Impact of Weather Factors on Traffic Conditions and Road Safety: A Case Study of Pan Island Expressway in Singapore

Submitted by

Jack Sim Jia Jing | U1920614D

School of Civil and Environmental Engineering

College of Engineering

Nanyang Technological university

A Final Year Project presented to the Nanyang Technological University
in partial fulfillment of the requirements for the
Degree of Bachelor of Engineering

2022/23

1. Abstract

This study examines the relationship between weather factors (rainfall, temperature, and wind speed) and the speed bands of vehicles on the Pan Island Expressway in Singapore during peak morning traffic hours. Using data from June, July, August, September, and December 2022, the analysis focuses on different road categories and conducts separate or combined month analyses. Traffic data was collected from the Land Transport Authority (LTA) DataMall, and weather data was acquired from Data.gov.sg. Multiple linear regression was performed to investigate the relationships between weather factors and average speed across different road categories.

The findings reveal that weather factors have limited explanatory power for average speed variations across different road categories, with weak correlations between average speed and weather factors. The R-squared values range from 0.5% to 14.5%, indicating that other factors not considered in this analysis might have a more significant impact on average speed. Further research is required to develop a more comprehensive understanding of the determinants of average speed on various road categories. This study contributes to the literature on the complex relationship between weather factors and road safety, informing traffic management strategies, road safety initiatives, and urban planning efforts.

2. Acknowledgements

The author would like to express his greatest appreciation and gratitude to his Final Year Project (FYP) supervisor, Assistant Professor Zhu Feng, for his continuous guidance and support throughout the project development phase despite his busy schedule. The author would also like to thank Panwar Shreyans for his mentorship since the beginning of the project. Finally, the author would also like to express heartfelt thanks to his family and friends for their unending support, help, and encouragement from the beginning to the end of the project.

Jack Sim Jia Jing

April 2023

3. Contents

4. Figures Contents

5. Table Contents

6. Introduction

Road safety and traffic management are critical concerns for urban planners, policymakers, and transportation agencies worldwide. The relationship between weather factors and road safety or traffic conditions has been extensively studied in the past, as adverse weather can have a significant impact on road conditions, driver behavior, and the likelihood of accidents (Bergel-Hayat et al., 2013; Koetse & Rietveld, 2009; Lee et al., 2015). Consequently, understanding the complex relationship between weather factors and traffic conditions is essential for designing effective traffic management strategies and improving road safety, particularly in the context of climate change and increased frequency of extreme weather events (Yu et al., 2021).

Several studies have investigated the relationship between weather factors, such as rainfall, temperature, and wind speed, and various aspects of road safety or traffic conditions. For example, Lee (Lee et al., 2015) developed a prediction model of traffic congestion using weather data, while Deb (Deb et al., 2019) examined travel time prediction using machine learning and the impact of weather on traffic conditions. These studies provide valuable insights into the influence of weather factors on traffic flow and congestion, informing traffic management and road safety strategies.

Furthermore, research has also focused on the impact of weather factors on specific road types and conditions (J. Andrey et al., n.d.; Chen et al., 2008; Eisenberg, 2004). For example, studies have analyzed the effects of weather on road accident risk (Bergel-Hayat et al., 2013; Fior & Cagliero, 2022), the impact of weather on fuel consumption and emissions (Shang et al., 2021), and the effect of weather on road capacity and traffic flow (Agarwal, n.d.). These studies highlight the importance of considering road type and context in understanding the relationship between weather factors and road safety or traffic conditions.

The current study aims to contribute to this body of literature by examining the relationship between weather factors, namely rainfall, temperature, and wind speed, and the speed bands of vehicles on the Pan Island Expressway in Singapore. By focusing on different road categories and conducting separate or combined month analyses, this study seeks to provide a more nuanced understanding of the complex relationship between weather factors and road safety in an urban context. The findings of this study can inform traffic management strategies, road safety initiatives, and urban planning efforts, ultimately contributing to the ongoing efforts to enhance road safety and traffic conditions in cities around the world

(Abdel-Aty et al., 2011; Antoniou et al., 2013; Malyshkina & Mannering, 2009; Theofilatos & Yannis, 2014).

In conclusion, this introduction has outlined the importance of studying the relationship between weather factors and road safety or traffic conditions, and the role of this research in informing traffic management strategies and urban planning. The following sections of this report will present a literature review, methodological approaches of this research, as well as the results, and conclusions.

7. Literature Review

Several studies have examined the relationship between weather factors and road safety or traffic conditions. Global-Similarity Local-Salience Network for Traffic Weather Recognition by Yu (Yu et al., 2021) and A Prediction Model of Traffic Congestion Using Weather Data by Lee (Lee et al., 2015) explored the use of machine learning and artificial intelligence techniques to predict traffic conditions and congestion based on weather data. Koetse (Koetse & Rietveld, 2009) provided an overview of empirical findings on the impact of climate change and weather on transport, emphasizing the need for better understanding and adaptation strategies in the face of changing weather patterns.

The effect of traffic and weather characteristics on road safety has been extensively studied by Theofilatos (Theofilatos & Yannis, 2014), who conducted a comprehensive review of the literature, identifying significant weather factors affecting road safety and the methodological approaches used in previous studies. Deb (Deb et al., 2019) investigated the use of machine learning to predict travel time based on weather data, demonstrating the potential of advanced modeling techniques to improve traffic management and safety under varying weather conditions.

Several studies have also explored the relationship between specific weather factors and road safety or traffic conditions. Rainfall has been identified as a significant factor affecting road safety and traffic flow in several studies, such as Andrey (J. Andrey & Yagar, 1993), Yannis (*Weather Effects on Daily Traffic Accidents and Fatalities.Pdf*, n.d.), Brijs (Brijs et al., 2008), and Abdel-Aty (Abdel-Aty et al., 2011). These studies have found that rain can lead to reduced visibility, increased braking distances, and other factors that contribute to increased crash risk and reduced traffic flow efficiency. Temperature has also been investigated in relation to road safety, with studies such as Hayat (Bergel-Hayat et al., 2013) and Keay (Keay & Simmonds, 2005) finding associations between temperature variations and road traffic volume or accident risk.

Wind speed is another weather factor that has been examined in the context of road safety. Studies such as Fior (Fior & Cagliero, 2022) and Chen C (Chen et al., 2008) have explored the relationship between wind speed and road traffic accidents or mode choice decisions, finding significant associations between wind speed and road safety outcomes. In addition, studies such as Usman (Usman et al., 2010) and Hammad (Hammad et al., 2019) have investigated the impact of weather conditions on road traffic accidents, identifying various environmental factors that can affect road safety.

The relationship between weather factors and traffic flow has also been studied in the context of urban freeway systems. Agarwal et al. (Agarwal, n.d.) examined the impacts of weather on urban freeway traffic flow characteristics and facility capacity, finding significant effects of weather conditions on traffic flow and capacity. Andrey et al. (D. J. Andrey, n.d.) investigated the role of weather information in road safety, emphasizing the importance of accurate and timely weather data for traffic management and safety strategies. Chen (Chen et al., 2008) studied the role of the built environment on mode choice decisions under varying weather conditions, highlighting the influence of weather on transportation choices and the need for urban planning that accounts for these factors.

Several studies have focused on the temporal aspects of weather-related road accidents and traffic conditions. Shin (Shin & Lee, 2020) conducted a temporal analysis of traffic accidents in South Korea, identifying patterns related to weather conditions and other factors. Cools (Cools et al., 2010) assessed the impact of weather on traffic intensity, finding significant temporal variations in the relationship between weather conditions and traffic flow.

The effect of weather on road accident severity has also been a focus of research. Jung (Jung et al., 2010) investigated the rainfall effect on single-vehicle crash severities using polychotomous response models, while Malyshkina (Malyshkina & Mannering, 2009) applied a Markov switching multinomial logit model to analyze accident-injury severities under different weather conditions. These studies provide insights into the complex relationship between weather factors and the severity of road accidents, informing traffic management and road safety strategies.

In summary, the literature on the relationship between weather factors and road safety or traffic conditions is extensive, encompassing various methodological approaches, weather factors, and road safety outcomes. The current study builds on this literature by examining the relationship between weather factors and speed bands of vehicles on the Pan Island Expressway in Singapore. By analyzing the data with a focus on different road categories and separate or combined month analyses, this study aims to provide valuable insights into the complex relationship between weather factors and road safety in an urban context.

This literature review has discussed key findings and methodological approaches in the research on weather factors and road safety or traffic conditions. The next sections of this report present the methodology, results and analysis, and the conclusion of this study.

8. Methodology

8.1. Data Collection

The datasets used in this study cover the peak morning traffic hours in Singapore, specifically from 6 am to 8 am. Traffic incident and speed band datasets were obtained from the Land Transport Authority (LTA) DataMall, which provides real-time traffic data collected by sensors distributed across Singapore. These sensors gather information at 5-minute intervals. Additionally, weather datasets, including rainfall, temperature, and wind speed data, were acquired from Data.gov.sg. While these weather datasets are updated every minute, this study utilized weather data collected at 5-minute intervals. The data analyzed in this study span the months of June, July, August, September, and December 2022.

To collect traffic data, an API request was sent using Python to 'http://datamall2.mytransport.sg/ltaodataservice/', followed by the specific data type for analysis (e.g., to extract Traffic Incidents data, the URL used was 'http://datamall2.mytransport.sg/ltaodataservice/TrafficIncidents'). A dedicated folder was created to store the data in JSON format. Given that LTA's sensors update the data every 5 minutes, the Python script was run continuously for a predetermined number of days to collect the necessary data.

Weather data was available for download directly from the Data.gov.sg website. To collect this data, an API request was sent using Python to 'https://api.data.gov.sg/v1/environment/', followed by the desired data type for analysis (e.g., the URL for extracting wind speed data was 'https://api.data.gov.sg/v1/environment/wind-speed?date='). The default format for data downloaded from Data.gov.sg is JSON.

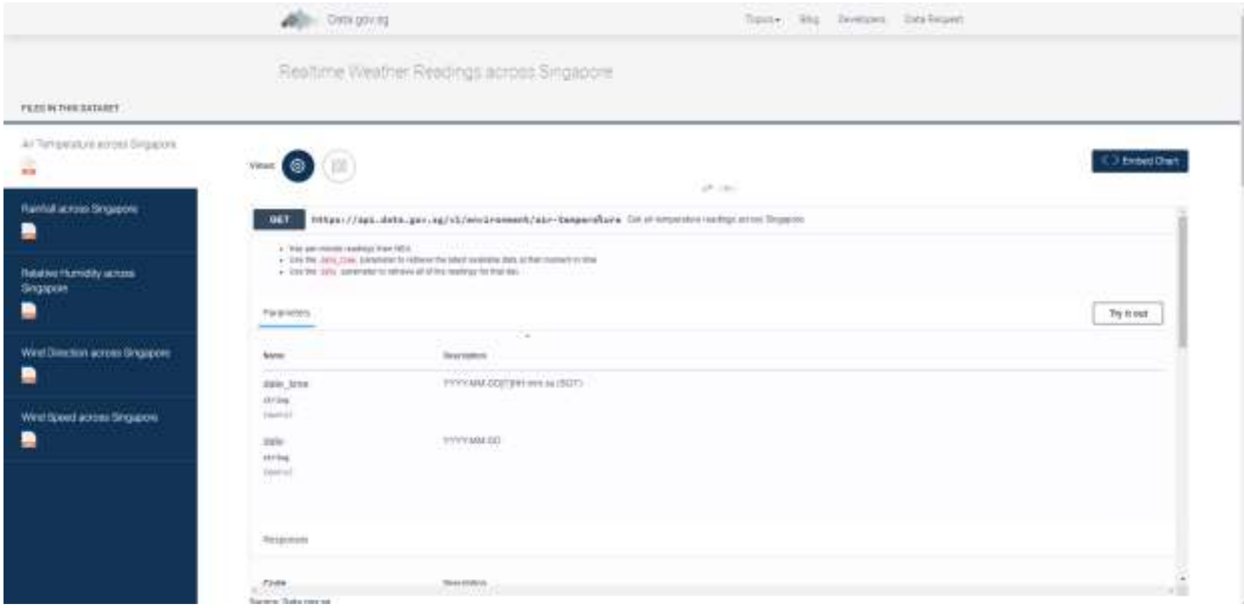

Figure 8.1 : Data.gov.sg API call

8.2. Data Cleaning and Preprocessing

Once the data were collected, they were extracted using Jupyter Notebook and converted into Pandas DataFrames for further processing. First, the speed band data were filtered to include only the observations between 6 am and 8 am, and then the Pan Island Expressway data were filtered using the road name. The traffic incidents dataset comprised eight different incident types: accidents, road works, vehicle breakdowns, obstacles, roadblocks, heavy traffic, roadblock, and unattended vehicles. Heavy traffic incidents were removed from the dataset. Subsequently, traffic incidents within 500 meters of any speed band sensor were also removed to reduce potential confounding factors that could contribute to traffic congestion.

Out[13]:

	date	Time	LinkID	RoadName	RoadCategory	SpeedBand	MinimumSpeed	MaximumSpeed	start_location	end_location	nea
29	2022-09-01	06-03-22	103000078	PAN ISLAND EXPRESSWAY	A	8	70	999	(1.3479138498233316, 103.6923894384768)	(1.3491170726830268, 103.69342360283092)	
30	2022-09-01	06-03-22	103000079	PAN ISLAND EXPRESSWAY	A	7	60	69	(1.3491170726830268, 103.69342360283092)	(1.3479138498233316, 103.6923894384768)	
31	2022-09-01	06-03-22	103000080	PAN ISLAND EXPRESSWAY	A	8	70	999	(1.3436196253925365, 103.78499221155269)	(1.3431444683048892, 103.78388724345461)	
48	2022-09-01	06-03-22	103000122	PAN ISLAND EXPRESSWAY	A	8	70	999	(1.350438528643087, 103.6944089247103)	(1.3517801862142893, 103.6953675592183 8)	
49	2022-09-01	06-03-22	103000123	PAN ISLAND EXPRESSWAY	A	7	60	69	(1.3517801862142893, 103.69536755921838)	(1.350438528643087, 103.6944089247103 3)	
...	...	...	...	...	...	...	...	...	...	...	
32333960	2022-09-28	07-57-09	115000336	PAN ISLAND EXPRESSWAY	F	5	40	49	(1.3329723955422337, 103.76342970083786)	(1.3337580137074072, 103.7620812510051)	
32335068	2022-09-28	07-57-09	117000002	PAN ISLAND EXPRESSWAY	A	6	50	59	(1.339143430149792, 103.9747625390361)	(1.3398504107917306, 103.97311185921735)	
32335069	2022-09-28	07-57-09	117000003	PAN ISLAND EXPRESSWAY	A	8	70	999	(1.3398504107917306, 103.97311185921735)	(1.339143430149792, 103.9747625390361)	
32335072	2022-09-28	07-57-09	117000024	PAN ISLAND EXPRESSWAY	F	8	70	999	(1.3387736443102318, 103.97618119422671)	(1.3382945440631218, 103.97719894282493)	
32335073	2022-09-28	07-57-09	117000028	PAN ISLAND EXPRESSWAY	F	8	70	999	(1.3393867622148272, 103.9748408621742)	(1.3387736443102318, 103.97618119422671)	

500872 rows × 11 columns

Figure 8.2: Speed band data frame

The rainfall data were filtered to cover the 6 am to 8 am timeframe. Next, the Haversine formula was used to identify the nearest rainfall station to the speed sensors. The Haversine formula is a mathematical equation that calculates the great-circle distance between two points on the Earth's surface, taking into account the Earth's curvature. This formula is particularly useful in determining accurate distances between points on a sphere, such as the Earth, and is often used in geospatial analysis and navigation.

Similarly, the wind and temperature data were filtered to include only the observations between 6 am and 8 am, and the Haversine formula was employed to determine the closest station to the speed sensors. Notably, the wind and temperature data collection stations were the same, with a limited number of stations primarily located in the western part of Singapore. Consequently, a threshold of 8 km was established; if the speed band sensors were more than 8 km away from the nearest temperature and wind stations, the data for that speed band were excluded from the analysis. The 8 km threshold was determined using the pairwise distance method, yielding an optimal distance of 8 km.

Out[12]:

	station_id	value	name	date	Time	Location
4757	S77	0.0	Alexandra Road	2022-06-01	06-00-00	(1.2937, 103.8125)
4758	S109	0.0	Ang Mo Kio Avenue 5	2022-06-01	06-00-00	(1.3764, 103.8492)
4759	S90	0.0	Bukit Timah Road	2022-06-01	06-00-00	(1.3191, 103.8191)
4760	S114	0.0	Choa Chu Kang Avenue 4	2022-06-01	06-00-00	(1.38, 103.73)
4761	S50	0.0	Clementi Road	2022-06-01	06-00-00	(1.3337, 103.7768)
...	...	...	...	...	...	...
4207504	S69	0.0	Upper Peirce Reservoir Park	2022-12-31	07-55-00	(1.37, 103.805)
4207505	S08	0.0	Upper Thomson Road	2022-12-31	07-55-00	(1.3701, 103.8271)
4207506	S116	0.0	West Coast Highway	2022-12-31	07-55-00	(1.281, 103.754)
4207507	S104	0.0	Woodlands Avenue 9	2022-12-31	07-55-00	(1.44387, 103.78538)
4207508	S100	0.0	Woodlands Road	2022-12-31	07-55-00	(1.4172, 103.74855)

350987 rows × 6 columns

Figure 8.3: Rainfall data frame

8.3. Data Visualization

After cleaning the data, all datasets were merged into a single DataFrame using date, closest station, and time as common keys. Before conducting the analysis, average speed was calculated using the minimum and maximum speeds. It is important to note that for rows with a speed band equal to 8, the maximum speed in the raw data is 999 km/h. In this project, the maximum speed for such rows was converted to 100 km/h, which is the design speed of the Pan Island Expressway. To reduce confounding factors, the data were further filtered based on road categories (A, B, and F) present on the Pan Island Expressway, as different road categories have different speed limits. Exploratory Data Analysis and multiple linear regression were performed on the final dataset.

The total monthly rainfall data for 2022 was analyzed, revealing that December 2022 had the highest total monthly rainfall among the five months of data, while September 2022 had the lowest total monthly rainfall. Consequently, separate analyses were conducted for these months to determine if the relationships between variables changed across different months.

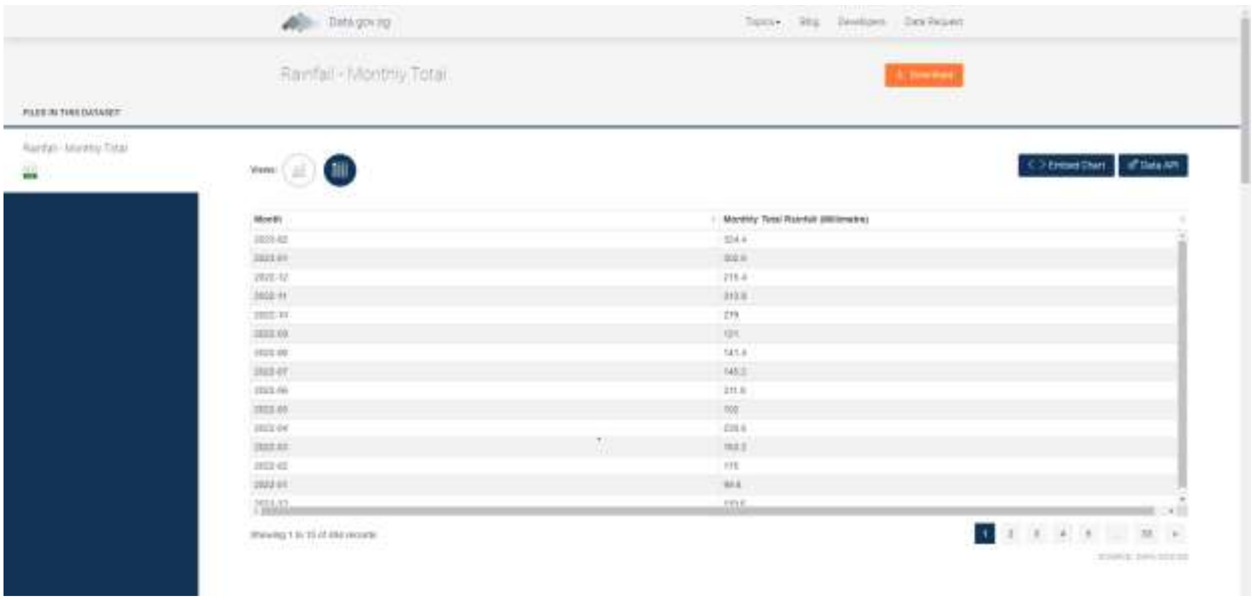

Figure 8.4: Total Monthly Rainfall

Table 8.1: Total Monthly Rainfall

Date	Total Monthly Rainfall (mm)
01-2022	99.8
02-2022	175
03-2022	163.2
04-2022	239.6
05-2022	102
06-2022	211.8
07-2022	145.2
08-2022	141.4
09-2022	121
10-2022	279
11-2022	313.8
12-2022	215.4

9. Results & Analysis

In the following chapter, the results and analysis of the investigation into the correlation between the average speed on the Pan Island Expressway (PIE) in Singapore and weather variables such as temperature, wind speed, and rainfall are presented. This analysis is vital in understanding the impact of weather conditions on traffic flow and road safety, enabling better planning and management of traffic during adverse weather events. To provide a comprehensive analysis, different road categories on the PIE have been considered, which include A, B, and F, each with its distinct speed limits and traffic characteristics.

