Knowledge Discovery from Commuter Flow Data with Cluster Analysis

Ranice TAN Hui Qi

Contents

1. Overview

Transportation analytics involves using data and statistical analytics to drive policy making. In Singapore, public transport is frequently used by residents to commute to work or school. By analysing data of the commuters travelling pattern, can give policy makers an insight on how to optimise service and plan routes to reduce congestion.

2. Objective

The objective of this study is to identify insights of MRT/LRT commuters by clustering analysis. In this analysis, origin-destination commuting interaction clusters are determined based on 3 peak traffic periods – weekday morning, weekday evening and weekend leisure. The clusters statistics are interpreted and characteristics of the clusters will be discussed.

3. Data

3.1. Data Used

The main dataset contains the hourly origin and destination of all MRT/LRT commuters throughout July 2021. It is the Passenger Volume by Train Stations in Singapore, which can be downloaded from the LTA Data Mall. A supplementary dataset used is the MRT/LRT codes and their corresponding names, which can be retrieved from data.gov.sg. The list of datasets can be found in Appendix A.

3.2. Data Preparation

The dataset was imported into JMP Pro to ensure all fields are filled, and all columns are appropriately formatted. A preliminary inspection of the data summary statistics and distribution was also conducted. The data was then filtered according to the 3 peak periods and transformed into origin-destination tables summarising the sum of total trips. All data tables were ensured to be filled and standardised before conducting cluster analysis. The data preparation log is accessible in Appendix B.

3.3. Data Quality

a. Renaming station codes to include names for easy reference

To identify stations names and their attributes, the passenger volume dataset was joined with the train station names data set to match the station codes and their names.

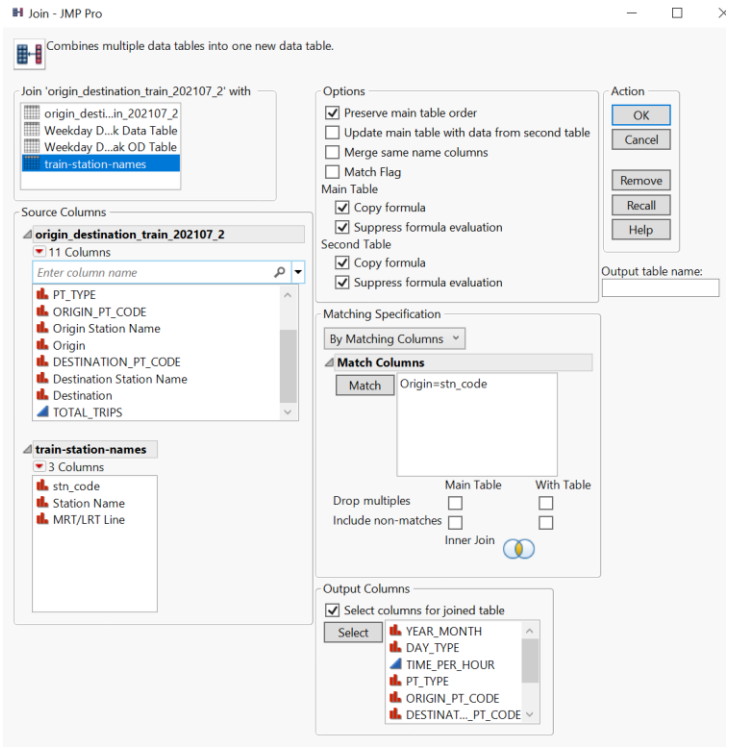

Figure 1 Join table function to include station names

b. Extracting relevant data

As the dataset contains data from across all timings and days in July 2021, it must be filtered for the timing in each peak period only. The conditions used are as follows:

Table 1 Row selection condition for each period

Peak Period	Conditions
Weekday Morning Peak 7am – 9am	DAY_TYPE equals WEEKDAY TIME_PER_HOUR is greater of equal to 7 TIME_PER_HOUR is less than 9
Weekday Evening Peak 5pm – 7pm	DAY_TYPE equals WEEKDAY TIME_PER_HOUR is greater of equal to 17 TIME_PER_HOUR is less than 19
Weekend Leisure Peak 11am – 1pm	DAY_TYPE contains WEEKEND TIME_PER_HOUR is greater of equal to 11 TIME_PER_HOUR is less than 13

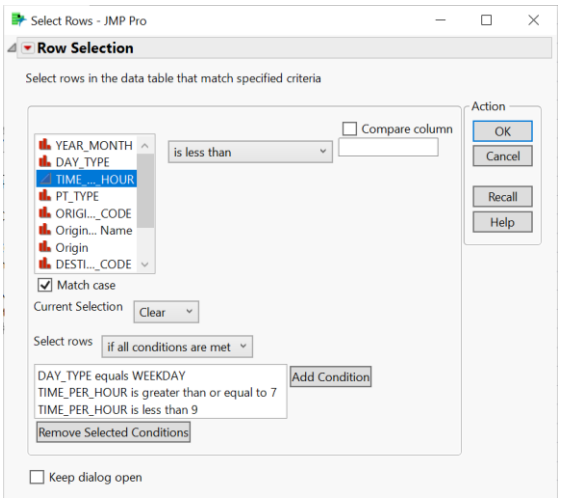

Figure 2 Row selection for weekday morning period

c. Transforming into origin-destination tables

After achieving the filtered data tables for each peak period, the data is further summarised into an origin-destination table in preparation for cluster analysis:

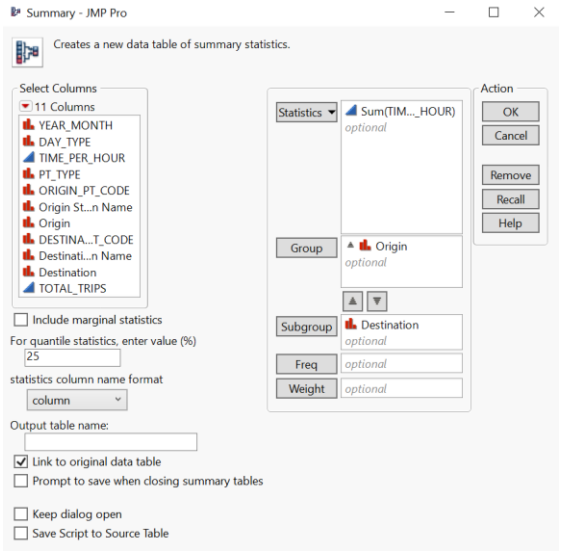

Figure 3 Transformation of data table into origin-destination table

d. Filling missing data

Missing values in the O-D table are replaced with numerical 0 before conducting clustering.

