

## User Guide

### Introduction Page

When you first visit the app, you'll be greeted by our landing page. In it, there is a short description of the app, as well as an overall view of the various tabs (ANOVA, UpSet Chart, Heat Map, Spatial Map and Network Graph) in the app.

Forensic Investigation App
Introduction
ANOVA
UpSet Chart
Heat Map
Spatial Map
Network Graph

### A Brief Introduction

This application explores the movement and tracking issues raised from VAST Challenge 2021: Mini Challenge 2.

With the aid of curated visualisations, we aim to forensically understand and investigate any suspicious activities, leading up to the disappearance of several GasTech employees.

This application is broken down into the different sections as per the image on the right.

Within each tab, you'll be greeted with a sidebar, as well as a main panel. While the sidebar allows easy customisation for each visualisation, the main panel is divided into various sub-tabs. The first sub-tab is always a guide on the visualisation objectives as well as steps on its use. Subsequent sub-tabs would showcase the visualisation itself.

As you explore the various tabs, do take advantage of the collapsible sidebar function. This is especially useful, when you want the visualisation to take up the whole width of your window. The collapsible sidebar function is denoted by the three horizontal white lines located beside the VAST 2021 Mini Challenge 2 title above.

Click here to find out more information about the [VAST Challenge 2021](#), as well as explore its accompanying dataset.

Alternatively, for more information on this project, do visit both our [Group Project site](#) and our [GitHub page](#).

Lastly, we would like to thank [Professor Kam Tin Seong](#) for his invaluable guidance and saintly patience in making this application a reality.

### App Overview

## ANOVA (Analysis of Variance)

This tab covers the analysis of variance tests on credit card data.

### 1. ANOVA – Guide

As you click on the ANOVA tab, you will land on the Guide section which explains the objective of ANOVA as well as general steps on its use.

Forensic Investigation App
Introduction
ANOVA
UpSet Chart
Heat Map
Spatial Map
Network Graph

ANOVA Plot Settings
Select Test Statistics
Non-Parametric
Select No. of Bins
3
Submit

### Understanding One-Way Analysis of Variance Test

Analysis of variance (ANOVA) test is conducted to check if the means of two or more groups are significantly different from each other. For example, we could use one-way ANOVA to understand if the test scores between 5 different classes are significantly different.

It is important to realize that one-way ANOVA is an omnibus test statistic. It is unable to tell us which specific groups were statistically significantly different from each other. Thus, we can make use of post-hoc test that can help us determine which of the groups differ from each other.

By using the `ggbetweenstats` function from `ggstatplot` package, we can easily visualize the results of one-way ANOVA. The visualization is built upon boxplot to show how the dependent variable distribute for each group. The function offers the flexibility to select the type of hypothesis testing, including parametric, non-parametric, robust of Bayes Factor. It also automatically selects the appropriate post-hoc test and displays the results of the pairwise comparisons on the chart.

In our use case, we made use of credit card transaction data to showcase how such a visualization can help us understand the difference in spend amount across locations and credit card users. The visualization aims to answer questions like 'Is the spending of user A significantly different from user B?' and 'Is the mean spend at this location significantly different from other cafes?'

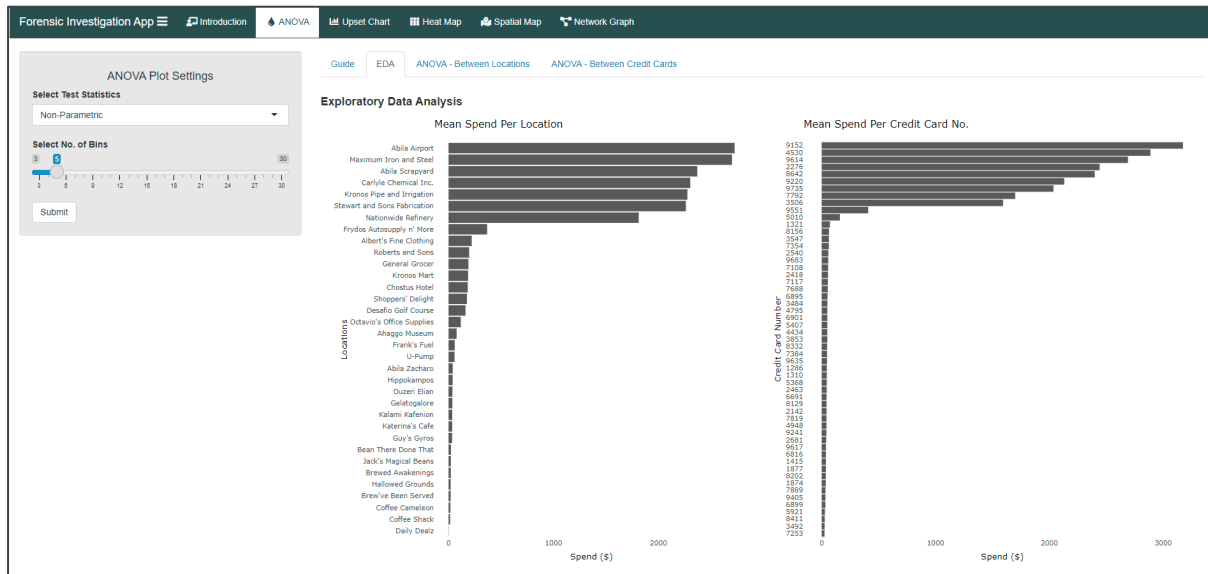
[Click here to find out more about ggstatplot R package.](#)

