



ISSS608 – Visual Analytics and Visualization Group Project

Course Year: AY2022-23, January Term
Group members: Aishwarya Maloo, Huo Da, Prachi Ashani

User Guide: Interactive SG Bus Commuting Pattern Exploration & Analysis



Table of Contents

1. Introduction Page	3
2. EDA (Exploratory Data Analysis)	3
2.1 Distribution of Commuters.....	3
2.2 Density of Commuters	4
3. ESDA (Exploratory Spatial Data Analysis)	5
3.1 Number of Commuters	5
3.2 Commuter Flow	6
4. Clustering Analysis.....	7
4.1 Hierarchical Clustering	7
4.2 Geospatial Constrained – SKATER Clustering	8
5. Regression Analysis	9
6. Data	10

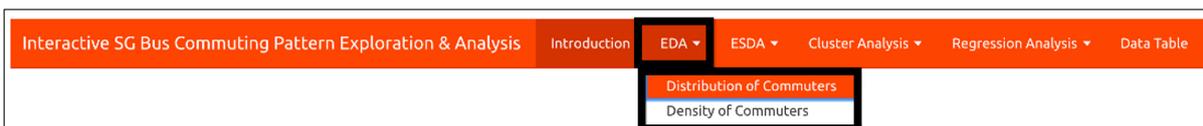
1. Introduction Page

As soon as the application launches, the user can view the 'Introduction' page. The page provides information on the project topic, authors, and their advisor. The navigation bar has tabs for EDA (Exploratory Data Analysis), ESDA (Exploratory Spatial Data Analysis), Cluster Analysis, Regression Analysis as well as the data tables for all the visualizations covered in this application. The user is required to navigate to downward arrow for (EDA, ESDA, Cluster Analysis and Regression Analysis) or merely click (Data Table) on the panel to explore them.



2. EDA (Exploratory Data Analysis)

The user is required to click on the drop-down menu to access the visualizations within 'EDA'. It has two tabs within – 'Distribution of Commuters' and 'Density of Commuters'

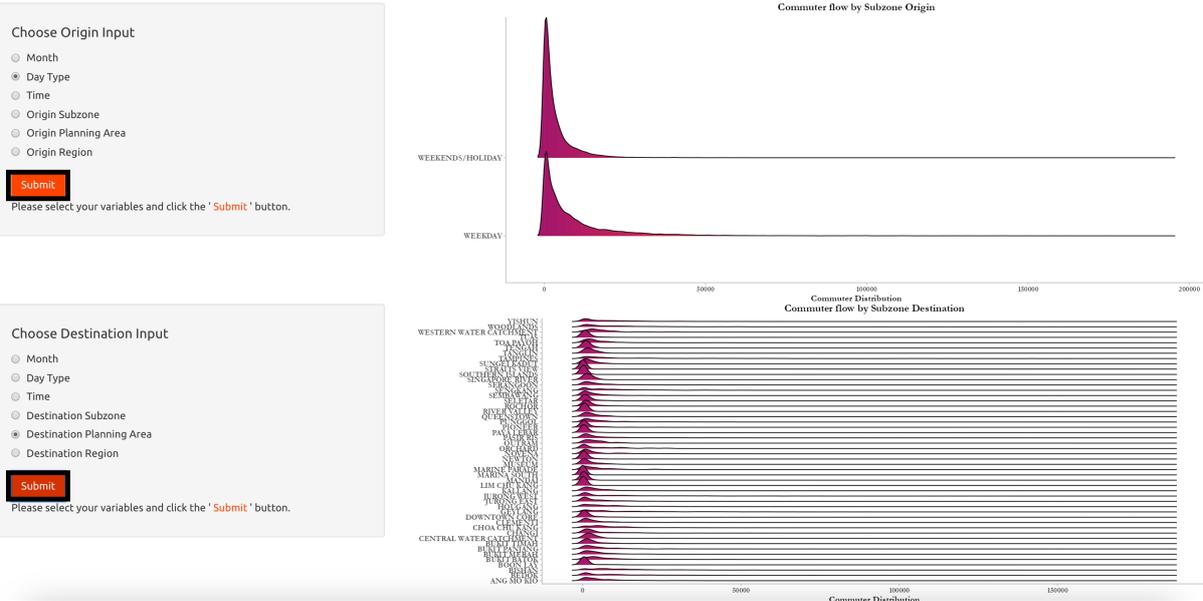


2.1 Distribution of Commuters

The user can simultaneously observe the commuter flow distribution by origin and destination.

