points.

Throughout the report, instances arise where the two metrics may present conflicting results. For instance, the activity percentage for a specific country and business vertical may display varying ranges in different years, while the activity quantile consistently stays within a range of 0 to 1. This potential discrepancy poses challenges in analysis, especially in the absence of prior knowledge or thorough research on a specific country. Understanding and reconciling such disparities require a nuanced approach to interpretation and a deeper exploration of the underlying factors influencing these metrics.

While this dataset enables the analysis of verticals in terms of different activity levels from 2020 to 2022, it falls short in comprehensively capturing the broader economic impact of the pandemic on different countries. For a more thorough analysis, the inclusion of revenue data before, during, and after the pandemic for various sectors would be invaluable. This additional information could help assess the extent to which growth or shrinkage in these sectors was caused by the pandemic.

Furthermore, a comprehensive analysis and accurate predictions necessitate more in-depth historical and present data, which the current data set lacks. Many countries implemented localized restrictions during the pandemic, with variations in the stringency of measures across regions. The absence of this data hinders a complete understanding of how the measures affected the entire country, leaving the analysis incomplete.

The data set offers a bird's eye view of the situation, providing an overview rather than a detailed picture that would elucidate the lasting effects of the COVID-19 pandemic or potential implications for the future, especially in the event of a similar pandemic.

References

- [1] timeline-coronavirus-lockdown-december-2021.
- [2] Alex Upadhyay @ Mymusiclife. Impact of Covid-19 on Indian economy, 7 2021.
- [3] Anna Sherrington. 2 Years of COVID-19 on GOV.UK, 7 2022.
- [4] Cassidy Knowlton. A timeline of Covid-19 in Australia, two years on, 2 2023.
- [5] S. E. Chang, C. Brown, J. Handmer, J. Helgeson, Y. Kajitani, A. Keating, I. Noy, M. Watson, S. Derakhshan, J. Kim, and A. Roa-Henriquez. Business recovery from disasters: Lessons from natural hazards and the COVID-19 pandemic. *International Journal of Disaster Risk Reduction*, 80, 10 2022.
- [6] David J. Sencer CDC Museum: In Association with the Smithsonian Institution. CDC Museum COVID-19 Timeline, 3 2023.
- [7] P. Lam, S. Iyer, E. Giraudy, and A. Pompe. FACEBOOK BUSINESS ACTIVITY TRENDS FOR DISASTER RECOVERY. Technical report, 2022.
- [8] Ralf Bosen and Jens Thurau. Chronology: How COVID has spread in Germany, 12 2021.
- [9] G. Samantha. The Impact of Natural Disasters on Micro, Small and Medium Enterprises (MSMEs): A Case Study on 2016 Flood Event in Western Sri Lanka. In *Procedia Engineering*, volume 212, pages 744–751. Elsevier Ltd, 2018.
- [10] P. Soni. Effects of COVID-19 lockdown phases in India: an atmospheric perspective. *Environment, Development and Sustainability*, 23(8):12044–12055, 8 2021.