### Steps

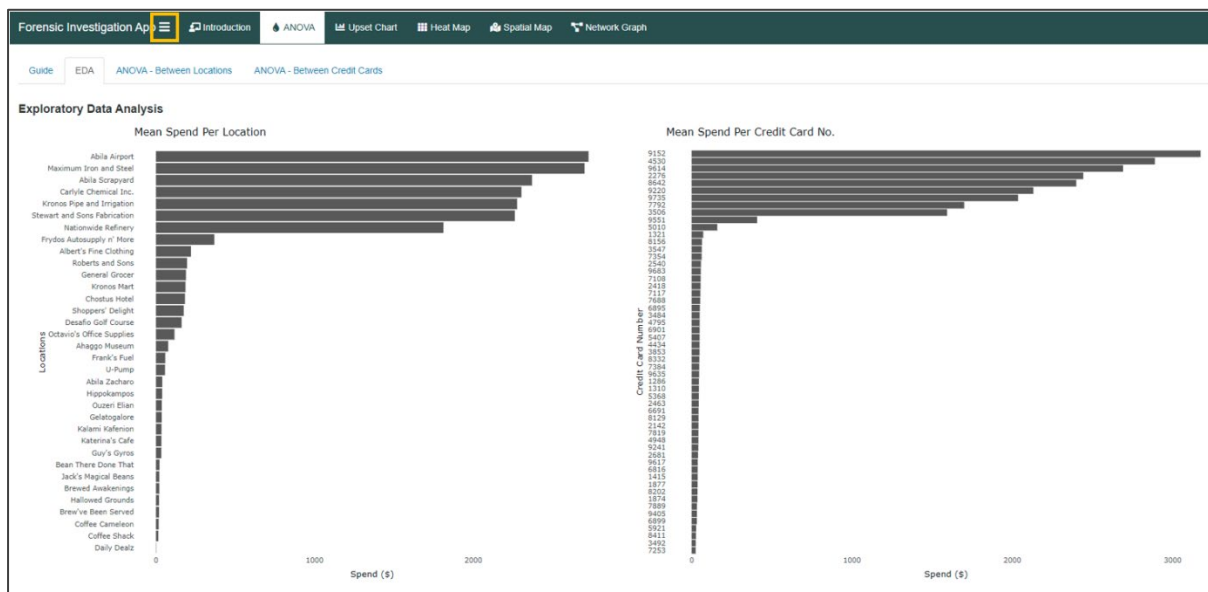
1. Begin by exploring the credit card transaction dataset under the 'EDA' tab.
2. Select the ANOVA tab based on the type of groups you are interested in comparing - 'ANOVA - Between Locations' or 'ANOVA - Between Credit Cards'.
3. Select the locations' or 'ANOVA - Between Credit Cards'.
4. You can play with the number of bins to understand the nature of the data to select the right test statistics.
5. For example, if the dataset is not normally distributed, you can select a non-parametric test.
6. Click on 'Submit' button and the plot will be updated with the one-way ANOVA and pairwise comparison results.

## 2. ANOVA – EDA (Exploratory Data Analysis)

Two horizontal histograms are displayed side-by-side to illustrate the mean spend per location and per credit card number.



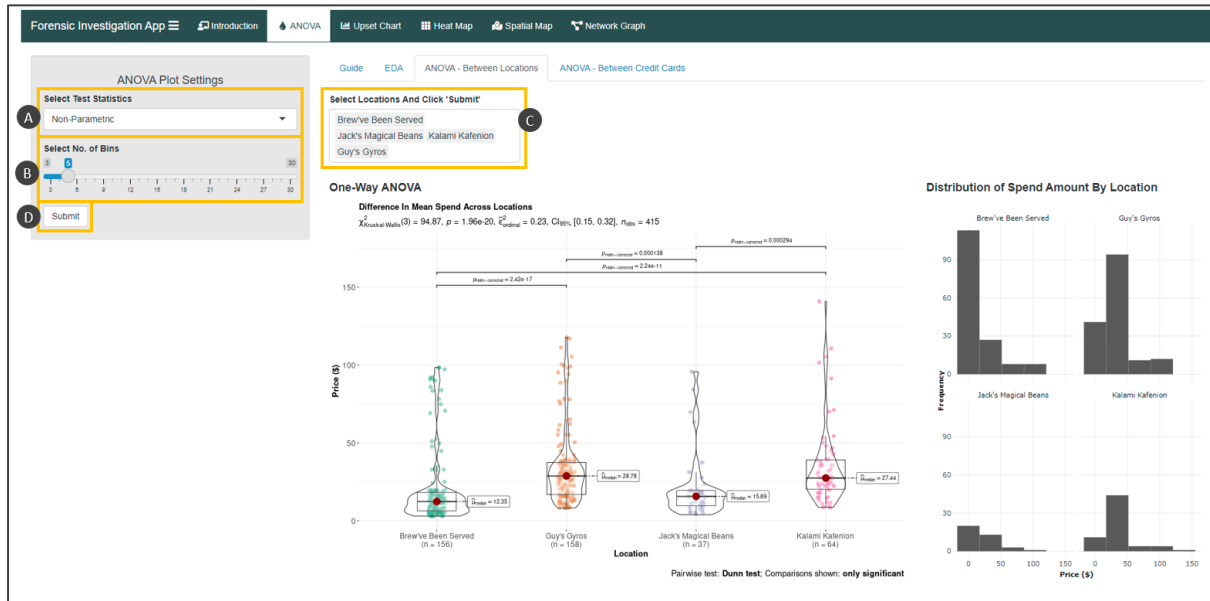
Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