1. The default page is set to no image.
2. The user must select the variable of their choice for origin/destination/both:
 - a. Month – the months present in the dataset are for October, November, December.
 - b. Day Type – it is bifurcated between weekdays and weekends & public holidays.
 - c. Time – ranges from 0 to 23:00 hours.
 - d. Origin/Destination Subzone – the subzones of Singapore.
 - e. Origin/Destination Planning Area – the planning areas of Singapore.
 - f. Origin/Destination Region – the broad regions of Singapore.
3. After selecting the parameter of their choice, the user must click the submit button to view the plot.

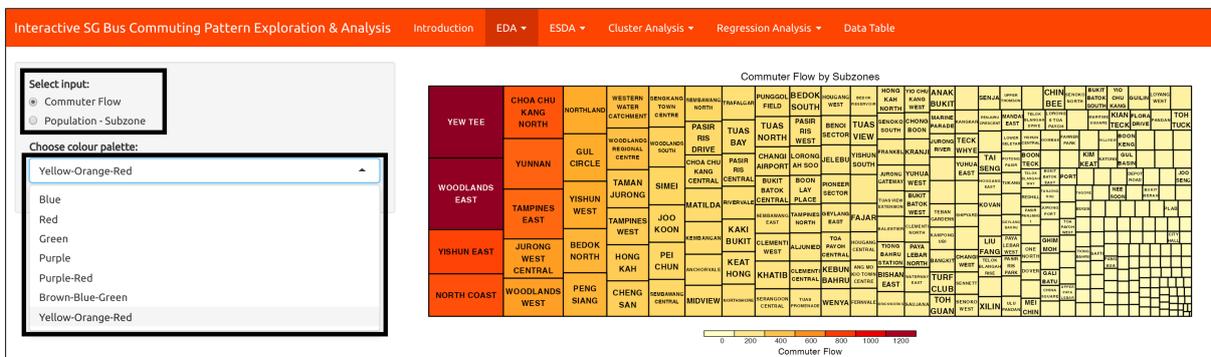
Commuter Flow Distribution by Origin and Destination, Singapore



2.2 Density of Commuters

Here, the users can see the density of commuters or the population density by subzones.

1. The default is set to commuter volume by subzone with the 'Yellow-Orange-Red' colour palette.
2. The user can select 'Commuter Flow' or 'Population - Subzone' from the 'Select Input' to see changes.
3. The user has the option of seven colour palettes to choose from:
 - a. Blue
 - b. Red
 - c. Green
 - d. Purple
 - e. Purple-Red
 - f. Brown-Blue-Green
 - g. Yellow-Orange-Red



3. ESDA (Exploratory Spatial Data Analysis)

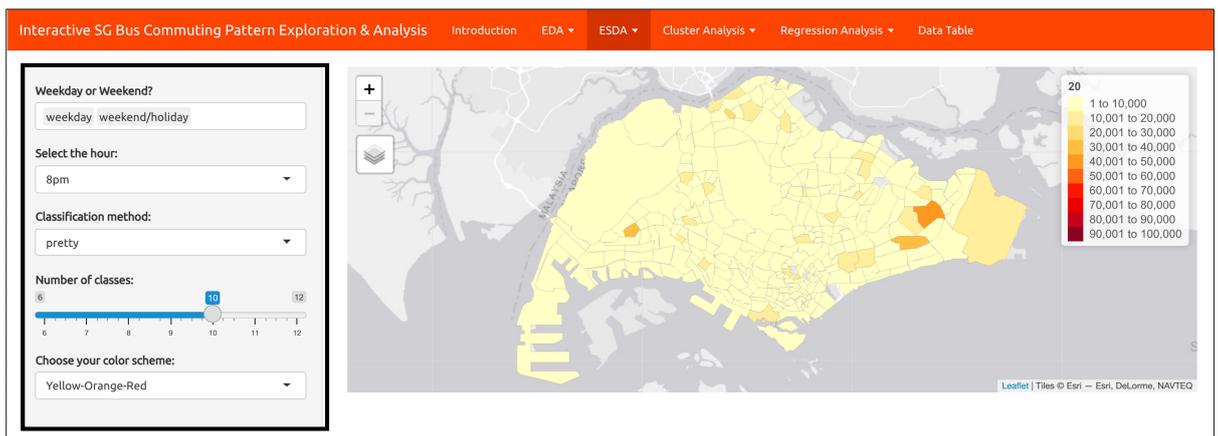
The user is required to click on the drop-down menu to access the visualizations within 'ESDA'. It has two tabs within – 'Number of Commuters' and 'Commuter Flow'