For each road category, separate analyses have been conducted focusing on September and December, as well as a combined analysis of all months. This approach allows the examination of any potential differences in the relationships between the average speed and weather variables across various months, which may experience different weather patterns. By investigating the correlations in this manner, a deeper understanding of the interplay between weather conditions and traffic flow on the Pan Island Expressway can be developed.

In the ensuing sections, the detailed results of the analyses for each road category and month will be presented, as well as insights into the underlying relationships between the average speed on the PIE and the weather variables. Furthermore, the implications of the findings for traffic management, road safety, and future research in this area will be discussed.

9.1. Road Category A

9.1.1. Combined Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	1.380260e+06	1.380260e+06	1.355029e+06	1.156847e+06
mean	7.763118e+01	4.336401e-02	2.629198e+01	2.992190e+00
std	1.190351e+01	3.700720e-01	1.363160e+00	2.083759e+00
min	4.500000e+00	0.000000e+00	2.170000e+01	3.000000e-01
25%	6.450000e+01	0.000000e+00	2.520000e+01	1.500000e+00
50%	8.500000e+01	0.000000e+00	2.630000e+01	2.400000e+00
75%	8.500000e+01	0.000000e+00	2.730000e+01	3.900000e+00
max	8.500000e+01	1.040000e+01	3.020000e+01	2.450000e+01

Figure 9.1: Descriptive Statistics

The count row displays the number of data points for each variable, with 1,380,260 data points for average speed and rainfall value, 1,355,029 for temperature, and 1,156,847 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 77.63, rainfall value at 0.043, temperature at 26.29, and wind value at 2.99.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 11.90, for rainfall value it is 0.37, for temperature it is 1.36, and for wind value it is 2.08. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

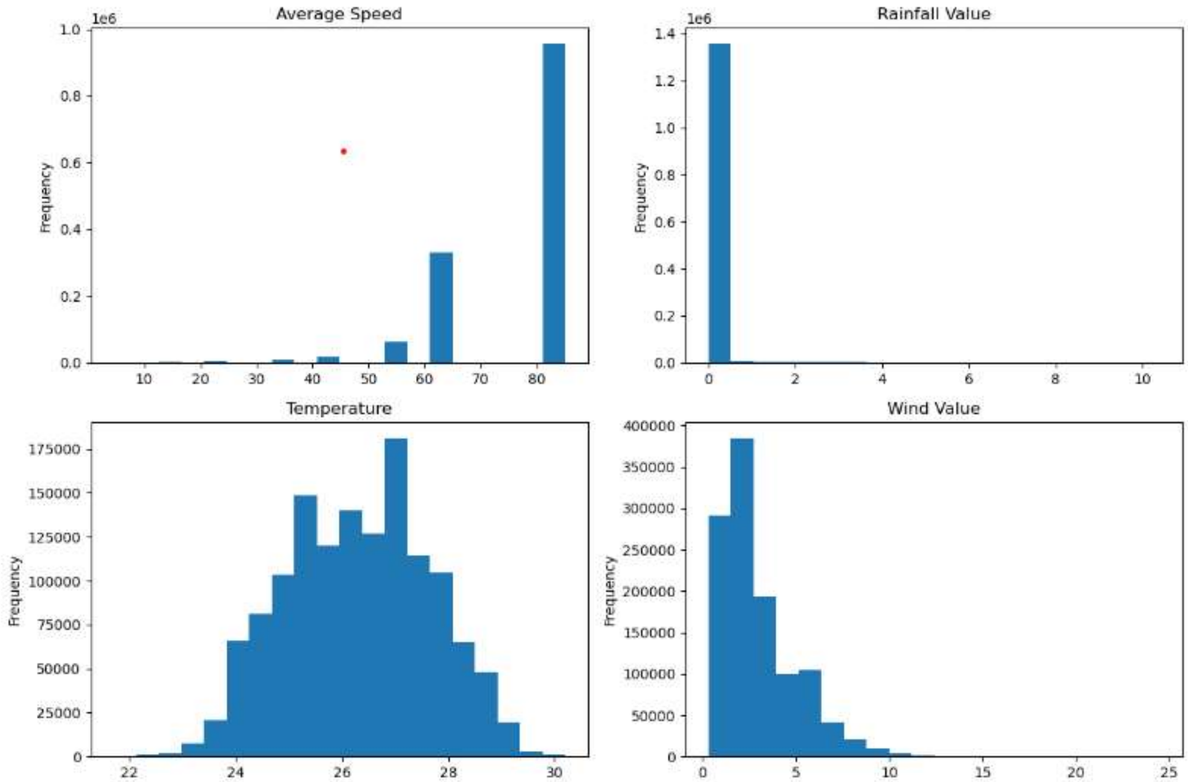

Figure 9.2: Frequencies of each data – Combined Analysis (A)

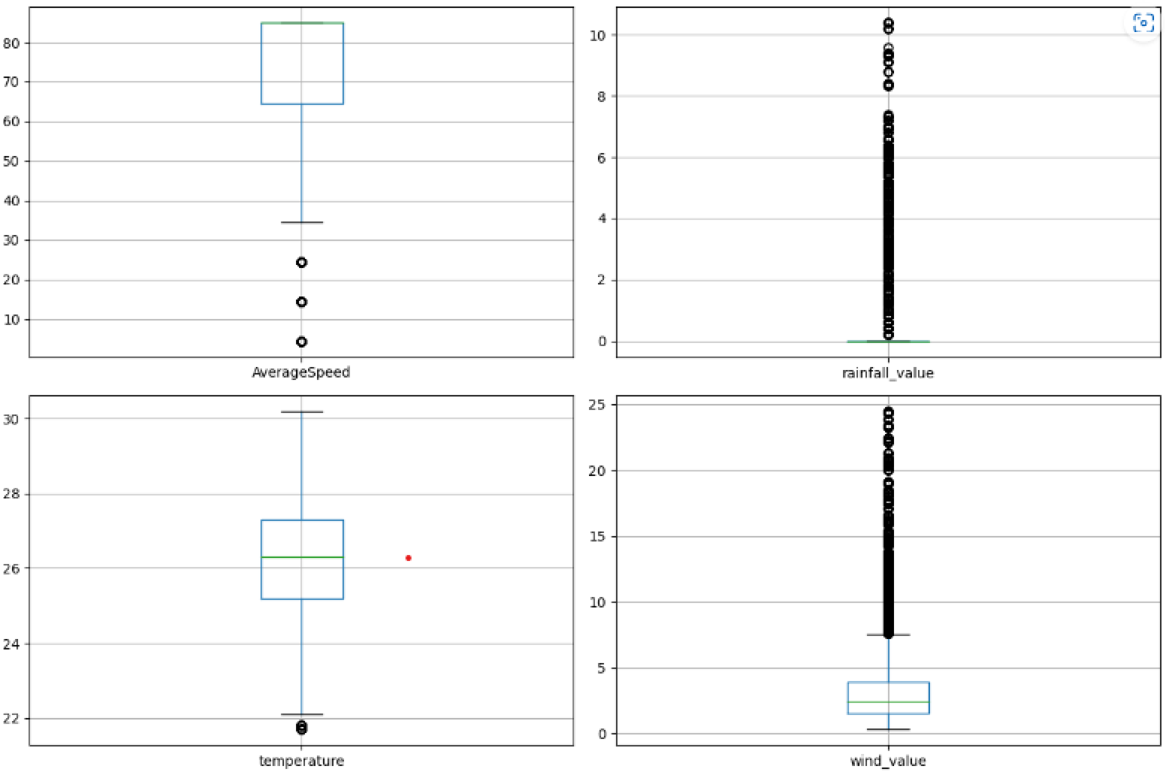

Figure 9.3: Boxplots of each data – Combined Analysis (A)

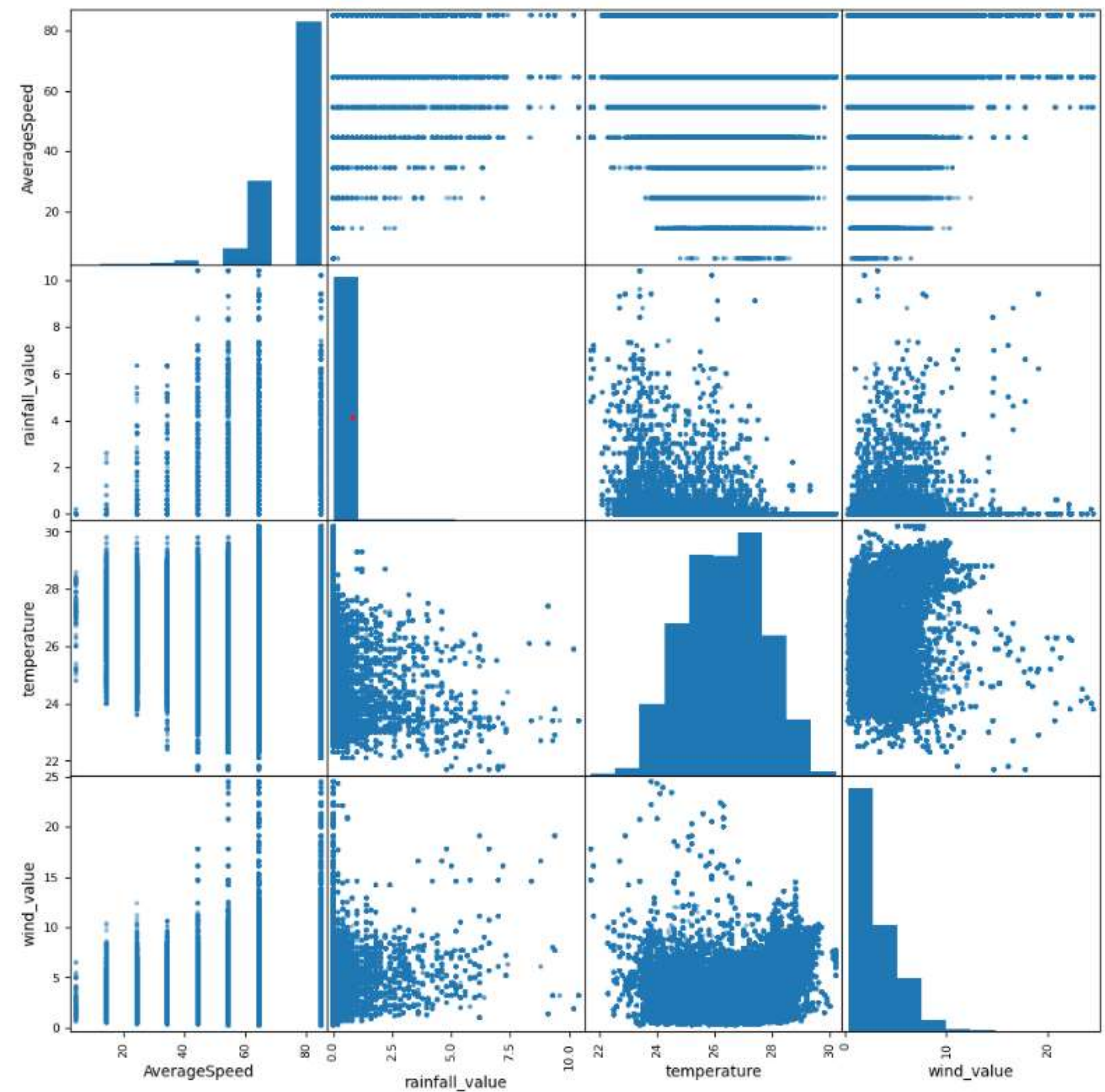

Figure 9.4: Scatter Matrix – Combined Analysis (A)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.060374	0.028814	0.032495
rainfall_value	-0.060374	1.000000	-0.170398	0.136501
temperature	0.028814	-0.170398	1.000000	0.084413
wind_value	0.032495	0.136501	0.084413	1.000000

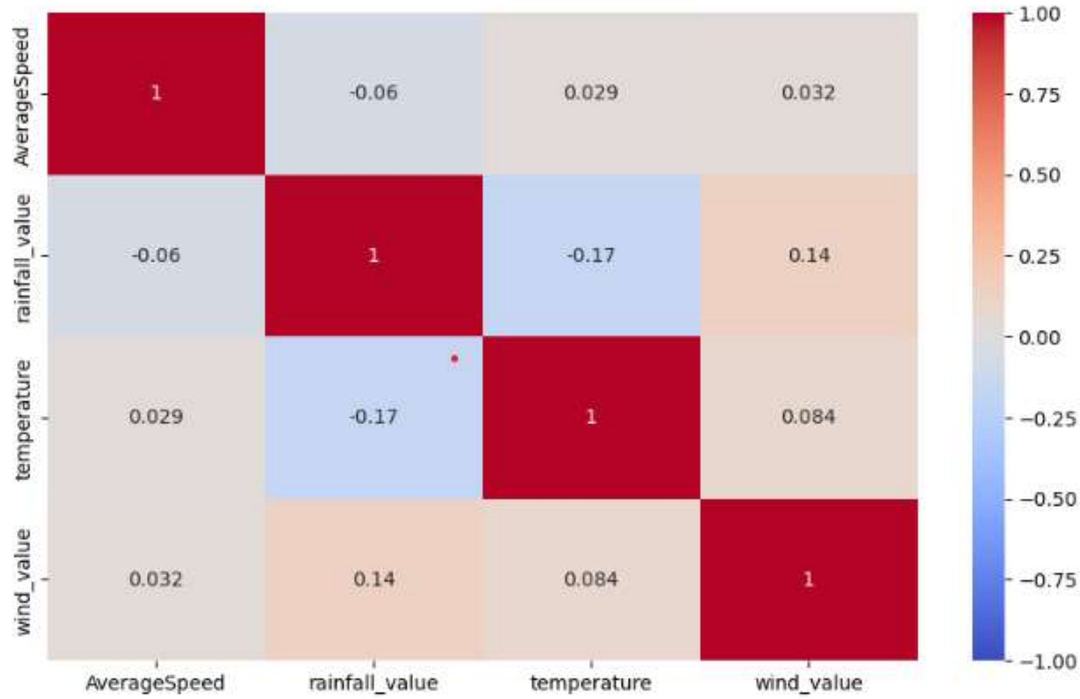

Figure 9.5: Correlation Matrix – Combined Analysis (A)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.060, indicating a weak negative relationship. This suggests that as rainfall increases, the average speed on the PIE may decrease slightly. The correlation between average speed and temperature is 0.029, suggesting a very weak positive relationship, while the correlation between average speed and wind value is 0.032, indicating another very weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) shows weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.006
Method:                 Least Squares   F-statistic:                     2465.
Date:                Fri, 07 Apr 2023   Prob (F-statistic):               0.00
Time:                        07:06:29   Log-Likelihood:            -4.5212e+06
No. Observations:             1156847   AIC:                         9.042e+06
Df Residuals:                 1156843   BIC:                         9.042e+06
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             70.5692      0.218    323.950      0.000      70.142      70.996
rainfall_value    -2.1944      0.033    -67.130      0.000      -2.258      -2.130
temperature        0.2344      0.008     28.206      0.000       0.218       0.251
wind_value         0.2260      0.005     41.390      0.000       0.215       0.237
==============================================================================
Omnibus:                   294408.023   Durbin-Watson:                   1.253
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           649193.623
Skew:                          -1.475   Prob(JB):                         0.00
Kurtosis:                       5.182   Cond. No.                         516.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.6: OLS Regression Results – Combined Analysis (A)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.006, indicating that only 0.6% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.006. These low R-squared values suggest that the model does not explain a substantial amount of the variation in average Speed.

The F-statistic is 2465, with a p-value (Prob (F-statistic)) of 0.00. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -2.1944, suggesting that for each unit increase in rainfall value, the average Speed decreases by approximately 2.19 units, holding the other variables constant. The coefficient for temperature is 0.2344, indicating that for each unit increase in temperature, the average speed increases

by approximately 0.23 units, holding the other variables constant. The coefficient for wind value is 0.2260, implying that for each unit increase in wind value, the average Speed increases by approximately 0.23 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.253, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 516, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.1.2. December Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	295779.000000	295779.000000	294699.000000	294643.000000
mean	77.597453	0.007029	24.970281	3.116993
std	11.451458	0.038263	0.773339	1.707235
min	4.500000	0.000000	22.900000	0.300000
25%	64.500000	0.000000	24.400000	1.800000
50%	85.000000	0.000000	24.900000	2.700000
75%	85.000000	0.000000	25.400000	4.300000
max	85.000000	0.402000	27.600000	11.200000

Figure 9.7: Descriptive Statistics – December Analysis (A)

Descriptive Statistics: The count row displays the number of data points for each variable, with 295,779 data points for average speed and rainfall value, 294,699 for temperature, and 294,643 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 77.60, rainfall value at 0.007, temperature at 24.97, and wind value at 3.12.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 11.45, for rainfall value it is 0.038, for temperature it is 0.77, and for wind value it is 1.71. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

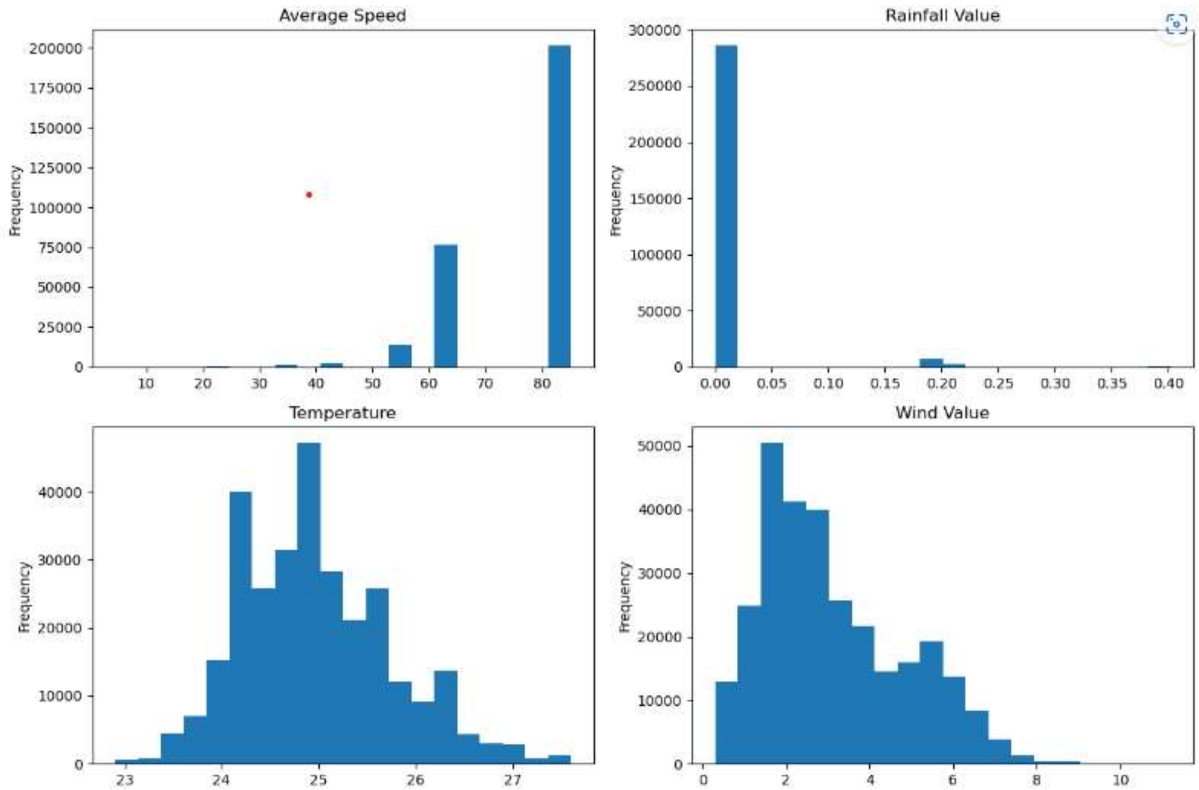

Figure 9.8: Frequencies of each data – December Analysis (A)

