Analyzing and Visualizing Traffic Volume Patterns through Data Mining Techniques

Higher National Diploma in Information System Management 24.1F

Data Warehousing and Business Intelligence – Coursework

Group Members Name	Index
A.DESHMIGA	KAHMDISM24.1F-002
D.H.A.GOONASEKERE	KAHMDISM24.1F-015
T.G.N.D.JAYASINGHE	KAHMDISM24.1F-016
H.P.A.S.M.KUMARI	KAHMDISM24.1F-017
A.M.WEERASINGHE	KAHMDISM24.1F-018

School of Computing and Engineering

National Institute of Business Management

No:2, Asgiri Vihara Mawatha, Kandy

Title of the project: Analyzing and Visualizing Traffic Volume Patterns through Data Mining Techniques.

Authors: A.DESHMIGA
D.H.A.GOONASEKERE
T.G.N.D.JAYASINGHE
H.P.A.S.M.KUMARI
A.M.WEERASINGHE

Name of the Program: Higher National Diploma in Information System Management.

Name of the lecturer: Ms. L.S. Chathurika

Name of the Module: Data Warehousing and Business Intelligence.

Name of Institute: Management Information system division, National Institute of Business Management.

Date: ……..…

The project is submitted in partial fulfillment of the requirements for the module Data Warehousing and Business Intelligence as part of the Higher National Diploma

in Information System Management at the National Institute of Business Management.

Declaration

"I certify that this project does not incorporate without acknowledgement, any material previously submitted for a Higher National Diploma in any institution and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my project report, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organizations"

Student Name	**Index**	**Signature**
A.DESHMIGA	KAHMDISM24.1F-002	
D.H.A.GOONASEKERE	KAHMDISM24.1F-015	
T.G.N.D.JAYASINGHE	KAHMDISM24.1F-016	
H.P.A.S.M.KUMARI	KAHMDISM24.1F-017	
A.M.WEERASINGHE	KAHMDISM24.1F-018	

Name of the lecturer: Ms. L.S. Chathurika

…… ……...

Signature Date

List of Figures

Contents

Chapter 1: Introduction

1.1 Overview of the Chosen Dataset

The Metro Interstate Traffic Volume dataset was acquired from a reputable online data repository. The dataset was selected because it was ideal for a regression problem whereby traffic volume is predicted versus a number of determinants. Some of the key attributes of the dataset include:

- Sensor Data: Hourly traffic volume measurements on the I-94 highway (westbound) were collected by the ATR 301 sensor.
- Temporal Attributes: Comprehensive time-related details, including hours, days, and dates, were provided, thereby enabling close scrutiny of daily and seasonal traffic trends.
- Weather Conditions: Relevant variables such as temperature, precipitation, and cloud cover were incorporated to determine their effects on traffic flow.
- Holidays and Special Events: Holiday indicators were incorporated to facilitate exploration of traffic volume patterns on non-typical days.

For this study, the Test Dataset (test.csv) was the sole file utilized. This subset, consisting of the latest year of traffic data collected, was utilized so that testing of the models could be conducted on data that had not been seen before, thereby enhancing the assessment of the predictive model's capacity to generalize.

1.2 Dataset Justification

The dataset was deemed exactly suitable for this research due to several reasons:

- It provided a full range of variables influencing traffic volume.
- The data's granularity (measurements per hour) allowed close temporal analysis.
- The inclusion of both weather conditions and holiday indicators allowed for a multifaceted analysis of traffic patterns.
- The split into training and test datasets, the latter of which represented the most recent data, was considered necessary to simulate real-world model evaluation conditions.

1.3 Data Cleaning and Preprocessing

The following are the key methods applied in the data cleaning and preprocessing process:

- Handling Missing Values:

  Missing data points were treated with appropriate imputation methods to ensure overall data integrity.

- Data Type Conversion:

  Timestamps were normalized into a uniform date_time format, and categorical variables were converted to numeric representations to enable temporal analysis.

- Outlier Detection and Correction:

  Outliers and anomalous data points were identified through exploratory analysis. Where data points were identified as sensor errors or irregularities, corrections or removals were performed. Normalization of Numerical Attributes: Continuous variables, especially those dealing with weather and traffic volume, were normalized to ensure that they were on the same scales and to avoid any one attribute disproportionately affecting the outcomes.

Chapter 2: Data Analysis and Visualization

2.1 Overview

The chapter presents a detailed analysis of traffic volume data using various visualization approaches to discover patterns and relationships among significant variables. The results are aimed at informing traffic management and urban planning initiatives.

2.2 Visualization Techniques

The following visualization techniques in Power BI have been utilized in analyzing the traffic volume information:

- Clustered Bar Chart: Analyzed distribution of traffic volumes at different times of a day
- Stacked Bar Chart: Volumes of traffic during holidays and non-holiday days compared
- Key Performance Indicator (KPI): Displayed total traffic volume trends over time.
- Scatter Chart: Examined the relation between temperature and traffic volume.
- Clustered Column Chart: Depicted monthly fluctuations in traffic volumes.
- Line Chart: Displayed trends in traffic volumes over time intervals.
- Area Chart: Depicts cumulative traffic volumes over years
- Tree Map: Illustrating distribution of traffic volumes under various weather conditions
- Donut Charts: Traffic volumes displayed in terms of holiday state, weather, date, and month.
- Pie Chart: Depiction of volumes of traffic under different principal weather conditions
- Slicer with Dropdown: Allowed filtering of data by specific date components (year, month, and day).
- Gauge: Displayed current traffic volume against predefined thresholds.

- Cards: Presented key information such as date/time, holiday information, weather, and general traffic volumes.

2.3 Detailed Descriptions and Interpretation of Visualizations

2.3.1 Clustered Bar Chart: Traffic Volume by Hour

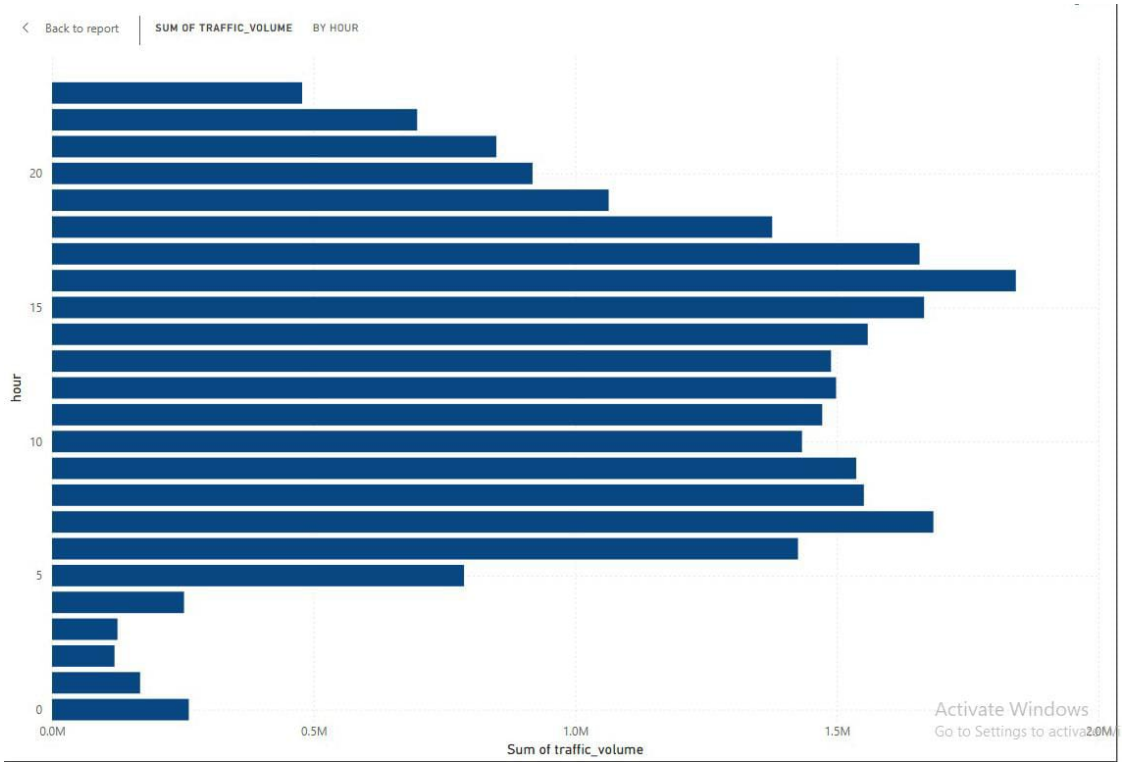


Figure 1: Clustered Bar Chart

- Description:

  This chart plots the distribution of hourly traffic volumes for each hour in a day. On the Y-axis, the hours (0–23) have been represented, and on the X-axis, the cumulative traffic volumes have been displayed.

- Interpretation:

  The visualization reveals most of the traffic during morning (around 8 AM) and evening (5–6 PM) times, indicative of common commuter peak times. Conversely,

a considerable drop in traffic during early morning (1 AM to 5 AM) times, indicative of reduced use of highways during off-peaking hours, can be seen.

2.3.2 Stacked Bar Chart: Traffic Volumes over Holidays

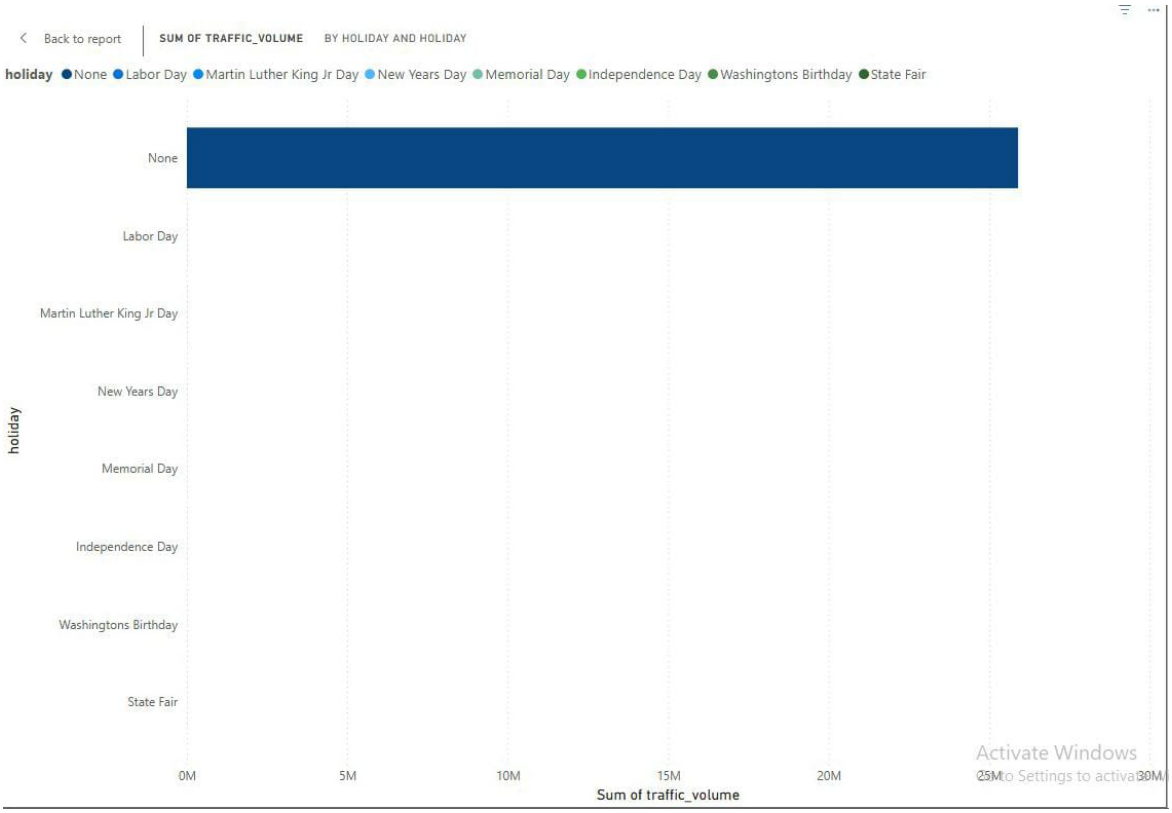


Figure 2: Stacked Bar Chart

- Description:

  This graph plots both holiday and non-holiday traffic volumes. Y-axis has holiday status, and X-axis has the sum of traffic volume, with legend separating holiday and non-holiday data by color-coding.

- Interpretation:

The chart shows that most of the traffic happens during non-holiday days. There is a lot less traffic during specific holidays such as Martin Luther King Jr. and New Year's Day, and it seems that commuting and commercial activity drop during these days.

2.3.3 KPI: Daily Traffic Volume

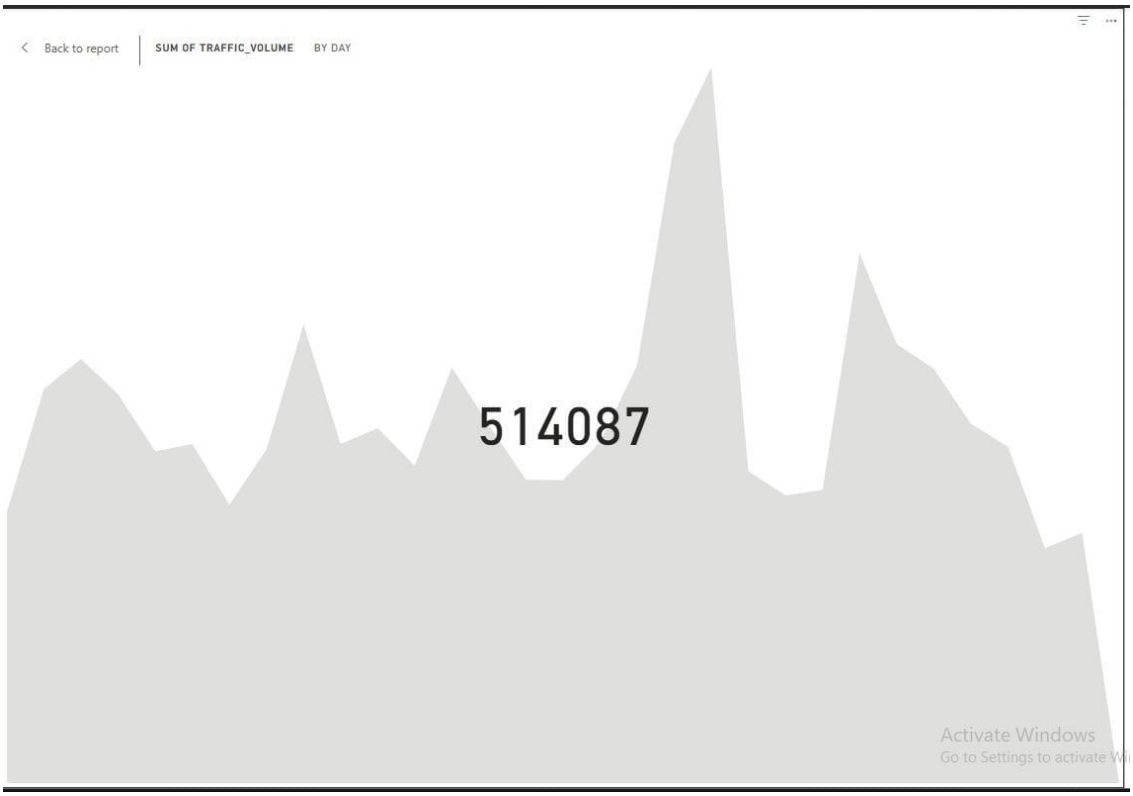


Figure 3: KPI: Daily Traffic Volume

- Description:

  The KPI visualization portrays a general view of daily traffic count and its trend over a specific period of time. It portrays the summation of traffic count with a trend marker in relation to the day.

- Interpretation:

This indicator reveals that weekday days have a larger volume of traffic than weekend days. Trend over a period of time aids in a rapid determination of whether traffic volumes are growing, constant, or shrinking.

2.3.4 Scatter Chart: Traffic Volume vs. Temperature

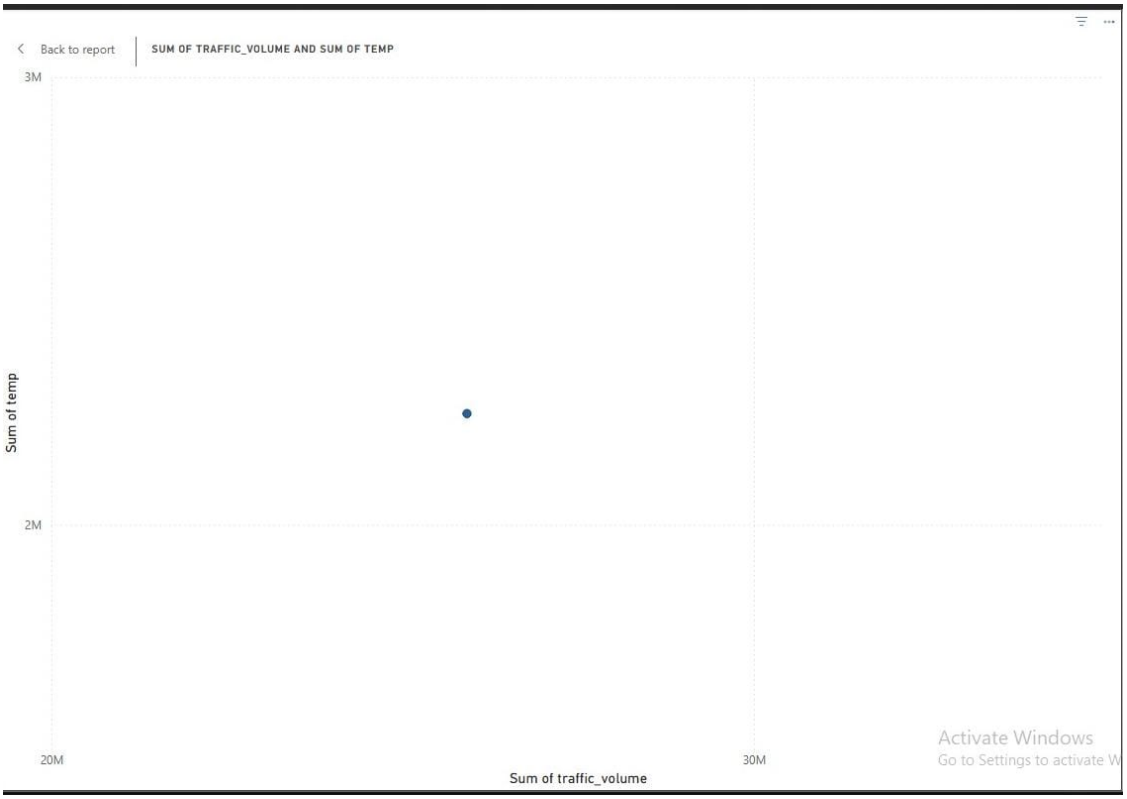


Figure 4: Scatter Chart

- Description: This scatter plot is an examination of temperature and traffic volume relation. X-axis is for summation of traffic volume, and Y-axis is for summation of temperature.

- Interpretation: The scatter plot infers a positive relationship in that extreme temperatures (high and low) can be seen to have an association with reduced volumes, and moderate temperatures with high volumes. This infers that poor weather could act as a deterrent to traveling.

2.3.5 Clustered Column Chart: Traffic Volume By Month

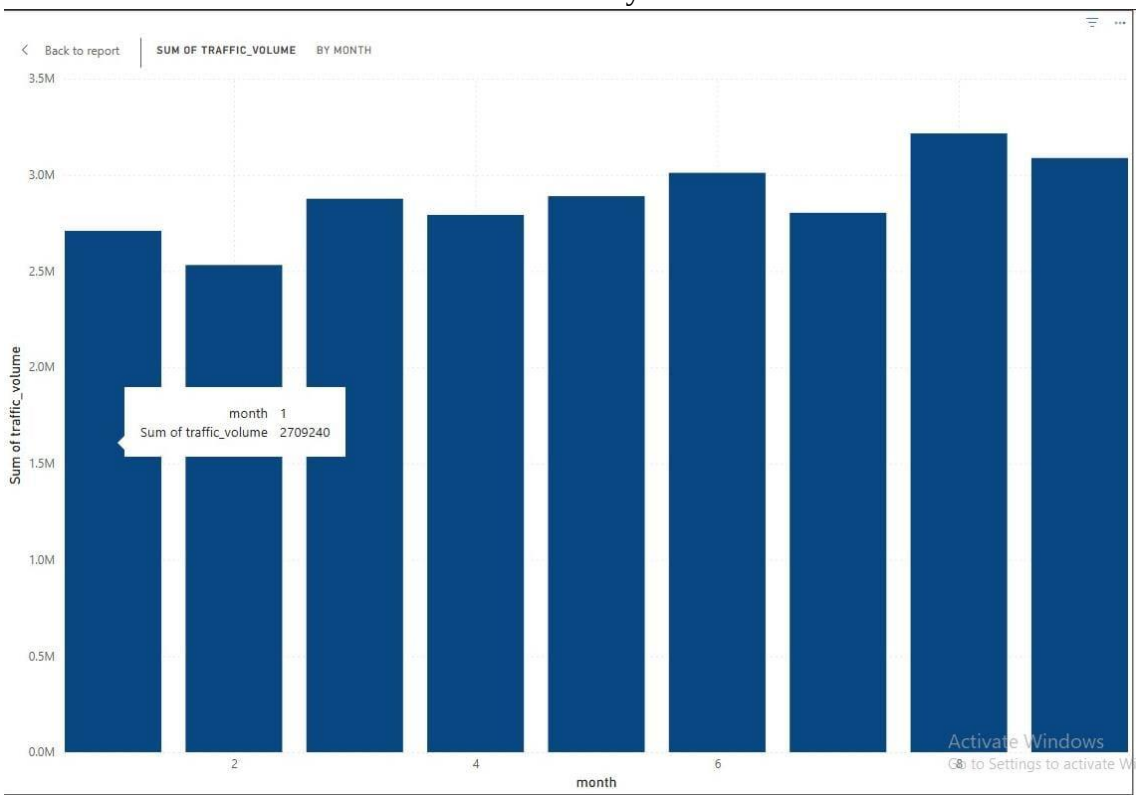


Figure 5: Column Chart: Traffic volume by month

- Description:

  This chart shows monthly variation in volumes of traffic. The X-axis is for representing the month, and Y-axis for representing the overall traffic volume.

- Interpretation:

The column grouped plot accentuates trends in terms of seasonality, with certain months (perhaps during the summer) having a larger volume of traffic. Trends such as these can inform planning for high-season use of resources.