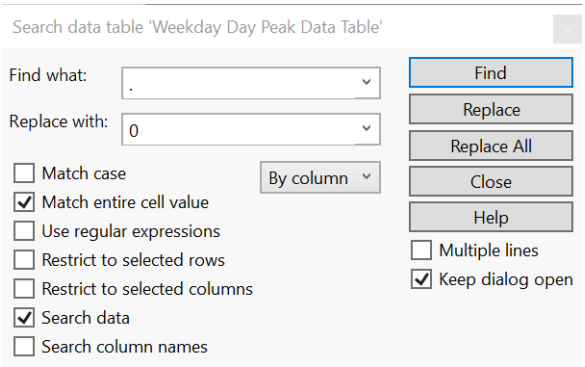

Figure 4 Filling in missing value in origin-destination table

e. Data standardisation

To achieve standardised data for analysis, new formula columns were inserted to distribute all data in the destination columns between the Range of 0-1.

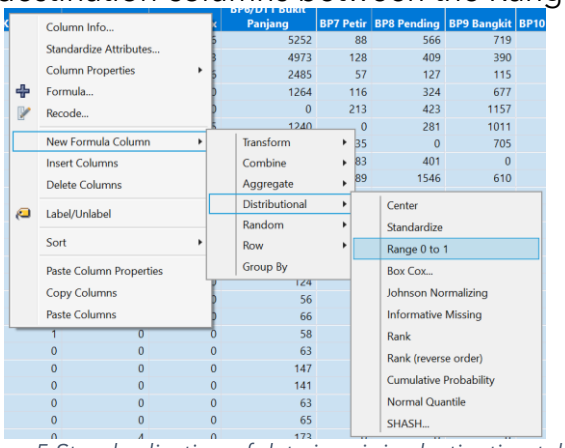

Figure 5 Standardisation of data in origin-destination table

f. Recode column names

Lastly, column names were also recoded to remove unnecessary string which may have been added through the summary table or new formula columns.

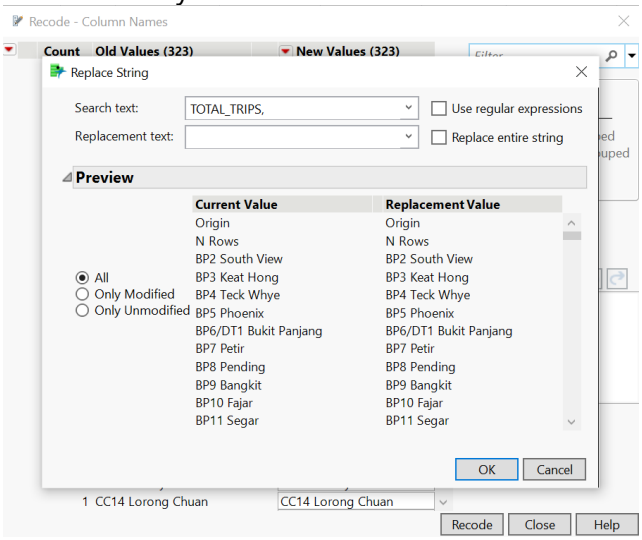

Figure 6 Mass recoding of column names by replacing string

4. Data Analysis

4.1. Weekday Day (7am-9am) Peak

a. Analysis Procedure

In this dataset, a 159 X 159 origin-destination matrix is explored. As the matrix is < 1000, and comprehensive description is required, Hierarchichal clustering method is used. The ward's method algorithm was used because some noise is expected between clusters.

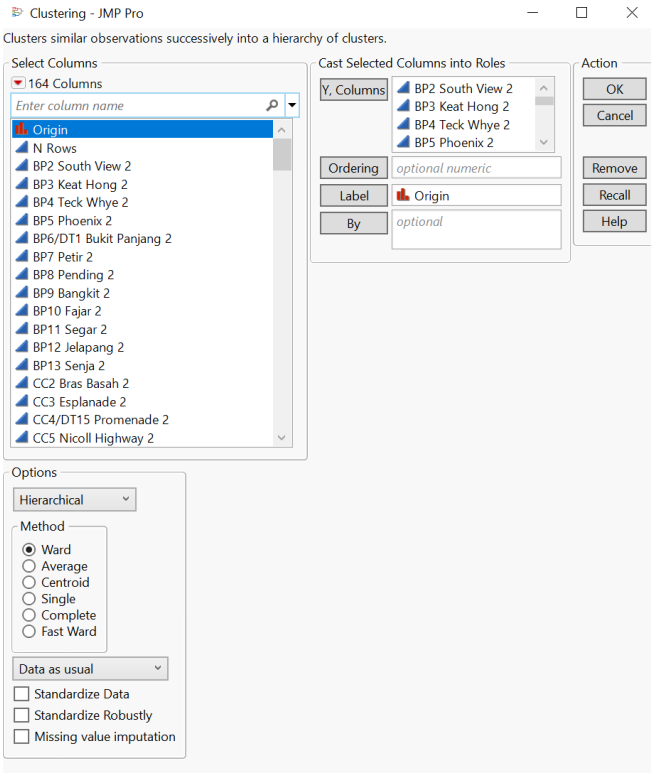

Figure 7 Hierarchichal ward's method clustering for weekday day O-D table

The cubic clustering criterion was generated based on the number of clusters, and the optimum 16 clusters was selected.

Cubic Clustering Criterion

	Number of Clusters	CCC
	1	0.000
	2	2.541
	3	3.348
	4	4.250
	5	5.358
	6	6.885
	7	7.921
	8	9.045
	9	10.097
	10	11.664
	11	12.351
	12	13.011
	13	14.299
	14	14.667
	15	15.086
->	16	15.512

Figure 8 CCC Table for weekday day clustering

The dendogram was coloured based on the numbers of clusters selected to further analysis the characteristics of the clusters.