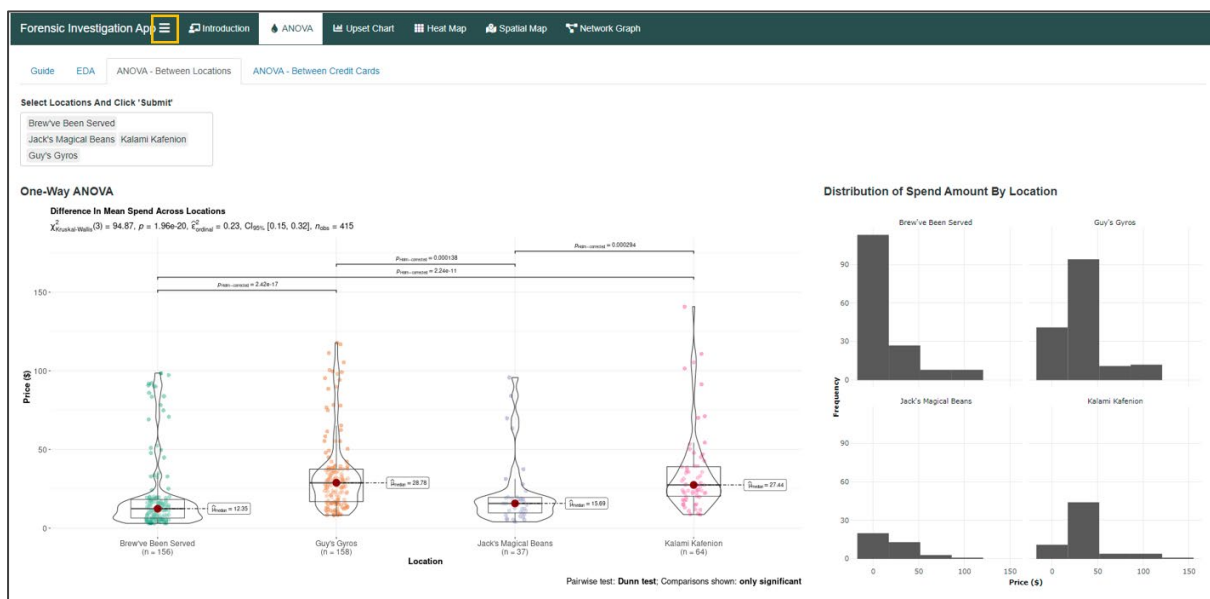
### 3. ANOVA – Between Locations

Here, we examine the ANOVA tests as well as the spending distributions for each selected location.

- Select Test Statistics (Parametric, Non-Parametric, Robust or Bayes Factor) to customize the ANOVA test method
- Select number of bins (between 3 to 30) to determine the bin sizes within the histograms
- Select Locations (from the drop-down list) to include as part of the visualization
- Once options have been selected, click 'Submit' to plot the One-Way ANOVA plot and Histogram Charts



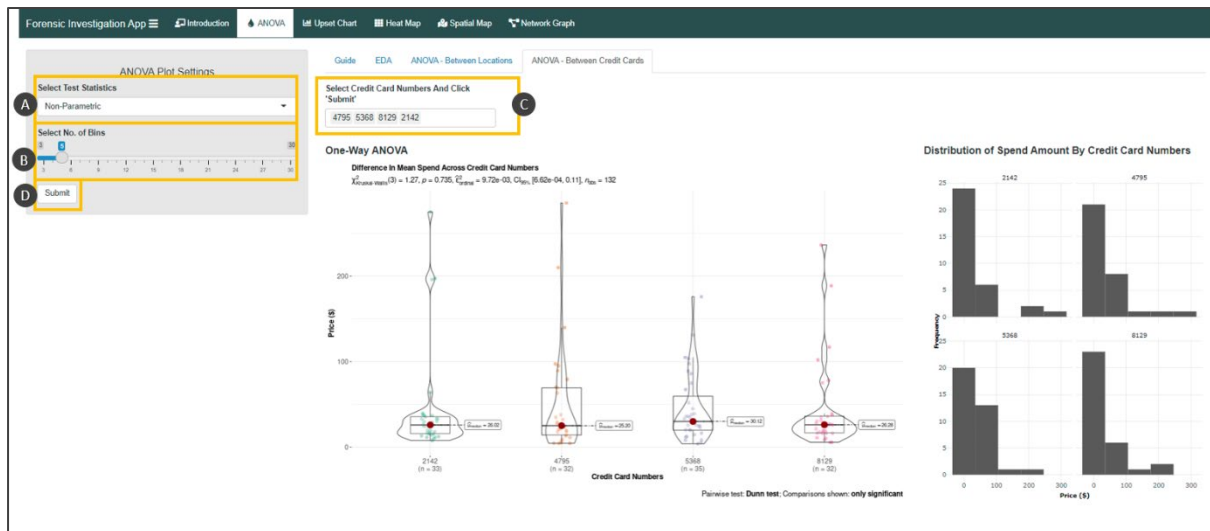
Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



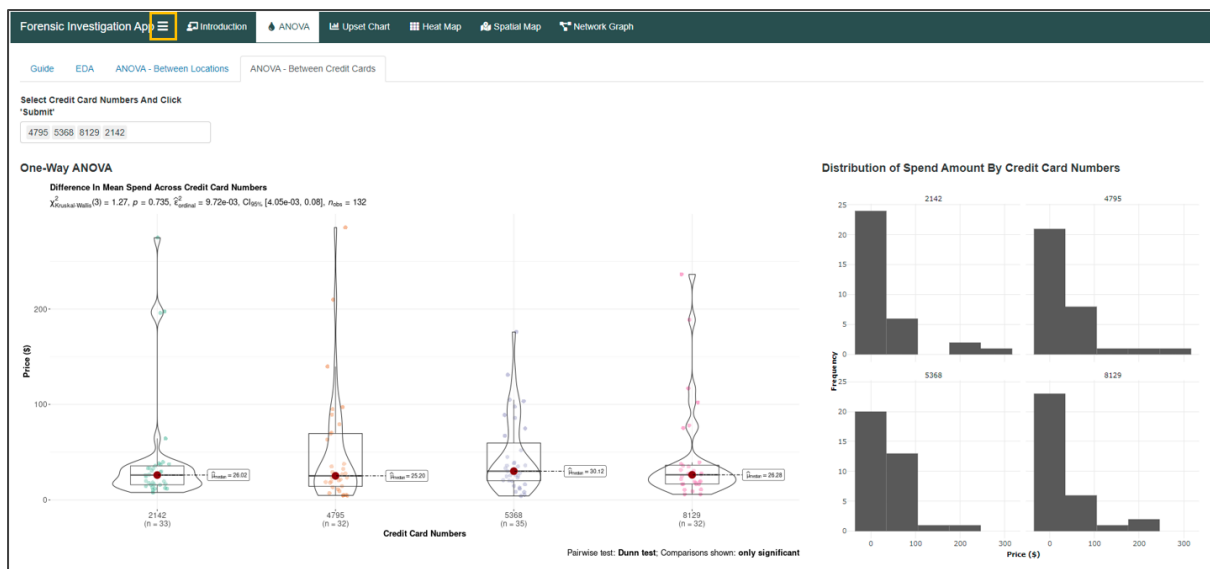
#### 4. ANOVA – Between Credit Cards

Here, we examine the ANOVA tests as well as the spending distributions for each selected credit card number.

- Select Test Statistics (Parametric, Non-Parametric, Robust or Bayes Factor) to customize the ANOVA test method
- Select number of bins (between 3 to 30) to determine the bin sizes within the histograms
- Select Credit Card Numbers (from the drop-down list) to include as part of the visualization
- Once options have been selected, click 'Submit' to plot the One-Way ANOVA plot and Histogram Charts



Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.