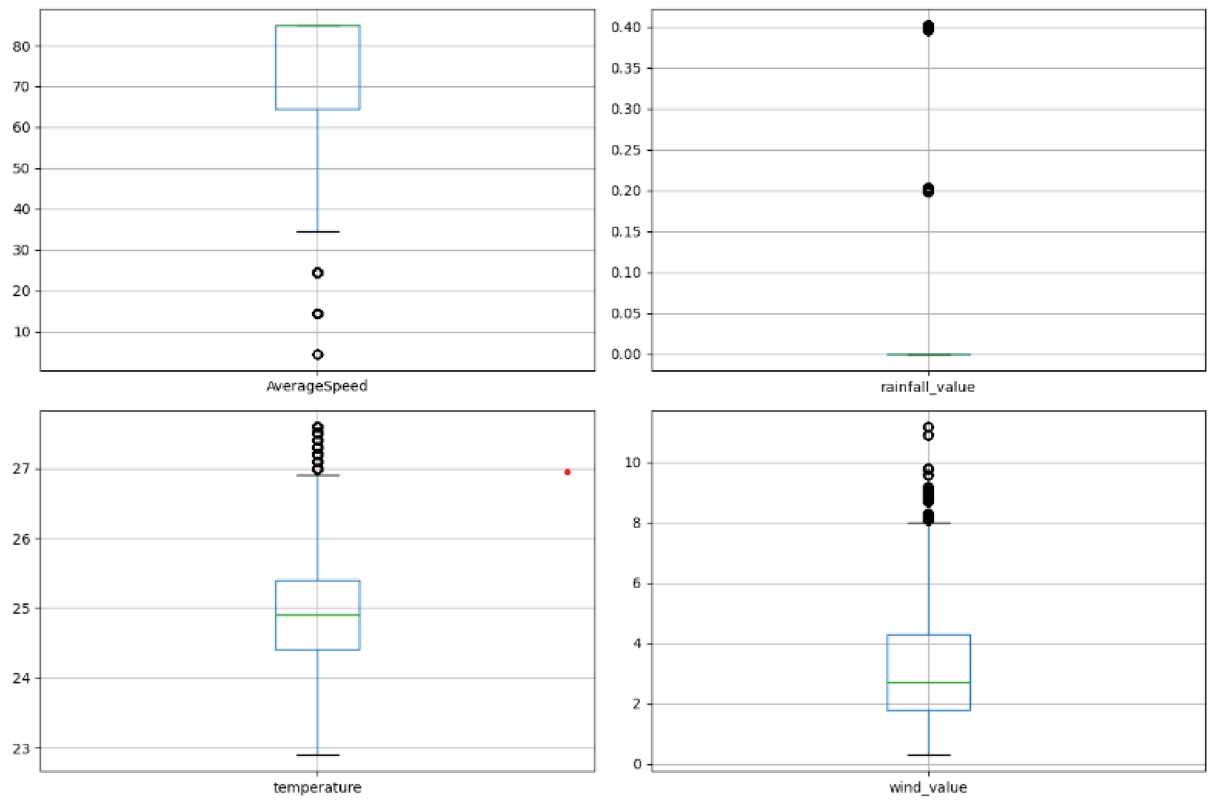

Figure 9.9: Boxplots of each data – December Analysis (A)

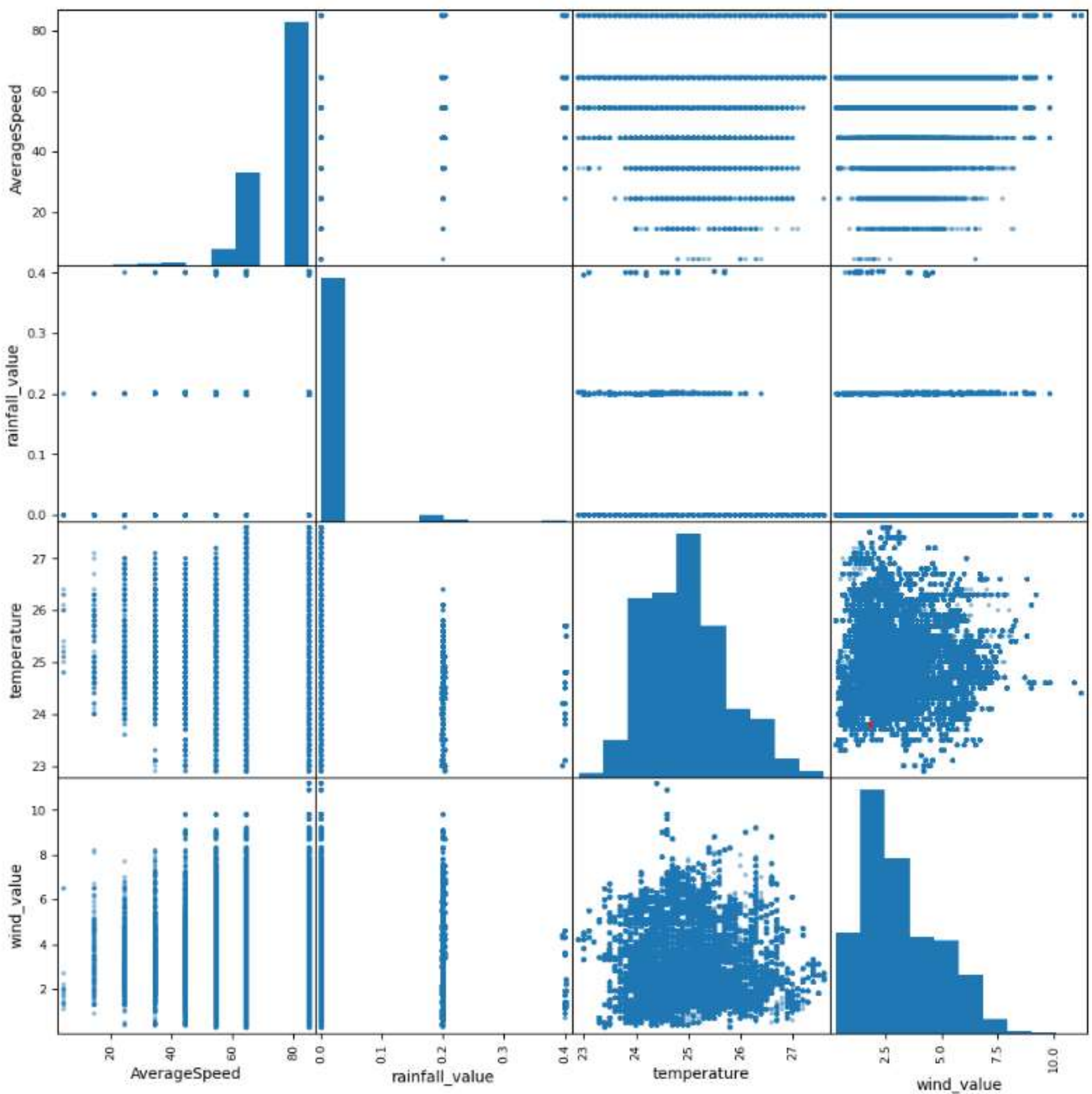

Figure 9.10: Scatter Matrix – December Analysis (A)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.114470	0.070383	0.039734
rainfall_value	-0.114470	1.000000	-0.114858	0.000807
temperature	0.070383	-0.114858	1.000000	-0.004751
wind_value	0.039734	0.000807	-0.004751	1.000000

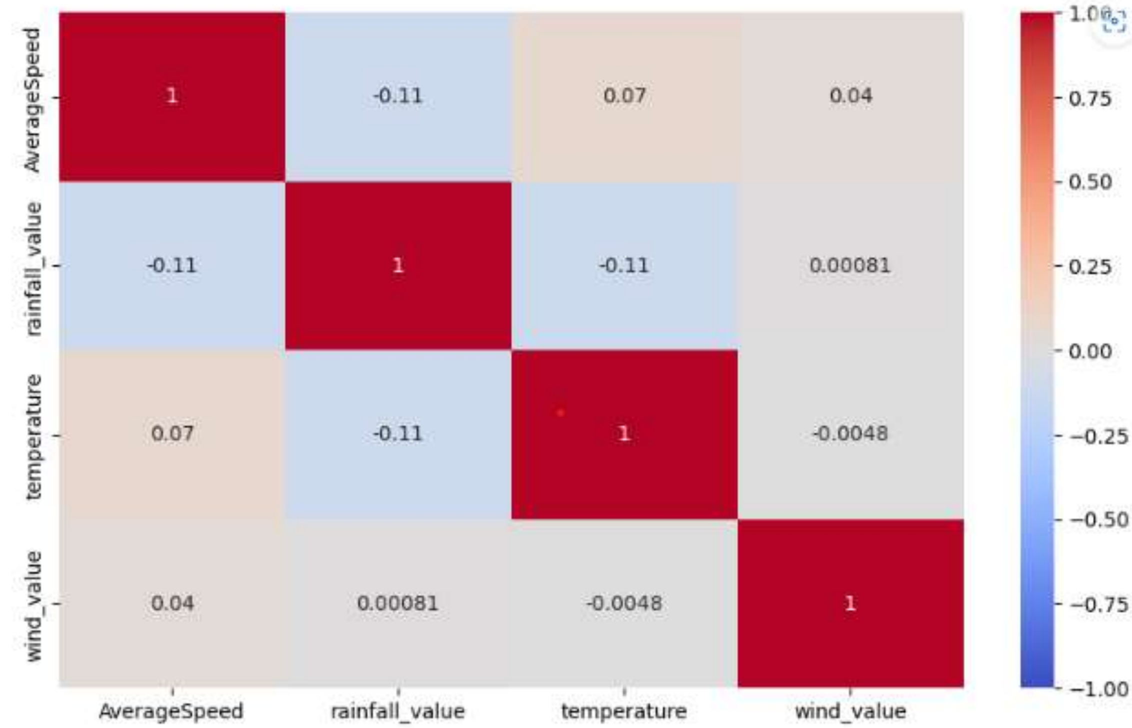

Figure 9.11: Correlation Matrix – December Analysis (A)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.114, indicating a weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is 0.070, suggesting a weak positive relationship, while the correlation between average speed and wind value is 0.040, indicating a very weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.018
Model:                            OLS   Adj. R-squared:                  0.018
Method:                 Least Squares   F-statistic:                     1825.
Date:                Fri, 07 Apr 2023   Prob (F-statistic):               0.00
Time:                        08:55:05   Log-Likelihood:            -1.1339e+06
No. Observations:              294643   AIC:                         2.268e+06
Df Residuals:                  294639   BIC:                         2.268e+06
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             55.5079      0.682     81.416      0.000      54.172      56.844
rainfall_value   -32.6082      0.551    -59.230      0.000     -33.687     -31.529
temperature        0.8600      0.027     31.591      0.000       0.807       0.913
wind_value         0.2691      0.012     21.966      0.000       0.245       0.293
==============================================================================
Omnibus:                    61191.839   Durbin-Watson:                   1.415
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           112184.662
Skew:                          -1.317   Prob(JB):                         0.00
Kurtosis:                       4.485   Cond. No.                         832.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.12: OLS Regression Results – December Analysis (A)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.018, indicating that only 1.8% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.018. These low R-squared values suggest that the model does not explain a substantial amount of the variation in average speed.

The F-statistic is 1825, with a p-value (Prob (F-statistic)) of 0.00. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average Speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -32.6082, suggesting that for each unit increase in rainfall value, the average Speed decreases by approximately 32.61 units, holding the other variables constant. The coefficient

for temperature is 0.8600, indicating that for each unit increase in temperature, the average speed increases by approximately 0.86 units, holding the other variables constant. The coefficient for wind value is 0.2691, implying that for each unit increase in wind value, the average speed increases by approximately 0.27 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.415, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 832, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.1.3. September Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	320994.000000	320994.000000	316469.000000	315043.000000
mean	77.914372	0.069783	26.627800	3.175052
std	11.482980	0.514011	1.028593	2.458434
min	4.500000	0.000000	21.700000	0.300000
25%	64.500000	0.000000	26.100000	1.500000
50%	85.000000	0.000000	26.700000	2.300000
75%	85.000000	0.000000	27.200000	4.800000
max	85.000000	10.400000	29.400000	24.500000

Figure 9.13: Descriptive Statistics – September Analysis (A)

The count row displays the number of data points for each variable, with 320,994 data points for average speed, rainfall value, 316,469 for temperature, and 315,043 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 77.91, rainfall value at 0.069, temperature at 26.63, and wind value at 3.18.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 11.48, for rainfall value it is 0.514, for temperature it is 1.03, and for wind value it is 2.46. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

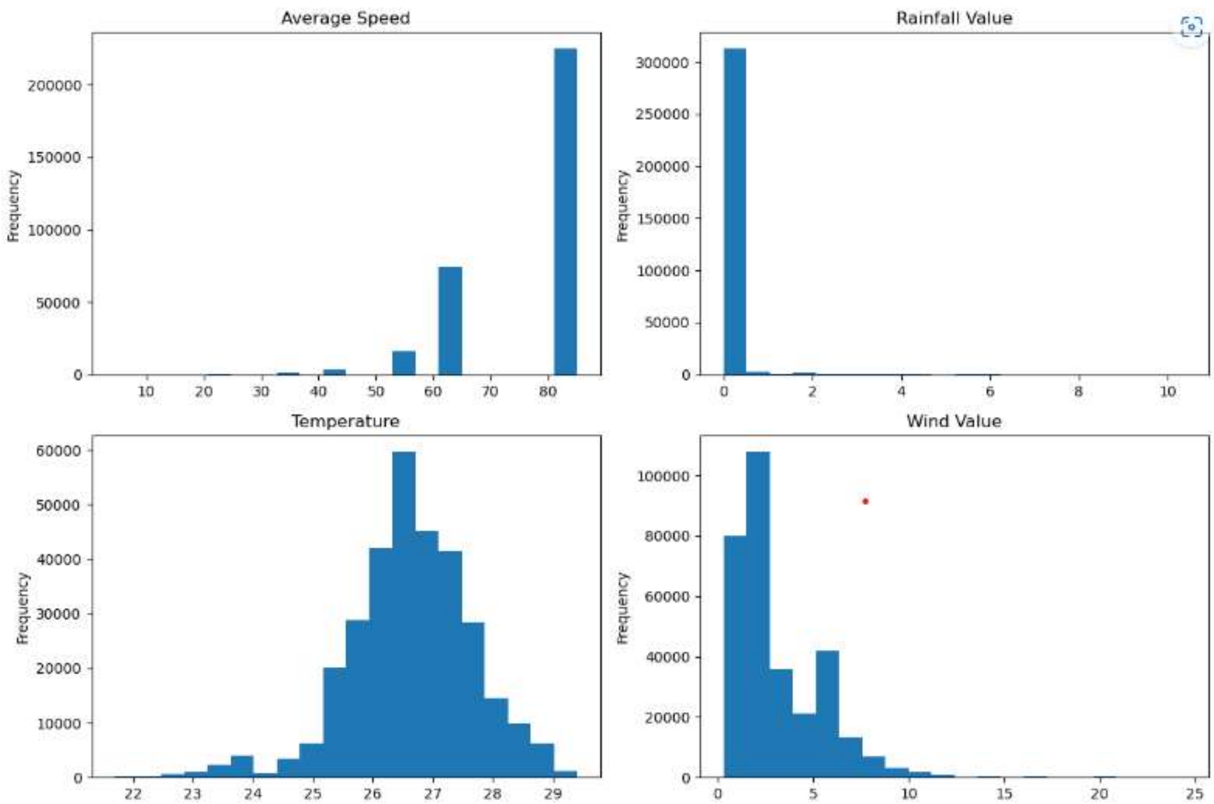

Figure 9.14: Frequencies of each data – September Analysis (A)

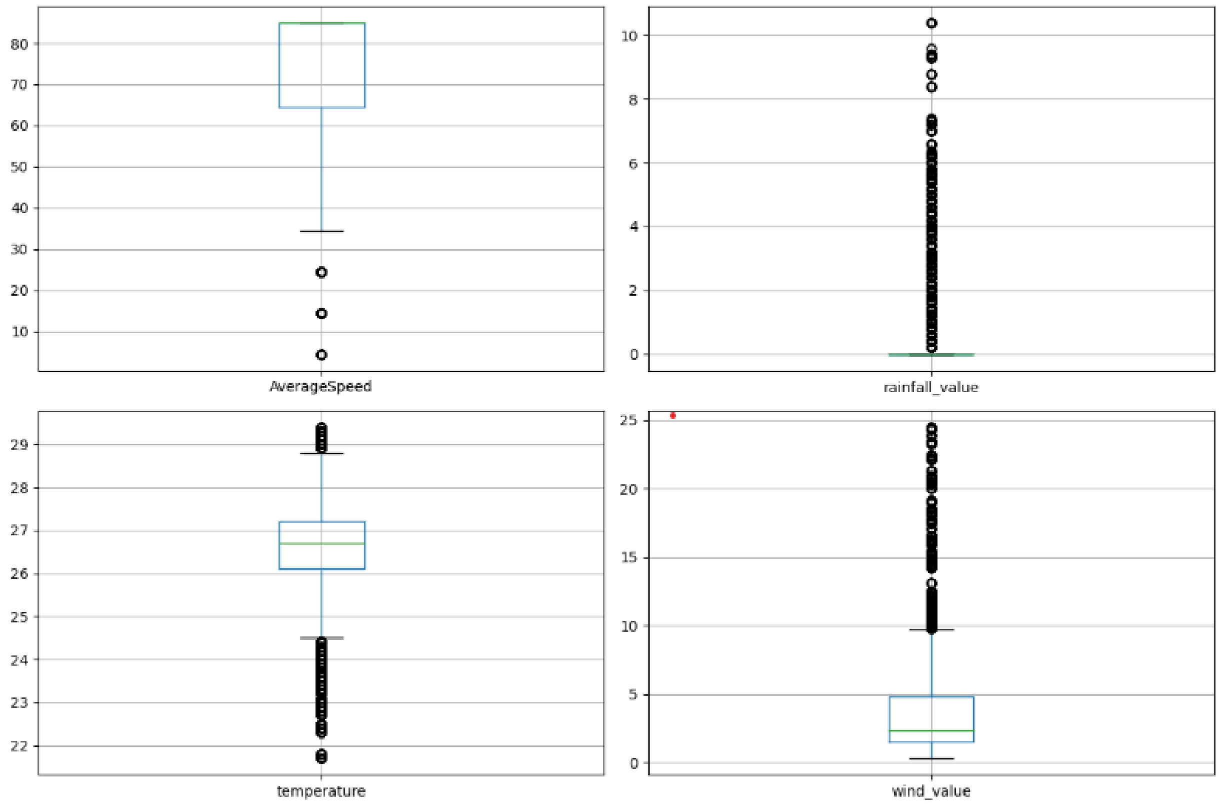

Figure 9.15: Boxplots of each data – September Analysis (A)

