Paper 013-2020

An Analysis of Flight Operations in Times of the COVID-19 Pandemic

Chung Wei Han Shaun, Son Yejin, Singapore Management University

ABSTRACT

The coronavirus disease 2019 (COVID-19) pandemic has taken a toll on the global aviation industry, with travel bans and a fall in demand for flights resulting in at least an 80% fall in flight numbers. Most available studies are currently focusing on the analysis of current flight trends and the financial and economic impact it will have. This paper however employs dynamic time warping (DTW) techniques for the comparison of temporal flight data among airlines. The study provides insight into similarities between the impact to flight patterns as a result of the COVID-19 Pandemic for different Airlines. In addition, it attempts to discover associations between similar flight patterns and other supplementary factors, such as COVID-19 cases in the headquartered country, geographical proximity, fleet size and star rating. The paper can provide insights that can be useful for policy planning by governments and corporations, or by investors and financial institutions.

INTRODUCTION

In response to the alarming rise in number of cases and fatalities, the World Health Organization officially declared the COVID-19 outbreak a pandemic. In a bid to stop cross-border spreading of the virus, governments across the globe have looked towards restricting travel. In addition, driven by fear of contagion, consumers are also deferring travel plans leading to a drastic fall in the demand for flights. This double pronged effect is taking a toll on the global aviation industry which has already seen 80% fall in the number of flights at the start of the Q2'20 (IATA, 2020).

OBJECTIVES

This paper examines the connections between daily flight numbers across global airlines in response to the worsening COVID-19 outbreak. Through Dynamic Time Warping (DTW), the paper aims to agglomerate airlines into clusters based on similarities within their flight patterns across Q1'20. Analysis is then carried out on these identified clusters to identify similar characteristics that could have contributed to the similar flight patterns. The paper will focus on airlines with a rating of between 5 stars to 3 stars as rated on Skytrax.

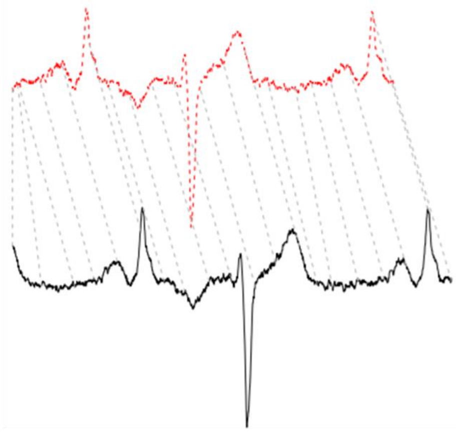
LITERATURE REVIEW

Previous literature analyzing the COVID-19 outbreak's impact on the airline industry has mainly been focused on analyzing current trends and its expected economic impacts. For example, the International Air Transport Association (IATA) focuses their analysis on changes in the number of flights at a regional level and the resulting actual and forecasted impact on passenger revenues and revenue passenger kilometers (IATA,2020). Other literature also covers the expected differences in recovery periods (VariFlight, 2020; IATA, 2020)

TIME-SERIES ANALYSIS

This paper takes a different approach and focuses on examining interconnectedness of flight patterns between airlines across 3 months. Given the temporal nature of the data, examining similarities using Euclidian distances could result in a less than optimal comparison as standard clustering techniques are not optimized to analyze non-static data.

Initially, introduced for isolated word recognition (Velichko and Zagoruyko,1970), DTW as demonstrated by Berndt & Clifford (1994), allows for better alignment and comparison of temporal data as the dimension of time is considered in the distance measure. An example of such alignment of is as follows:



Source: Giorgino, 2009

Recent literature has also demonstrated the ability of DTW in the adjustment of nonsynchronous effects present in temporal data (Huang & Lu, 2020). DTW has since been applied in time-series classification for many other purposes (Chen, Tseng, Ke & Sun, 2011; Jeong, Jeong & Omitaomu, 2011).

Hence, given that there may be variation in airlines' response speed, DTW will be used for the formation of clusters.

DATASETS

The main dataset used in this paper is obtained from Flightstats. It contains time-stamped information of flights between January 2020 to March 2020, with useful information such as:

- Operating airline IATA & ICAO Code for airline identification
- Flight Status:

Value	Description
A	Active
C	Cancelled
D	Diverted
DN	Data Source Needed
L	Landed
NO	Not Operational
R	Redirected
S	Scheduled
U	Unknown

- Flight scheduled departure information

In addition, other supplementary data sources were referred to which include:

- IATA website and AvCodes - provides information regarding IATA & ICAO codes (e.g. airline name, country)
- Skytrax – provides the star-rating of airlines
- John Hopkins University Coronavirus Resource Center – provides time-series data of confirmed COVID-19 cases

DATA PREPARATION

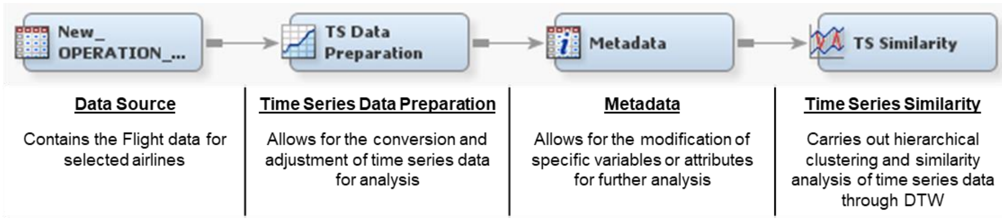
The FlightStats data from January to March 2020 contains each flight's flight status, which can be categorized into 9 different status. To focus on operated flights, the data with "C (Cancelled)", "NO (Not Operational)", and "U (Unknown)" status were removed from the dataset.

