**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**

**CONTENT**

|  |  |  |
| --- | --- | --- |
|  |  | **Page** |
| **1.** | Introduction  1.1 Objective | **3** |
| **2.** | Method  2.1 Descriptive Analysis  2.2 Histograms  2.3 Statistical Analysis I  2.4 Statistical Analysis II | **4**  **5**  **5**  **6** |
| **3.** | Results  Table 1.0  Table 2.0  Table 3.0  Table 4.0  Figure 1-4  Figure 5 | **8**  **8**  **10**  **10**  **9**  **11** |
| **4.** | Discussion | **12** |
| **5.** | Conclusion | **13** |
| **6.** | References | **13** |

**1.0 Introduction**

Covid-19 cases

A descriptive analysis of the data has been collected and analyzed. The attributes of the data are listed below:

1. Age: age of a subject
2. Gender: sex type
3. Occupation: Occupation of a subject
4. Country: country of a subject
5. Mental health condition: presence or absence of mental health in a subject
6. Severity: Severity condition of a subject
7. Consultation history: For individuals who have mental health conditions, the dataset notes whether they have consulted a mental health professional.
8. Stress level: Individual’s stress level
9. Sleep hours: Individual’s sleep hours per day
10. Work hours: Individual’s work hours per week
11. Physical Activity Hours: Individual’s physical activity hour per week

**1.1 Objective**

The dataset for the study is provided from MoH Malaysia Github website, (link : <https://github.com/MoH-Malaysia/>) under the repository of Covid-19 for public.

**Research question**:

**Objective**:

**2.0 Method**

Based on the data collected

**2.1**

**2.2**

**2.3**

**2.4 Statistical Analysis I**

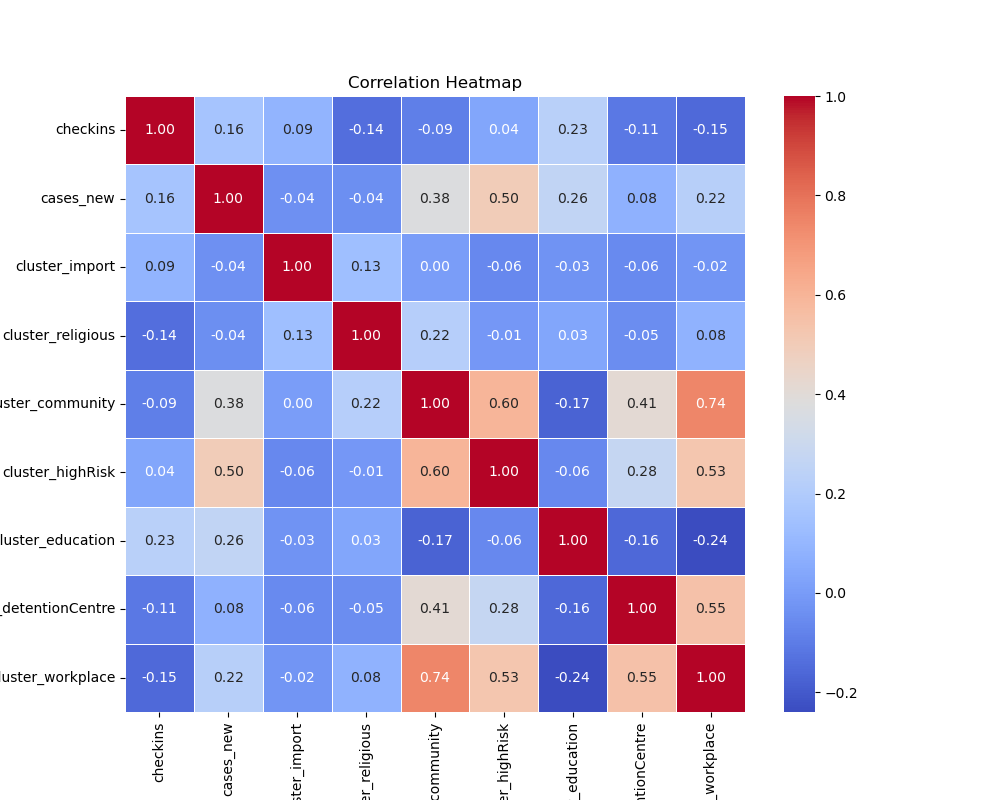
**2.4.1 MySejahtera Apps Daily Check-in usage and Daily new Covid-19 cases**

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 objective 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 1.0. 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 2.0. 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.

**2.4.2 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 3.0 showed that there was only 0.5% of the variation in unvaccinated cases daily explainable by the daily check-ins (p-value >0.05).