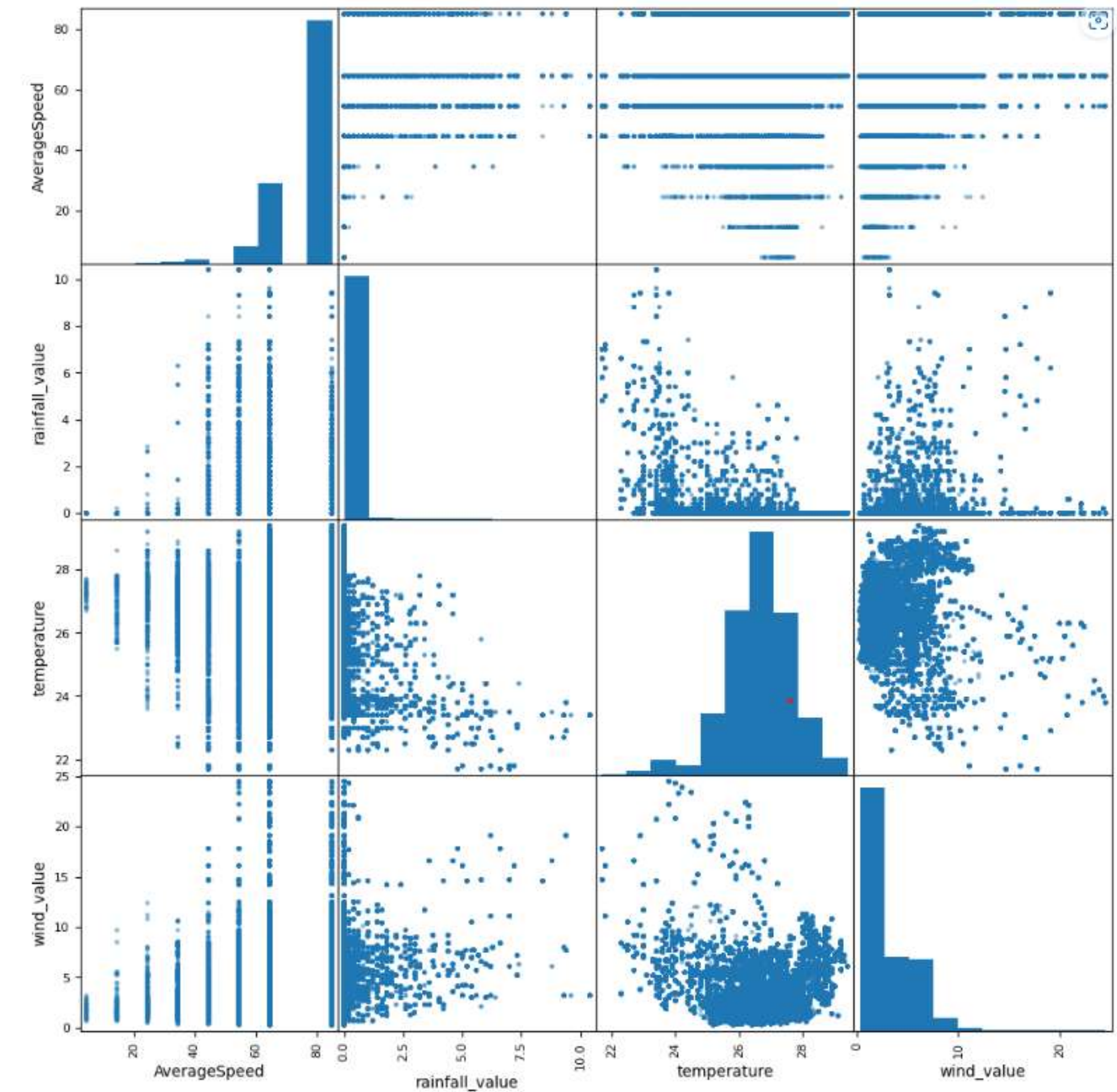

Figure 9.16: Scatter Matrix – September Analysis (A)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.132533	0.139912	0.020425
rainfall_value	-0.132533	1.000000	-0.346370	0.191920
temperature	0.139912	-0.346370	1.000000	-0.114775
wind_value	0.020425	0.191920	-0.114775	1.000000

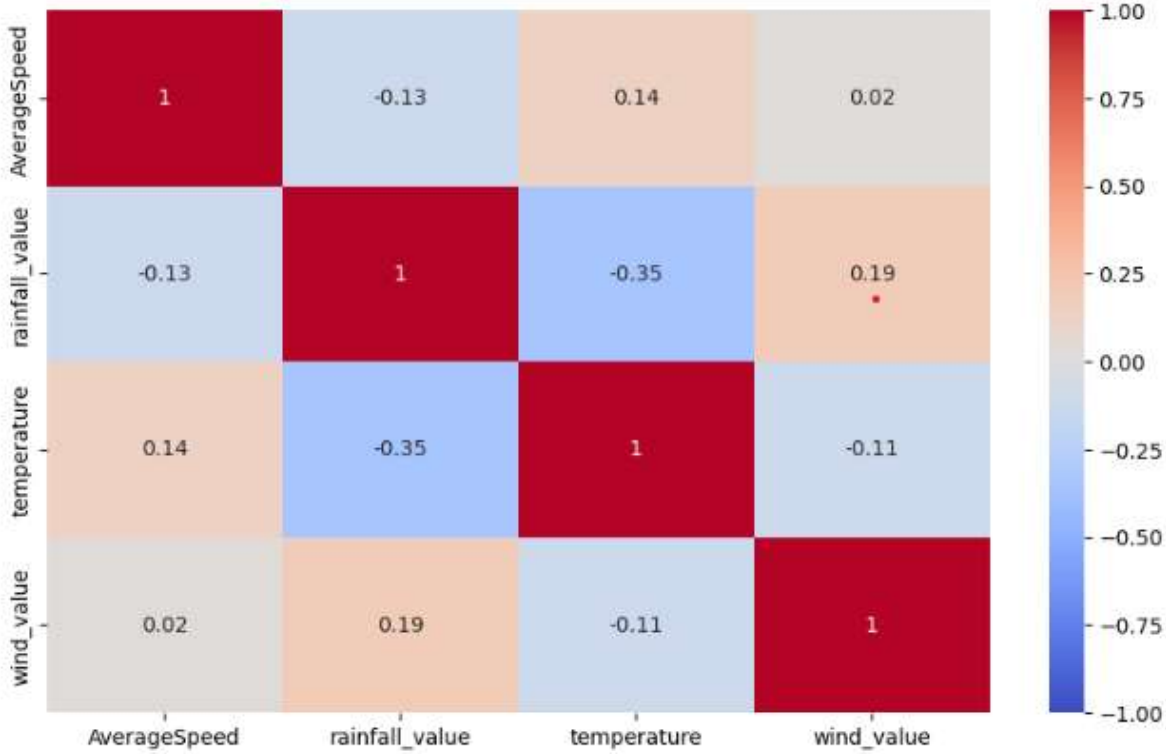

Figure 9.17: Correlation Matrix – September Analysis (A)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.133, indicating a weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is 0.140, suggesting a weak positive relationship, while the correlation between average speed and wind value is 0.020, indicating a very weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.030
Model:                            OLS   Adj. R-squared:                  0.030
Method:                 Least Squares   F-statistic:                     3232.
Date:                Fri, 07 Apr 2023   Prob (F-statistic):               0.00
Time:                        09:43:00   Log-Likelihood:            -1.2121e+06
No. Observations:              315043   AIC:                         2.424e+06
Df Residuals:                  315039   BIC:                         2.424e+06
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             45.2236      0.566     79.890      0.000      44.114      46.333
rainfall_value    -2.3531      0.042    -55.852      0.000      -2.436      -2.270
temperature        1.2032      0.021     56.917      0.000       1.162       1.245
wind_value         0.2483      0.008     29.607      0.000       0.232       0.265
==============================================================================
Omnibus:                    70358.377   Durbin-Watson:                   1.399
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           134206.789
Skew:                          -1.389   Prob(JB):                         0.00
Kurtosis:                       4.585   Cond. No.                         753.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.18: OLS Regression Results – September Analysis (A)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.030, indicating that only 3% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.030. These R-squared values suggest that the model does not explain a substantial amount of the variation in average speed.

The F-statistic is 3232, with a p-value (Prob (F-statistic)) of 0.00. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -2.3531, suggesting that for each unit increase in rainfall value, the average speed decreases by approximately 2.35 units, holding the other variables constant. The coefficient

for temperature is 1.2032, indicating that for each unit increase in temperature, the average speed increases by approximately 1.20 units, holding the other variables constant. The coefficient for wind value is 0.2483, implying that for each unit increase in wind value, the average speed increases by approximately 0.25 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.399, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 753, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.2. Road Category B

9.2.1. Combined Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	14336.000000	14336.000000	14202.000000	12871.000000
mean	50.535261	0.054269	26.039199	2.201453
std	17.988071	0.435887	1.228795	1.461527
min	4.500000	0.000000	22.300000	0.300000
25%	34.500000	0.000000	25.100000	1.100000
50%	54.500000	0.000000	26.200000	1.900000
75%	64.500000	0.000000	27.000000	2.900000
max	85.000000	10.400000	29.800000	19.100000

Figure 9.19: Descriptive Statistics – Combined Analysis (B)

The count row displays the number of data points for each variable, with 14,336 data points for average speed and rainfall value, 14,202 for temperature, and 12,871 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 50.54, rainfall value at 0.054, temperature at 26.04, and wind value at 2.20.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 17.99, for rainfall value it is 0.44, for temperature it is 1.23, and for wind value it is 1.46. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

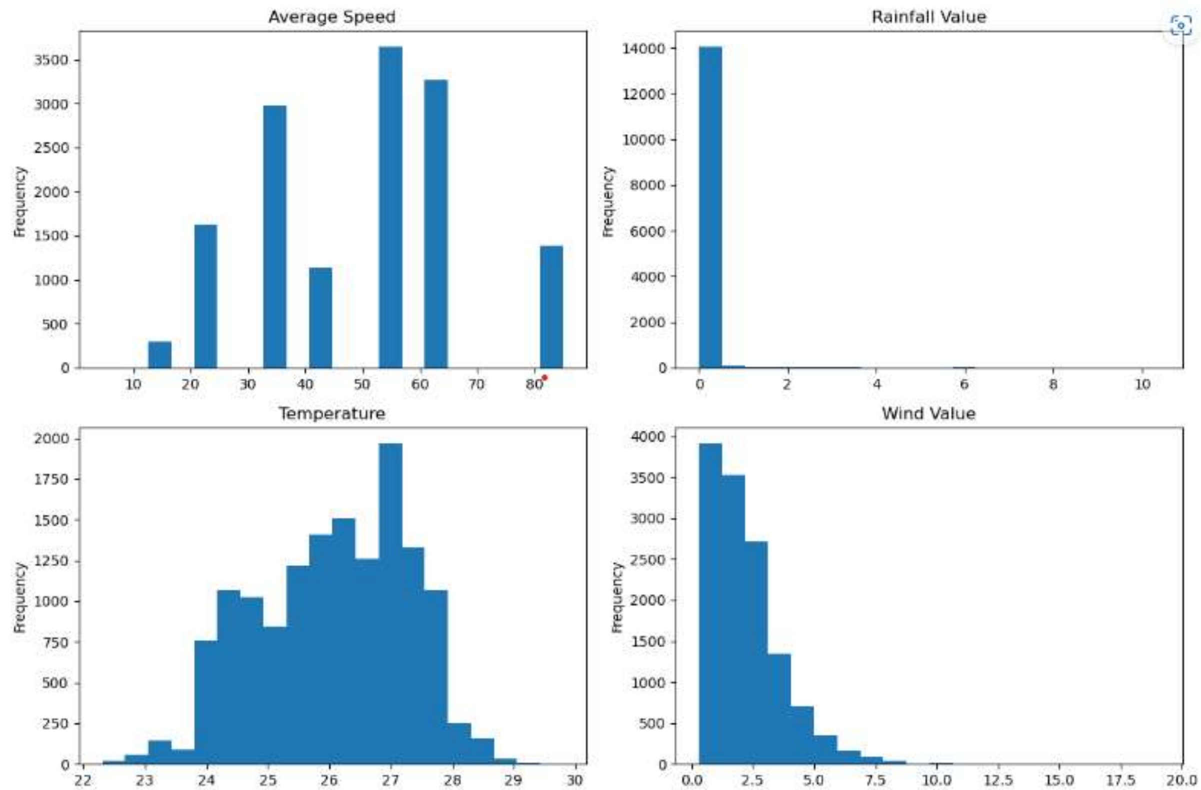

Figure 9.20: Frequencies of each data – Combined Analysis (B)

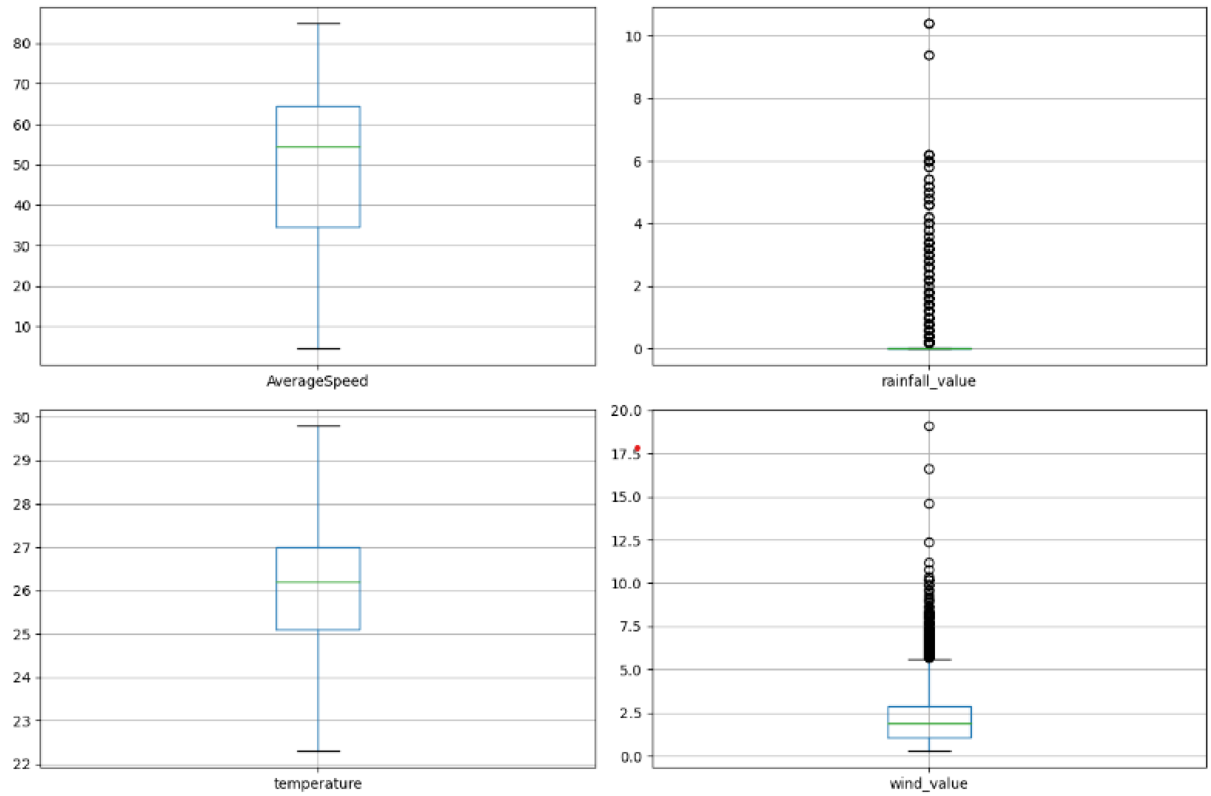

Figure 9.21: Boxplots of each data – Combined Analysis (B)

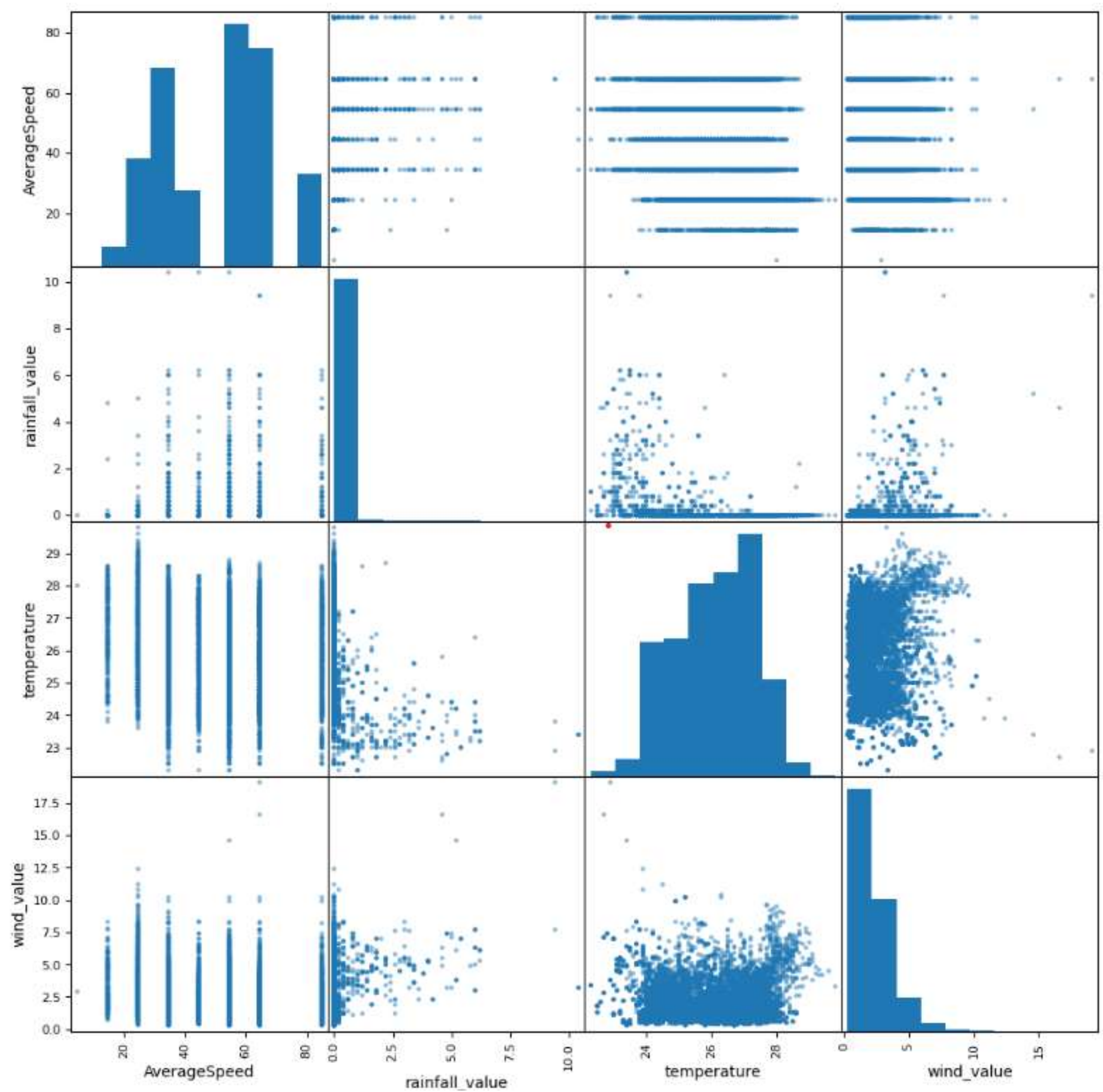

Figure 9.22: Scatter Matrix – Combined Analysis (B)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	0.024601	-0.108307	-0.200165
rainfall_value	0.024601	1.000000	-0.223591	0.240270
temperature	-0.108307	-0.223591	1.000000	0.062291
wind_value	-0.200165	0.240270	0.062291	1.000000