Moreover, to focus on airlines with robust operational schedule, airlines that went bankrupt, airlines with lower than 3-star rating from Skytrax or airlines with less than 50 fleet size are removed from the dataset. As a result of the data cleaning, we have number of daily operated flights for 91 days, from 1st January 2020 to 31st March 2020 of 67 airlines.

Data cleaning was performed using JMP Pro 15 and the final dataset was saved as SAS data table (.sas7bdat) for easier analysis with SAS Enterprise Miner 14.

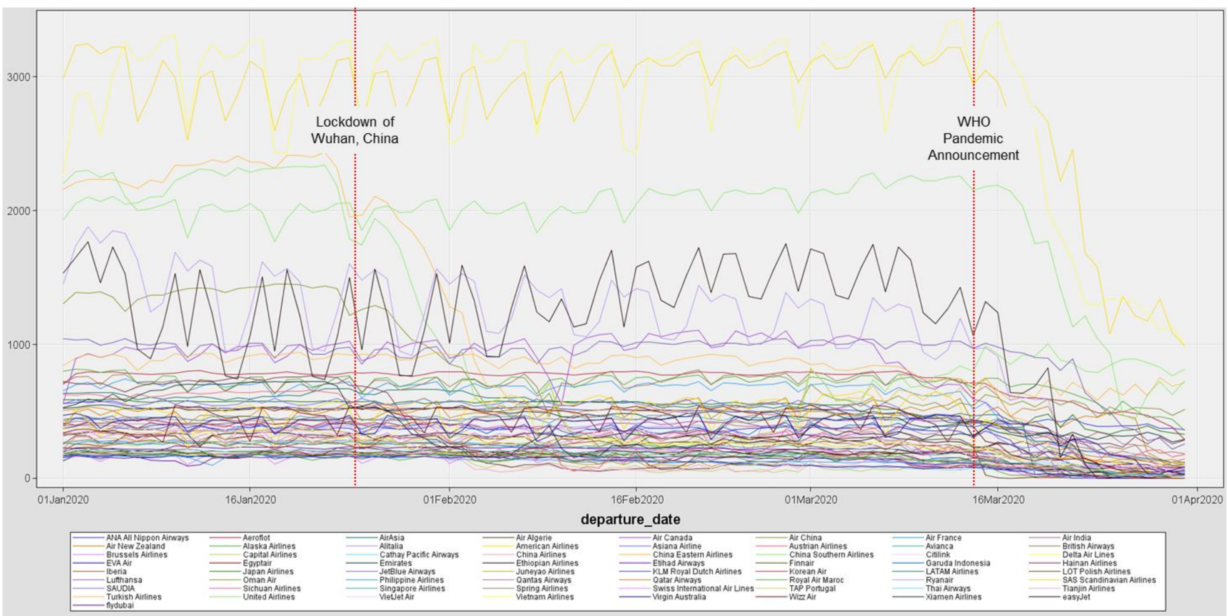
METHODOLOGY

DTW was carried out using the TS Similarity node in SAS Enterprise Miner Workstation 14.1 ("EM"). The TS Similarity node calculates similarity measures for timestamped data and performs hierarchical clustering to form clusters from the dataset. The following diagram depicts the nodes used in our data analysis in EM:



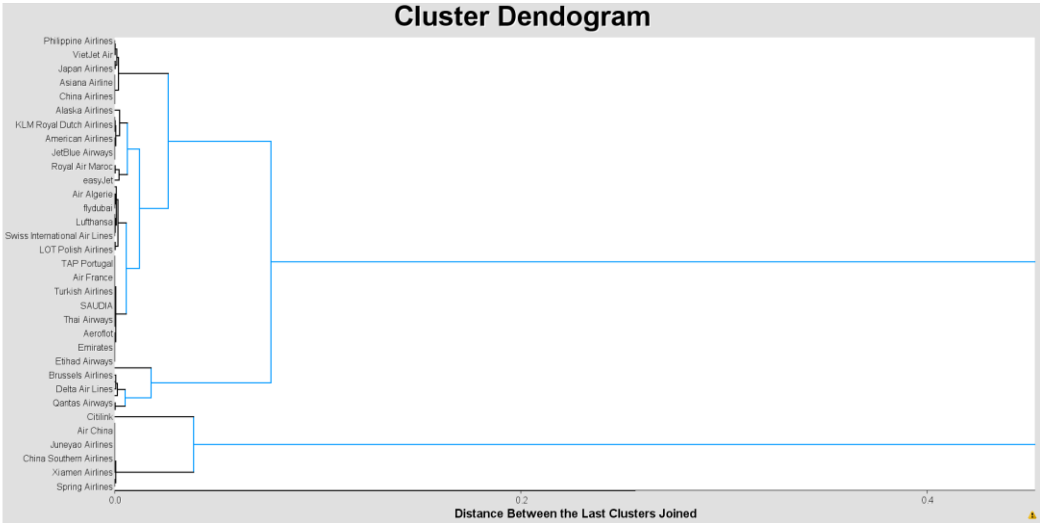
FINDINGS AND DISCUSSIONS

Through the TS Data Preparation Node, one single graphical window was generated to show the time series pattern for all airlines. Aside from cyclical changes in the number of operated flights, there are two points in time that shows a significant drop in flight numbers. The first occurred around 23rd January 2020 - the first day of the lockdown of the Chinese city of Wuhan, in an attempt to quarantine the city due to the outbreak of COVID-19. The second fall occurred around 12th March 2020 - the day WHO officially declared the COVID-19 outbreak a pandemic.

