INTERNATIONAL MEDICAL UNIVERSITY
MALAYSIA

MASTER IN HEALTH DATA COLLECTIONS

Intake : MHIA124

HIA 322 GROUP PROJECT

TITLE :

EXPLORING THE ASSOCIATION OF MYSEJAHTERA APPS EFFECTIVENESS WITH COVID-19 CASES IN MALAYSIA (Jan 2020 – June 2022)

KHAIRUL IZWAN BIN MOHAMED @ HASHIM 00000047728

CHOW TAK KUAN 00000044142

AN JIFENG 00000047312

CONTENT

1.0 Introduction

The COVID-19 pandemic has affected the whole wide world. It brought challenges that required countries to adapt quickly with creative solutions to manage and control the spread of the novel virus. In Malaysia, the MySejahtera app has been one of the key tools in this effort. It was launched in April 2020, designed to help the public and authorities to track COVID-19 cases, facilitate contact tracing, manage vaccinations, and allow check-ins at locations through QR codes. These features made it essential for monitoring and managing daily activities during the pandemic.

As the pandemic emerged into national crisis, Movement Control Order (MCO) was introduced in March 2020 to enforce travel restrictions and limiting assemblies across Malaysia to minimize the virus spread. The MCO was introduced in phases followed by a National Recovery Plan to gradually reopen the country. The phases of MCO are as follows:

1. MCO: 18th March 2020 – 3rd May 2020
2. Conditional MCO: 4th May 2020 – 9th June 2020
3. Recovery MCO: 10th June 2020 – 31st March 2021
4. MCO by states: 13th January 2021 – 31st May 2021
5. Total Lockdown: 1st June 2021 – 28th June 2021

In April 2020, Malaysia launched the MySejahtera application, a mobile application, to help authorities and health officials monitor population movements and identify potential hotspots and clusters. This link between mobility data and rising or falling COVID-19 cases during various MCO phases provides an opportunity to study the patterns and impacts of these measures.

This study aims to explore the relationship between MySejahtera check-ins and the number of COVID-19 cases reported, with a focus on how these patterns were influenced by the different MCO phases. By looking into the data, this research hopes to reveal how public movement and interactions contributed to the spread of the virus. It also seeks to evaluate whether digital check-in systems like MySejahtera can serve as effective tools for managing and predicting outbreaks in the future.

By understanding how MySejahtera data ties into case trends and MCO policies, this study could offer valuable insights for future public health planning. Identifying key trends, areas at higher risk, and patterns in case surges may help authorities respond more effectively and plan targeted interventions in the future. A descriptive analysis of the data has been collected and analyzed. The attributes of the data are listed below:

1. **date**: yyyy-mm-dd format; data correct as of 2359hrs on that date
2. **checkins**: number of checkins at all locations registered on MySejahtera
3. **unique_ind**: number of unique accounts which checked in
4. **unique_loc**: number of unique premises checked into
5. **i**: in the time density file, checkins are aggregated by half-hour buckets, giving 48 in total; bucket i corresponds to the ith half-hour slot of the day. for instance, i = 0 corresponds to 0000 - 0029; i = 31 corresponds to 1500 - 1529.
6. **cases_new**: cases reported in the 24h since the last report
7. **cases_import**: imported cases reported in the 24h since the last report
8. **cases_active**: Covid+ individuals who have not recovered or died
9. **cases_recovered**: recovered cases reported in the 24h since the last report
10. **cases_cluster**: number of cases attributable to clusters; the difference between cases_new and the sum of cases attributable to clusters is the number of sporadic cases
11. **cluster_x**: cases attributable to clusters under category x; possible values for x are import, religious, community, highRisk, education, detentionCentre, and workplace
12. **cases_agecat**: cases falling into one of 4 age categories, i.e. child (0-11), adolescent (12-17), adult (18-59), elderly (60+); note that the sum of cases by age may not equal the total cases for that day, as some cases are registered without ages or with unverifiable age data
13. **cases_pvax**: number of partially-vaccinated individuals who tested positive for Covid (perfect subset of cases_new), where "partially vaccinated" is defined as receiving at least 1 dose of a 2-dose vaccine at least 1 day prior to testing positive, or receiving the Cansino vaccine between 1-27 days before testing positive
14. **cases_fvax**: number of fully-vaccinated who tested positive for Covid (perfect subset of cases_new), where "fully vaccinated" is defined as receiving the 2nd dose of a 2-dose vaccine at least 14 days prior to testing positive, or receiving the Cansino vaccine at least 28 days before testing positive

1.1 Data source

The data for this study was obtained from the official **COVID-19 Malaysia GitHub repository**, which previously and currently still maintained by the Ministry of Health (MOH) Malaysia. This repository provides comprehensive and publicly accessible datasets on COVID-19 cases, testing, vaccinations, and related statistics. The data includes daily updates on the number of reported cases, recoveries, deaths, and other metrics, as well as information categorized by states and districts.

For this analysis, the dataset utilized was last accessed on 8th December 2024. The specific files used include "cases_malaysia.csv", "cases_state.csv", "checkin_malaysia.csv", "checkin_state.csv", "trace_malaysia.csv" and "clusters.csv". These CSV files provide detailed information regarding MySejahtera total check-in per day for the whole country and states, and detailed information regarding Covid-19 cases daily update.

The GitHub repository is available at: https://github.com/MoH-Malaysia/. All data was cleaned, processed, and analysed in accordance with ethical standards, and no personal or identifiable information was used.

1.2 Objective

Research question:

- What is the check-in pattern recorded via MySejahtera app associated with the number of Covid-19 daily new cases in Malaysia?
- How does the MCO affect the reporting of new daily Covid-19 cases in the whole country and the states?
- How effective is the MySejahtera app in identifying the geolocation of new clusters or outbreaks?
- Can check in data trend by MySejahtera app being used as an early indicator for prevention of contagious cases?
- How can the check-ins data collected be improved the current MySejahtera app to better predict and manage future pandemic?

Objective:

- To determine the association of daily check-in data by MySejahtera app with that of reported new daily Covid-19 cases
- To determine the association of daily check-in data by MySejahtera app with reported daily new unvaccinated Covid-19 cases

- To determine the significant association of detecting new hotspot or cluster of Covid-19 cases via MySejahtera check-in pattern

2.0 Methodology

The study aims to fulfill the objectives, which mainly to examine the relationship of daily MySejahtera app's usage and the Covid-19 trend cases. The data collected and used was during the period of 1^{st} December 2020 till 11^{th} June 2022. This period was selected as it was commonly available in all the files and covers MCOs 2 and 3 as well as the entire National Recovery Plan.

2.1 Data collection

Usage of MySejahtera app became crucial nationwide. The check-ins data were collected upon entering facilities and places were obtained from official Ministry of Health, Malaysia thru reliable data source.