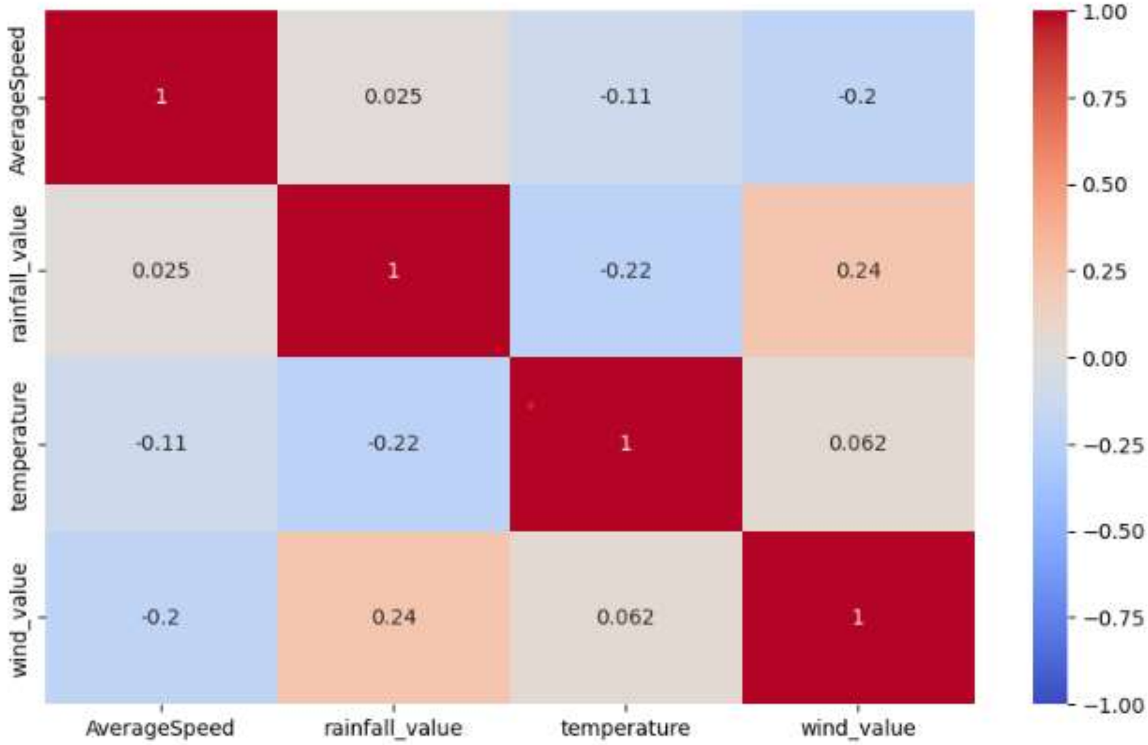

Figure 9.23: Correlation Matrix – Combined Analysis (B)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is 0.025, indicating a weak positive relationship. This suggests that as rainfall increases, the average speed on the PIE may slightly increase. The correlation between average speed and temperature is -0.108, suggesting a weak negative relationship, while the correlation between average Speed and wind value is -0.200, indicating a weak to moderate negative relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak to moderate correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.054
Model:                            OLS   Adj. R-squared:                  0.054
Method:                 Least Squares   F-statistic:                     243.5
Date:                Fri, 07 Apr 2023   Prob (F-statistic):          1.09e-153
Time:                        07:47:10   Log-Likelihood:                -55505.
No. Observations:               12871   AIC:                         1.110e+05
Df Residuals:                   12867   BIC:                         1.110e+05
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             90.5089      3.470     26.086      0.000      83.708      97.310
rainfall_value     2.6221      0.391      6.708      0.000       1.856       3.388
temperature       -1.3257      0.134     -9.897      0.000      -1.588      -1.063
wind_value        -2.6592      0.113    -23.526      0.000      -2.881      -2.438
==============================================================================
Omnibus:                      897.543   Durbin-Watson:                   1.608
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              408.654
Skew:                           0.241   Prob(JB):                     1.83e-89
Kurtosis:                       2.272   Cond. No.                         570.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.24: OLS Regression Results – Combined Analysis (B)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.054, indicating that 5.4% of the variation in average Speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.054. These low R-squared values suggest that the model does not explain a substantial amount of the variation in average Speed.

The F-statistic is 243.5, with a p-value (Prob (F-statistic)) of 1.09e-153. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is 2.6221, suggesting that for each unit increase in rainfall value, the average speed increases by approximately 2.62 units, holding the other variables constant. The coefficient for temperature is -1.3257, indicating that for each unit increase in temperature, the average speed decreases by approximately 1.33 units, holding the other variables constant. The coefficient for wind value is -2.6592, implying that for each unit increase in wind value, the average speed decreases by approximately 2.66 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.608, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 1.83e-89, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 570, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average Speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.2.2. December Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	2975.000000	2975.000000	2975.000000	2975.000000
mean	49.244706	0.007462	24.706857	2.279697
std	16.934879	0.038961	0.605289	1.172728
min	14.500000	0.000000	23.300000	0.300000
25%	34.500000	0.000000	24.200000	1.400000
50%	54.500000	0.000000	24.700000	2.200000
75%	64.500000	0.000000	25.100000	3.000000
max	85.000000	0.400000	27.000000	8.200000

Figure 9.25: Descriptive Statistics – December Analysis (B)

The count row displays the number of data points for each variable, with 2,975 data points for average speed, rainfall value, temperature, and wind value. The mean row shows the average value for each variable, with average speed having a mean of 49.24, rainfall value at 0.007, temperature at 24.71, and wind value at 2.28.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 16.93, for rainfall value it is 0.039, for temperature it is 0.61, and for wind value it is 1.17. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

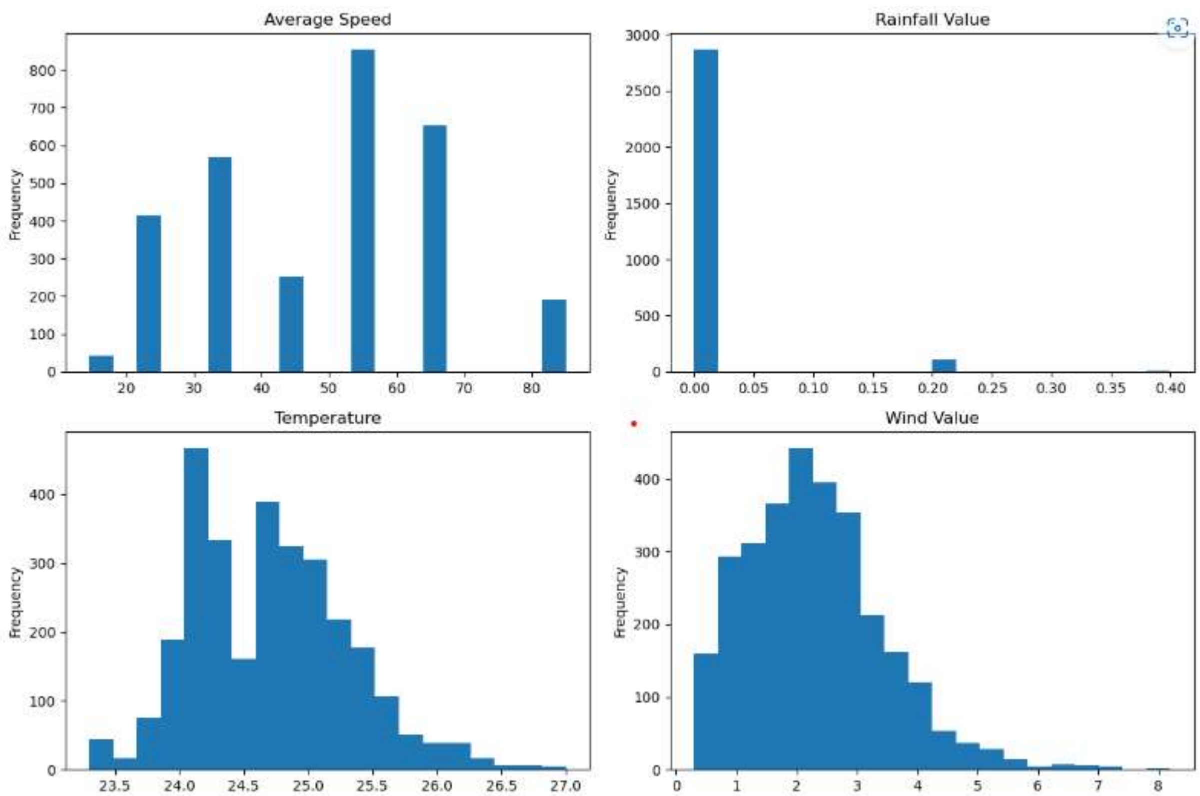

Figure 9.26: Frequencies of each data – December Analysis (B)

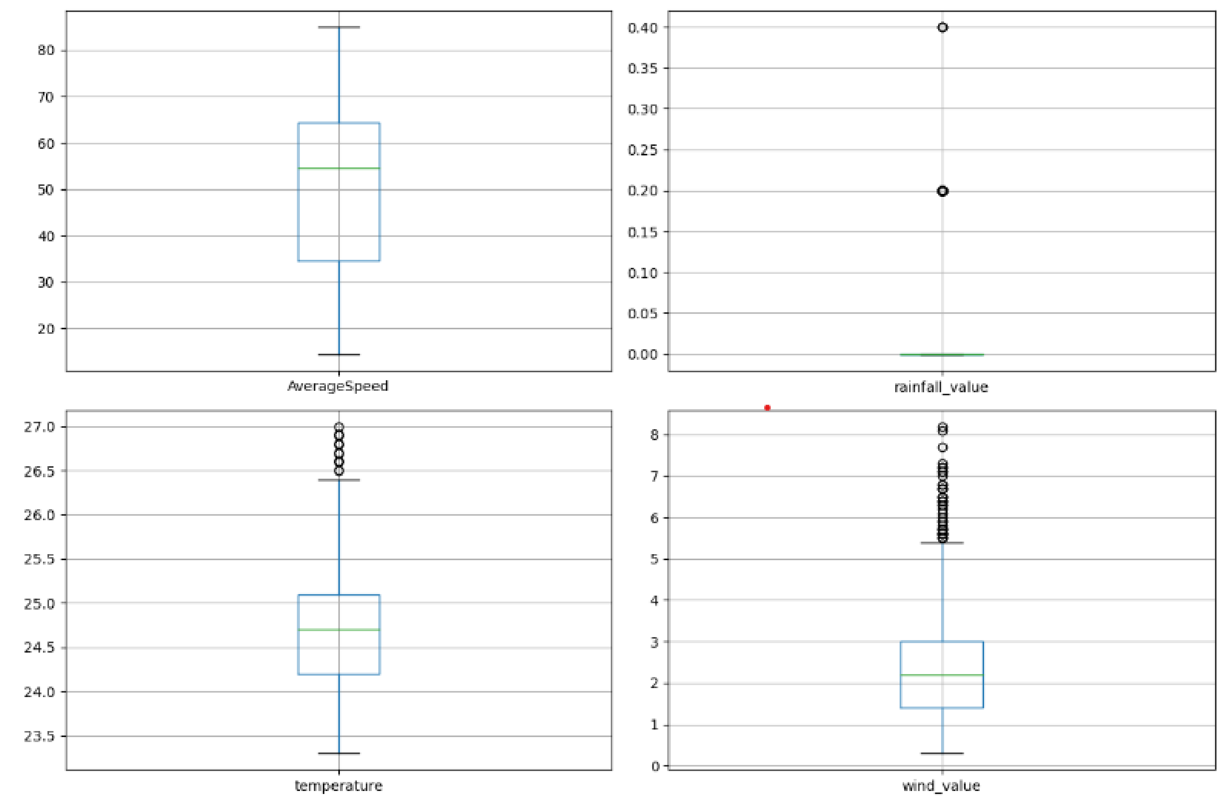

Figure 9.27: Boxplots of each data – December Analysis (B)

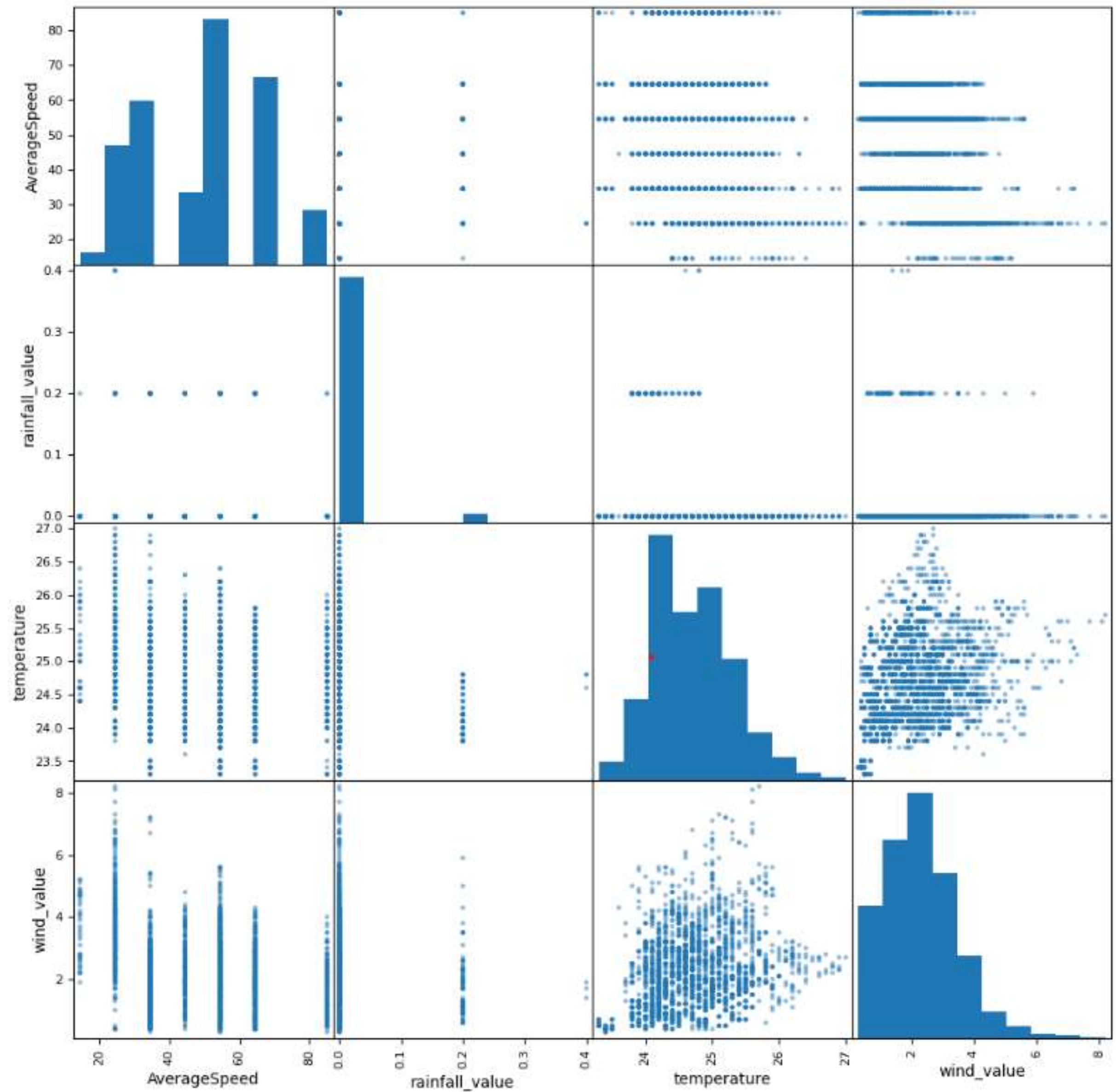

Figure 9.28: Scatter Matrix – December Analysis (B)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.040328	-0.208835	-0.348334
rainfall_value	-0.040328	1.000000	-0.168138	-0.086319
temperature	-0.208835	-0.168138	1.000000	0.249241
wind_value	-0.348334	-0.086319	0.249241	1.000000

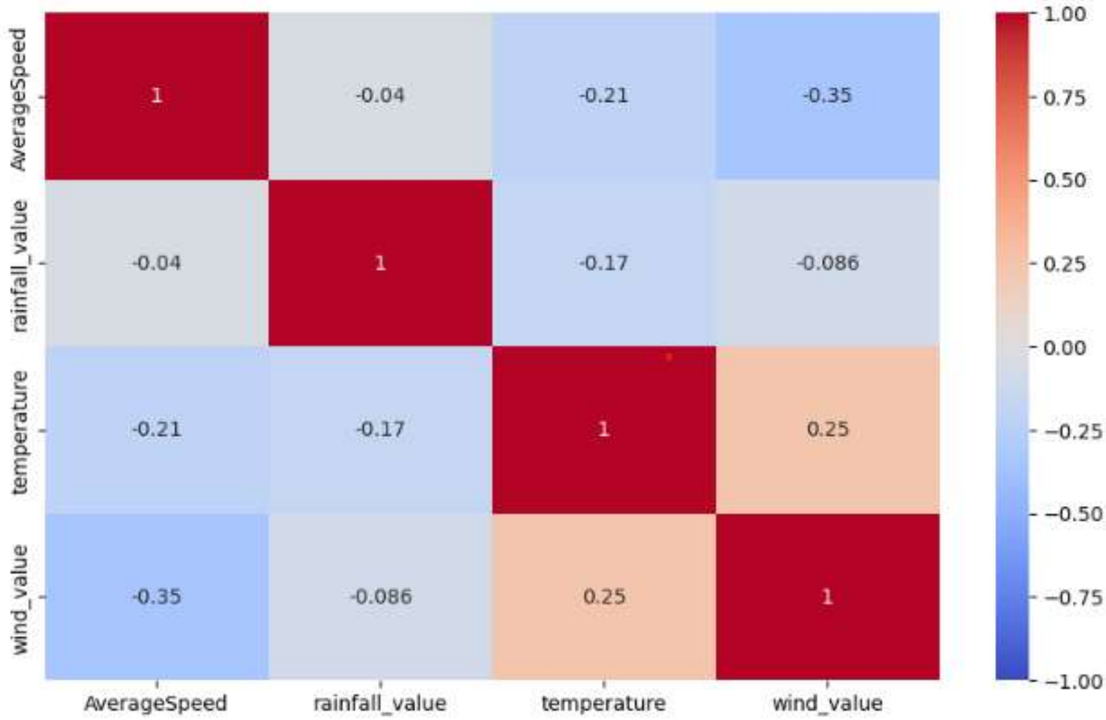

Figure 9.29: Correlation Matrix – December Analysis (B)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.040, indicating a very weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is -0.209, suggesting a weak negative relationship, while the correlation between average speed and wind value is -0.348, indicating a weak negative relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.145
Model:                            OLS   Adj. R-squared:                  0.145
Method:                 Least Squares   F-statistic:                     168.6
Date:                Fri, 07 Apr 2023   Prob (F-statistic):          6.37e-101
Time:                        09:11:42   Log-Likelihood:                -12404.
No. Observations:                2975   AIC:                         2.482e+04
Df Residuals:                    2971   BIC:                         2.484e+04
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const            159.9898     12.135     13.184      0.000     136.196     183.783
rainfall_value   -40.1083      7.486     -5.358      0.000     -54.787     -25.429
temperature       -4.0435      0.496     -8.157      0.000      -5.015      -3.072
wind_value        -4.6250      0.253    -18.269      0.000      -5.121      -4.129
==============================================================================
Omnibus:                       67.880   Durbin-Watson:                   1.214
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               48.421
Skew:                           0.207   Prob(JB):                     3.06e-11
Kurtosis:                       2.531   Cond. No.                     1.06e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.06e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 9.30: OLS Regression Results – December Analysis (B)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.145, indicating that 14.5% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.145. These R-squared values suggest that the model explains a moderate amount of the variation in average speed.