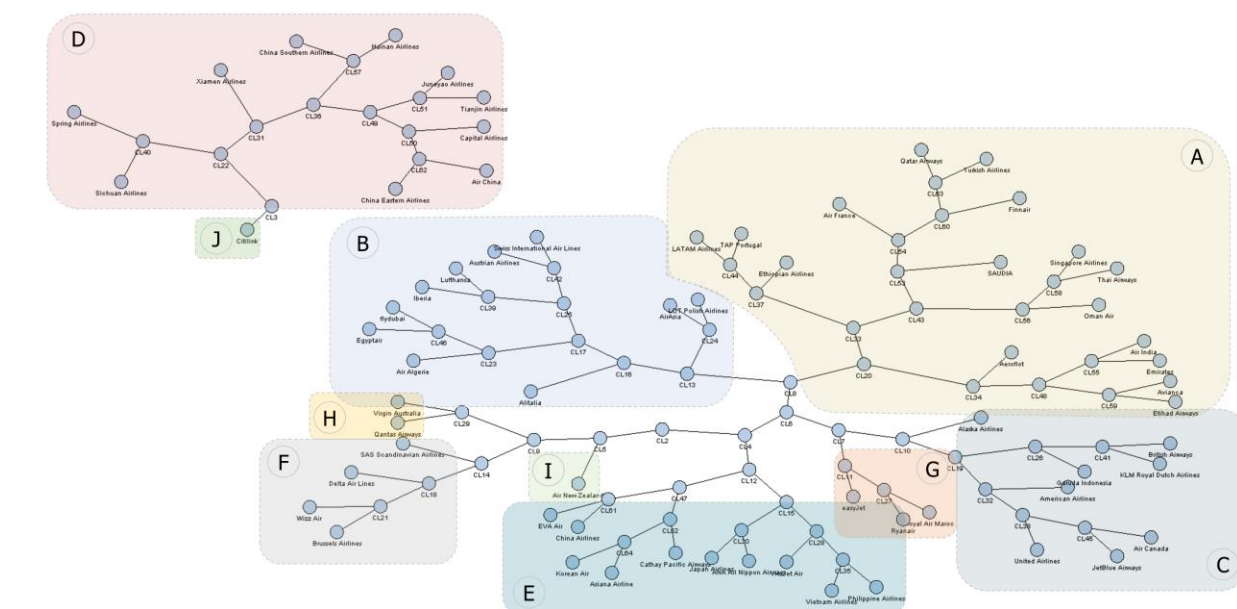


TIME SERIES CLUSTERING RESULTS

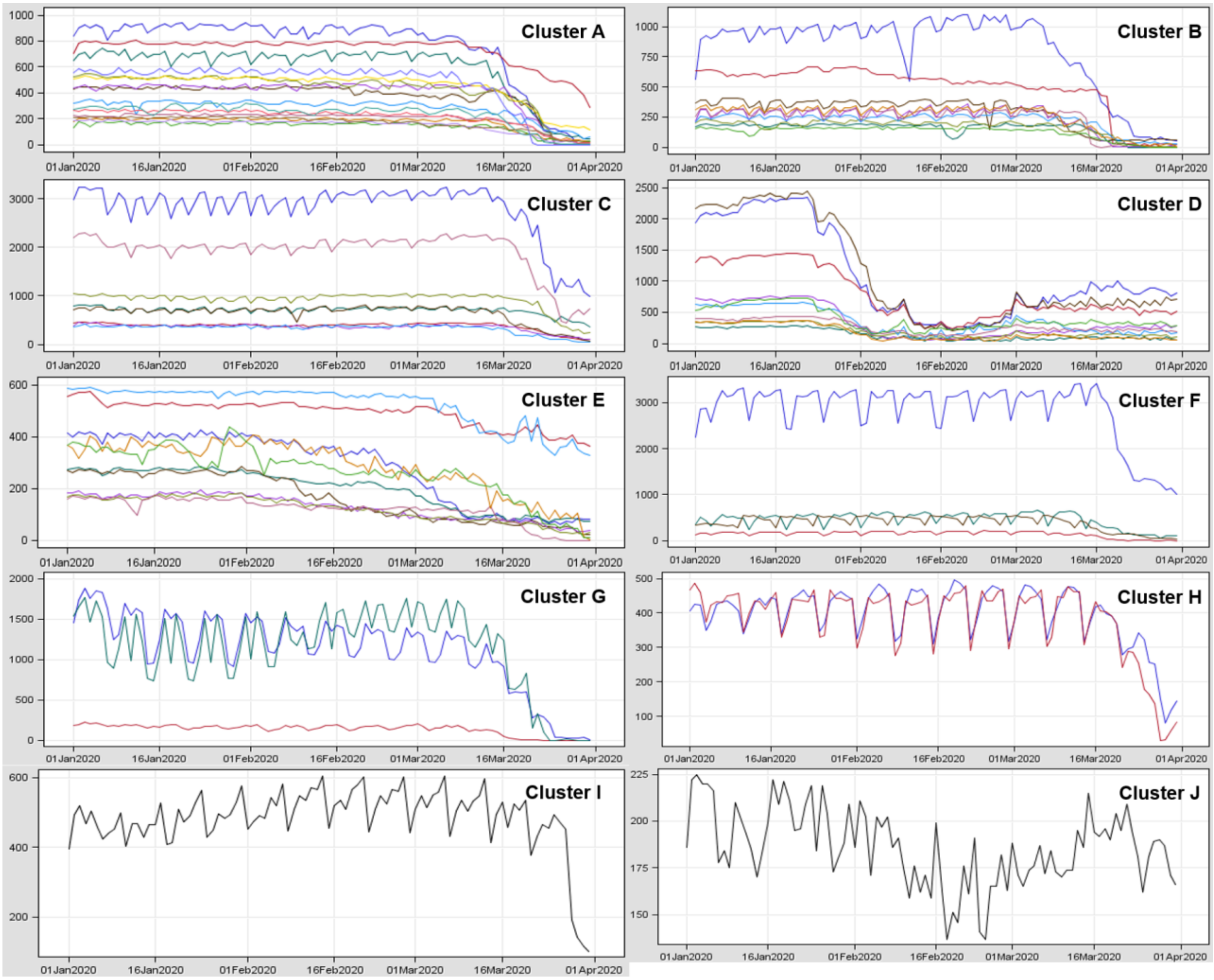
Using the TS Similarity Node, time-series clustering analysis was performed through hierarchical clustering. In determining the optimal number of clusters for the analysis, multiple runs were carried out with varying the number of clusters between 5 to 12. It was observed that any further splitting from 10 clusters would only result in the formation of clusters consisting of only 1 airline. This would not provide any material insight to our analysis and hence it was decided that 10 clusters would be the optimal number of clusters.



The following constellation plot shows there are two main clusters, with each cluster generating more branches to reduce distance between the clusters joined. In total, 10 clusters were identified with member airlines demonstrating specific similar characteristics with other airlines in the cluster.



The following graphs show the different flight numbers across the 3 month time period of airlines within their respective clusters. Do note that although it might appear to the naked eye that some airlines are not demonstrating similar flight patterns, it is mainly due to the difference in scale of the number of flights.



This paper also attempts to provide a description of observable similarities between airlines within the same cluster. The following table contains a list of airlines within each cluster, with a description of the observed similarities.

Cluster	Composition	Description
A	Aeroflot, Air France, Air India, Avianca, Emirates, Ethiopian Airlines, Etihad Airways, Finnair, LATAM, Oman Air, Qatar Airways, SAUDIA, Singapore Airlines, TAP Portugal, Thai Airways, Turkish Airlines	Cluster A consists of airlines that have wide network globally. Compared to other airlines that are focusing on one continent or regional network, these airlines are legacy airlines of the country that will focus more on international routes and intercontinental flights. As a result, we are not able to see the cyclic patterns that appeared during January and February after the significant drop around 18th March.
B	AirAsia, Air Algerie, Alitalia, Austrian Airlines, Egyptair, Iberia, LOT Polish Airlines, Lufthansa, Swiss International Airlines, flydubai	Cluster B consists of airlines that showed a drop almost to 0 after around 18th March. While airlines in Cluster A would still operate several key routes, airlines in cluster B shows the drop to almost 0 after the drop started such as Austrian Airlines, Air Algerie, Egyptair and flydubai.
C	Air Canada, Alaska Airlines, American Airlines, British Airways, Garuda Indonesia, JetBlue Airways, KLM Royal Dutch Airlines, United Airlines	Cluster C consists of most airlines headquartered in American continent and other legacy carriers such as British Airways, Garuda Indonesia and KLM. The significant drop started around mid-March, like Cluster A or Cluster B. However, the cyclic pattern during January and February appears to be different from these other clusters.
D	Air China, Capital Airlines, China Eastern Airlines, China Southern Airlines, Hainan	Cluster D consists of all airlines from mainland China. Pattern shows that there is a huge drop starting from end of January and the drop of