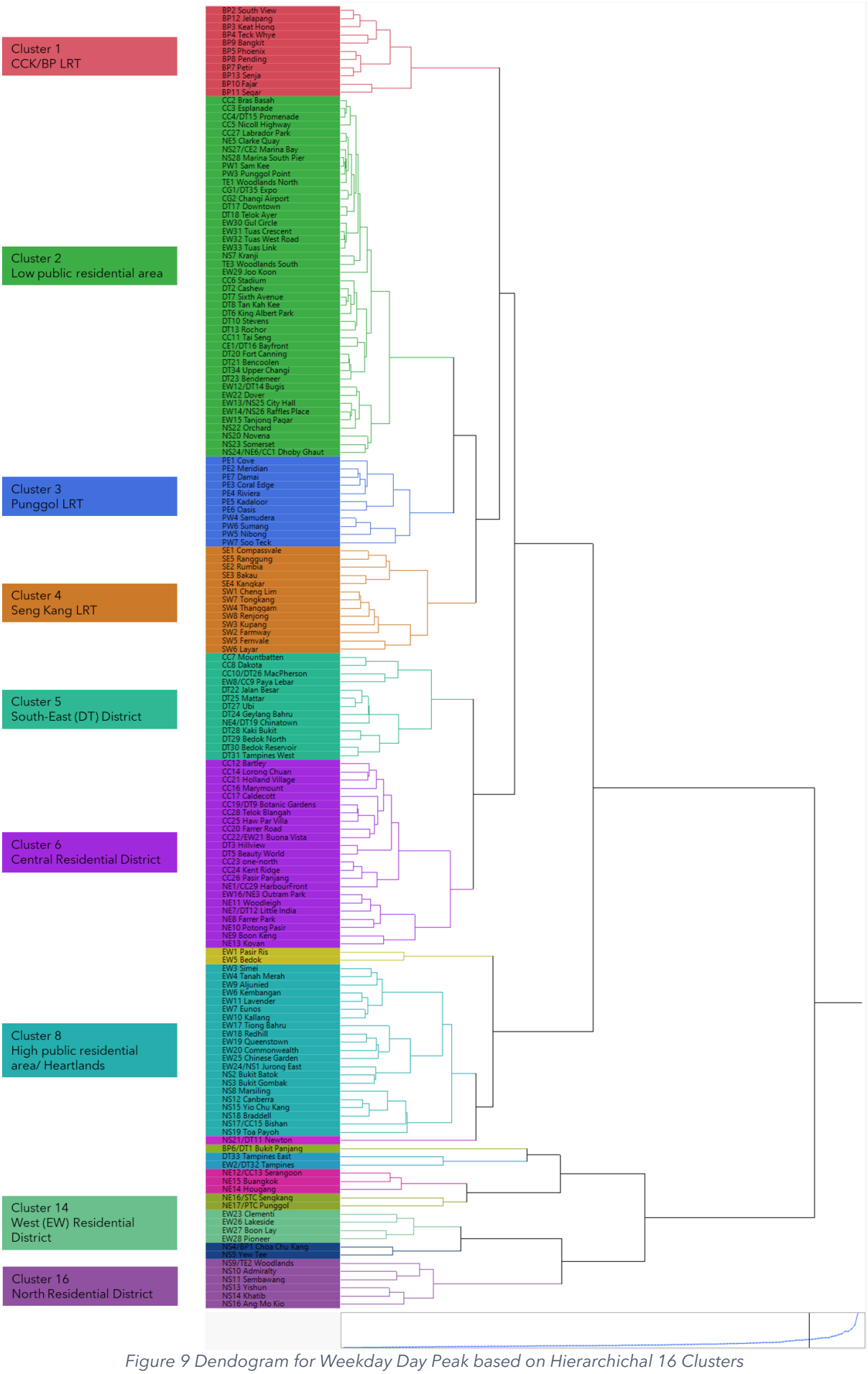

Figure 9 Dendogram for Weekday Day Peak based on Hierarchichal 16 Clusters

b. Insights

The dendogram above shows Weekday day commuting pattern of MRT passengers. By hierarchichal clustering, 16 clusters have been identified.

Cluster 1,3,4 showcase the correlation of the LRT feeder services in Bukit Panjang, Punggol and Seng Kang respectively. This may be because the commuters of these services are typically headed to the closest MRT station to transit to their workplace, which are more likely to located along the MRT lines.

Cluster 2 shows the correlation between the MRT stations dominantly with lower public residential districts as follows

- Central business district e.g. Brash Basah, Marina Bay, Downtown, Telok Ayer, City Hall, Raffles Place, Tanjong Pagar, Orchard, Dhoby Ghaut.
- Bukit Timah enclave – Cashew, Sixth Avenue, Tan Kah Kee, King Albert Park, Stevens
- West Industrial – Joo Koon, Gul Circle, Tuas Crescent, Tuas West Road, Tuas Link
- Northern Shore – Kranji, Woodlands South, Woodlands North
- Changi Region - Expo, Changi Airport, Upper Changi
- Punggol non-residential region: Sam Kee (Park, SAFRA), Punggol Point Jetty

The commonality of these areas is that low volumes of commuters is expected to originate from businesses/industrial areas during weekday mornings. Residents of the Bukit Timah may use other forms of transport, as not all residents stay close to the MRT stations along Bukit Timah Road. Personal preference and purchasing power may also come into play for Bukit Timah residents.

Cluster 5, 14, 16 depicts the correlation in the South-East, West and North heartlands respectively.

- South-East e.g. Tampines West, Bedok Reservoir, Mattar, MacPherson, Paya Lebar
- West – Clementi, Lakeside, Boon Lay and Pioneer
- North – Woodlands, Admiralty, Sembawang, Yishun, Khatib and Ang Mo Kio

These clusters show that commuters from these locations within their cluster may be headed to similar destinations, such as business/industrial parks along their respective MRT lines. The commuters from these heartlands, especially mature estates like those in the North or West are also expected to be higher.