The F-statistic is 168.6, with a p-value (Prob (F-statistic)) of 6.37e-101. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average Speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -40.1083, suggesting that for each unit increase in rainfall value, the average Speed decreases by approximately 40.11 units, holding the other variables constant. The coefficient for temperature is -4.0435, indicating that for each unit increase in temperature, the average speed decreases by approximately 4.04 units, holding the other variables constant. The coefficient for wind value is -4.6250, implying that for each unit increase in wind value, the average speed decreases by approximately 4.63 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.214, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 3.06e-11, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 1.06e+03, which is relatively high, suggesting that there might be strong multicollinearity or other numerical problems in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a moderate portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.2.3. September Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	3216.000000	3216.000000	3130.000000	3129.000000
mean	51.137593	0.102736	26.360511	2.121093
std	16.403579	0.688546	0.831318	1.554684
min	14.500000	0.000000	22.300000	0.300000
25%	34.500000	0.000000	26.100000	1.000000
50%	54.500000	0.000000	26.600000	1.700000
75%	64.500000	0.000000	26.875000	2.700000
max	85.000000	10.400000	28.600000	19.100000

Figure 9.31: Descriptive Statistics – September Analysis (B)

The count row displays the number of data points for each variable, with 3,216 data points for average speed and rainfall value, 3,130 for temperature, and 3,129 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 51.14, rainfall value at 0.103, temperature at 26.36, and wind value at 2.12.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 16.40, for rainfall value it is 0.689, for temperature it is 0.83, and for wind value it is 1.55. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

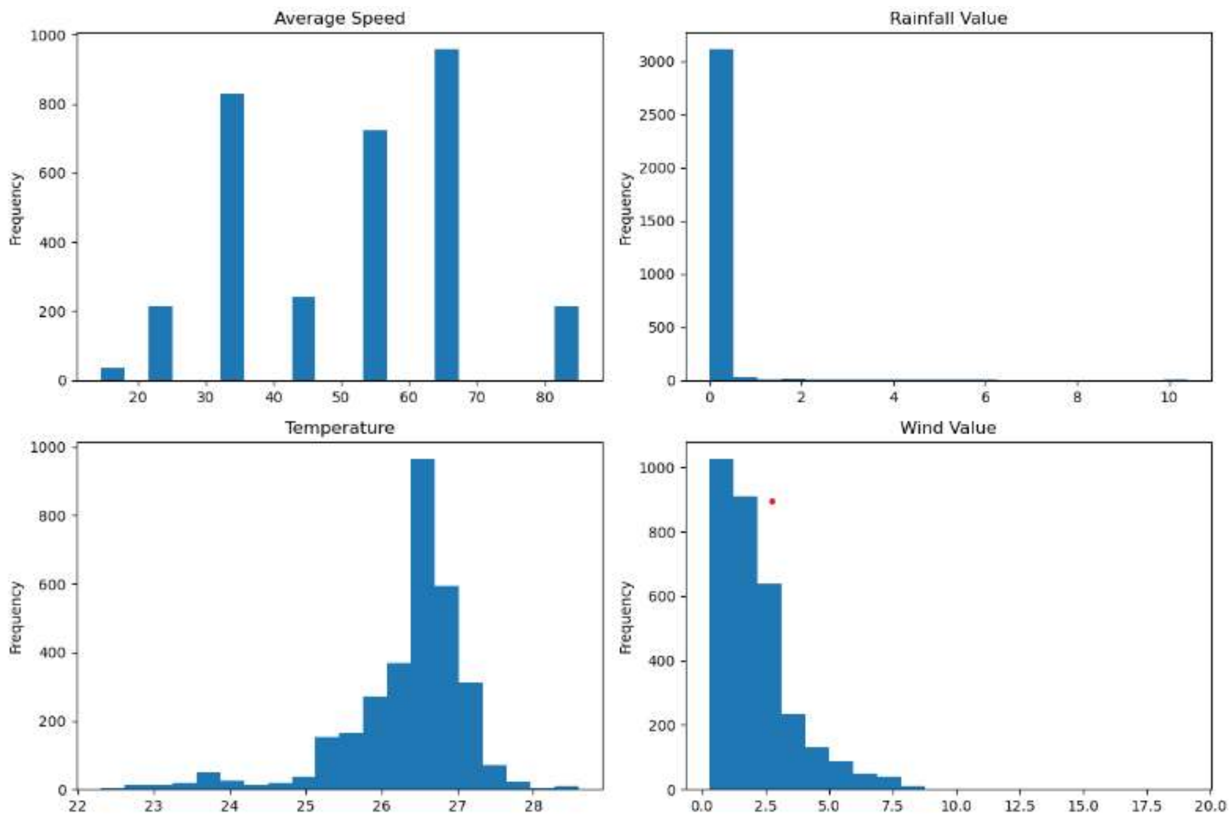

Figure 9.32: Frequencies of each data – September Analysis (B)

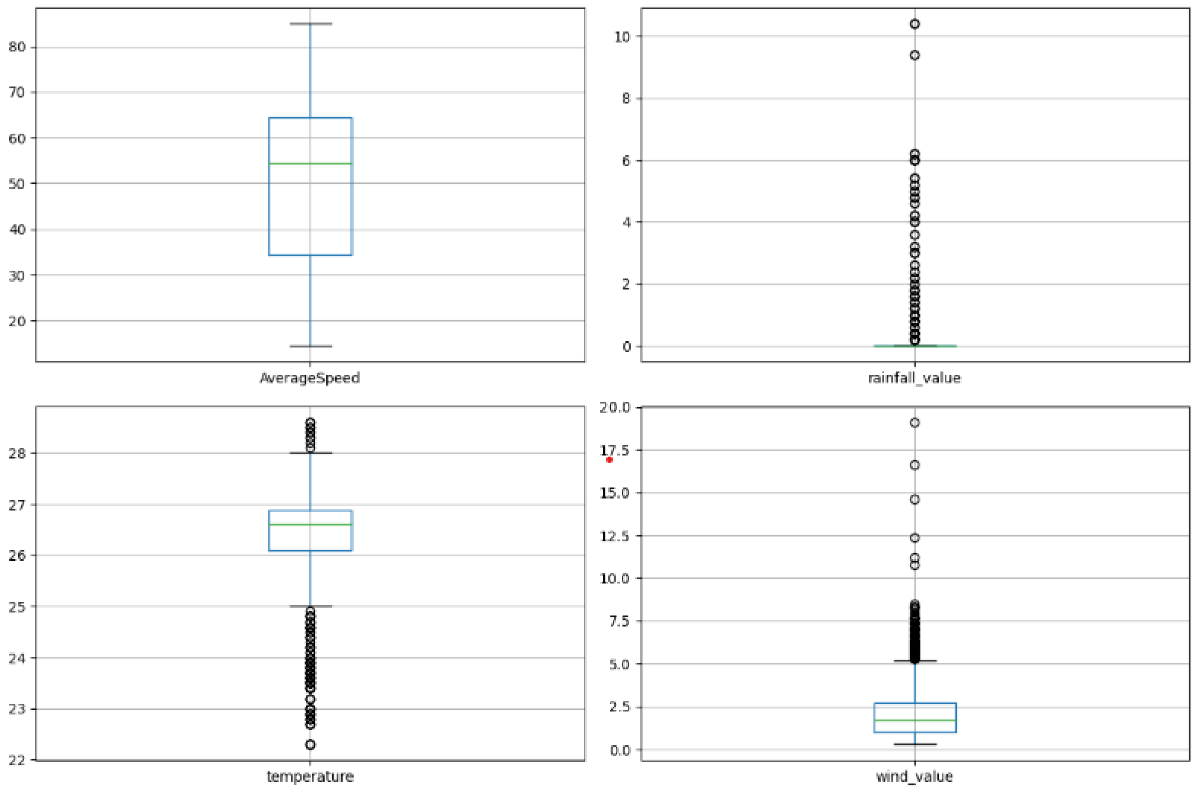

Figure 9.33: Boxplots of each data – September Analysis (B)

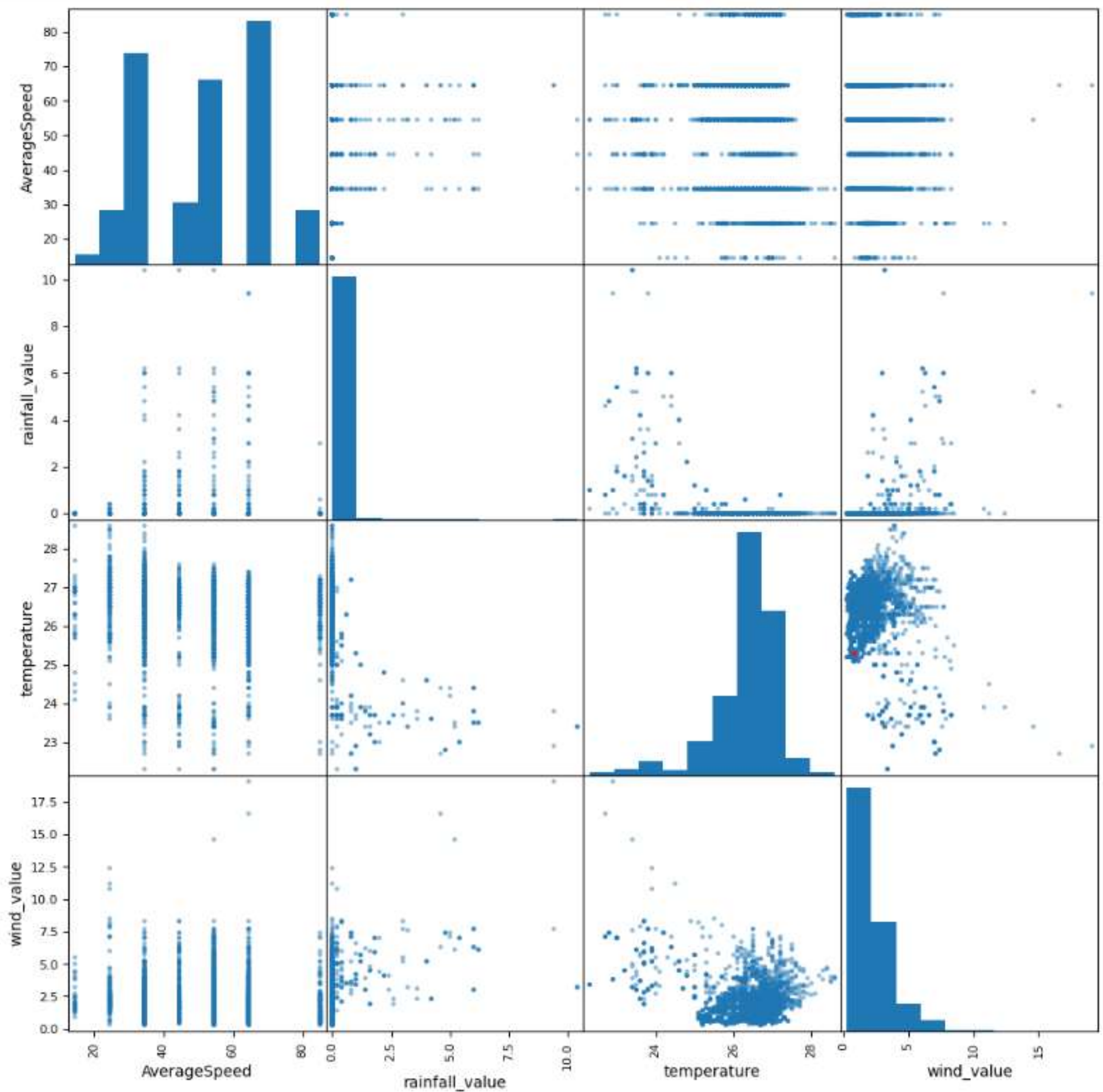

Figure 9.34: Scatter Matrix – September Analysis (B)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.001574	-0.067029	-0.068108
rainfall_value	-0.001574	1.000000	-0.480896	0.367386
temperature	-0.067029	-0.480896	1.000000	-0.264470
wind_value	-0.068108	0.367386	-0.264470	1.000000

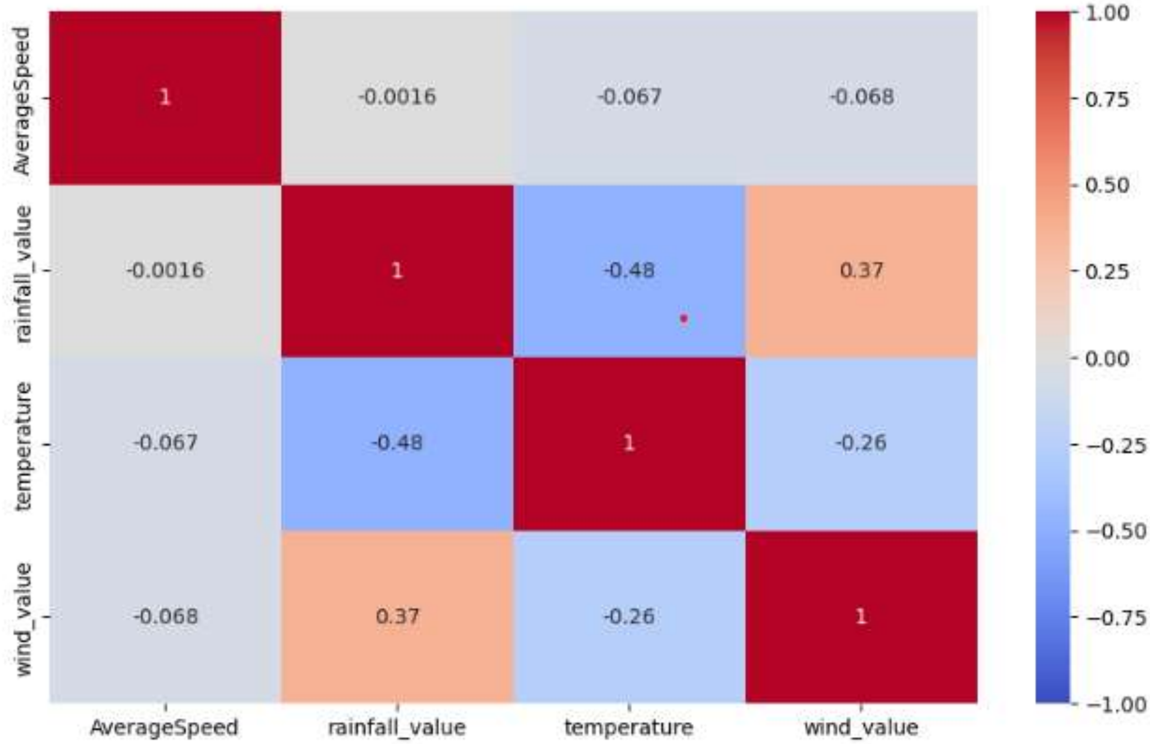

Figure 9.35: Correlation Matrix – September Analysis (B)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.001574, indicating a very weak negative relationship. This suggests that as rainfall increases, the average speed may decrease very slightly. The correlation between average speed and temperature is -0.067029, suggesting a weak negative relationship, while the correlation between average speed and wind value is -0.068108, indicating a weak negative relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak to moderate correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.013
Model:                            OLS   Adj. R-squared:                  0.012
Method:                 Least Squares   F-statistic:                     13.25
Date:                Fri, 07 Apr 2023   Prob (F-statistic):           1.37e-08
Time:                        10:03:46   Log-Likelihood:                -13181.
No. Observations:                3129   AIC:                         2.637e+04
Df Residuals:                    3125   BIC:                         2.639e+04
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const            103.7900     10.716      9.685      0.000      82.779     124.801
rainfall_value    -0.3418      0.498     -0.686      0.493      -1.319       0.635
temperature       -1.9264      0.403     -4.775      0.000      -2.717      -1.135
wind_value        -0.9366      0.203     -4.605      0.000      -1.335      -0.538
==============================================================================
Omnibus:                      129.413   Durbin-Watson:                   1.350
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               58.089
Skew:                           0.084   Prob(JB):                     2.43e-13
Kurtosis:                       2.354   Cond. No.                         971.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.36: OLS Regression Results – September Analysis (B)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.013, indicating that only 1.3% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is 0.012. These R-squared values suggest that the model does not explain a substantial amount of the variation in average Speed.

The F-statistic is 13.25, with a p-value (Prob (F-statistic)) of 1.37e-08. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. Not all independent variables have statistically significant effects on average speed, as the p-value for rainfall value is greater than 0.05.

The coefficient for rainfall value is -0.3418, suggesting that for each unit increase in rainfall value, the average speed decreases by approximately 0.34 units, holding the other variables constant. However, this effect is not statistically significant (p-value = 0.493). The coefficient for temperature is -1.9264, indicating that for each unit increase in temperature, the average speed decreases by approximately 1.93 units, holding the other variables constant. The coefficient for wind value is -0.9366, implying that for each unit increase in wind value, the average speed decreases by approximately 0.94 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.350, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 2.43e-13, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 971, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that temperature and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.3. Road Category F

9.3.1. Combined Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	554203.000000	554203.000000	543553.000000	432932.000000
mean	56.005814	0.039492	26.336877	3.012557
std	16.744794	0.342657	1.362197	2.106239
min	4.500000	0.000000	21.700000	0.300000
25%	44.500000	0.000000	25.300000	1.600000
50%	54.500000	0.000000	26.400000	2.400000
75%	64.500000	0.000000	27.400000	3.900000
max	85.000000	10.400000	30.200000	24.500000

Figure 9.37: Descriptive Statistics – Combined Analysis (F)

The count row displays the number of data points for each variable, with 554,203 data points for average speed and rainfall value, 543,553 for temperature, and 432,932 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 56.01, rainfall value at 0.039, temperature at 26.34, and wind value at 3.01.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 16.74, for rainfall value it is 0.34, for temperature it is 1.36, and for wind value it is 2.11. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

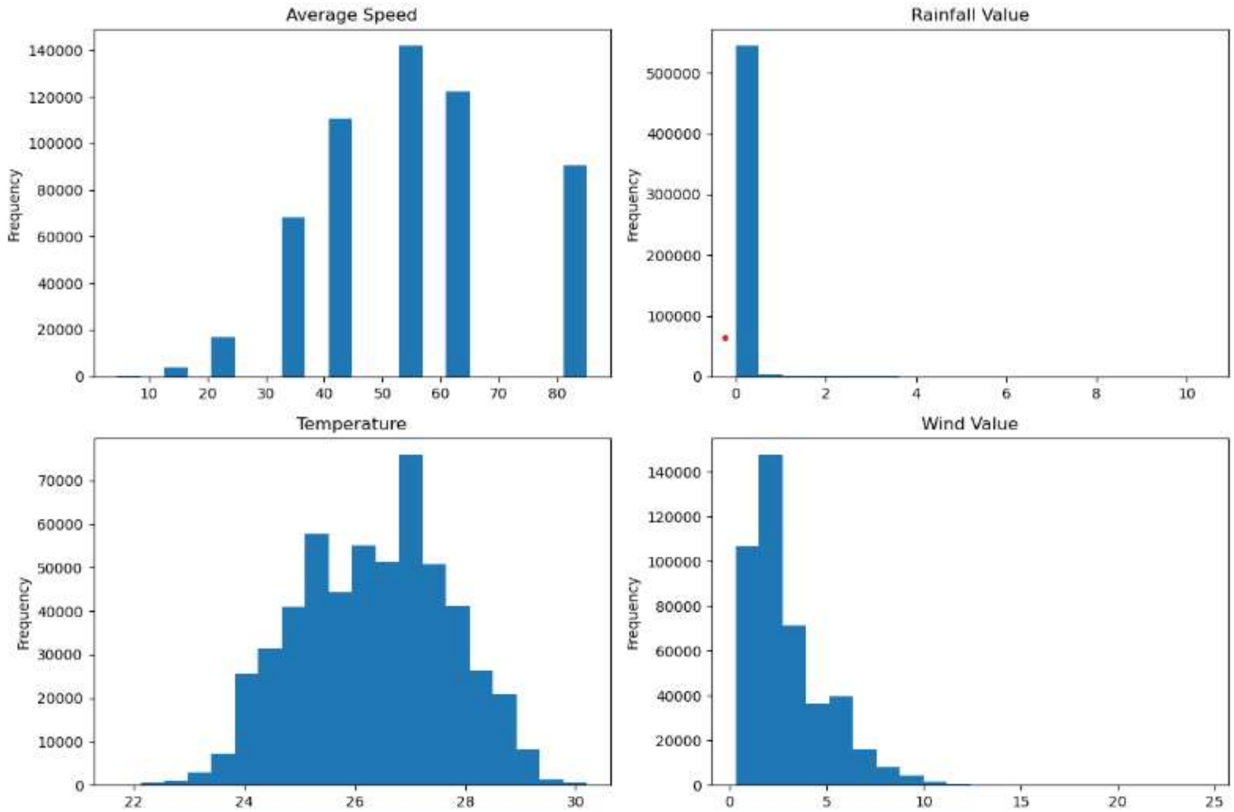

Figure 9.38: Frequencies of each data – Combined Analysis (F)