2.2 Data preprocessing

Data integration

Upon gathering the data collected from the data source, there were 6 comma separated value (CSV) files intended to use for this study. Initially all the CSV files were standardized in terms of the date format. Then, we identified the period of dates consists of row to use as the data. Unintended data was excluded from the main merged CSV files. Two merged CSV files created focusing on national check-ins data usage with Covid-19 trends and states check-ins data as well. We aligned the data points from different sources to ensure consistency in the merged data files created.

Data cleaning

As the merged files came from different format, we addressed missing values and duplicated data. All the processes were done via phyton codes, ran either from Spyder software or Jupyter Notebook via Anaconda. We found that missing values were reported at the Covid-19 cluster data, probably the data was not recorded in the initial phase before the case surge. All missing values were already filled with appropriate values. 558 data rows of check-ins data were taken for this study.

Ethical considerations

In this study, data privacy of all MySejahtera app user were not affected. No private data were exposed. The data source also mentioned that the data was available for data processing. Necessary credit already mentioned in this study document.

Tools and Software

Data analysis and calculations were made using python codes via Spyder and Jupyter Notebook. Apart from the base library, we also employed various libraries in Python for statistical analysis and visualizations such as pandas, numpy, Matplotlib, Tableau or statsmodel.

2.3 Descriptive Analysis of Malaysia's scenario

The dataset contains **559 observations** (date in rows, starting from 1st December 2020) and spans of **85 variables column**. There were no missing values after data cleaning done. These variables include numerical data, such as daily MySejahtera check-ins, counts of unique individual check-ins (unique_ind), location-based check-ins (unique_loc), Covid-19 case numbers and its trend, and Covid-19 cluster-related variables. Additionally, the dataset features a date column that helps track trends over time.

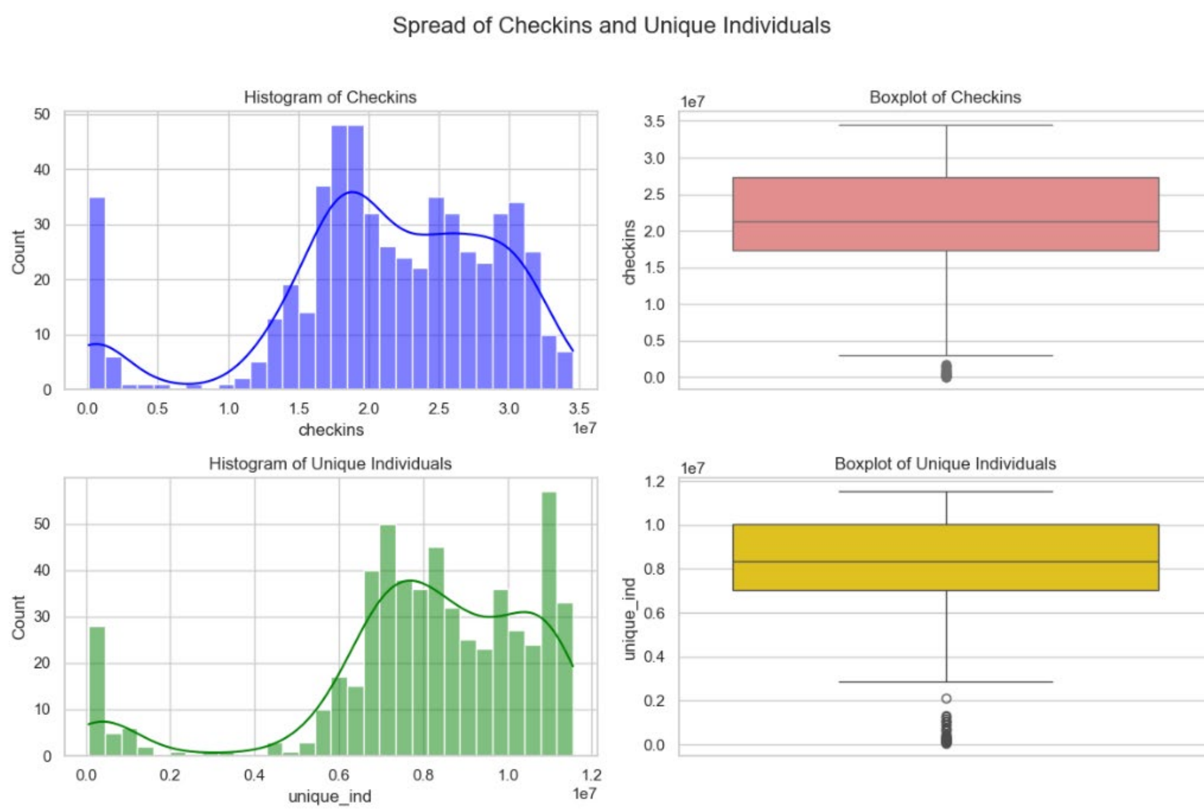
To gain an understanding of the dataset, we examined the descriptive statistics for key variables. From Table 1.0, the check-ins variable has an average of approximately 21 million daily check-ins, with a wide variability as reflected by its standard deviation of over 8 million. The minimum number of daily check-ins recorded was 71,660, while the maximum reached a staggering 34.5 million.

Table 1.0 : Descriptive statistics for daily MySejahtera check-ins, unique individual check-ins, and unique location check-ins.

	Mean	Std Dev	Min	25%	Median	75%	Max
Daily check-ins	29,096,470	8,082,974	71,660	17,370,543	21,277,0389	17,334,417	34,546,300
Unique-individual check-in	8,048,036	2,734,497	64,666	7,026,666	8,340,927	10,033,049	11,555,205
Unique-location check-ins	647,926	196,738	29,048	576,424	671,537	787,073	893,008
Daily new Covid-19 case	523	346	50	-	410	-	410
Case recovered	450	310	30	-	370	-	370

