

The Impact of COVID-19: An analysis on the Unemployment and Underemployment Rates of Hong Kong Special Administrative Region (HKSAR)

Sicat, Ma. Gracia Corazon E.¹
mmgcsicat@tip.edu.ph

Marinduque, Ric Russel²
mrrrmarinduque@tip.edu.ph

Salacup, Severino Jr.³
msasalacup@tip.edu.ph

Professional Science Masters in Data Science¹²³
Technological Institute of the Philippines
1338 Arlegui St. Quiapo, Manila

ABSTRACT

In this paper, we studied the effect of the corona virus disease 2019 (COVID-19) on the unemployment and underemployment of the Hong Kong Special Administrative Region, aiming to provide precautionary measures and suggestions to policy-makers, companies and individuals for better mitigation strategies of future pandemics. We studied and analyzed the time series data from the Census and Statistics Department of the Hong Kong Special Administrative Region starting from January to September 2020. We compared the results of the two (2) evaluation methods: Mean Score Evaluation and Clustering Methods, specifically, employing K-Means Clustering and Agglomerative Clustering to the time series datasets. We discovered that the most affected industries are the 'Construction' and 'Retail, Accommodation and Food Services'; most affected local citizens are from the age 15 to 19; and, affected occupations are 'Services and Sales Workers' and 'Craft and related workers'. While the least affected industry is 'Financing, Insurance, Real Estate and Professional and Business Services'; the least affected age group is 30 to 49; and in occupations are 'Managers, Administrators and Professionals'.

The result shows that the groups that were mostly affected and least affected are directly correlated to one another and this result is due to quarantines, lockdowns, stay-at-home policies implemented by the government to prevent the spread of the virus. Furthermore, this paper provided insights and fundamental knowledge for policy-making and pandemic response and recovery programs for unemployment and underemployment in HKSAR.

Keywords: COVID-19, job losses, k-means clustering, agglomerative clustering

1. INTRODUCTION

The COVID-19 is a global health crisis that has disrupted the world economy and caused shocking employment losses due to quarantines, business closings, and consumer demand reduction. The economic pain is widespread but not proportionately distributed. Many workers have been severely distressed by mass redundancies while others have shifted to work from home setup while still receiving their regular salary. These effects vary systematically with individual demographic and socioeconomic characteristics. The Statistics and Census Department Hong Kong Special Administrative Region (HKSAR) report that the national employment rate surged to 5.9% in the three months ending May 31, surpassing the 5.5% highest during the global financial crisis in 2009

and marking the highest level since 2005 in the aftermath of the SARS outbreak (Chan, 2020). The hike in unemployment is the latest indicator of the financial hub's economic doldrum.

This effect continues to grow in number as the pandemic prolongs. And, one should be mindful of the duration, the coverage and the depth of this effect. This project aims to understand the effects of the Covid-19 pandemic to the unemployment and underemployment rates of the HKSAR. This project studied the labour force, unemployment and underemployment data provided by the Census and Statistics Department of HKSAR where it is composed of the following groups: (1) age and gender group; (2) occupation group and (3) industry group.

Specifically, this project answered the following research problems:

- What is the effect of COVID-19 pandemic to the unemployment and underemployment of the HKSAR?
- What is the highly affected and least affected by the COVID-19 pandemic among the following data groups:
 - Age and gender group
 - Occupation group
 - Industry group
- What are the precautionary measures and/or strategies that the government, industrial companies and individuals can do in order to cope and survive in a pandemic event, such as COVID-19, if such will happen again in the future?

2. MATERIAL AND METHODS

This section discusses the various techniques such as the data collection, data preprocessing, data evaluation and analysis techniques. The first subsection presents the characteristics of the project's dataset. Then, the methods that were used to analyze the dataset are described.

2.1. Dataset

The dataset of this project is comprised of the unemployment and underemployment rates provided by the Census and Statistics Department of HKSAR from the year 1983 to 2020, with a combination of monthly and yearly time series data.

The dataset is composed of six (6) excel files and are grouped by age and gender, occupation and industry.

Combining all the files, it has a total of 3,110 rows and 64 columns.

Figure 1 shows the actual data set of the six (6) excel files.

Figure 1 displays three screenshots of Excel spreadsheets showing unemployment and underemployment rates by sex and age. The spreadsheets are titled 'TableA01 Unemployment Rate by Sex and Age', 'TableA02 Unemployment Rate by Sex and Age', and 'TableA03 Unemployment Rate by Sex and Age'. Each spreadsheet contains columns for Period, Sex, Age Group, and Unemployment rate (%).

Figure 1 displays two screenshots of Excel spreadsheets showing unemployment and underemployment rates by sex and age. The spreadsheets are titled 'TableA04 Unemployment Rate by Sex and Age' and 'TableA05 Unemployment Rate by Sex and Age'. Each spreadsheet contains columns for Period, Sex, Age Group, and Unemployment rate (%).

Figure 1. Actual Dataset

2.2. Methodology Framework

This subsection presents the main processes and sub-processes implemented in analyzing the dataset. Figure 2 shows the methodology framework.



Figure 2. Methodology Framework

There are two (2) major processes that this project implemented. These are: (1) Data Preprocessing and (2) Data Mining.

Data Preprocessing. This ensures that the data meets its standard characteristics. These characteristics as enumerated by Delvecchio(2020) are timeliness, relevance, accuracy, completeness, and consistency. The following sub-processes were performed to complete the data preprocessing:

- **Data Cleaning.** Is the removal of inaccurate, incomplete and 'dirty' (erroneous) values from the data set. In the case of the project's data set, the removal of the unnecessary rows and columns with all NaN values, setting the header of the dataframe and skipping of headers and footers.
- **Data Reduction.** This refers to obtaining the most relevant information in the data set. The reduction was done through selecting data from the whole data frame through regular expressions, str.contains() function and row and column indexing.
- **Data Integration.** This process refers to the collection and combining the sorted data into a one single data frame where this single data frame is suitable for analysis and manipulation.. The integration processes were transposition of the data frames, including mean column values, and extraction of data frames from the single data frame.
- **Data Transformation.** This is the last process in the data preprocessing where the cleaned, optimized and reduced data were transformed for better analysis and manipulation. The transformation processes were renaming of columns relevant to the values that were cleaned, optimized and reduced, replacement of non-numeric values, value indexing and the use of the Standard Scaler library in transforming the data for modeling.

Figure 3 shows the outcome of the dataset after implementation of the processes in data preprocessing.

First rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

AGE AND GENDER

Last rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

First rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

OCCUPATION

Last rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

First rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

INDUSTRY

Last rows

Year	Age	Gender	Occupation	Industry
2010-1000	25	Male	Software Engineer	Technology
2011-1000	28	Female	Marketing Specialist	Retail
2012-1000	30	Male	Finance Analyst	Finance
2013-1000	32	Female	Healthcare Worker	Healthcare
2014-1000	35	Male	Education Specialist	Education
2015-1000	38	Female	Operations Manager	Manufacturing
2016-1000	40	Male	Project Manager	Construction
2017-1000	42	Female	Human Resources	Human Resources
2018-1000	45	Male	Business Development	Business Development
2019-1000	48	Female	Quality Assurance	Quality Assurance
2020-1000	50	Male	Systems Administrator	Information Technology

Figure 3. Dataset after Data preprocessing

Data Mining. After the conduct of the processes in the data preprocessing, the next process, is the Data Mining. Data Mining is incorporating relevant and important steps to generate incremental insights where it started from prior knowledge and ends with a posterior knowledge. (Kotu and Deshpande, 2015) In performing this, the sub-processes in data mining that this project incorporated were the following:

- **Data Evaluation.** This is the process of understanding the data set, its 'ground' truth, its nature, its characteristics and values. In evaluating the data of this project, describe() and isna() functions were used. Also, the Pandas Profiling Report was incorporated. The Pandas Profiling Report is a library in Python that generates a basic report on the input data frame. The report is exported into a HTML file for better visualization and document keeping.
- **Data and Pattern Analysis.** The heart of the data mining process lies in this sub-process, the data and pattern analysis. The objective of this is to generate the 'posterior' knowledge, the insight that is useful and relevant. In the data and pattern analysis, the following were used:
 - Correlation Heat Map
 - Elbow Method and Silhouette Coefficient Method.
 - Mean Score and Clustering Methods

- **Data Presentation.** To communicate the results, this project made use of dendrogram, scatter plots, heat maps, and bar graphs / bar plots.

Python Libraries. The selected libraries that were used in the project are: (1) the Numpy Library, (2) Pandas Library, (3) Scikit Learn, (4) Matplotlib, (5) Seaborn and (6) Pandas Profiling.

2.3. Determining the Value of K

The value of K is a prerequisite to the k-means clustering algorithm. In determining the value of k, this project used two (2) methods: *the Elbow Method and the Silhouette Method*

- **Elbow method.** This plots the number of clusters according to the variation (dispersion) of the given data set. In the elbow method, determining the K value is through the intuition of picking the 'elbow' of the curve.
- **Silhouette Coefficient Method.** Is the method of determining the similarity of the values within the cluster (cohesion) and outside the cluster (separation). The value of this method ranges from -1 to 1. And, the closer the value to one(1), the higher the probability that the features are matched with the features within its cluster and poorly matched with the features of other clusters.

Figure 4 shows the optimal k value using elbow method and silhouette coefficient method in the data set.

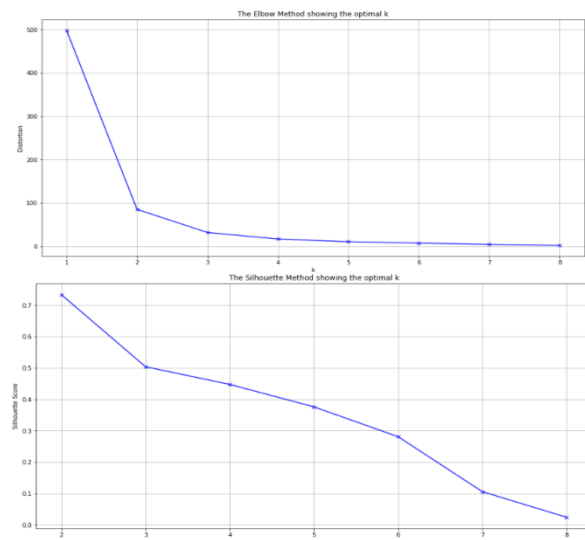


Figure 4. Optimal k Value Result

As shown in figure 4, the k value from the elbow method is between two (2) and three (3) where the sharp edge of the curve is seen. While, the silhouette score that is near the value of one (1) is also two (2) and three (3). Both the elbow method and silhouette coefficient method show the same value for k.

Between two(2) and three(3) the appropriate value for k is two(2). However, the value for the optimal k used in this project in all of the data set is three (3) which considered not only the result of the two (2) methods but also the 'ground' truth behind the data set being used.

2.4. Calculated Mean Score and Clustering Algorithms

These methods were used to evaluate the data set in order to answer the research questions. The result of these methods were analyzed and compared.

- **Calculated Mean Score.** The mean score is the average of the values. Also, referred to as the 'centroid'. The drawback of mean score is that it can be greatly influenced by outliers. However this drawback is not a problem in this dataset, because we used a summarized data set. The project made use of the calculated mean score in sorting the selected features to determine the highly and least affected groups in the dataset.

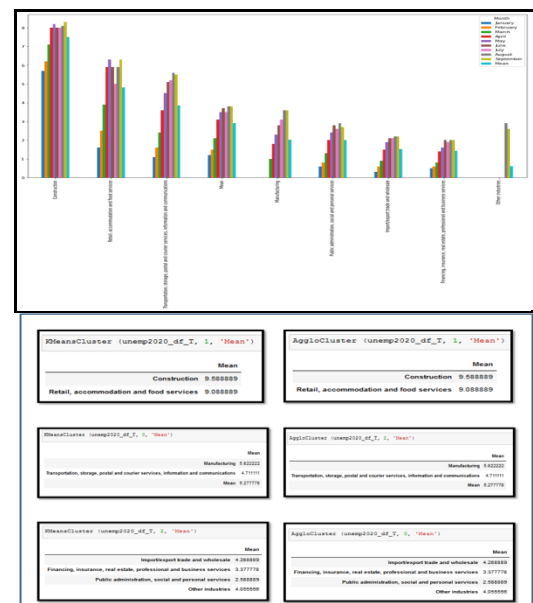


Figure 5. Use of the Calculated Mean Score in Evaluation

Figure 5 shows that the calculated mean score is used in evaluating the result of the bar graphs and clustering methods. As seen in the bar graph and the cluster lists, the mean score was used to sort the data of the data set and was also used to determine highly and least affected groups. The use of the mean score mostly aided in the insight formulation

- **Clustering Algorithms.** These methods find and group objects according to their relationship and dimensions, and these objects are different to other groups. (Sharma, 2020) In this project, two(2) kinds of clustering methods were used, these are:
 - **K-Means Clustering.** Is a partitioning clustering method where it partitions or separates features in the dataframe according to the given value of K (number of clusters) (Ciaburro, 2017). The objective of this method is to group the features where it will only belong to one(1) group and every group should have a purpose why it is grouped (Sharma, 2020). Scatter plot is used to present this grouping of k-means.