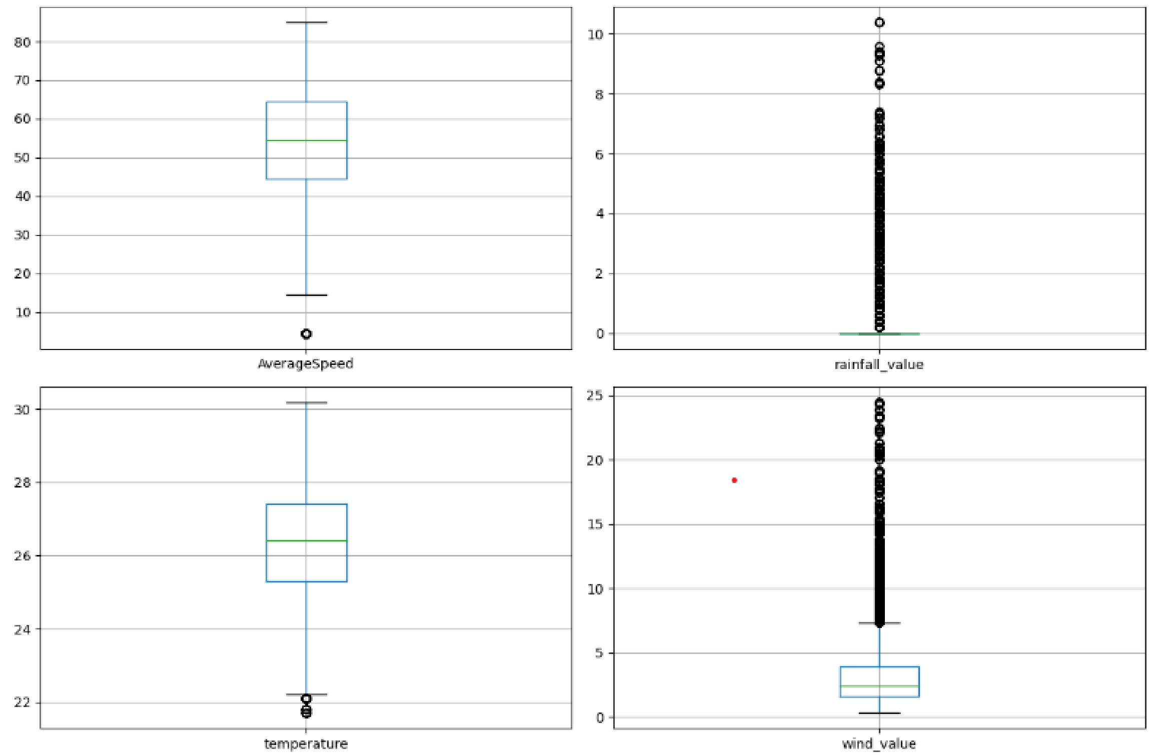

Figure 9.39: Boxplots of each data – Combined Analysis (F)

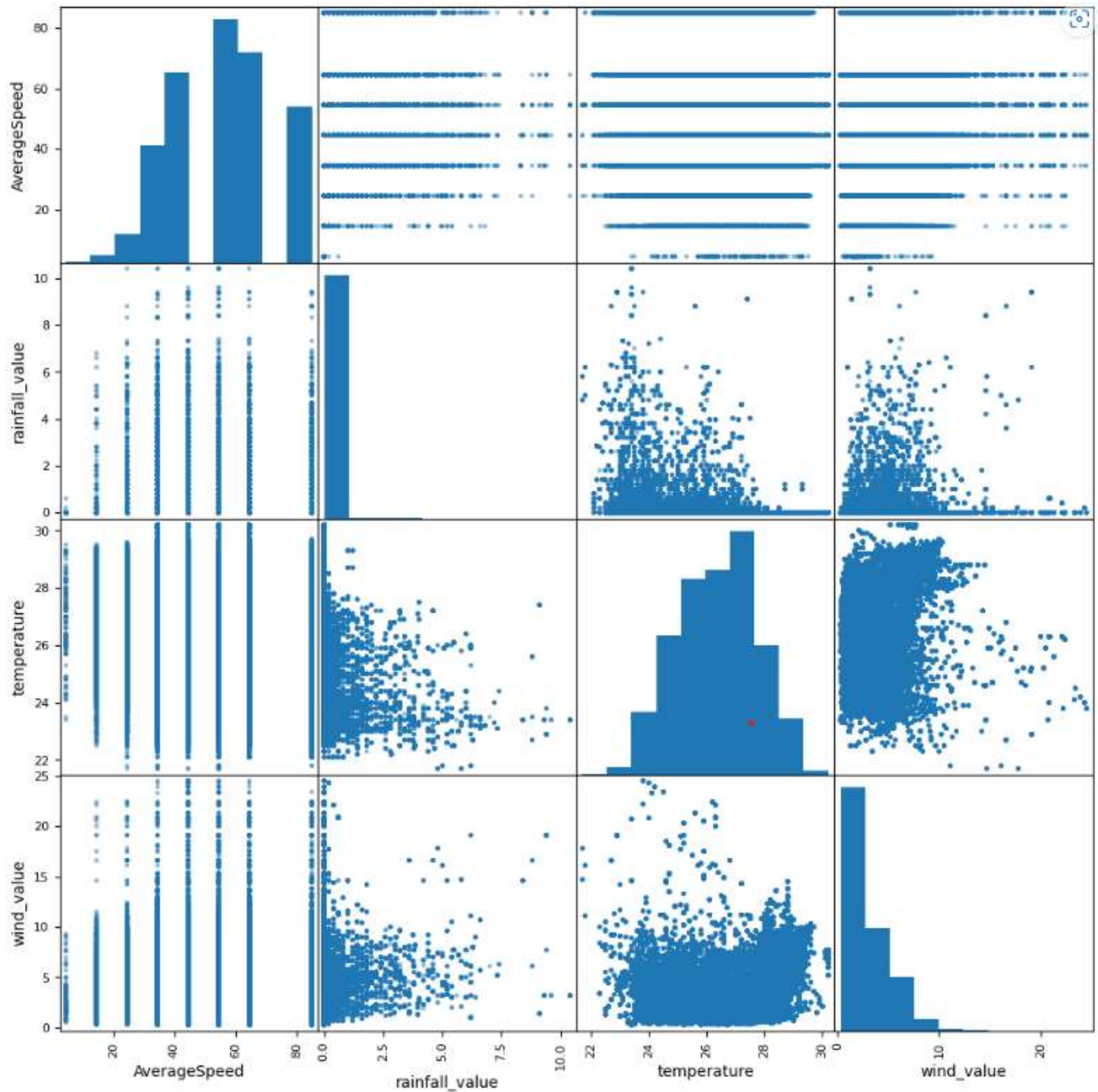

Figure 9.40: Scatter Matrix – Combined Analysis (F)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.016349	-0.056926	0.033456
rainfall_value	-0.016349	1.000000	-0.165098	0.120441
temperature	-0.056926	-0.165098	1.000000	0.079264
wind_value	0.033456	0.120441	0.079264	1.000000

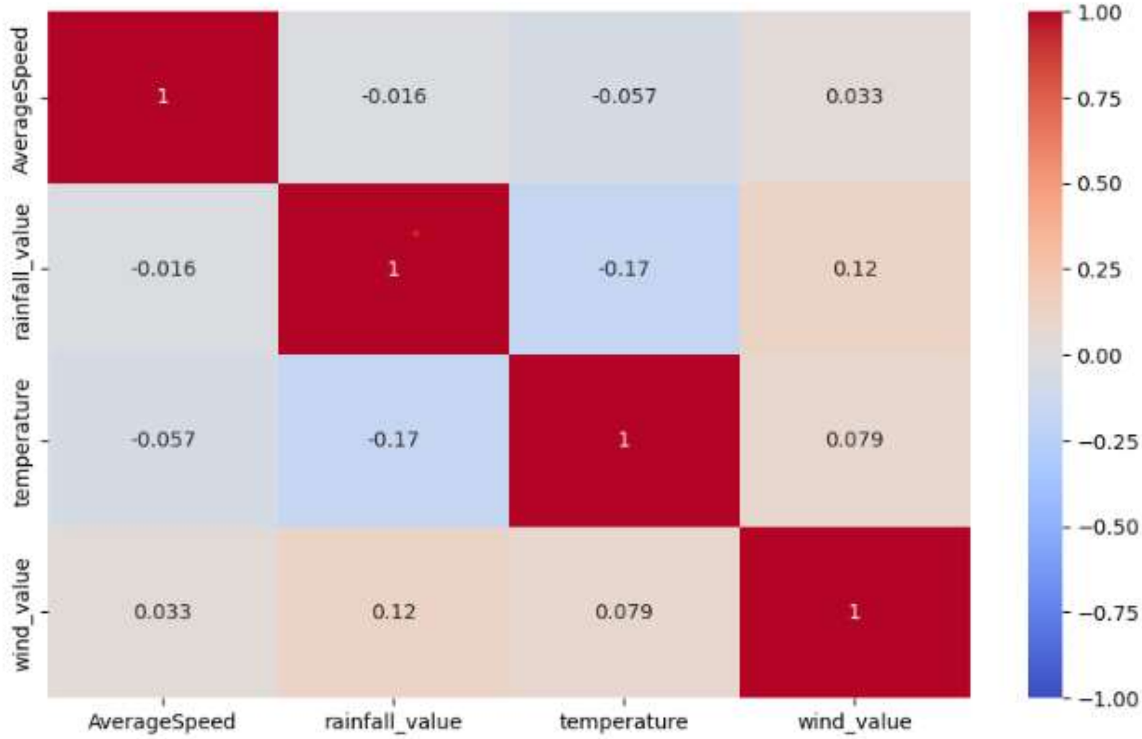

Figure 9.41: Correlation Matrix – Combined Analysis (F)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.016, indicating a very weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is -0.057, suggesting a weak negative relationship, while the correlation between average speed and wind value is 0.033, indicating a very weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                  0.005
Method:                 Least Squares   F-statistic:                     692.6
Date:                Fri, 07 Apr 2023   Prob (F-statistic):               0.00
Time:                        08:37:01   Log-Likelihood:            -1.8185e+06
No. Observations:              432932   AIC:                         3.637e+06
Df Residuals:                  432928   BIC:                         3.637e+06
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             72.6686      0.473    153.719      0.000      71.742      73.595
rainfall_value    -1.4908      0.078    -19.180      0.000      -1.643      -1.338
temperature       -0.6769      0.018    -37.627      0.000      -0.712      -0.642
wind_value         0.3198      0.012     27.122      0.000       0.297       0.343
==============================================================================
Omnibus:                     7547.615   Durbin-Watson:                   1.562
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6546.495
Skew:                           0.244   Prob(JB):                         0.00
Kurtosis:                       2.646   Cond. No.                         513.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.42: OLS Regression Analysis – Combined Analysis (F)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.005, indicating that only 0.5% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.005. These low R-squared values suggest that the model does not explain a substantial amount of the variation in average Speed.

The F-statistic is 692.6, with a p-value (Prob (F-statistic)) of 0.00. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -1.4908, suggesting that for each unit increase in rainfall value, the average speed decreases by approximately 1.49 units, holding the other variables constant. The coefficient for temperature is -0.6769, indicating that for each unit increase in temperature, the average speed decreases

by approximately 0.68 units, holding the other variables constant. The coefficient for wind value is 0.3198, implying that for each unit increase in wind value, the average speed increases by approximately 0.32 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.562, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 513, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.3.2. December Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	117822.000000	117822.000000	117392.000000	117368.000000
mean	56.074668	0.007197	24.990311	3.097681
std	16.262962	0.038320	0.769184	1.698402
min	4.500000	0.000000	22.900000	0.300000
25%	44.500000	0.000000	24.400000	1.800000
50%	54.500000	0.000000	24.900000	2.600000
75%	64.500000	0.000000	25.500000	4.200000
max	85.000000	0.402000	27.600000	11.200000

Figure 9.43: Descriptive Statistics – December Analysis (F)

The count row displays the number of data points for each variable, with 117,822 data points for average speed, rainfall value, 117,392 for temperature, and 117,368 for wind value. The mean row shows the average value for each variable, with average speed having a mean of 56.07, rainfall value at 0.007, temperature at 24.99, and wind value at 3.10.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 16.26, for rainfall value it is 0.038, for temperature it is 0.77, and for wind value it is 1.70. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

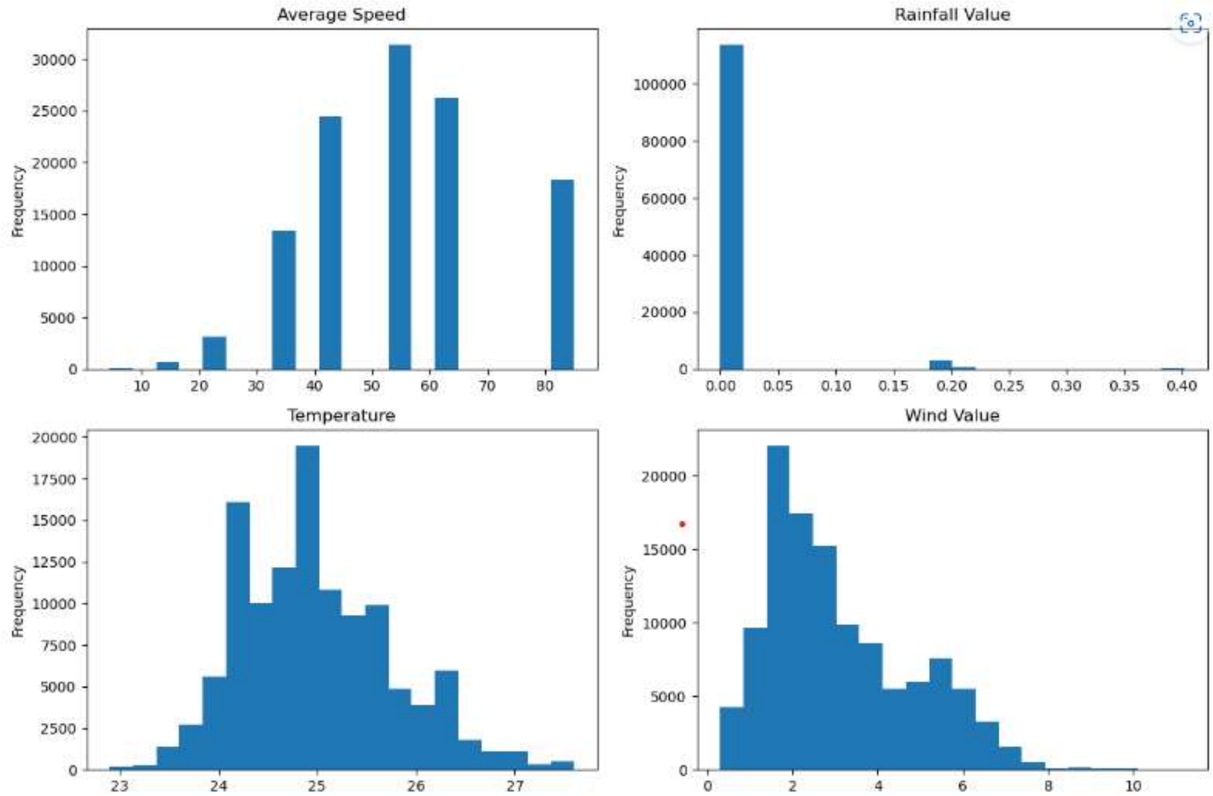

Figure 9.44: Frequencies of each data – December Analysis (F)

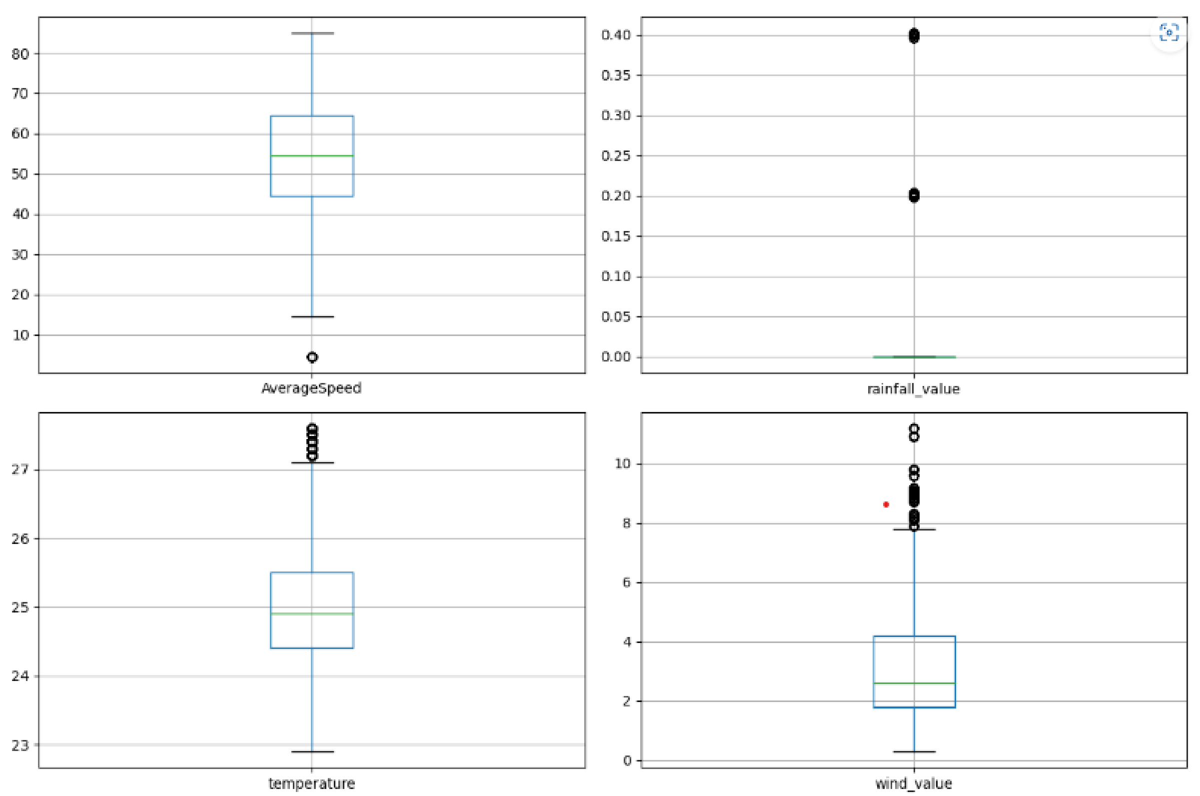

Figure 9.45: Boxplots of each data – December Analysis (F)