2.3.6 Line Chart: Traffic Over Time

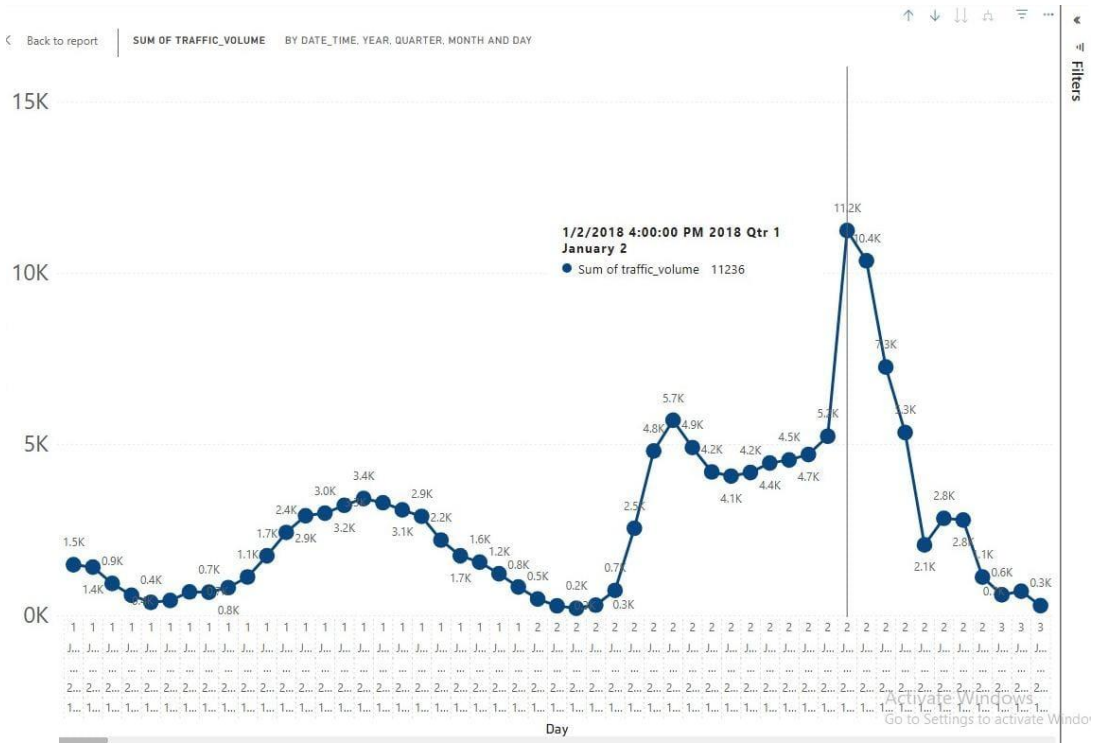


Figure 6: Line Chart : Volume by DATE_TIME,YEAR,QUARTER,MONTH AND DATE

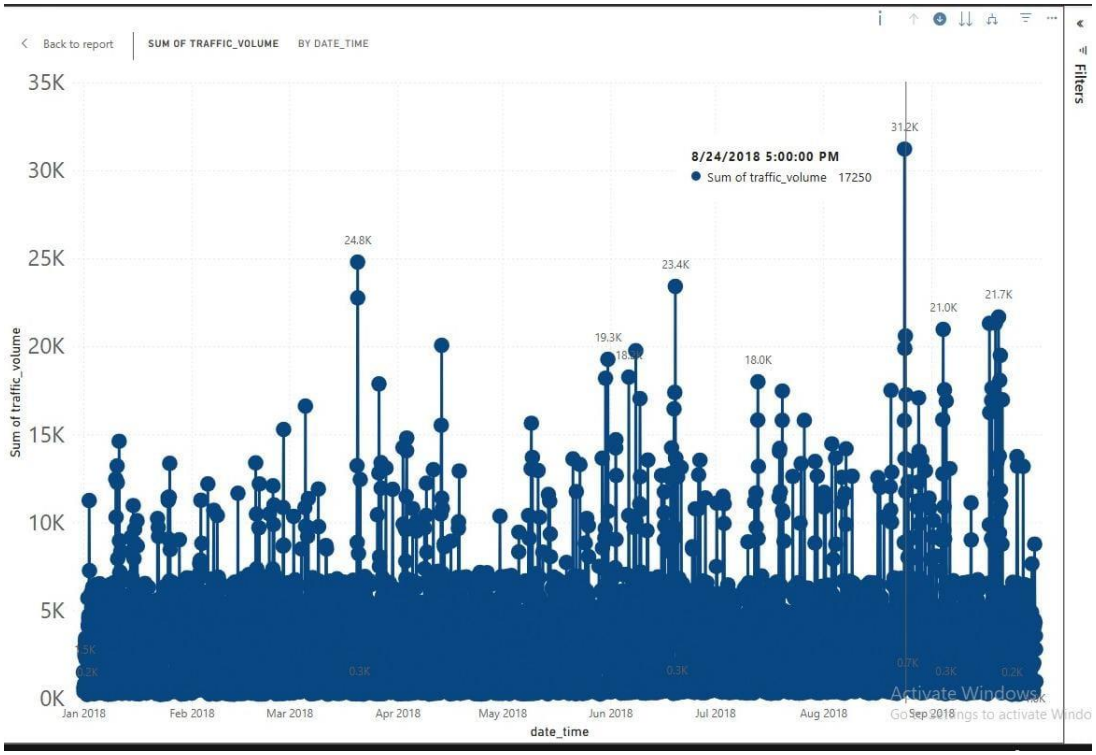


Figure 7: Line Chart : Volume by DATE_TIME

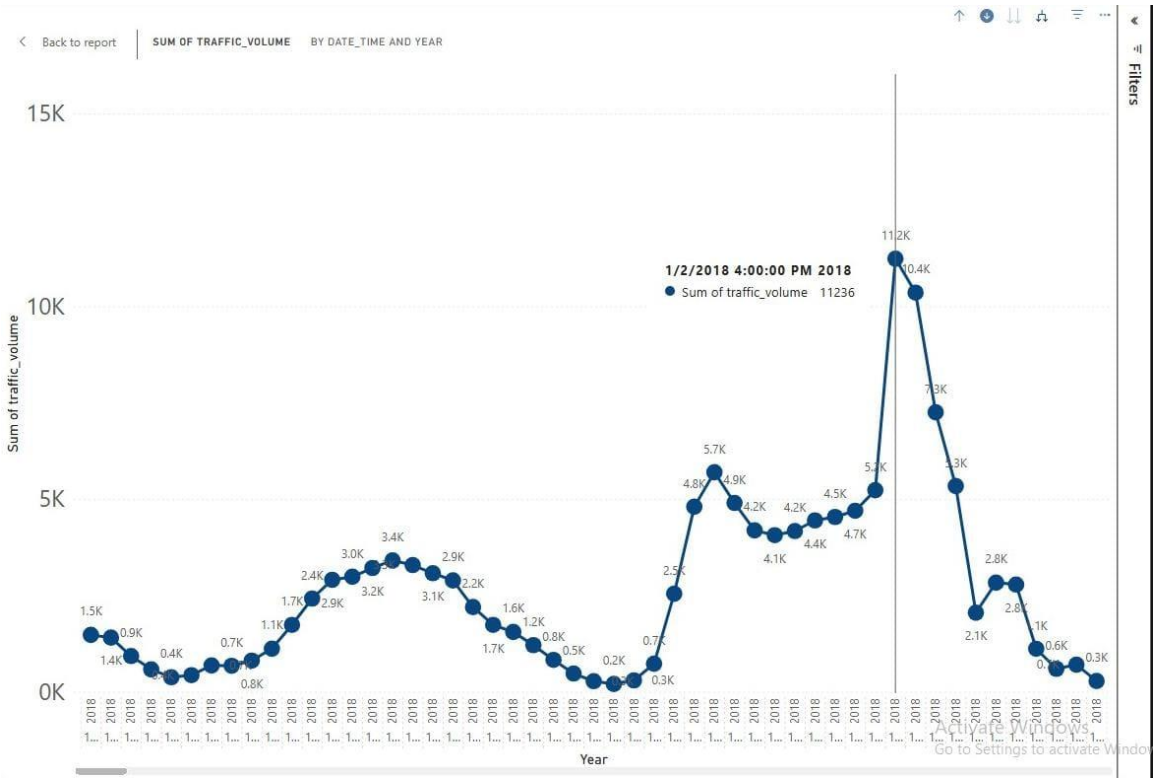


Figure 8: Line Chart : Volume by DATE_TIME AND YEAR

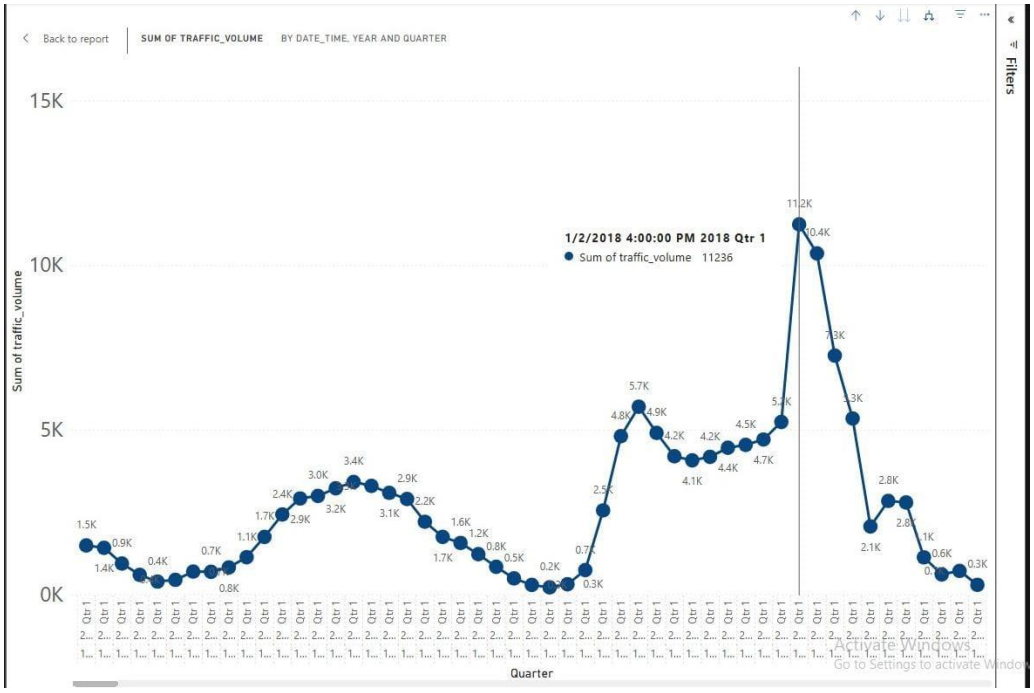


Figure 9: Line Chart : Volume by DATE_TIME,YEAR AND QUARTER

- Description:

  The line chart gives a continuous representation of traffic volume trends over a period of time, segmented by date/time (year, quarter, month, day).

- Interpretation:

  This visualization reveals both daily variation and long-term trends. Repeated high and low values represent predictable behavior in terms of traffic, and any outliers can represent one-time events or disruptions worth investigating.

2.3.7 Area Chart: Traffic Over Time

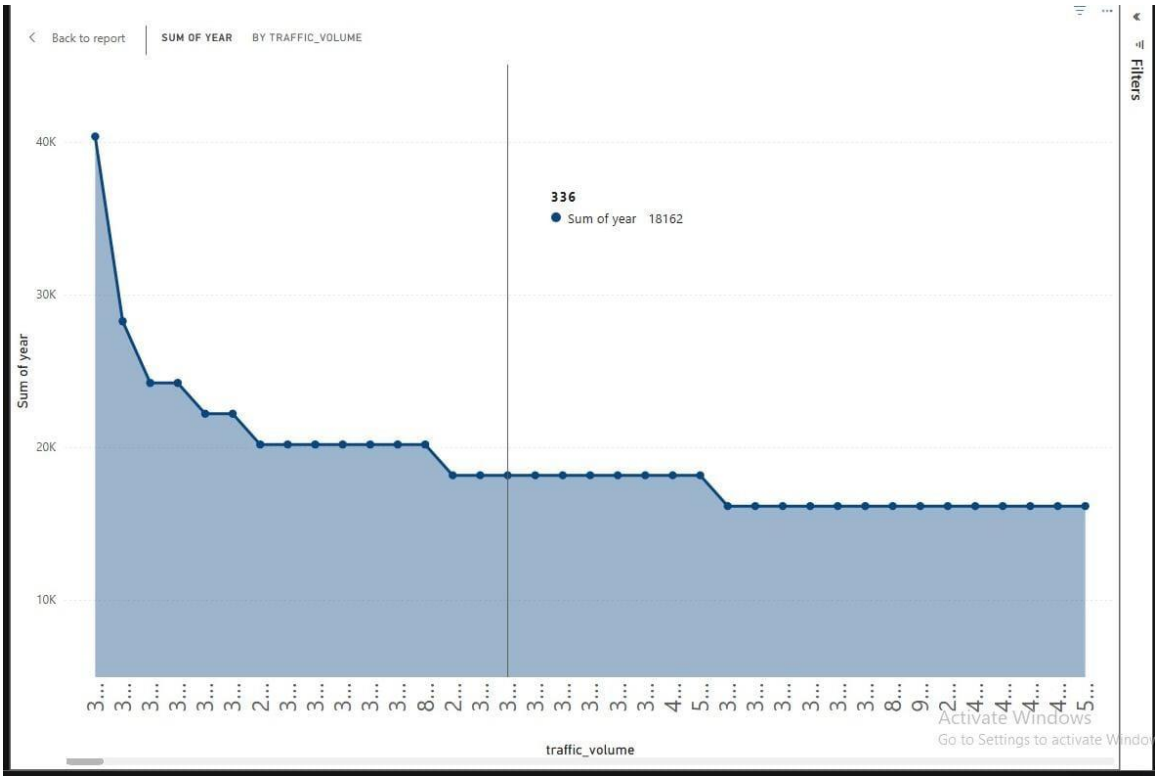


Figure 10: Area Chart: Traffic Over Time

- Description:

  The area chart illustrates cumulative traffic volume trends over the years. The X-axis shows the sum of traffic volume, and the Y-axis shows the year.

- Interpretation:

  The chart reflects an overall increase in volumes of traffic over years, with spikes in individual years being apparent. Trends of this kind can illustrate factors such as urban development, population growth, or a change in collection methodologies for traffic.

2.3.8 Tree Map: Traffic Volumes By Weather Description

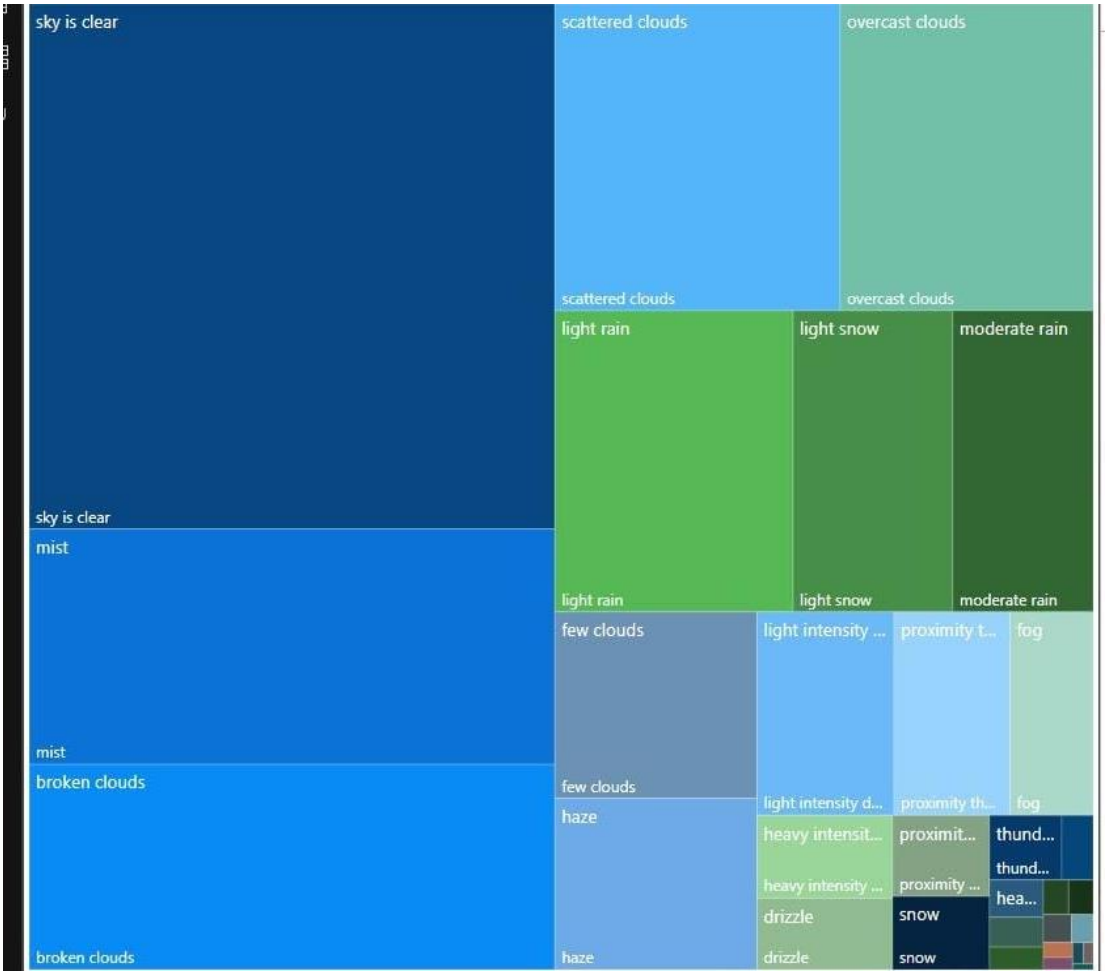


Figure 11: Traffic Volumes By Weather Description

- Description:

  This tree map illustrates the distribution of volumes of traffic over a range of weather descriptions. The proportion of each block is equivalent to the summed volumes of traffic for each weather state.

- Interpretation:

  The tree map clearly shows that largest volumes of traffic under ordinary weather conditions prevail. In contrast, weather including rain, and even snow, yield smaller blocks, and therefore less traffic under poor weather conditions.