Figure 6 shows the features of the unemployment and underemployment dataset sgrouped by k-means, where k = 3. As seen in the scatter plots, k-means algorithm clustered the groups according to its dimensions where the data points are within the same range of both x and y.

One (1) cluster in both data sets are far from the other two clusters, which indicates its weak relationship to the other two clusters, whose plots are more closely grouped. Moreover, this also indicates that the far plotted cluster do not directly affects the other two groups as compared to these two (2) whose plots are near each other.

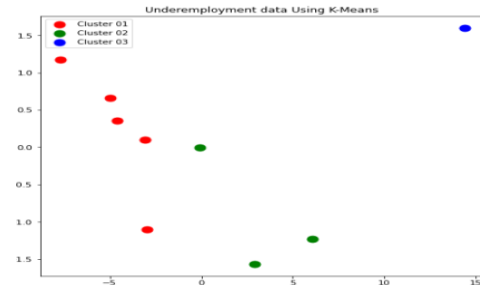


Figure 6.K-Means Clustering on the Datasets

- **Agglomerative Clustering.** Is a hierarchical clustering method which separates groups through bottom-up approach. This is done through separating all features, first, and then, merge it altogether. Dendrogram is used to present the grouping of agglomerative clustering.

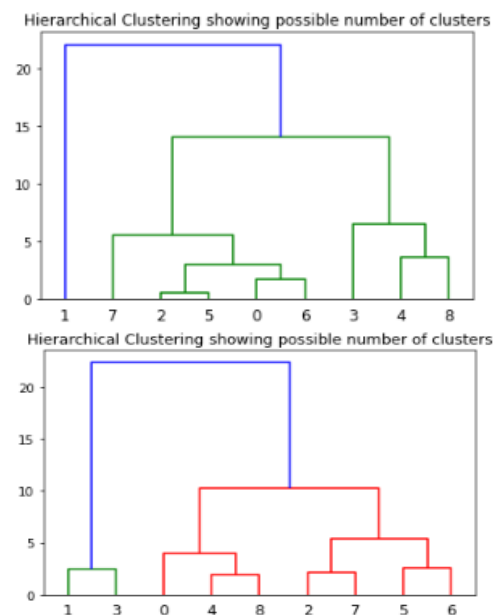


Figure 7. Hierarchical Clustering on the Datasets

Figure 7 shows the clustering of the unemployment and underemployment dataset being grouped by the hierarchical clustering. As shown in the dendrograms, the best number of clusters is three (3). And, there is also one (1) cluster that does not have correlation to the other two (2) clusters, while these two (2) can be grouped into one (1) cluster as they are closely correlated to each other.

Both the k-means and hierarchical clustering have the same presentation of the datasets. And, these results justified the intuition of choosing the number of clusters for both the k-means and agglomerative algorithm which is three (3).

3. RESULTS AND DISCUSSIONS

This section presents the results of the evaluation and analysis in the dataset. The first part presents the result of the dataset evaluation and then, followed by the answers to the research questions.

3.1. Data Evaluation Result

This subsection presents the result of the data preprocessing. A meticulous study and analysis were conducted to the dataset, aiming to deepen the understanding on the data and on the techniques used.

This project made use of the Pandas Profiling Report for a rapid and easy way to understand the features, its values and relationships. Figure 8 shows the result of the Pandas Profiling Report.

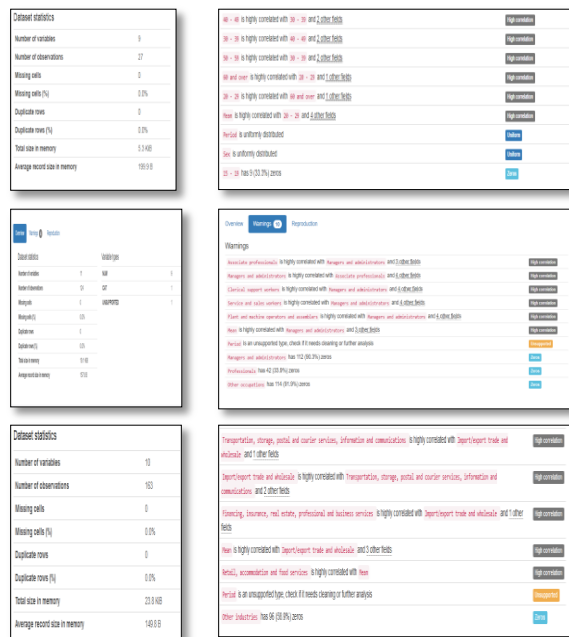


Figure 8. Pandas Profiling Report on the Datasets

As shown in figure 8, most of the features in the datasets are highly correlated to each other. And, only few features were marked with zero values.

Another way to understand this correlation is through heat map. Figure 9 shows the correlation of the features in all data sets through heat map.

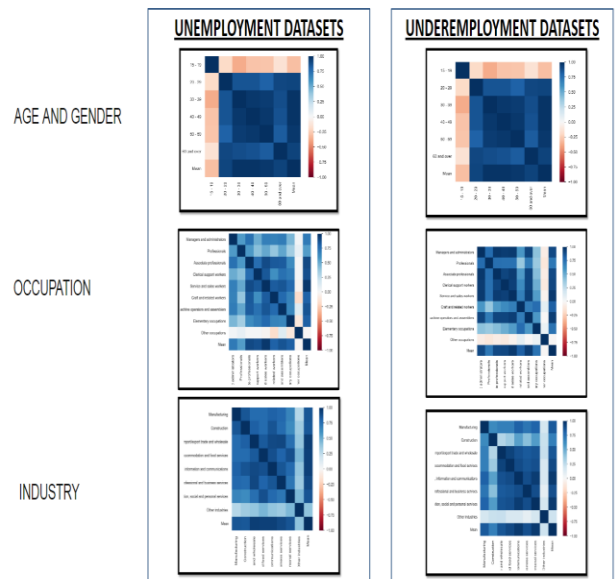


Figure 9. Correlation Heat maps of the Datasets

As shown in figure 9, most of the features in the data set are highly correlated having strong shades of color. But some features, at least one (1) or two (2), shows light shades of color indicating a weak correlation with other features.

As mentioned, some of the features have zero values. This is due to the very high or very low computed result by the Census and Statistics department of HKSAR. And, they considered these as erroneous values. The dataset that was gathered was already a summarized data. So, these columns with zero values were still included in the analysis and studied for possible insight.

Overall, the datasets, even they are in six (6) separate files, it is shown both in the report and in the correlation heat maps that they have strong relationships. Having these results, the following questions were in mind: What do these 'strong' correlations mean? What is the effect of these relationships to the overall result and finding?

These questions were answered in the next subsection of the paper.