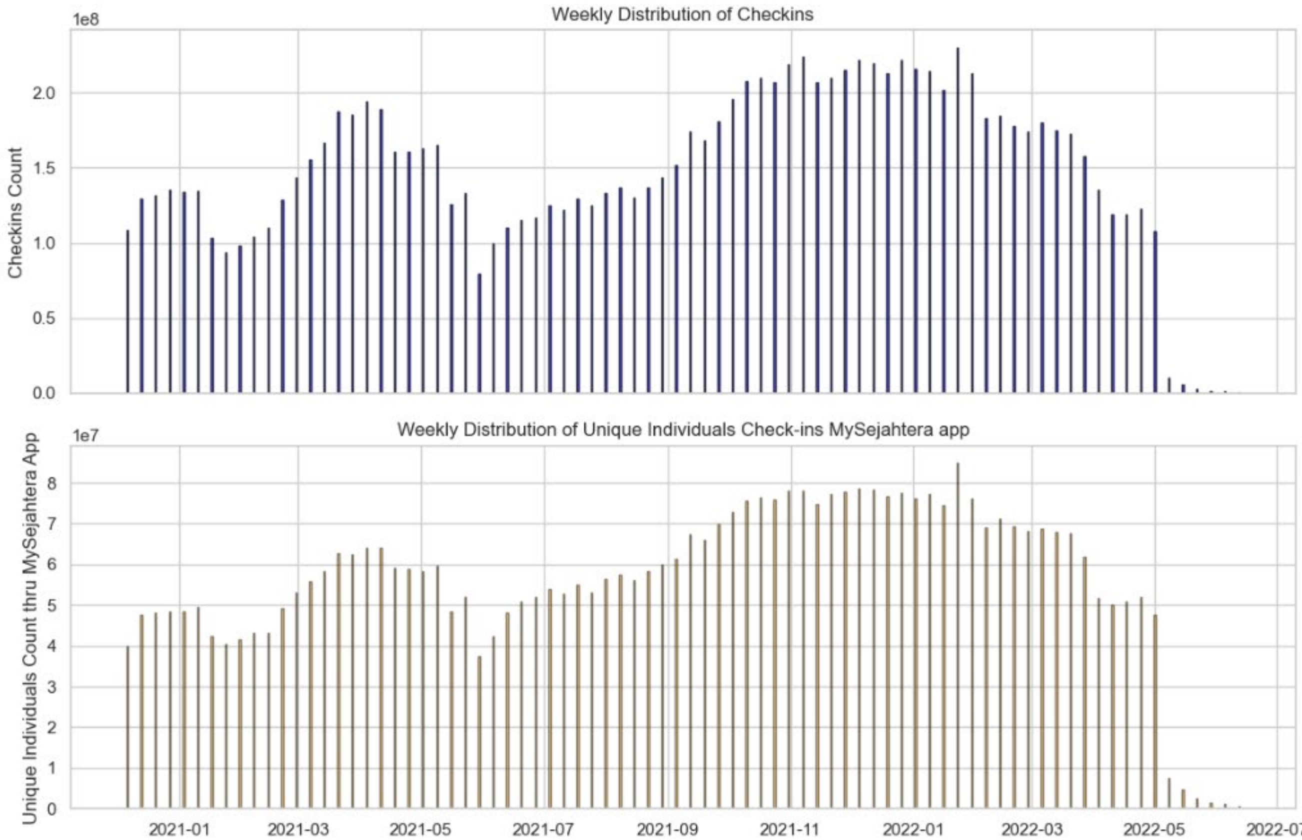
Similarly, the unique_ind column (Table 1.0), representing individual users, had a mean of 8 million individuals daily, with values ranging between 64,666 and over 11.5 million. Such variation highlights substantial fluctuations in daily activity levels, likely influenced by external events or specific user behaviors.

When analyzing the distribution of checkins and unique_ind, both variables showed a right-skewed pattern (Figure 1.0). Most observations cluster around their mean values, but there are several high outliers. This suggests that while typical daily user activity remains stable, certain days experienced a significant surge in check-ins and user counts, which may correspond to special events or specific interventions.

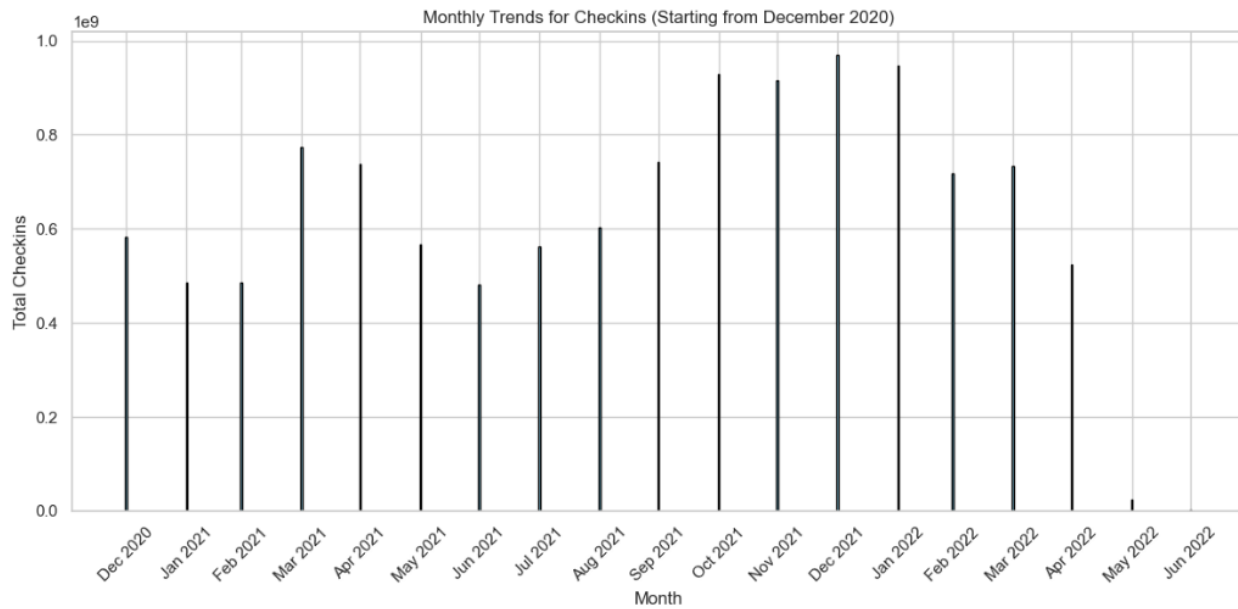


Graph 1: Trends of MySejahtera check-in and its count frequency; shown in histogram with smooth curve and boxplot.

To explore long-term trends, we aggregated the data weekly and monthly starting from 1st December 2020. Both trends analysis revealed fluctuations in check-ins, with noticeable peaks during particular months. These peaks suggest higher engagement periods that could align with external factors like occurrence of MCOs, surge or declined of Covid-19 cases, holidays, policy changes, or other time-specific influences. The visualization of these trends clearly shows a dynamic pattern, where engagement levels are not uniform over time but instead follow a wave-like movement (Figure 1.1 & Figure 1.2).



Graph 1.1: Histogram of weekly distribution of MySejahtera app check-ins and unique-individuals check-ins.



Graph 1.2: Histogram of monthly distribution of MySejahtera app check-ins.

Beyond the checkins and unique_ind variables, several other fields such as cases_new, cases_recovered, cases_active, and specific case counts by age groups were analyzed to uncover patterns.

The dataset also includes cluster-related variables, such as cluster_workplace, cluster_community, cluster_education, and cluster_highRisk. These columns highlight sources of infection spread. For example, cluster_workplace and cluster_community likely represent common transmission points due to their frequent interaction potential. The variables show relatively low values overall but may spike during localized outbreaks, reflecting an increased focus on testing and tracing efforts in specific clusters. Similarly, columns like cluster_import and cluster_religious reflect niche sources of case clusters, though their frequency in the dataset appears limited.