	Airlines, Juneyao Airlines, Sichuan Airlines, Spring Airlines, Tianjin Airlines, Xiamen Airlines	the number continues for more than one month. Starting from March onwards, it shows a slight recovery. It is due to the airlines in mainland China slowly started to operate domestic flights, while maintaining the direction of minimizing international flights.
E	Asiana Airlines, Korean Air, ANA All Nippon Airways, Japan Airlines, Cathay Pacific Airways, China Airlines, EVA Air, Philippine Airlines, VietJet Air, Vietnam Airlines	Cluster E consists of airlines that are closely related to the Chinese market. Airlines in Korea, Japan, Vietnam, Philippines and Greater China area operate significant number of routes in and out of China to connect passengers. Due to this relationship, number of operated flights in this cluster starts decreasing from early February. It is not as fast as Cluster D (Mainland China) airlines showed, but the decreasing started earlier than other clusters.
F	Brussels Airlines, Delta Air Lines, SAS Scandinavian Airlines, Wizz Air	Cluster F is mainly clustered together due to the similar cyclic pattern for the normal operation that can be observed from January to February pattern. Other than the cyclic pattern, it is showing similar drop around 18th March as Cluster A, Cluster C.
G	Royal Air Maroc, Ryanair, easyJet	Cluster G shows big cyclic pattern throughout January and February, followed by a significant drop starting from 16th March. Ryanair and easyJet are leading low-cost carriers in Western Europe, which show the fact that number of operated flights is decreased by more than 95% in less than 2 weeks.
H	Qantas Airways, Virgin Australia	Cluster H consists of Australian airlines. Cluster H pattern is different from other cluster as the pattern shows cyclical increase and decrease even after the significant drop around 20th March. As these two airlines' cyclic pattern was mainly caused by domestic flight schedule rather than international flight schedule, the cyclic pattern has not been changed by reducing the international flights. Also, the dropping point of the operated flights is slightly later than other Clusters such as Cluster A and Cluster C.
I	Air New Zealand	Air New Zealand shows a similar pattern as Cluster H and Cluster C, but it has been separated out as the number of operated flights started to reduce around 20th March, which is slightly later than other Clusters that shows similar pattern during January and February such as Cluster A and Cluster C. It shows a high similarity with Qantas Airways pattern from the analysis, but Air New Zealand became a stand-alone cluster as it does not show cyclic pattern yet after the decrease in number of operated flights.
J	Citilink	Citilink is an airline in Indonesia mainly operating domestic and regional routes. Compared to Garuda Indonesia (Cluster C), the pattern shows a slight drop in operation in mid-February, then started to recover the similar level of operation in March. We can see that it does not follow any of patterns in data.

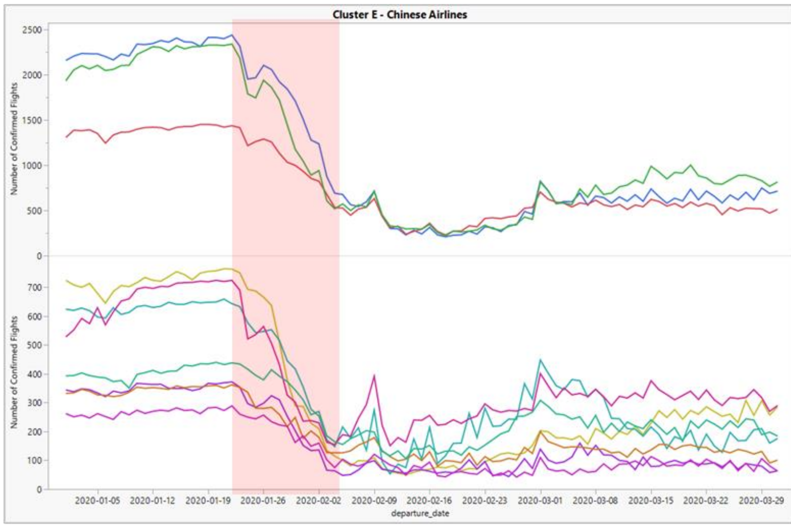
IMPACT OF SUPPLEMENTARY CHARACTERISTICS

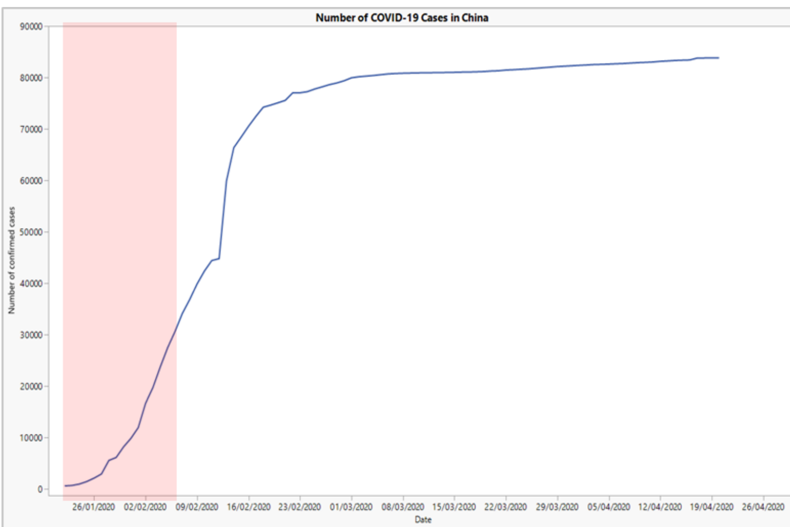
The identified clusters from the TS Similarity node provided an insight into the similarities between flight operations. However, this paper further attempts to understand associations between variations in flight operations and other supplementary characteristics of airlines.