3.2. Effect of Covid 19 to the Unemployment and Underemployment of HKSAR

COVID-19 started in December 2019. Being close to China, Hong Kong was one of the first countries who recorded the first few cases of COVID19. And, up until

now, the HKSAR government is in close monitoring of the status and number of COVID19 cases.

Figure 10 shows the status of the COVID 19 cases in HKSAR. Data was updated last 21 November 2020.

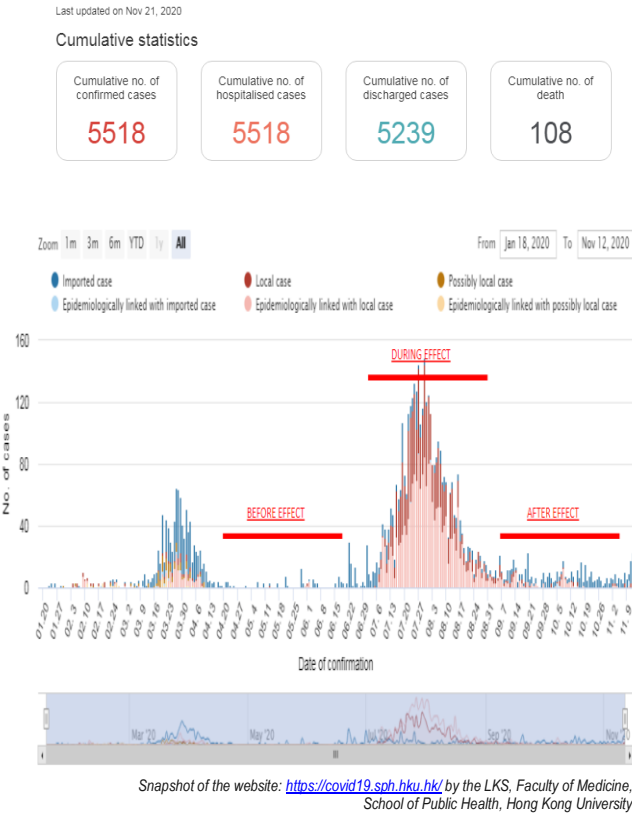
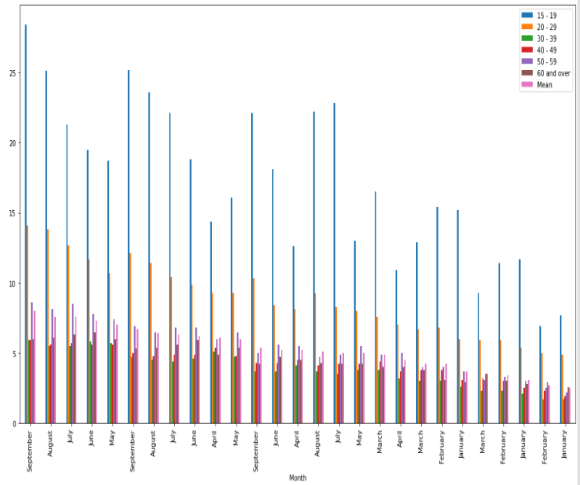


Figure 10. COVID 19 Status in HKSAR

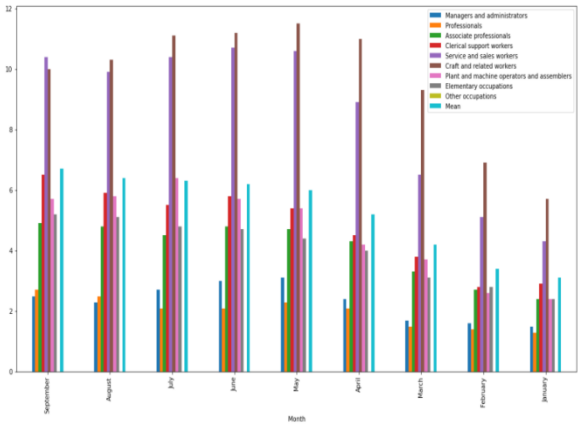
As shown in the figure above, as of the writing of this paper, the month with high number of cases falls in the later two (2) weeks of JUNE and the whole month of JULY. Studying this peak month of COVID19, this project analyzed whether this has a direct effect on the unemployment and underemployment of HKSAR. This effect is categorized into three (3) types: (1) Before effect, (2) During effect, and (3) After effect.

To understand this effect, the datasets was transposed and arranged according to months, from January to September 2020. Then, the columns of the datasets were sorted by the calculated mean score. Figure 11 shows the output.

AGE AND GENDER



OCCUPATION



INDUSTRY

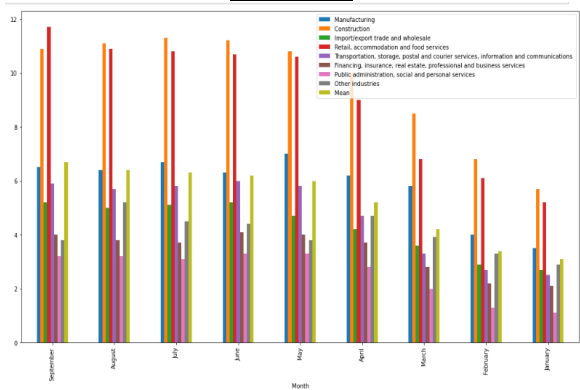


Figure 11. Datasets sorted by Months

Figure 11 shows the sorted datasets by months. As shown in the figure, all the datasets have the same months where rates in unemployment and underemployment are high. The top three (3) months that the unemployment and underemployment rates are high are the months of September, August and July, respectively.

Based on the 'peak' month of COVID19, the effect of the pandemic to the unemployment and underemployment of HKSAR is an AFTER-EFFECT where the 'peak' months of unemployment and underemployment were recorded after the peak month of COVID 19 which is July.

Overall, as the number of cases in COVID19 starts to increase and with a more strict quarantine and lockdown policies of the HKSAR government, the increase of the unemployment and underemployment rates happens, thereafter.

3.3. Effect of COVID19 to the dataset groups

This subsection presents the result of the after-effect of COVID19 to unemployment and underemployment of HKSAR. Specifically, in these three (3) groups: Age and Gender Group, Occupation Group, and Industry Group.

Moreover, there are two(2) kinds of evaluation this project conducted in order to determine the highest and least affected groups by COVID 19. These two(2) kinds of evaluation are: (1) Evaluation by Calculated Mean Score; and (2) Evaluation by Clustering Algorithms.

In the evaluation by Calculated Mean Score, the rates in the unemployment and underemployment data set were ranked according to the total mean score from the months of January to September, 2020.

While in the evaluation by Clustering Algorithms, there are two(2) methods that were used: the K-means Clustering and the Agglomerative Clustering. As discussed in the methodology section, these two methods have different approaches in grouping objects. K-means clustering is dependent on the defined k value while the agglomerative uses a top-down approach. (See Methodology)The next section presents both the result of these two(2) kinds of evaluation.