Cluster 6 illustrates the relationship between the predominantly-central districts like Potong Pasir, Pasir Panjang, Buona Vista, Telok Blangah, Outram Park and Lorong Chuan. Most of these areas contain some mix of industrial and residential developments. The commuters from these areas are expected to be moderate but less than predominantly residential areas.

Cluster 8 shows correlation of districts with higher public residential density from all over Singapore e.g. Simei, Aljunied, Redhill, Queenstown, Chinese Garden, Bukit Batok, Yio Chu Kang, Bishan and Toa Payoh. The traffic from these locations are expected to be high and may be headed towards a more central location as they come from all-around Singapore.

4.2. Weekday Evening (5pm-7pm) Peak

a. Analysis Procedure

In this dataset, a 159 X 159 origin-destination matrix is explored. Hierarchichal ward's clustering method is used.

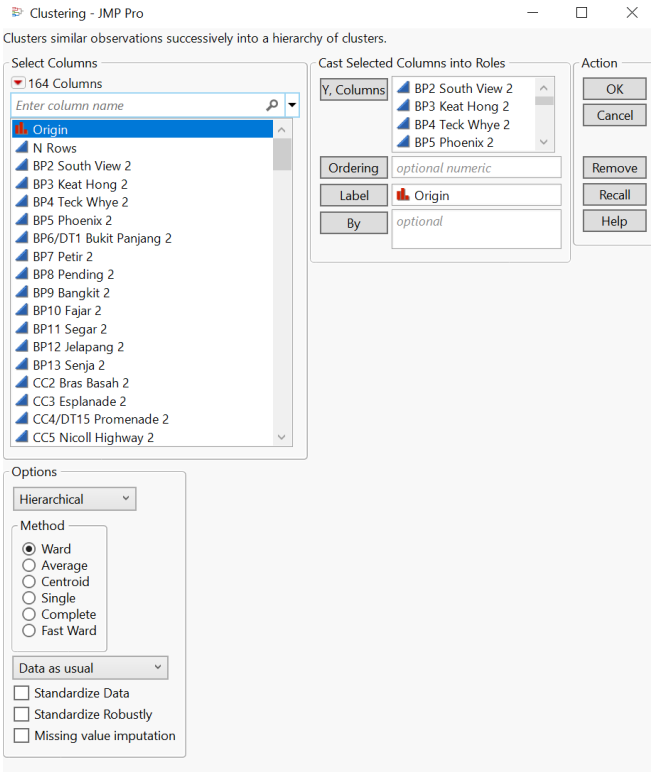

Figure 10 Hierarchichal ward's clustering method for weekday evening O-D table

The cubic clustering criterion was generated based on the number of clusters, and the optimum 16 clusters was selected.

Cubic Clustering Criterion

	Number of Clusters	CCC
	1	0.000
	2	2.176
	3	3.602
	4	3.469
	5	4.647
	6	6.601
	7	7.902
	8	9.363
	9	10.897
	10	13.074
	11	14.402
	12	15.504
	13	16.384
	14	17.419
	15	18.445
->	16	19.265

Figure 11 CCC Table for weekday evening peak clusters

The dendogram was coloured based on the numbers of clusters selected to further analysis the characteristics of the clusters.