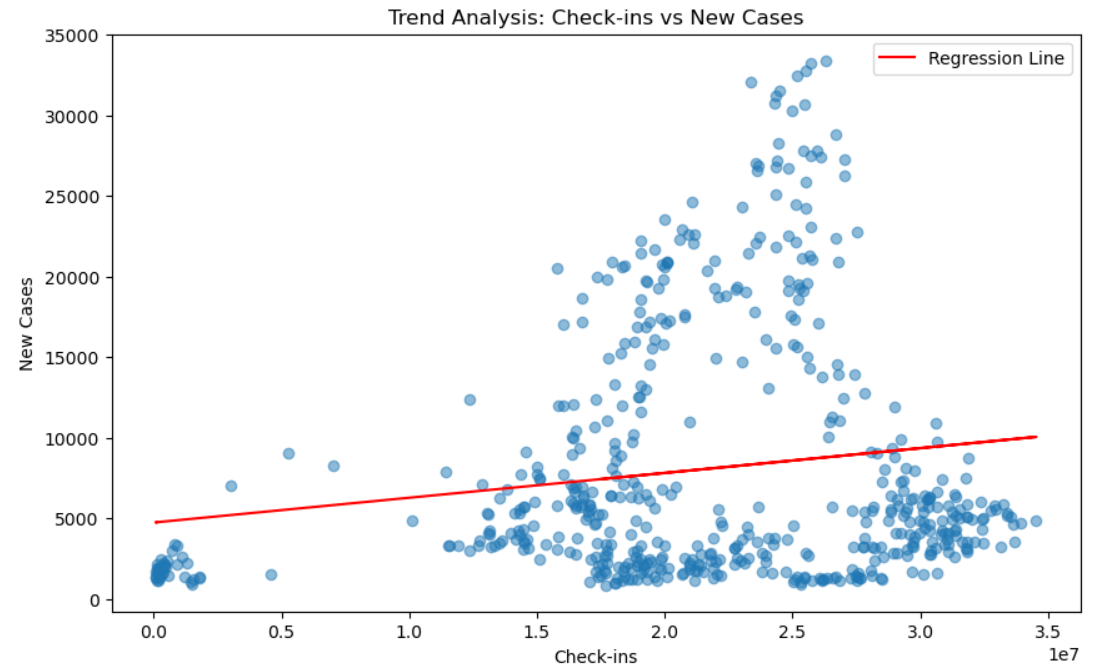
**2.4.3 MySejahtera Apps Daily Check-in with cluster of Covid-19 cases**

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 4.0.

**3.0 Results**

**Table 1.0 : OLS 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] |



**Table 2.0 : OLS 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] |

**Table 3.0 : 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] |

**Table 4.0 : 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] |
|  |  |  |  |  |  |

**4.0 Discussion**

**Descriptive Analysis**

Based on the presented dataset, it is known that the data given comprises of categorical and continuous data. Among the continuous data, all the data was rejected for normality testing as p-value >0.05 according to Kolmogorov-Smirnov tests. The visualization of these data also proves that the data normality (Figure 1 – 4).

The details of mean, median, IQR were gathered further in Table 2.0. It shows that the age range is from 18-16 years, with a mean subject's age is 42 years old. The data is fairly spread out with a relatively high standard of deviation. Most subjects sleep between 4 – 10 hours, with a mean close to 7 hours. For work hours, they have a larger spread, ranging from 30 – 80 hours (about 7 days) per week, with a mean around 55 hours (about 5 days). This indicates that some subjects may have been working for long hours. There is also a variation with physical activity hours in a week, with a range from 0 – 10 hours per week.

**Statistical Analysis**

On the earlier step of analysis, the logistic regression method was used to determine any significant association between variables and mental health conditions. However, based on Table 3.0, none of the variables were tested significantly affecting the presence of mental health illness in the subjects (p-value > 0.05). *Support journal ----*

On the contrary, further tests were done to check the association among the variables. We have found that stress level (High, medium, low) was influenced by work category. On the earlier part, working hours per week data has been grouped into 3 groups; Part-time (<20 hours), Full-time (20-45 hours), Overtime(45-60hours) and Excessive (>60 hours).

Since the p-value (0.04223) is less than the common significance level of 0.05, we can reject the null hypothesis. Among other variables, none showing significant p-value as compared to work category. Figure 5 shows that Excessive group (>60 hours/week) having the highest count in high level of stress.

Combining both logistic regression which showed no association among variables with presence of mental health and significant contribution of work categories with level of stress. We can deduce that high stress level is contributed by higher working hours per week, however, it does not necessarily cause mental illness in these samples of subjects. In addition, possible coping mechanism among individuals in this samples that have prevent the presence of mental illness altogether. In the study conducted by Rodrigues et al2 among Portuguese workers, they found that by applying the adaptive coping strategies may help mitigate symptoms of mental illness and increase satisfaction in life.

**Conclusion**

Stress level can be attributed to excessive or higher working hours per week. Suitable working hours per week should be kept at less than 45 hours per week. Nonetheless, good capable adaptive coping strategies can prevent individuals from having mental health illness.

**Reference**

1. Office of the Surgeon General (US); Center for Mental Health Services (US); National Institute of Mental Health (US). Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General. Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2001 Aug. Chapter 2 Culture Counts: The Influence of Culture and Society on Mental Health.
2. Rodrigues, F., Morouço, P., & Santos, T. (2023). Testing the Associations between Coping, Mental Health, and Satisfaction with Life in Portuguese Workers. *European Journal of Investigation in Health, Psychology and Education*, *13*(7), 1245-1256. <https://doi.org/10.3390/ejihpe13070092>
3. https://www.geeksforgeeks.org/r-tutorial/