2.3.9 Donut Charts: Traffic Volumes in Varying Categories

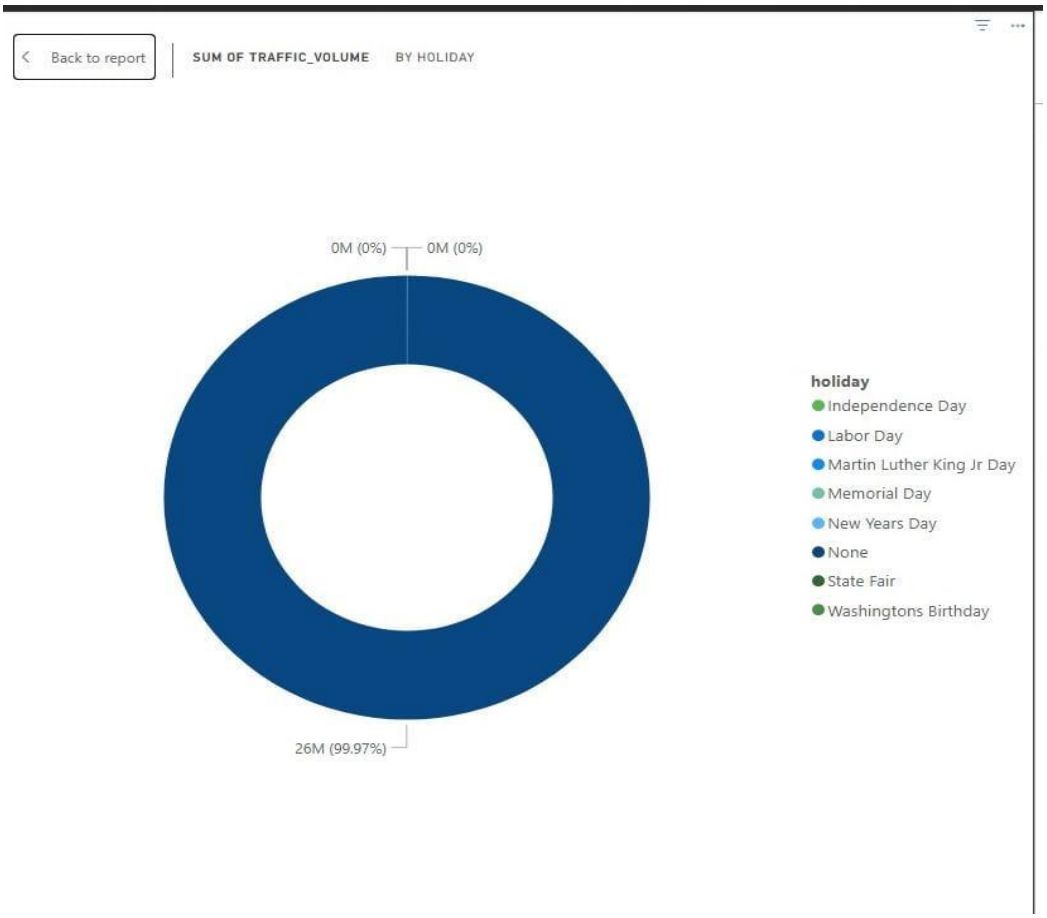


Figure 12: Chart 1: Holiday Traffic Volumes

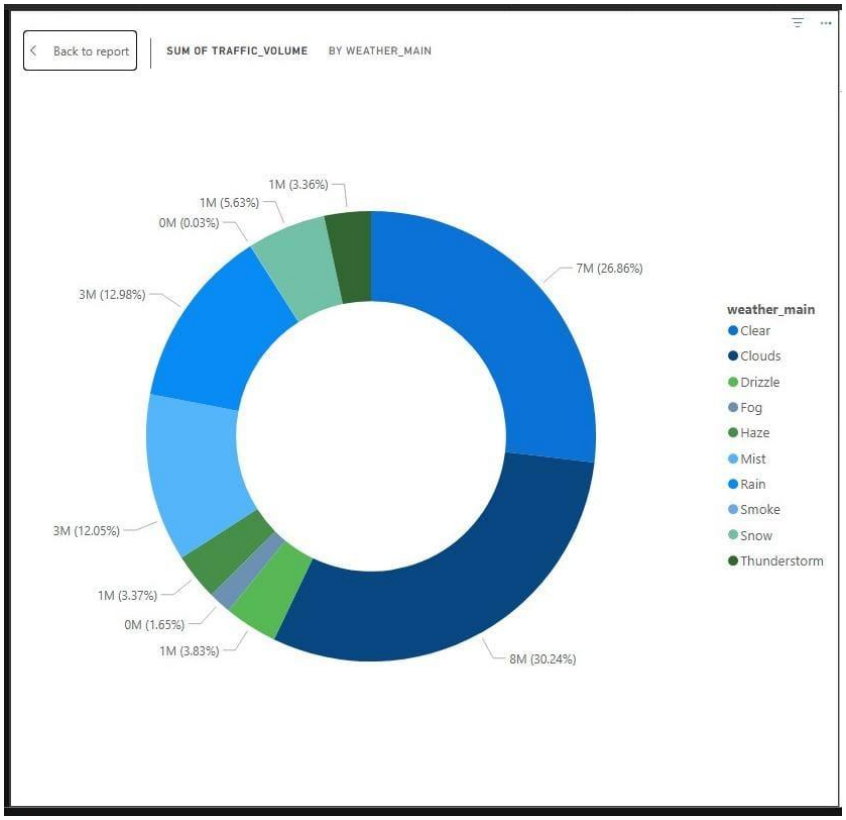


Figure 13: Chart 2: Traffic Volumes according to Weather Main

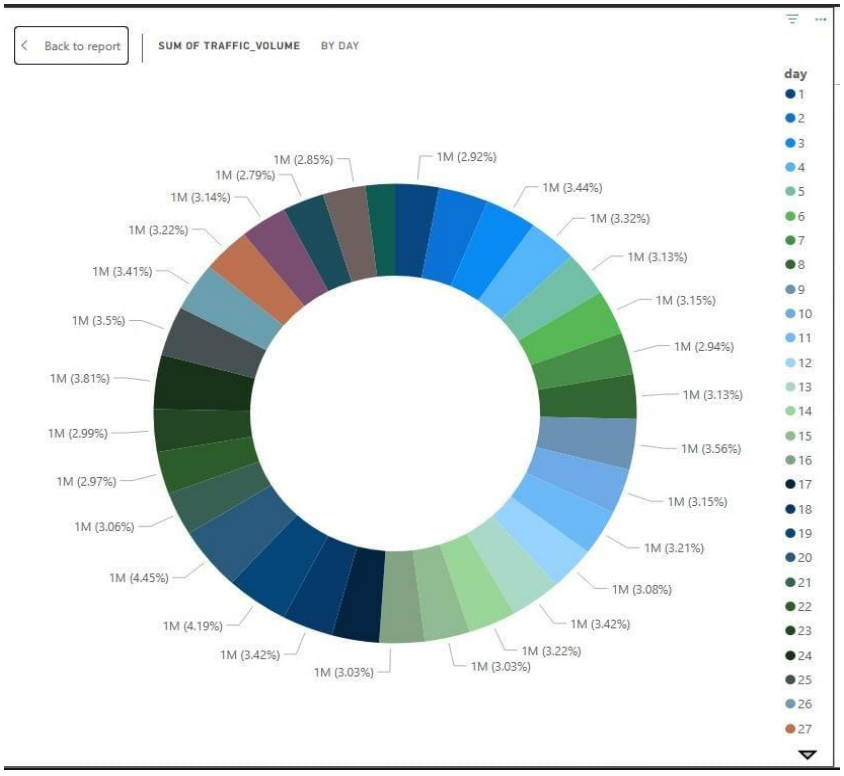


Figure 14: Traffic Volumes by Day

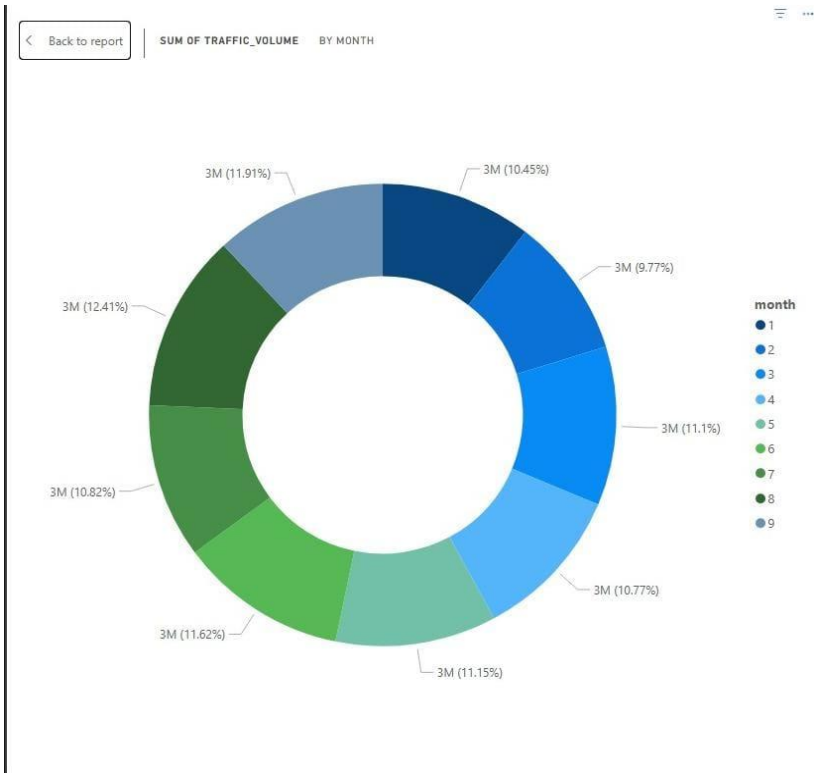


Figure 15: Chart 4: Traffic Volumes Per Month

- Description:

  Several donut charts are utilized to illustrate the proportion of traffic volumes for different categories:

  - Chart 1: Holiday Traffic Volumes
  - Chart 2: Traffic Volumes according to Weather Main
  - Chart 3: Traffic Volumes by Day
  - Chart 4: Traffic Volumes Per Month

- Interpretation:

  All of the donuts represent the proportional make-up of overall volume of traffic in its respective category. They reveal, for instance, that off holidays dominate overall traffic, specific weather has a significant impact on traffic, and trends during days/months represent an additional analysis of usage behavior.

2.3.10 Pie Chart: Traffic Volume By Weather

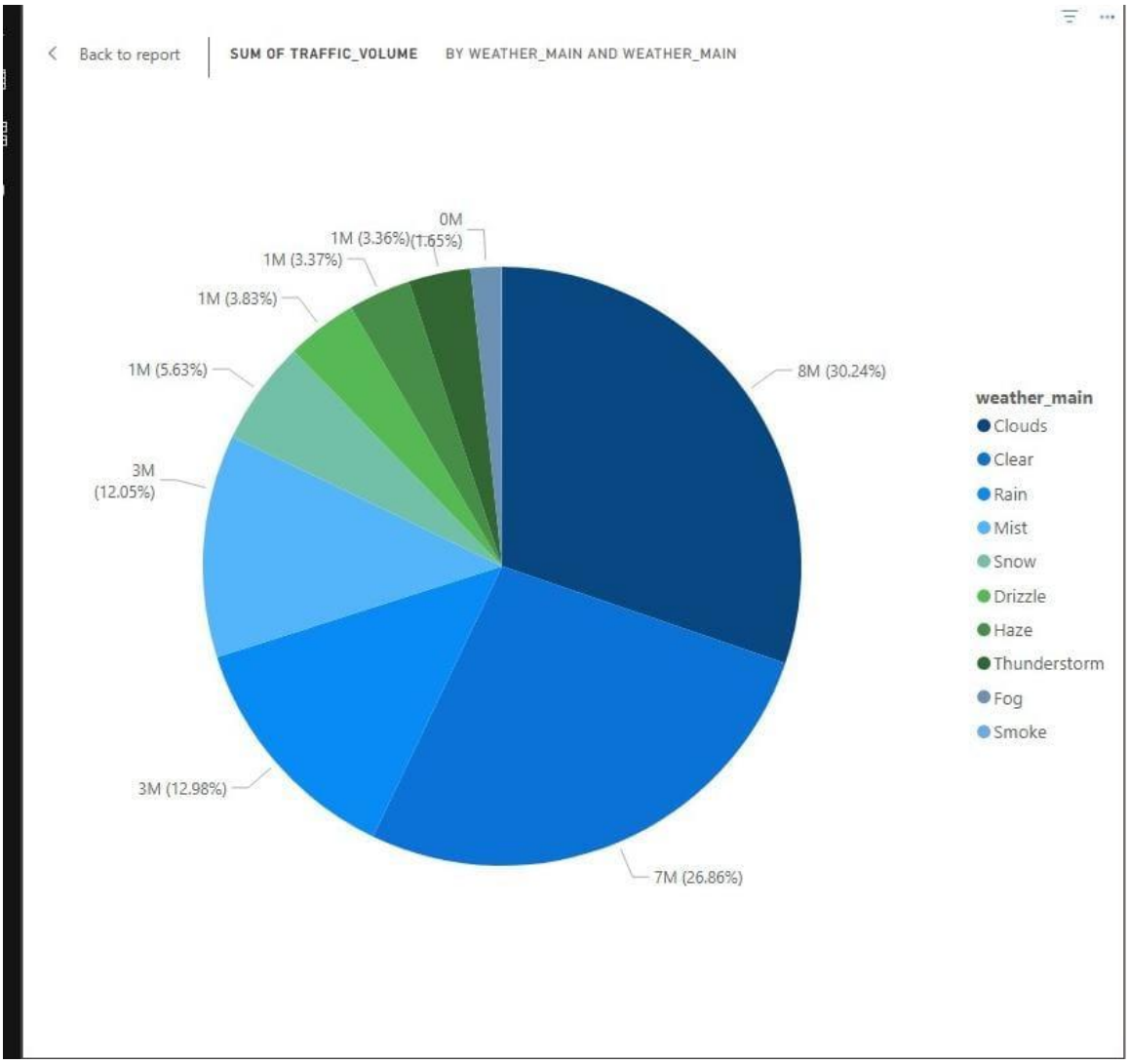


Figure 16: Pie Chart: Traffic Volume By Weather

- Description:

  This pie chart illustrates distribution of traffic volumes according to principal weather conditions (weather main). Legend separates various weather conditions, and each portion's proportion corresponds to overall traffic volume for a respective category.

- Interpretation:

  The pie chart successfully communicates which types of weather (e.g., cloudiness, clear) are most contributing to traffic volumes. Larger wedges for cloudiness and clear communicate that favorable weather is accountable for higher levels of traffic.

2.3.11 Slicer with Dropdown: Date and Time Filtering

Figure 17: Slicer with Dropdown

- Description:

  The interactive slicer allows you to filter the dashboard by each individual date part (year, month, day).

- Interpretation:

  By enabling filtering, this slicer allows for focused analysis. Users can target specific timeframes, and in doing so, gain a more specific view of temporal traffic trends.

2.3.12 Gauge: Current Traffic Volumes

Figure 18: Gauge: Current Traffic Volumes

- Description:

  The gauge graphical view presents current volumes of traffic in terms of predefined targets and thresholds.

- Interpretation:

  This visualization can make an immediate determination of whether present volumes of traffic fall in forecasted ranges, providing for quick decision-making in response to volumes of traffic.

2.3.13 Cards: Displaying Key Information

Figure 19: CARD date/time

Figure 19: CARD holiday information

Figure 21: CARD weather_main

Figure 22: CARD: weather description

Figure 23:CARD total traffic volume

- Description:

  Cards are utilized to present key information points, such as current date/time, holiday information, weather main, weather detail, and total traffic volume.

- Interpretation:

  The use of cards aids in providing key statistics at a glance for the observer. The cards serve as a quick-reference tool that complements the deeper charts and graphs.

2.4 Full Dashboard Image

Figure 20: 2.4 Full Dashboard Image

- Full Dashboard Description:

  The full dashboard integrates all of the aforementioned visualizations in one coherent interface. It shows an overall picture of the traffic volume information through a mix of temporal, weather, and holiday analysis in one format. Interactive elements such as drop-downs and slicers allow for information filtering dynamically, and thus, for customizability in terms of individual insights to be highlighted in the dashboard. The dashboard is constructed for ease of use, with each visualization labelled appropriately and timed to deliver an integrated analysis platform. Overall view allows for quick assessments and sound decision-making through an aggregated picture of trends in traffic and driving factors.

Chapter 3: Selection of Data Mining Algorithm and Data Preprocessing

3.1 Overview

Traffic volume prediction is important for urban planning, traffic management, and reducing congestion. In this work, Metro Interstate Traffic Volume is analyzed using Weka. Analysis consists of preprocessing, clustering, classification, feature selection, and visualization for insightful and meaningful extraction of patterns and trends.

3.2 Data Preprocessing

3.2.1 Overview

Data preprocessing is an important part of data analysis and machine learning for ensuring data usability, uniformity, and quality. Missing values, outliers, and duplicates in raw data must be handled in preparation for using analysis techniques.

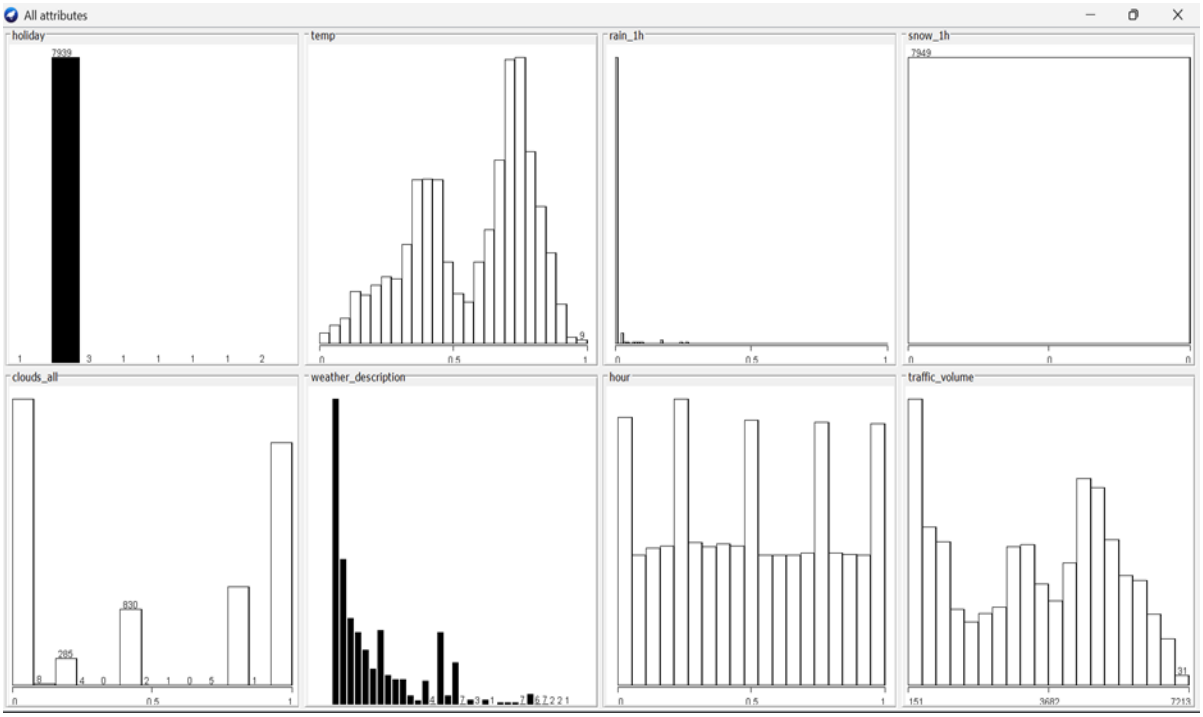


Figure 21:Pre-Process visualization

3.2.2 Steps Adopted

- Handling Missing Value: Missing values have been replaced with mean/mode substitution in order to preserve data integrity.
- Data Normalization: All the numerical features (i.e., temp, rain_1h, snow_1h, clouds_all, and traffic_volume) are normalized between 0 and 1 for equal weightage and for the better performance of the model.
- Attribute Removal: Unrelated attributes, including timestamps and unnecessary features, have been removed for enhancing model efficiency.
- Data Type Conversion: Categorial values have been converted to numerical values using encoding techniques for compatibility with requirements for a machine learning model.

3.3 Clustering

3.3.1 Overview

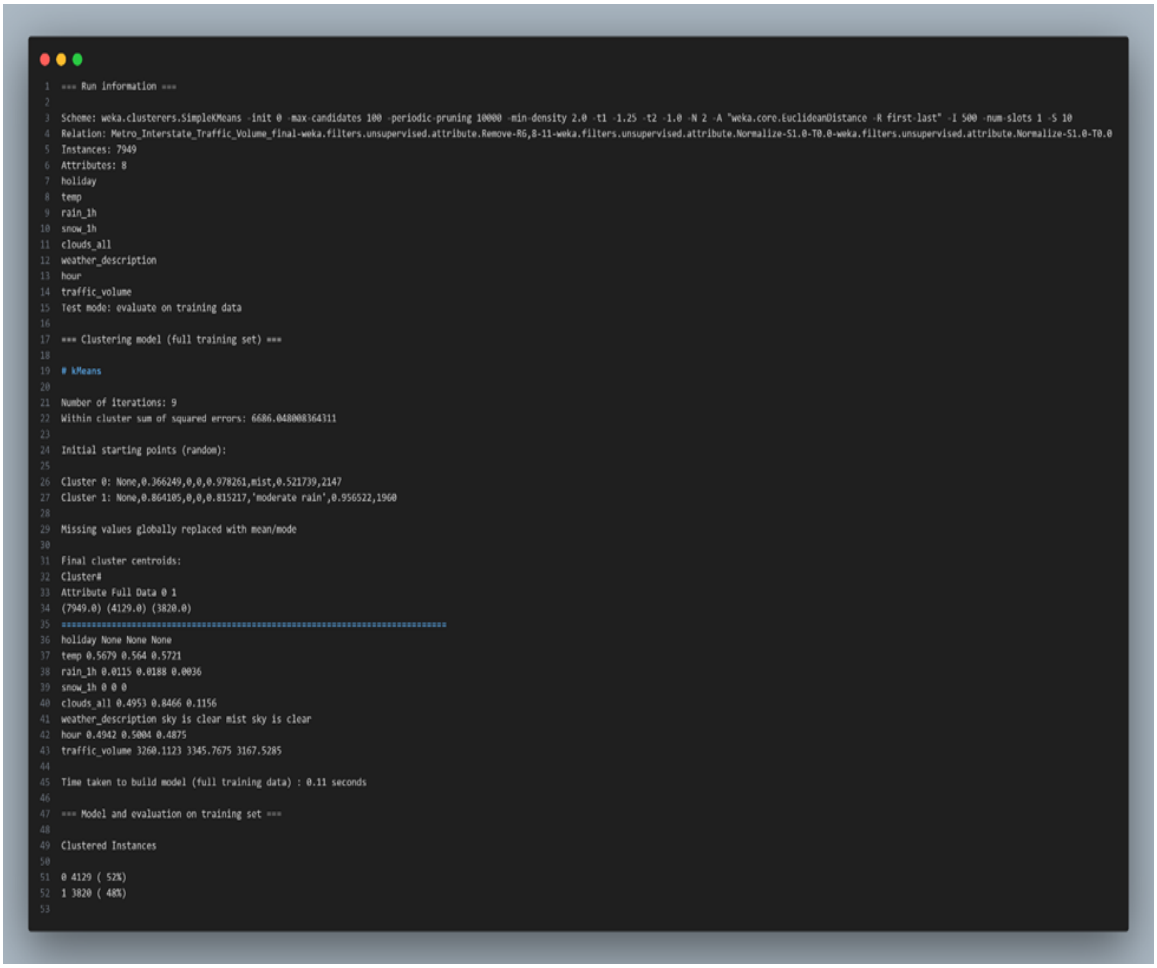
Clustering is an unsupervised learning technique used to group similar data points based on common patterns. In this project, we used K-Means Clustering to segment the traffic data.

3.3.2 Steps Followed

- Algorithm Selection: The Simple K-Means algorithm was chosen with K=2 clusters.
- Distance Measure: Euclidean Distance was used to measure similarity.
- Cluster Evaluation:
    - The number of iterations required to converge was 9.
    - The within-cluster sum of squared errors (WCSS) was 6686.04.
    - The two clusters identified different traffic volume patterns based on attributes like weather_description, hour, and clouds_all.

3.3.3 Interpretation

- Cluster 0: Represents traffic conditions during clear sky conditions with high traffic volume.
- Cluster 1: Represents traffic conditions during rainy/misty conditions with slightly lower traffic volume.



```
=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation: Metro_Interstate_Traffic_Volume_final-weka.filters.unsupervised.attribute.Remove-R6,8-11-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0
Instances: 7949
Attributes: 8
holiday
temp
rain_1h
snow_1h
clouds_all
weather_description
hour
traffic_volume
Test mode: evaluate on training data

=== Clustering model (full training set) ===

# kMeans

Number of iterations: 9
Within cluster sum of squared errors: 6686.048008364311

Initial starting points (random):

Cluster 0: None,0.366249,0,0,0.978261,mist,0.521739,2147
Cluster 1: None,0.864105,0,0,0.815217,'moderate rain',0.956522,1960

Missing values globally replaced with mean/mode

Final cluster centroids:
Cluster#
Attribute Full Data 0 1
(7949.0) (4129.0) (3820.0)
==========================================================================
holiday None None None
temp 0.5679 0.564 0.5721
rain_1h 0.0115 0.0188 0.0036
snow_1h 0 0 0
clouds_all 0.4953 0.8466 0.1156
weather_description sky is clear mist sky is clear
hour 0.4942 0.5004 0.4875
traffic_volume 3260.1123 3345.7675 3167.5285

Time taken to build model (full training data) : 0.11 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 4129 ( 52%)
1 3820 ( 48%)
```

Figure 22: Clustering

3.4 Classification

3.4.1 Overview

Classification is a supervised learning approach to forecasting traffic volume based on more than one attribute.

3.4.2 Algorithm selection

- We employed Linear Regression to predict traffic volumes.

3.4.3 Model training

- 10-Fold Cross-Validation was used to evaluate model performance.
- The correlation coefficient of 0.3939 reflects a moderate relationship between input factors and traffic volumes.