2.4 Descriptive Analysis of States and Territories

The dataset provides an overview of 13 states and 3 territories in Malaysia, detailing check-ins, unique individuals, locations, active COVID-19 cases, and new cases. It tracks daily check-ins, the number of distinct individuals involved, and the proportion of individuals relative to the population. It also includes data on the number of unique locations visited and the current and daily increases in COVID-19 cases across the different regions. This data is presented in Table 2.0.

Table 2.0 : Average number of Check-ins, unique individual users, unique check-in locations, active and new cases of Covid-19 from 1st Dec 2020 to 11 Jun 2022.

State	Check-in per day	Unique individual	Unique Individual (%)	Unique Location	Active Cases	New Cases
Johor	2302362.35	1053294.33	26.19	80575.15	8782.55	687.23
Kedah	786808.36	415801.82	19.31	32398.49	5184.75	534.51
Kelantan	301059.43	168817.09	9.30	18371.99	4351.24	443.83
Melaka	581772.75	275643.94	27.42	20224.85	2825.70	223.05
Negeri Sembilan	694733.97	344591.13	2860	22817.81	3997.95	360.03
Pahang	805028.52	357453.46	22.28	28880.00	3720.13	307.33
Perak	1324160.65	601139.52	23.91	49060.89	3568.94	363.93
Perlis	203130.68	121816.84	42.30	8237.56	357.03	31.24
Pulau Pinang	1372577.49	623215.30	35.81	43779.85	4562.00	502.58
Sabah	1354646.80	502348.01	14.72	44993.61	8703.66	616.34
Sarawak	1453895.87	556055.10	22.54	41413.77	6284.53	548.26
Selangor	4885762.92	2324556.93	33.10	131455.73	29197.24	2437.53
Terengganu	333083.89	177310.16	15.11	16196.30	2268.12	225.97
W.P Kuala Lumpur	3603656.92	1699940.06	86.55	76993.43	7903.19	635.96
W.P. Labuan	63367.65	26360.61	27.41	1729.68	350.99	35.97
W.P Putrajaya	192953.94	117648.30	102.04	2256.50	482.82	37.33

Overall, the check-in data reveals notable regional variations that are influenced by factors such as population size, urbanization, tourism, and economic activity. High check-in numbers in more urbanized or economically active regions suggest higher levels of mobility and engagement. For instance, Selangor exhibits the highest number of check-ins per day, with 4,885,762.92, which is likely due to its large population and urbanized environment. Similarly, W.P. Kuala Lumpur, with 3,603,656.92 check-ins, mirrors this trend, indicating a high concentration of activity in the federal capital. Conversely, Perlis shows the lowest check-in frequency with 203,130.68 per day. This lower number may reflect the state's smaller population size and less urbanized nature compared to other regions. Other states, such as Johor (2,302,362.35) and Sabah (1,354,646.80), report moderate check-in rates. Johor’s check-ins possibly reflects its proximity to Singapore and its status as a commercial and tourism hub in the later stages of the lockdown.

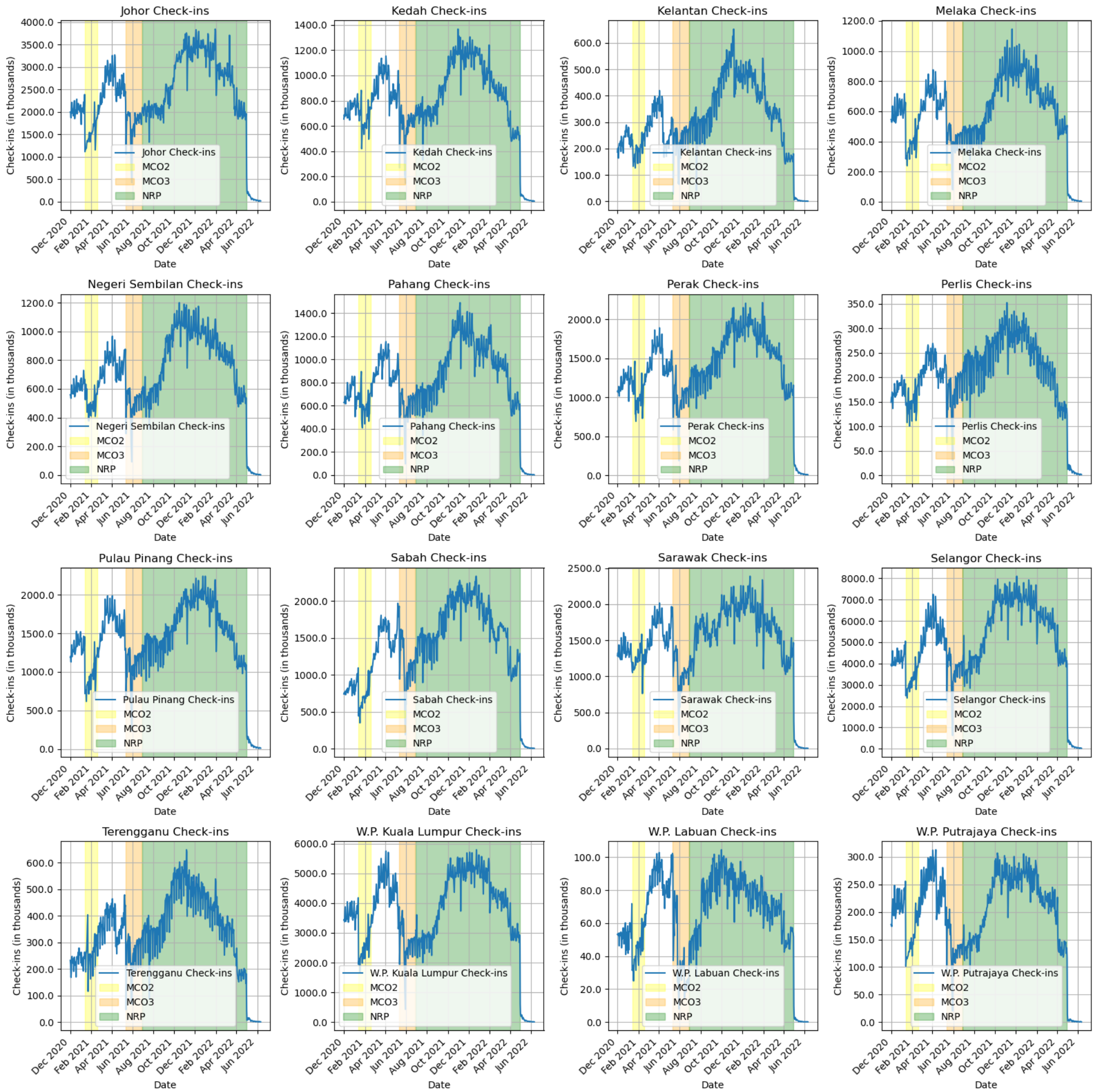
Notably, W.P. Putrajaya and Kuala Lumpur reports the highest percentage of unique individuals (102.04% and 86.55%), reflecting its role as an urban center with high mobility and frequent check-ins, likely driven by its status as the economic and administrative hub of Malaysia. Additionally, being developed urban areas, health care facilities such as secondary and tertiary hospitals, are also more concentrated in these areas. In a lockdown due to deadly viral pandemic, visitations to hospitals would have increased and thereby increasing the check-ins. This is further borne by the greater number of unique locations reported in urbanized states like Selangor and Pulau Pinang (131,455.73 and 43,779.85, respectively).