3.3.1. Age and Gender Group

The age was grouped in the data set as follows:

1. 15 – 19
2. 20 – 29
3. 30 – 39
4. 40 – 49
5. 50 – 59
6. 60 and over

While the gender was grouped in the data set as follows:

1. Male
2. Female
3. Both Sexes (computed mean score of both male and female values in the data set)

I. Evaluation by Calculated Mean Score (Age)

Figure 12 and 13 show the ranking of age that are affected by the COVID 19 pandemic by calculated mean score.

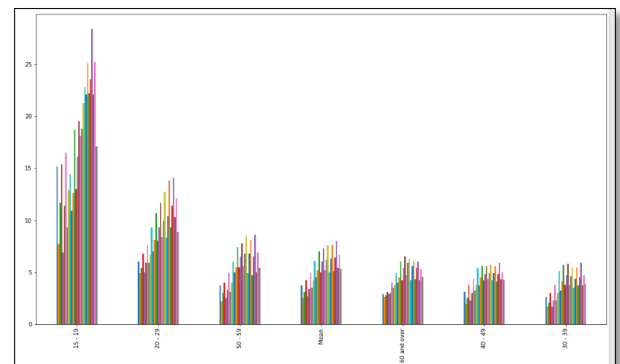


Figure 12. Unemployment Rates by Age Group Using Mean Score Evaluation

As shown in figure 12, the top three (3) age brackets that were *highly affected* by COVID 19 in the unemployment data set are:

1. 15-19;
2. 20-29; and
3. 50-59

While the *least affected* by COVID 19 in the unemployment data set are:

1. 30-39;
2. 40-49; and
3. 60 and over

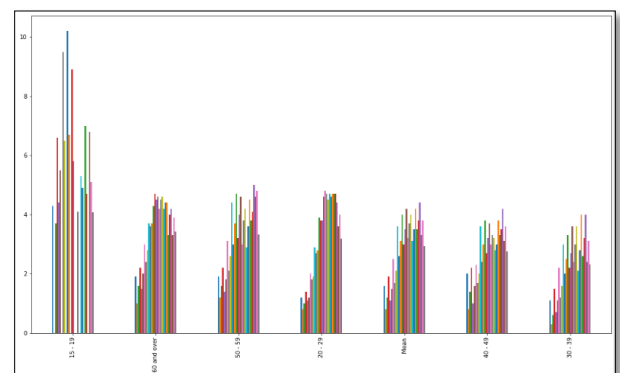


Figure 13 . Underemployment Rates by Age Group

Using Mean Score Evaluation

As shown in figure 13, the top three (3) age brackets that were *highly affected* by COVID 19 in the underemployment data set are:

1. 15-19;
2. 60 and over; and
3. 50-59

While the *least affected* by COVID 19 in the underemployment data set are:

1. 30-39; and
2. 40-49

II. Evaluation by Clustering Algorithms (Age)

Figure 14 and 15 show the result of the clustering algorithms (K-Means and Agglomerative Clustering) on the datasets.



Figure 14. Unemployment data set by Age Group using Clustering Algorithms

As shown in figure 14, in the unemployment data set, the highly affected age bracket is:

1. 15-19

While the least affected age brackets are:

1. 30-39; and
2. 40-49
3. 50-59
4. 60 and over

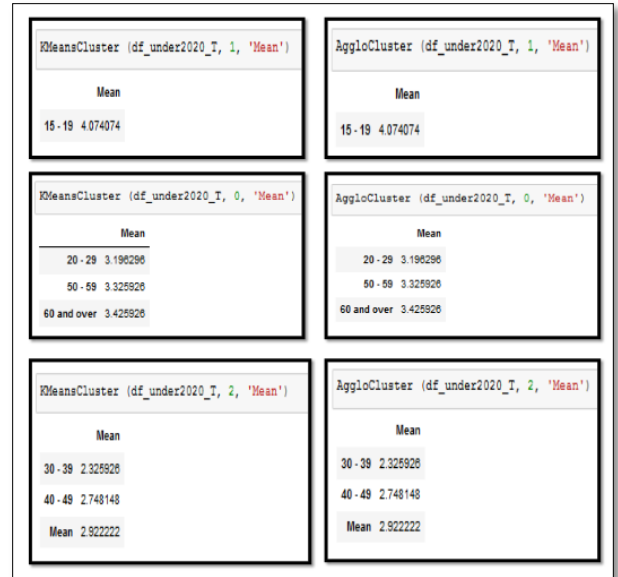


Figure 15. Underemployment data set by Age Group using Clustering Algorithms

As shown in figure 15, in the unemployment data set, the highly affected age bracket is:

1. 15-19

While the least affected age brackets are:

1. 30-39; and
2. 40-49

III. Evaluation by Mean Score (Gender)

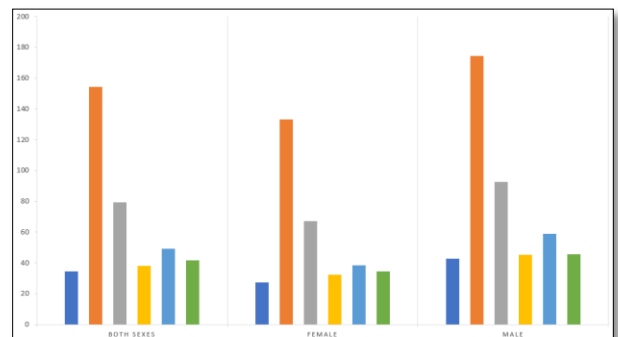


Figure 16. Unemployment Rates by Gender

As shown in figure 16, in the unemployment data set, the highly affected gender is Female while the least affected is Male.

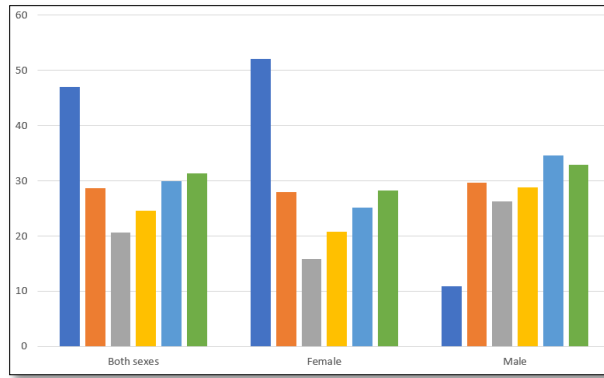


Figure 17. Underemployment Rates by Gender

As shown in figure 17, in the underemployment data set, the highly affected gender is also, Female while the least affected is also, Male.