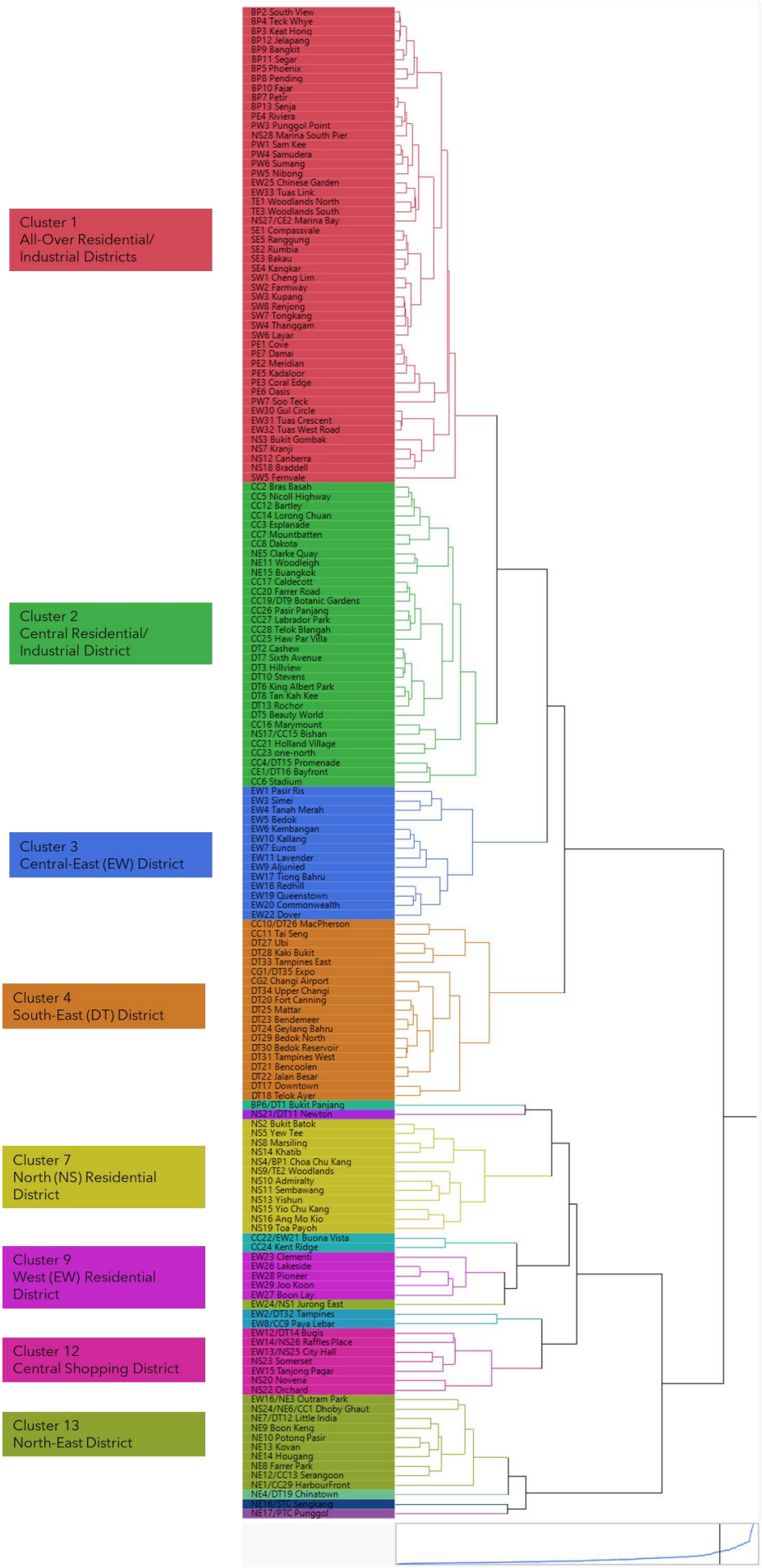

Figure 12 Dendogram for weekday evening peak based on Hierarchichal 16 clusters

b. Insights

The dendogram above shows weekday evening commuting pattern of MRT passengers. By hierarchichal clustering, 16 clusters have been identified.

Cluster 1 shows the correlation between the MRT stations dominantly with high residential/industrial districts as follows:

- Bukit Panjang LRT
- Punggol LRT
- Seng Kang LRT
- West Industrial – Gul Circle, Tuas Crescent, Tuas Link, Tuas West Road
- North-West – Chinese Garden, Bukit Gombak, Woodlands, Kranji, Canberra, Braddell
- Marina South – Marina Bay, Marina South Pier

The Bukit Panjang, Punggol, Seng Kang areas are dominantly residential, hence low volumes of commuters expected to board from these stations. The data may also show low commuters from Tuas and some areas in the North, as workers in manufacturing industry may be required for shift, and their knock-off time may not coincide from 5pm-7pm. The north-west area consists of mixture of residential and industrial which may be shift-based. The Marina South district may experience less commuters due to more conveniently located Downtown MRT stations for CBD commuters.

Cluster 2 showcases the relationship in the following central areas:

- Central residential area – Bartley, Lorong Chuan, Mountbatten, Dakota, Woodleigh, Buangkok, Caldecott, Farrer Road, Telok Blangah, Marymount, Bishan, Holland Village
- Bukit Timah enclave – Botanic Gardens, Cashew, Sixth Avenue, Tan Kah Kee, King Albert Park, Stevens, Beauty World
- Central business/industrial area – Bras Basah, Nicoll Highway, Esplanade, Clarke Quay, Pasir Panjang, Labrador Park, Haw Par Villa, one-north, Promenade, Bayfront Stadium

Some commuter traffic may be expected from these areas which contain mix of residential business and industrial sectors, though not as high as the CBD. It also shows that the congestion in MRT stations in the central business/industrial is correlated to those of the Bukit Timah and central residential areas, which should be manageable. These may be due to work from home conditions introduced during COVID-19 as well.

Clusters 3, 4, 7, 9, 13 depicts the correlation in the Central-East(EW), South-East(DT), North, West and North-East residential areas respectively.

- Central-East(EW) e.g. Pasir Ris, Bedok, Kallang, Tiong Bahru, Queenstown
- South-East(DT) e.g. Tampines East, Ubi, Expo, Changi Airport, Downtown, Telok Ayer
- North e.g. Bukit Batok, Woodlands, Yishun, Ang Mo Kio, Toa Payoh
- West – Clementi, Lakeside, Pioneer, Joo Koon, Boon Lay
- North-East e.g. Outram Park, Little India, Boon Keng, Serangoon, Hougang

These clusters show that commuters within their clusters are potentially from neighbouring industrial/business parks and headed home to similar destinations, probably along the same line. Traffic may be high, especially so when two-way traffic from industrial workers and residents heading home.

Cluster 12 illustrates the central shopping and business district of Singapore like Raffles City, City Hall, Orchard, Somerset. High volume of commuters are expected at these stations, as they may include people who knock-off from work, and those who went to town to run errands after work.

4.3. Weekend Leisure (11am-1pm) Peak

a. Analysis Procedure

In this dataset, a 159 X 159 destination-origin matrix is explored. Hierarchichal ward's clustering method is used.

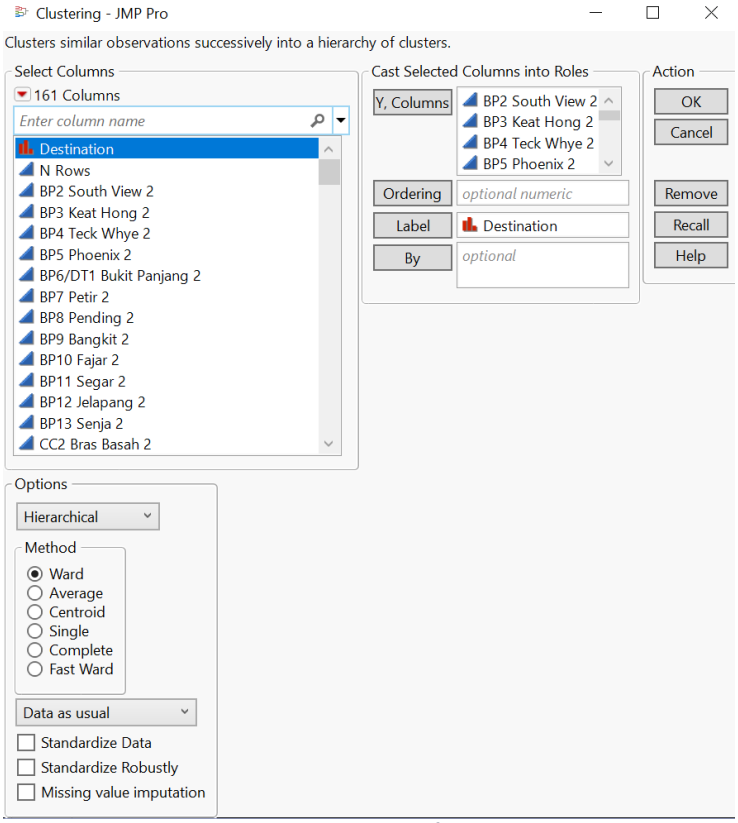

Figure 13 Hierarchichal ward's clustering for weekend leisure D-O table

The cubic clustering criterion was generated based on the number of clusters, and the optimum 16 clusters was selected.