Confirmed COVID-19 Cases:

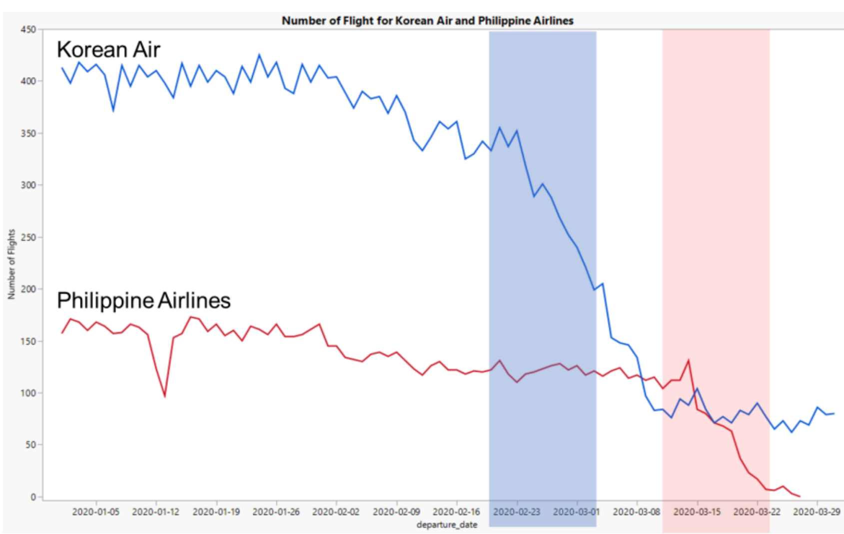
With an increase in COVID-19 Cases in each individual country, governments globally have been observed to restrict travel be it in or out of the country. Travelers were also observed to steer away from travel to such countries. An example is the restriction of flights with China and Korea with rising confirmed COVID-19 cases in those 2 countries.

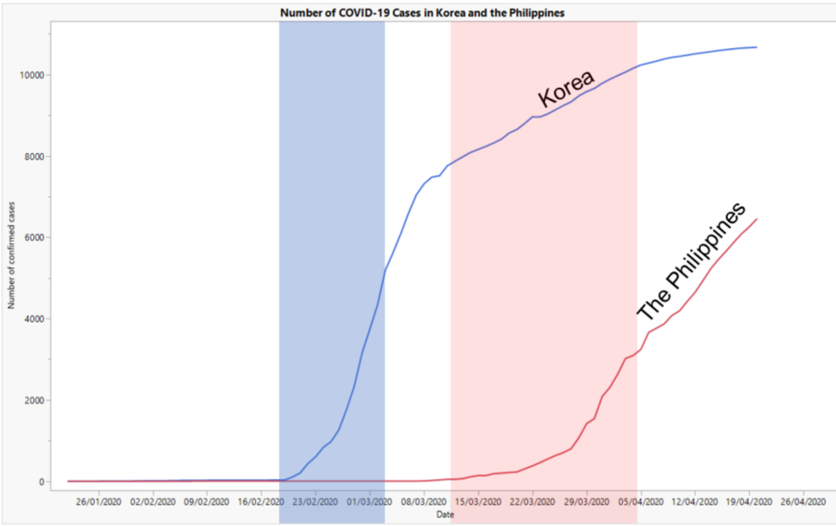
To identify any associations, a comparison of airlines within the same clusters was conducted:





As seen in the 2 diagrams above, the sudden decrease in flights by Chinese airlines (Cluster D) was observed to coincide with the increase in reported confirmed COVID-19 cases in late January to early February.