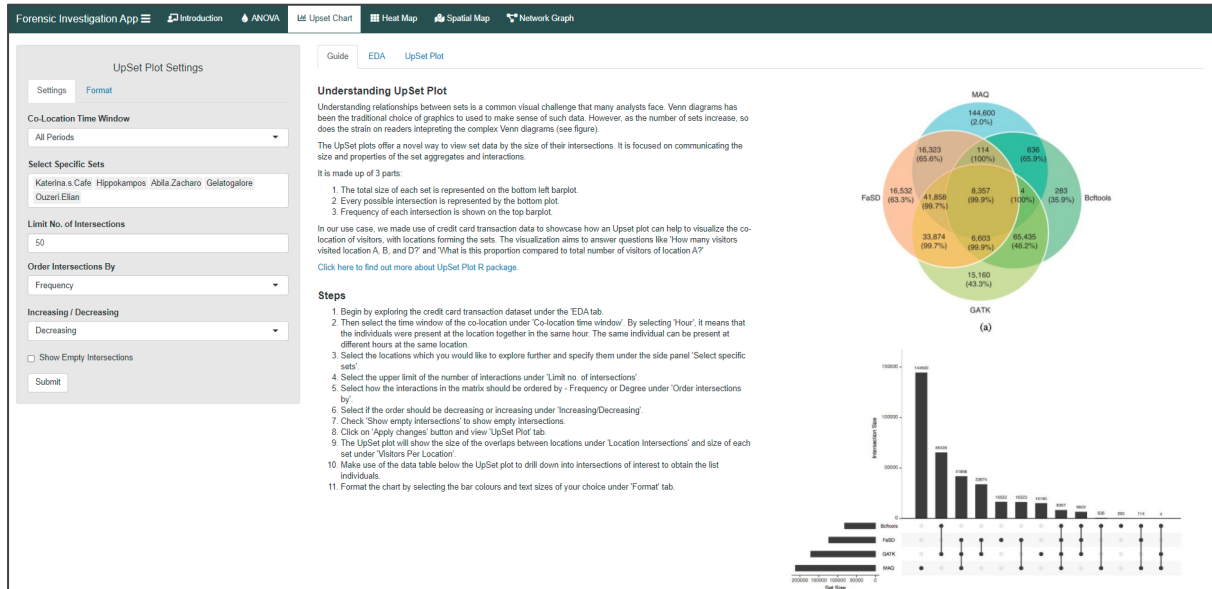


## UpSet Chart

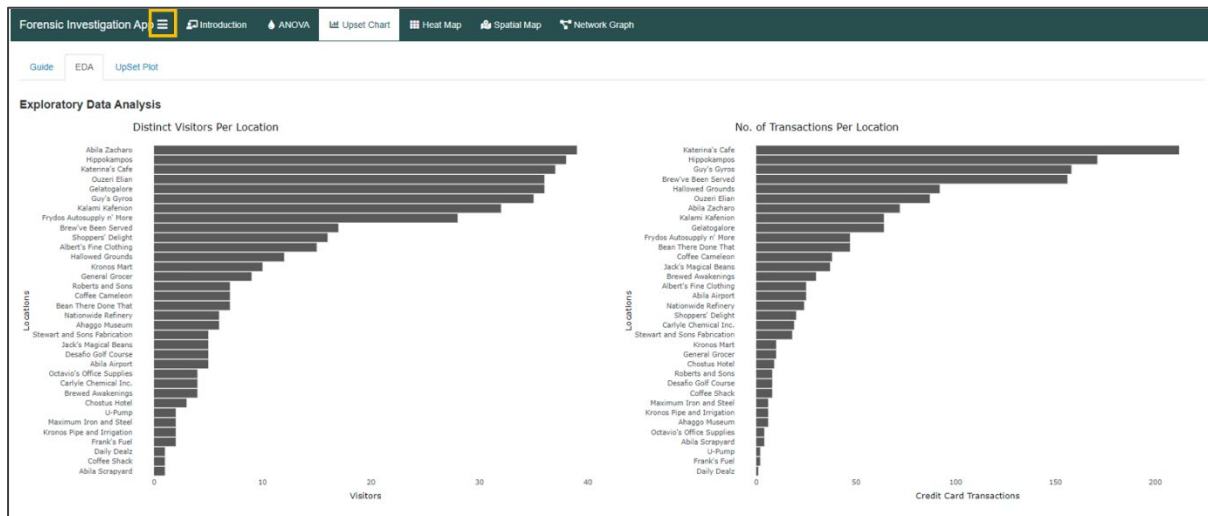
This tab covers the UpSet plot on credit card data.

### 1. UpSet Chart – Guide

As you click on the UpSet Chart tab, you will land on the Guide section which explains the objective of UpSet plot as well as general steps on its use.



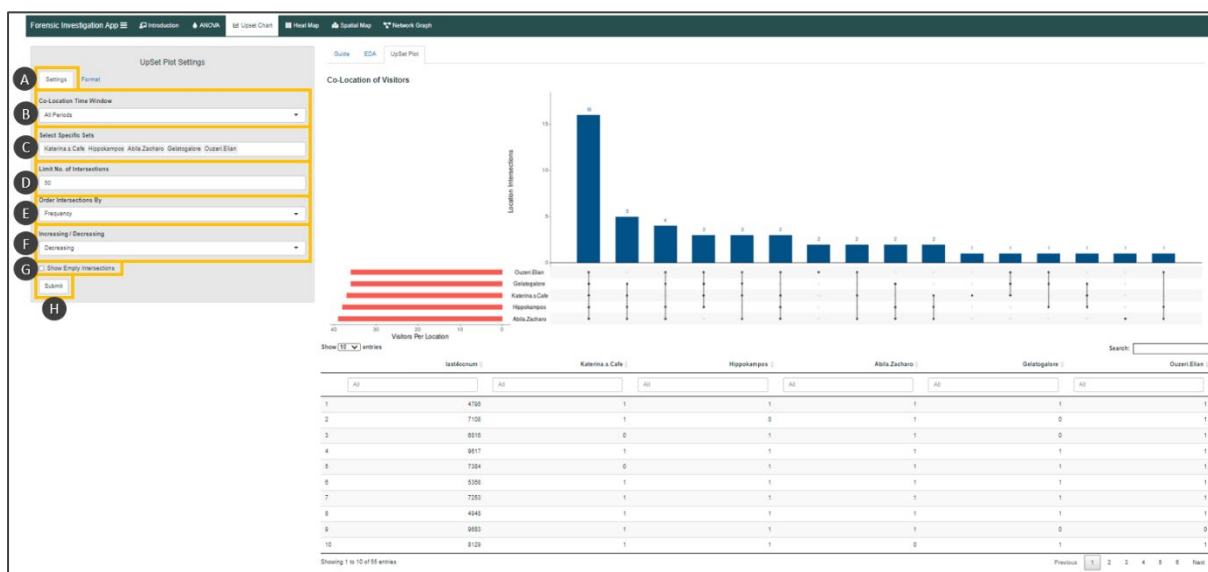
Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