Cubic Clustering Criterion

	Number of Clusters	CCC
	1	0.000
	2	2.541
	3	3.348
	4	4.250
	5	5.358
	6	6.885
	7	7.921
	8	9.045
	9	10.097
	10	11.664
	11	12.351
	12	13.011
	13	14.299
	14	14.667
	15	15.086
->	16	15.512

Figure 14 CCC Table for weekend leisure peak

The dendogram was coloured based on the numbers of clusters selected to further analysis the characteristics of the clusters.

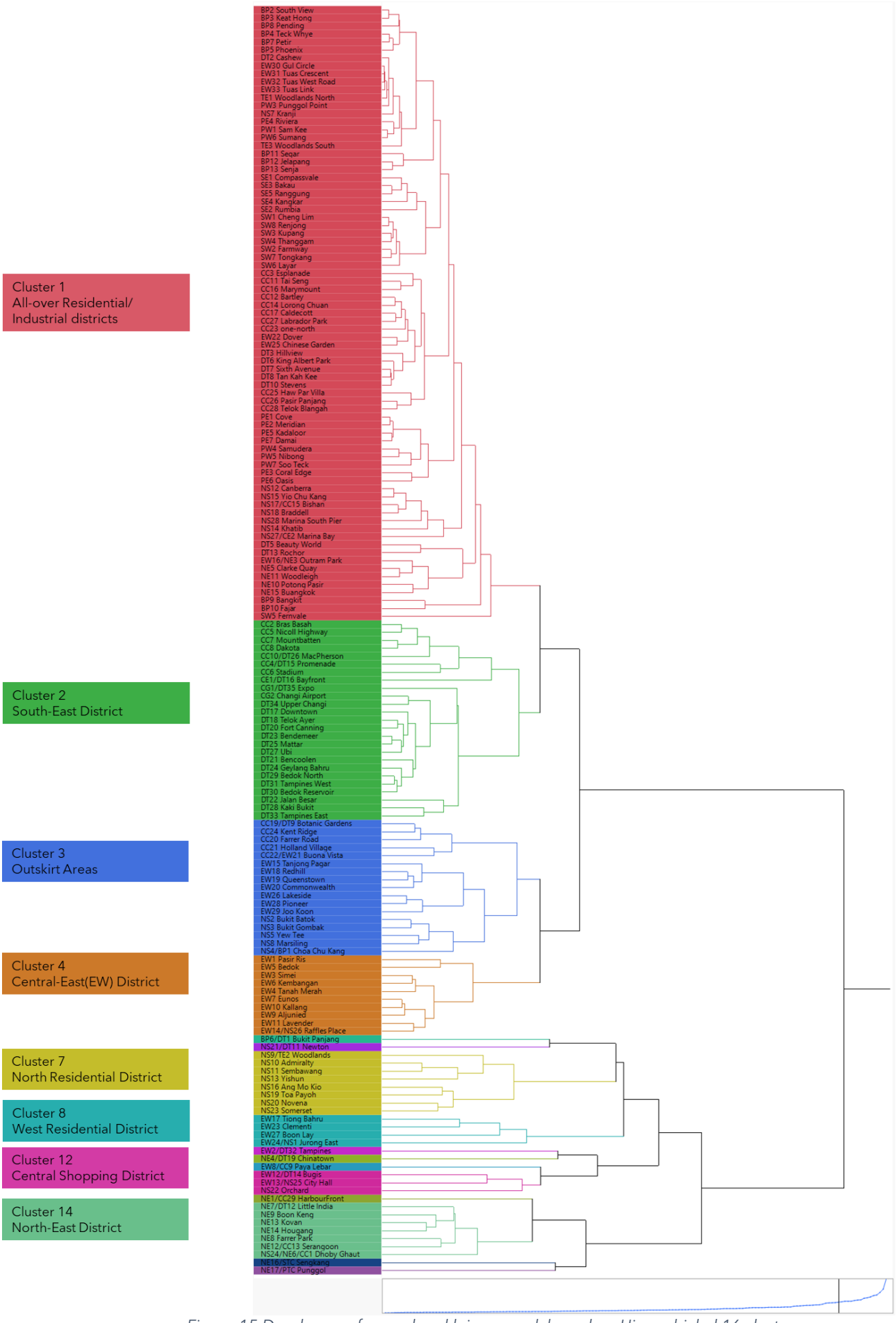

Figure 15 Dendogram for weekend leisure peak based on Hierarchichal 16 clusters

b. Insights

The dendogram above shows weekend leisure commuting pattern of MRT passengers. By hierarchichal clustering, 16 clusters have been identified.

Cluster 1 shows the correlation between the LRT/MRT stations as follows

- Bukit Panjang LRT
- Punggol LRT
- Seng Kang LRT
- Residential - Marymount, Lorong Chuan, Bartley, Caldecott, Hillview, King Albert Park, Sixth Avenue, Stevens, Telok Blangah, Outram Park, Woodleigh, Potong Pasir, Buangkok, Beauty World, Rochor, Khatib, Yio Chu Kang, Canberra, Chinese Garden
- Industrial/Business – Gul Circle, Tuas Crescent, Tuas Link, Tuas West Road, Labrador Park, Haw Par Villa, Pasir Panjang, Tai Seng, Kranji, Woodlands South, Woodlands North
- Marina South – Marina Bay, Marina South Pier, Clarke Quay

The commuters headed to the above destinations is expected to be lower, as they are mainly residential districts. Other than visiting family/friends, which has been reduced since COVID-19, not much activities may be expected. Furthermore, the industrial/business commuters also should be low during 11am-1pm as it is a weekend and people on shift probably do not coincide with this timing. Lastly, the marina south sector may experience lower commuters as it is very hot in the afternoon and people may seek more conveniently located MRT stations like Bayfront to visit the Barrage area.