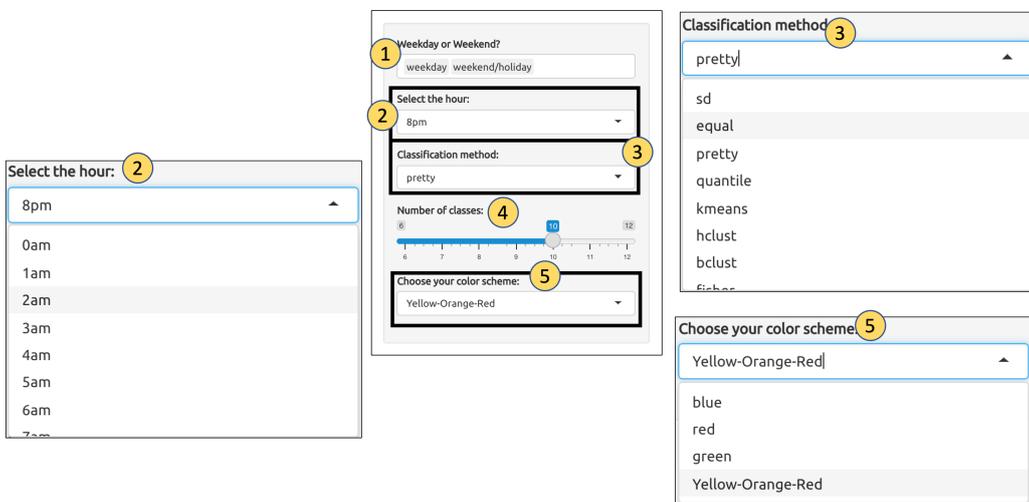
3.1 Number of Commuters

This tab shows a choropleth map of Singapore with inputs as provided by the user. A choropleth map is a thematic map that represents statistical data using colour mapping symbology technique¹.

1. The user is given 5 options for modification of the default map. The default is set to the selections seen below in the screenshot.



2. The choropleth tab below provides a sample of the dropdown fields:



¹ [Choropleth Map](#)

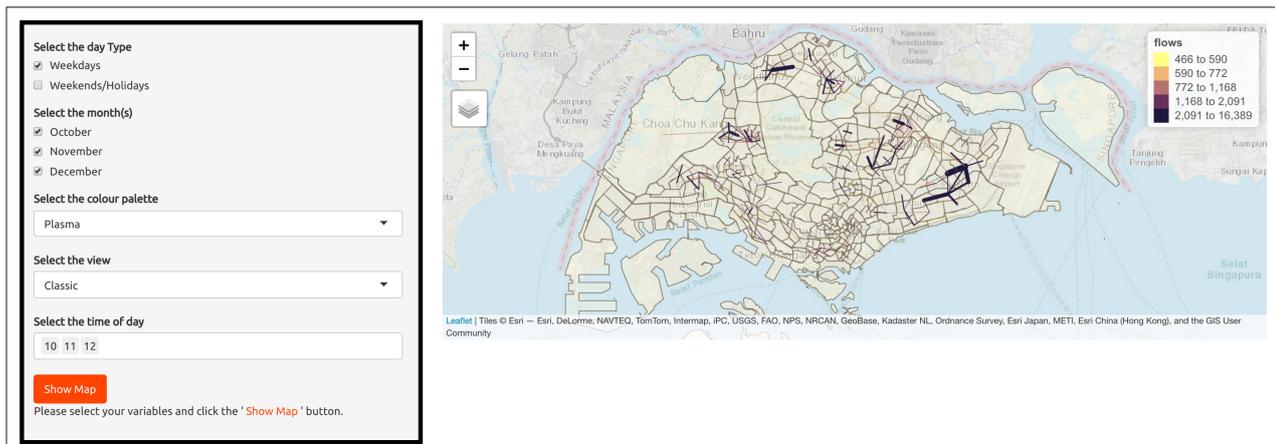
The user gets to select from:

- Weekday or Weekend? – the user gets to choose either ‘weekday’, ‘weekend’ or both options.
- Select the hour – from the dropdown the user can select the hours from 1 am to 12 pm.
- Classification method – the user gets an option to select any one of the nine classification methods (‘sd’, ‘equal’, ‘pretty’, ‘quantile’, ‘kmeans’, ‘hclust’, ‘bclust’, ‘fisher’, and ‘jenks’) provided in the drop down list.
- Number of Classes – the user gets the option to select the number of classes for clustering, the minimum value is six and the maximum is twelve.
- Choose your color scheme – the user gets the option to select from four colour palettes (‘blue’, ‘red’, ‘green’ and ‘Yellow-Orange-Red’).

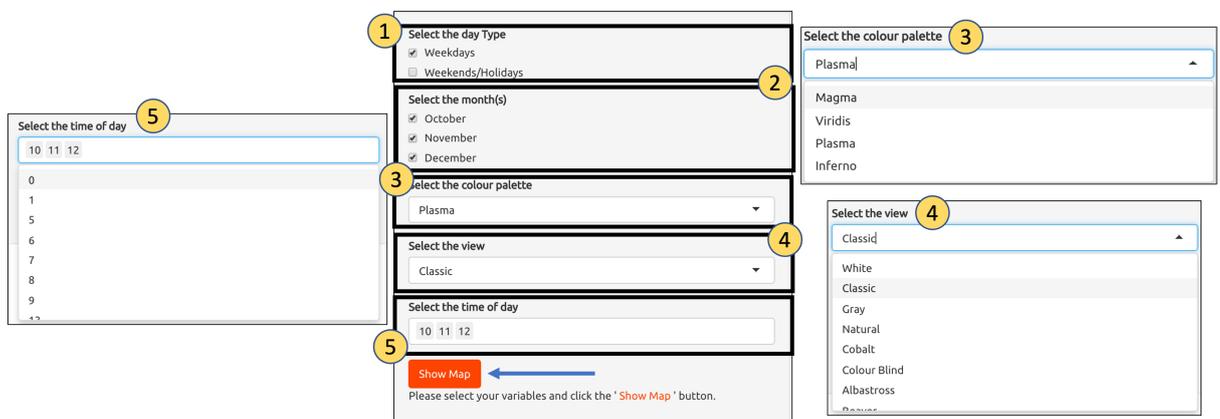
3.2 Commuter Flow

The user can toggle around with the options to see the desire lines or the commuter flow between the origin and the destination subzone.

- The user is given five options for modification of the default map. The default is set to the selections seen below in the screenshot. Note that, unless the submit button has been clicked, the map will not appear.



- The commuter flow tab below provides a sample of the dropdown fields:



The user gets to select from:

- a. Select the day type – the user gets to choose either ‘Weekdays’, ‘Weekends/Holidays’ or both.
 - b. Select the month(s) – the user gets the option to zoom into any of the given months (‘October’, ‘November’, and ‘December’)
 - c. Select the colour palettes – the user gets to change the colour palette of the flow lines from one of the four options available (‘Plasma’, ‘Magma’, ‘Viridis’, and ‘Inferno’).
 - d. Select the view – the user has the option to change the view of the map as per their liking from 10 options (‘White’, ‘Classic’, ‘Gray’, ‘Natural’, ‘Cobalt’, ‘Colour Blind’, ‘Albastross’, ‘Beaver’, ‘Black-White’, and ‘Watercolour’)
 - e. Select the time of the day – the user has the option to select the time of the day from 5 am to 1 am. This has multiple selection, so the user can group timings to observe the morning or evening peak as they choose to define the peak hours.
3. After making the selections, the user must click on the ‘Show Map’ button and be patient till the map appears on the screen.