### 3. UpSet Chart – UpSet Plot

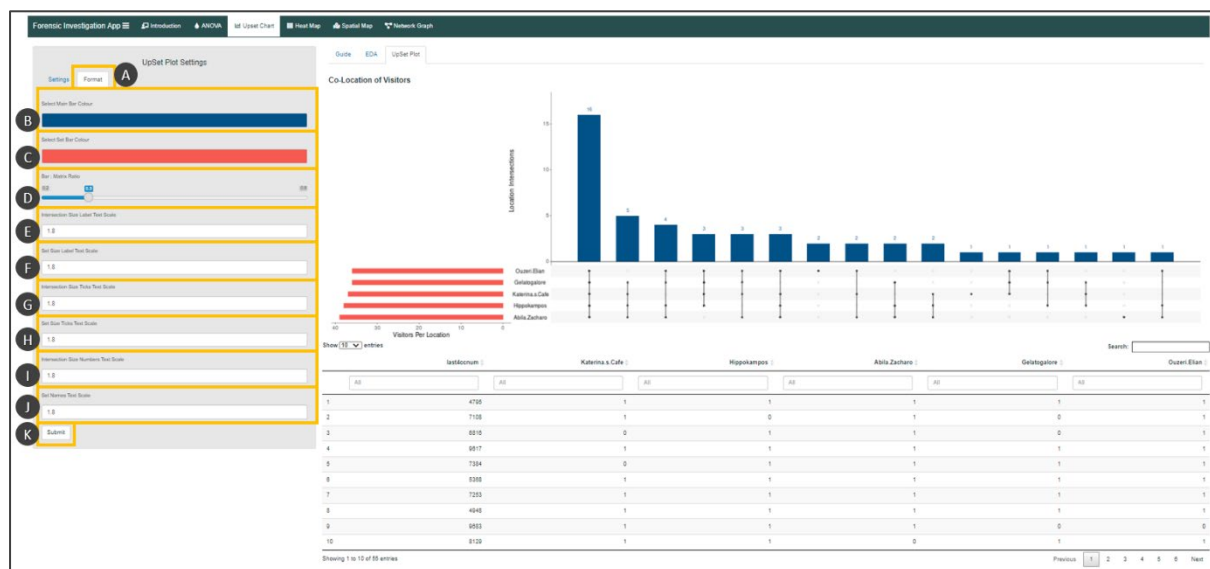
Here, we examine the UpSet Plot on Co-Location of Visitors, by exploring its parameter settings.

- Select the Settings Tab (This gives you the UpSet Plot parameter settings)
- Select Co-Location Time Window (All Periods, Day or Hour)
- Select Specific Sets (from the drop-down list of locations) to include as part of the visualization
- Limit the number of intersections
- Order intersections by either Degree or Frequency
- Choose to rank either by Increasing or Decreasing values
- Check this box if you intend to show empty intersections
- Once parameter options have been selected, click 'Submit' to plot the UpSet plot and update the table

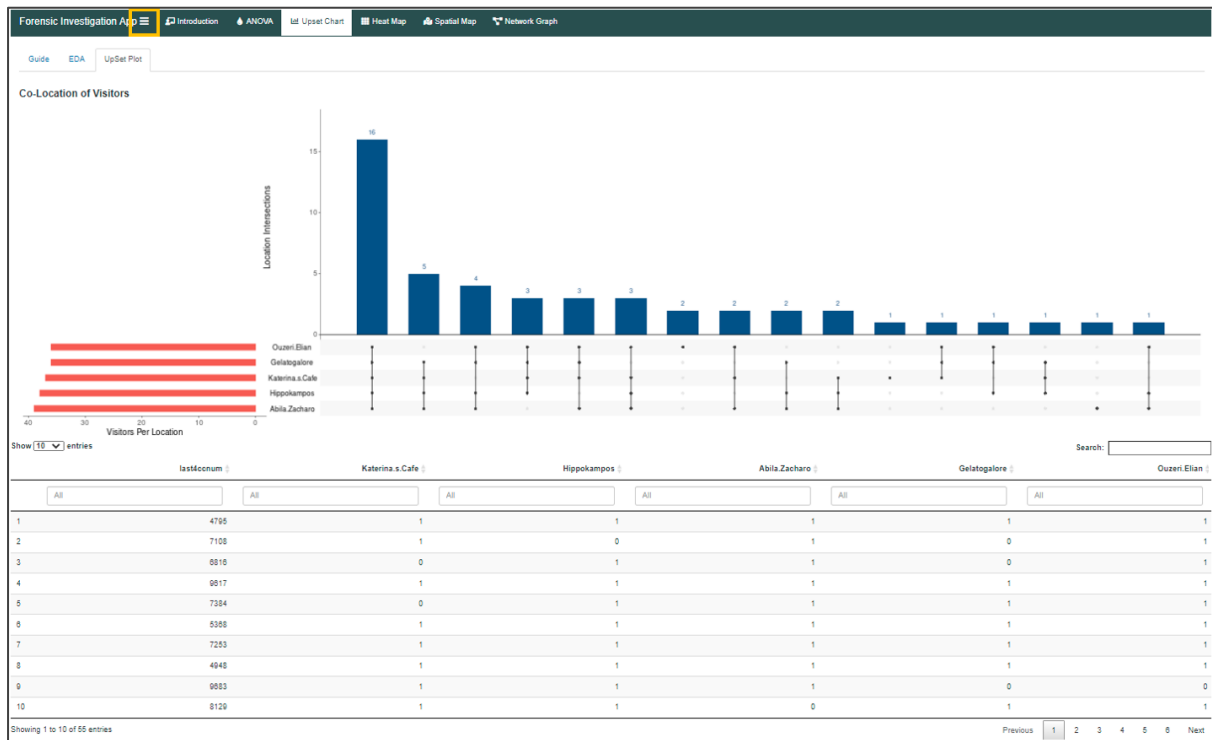


Here, we examine the UpSet Plot on Co-Location of Visitors, by exploring its format settings.