Cluster 2 illustrates the relationship between the MRT stations in South-East cluster below:

- Central – Bras Basah, Nicoll Highway, Dakota, Downtown, Telok Ayer, Fort Canning, Bencoolen, Jalan Besar
- Marina South – Stadium, Bayfront, Downtown, Promenade
- Changi – Expo, Changi Airport, Upper Changi
- South-East – Bendemeer, Mattar, Ubi, Geylang Bahru, Bedok North, Tampines West, Bedok Reservoir, Kaku Bukit, Tampines West

Notable places of interests from the MRT stations above include central museums/heritage area, Marina Bay, Changi Jewel and Bedok Reservoir Park which may attract visitors. The crowd to these locations are expected to be mild to moderate, particularly with some apprehension to Changi Jewel reopening in mid-Jun post its COVID-19 outbreak.

Cluster 3 showcases the correlation between MRT stations from several outskirt areas with unique activities to do as follows:

Central - Botanic Gardens, Farrer Road, Holland Village, Buona Vista, Kent Ridge, Tanjong Pagar

South – Redhill, Queenstown, Commonwealth

West – Lakeside, Pioneer, Joo Koon, Bukit Batok, Bukit Gombak, Yew Tee, Marsiling, Choa Chu Kang

The central area contains large parks like Botanic Gardens and heritage Tanjong Pagar for locals seeking for a touristy adventure. The South area contains many hawker centres (ABC Brickworks, Mei Ling Food Centre, Redhill Food Centre) which may be popular amongst the locals living along the South. The west area contains of some nature reserves such as Jurong Lake Park, Little Guilin and Dairy Farm which may be popular amongst people seeking an active lifestyle. Hence, these destinations may be more

popular with locals seeking unique things to do during the COVID-19 travel restrictions. The congestion in these areas are expected to be moderate to high as they are spread out across many regions.

Clusters 4, 7, 8, 14 depicts the correlation in the Central-East(EW), North, West and North-East residential areas respectively.

- Central-East(EW) – Pasir Ris, Bedok, Simei, Kembangan, Tanah Merah, Eunos, Kallang, Aljunied, Lavender, Raffles Place
- North – Woodlands, Admiralty, Sembawang, Yishun, Ang Mo Kio, Toa Payoh, Novena, Somerset
- West – Tiong Bahru, Clementi, Boon Lay, Jurong East
- North-East – Little India, Boon Keng, Kovan, Hougang, Farrer Park, Serangoon, Dhoby Ghaut

These correlations show that there is some demand to travel to regional shopping districts like Bedok, Woodlands, Toa Payoh, Jurong East, Serangoon during the weekends. The commuters likely originate from the same region, along the same MRT lines. This may also be driven by COVID-19, where more people avoid crowded town areas as compared to pre-COVID period.

Cluster 12 illustrates the central shopping and business district of Singapore like City Hall, Orchard, Bugis. High volume of commuters are expected at these stations, as people tend to head to town for personal errands or social activities.

5. Interpretation of Analysis Results

5.1. Potentially high levels of commuters from mature residential estates during the weekday morning peak

The commuters from heartlands like Ang Mo Kio, Woodlands, Clementi is expected to be high from Cluster 14 and Cluster 16 in the weekday morning peak analysis. These commuters are potentially travelling to neighbouring industrial estates or major business parks. Furthermore, these train stations are the older configuration which may be less spacious than the newer stations. Hence, may recommend stakeholders to pay attention to stations in these clusters and apply the appropriate crowd control measures to alleviate potential congestion during the morning peak.

5.2. Introduction of Downtown Line may help to ease some commuter from East-West line during weekday morning peak

Majority of the stations in Cluster 5 South-East district of weekday morning peak is from the Downtown Line. Prior to which, commuters from locations like Bedok North, Tampines West may have to board from Bedok or Tampines on the East-West line to reach their destination. This can potentially debottleneck congestion in the Eastern stations of the East-West line, allowing passengers to enjoy a more pleasant journey.

5.3. Central shopping/business district potentially congested during weekday evening peak

The central shopping/business district includes stations Bugis, Raffles Place, City Hall, Somerset, Tanjong Pagar, Novena and Orchard (Cluster 12 of weekday evening). Other than offices in these areas for businesses, it also serves as a central destination for people to run errands or socialise with their friends after work. With the two-way traffic of office workers heading home, and other people heading to town for their personal needs, it can potentially become very congested during the weekday evening peak period. Hence, will suggest stakeholders to pay attention and apply the appropriate crowd control measures during weekday evening.

5.4. North-East, West and North high density residential locations may see high crowds during evening peak

Mix of residential and industrial areas like Hougang, Boon Lay, Yishun in Clusters 13, 9 and 7 of weekday evening dendogram respectively may potentially see high levels of commuters during the evening peak period, as many workers are heading back home. Therefore, it is advisable to also work out some decongestion measures in these locations to ease crowds.

5.5. Central shopping district stations expected to be congested during weekend leisure

The central shopping district in Bugis, City Hall, Somerset and Orchard of Cluster 12 of weekend leisure dendogram. This is expected as many locals head to town for

shopping, social or leisure activities during the weekend, hence, crowds here are expected to be on the high side.

5.6. Potential visitorship to unconventional activities in Singapore during weekends with COVID-19 restrictions

An interesting cluster in the weekend leisure dendogram is Cluster 3, which depicts several areas all over the country. A commonality of these locations is that they contain unique activities for locals to embark on. For example, enjoying a walk at the Botanic Gardens, visiting heritage areas like Tanjong Pagar, enjoying lunch at ABC Brickworks, or even heading to nature reserves like Dairy Farm, Little Guilin in the West for a hike. These may be due to COVID-19 travel restrictions, which encourage locals to seek more unique things to do within the country, other than conventional leisure activities like shopping.