Table 1. Evaluation Summary of Age Group

Weight	Evaluation by Calculated Mean Score		Evaluation by Clustering Algorithms	
	Unemployment	Underemployment	Unemployment	Underemployment
Highly Affected	1. 15-19 2. 20-29 3. 50-59	1. 15-19 2. 60 and over 3. 50-59	1. 15-19	1. 15-19
Least Affected	1. 30-39 2. 40-49 3. 60 and over	1. 30-39 2. 40-49	1. 30-39 2. 40-49 3. 50-59 4. 60 and over	1. 30-39 2. 40-49

Based on the results of the two(2) evaluations, the highly affected age group on both Unemployment and Underemployment data sets is **Ages 15-19**

While, the least affected age groups on both unemployment and underemployment are the following:

1. **Ages 30-39**
2. **Ages 40-49**

Moreover, in the gender group data, the most affected is **Female** and least affected is **Male**.

3.3.2. Occupation Group

The occupation was grouped in the data set as follows:

1. Managers and administrators
2. Professionals
3. Associate professionals
4. Clerical support workers
5. Service and sales workers
6. Craft and related workers
7. Plant and machine operators and assemblers
8. Elementary occupations
9. Other occupations

I. Evaluation by Calculated Mean Score (Occupation)

Figure 18 and 19 show the ranking of occupation that are affected by the COVID 19 pandemic by calculated mean score.

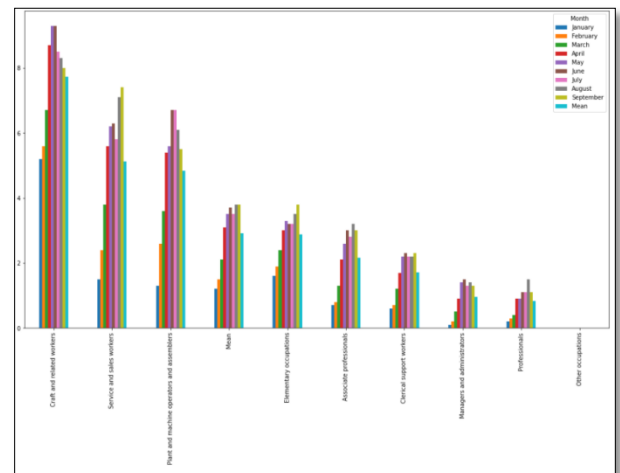


Figure 18. Unemployment Rates by Occupation Group Using Mean Score Evaluation

As shown in figure 18, the top three (3) occupations that were highly affected by COVID 19 in the unemployment data set are:

1. Craft and related workers;
2. Service and sales workers; and
3. Plant and machine operators and assemblers

While the least affected by COVID 19 in the unemployment data set are:

1. Managers and administrators
2. Professionals
3. Other occupations

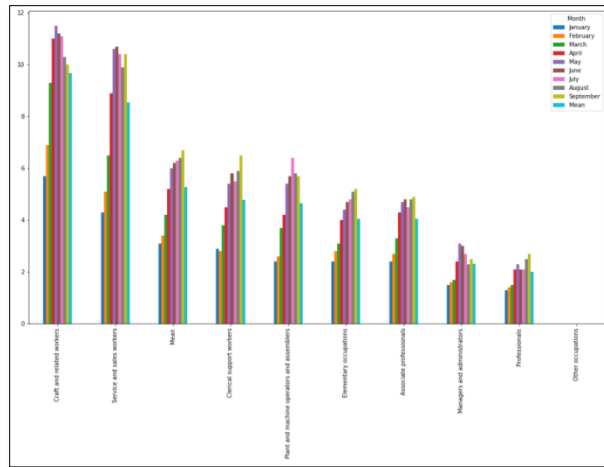


Figure 19. Underemployment Rates by Occupation Group Using Mean Score Evaluation

As shown in figure 19, the top two (2) occupations that were highly affected by COVID 19 in the underemployment data set are:

1. Craft and related workers; and
2. Service and sales workers

While the least affected by COVID 19 in the underemployment data set are:

1. Managers and administrators
2. Professionals
3. Other Industries

II. Evaluation by Clustering Algorithms (Occupation)

Figure 20 and 21 show the result of the clustering algorithms (K-Means and Agglomerative Clustering) on the datasets.

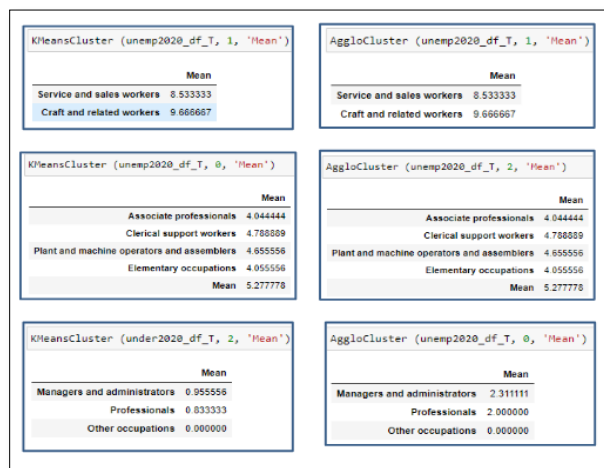


Figure 20. Unemployment data set by Occupation Group using Clustering Algorithms

As shown in figure 20, in the unemployment data set, the highly affected occupations are:

2. Craft and related workers; and
3. Service and sales workers

While the least affected occupations are:

1. Managers and administrators
2. Professionals
3. Other Industries

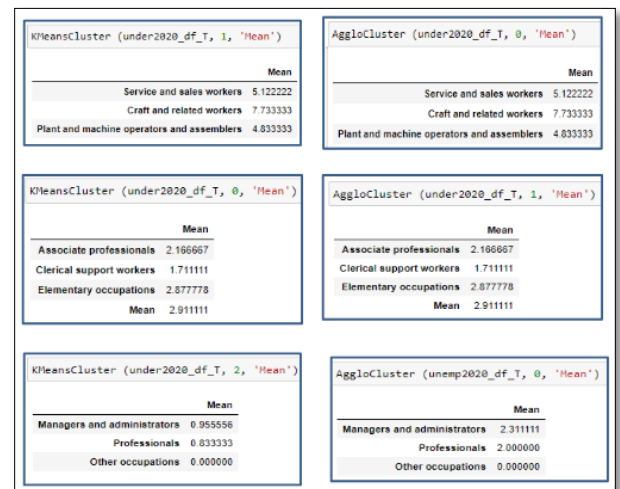


Figure 21. Underemployment data set by Occupation Group using Clustering Algorithms

As shown in figure 21, in the underemployment data set, the highly affected occupations are:

1. Craft and related workers;
2. Service and sales workers; and
3. Plant and machine operators and assemblers

While the least affected occupations are:

1. Managers and administrators
2. Professionals
3. Other Industries

Table 2. Evaluation Summary of Occupation Group

Weight	Evaluation by Calculated Mean Score		Evaluation by Clustering Algorithm	
	Unemployment	Underemployment	Unemployment	Underemployment
Highly Affected	1. Craft and related workers	1. Craft and related workers	1. Craft and related workers	1. Craft and related workers
	2. Service and Sales workers	2. Service and Sales workers	2. Service and Sales workers	2. Service and Sales workers
	3. Clerical support workers	3. Plant and Machine operators and assemblers		3. Plant and Machine operators and assemblers
Least Affected	1. Professionals	1. Professionals	1. Professionals	1. Professionals
	2. Managers and administrators	2. Managers and administrators	2. Managers and administrators	2. Managers and administrators
	3. Associate professionals	3. Clerical support workers	3. Other Occupations	3. Other Occupations

Based on the results of the two (2) evaluations, the highly affected occupations are the following:

1. On Unemployment data set: 'Craft and related workers' and 'Service and sales workers'
2. On Underemployment data set: 'Craft and related workers'; 'Service and sales workers'; and 'Plant and machine operators and assemblers'
3. On both Unemployment and Underemployment data sets: **'Craft and related workers' and 'Service and sales workers'**

While, the least affected occupations on both unemployment and underemployment are the following:

3. **Managers and administrators**
4. **Professionals**

Other industries was listed as least affected occupation but this feature has 91.1% zero values. Thus, it was removed from the list.

3.3.3. Industry Group

The groups of industries in the HKSAR were clustered and enumerated by the Census and Statistics Department in their published document entitled, Hong Kong Standard Industrial Classification Version 2.0. In this document, the industries were enumerated into eight (8) groups. These are:

1. Manufacturing
2. Construction
3. Import/export trade and wholesale
4. Retail, accommodation and food services
5. Transportation, storage, postal, and courier services, information and communications
6. Financing, insurance, real estate, and professional and business services
7. Public administration, social and personal services
8. Other Industries

I. Evaluation by Calculated Mean Score (Industry)

Figure 22 and 23 show the ranking of the group of industries that are affected by the COVID 19 pandemic.

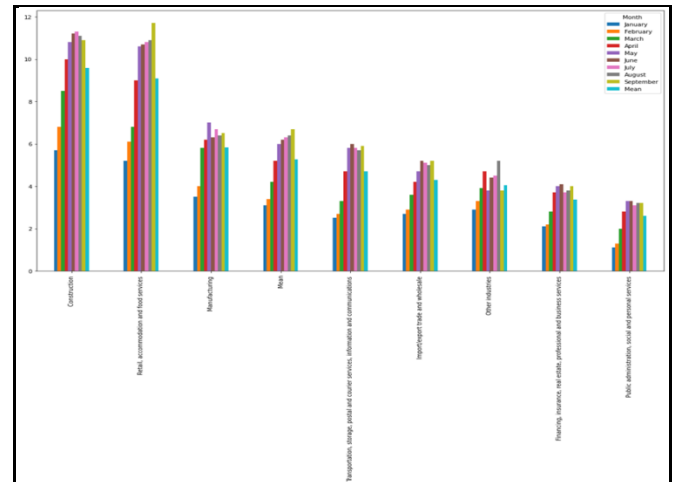


Figure 22. Unemployment Rates by Industry Group Using Mean Score Evaluation

As shown in figure 22, the top three(3) industry groups that were *highly affected* by COVID 19 in the unemployment data set are:

1. Construction;
2. Retail, accommodation and food services; and
3. Manufacturing

While the *least affected* by COVID 19 in the unemployment data set are:

1. Public administration, social and personal services
2. Financing, insurance, real estate, and professional and business services
3. Other Industries

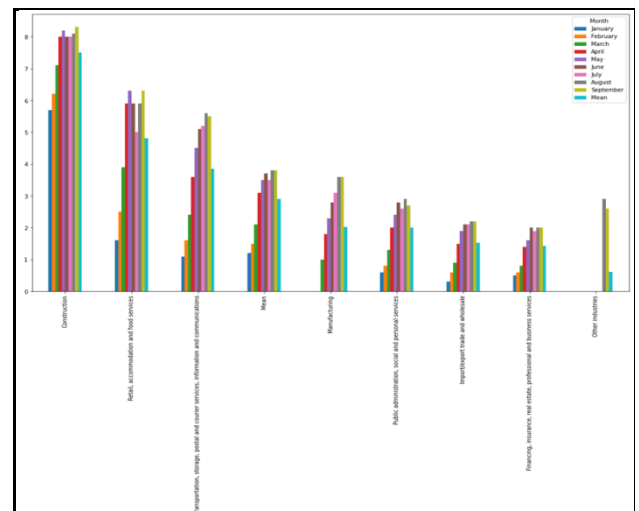


Figure 23. Underemployment Rates by Industry Group using Mean Score Evaluation

In the underemployment data set, the top three(3) industry groups that were highly affected by COVID 19, as shown in figure_ are:

1. Construction;
2. Retail, accommodation and food services; and
3. Transportation, storage, postal, and courier services, information and communications

While the least affected groups are:

1. Other Industries
2. Financing, insurance, real estate, and professional and business services
3. Import/export trade and wholesale

II. Evaluation by Clustering Algorithms (Industry)

Figure 24 and 25 show the result of the clustering algorithms (K-Means and Agglomerative Clustering) on the datasets.



Figure 24. Unemployment data set by Industry Groups using Clustering Algorithms

As shown in figure 24, in the unemployment data set, the highly affected industry groups are:

1. Construction
2. Retail, accommodation and food services

While the least affected industry groups are:

1. Import/export trade and wholesale
2. Financing, insurance, real estate, and professional and business services
3. Public administration, social and personal services
4. Other industries



Figure 25. Underemployment data set by Industry Groups using Clustering Algorithms

As shown in figure 25, In the underemployment data set, the highly affected industry group is:

1. Construction

While the least affected industry groups are:

1. Manufacturing
2. Import/export trade and wholesale
3. Financing, insurance, real estate, and professional and business services
4. Public administration, social and personal services
5. Other industries

Table 3. Evaluation Summary of Industry Group

Weight	Evaluation by Calculated Mean Score		Evaluation by Clustering Algorithms	
	Unemployment	Underemployment	Unemployment	Underemployment
Highly Affected	1. Construction; 2. Retail, accommodation and food services; and 3. Manufacturing	1. Construction; 2. Retail, accommodation and food services; and 3. Transportation, storage, postal, and courier services, information and communications	1. Construction 2. Retail, accommodation and food services	1. Construction
Least Affected	1. Public administration, social and personal services; 2. Financing, insurance, real estate, and professional and business services; and 3. Other Industries	1. Other Industries; 2. Financing, insurance, real estate, and professional and business services; 3. Import/export trade and wholesale	1. Import/export trade and wholesale 2. Financing, insurance, real estate, and professional and business services 3. Public administration, social and personal services 4. Other industries	1. Manufacturing 2. Import/export trade and wholesale 3. Financing, insurance, real estate, and professional and business services 4. Public administration, social and personal services 5. Other industries