- Select the Format Tab (This gives you the UpSet Plot format settings)
- Select main bar colour
- Select the set bar colour
- Select the bar to matrix ratio (between 0.2 to 0.8)
- Select the Intersection Size Label Text Scale
- Select the Size Label Text Scale
- Select the Intersection Size Ticks Text Scale
- Select the Set Size Ticks Text Scale
- Select the Intersection Size Numbers Text Scale
- Select the Set Names Text Scale
- Once parameter options have been selected, click 'Submit' to update the format of both the UpSet plot and table.



Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



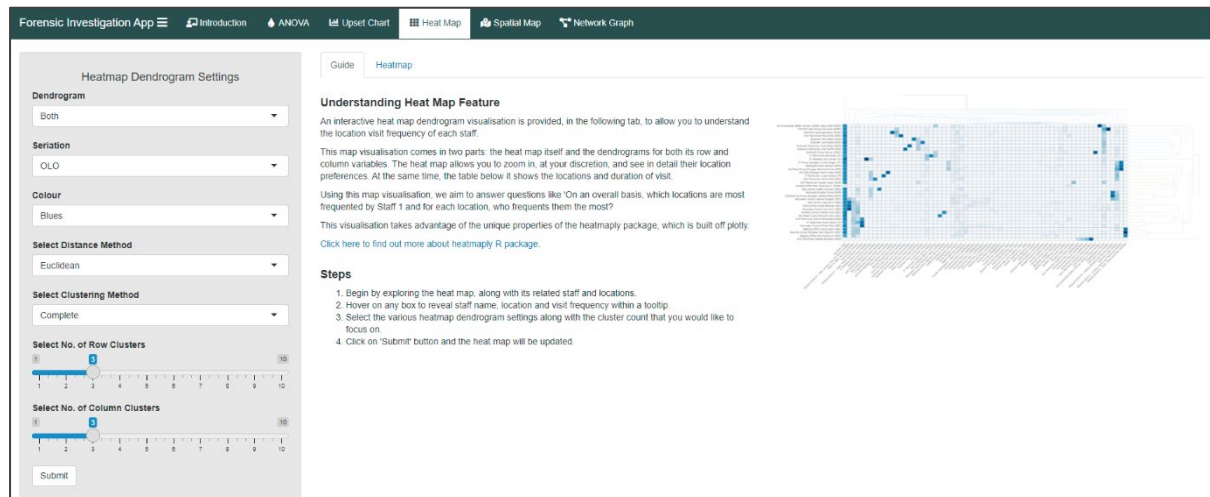


## Heat Map

This tab covers the Heat Map with Hierarchical Clustering Dendrograms on GPS data.

### 1. Heat Map – Guide

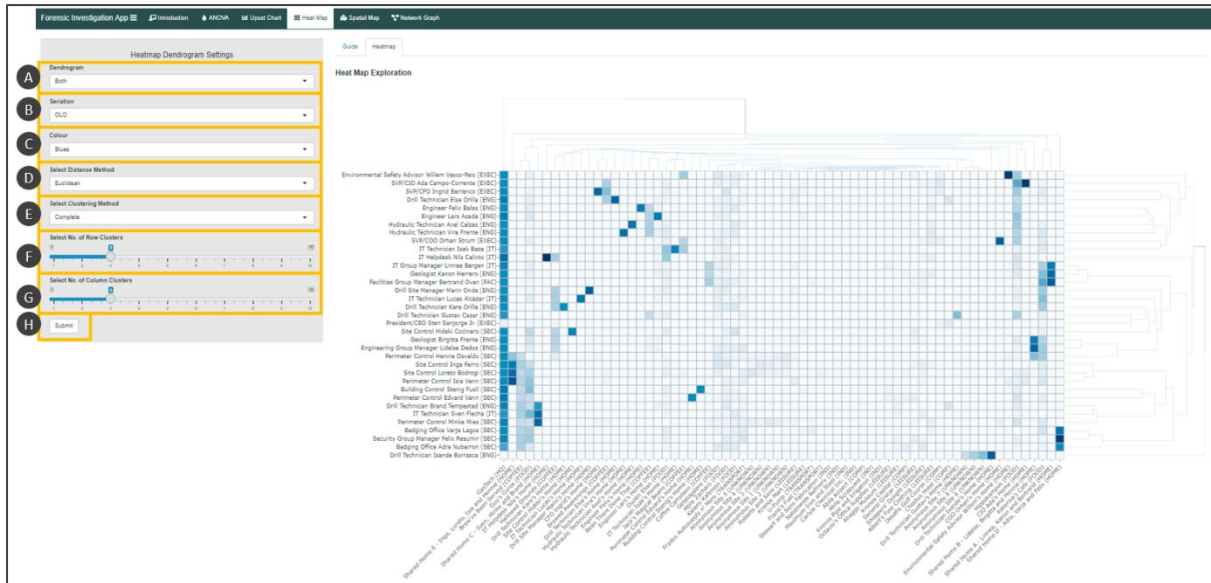
As you click on the Heat Map tab, you will land on the Guide section which explains the objective of Heatmap with Hierarchical Clustering as well as general steps on its use.



### 2. Heat Map – Heatmap

Here, we examine the Heat Map on Staff and Locations.