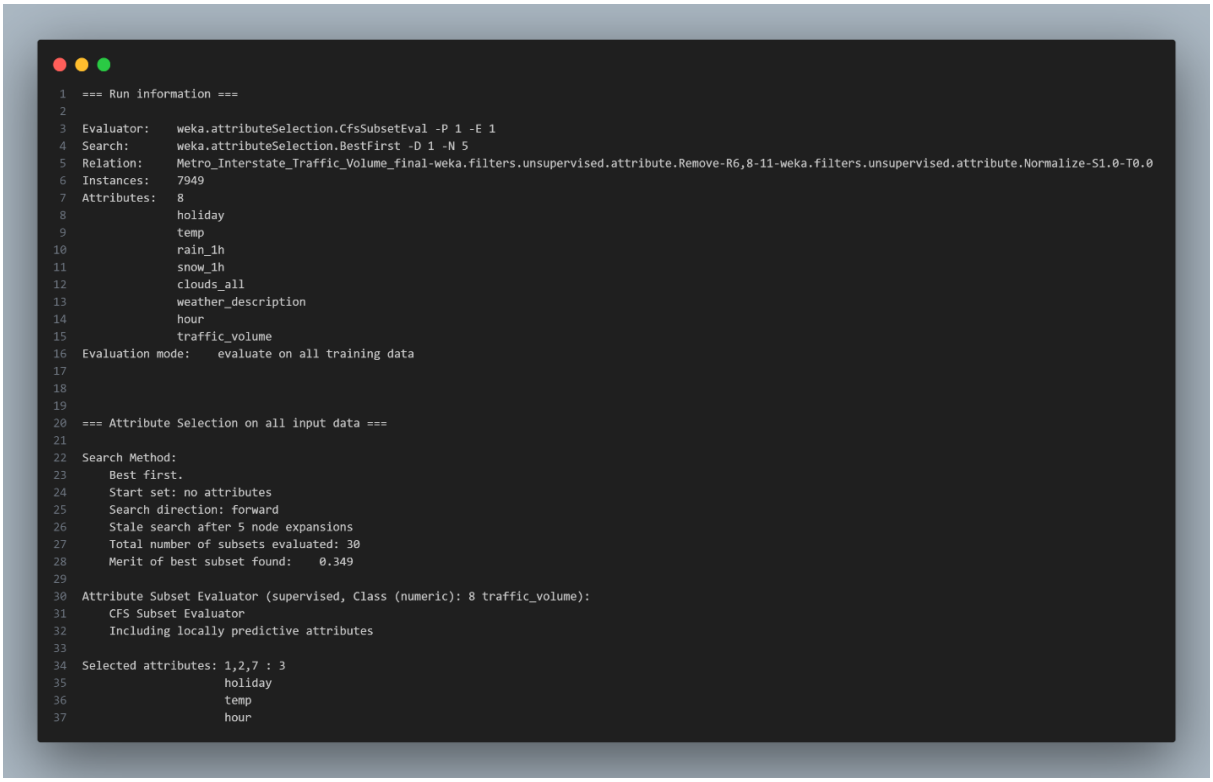
- The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values were 1587.18 and 1813.54, respectively. 4.4 Interpretation
- The model postulates that traffic count is a function of factors including holidays, weather, and hour. The presence of holidays significantly impacts traffic volume, increasing congestion on special days.

Figure 23: Classification

3.5 Attribute Selection

3.5.1 Overview

Attribute selection helps in identifying the most important features that contribute to traffic volume prediction.



```
=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     Metro_Interstate_Traffic_Volume_final-weka.filters.unsupervised.attribute.Remove-R6,8-11-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0
Instances:    7949
Attributes:   8
              holiday
              temp
              rain_1h
              snow_1h
              clouds_all
              weather_description
              hour
              traffic_volume
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
    Best first.
    Start set: no attributes
    Search direction: forward
    Stale search after 5 node expansions
    Total number of subsets evaluated: 30
    Merit of best subset found:    0.349

Attribute Subset Evaluator (supervised, Class (numeric): 8 traffic_volume):
    CFS Subset Evaluator
    Including locally predictive attributes

Selected attributes: 1,2,7 : 3
                     holiday
                     temp
                     hour
```

Figure 24:Selection

3.5.2 Method Used

- Evaluator: CfsSubsetEval (Correlation-based Feature Selection)
- Search Method: BestFirst (Forward selection)
- Selected Attributes:
    - Holiday
    - Temp
    - hour

3.5.3 Interpretation

These three attributes were identified as the most relevant for traffic volume prediction. Removing irrelevant features improved computational efficiency while retaining predictive power.

3.6. Data Visualization

3.6.1 Overview

Data visualization helps in understanding patterns and relationships among variables.