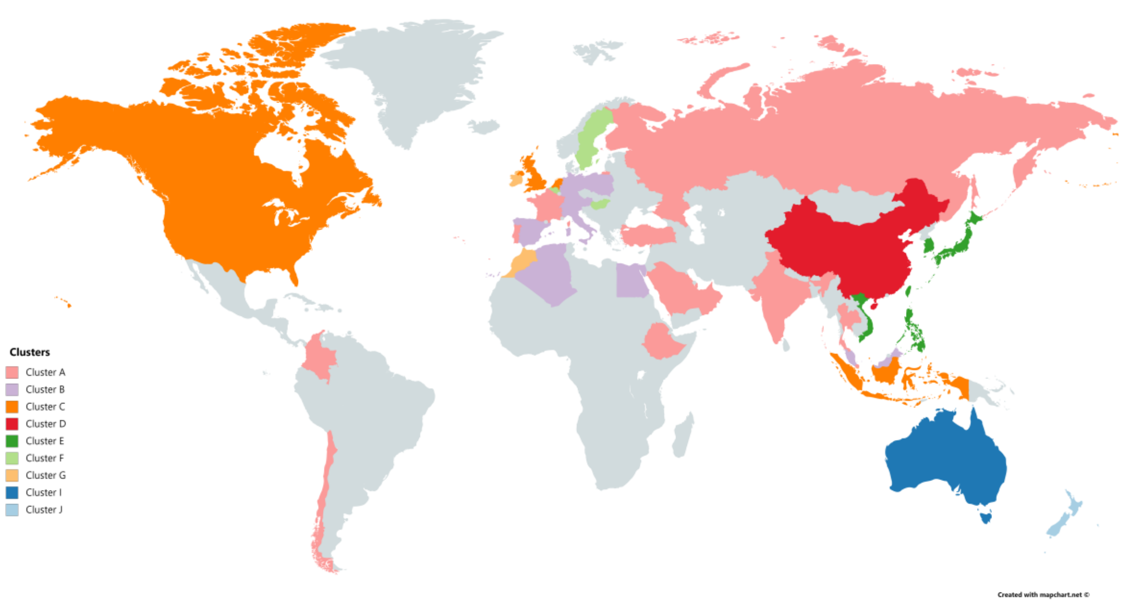
This was also observed between airlines from different countries in Cluster E. As observed in the 2 diagrams above, the number of flights for Korean Air started to decrease as the number of confirmed COVID-19 cases started to increase sometime in Mid-February. On the other hand, flights for Philippine Airlines also started to decrease at a later time in early March. This also demonstrates the ability of DTW in detecting similarities over time-periods. Such observations were also present for the other airlines within the same cluster.

These observations are in line with the assumption that there is an association between the number of flights and the rise of COVID-19 cases in the airline's headquartered country. In particular it demonstrates a potential association between the timing of rise in COVID-19 cases and fall in flight numbers.

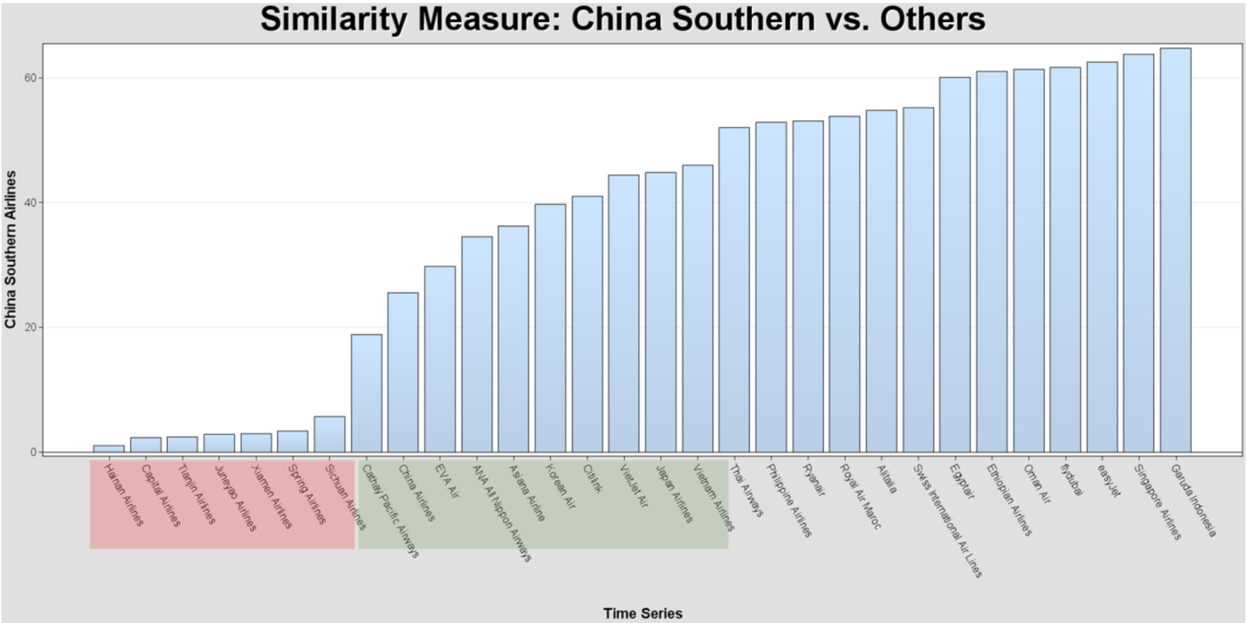
However, even with differences in the rate and number of reported confirmed COVID-19 cases between countries, airlines within the same cluster still demonstrate similarities in their flight patterns regardless of their headquartered country. It is therefore unlikely that number of confirmed COVID-19 cases has an association with the flight patterns.

Headquartered Country and Geographic Proximity:

Another perspective that the study focused on throughout the analysis is geographic proximity amongst airlines within the same cluster. The following colored map breaks down the identified clusters based on the geographical headquartered location of member airlines. Cluster D (mainland China), Cluster E (Near China) and Cluster I (Australia) demonstrates a congregation of member airlines within a specific region. On the other hand, Cluster A and Cluster B show that cluster is not distributed in specific region.