- Select whether you intend to show both the row and column dendrograms, either just the row or column dendrogram, or neither of them
- Select Seriation method (OLO, Mean, None or GW)
- Select Heat Map colour (based on Colour Brewer's colour ramp)
- Select Distance Method (Euclidean, Maximum, Manhattan, Canberra, Binary or Minkowski)
- Select Clustering Method (Ward.D, Ward. D2, Single, Complete, Average, Mcquitty, Median or Centroid)
- Select number of row clusters (between 1 to 10)
- Select number of column clusters (between 1 to 10)
- Once options have been selected, click 'Submit' to update the heatmap and dendrograms



Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



## Spatial Map

This tab covers the Spatial Map on GPS data.

### 1. Spatial Map – Guide

As you click on the Spatial Map tab, you will land on the Guide section which explains the objective of Spatial Map as well as general steps on its use.

Forensic Investigation App

Introduction

ANOVA

Upset Chart

Heat Map

Spatial Map

Network Graph

Select Date

6 Jan

Compare Max Three Staff

Select Staff ID 1

Drill Technician Isande Borrasca

Select Line Colour 1

Select Line Type 1

Solid

Select Staff ID 2

Drill Technician Elsa Orilla

Select Line Colour 2

Select Line Type 2

Dashed

Select Staff ID 3

Drill Technician Brand Tempestad

Select Line Colour 3

Select Line Type 3

Dotted

Submit

Guide

Map

#### Understanding Spatial Map Feature

An interactive map visualisation is provided, in the following tab, to allow you to explore the day-by-day movements of three chosen staff, as they traverse the city.

This map visualisation comes in two parts, the map itself and the table below it. The map allows you to zoom into the map, at your discretion, and understand the movements made by the chosen staff. At the same time, the table below it shows the locations and duration of visit.

Using this map visualisation, we aim to answer questions like 'On a certain day, which locations were visited by Staff 1 and for how long? What was the route taken by Staff 1?'

This visualisation takes advantage of the unique properties of the tmap package, where it allows layer by layer customisation.

[Click here to find out more about tmap R package.](#)

#### Steps

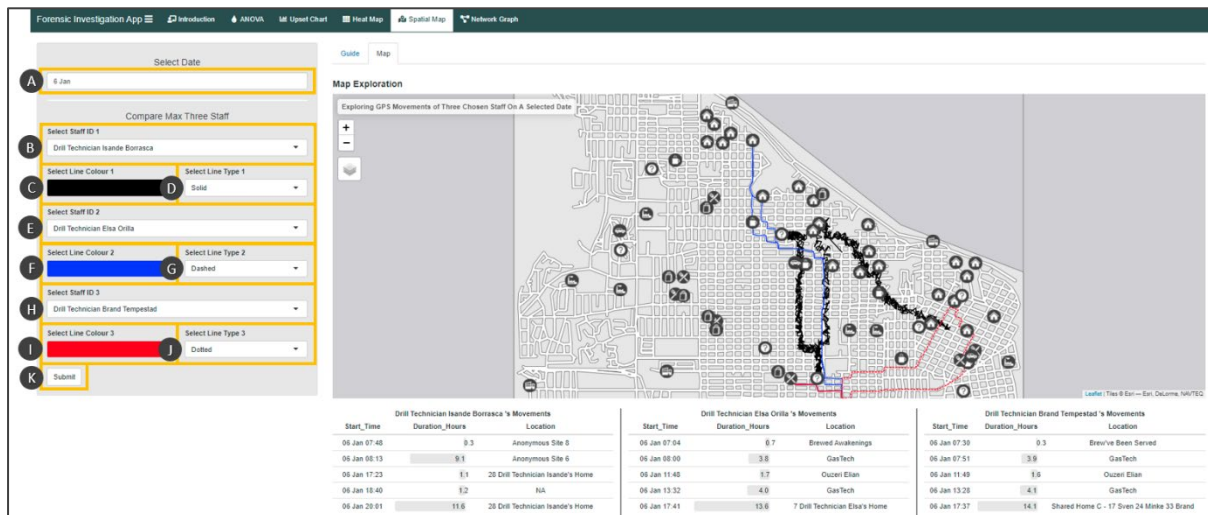
1. Begin by exploring the map, along with its related paths and locations under the 'Map' tab.
2. Click on the lines and locations to reveal more information within a tooltip.
3. Select the date of interest, followed by the three staff you would like to focus on.
4. Customise and differentiate each staff's path by selecting its respective line colour and line type.
5. Click on 'Apply Changes' button and the map will be updated with the respective paths.
6. View the table below the map to see the locations visited by each staff, as well as its start time and duration of visit.

Staff	Location	Start Time	End Time	Duration
Isande Borrasca	Location 1	08:00:00	08:05:00	5 min
	Location 2	08:05:00	08:10:00	5 min
	Location 3	08:10:00	08:15:00	5 min
	Location 4	08:15:00	08:20:00	5 min
Elsa Orilla	Location 5	08:20:00	08:25:00	5 min
	Location 6	08:25:00	08:30:00	5 min
	Location 7	08:30:00	08:35:00	5 min
	Location 8	08:35:00	08:40:00	5 min
Brand Tempestad	Location 9	08:40:00	08:45:00	5 min
	Location 10	08:45:00	08:50:00	5 min
	Location 11	08:50:00	08:55:00	5 min
	Location 12	08:55:00	09:00:00	5 min