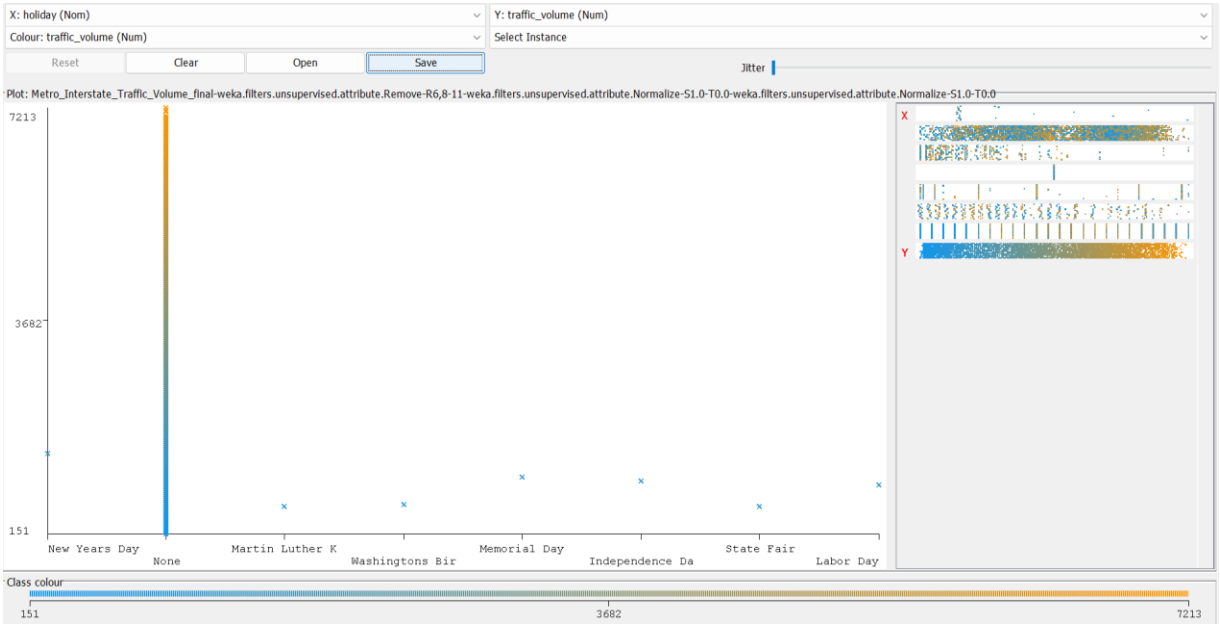


Figure 25: Visual 01

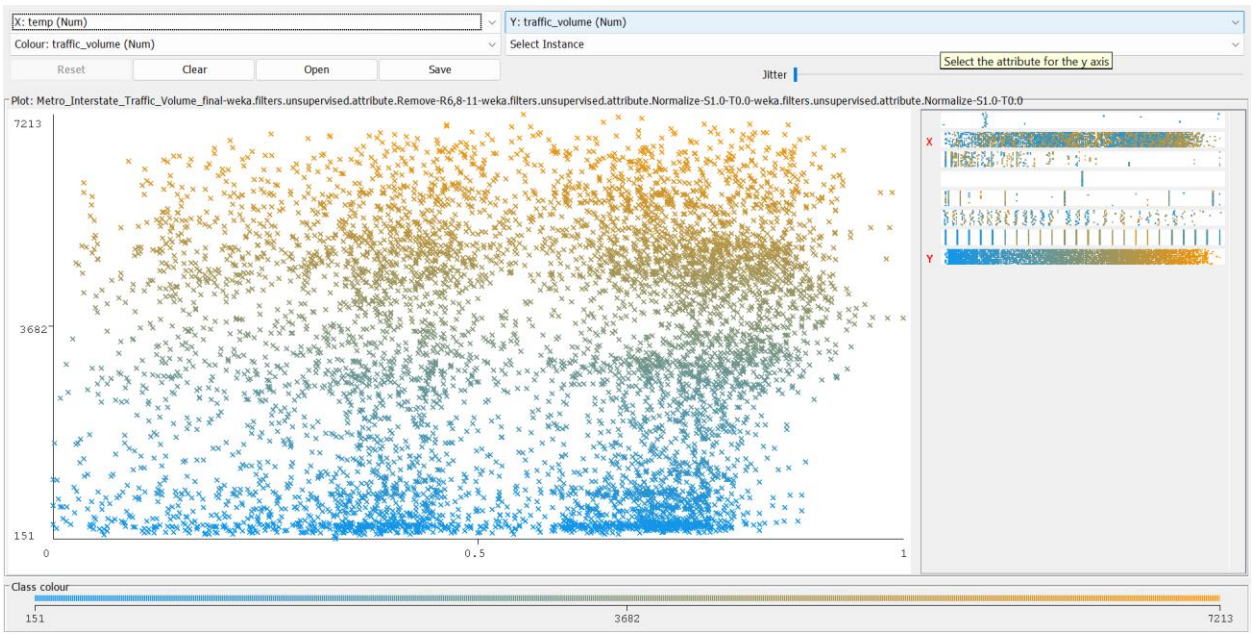


Figure 26; Visual 02

Figure 27: Visual 03

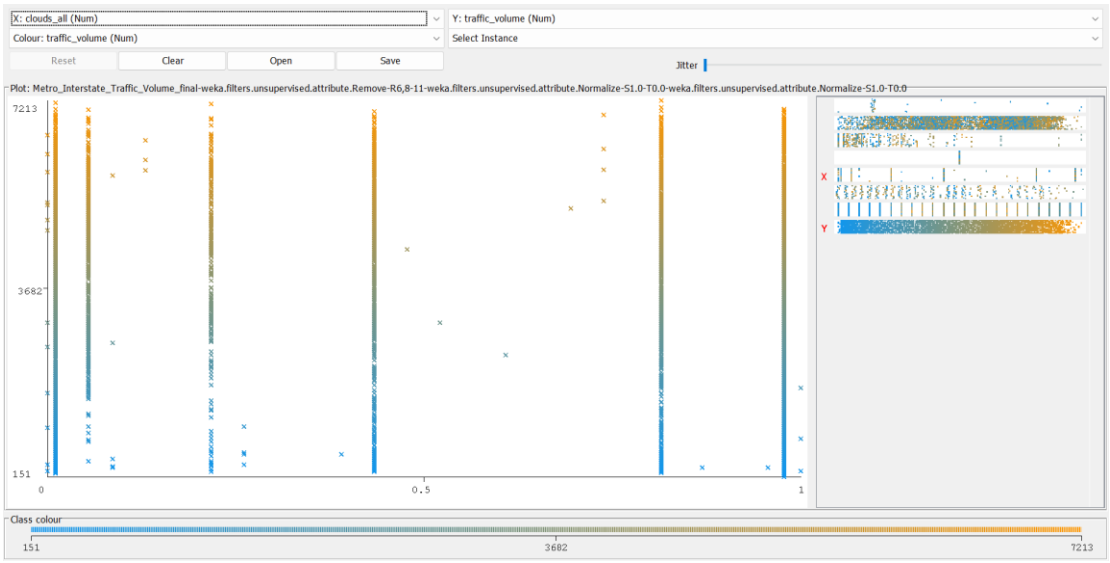


Figure 28: Visual 04

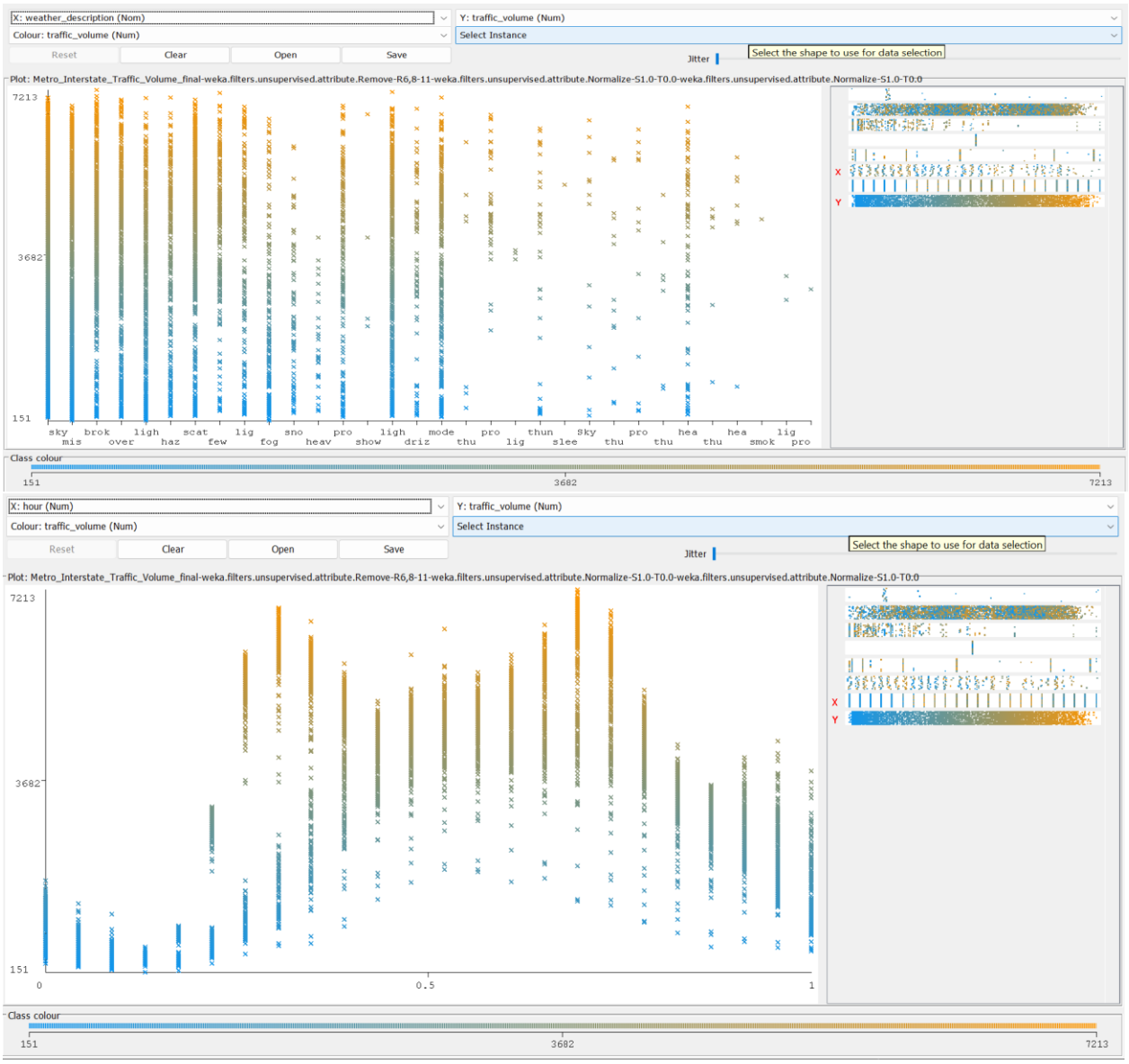


Figure 29: Visual 05 & 06

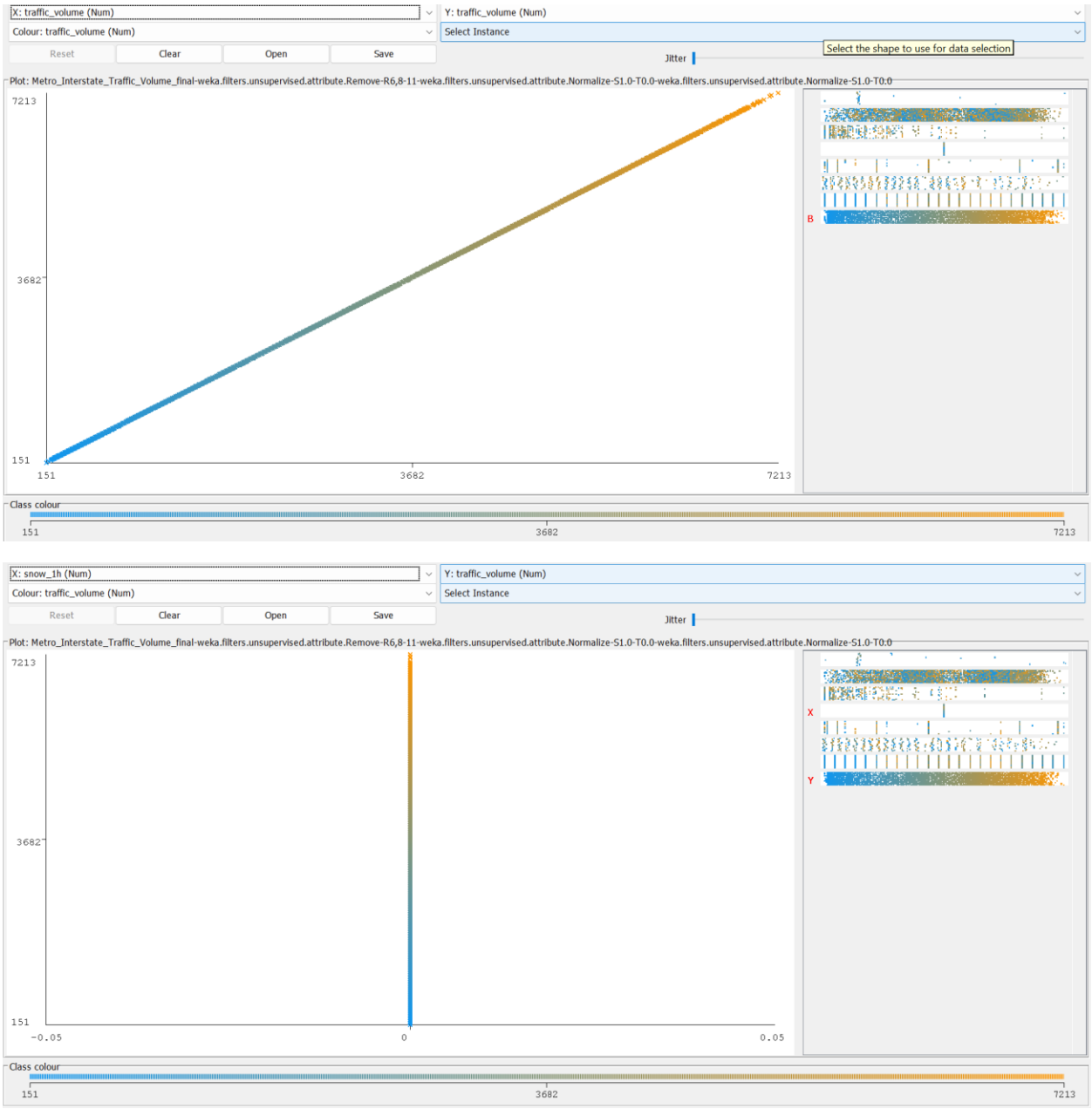


Figure 30: Visual 07 & 08

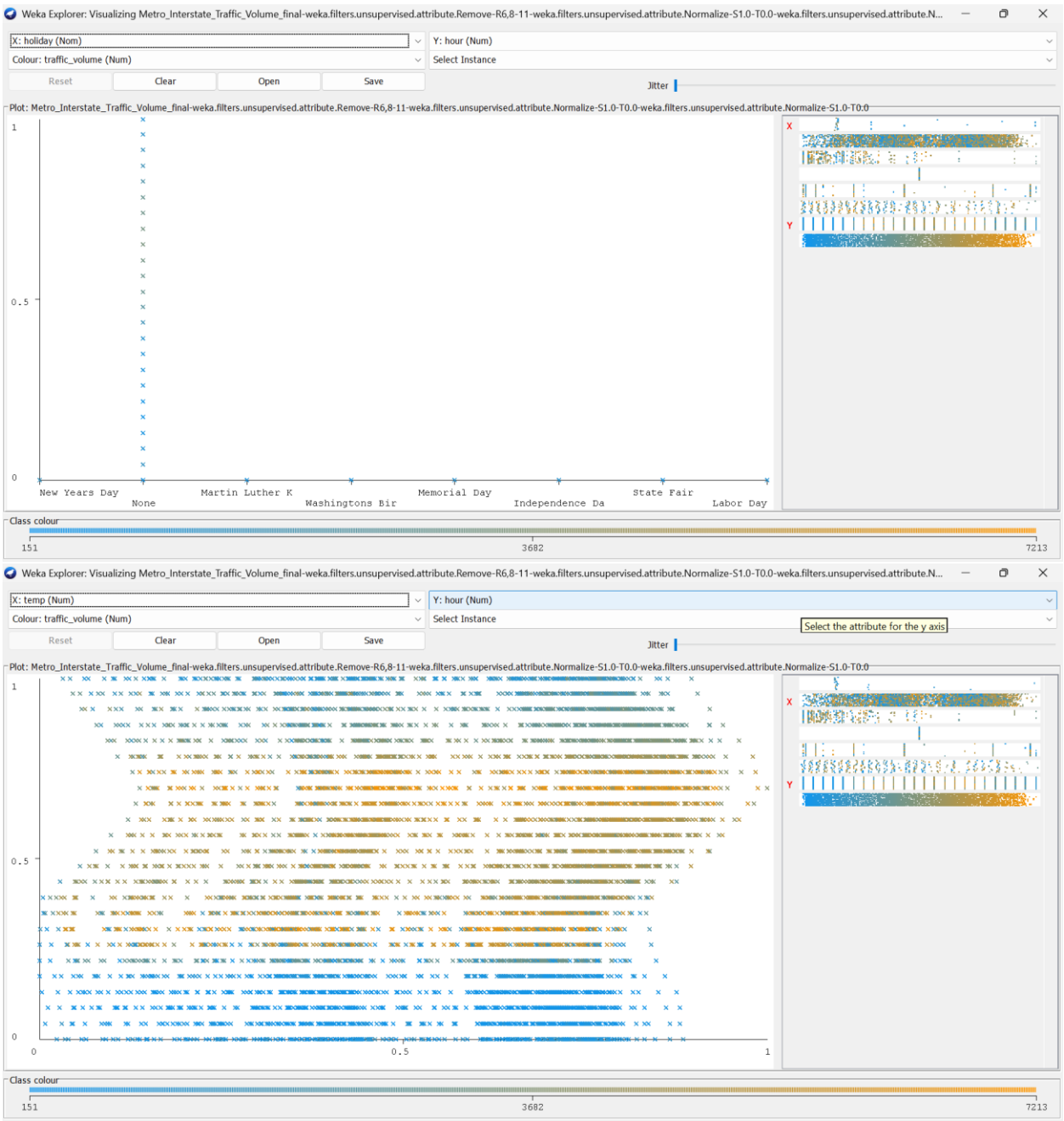


Figure 31: Visual 09 & 10

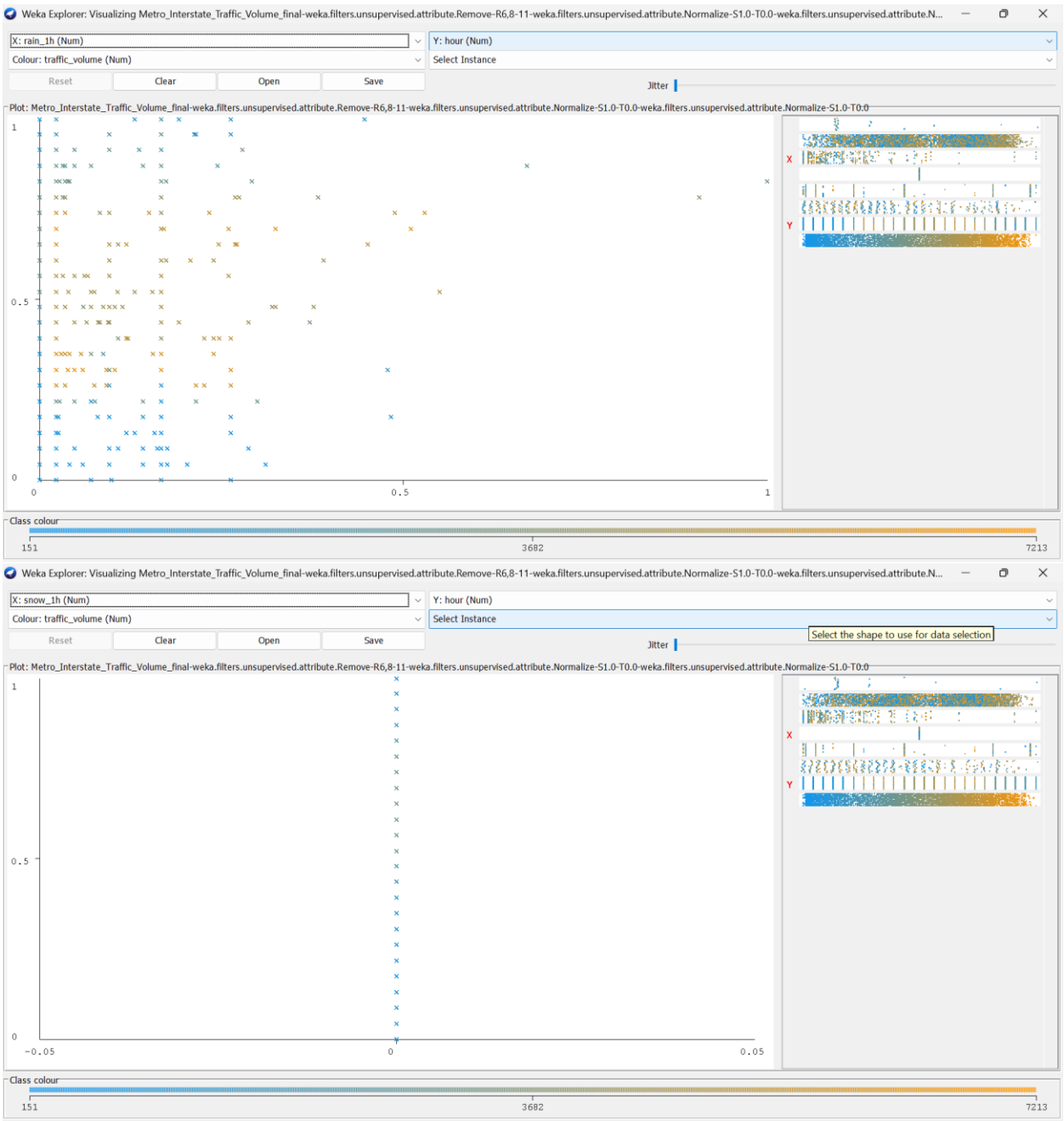


Figure 32: Visual 11 & 12

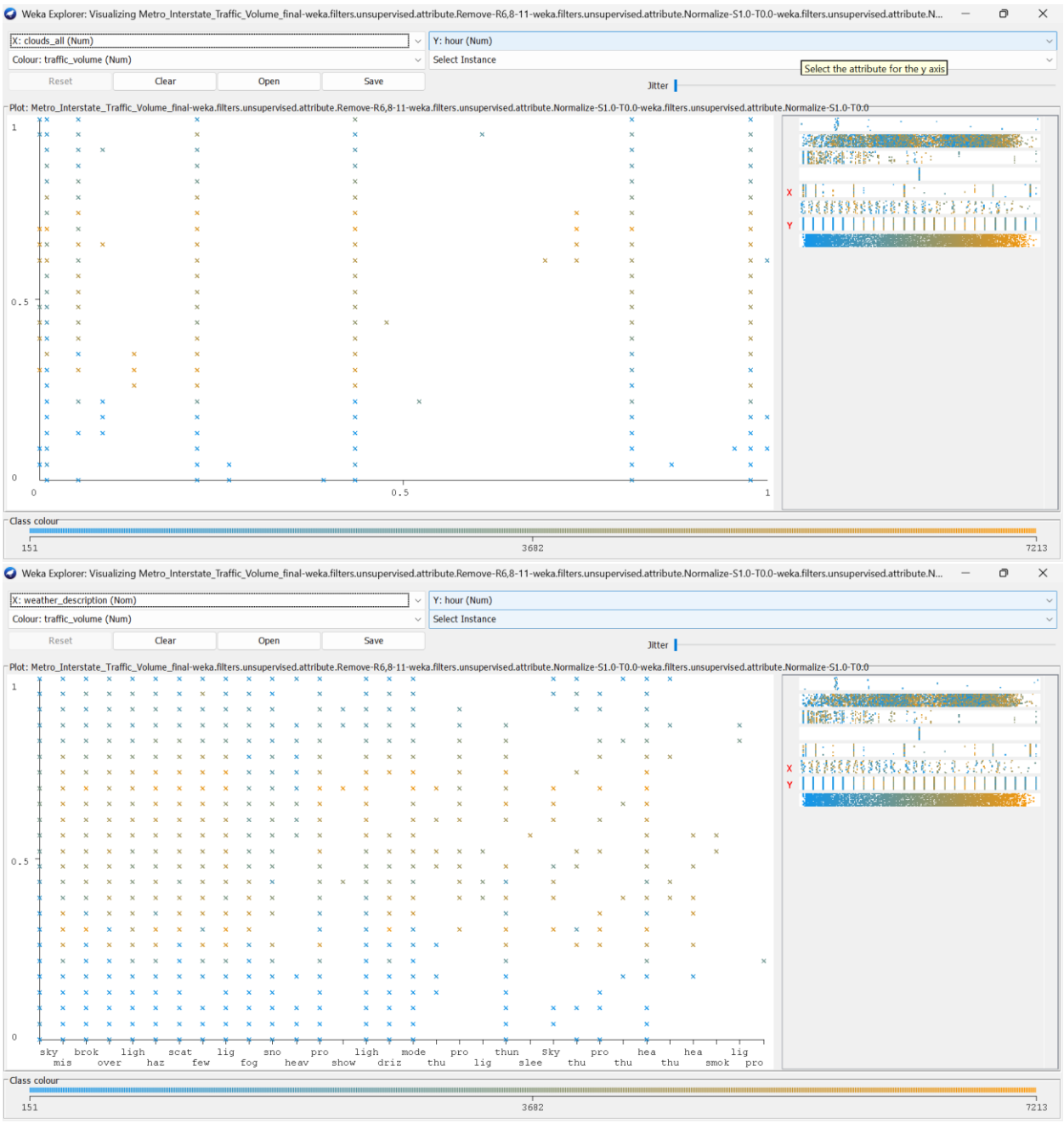


Figure 33: Visual 13 & 14