In contrast, less urbanized regions such as Kelantan and Terengganu have significantly lower unique individual percentages (9.30% and 15.11%, respectively), suggesting limited population mobility or fewer check-ins relative to their size. This may be attributed to differences in population density, the availability of public spaces and healthcare facilities. Additionally, lack of technological access or poor familiarity with mobile applications may have also contributed to the low individual reporting rates.

The variation in the number of unique locations further emphasizes the contrast between urban and rural regions. States like Johor and Selangor have substantial numbers of unique locations (80,575.15 and 131,455.73), which may be reflective of larger, more developed commercial and recreational infrastructures. In contrast, regions such as Kelantan and Perlis report fewer unique locations (18,371.99 and 8,237.56, respectively), likely due to smaller populations and less urban development. These discrepancies in both mobility and location diversity could influence the spread of COVID-19, as higher movement between unique locations in more urbanized areas might increase exposure and transmission rates. This is further suggested by the high numbers of active (29,197 and 8,782) and new cases (2,437 and 687) in Selangor and Johor respectively, likely reflecting their larger populations and urban environments, which facilitate the virus's spread. In contrast, less populous areas like Perlis and W.P. Putrajaya report lower figures for active and new cases, possibly due to successful containment strategies or smaller populations with fewer interactions. This regional variability underscores the importance of tailored public health interventions, considering both the level of mobility and the density of unique locations in each area. The combination of high check-ins, large numbers of unique locations, and active case counts highlights the complex relationship between mobility, social behavior, and disease transmission, necessitating context-specific strategies to mitigate the spread of COVID-19 across diverse regions

When visualizing the "Check-in per day" data in a line graph, all the 13 states and 3 territories appear to follow the same bimodal distribution, indicating a potential double normal distribution. However, when the Shapiro-Wilk test was conducted, the data does not follow a normal distribution, even after cutting and altering the start and end points.

Graph 2 : Line graph of Check-ins by state during the study period with MCO2, MCO3 and NRP highligjted.



Comparisons were then performed between geographical groups – East Malaysia vs West Malaysia, East Coast vs West Coast, Territories vs States, and least populous vs most populous states using the Mann-Whitney U test, to assess differences in key variables such as check-in per day, unique individuals, and unique locations. The results from the groups indicate differing levels of significance across the comparisons. For the comparisons between East and West Malaysia, East Coast and West Coast, and Territories and States, the p-values for all tested variables (check-in per day, unique individual logins, and unique locations registered) were consistently above the conventional significance threshold of 0.05 (p-values ranged from 0.1939 to 1.000), suggesting that there were no statistically significant differences between these geographical groupings. However, the comparison between the five least populous and five most populous states yielded significant results, with p-values of 0.0285 for all variables, indicating that the level of population density has a notable impact on check-in behaviors, unique individual logins, and the diversity of check-in locations. This suggests that more populous states exhibit higher mobility, engagement, and spatial variation in activity, likely due to increased social interactions, economic activity, and healthcare access, which could influence the spread of COVID-19.

Table 2.1 : Mann-Whitney-U tests comparing state groups.

Comparison	**U-Score**	**p-value**
East Malaysia vs West Malaysia		
Average check-in per day	19.0	1.000
Average unique individual login	16.0	0.7036
Average unique location registered	17.0	0.8000
East Coast vs West Coast		
Average check-in per day	18.0	0.2787
Average unique individual login	19.0	0.1939
Average unique location registered	19.0	0.1939
Territories vs States		
Average check-in per day	12.0	0.3643
Average unique individual login	11.0	0.3643
Average unique location registered	18.0	0.2964
5 least populous vs 5 most populous states		
Average check-in per day	0.0	0.0285
Average unique individual login	0.0	0.0285
Average unique location registered	0.0	0.0285

2.3 Statistical Analysis

By using the merged data, we are keen to assess the strength and direction of relationship between the MySejahtera check-ins with daily new Covid-19 cases, unvaccinated cases, and reported clusters trend. We have applied the Pearson Correlation method via python codes. The correlation coefficient with significance value were generated.

To explore predictive relationships among the said data, we have used the Ordinary Least Squares (OLS) logistic regression model. The details of the model as below:

- Dependant variables: Daily new Covid-19 cases, daily unvaccinated Covid-19 cases, number of clusters daily report
- Independent variables: Daily MySejahtera check-ins

All the generated report and tables consists of significance value. We take p-value<0.05 as significant relationship and association among the variables. Regression model was created based on the following equation:

$$new\ daily\ cases = \beta_0 + \beta_1. + \beta_2\ .cluster_{import} + \beta 3.cluster_{religious} \ldots$$

3.0 Results

3.1 MySejahtera Apps Daily Check-in usage and Daily new Covid-19 cases

The dataset dated from Jan 2020 till June 2022 comprises from a lot of numbers as it records all the check-ins done by the Malaysians during Covid period, albeit pre, during or post lockdown time. One of the objectives of this study is to look into the correlation of this check-ins data and new covid cases reported.

Based on the dataset, there is significant correlation between daily check-in and new Covid-19 cases day to day. It was both verified by Pearson Correlation method (p-value = 0.0001) and Ordinary Least Squares (OLS) logistic regression model (p-value = 0.0001). However, both methods showed there is very minimal changes in the new cases recorded daily affected by the check-ins. Only 2.6% variation was found in Table 3.0. Thus, even though the correlation among them is statistically significant, but the overall performance to explain the changes is very weak. The same phenomenon also being reflected in Covid-19 daily active cases, as per Table 3.1. Furthermore, the scatter plot in Figure 1.0 also shown only slight upward of the regression line which indicated weak positive relationship among daily Covid-19 cases and check-in apps usage.