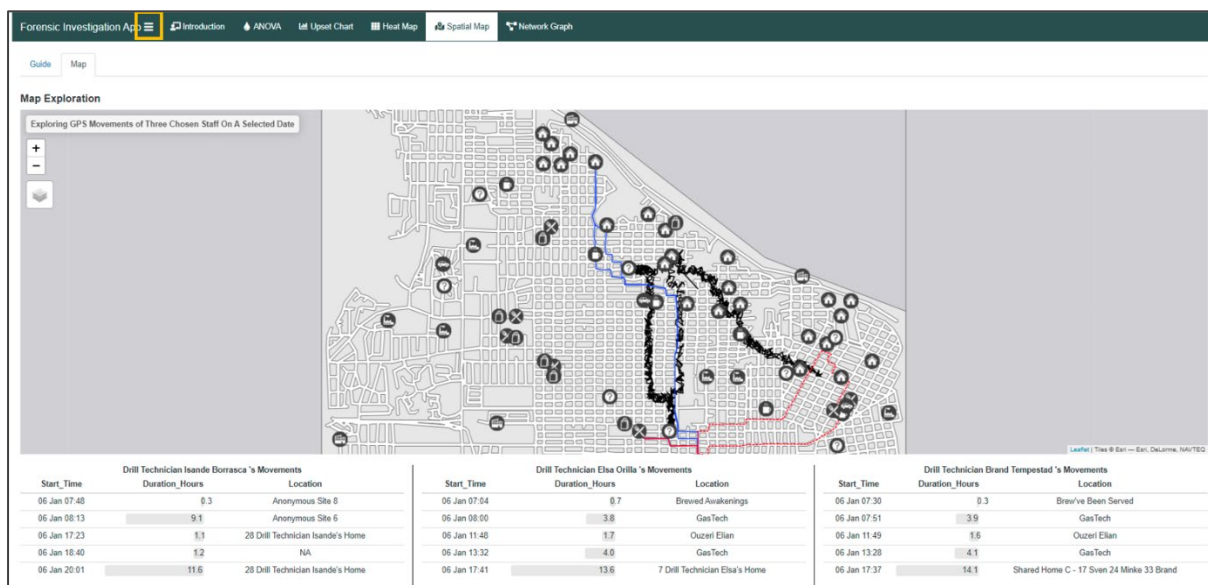
## 2. Spatial Map – Map

Here, we examine the Heat Map on Staff and Locations.

- Select date of GPS (between 6<sup>th</sup> Jan and 19<sup>th</sup> Jan)
- Select first staff of interest (drop-down list excludes second and last selected staff)
- Select line colour on map of first staff
- Select line type on map of first staff (drop-down list excludes second and last selected line type)
- Select second staff of interest (drop-down list excludes second and last selected staff)
- Select line colour on map of second staff
- Select line type on map of second staff (drop-down list excludes second and last selected line type)
- Select last staff of interest (drop-down list excludes first and second selected staff)
- Select line colour on map of last staff
- Select line type on map of last staff (drop-down list excludes first and second selected line type)
- Once options have been selected, click 'Submit' to update the spatial map and table



Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.



## Network Graph

This tab covers the Network Graph on credit card and loyalty data.

### 1. Network Graph – Guide

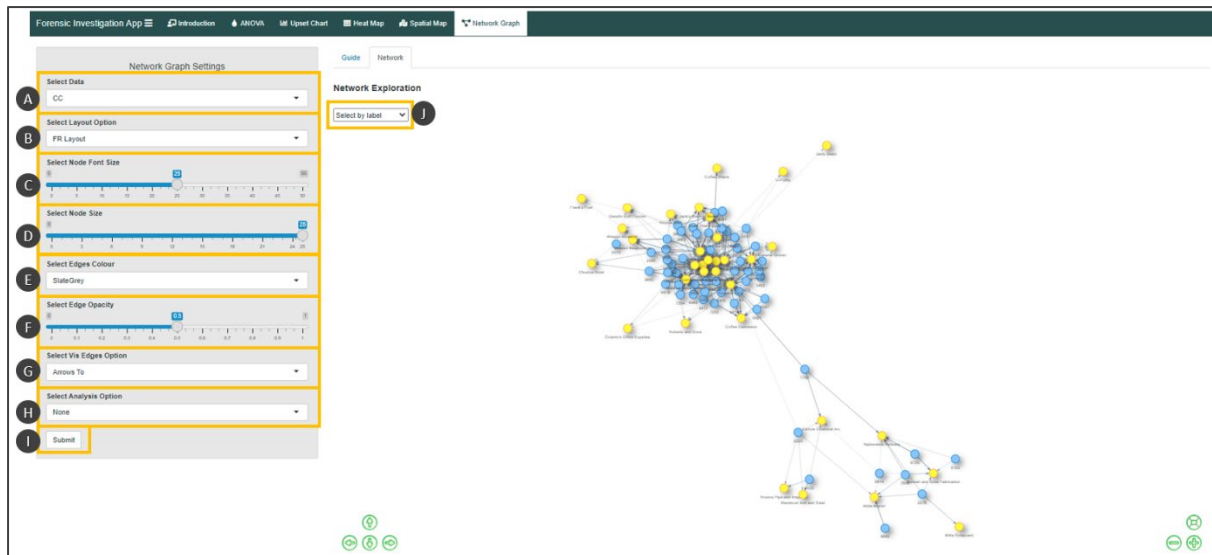
As you click on the Network Graph tab, you will land on the Guide section which explains the objective of Network Graph as well as general steps on its use.

The screenshot displays the 'Forensic Investigation App' interface. The top navigation bar includes 'Introduction', 'ANOVA', 'Mid Upset Chart', 'Heat Map', 'Spatial Map', and 'Network Graph'. The 'Network Graph' tab is active, showing a 'Guide' section. On the left, the 'Network Graph Settings' panel includes dropdowns for 'Select Data' (set to 'CC'), 'Select Layout Option' (set to 'FR Layout'), 'Select Edges Colour' (set to 'SlateGrey'), and 'Select Analysis Option' (set to 'None'). It also features sliders for 'Select Node Font Size' (set to 25) and 'Select Edge Opacity' (set to 0.5). A 'Submit' button is at the bottom of the settings panel. The main content area, titled 'Understanding the Network Graph Feature', explains that network graphs use graph theory to transform datasets into nodes and edges, highlighting relationships and importance. It lists steps for using the feature, including selecting data, layout, and analysis options, and clicking 'Submit' to update the graph. A network graph visualization is shown on the right, with nodes and edges representing relationships. A 'Select by label' dropdown is visible above the graph.

### 2. Network Graph – Network

Here, we examine the Heat Map on Staff and Locations.

- A. Select either CC(Credit Card) or Loyalty data
- B. Select layout option (Random, FR, Circle, Nicely or KK)
- C. Select node font size (between 0 to 50)
- D. Select node size (between 0 to 25)
- E. Select edges colour (either Light Grey, Dark Grey or Slate Grey)
- F. Select edges opacity (between 0 to 1)
- G. Select vis edges option (either Arrows From, Arrows To, or Middle)
- H. Select Analysis Option (Either None, Eigen or Degree)
- I. Once options have been selected, click 'Submit' to update the network graph
- J. Select label number to focus on a particular number



Clicking on the collapsible button (denoted by the three white lines) hides the sidebar panel allowing for the visualization to take up the whole width of the window.