Based on the results of the two(2) evaluations, the highly affected industry groups are the following:

1. On Unemployment data set: Construction and Retail, accommodation and food services
2. On Underemployment data set: Construction

- On both Unemployment and Underemployment data sets: **Construction**

While, the least affected industry groups are the following:

- On Unemployment data set: Public administration, social and personal services; and Financing, insurance, real estate, and professional and business services.
- On Underemployment data set: Financing, insurance, real estate, and professional and business services; and Import/export trade and wholesale
- On both Unemployment and Underemployment data sets: **Financing, insurance, real estate, and professional and business services.**

3.4. Overall Summary

This subsection presents the overall insight and business value formulated based on the result of the analysis and evaluation.

Table 4. Summary Result of the Effect of Covid-19 on Groups

Weight	Age and Gender	Occupation	Industry
Highly Affected	15 - 19 / Female	Service and sales workers Craft and related workers	Construction Retail, accommodation and food services
Least Affected	30 - 39 / Male 40 - 49	Managers and administrators Professionals	Financing, insurance, real estate, and professional and business services

Highly Affected Groups

1. Ages 15 - 19 (Students)

Students in Hong Kong are allowed to take part time jobs on summer months, June to August. These part time jobs available for students are in the 'Retail, accommodation and food services', and roles under 'Service and Sales Workers'. Students were asked to stay-at-home during and latter months of COVID19 peak month, which is why they were listed as highly affected by COVID19.

2. Construction Industry and Craft Related Workers

Occupations that are available in Construction industry is categorized under 'Craft Related Workers'.

Both of which are placed as highly affected by COVID19.

3. Female

Majority of the labour force in Hong Kong are females. With this claim by the Census and Statistics Department (2012) of Hong Kong, women were listed as highly affected by COVID19.

Least Affected Groups

1. Ages 30-39 (Professionals) and Ages 40-49 (Managers and Administrators)

Professionals, managers and administrators who are mostly under the 'Financing, insurance, real estate, and professional and business services' are least affected by COVID19. Not much impact happened to these types of workers as offices in Hong Kong continue to operate and they are more capable of working remotely.

2. Male

Not only that the labour force is majority comprised of women but also, most of the Professionals, Managers and Administrators were recorded to be male as reported by the Census and Statistics Department of Hong Kong (2012). With this, male is least affected by COVID19.

Business Value

1. HKSAR Government

In light of the results, the HKSAR government may do the following:

- Schedule assistance programs after the peak month of a pandemic because its effect to the unemployment and underemployment happens after a high record of pandemic cases.
- Prioritize the Construction (Craft and related workers) and 'Retail, accommodation and food services' (Service and sales workers) industries in designing support programs and in release of relief funds.
- The HKSAR government released relief funds not only to local employees but also to students in this COVID19 pandemic which amounts to 5,000 HKD (30,000 Php). The giving of relief funds to students must be continued during pandemic events.

2. Companies

Based on the results, the companies are recommended to do the following:

- If under 'Retail, accommodation and food services', specifically food services, companies may entice their customers through discounts and/or freebies when they order food through delivery and take out services.
- If under 'Construction' industry, one of the crucial things that the company considers is the completion of their projects. An extension of time can be included to the contract in the future if an unexpected pandemic might happen.
- Also, companies under these industries may also consider acquiring unemployment insurance for its employees.

3. Individuals

Based on the results, the employees are recommended to do the following:

- If the company does not provide unemployment insurance, the employee may acquire for himself.
- The employee should also start considering advancement to his/her career, especially if position is in the 'Service and Sales Workers' and 'Craft and related workers'. Because those who are least affected by a pandemic event, such as COVID19, are those in the high rank positions such as Professionals and 'Managers and Administrators'.

4. CONCLUSION

4.1. On Methods

Based on the results from the two (2) clustering algorithms, there is no difference between how the dataset groups (age and gender, occupation, and industry) were clustered. Both the K-means Clustering and Agglomerative Clustering have identical sets of clustered groups. Thus, this shows that the groups are with strong relationship where their dimensions and values are correlated.

Based on the results from the Evaluation Using Computed Mean Score and Evaluation Using Clustering Algorithms, both methods provided different sets of results. But the highest among the highly affected and highest among the least affected are present in the results of these two (2) methods. With that, the findings and insights of this paper were based on a strong level

of evidence as these are reflected on both evaluation methods.

4.2. On the Findings

COVID19 has an after-effect on the unemployment and underemployment of HKSAR. This effect is greatly felt specifically on the Construction and 'Retail, accommodation and food services' industries; employees with roles as service, sales, craft and related workers; and, students.

However, those who are in the 'Financing, insurance, real estate, and professional and business services' industry and having a role as professionals, managers and administrators are more safe as compared to other industries and positions.

4.3. On Business Value

Ages 15-19 are listed to be as highly affected by COVID19, but students are not mainly part of the labour force. So, mitigating strategies and programs can be focused on the highly affected occupation group and industry group.

5. REFERENCES

1. Census and Statistics Department (2012). Women and Men in Hong Kong Key Statistics. General Household Survey Section
2. Chan, M. (2020, June 16). *Nikkei Asia*. Retrieved from Hong Kong's unemployment rate jumps to highest level in 15 years: <https://asia.nikkei.com/Economy/Hong-Kong-s-unemployment-rate-jumps-to-highest-level-in-15-years>
3. Ciaburro, G. (2017). *Partitioning-based clustering methods - K-means algorithm*. MATLAB for Machine Learning.
4. Delvecchio L. (2020, Apr 12). *Understanding the Process of Data Mining. Digging Deep for Valuable Business Insights*. Purchase Control. Business Strategy.
5. Galant, J. et.al. (2020). *Temporary Unemployment and Labor Market Dynamics during the COVID-19 Recession*. National Bureau of Economic Research Working Paper Series.
6. Kotu and Deshpande (2015). *Data Mining Process. Predictive Analytics and Data Mining*. ScienceDirect
7. *Moody's Analytics*. (2020). Retrieved from Hong Kong - Unemployment Rate: <https://www.economy.com/hong-kong/unemployment-rate>
8. Sharma, R. (2020, Jan 20). *Cluster Analysis in Data Mining: Applications, Methods & Requirements*. UpGradblog
9. Yau, C. (2020, Oct 21). *Coronavirus: Hong Kong jobless rate climbs to 6.4 per cent, highest in nearly 16 years*. South China Morning Post: Hong Kong Economy.