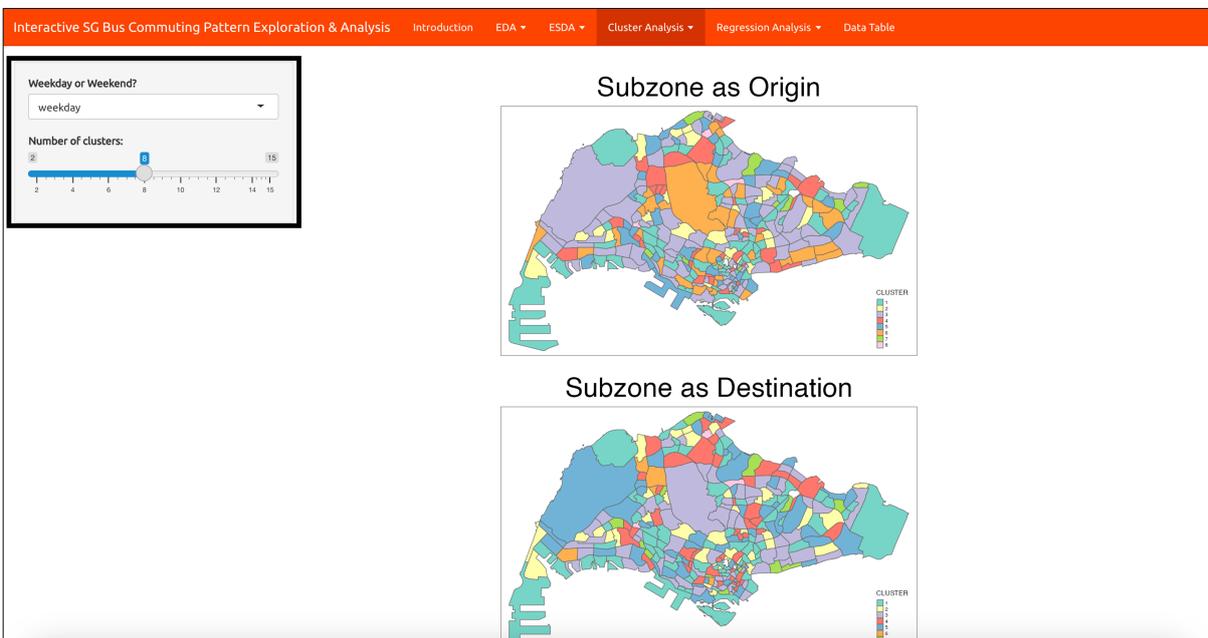
4. Clustering Analysis

The user is required to click on the drop-down menu to access the visualizations within ‘Cluster Analysis’. It has two tabs within – ‘Hierarchical Clustering’ and ‘Geospatial Constrained – SKATER Clustering’.



4.1 Hierarchical Clustering

The user can visualise hierarchical clustering on the Singapore map by subzone for origin and destination.