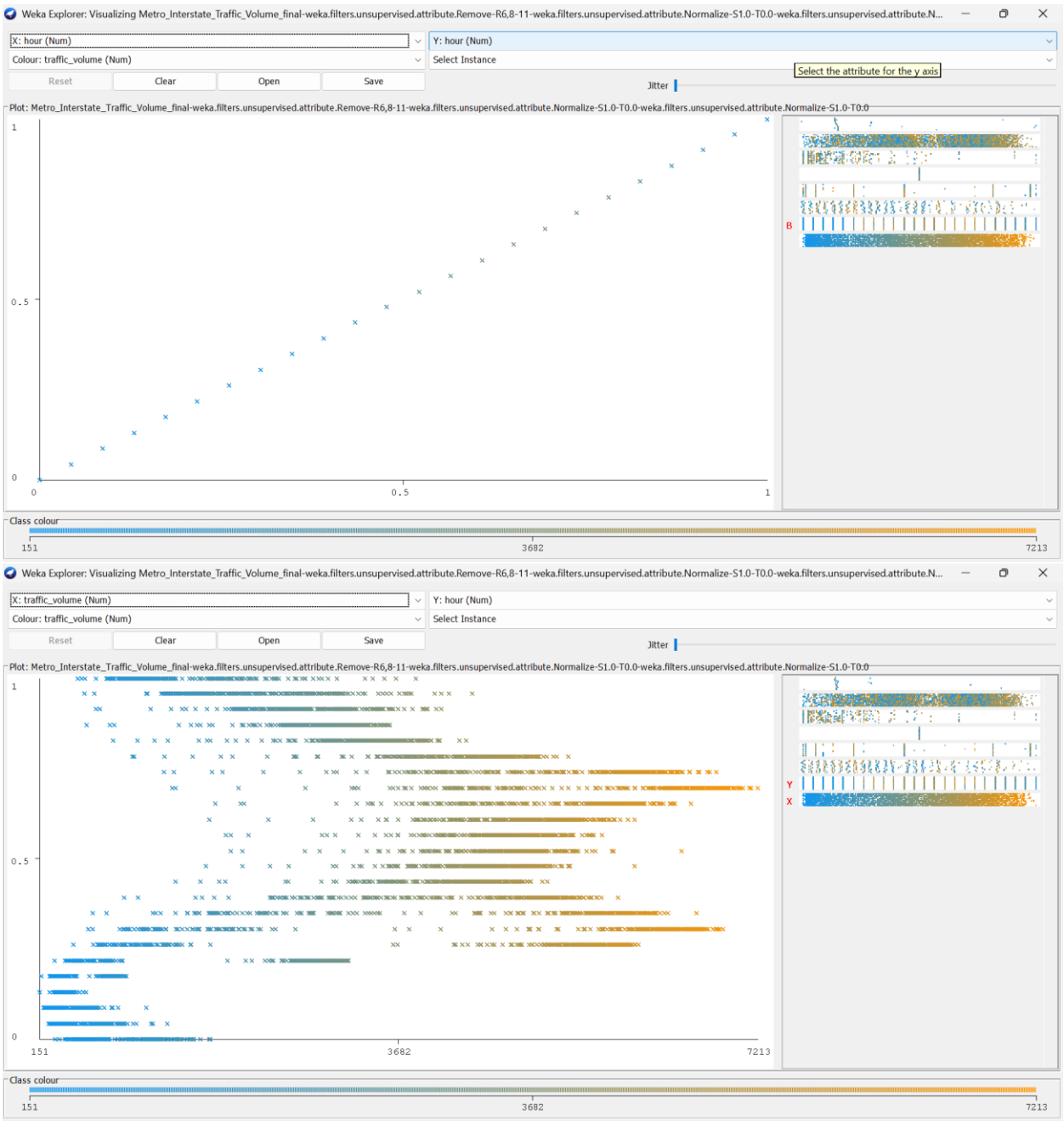


Figure 34: Visual 15 & 16

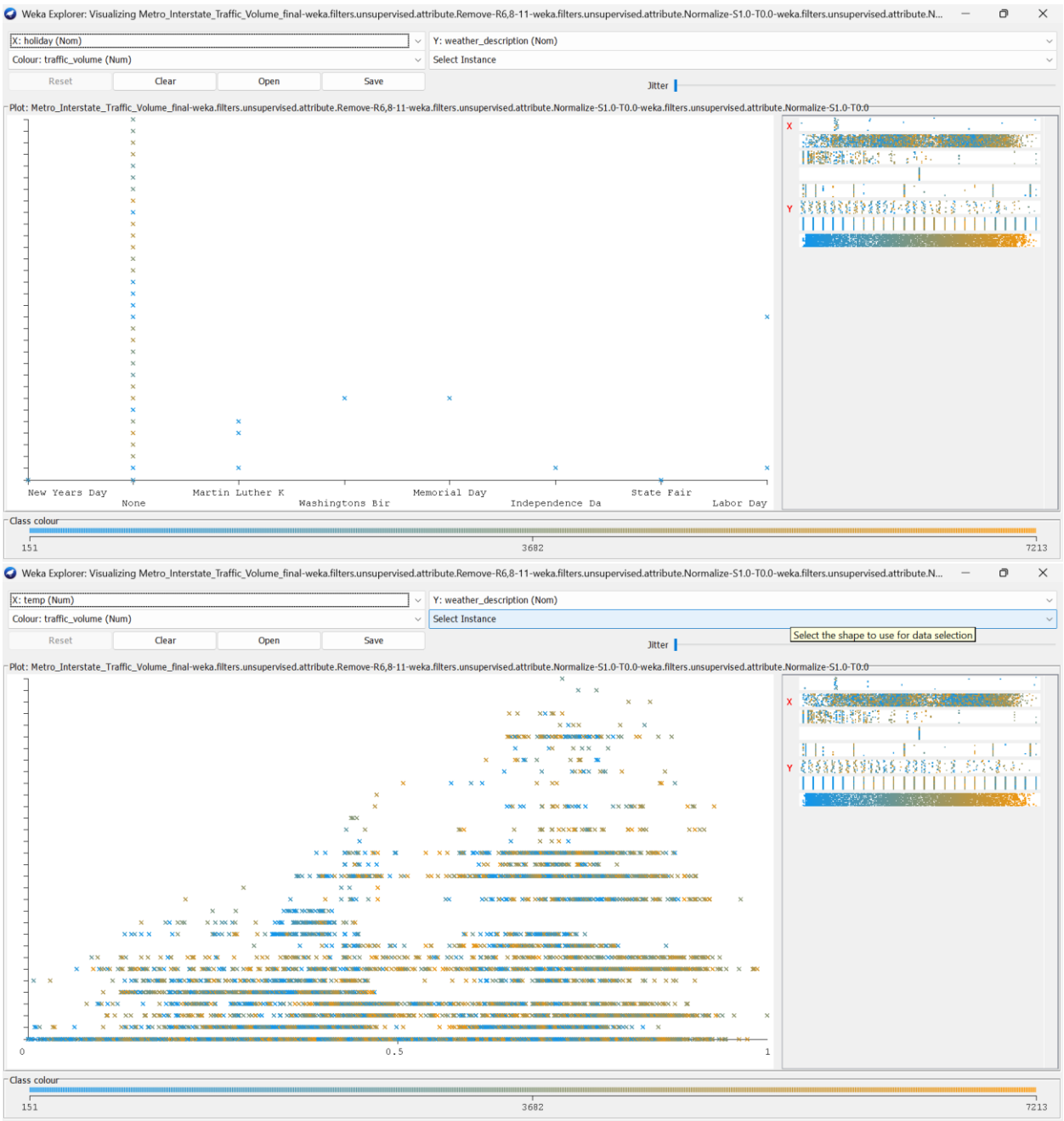


Figure 35: Visual 17 & 18

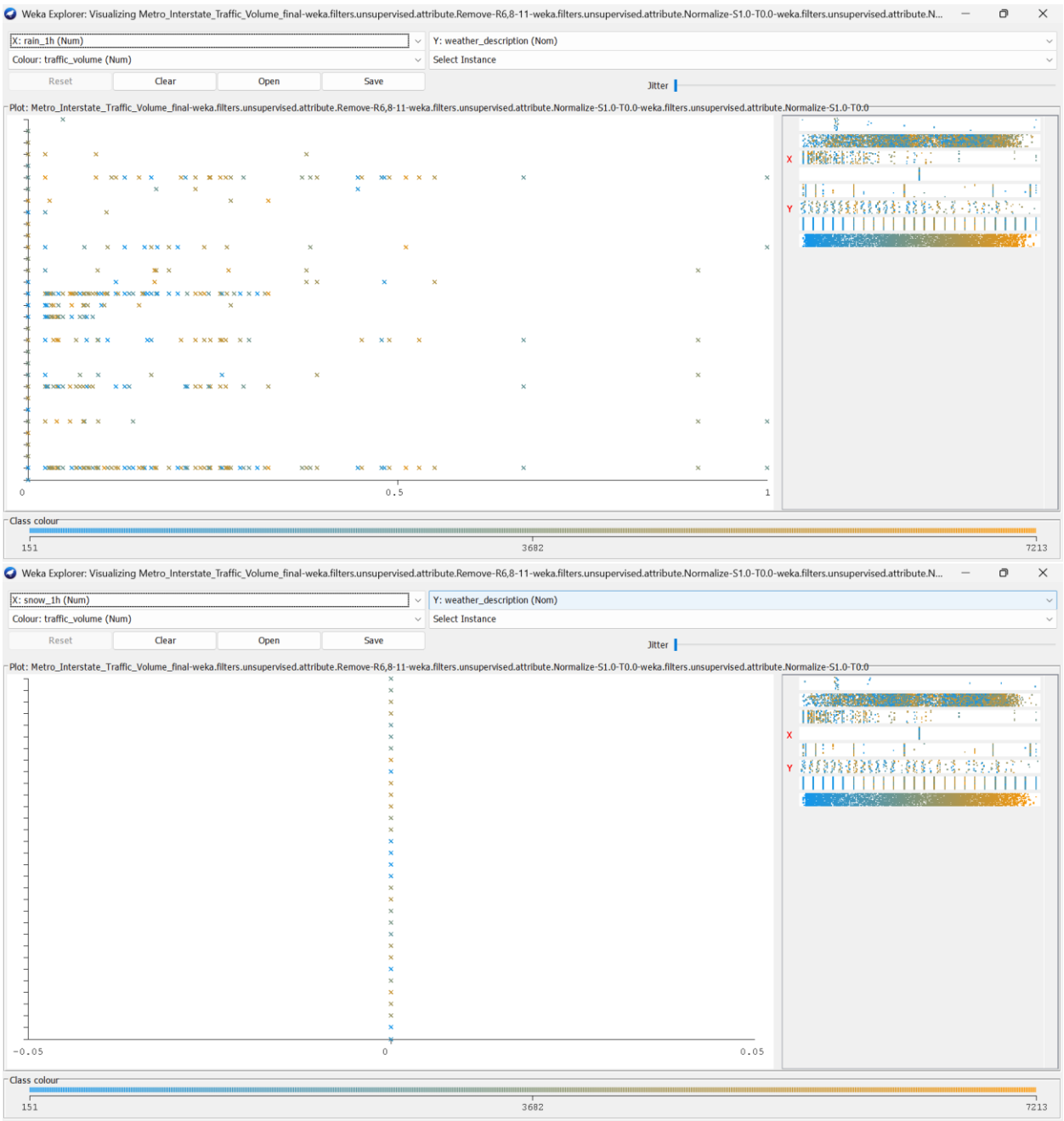


Figure 36: Visual 19 & 20

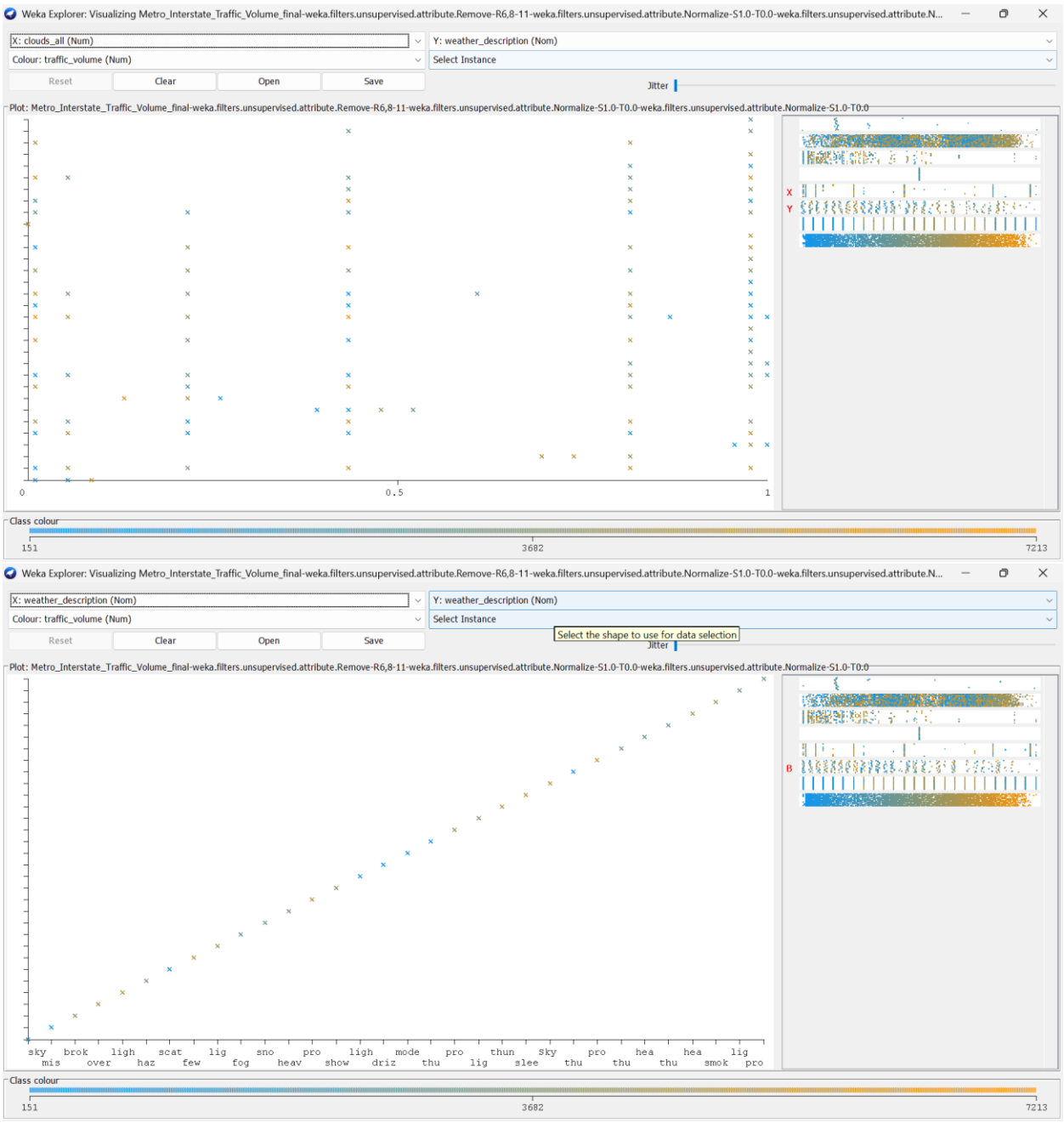


Figure 37: Visual 20 & 21

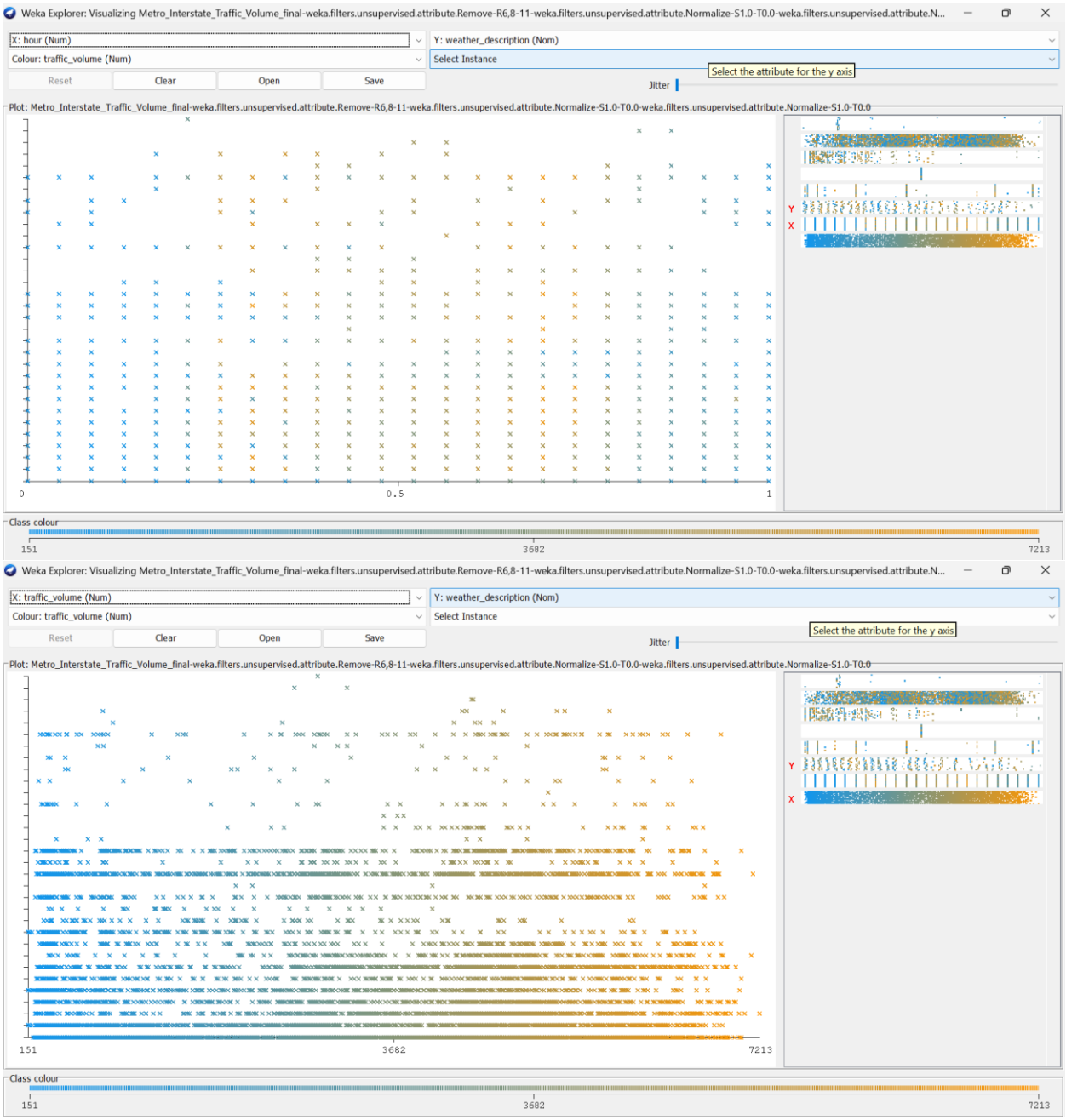


Figure 38: Visual 22 & 23

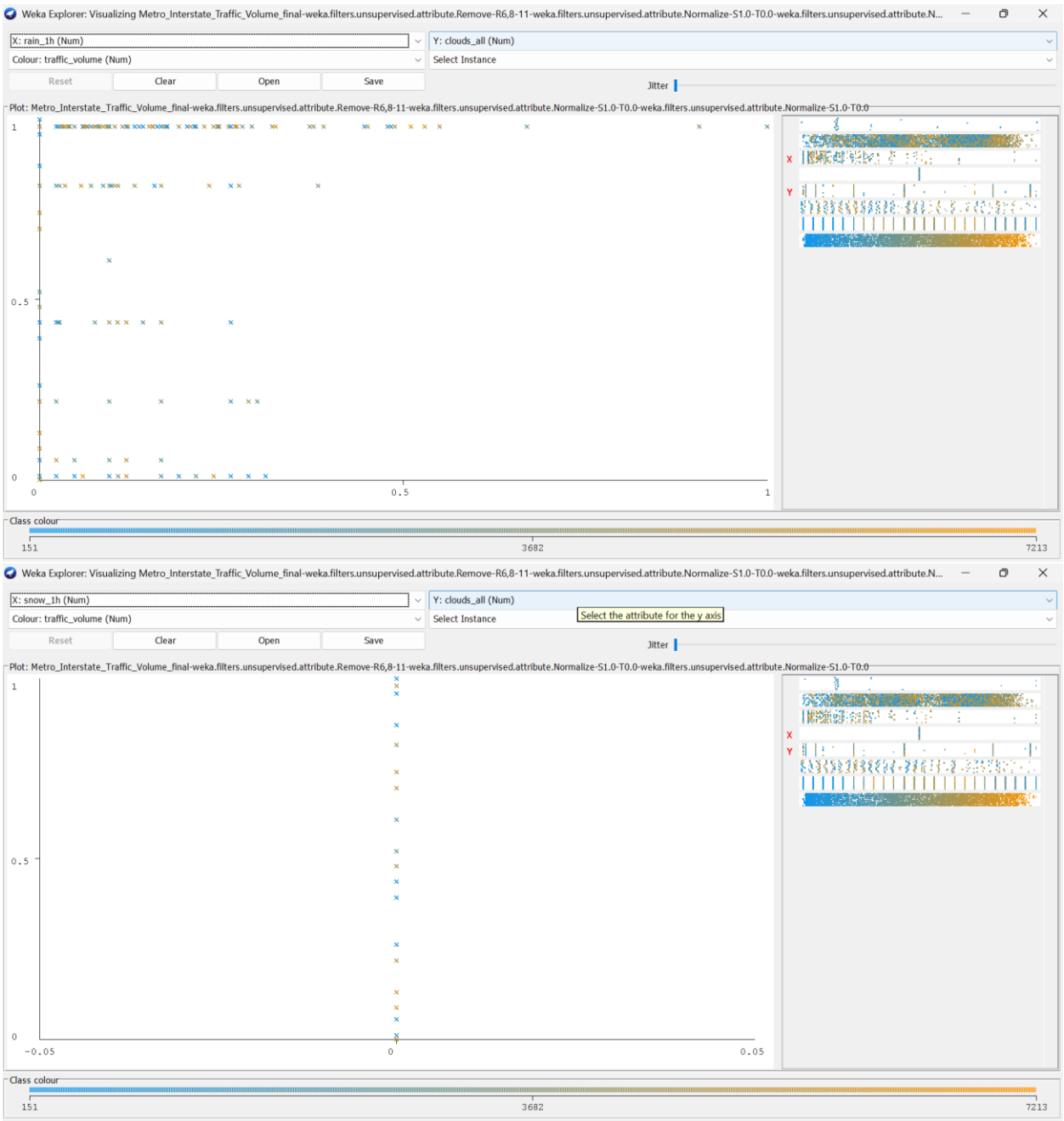


Figure 39: Visual 24 & 25

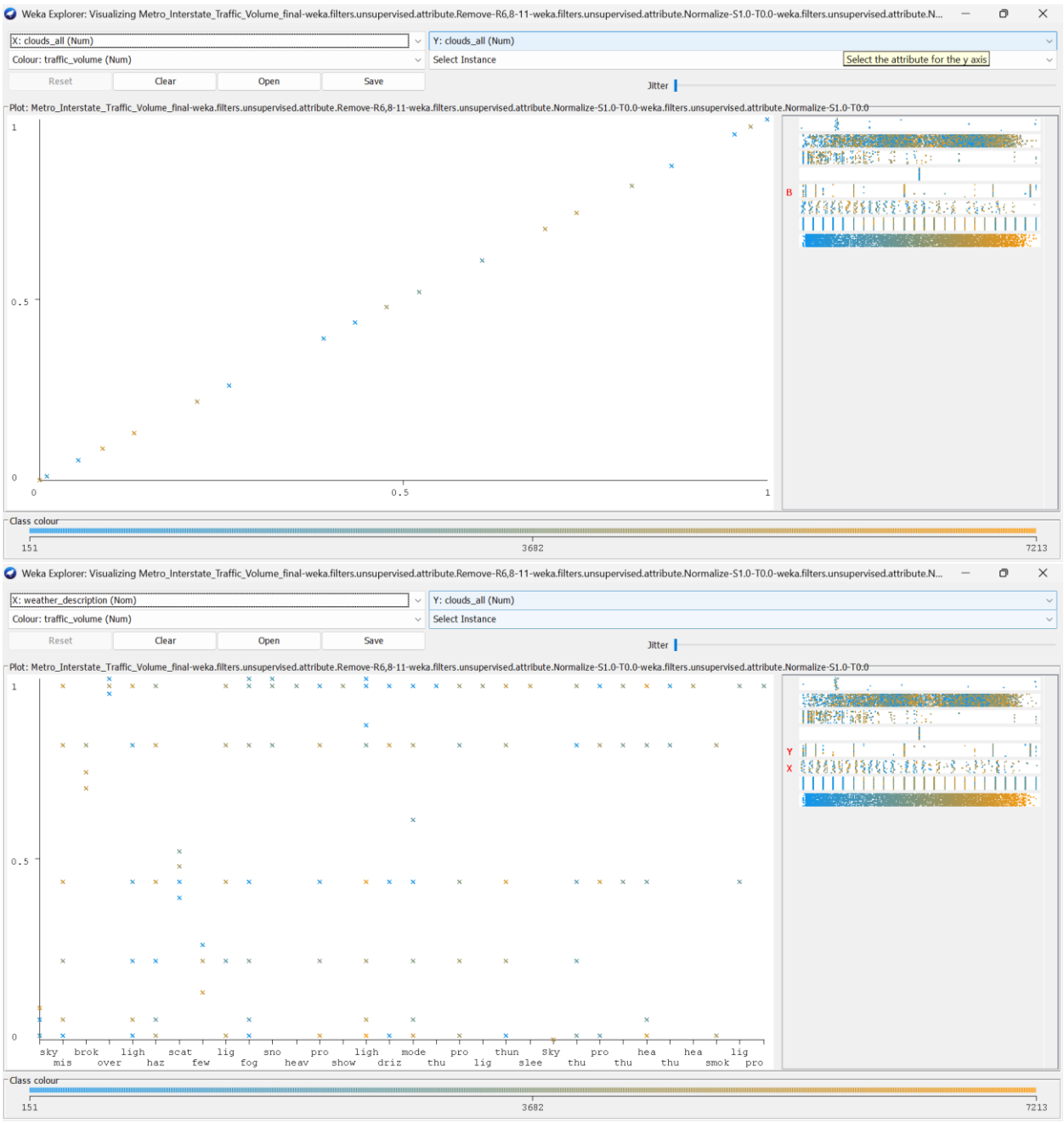


Figure 40: Visual 26 & 27

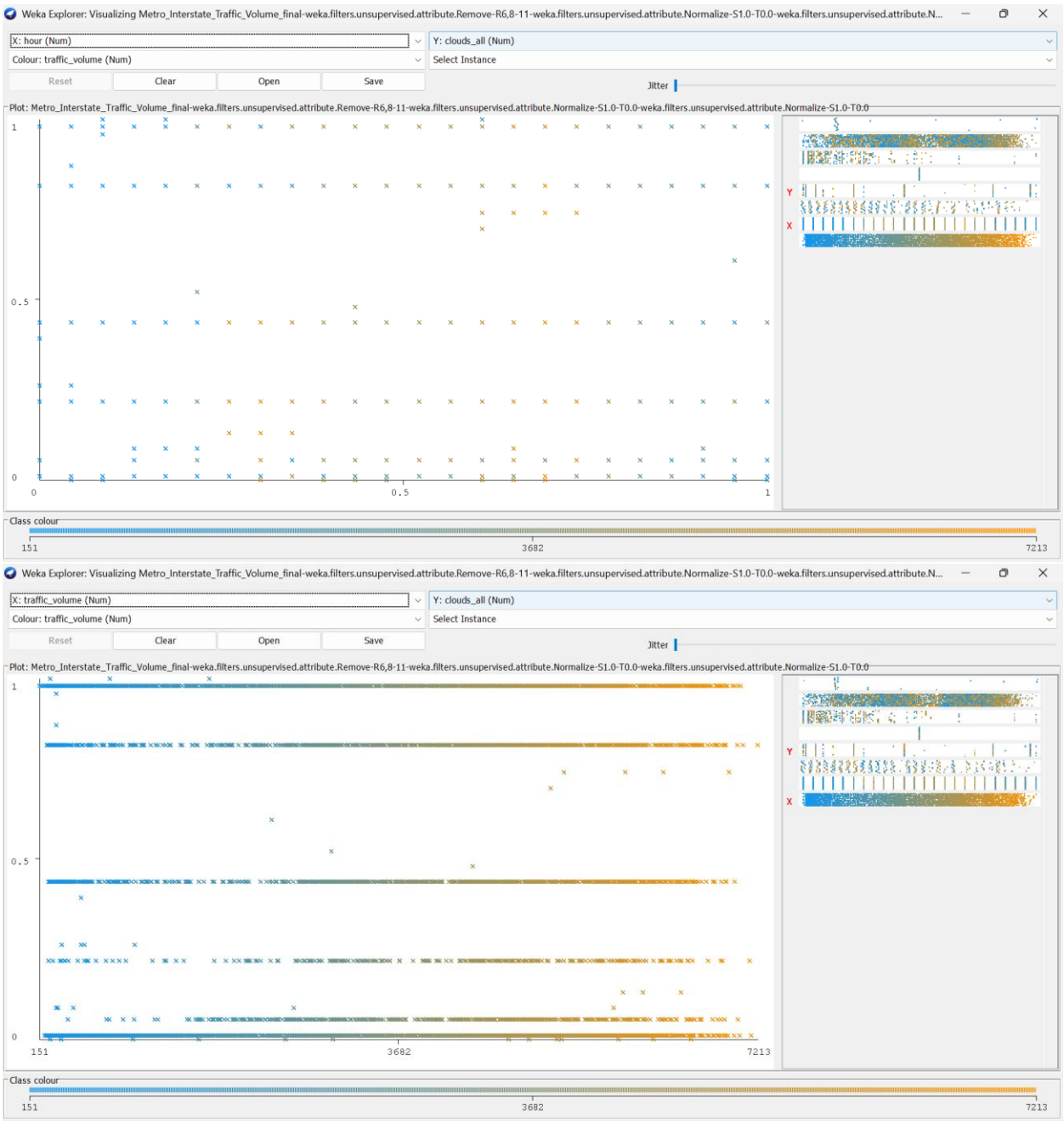


Figure 41: Visual 28 & 29