Table 3.0 : OLS model summary & regression table for association of dependant variable (new daily Covid-19 cases) and independent variable (daily MySejahtera Check-ins)

Variable	Coefficient	Std. Error	t-statistic	p-value	95% CI
Intercept	4738.05	899.85	5.265	0.000	[2970.54, 6505.56]
MySejahtera Apps Daily Check-ins	0.0002	0.00004	3.861	0.000	[0.00008, 0.0003]
R-squared	0.026				
Adj. R-squared	0.024				
F-statistic	14.91				
AIC	1.158e+04				
BIC	1.159e+04				

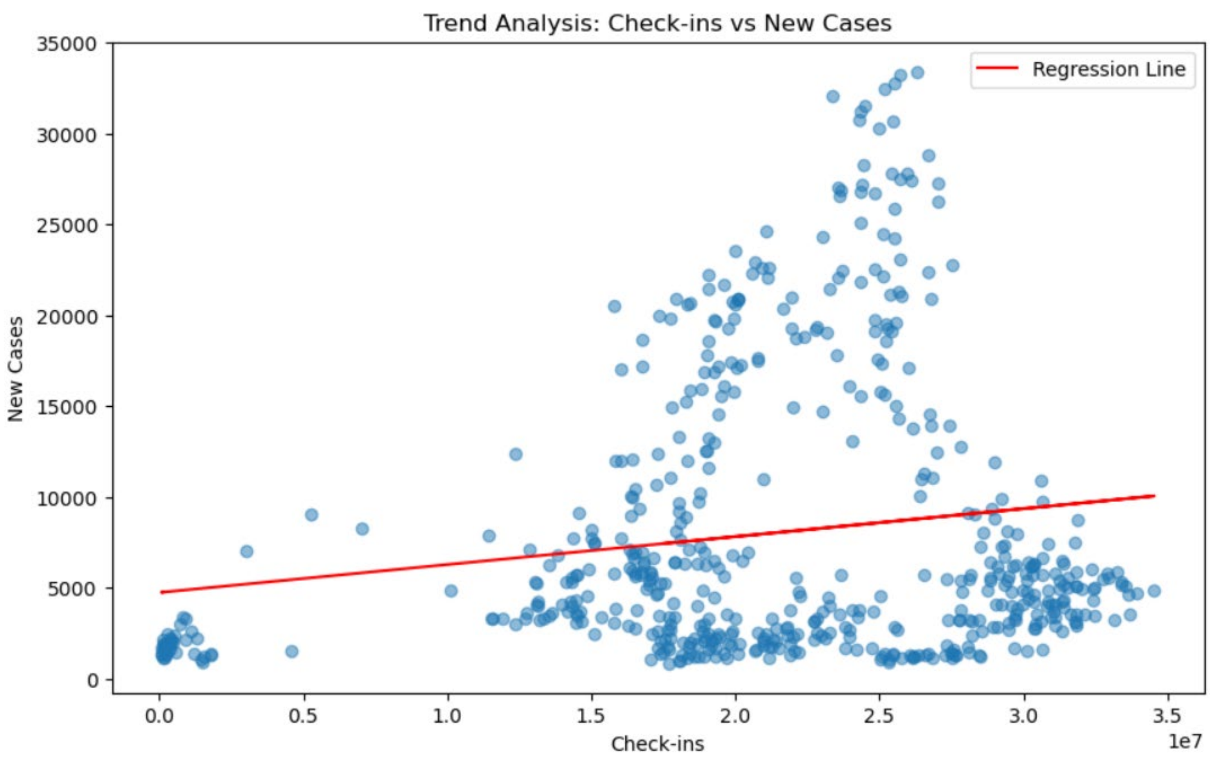


Figure 1: Scatter plot of trend analysis – Daily MySejahtera check-ins and new Covid-19 cases

3.2 MySejahtera Apps Daily Check in and Daily new Covid-19 unvaccinated cases

On the other angle, the apps which intended to detect, assist and boost the Covid-19 vaccination program in Malaysia was also checked for statistically related to new daily unvaccinated cases. The dataset did not reveal a good relation between daily check-ins of MySejahtera apps and the daily unvaccinated cases. Table 3.2 showed that there was only 0.5% of the variation in unvaccinated cases daily explainable by the daily check-ins (p-value >0.05).

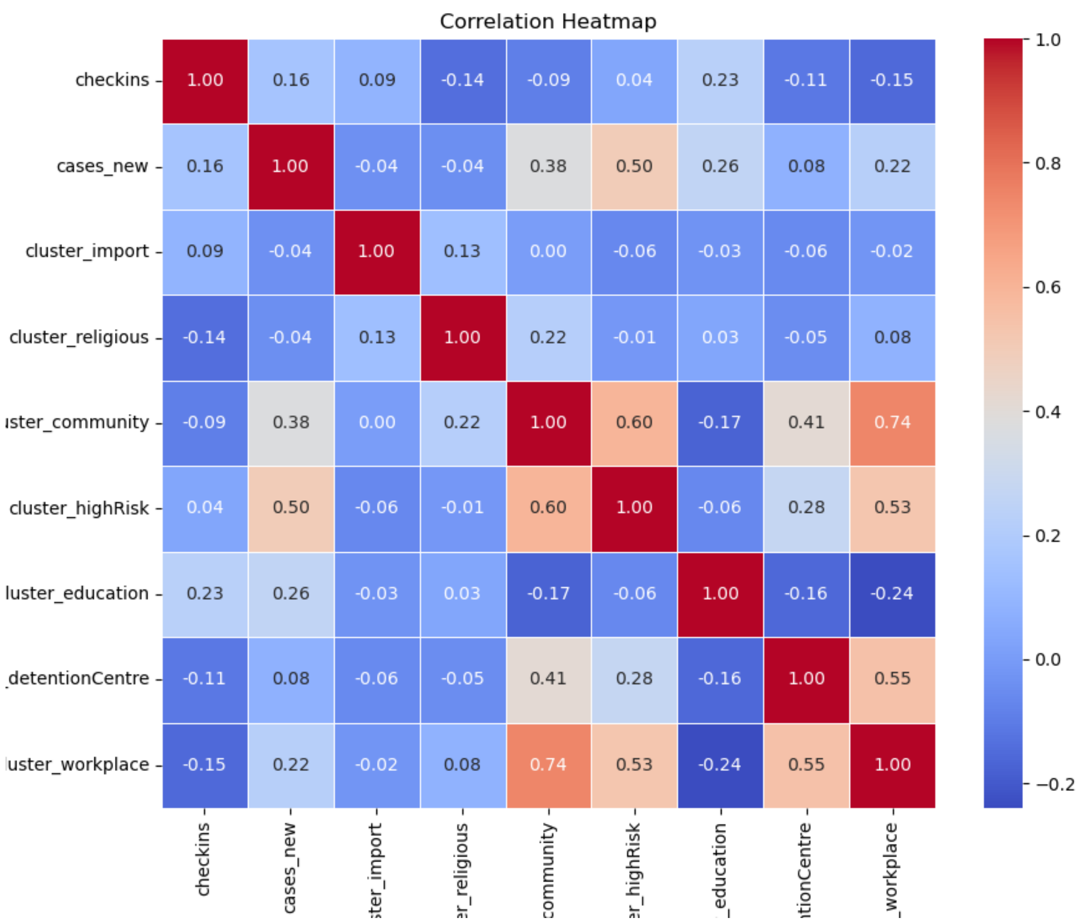


Figure 2: Correation heatmap of daily MySejahtera check-ins

Table 3.1: OLS model summary & regression table for association of dependant variable (daily Covid-19 active cases) and independent variable (daily MySejahtera Check-ins)