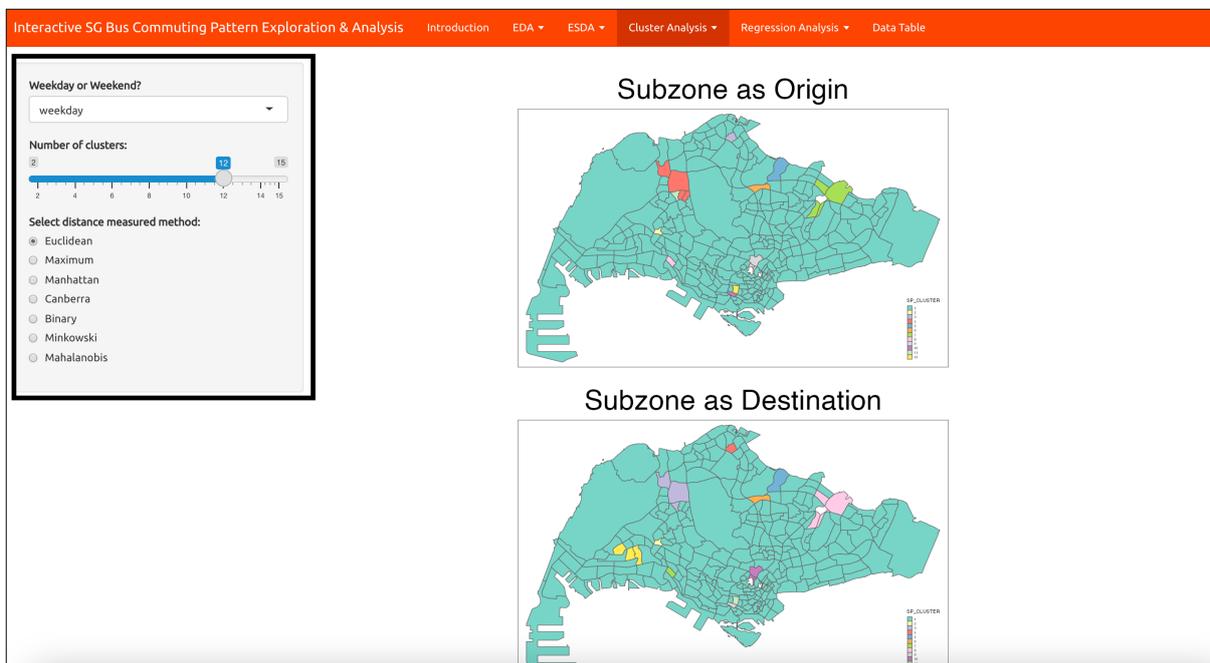


The user gets to select from:

1. Weekend or Weekday? – the user can choose to see the clusters for either weekend or weekday. Note that this selection will be applicable to both the subzone as origin and subzone as destination.
2. Number of Clusters – the user can pick the number of clusters they want to see. The minimum number of clusters are two and the maximum number of clusters are fifteen. This is a slider option, so the user needs to move the scale in order to see the modifications applied.

4.2 Geospatial Constrained – SKATER Clustering

The user can visualise geospatial constrained – skater clustering on the Singapore map by subzone for origin and destination.



The user gets to select from:

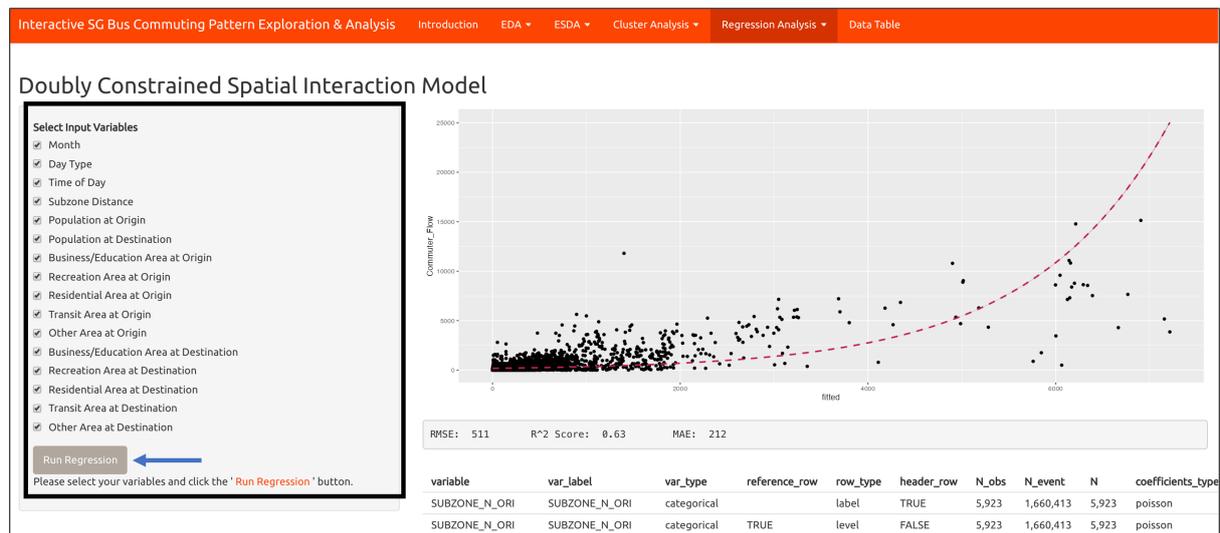
1. Weekend or Weekday? – the user can choose to see the clusters for either weekend or weekday. Note that this selection will be applicable to both the subzone as origin and subzone as destination.
2. Number of Clusters – the user can pick the number of clusters they want to see. The minimum number of clusters are two and the maximum number of clusters are fifteen. This is a slider option, so the user needs to move the scale in order to see the modifications applied.
3. Select distance measured method – the user can select from seven different distance measures ('Euclidean', 'Maximum', 'Manhattan', 'Canberra', 'Binary', 'Minkowski' and 'Mahalanobis').