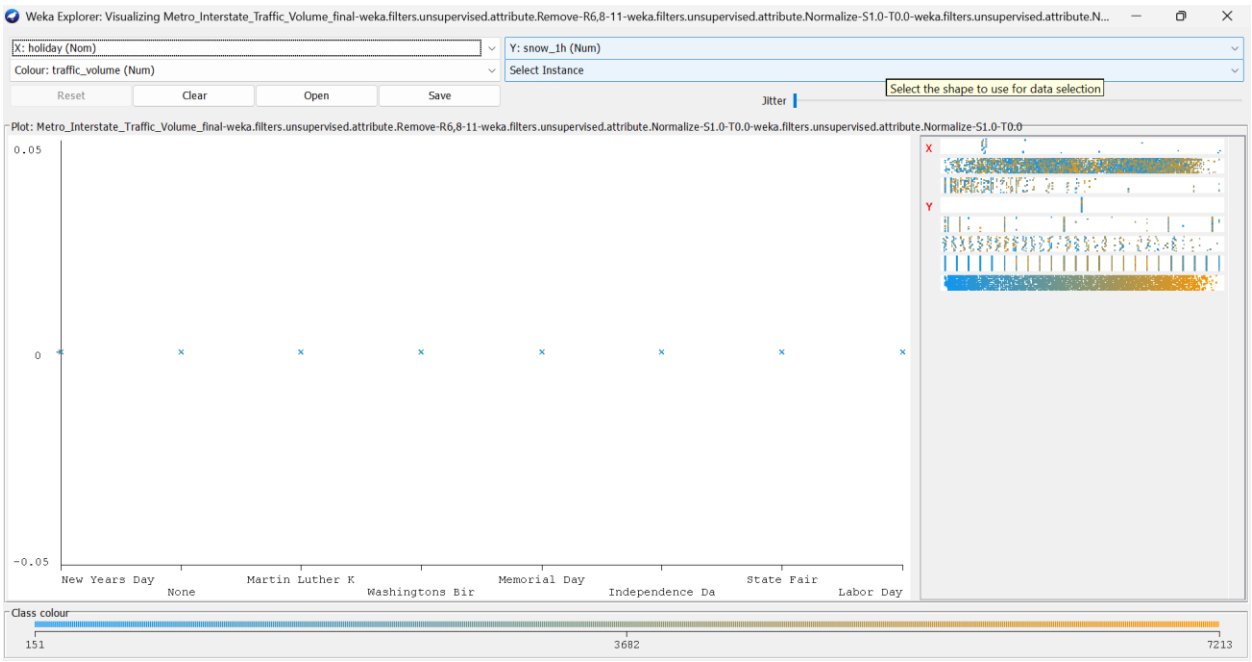


Figure 42: Visual 30 & 31

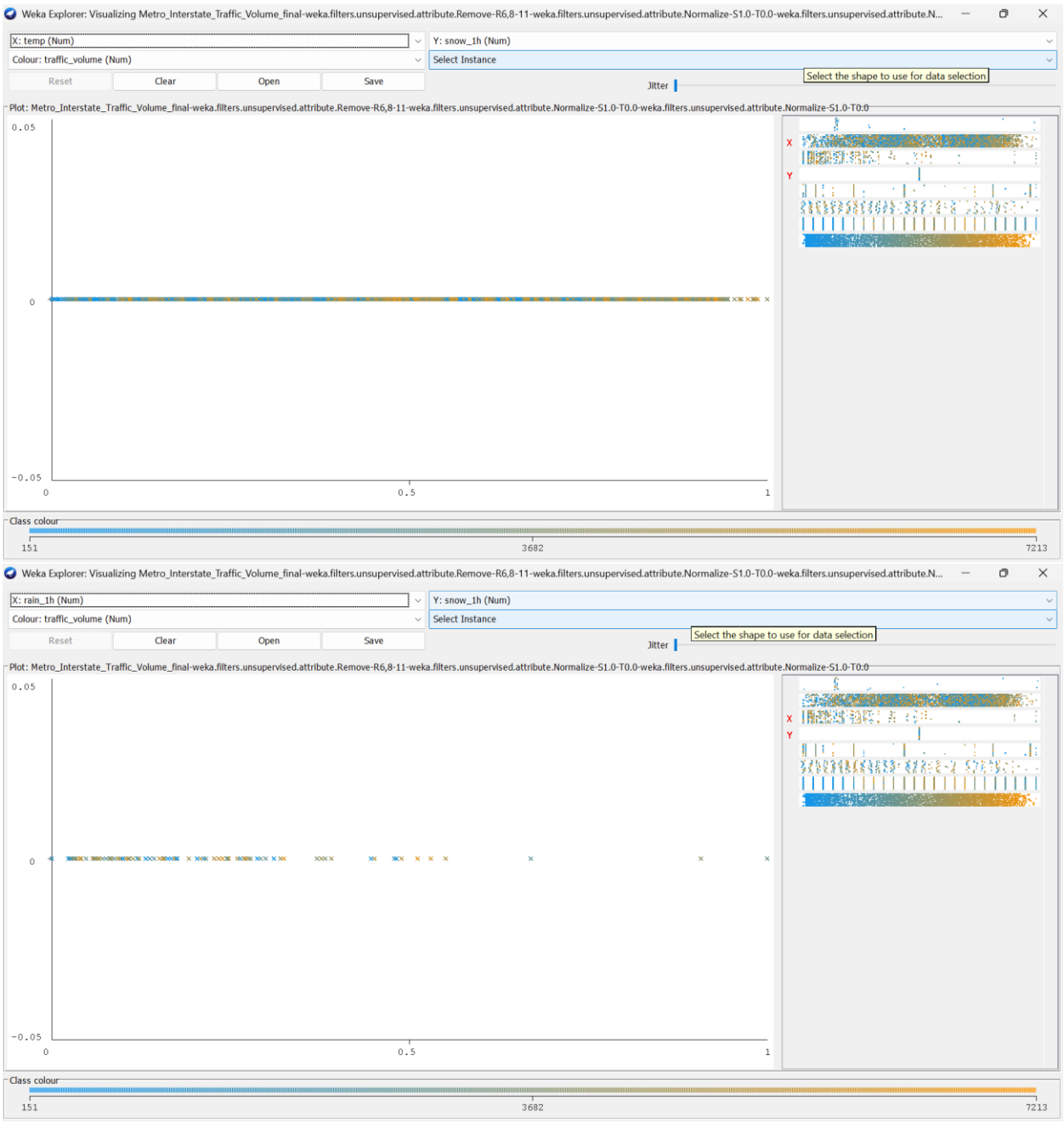


Figure 43: Visual 32 & 33

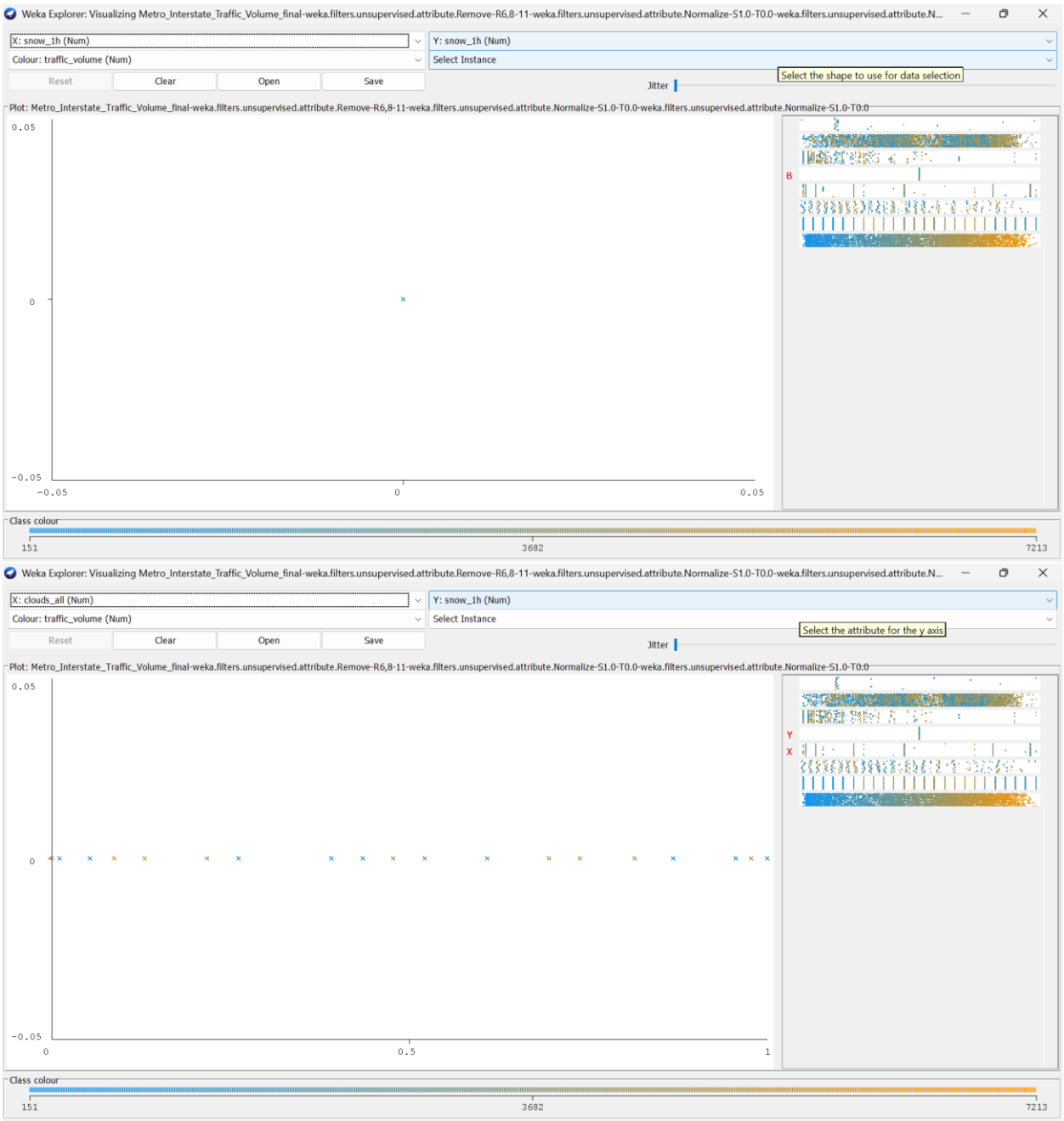


Figure 44: Visual 34 & 35

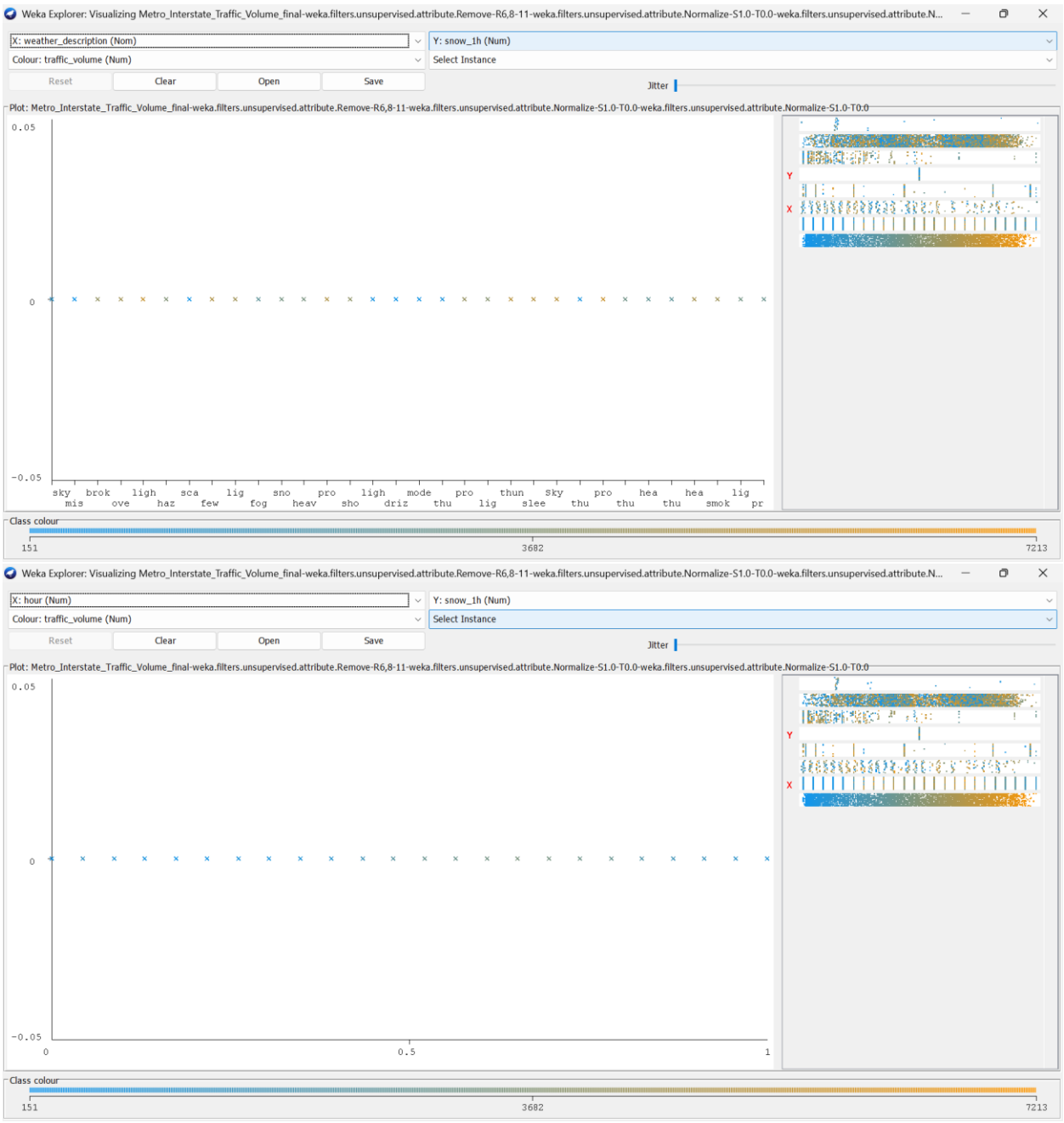


Figure 45: Visual 36 & 37

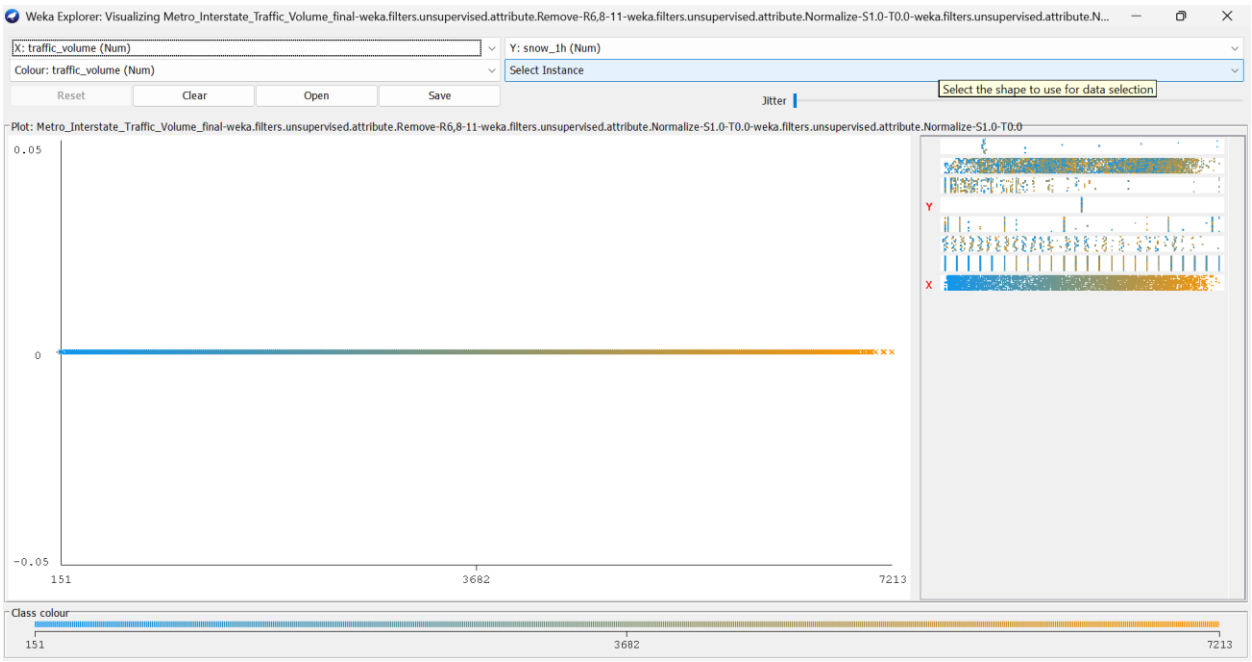


Figure 46: Visual 38 & 39

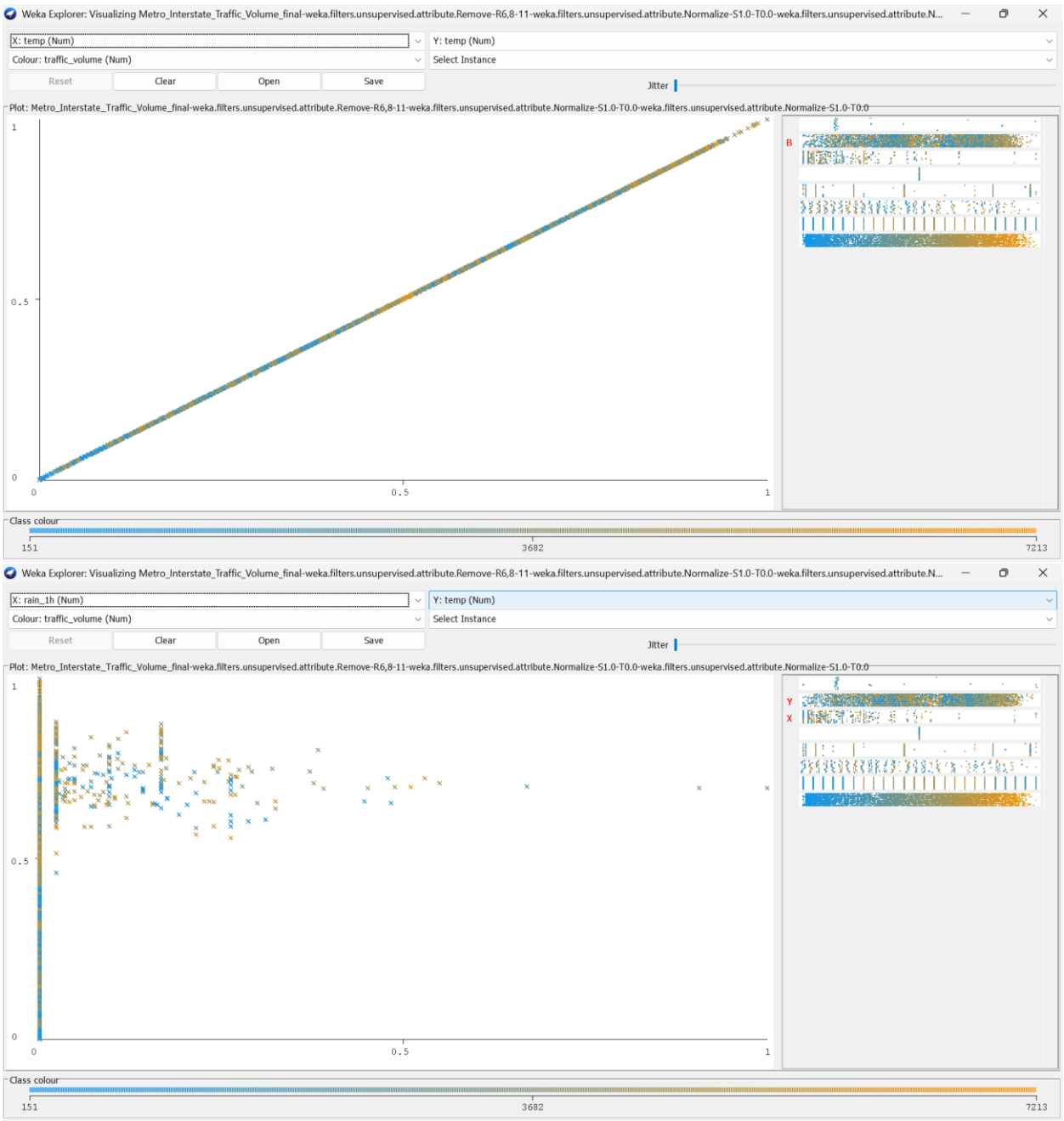


Figure 47: Visual 40 & 41

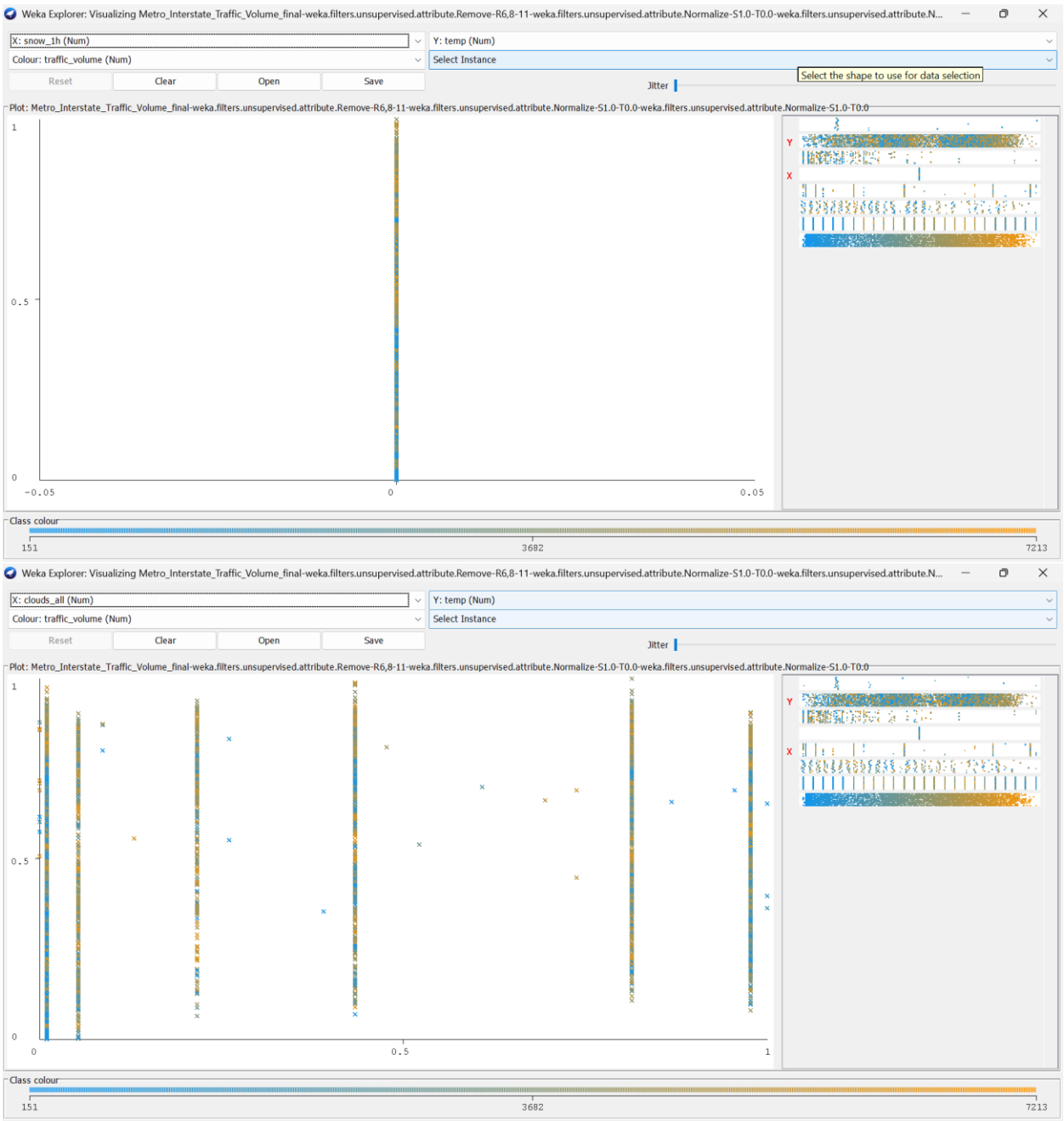


Figure 48: Visual 42 & 43

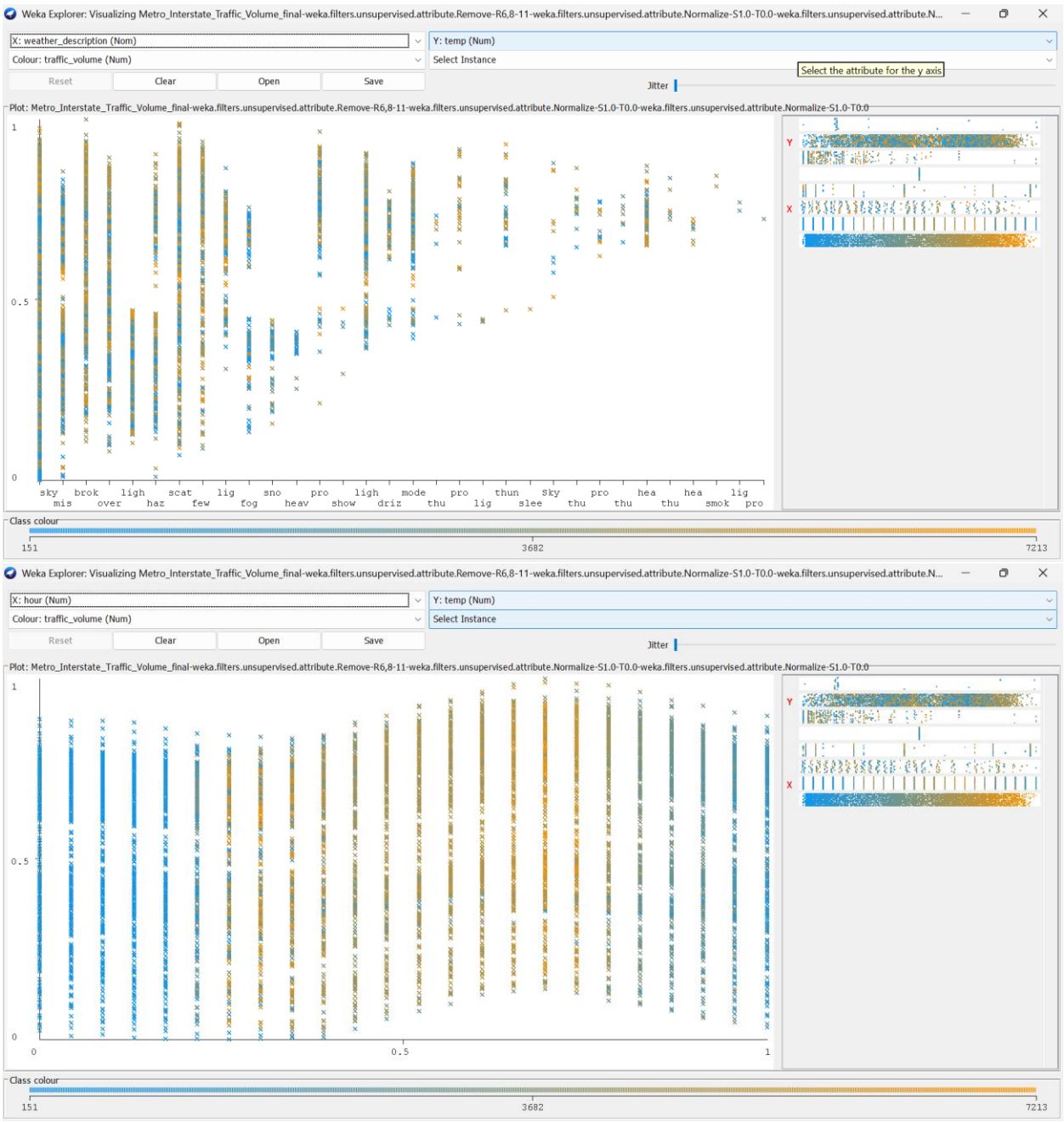
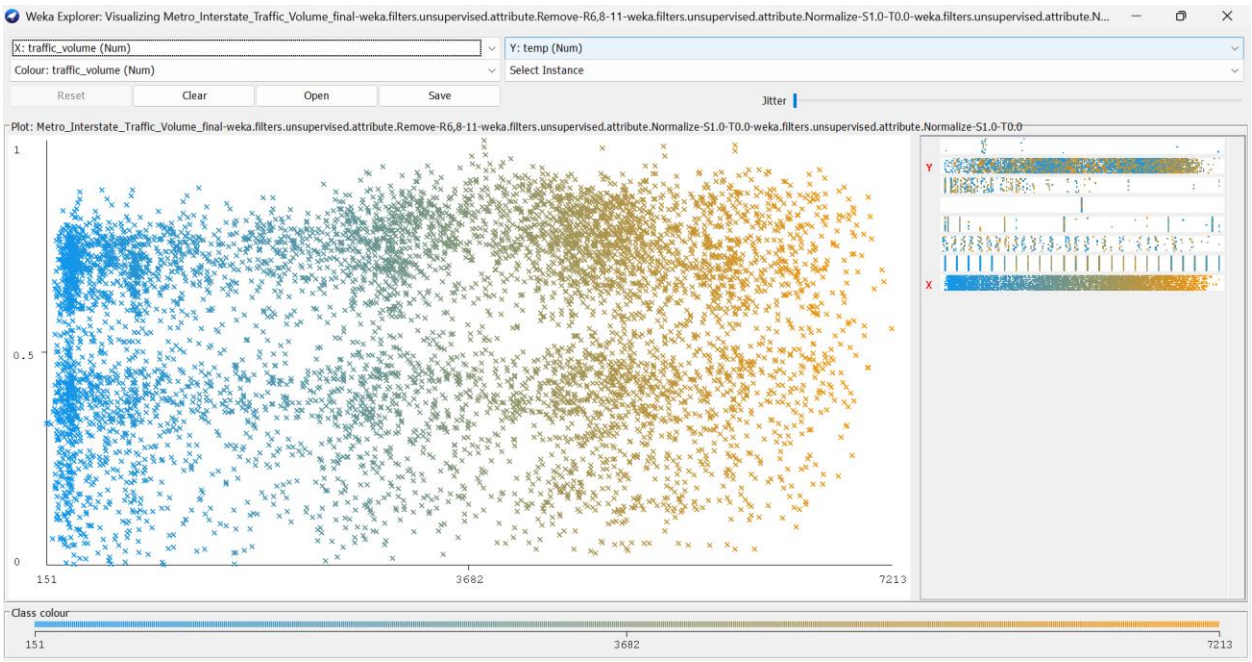