Variable	Coefficient	Std. Error	t-statistic	p-value	95% CI
Intercept	56590	10000	5.659	0.000	[36900, 76200]
MySejahtera Apps Daily Check-ins	0.0017	0.0004	3.840	0.000	[0.001, 0.003]
R-squared	0.026				
Adj. R-squared	0.024				
F-statistic	14.74				
AIC	1.427e+04				
BIC	1.428e+04				

Table 3.2: OLS regression table for association of dependant variable (daily Covid-19 unvaccinated cases) and independent variable (daily MySejahtera Check-ins)

Variable	Coefficient	Std. Error	t-statistic	p-value	95% CI
Intercept	3961.01	367.58	10.776	0.000	[3238.99, 4683.02]
MySejahtera Apps Daily Check-ins	-0.000	0.000	-1.632	0.103	[-0.0000585, 0.0000054]
R-squared	0.005				
Adj. R-squared	0.003				
F-statistic	2.665				
AIC	1.058e+04				
BIC	1.059e+04				

3.3 MySejahtera Apps Daily Check-in with cluster of Covid-19 cases

There are few clusters of emerging Covid-19 cases during the said duration of endemic time in Malaysia. The heatmap shown as Figure 2.0 that the nearest correlation with were cluster cases within the community (of the infected cases), followed by high risk cluster and cluster at Covid-19 detention centre. Further correlation using OLS regression with multiple independent variables also revealed that there were positive and great significant of new daily Covid-19 cases with cluster of cases in the community, high risk group and education. These clusters were associated with higher new Covid-19 cases daily (p-value<0.05), as per Table 3.3.

Table 3.3: OLS regressions table for association of dependant variable (daily Covid-19 new cases) with independent variable (daily MySejahtera Check-ins, clusters of Covid-19 cases)

Dependant Variable	**R-squared**	**F-statistic**	**p-value**	**Log-likelihood**	**AIC/BIC**
New Covid-19 cases daily	0.384	42.90	0.000	-5659	11340/11380

Independent Variable	**Coefficient**	**Std. Error**	**t-statistic**	**p-value**	**95% CI**
Intercept	3961.01	367.58	10.776	0.000	[3238.99, 4683.02]
MySejahtera Apps Daily Check-ins	0.00007	0.00003	2.050	0.041	[0.000003, 0.000138]
Cluster import	-16.38	80.28	-0.204	0.838	[-174.06, 141.31]
Cluster Religious	-16.61	5.93	-2.801	0.005	[-28.26, -4.96]
Cluster community	13.75	2.40	5.725	0.000	[9.03, 18.47]
Cluster high-risk	103.26	11.37	9.081	0.000	[80.92, 125.59]
Cluster education	18.54	2.23	8.308	0.000	[14.16, 22.92]
Cluster detention centre	-6.56	4.58	-1.434	0.152	[-15.55, 2.43]
Cluster workplace	-1.65	0.88	-1.872	0.062	[-3.39, 0.08]
R-squared	0.384				
Adj. R-squared	0.375				
F-statistic	42.90				
AIC	1.134e+04				
BIC	1.138e+04				

4.0 Discussion

Descriptive Analysis – National Scenario

The Covid-19 pandemic in Malaysia has progressed through several distinct phases, each marked by unique trends in case numbers, severity, and public health responses. By June 2022, Malaysia had recorded high number of cases, starting with the first wave (January–February 2020), which primarily consisted of imported cases with minimal local spread[7]. The situation changed during the second wave (March–June 2020), as significant local clusters emerged, prompting the implementation of the Movement Control Order (MCO). The third wave (late 2020 to early 2021) brought a significant rise in cases due to community and workplace outbreaks. By mid-2021, the Delta variant triggered severe consequences, with daily cases peaking in late 2021. The Omicron wave in 2022 caused another surge in case numbers but with fewer severe outcomes, thanks to widespread vaccination efforts.

When examining the demographics of the pandemic, certain patterns emerge. Working-age adults, particularly those between 20 and 49 years old, recorded the highest infection rates. This trend likely due to their higher movements with exposure at workplaces[3]. Conversely, the elderly population (over 60 years old) was disproportionately affected by severe cases and fatalities, highlighting their vulnerability. Analyzing gender differences revealed no major disparities in infection rates; however, mortality was slightly higher among men.

Further data analysis covering the period from December 1, 2020, to June 11, 2022, reveals several critical insights. The use of MySejahtera app program for check-ins initially showed a rising trend. However, after April 30, 2022, check-in rates began to declined, likely due to changes in policies and reduced enforcement. Imported case numbers surged on December 14, 2020, following the easing of border restrictions but dropped again after April 4, 2022. This decline reflects the virus's weakening impact, the rise of asymptomatic cases, and the success of vaccination programs worldwide.

Malaysia's Covid-19 data demonstrates a complex interaction between public health policies, virus mutations, and vaccination coverage. The phased implementation of the MCO, alongside the vaccination campaign, played key roles in managing infection rates and reducing severity. Moving forward, understanding the factors behind these trends—such as cultural behaviors, economic conditions, and the broader global response—will be crucial for future decision-making.

Descriptive Analysis – Comparison of States and Territories

The data analysis reveals notable regional disparities in check-in activity, unique individual users, and unique locations, which are influenced by factors such as urbanization, population size, and regional development. Urban centers like Selangor, with 4,885,762.92 check-ins per day, and W.P. Kuala Lumpur, with 3,603,656.92 check-ins, exhibit significantly higher check-in frequencies, unique individual percentages (33.10% and 86.55%, respectively), and numbers of unique locations (131,455.73 and 76,993.43, respectively). These figures reflect the greater levels of mobility and social engagement typically found in highly urbanized regions, likely due to the concentration of commercial, recreational, and healthcare facilities. On the other hand, less urbanized regions such as Perlis (203,130.68 check-ins per day, 42.30% unique individuals) and Kelantan (301,059.43 check-ins per day, 9.30% unique individuals) exhibit lower check-in numbers and unique individual percentages. This suggests limited population mobility and fewer interactions across diverse locations, which may be attributed to smaller population sizes, less-developed infrastructure, and lower technological accessibility in these areas.

The observed disparities in the number of unique locations could be further elucidated with the availability of more granular data, such as the categorization of locations where check-ins occurred. It is likely that this data is already captured or embedded within the underlying database, as most mobile devices utilize GPS tracking functionality. However, the decision to make such data publicly accessible requires careful consideration. Presently, many states exhibit relatively low unique user registration rates (as low as 9%), which may be attributed to concerns regarding privacy and trust. While the release of location-based data could provide valuable insights into human mobility and behavior, it is essential to balance these potential benefits with the risk of diminishing user engagement. If users perceive a threat to their privacy, they may choose to disengage from the application, which could ultimately undermine its utility for public health monitoring and behavioral analysis.