6. Recommendation

6.1. Introduction crowd control measures to potential high congestion areas below:

Weekday Day Peak:	• Cluster 14 Clementi, Lakeside, Boon Lay, Pioneer • Cluster 16 Woodlands, Admiralty, Sembawang, Yishun, Khatib, Ang Mo Kio
Weekday Evening Peak:	• Cluster 7 Bukit Batok, Yew Tee, Marsiling, Khatib, Choa Chu Kang, Woodlands, Admiralty, Sembawang, Yishun, Yio Chu Kang, Ang Mo Kio, Toa Payoh • Cluster 9 Clementi, Lakeside, Pioneer, Joo Koon, Boon Lay • Cluster 12 Bugis, Raffles Place, City Hall, Somerset, Tanjong Pagar, Novena, Orchard • Cluster 13 Little India, Boon Keng, Kovan, Hougang, Farrer Park, Serangoon, Dhoby Ghaut
Weekend Leisure:	• Cluster 12 Bugis, City Hall, Orchard • Cluster 14 Little India, Boon Keng, Kovan, Hougang, Farrer Park, Serangoon, Dhoby Ghaut

6.2. Train service can consider advertisement placement and pricings based on potential crowd and its exposure rate

6.3. Continue to monitor the effect of Downtown Line and ease of congestion of the East stations on the East-West line

6.4. Anticipate potentially more commuters to unique and interesting locations during weekend leisure peak

7. Appendix

7.1. Appendix A: Datasets

Item	Dataset Name	Brief Description	Usage
1	Origin_destination_train_202107	Consists of all MRT/LRT ridership data from origin to destination in July 2021	Used for retrieving selected data
2	Train-station-names	Consists of all MRT/LRT code and their corresponding names	Used to combine with station code in original dataset to get better understanding of stations
3	Weekday Day Peak Data Table	Filtered from origin_destination_train_202107 dataset	Used to generate OD table
4	Weekday Day Peak OD Table	Origin-Destination table of July 2021 MRT/LRT ridership	Used for clustering analysis
5	Weekday Evening Peak Data Table	Filtered from origin_destination_train_202107 dataset	Used to generate OD table
6	Weekday Evening Peak OD Table	Origin-Destination table of July 2021 MRT/LRT ridership	Used for clustering analysis
7	Weekend Leisure Peak Data Table	Filtered from origin_destination_train_202107 dataset	Used to generate DO table
8	Weekend Leisure Peak DO Table	Destination-Origin table of July 2021 MRT/LRT ridership	Used for clustering analysis

7.2. Appendix B: Data Preparation Change Log

a. Dataset: Origin_destination_train_202107

Item	Variable Name	Issue	Action
1	ORIGIN_PT_CODE	Unable to identify station names through code	Join with train-station-names table to generate origin station name column. Then, concatenate origin_pt_code and origin station name column.
2	DESTINATION_PT_CODE	Unable to identify station names through code	Join with train-station-names table to generate destination station name column. Then, concatenate destination station name and destination_pt_code columns.
3	TIME_PER_HOUR	Contains all timings throughout the day. Unable to identify weekday day peak data easily.	Select rows and generate table that correspond to weekday morning peak timing
4	TIME_PER_HOUR	Contains all timings throughout the day. Unable to identify weekday evening peak data easily.	Select rows and generate table that correspond to weekday evening peak timing
5	TIME_PER_HOUR	Contains all timings throughout the day. Unable to identify weekend leisure peak data easily.	Select rows and generate table that correspond to weekend leisure peak timing

b. Dataset: Weekday Day Peak Data Table

Item	Variable Name	Issue	Action
1	TIME_PER_HOUR	Unable to conduct clustering due to formatting of table	Summarise sum of total_trips, grouped by origin and subgrouped by destination to generate OD table

c. Dataset: Weekday Day Peak OD Table

Item	Variable Name	Issue	Action
1	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Some cells contain missing data	Search for missing data and replace '.' with '0'
2	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Column names contains additional string.	Remove extra string by recoding column names replacing string 'TOTAL TRIPS, ' with ''
3	BP2 South View, BP3 Keat Hong etc..	Data is not standardised	Standardise data by creating new formula column and selecting range between 0 – 1.
4	BP2 South View, BP3 Keat Hong etc..	Too many columns for analysis	Hide and exclude non-standardised columns not used for clustering analysis
5	Range Scale (BP2 South View), Range Scale (BP3 Keat Hong), Range Scale (etc.	Column names contains additional string.	Remove extra string by recoding column names replacing string 'Range Scale (' and ')' with ''

d. Dataset: Weekday Evening Peak Data Table

Item	Variable Name	Issue	Action
1	TIME_PER_HOUR	Unable to conduct clustering due to formatting of table	Summarise sum of total_trips, grouped by origin and subgrouped by destination to generate OD table

e. Dataset: Weekday Evening Peak OD Table

Item	Variable Name	Issue	Action
1	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Some cells contain missing data	Search for missing data and replace '.' with '0'
2	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Column names contains additional string.	Remove extra string by recoding column names replacing string 'TOTAL TRIPS, ' with ''

3	BP2 South View, BP3 Keat Hong etc..	Data is not standardised	Standardise data by creating new formula column and selecting range between 0 – 1.
4	BP2 South View, BP3 Keat Hong etc..	Too many columns for analysis	Hide and exclude non-standardised columns not used for clustering analysis
5	Range Scale (BP2 South View), Range Scale (BP3 Keat Hong), Range Scale (etc.	Column names contains additional string.	Remove extra string by recoding column names replacing string 'Range Scale (' and ')' with ''

f. Dataset: Weekend Leisure Peak Data Table

Item	Variable Name	Issue	Action
1	TIME_PER_HOUR	Unable to conduct clustering due to formatting of table	Summarise sum of total_trips, grouped by destination and subgrouped by origin to generate DO table

g. Dataset: Weekend Leisure Peak OD Table

Item	Variable Name	Issue	Action
1	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Some cells contain missing data	Search for missing data and replace '.' with '0'
2	TOTAL TRIPS, BP2 South View, TOTAL TRIPS, BP3 Keat Hong, TOTAL TRIPS, etc..	Column names contains additional string.	Remove extra string by recoding column names replacing string 'TOTAL TRIPS, ' with ''
3	BP2 South View, BP3 Keat Hong etc..	Data is not standardised	Standardise data by creating new formula column and selecting range between 0 – 1.
4	BP2 South View, BP3 Keat Hong etc..	Too many columns for analysis	Hide and exclude non-standardised columns not used for clustering analysis
5	Range Scale (BP2 South View), Range Scale (BP3 Keat Hong), Range Scale (etc.	Column names contains additional string.	Remove extra string by recoding column names replacing string 'Range Scale (' and ')' with ''