Figure 49: Visual 44 & 45



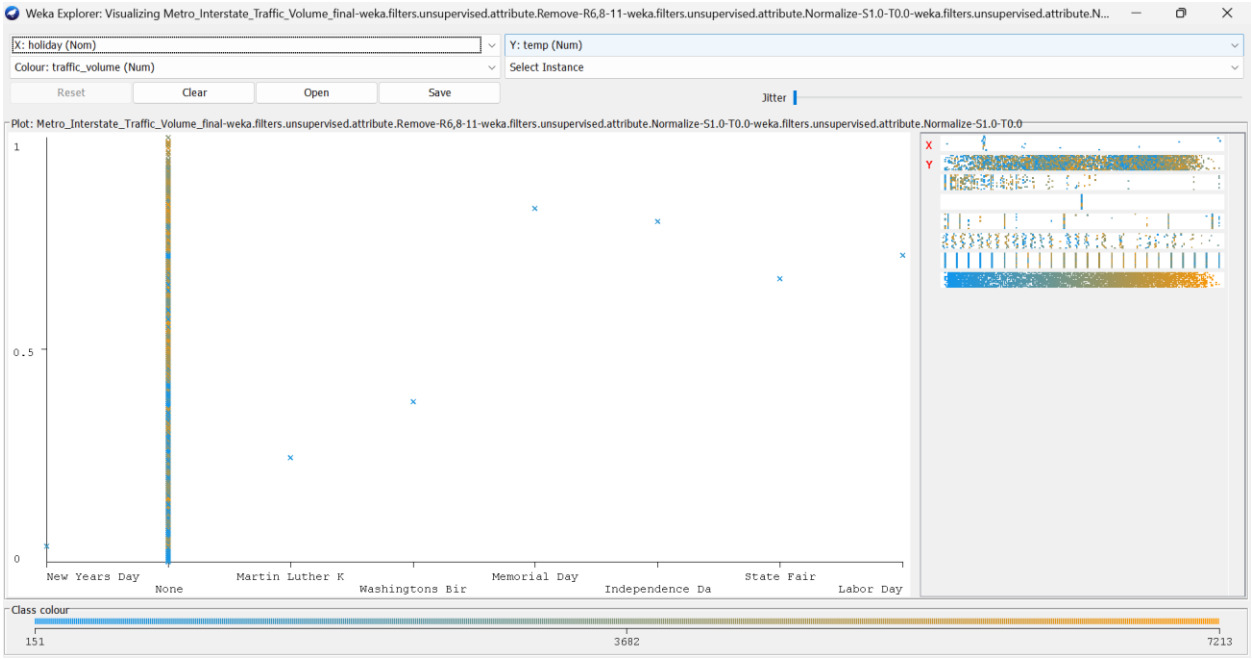


Figure 50: Visual 45 & 46

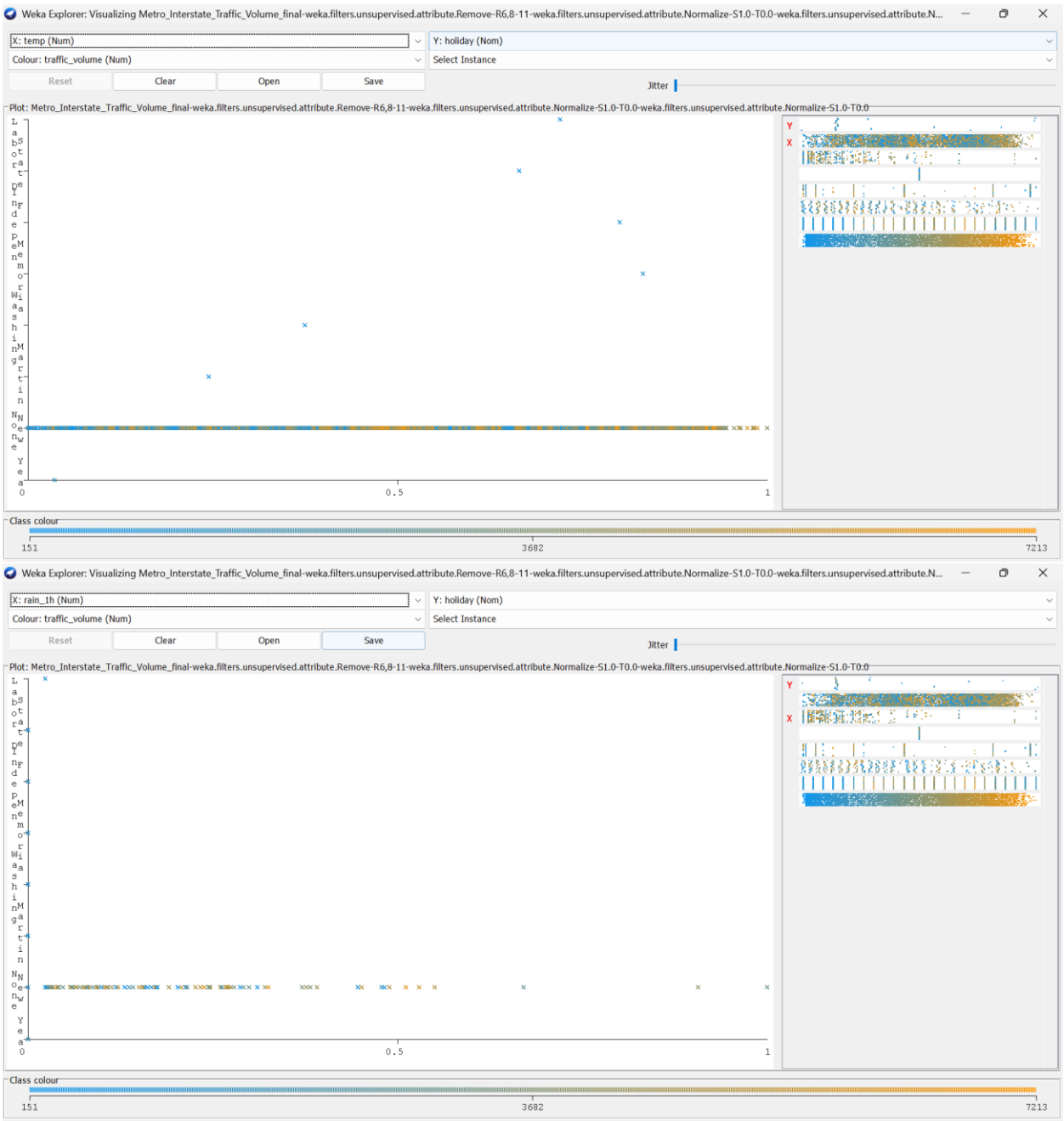


Figure 51: Visual 47 & 48

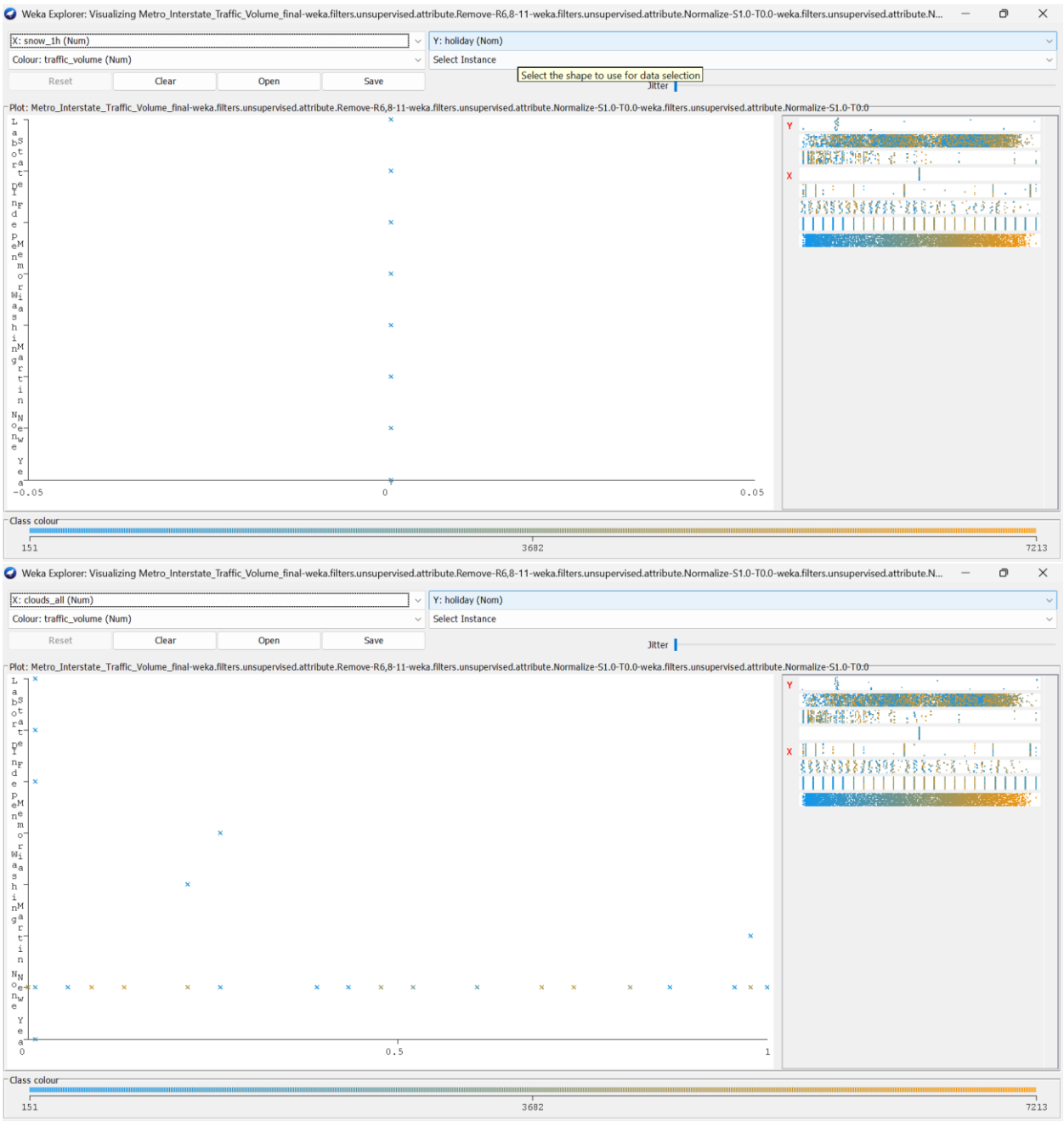


Figure 52: Visual 49 & 50

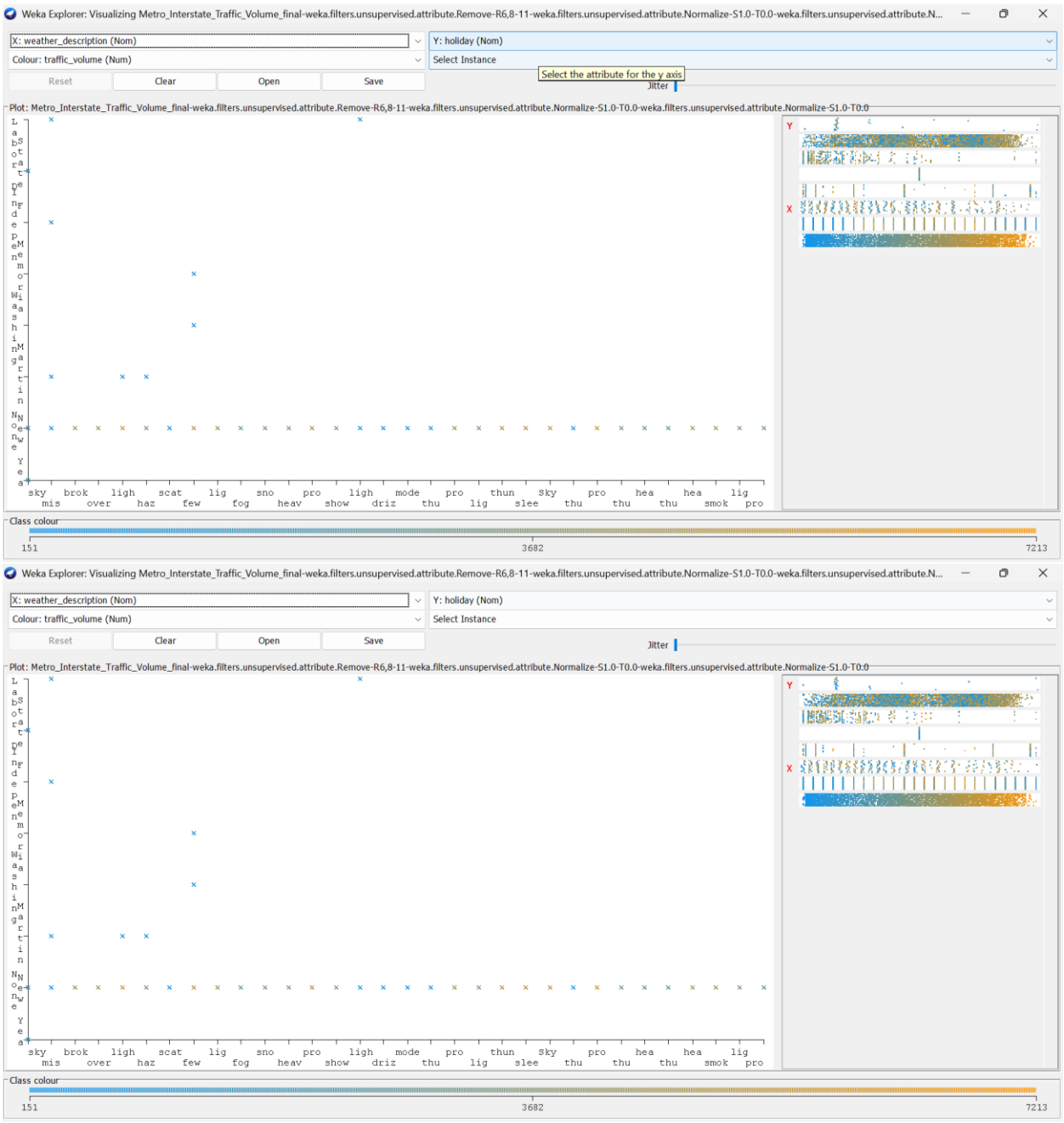


Figure 53: Visual 51 & 52

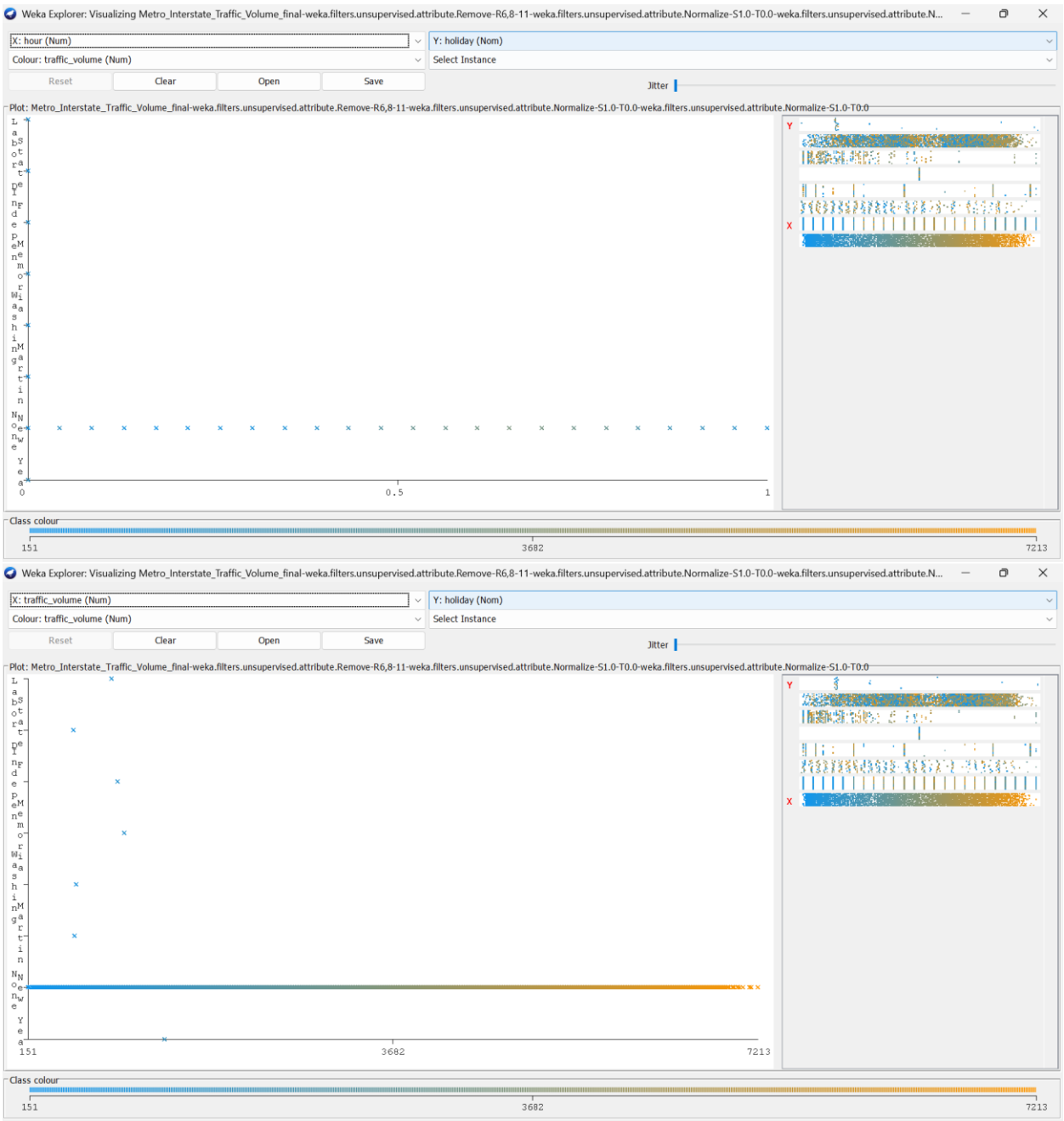


Figure 54: Visual 53 & 54

3.6.2 Techniques Used

- Histogram: Used to analyze the distribution of traffic_volume over different times of the day.
- Scatter Plot: Showed the correlation between temp and traffic_volume, indicating that extreme temperatures impact traffic volume. Cluster Visualization: Displayed the two traffic behavior groups identified in clustering.

3.6.3 Interpretation

Traffic Volume Trends: Higher traffic volume was observed during peak hours (morning and evening).Weather Influence: Clear weather conditions resulted in higher traffic volumes, while rain/mist led to lower traffic.

Chapter 4: Data Ethics

4.1 Overview

Ethical use of business information is critical in providing transparency, fairness, and security. The following ethical issues involve use and analysis of business information in this project:

4.2 Privacy and Data Protection

Traffic data, when combined with location tracking, can lead to privacy concerns when improperly utilized. Providing anonymization and GDPR and other data protection legislation compliance is essential.

4.3 Bias and Fairness

The dataset can become unbalanced for any region and hence can yield biased estimates. If predictive algorithms prefer one site over another, urban planning and budget allocation can become unequal.

4.4 Data Accuracy and Integrity

- Data preprocessing is a meticulous exercise to avert incorrect imputations and spurious trends. Errors in collection, such as incorrect timestamps and missing values, can impact prediction accuracy. Transparency in Data Usage Stakeholders ought to have a sense of how information is gathered, processed, and used.
- Offering clear documentation of model decisions can increase trust. Security Risks If real-time traffic information is disclosed, it can be utilized for illicit use (e.g., for attacking specific locations with cyber or physical attacks). It is imperative to deploy appropriate security controls, such as access controls and encryptions.

Chapter 5: Conclusion

5.1 Summary of Overall Findings, Trends, and Patterns

Peak Traffic Volumes: Peak volumes of traffic have been experienced during morning (8 AM) and evening (5–6 PM) rush hours, when commuting times occur. Weather Effect: Clear weather experienced higher volumes, while rain, snow, and fog played a major role in reducing volumes of traffic .Holidays vs. Not Holidays: Traffic during holidays was much lighter in contrast to work days.

5.2 Data Mining Outcomes and Model Fit

The Linear Model exhibited a 0.3939 correlation coefficient, indicative of a moderate level of relation between the independent factors (holidays, weather, and time) and traffic flow.

The K-Means Algorithm identified two dominant clusters:

- Cluster 0: Large traffic volumes under clear weather conditions.
- Cluster 1: Lower traffic volumes in adverse weather conditions.

The chosen prediction key factors (hour, temp, and holiday) increased model efficiency and accuracy.

5.3 Business Intelligence Analysis

Traffic Management: All these can be utilized by city planners and transport authorities to make infrastructure less congested and efficient. Weather-Dependent Planning: Traffic lights and emergency service can be optimized in relation to weather-related fluctuations in traffic. Holiday Traffic Forecasts: Enterprises can utilize such information for focused marketing, logistics reorientation, and enhancing customer experiences during peak and off-peak times of traffic.

This project successfully analyzed Metro Interstate Traffic Volume using Weka by applying preprocessing, clustering, classification, attribute selection, and visualization. The results indicate that traffic volume is primarily influenced by time, holidays, and weather conditions. Future work can focus on integrating real-time traffic updates for better predictions.

References

1. YouTube. (2024). *Apriori Algorithm with WEKA*. [online] Available at: https://youtu.be/ftUM9v4kdEk?si=TxSFDLD9zns-xFud [Accessed 8 Feb. 2025].

2. Power, in (2022). *Top 10 Most Important Data Cleaning Methods in Power BI | Power BI*. [online] YouTube. Available at: https://youtu.be/6DopXivHmP4?si=-i8rpina675BNYz- [Accessed 8 Feb. 2025].

3. Youtu.be. (2025). Available at: https://youtu.be/v6fP8gyCLLc?si=a8NORMftaw8aCfvI [Accessed 8 Feb. 2025].

4. Galassi, D. (2023). Traffic Volume. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/damianogalassi/traffic-volume?resource=download [Accessed 8 Feb. 2025].