Regional variations in active and new COVID-19 cases further illustrate the connection between mobility patterns and disease transmission. Selangor, with 29,197 active cases and 2,437 new cases, and Johor, with 8,782 active cases and 687 new cases, report higher numbers of active and new COVID-19 cases, likely driven by their larger populations and frequent mobility. This is consistent with the higher number of unique locations in these states, which can increase the likelihood of virus transmission. In contrast, less densely populated states like Perlis (357 active cases and 31 new cases) and W.P. Putrajaya (482 active cases and 37 new cases) report lower active and new case counts, which could be indicative of successful containment strategies or fewer opportunities for virus transmission due to lower mobility. These findings emphasize the need for context-specific public health interventions, as urban areas with higher levels of mobility and more complex social interactions may require more targeted strategies to mitigate the spread of COVID-19, compared to less urbanized regions with lower levels of viral exposure.

The comparison between the various state groups indicate there was no significant difference in between these geographical groupings. This lack of significance may imply that, despite regional variations in infrastructure or population size, mobility patterns and user engagement with the app are relatively uniform across these regions. This is probably due to combination of government regulation and enforcement coupled by the fear of the disease. This is further confirmed by the similar check-in trends of each state visualized on the line-chart. This is especially so as the number of check-ins drastically drop on 1st May 2022 when the government announced that registration on MySejahtera is no longer compulsory to enter building premises[6]. However, a small residual number of individuals continue to do so, most likely because certain facilities such as hospitals, still continue to insist on registering individuals.

Statistical Analysis

Few practical concerns and highlights can be learn and unlearn after the Covid-19 pandemic in Malaysia. This studyclearly shows that certain steps can be taken to mitigate the spread of the disease. The arrival of super app such as MySejahtera app have a great potential towards the advantage of controlling the disease – Covid-19.

Awareness of Covid-19 cases and MySejahtera app

There was significant correlation between daily check-ins data by MySejahtera app recorded and new Covid-19 cases reported during the study period. For every additional check-in recorded in the data, there is estimated average increase of 0.00002 in a new daily Covid-19 cases. It could probably due to increase awareness and compliance to MySejahtera app in daily routine of Malaysians during that period. The detection of new Covid-19 cases can be seen and reported via the app. It also showed the public compliance and confidence towards the government policies. South Korea also used a detecting tracking app during the surge of Covid-19 cases[4]. The implications are valuable for policy makers such as Ministry of Health to encourage safety and self-coping behaviour.

The small increase in relation to daily check-ins but indeed a significant relations (p-value<0.05) may be contributed by lagged effect of cases. Due to increase mobility of public during and after the MCOs, individuals move more and check-in more frequently. This will lead to increase exposure risk to Covid-19 infection, leading to rise in detected cases after certain incubation period. The incubation period known for Covid-19 infection is around 2-14 days after first contact[5]. It is also depending on the variant that individuals exposed to. Hence, the rise or surge of cases may differ because it will undergo incubation period first. Patient to patient may differ as unique defensive immune system in every individual.

Other contributing factors, such as public health interventions towards Covid-19, government policies, raised vaccination rates and demographic dynamic also lead to a very small coefficient change (0.00002).

Unvaccinated – special population

It was challenging to get the whole public to adhere to a new policy – such as vaccination program. When vaccination introduced to the world, it was brought in to Malaysia by the Health minister on February 2021[2]. The program establish more with the help of MySejahtera app. The data of vaccinated versus unvaccinated also obtained from this app. One of the reason the association of check-ins with unvaccinated Covid-19 cases may be attributed by the dynamics of unvaccinated population itself. Many reasons why this population refused for vaccination, or they not even captured by the apps itself. They also have reduced access to testing facilites. Given the lack of statistical significance, MySejahtera check-in data may not be a reliable tools for tracking the unvaccinated individuals. In future, government should tackle this issue with different method to better understand this type of group dynamics. An article published in Elsevier, Science Direct[8], exposed that many of the unvaccinated personnel are victims of fake news. Those individuals who refused or delayed vaccinations more likely to have looked for vaccine information from the internet, where negative content is predominant.

Cluster detection

Direct impact of public mobility and high usage of MySejahtera app check-ins correlates with higher transmissions of Covid-19 infection during the study period. High risk clusters group, community clusters and education clusters group have significant coefficient and impact on daily reported Covid-19 cases. On the other hand, religious clusters have negative coefficient, maybe indicated by reduced religious activity in the community. Policy maker can use this data to mitigate new infection if ever to happen again. Focus on high risk population, education planning can reduce the spread of an infection, such as Covid-19. Closing the religious temple or mosque can also reduce the rate of infection via touch and airborne[0].

5.0 Conclusion

There are many ways of tackling the spread of Covid-19 infection. Using MySejahtera check-in alone can reveal a lot of information that is useful. Increasing awareness among the public but adherence to a specific policy of using a simple app such as MySejahtera can give many advantages. Future improvement can be done by incorporate other demographic variables to enhance model efficacy. Establishing a data-driven analysis framework will ensure preparedness for similar challenges in the future.

Reference

1. https://github.com/MoH-Malaysia
2. https://en.wikipedia.org/wiki/COVID-19_vaccination_in_Malaysia#:~:text=of%20Medical%20Biology.-,February,a%20period%20of%20two%20weeks.
3. Morawska, L., & Cao, J. (2020). Airborne transmission of SARS-CoV-2: The world should face the reality. *Environment international*, *139*, 105730.
4. Kim, H. (2021). COVID-19 apps as a digital intervention policy: a longitudinal panel data analysis in South Korea. *Health Policy*, *125*(11), 1430-1440.
5. How long is the Covid-19 Incubation period?, WebMD source website, article written by Sonya Collins, Even Starkman. https://www.webmd.com/covid/coronavirus-incubation-period
6. Ministry of Health Malaysia. (2022, April). *Daftar masuk MySejahtera tidak lagi diperlukan mulai 1 Mei 2022*. COVID-19 Malaysia. https://covid-19.moh.gov.my/reopeningsafely/semasa/2022/04/daftar-masuk-mysejahtera-tidak-lagi-diperlukan-mulai-1-mei-2022#:~:text=Daftar%20masuk%20MySejahtera%20tidak%20lagi%20diperlukan%20untuk%20memasuki,di%20bawah%20HSO%20masih%20tidak%20dibenarkan%20memasuki%20premis.
7. Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Abd Rahman, A., Sekawi, Z., ... & Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International journal of infectious diseases*, *97*, 108-116.
8. Zamir, E., & Gillis, P. (2023). The pandemic of the unvaccinated: a Covid-19 ethical dilemma. *Heart & Lung*, *57*, 292-294.