5. Regression Analysis

The user is required to click on the drop-down menu to access the visualizations within 'Regression Analysis'. It has four tabs within – 'Unconstrained', 'Origin Constrained', 'Destination Constrained' and 'Double Constrained'.



The tabs within regression analysis operate in the same way. They differ in terms of the variables used within, given the choice of the model. For instance, an origin constrained model will have the origin subzone code with all the destination variables used in the regression. For this purpose only one tab – 'Double Constrained' model will be shown as an example:



These tabs will populate the regression plot, model summary and summary statistics for regression.

1. The user gets to select any of the input variables of their choice. They must select at least one for the regression to run.
2. Once the user has made their selection, they must click on the 'Run Regression' button to see the results. The user must be patient as this may take some time to run.
3. The variables are not listed down here as they will be slightly different for each model. All the regression tabs have all the variables selected as default as they garner the best regression results. The user can choose to play around with the variables to see if their results improve with opting for one over the other.

6. Data

The user can click on the 'Data Table' to access the data used for all the visualisations present in this application.

	IR_MONTH	DAY_TYPE	TIME_PER_HOUR	SUBZONE_N_ORI	PLN_AREA_N_ORI	REGION_N_ORI	flow
1	2022-10-01	WEEKDAY	0	ADMIRALTY	SEMBAWANG	NORTH REGION	45
2	2022-10-01	WEEKDAY	0	ALEXANDRA HILL	BUKIT MERAH	CENTRAL REGION	396
3	2022-10-01	WEEKDAY	0	ALEXANDRA NORTH	BUKIT MERAH	CENTRAL REGION	22
4	2022-10-01	WEEKDAY	0	ALJUNIED	GEYLANG	CENTRAL REGION	3588
5	2022-10-01	WEEKDAY	0	ANAK BUKIT	BUKIT TIMAH	CENTRAL REGION	1317
6	2022-10-01	WEEKDAY	0	ANCHORVALE	SENGKANG	NORTH-EAST REGION	334
7	2022-10-01	WEEKDAY	0	ANG MO KIO TOWN CENTRE	ANG MO KIO	NORTH-EAST REGION	2336
8	2022-10-01	WEEKDAY	0	BALESTIER	NOVENA	CENTRAL REGION	867
9	2022-10-01	WEEKDAY	0	BANGKIT	BUKIT PANJANG	WEST REGION	530
10	2022-10-01	WEEKDAY	0	BAYSHORE	BEDOK	EAST REGION	24

1. The user has six different panels available. The panel header informs the user of the source file for the visualisation used.
2. The user can choose the number of entries they wish to see on the screen. The options are 10, 25, 50 and 100.
3. The user can search for values/columns from the 'Search' option.
4. The user can navigate to different pages in the data table chosen.

We hope that the user guide made it easier to navigate the application.