To look closer to the China and nearby countries, a comparison was done using the similarity measure plot generated by the TS Similarity node. Using China Southern from Cluster D as a benchmark, it was observed that airlines in mainland China show the highest similarity followed by Asian airlines from Cluster E.



Hence, it can be observed that there is a potential association between the geographical proximity and flight patterns of some airlines and clusters. However, this observation was not prevalent in all clusters and airlines.

Fleet Size & Star Rating:

Airline	Cluster	Fleet Size	Star Rating
Lufthansa	B	300	5 stars
Swiss International Air Lines		91	4 stars
Iberia		88	4 stars
Austrian Airlines		83	4 stars
AirAsia		255	3 stars
Alitalia		113	3 stars
LOT Polish Airlines		98	3 stars
Egyptair		66	3 stars
Air Algerie		55	3 stars
flydubai		54	3 stars
ANA All Nippon Airways	E	241	5 stars
Japan Airlines		175	5 stars
Cathay Pacific Airways		133	5 stars
Asiana Airline		85	5 stars
EVA Air		82	5 stars
Korean Air		180	4 stars
Vietnam Airlines		102	4 stars
Philippine Airlines		97	4 stars
China Airlines		88	4 stars
VietJet Air		75	3 stars

The table above compares the fleet size and star-rating of airlines within clusters B and E. Fleet size and star rating varies across airlines regardless of the cluster they belong to. Therefore, there are no observable associations between the 2 characteristics and the clustering results.

FINDINGS AND POTENTIAL DEVELOPMENTS

This paper employed dynamic time warping techniques to identify and group airlines into 10 clusters based on their flight patterns between January to March 2020, in light of the COVID-19 pandemic. Clusters were formed based on similarities in falling flight numbers between airlines. Additionally, the cyclical operating patterns of airlines were observed to have also influenced cluster formation.

It was observed that International Airlines globally have all been impacted drastically regardless of geographical location. Chinese Airlines were the earliest to be hit with flight operations falling since early-February. However, these airlines showed some recovery in late-March. Additionally, airlines from countries with a high proportion of flights in and out of China (i.e. Cluster E) started a gradual decline since Early-February. Majority of the other airlines demonstrated a sharp decrease in flights from around 12th March – the day the WHO declared the COVID-19 outbreak a pandemic.

Further analysis also observed a potential association between the timing of the fall in airline flights and rise of COVID-19 cases in their respective headquartered country. Geographical location demonstrated some associations to cluster formation for some airlines and clusters but was not observed in others. However, no associations were observed with other characteristics such as star ratings and fleet size on cluster formation

The paper also demonstrated that despite the fall in flight numbers globally, domestic flights still provide some respite for certain airlines as demonstrated by the Australian Airlines and Citilink.

The findings from this paper provides some insights into the similarities between flight patterns of different airlines during this pandemic. It could potentially shed some light into the financial and non-financial impacts to airlines within similar clusters. This information could prove useful for policy planning by governments and corporations, or by investors and financial institutions to name a few.

Potential developments to the paper could include observing other factors to provide more insight behind the driving force behind similarities within clusters. One example could be the analysis of flight networks and its association. Further studies can be done in the future with an extended time series to also understand recovery patterns once available. By also comparing it with previous global pandemics such as SARS, this paper can also allow for a greater understanding of the impacts of pandemics on the airline industry and potentially be used in the predicting of impact for future pandemics.

REFERENCES

- Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series, vol 16. KDD workshop, Seattle, pp 359–370
- Chen, S., Tseng, T., Ke, H., & Sun, C. (2011). Social trend tracking by time series based social tagging clustering. Expert Systems With Applications, 38(10), 12807–12817. https://doi.org/10.1016/j.eswa.2011.04.073
- Huang, S., & Lu, H. (2020). Classification of temporal data using dynamic time warping and compressed learning. *Biomedical Signal Processing and Control*, *57*. https://doi.org/10.1016/j.bspc.2019.101781
- IATA (2020). *COVID fourth impact assessment.* Retrieved from the International Air Transport Association (IATA) website: https://www.iata.org/en/iata-repository/publications/economic-reports/covid-fourth-impact-assessment/
- Jeong, Y., Jeong, M., & Omitaomu, O. (2011). Weighted dynamic time warping for time series classification. Pattern Recognition, 44(9), 2231–2240. https://doi.org/10.1016/j.patcog.2010.09.022
- VariFlight (2020). *COVID-19 Outbreak: Analysis of Global Aviation Operation.* Retrieved from the Airsavvi website: http://www.airsavvi.com/newsDetails/3_33.html?nidnex=1?AE71649A58c77
- Toni Giorgino. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. Journal of Statistical Software, 31(7), 1–24. https://doi.org/10.18637/jss.v031.i07

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Chung Wei Han, Shaun
E-mail: shaun.chung.2019@mitb.smu.edu.sg

Name: Son Yejin
E-mail: yejin.son.2019@mitb.smu.edu.sg