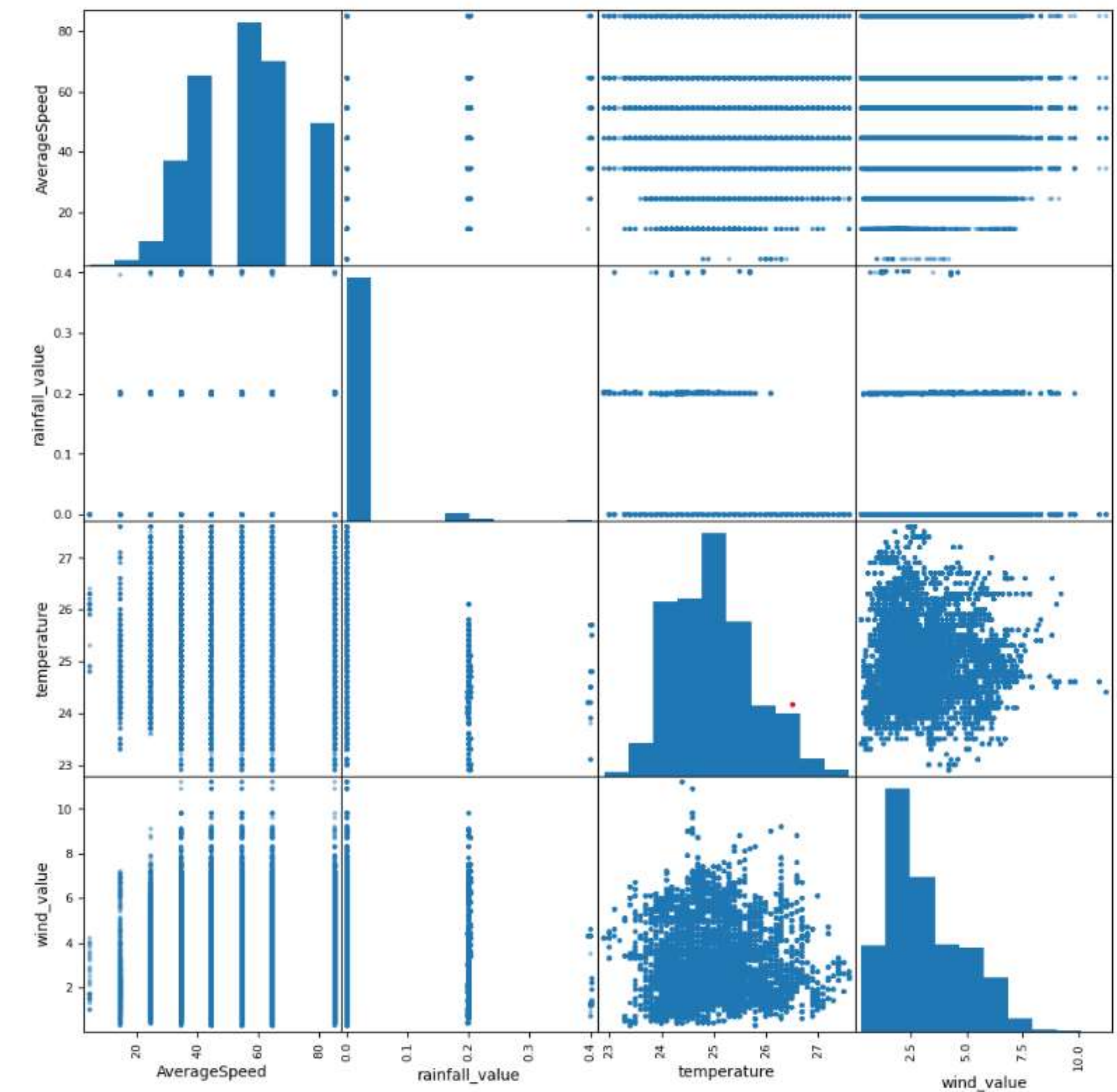

Figure 9.46: Scatter Matrix – December Analysis (F)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.046163	-0.043378	0.106073
rainfall_value	-0.046163	1.000000	-0.120411	0.020849
temperature	-0.043378	-0.120411	1.000000	-0.015641
wind_value	0.106073	0.020849	-0.015641	1.000000

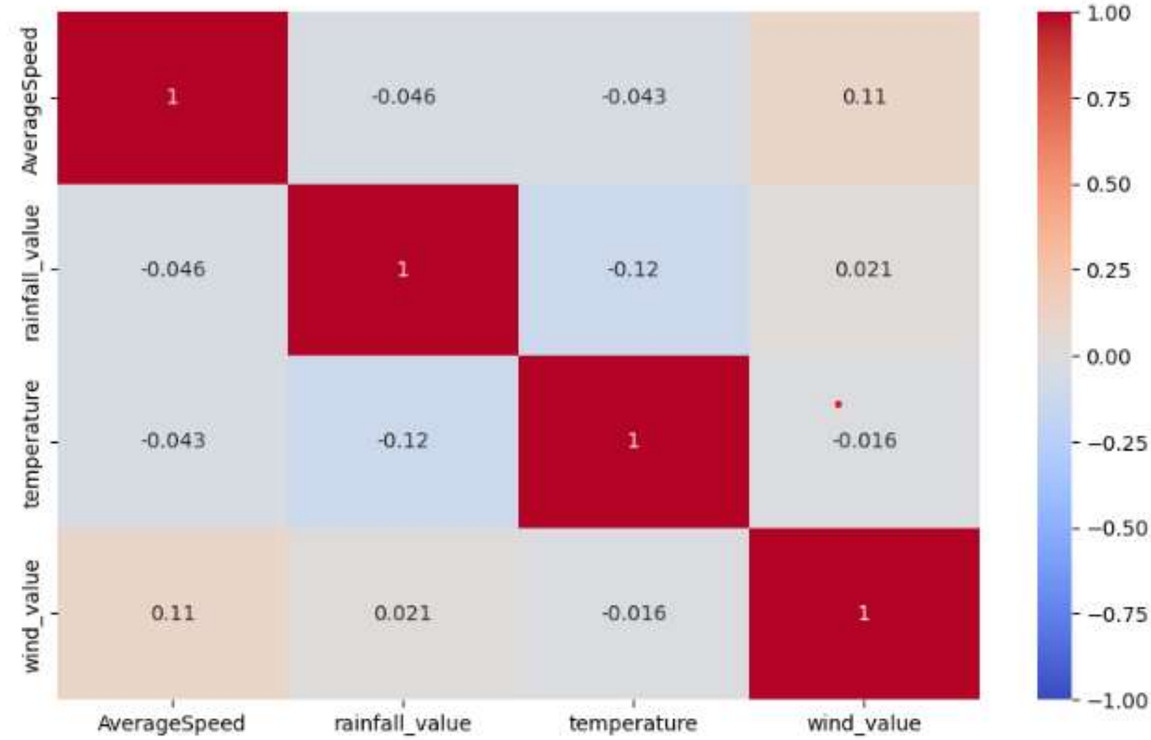

Figure 9.47: Correlation Matrix – December Analysis (F)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.046, indicating a weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is -0.043, suggesting a weak negative relationship, while the correlation between average speed and wind value is 0.106, indicating a weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     633.1
Date:                Fri, 07 Apr 2023   Prob (F-statistic):               0.00
Time:                        09:25:32   Log-Likelihood:            -4.9301e+05
No. Observations:              117368   AIC:                         9.860e+05
Df Residuals:                  117364   BIC:                         9.861e+05
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             78.6428      1.548     50.812      0.000      75.609      81.676
rainfall_value   -23.1801      1.240    -18.694      0.000     -25.611     -20.750
temperature       -1.0224      0.062    -16.566      0.000      -1.143      -0.901
wind_value         1.0201      0.028     36.752      0.000       0.966       1.075
==============================================================================
Omnibus:                     2168.699   Durbin-Watson:                   1.553
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1808.711
Skew:                           0.237   Prob(JB):                         0.00
Kurtosis:                       2.620   Cond. No.                         839.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.48: OLS Regression Results – December Analysis (F)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.016, indicating that only 1.6% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.016. These R-squared values suggest that the model does not explain a substantial amount of the variation in average Speed.

The F-statistic is 633.1, with a p-value (Prob (F-statistic)) of 0.00. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -23.1801, suggesting that for each unit increase in rainfall value, the average speed decreases by approximately 23.18 units, holding the other variables constant. The coefficient for temperature is -1.0224, indicating that for each unit increase in temperature, the average speed decreases

by approximately 1.02 units, holding the other variables constant. The coefficient for wind value is 1.0201, implying that for each unit increase in wind value, the average speed increases by approximately 1.02 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.553, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 839, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average Speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

9.3.3. September Analysis

	AverageSpeed	rainfall_value	temperature	wind_value
count	129649.000000	129649.000000	127594.000000	126995.000000
mean	56.570467	0.058664	26.691842	3.239996
std	16.388921	0.452148	1.005452	2.460904
min	4.500000	0.000000	21.700000	0.300000
25%	44.500000	0.000000	26.200000	1.600000
50%	54.500000	0.000000	26.700000	2.400000
75%	64.500000	0.000000	27.300000	4.900000
max	85.000000	10.400000	29.400000	24.500000

Figure 9.49: Descriptive Statistics – September Analysis (F)

The count row displays the number of data points for each variable, with 129,649 data points for average speed and rainfall value, 127,594 for temperature, and 126,995 for wind value. The mean row shows the average value for each variable, with average Speed having a mean of 56.57, rainfall value at 0.059, temperature at 26.69, and wind value at 3.24.

The standard deviation (std) row indicates the spread of the data points around their respective means. For average speed, the standard deviation is 16.39, for rainfall value it is 0.452, for temperature it is 1.01, and for wind value it is 2.46. The minimum (min), 25th percentile (25%), 50th percentile (50%, also known as the median), 75th percentile (75%), and maximum (max) values for each variable are also provided, giving a more detailed understanding of the distribution of the data.

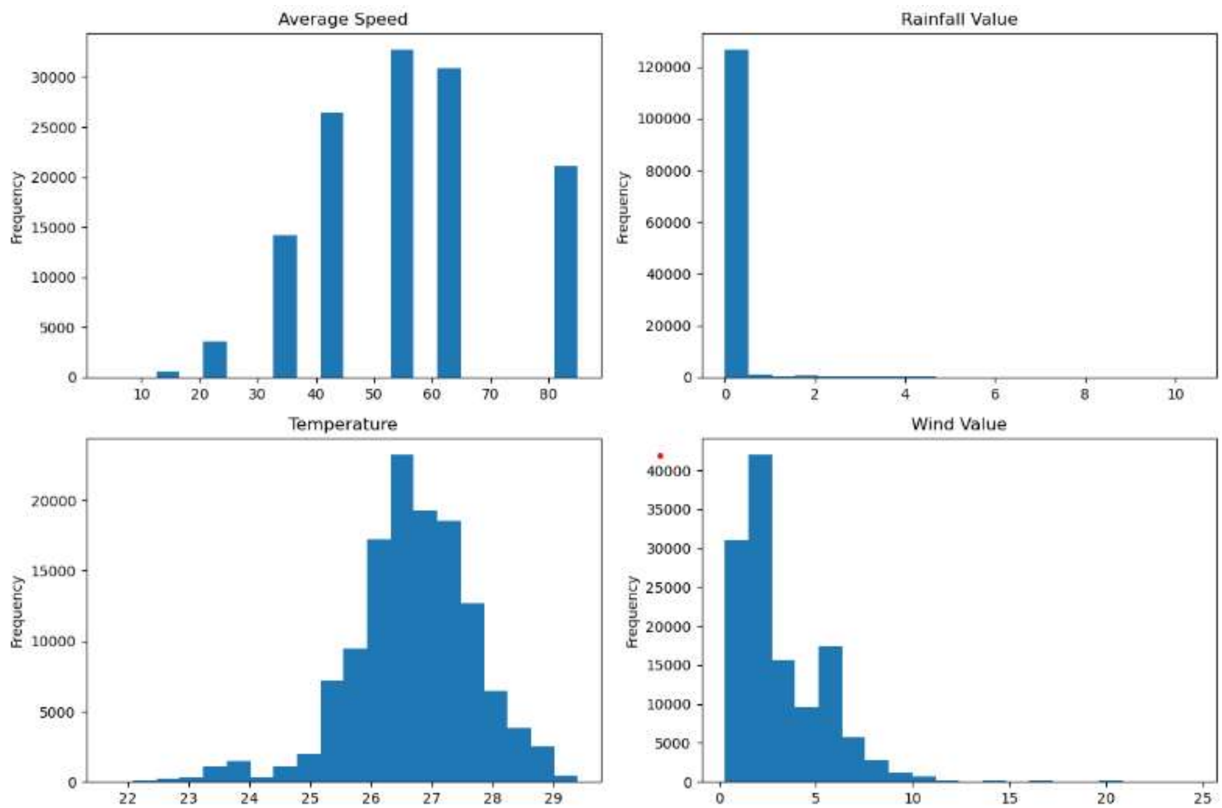

Figure 9.50: Frequencies of each data – September Analysis (F)

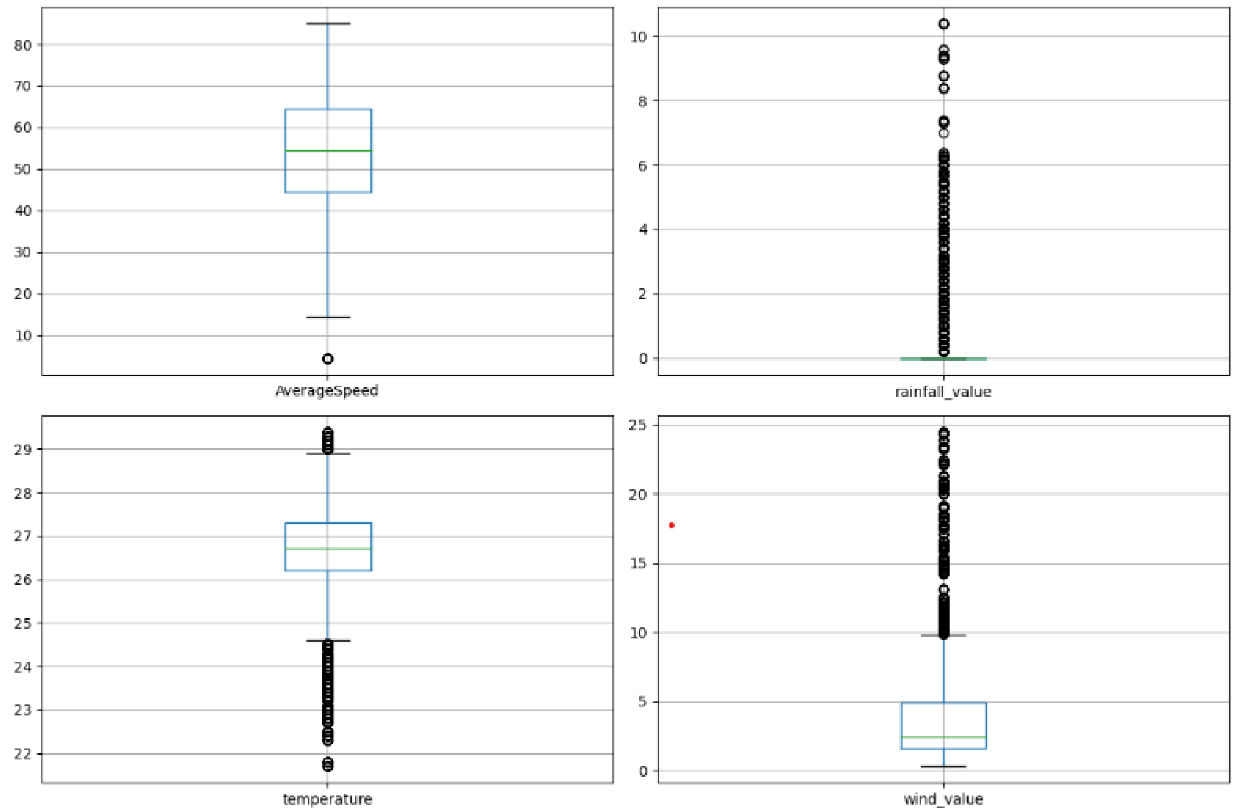

Figure 9.51: Boxplots of each data – September Analysis (F)

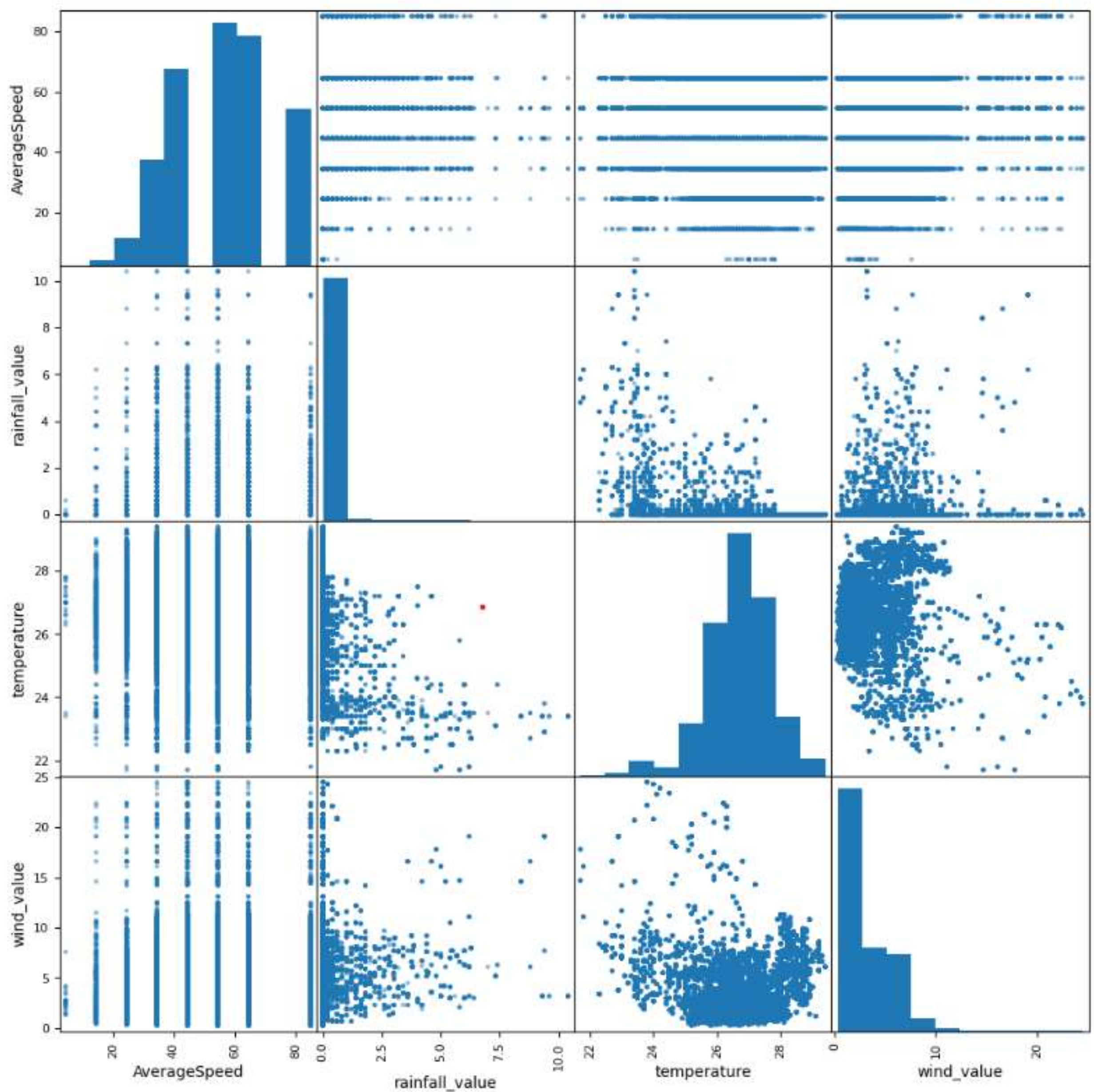

Figure 9.52: Scatter Matrix – September Analysis (F)

	AverageSpeed	rainfall_value	temperature	wind_value
AverageSpeed	1.000000	-0.029291	-0.059072	0.050621
rainfall_value	-0.029291	1.000000	-0.330708	0.166681
temperature	-0.059072	-0.330708	1.000000	-0.124251
wind_value	0.050621	0.166681	-0.124251	1.000000

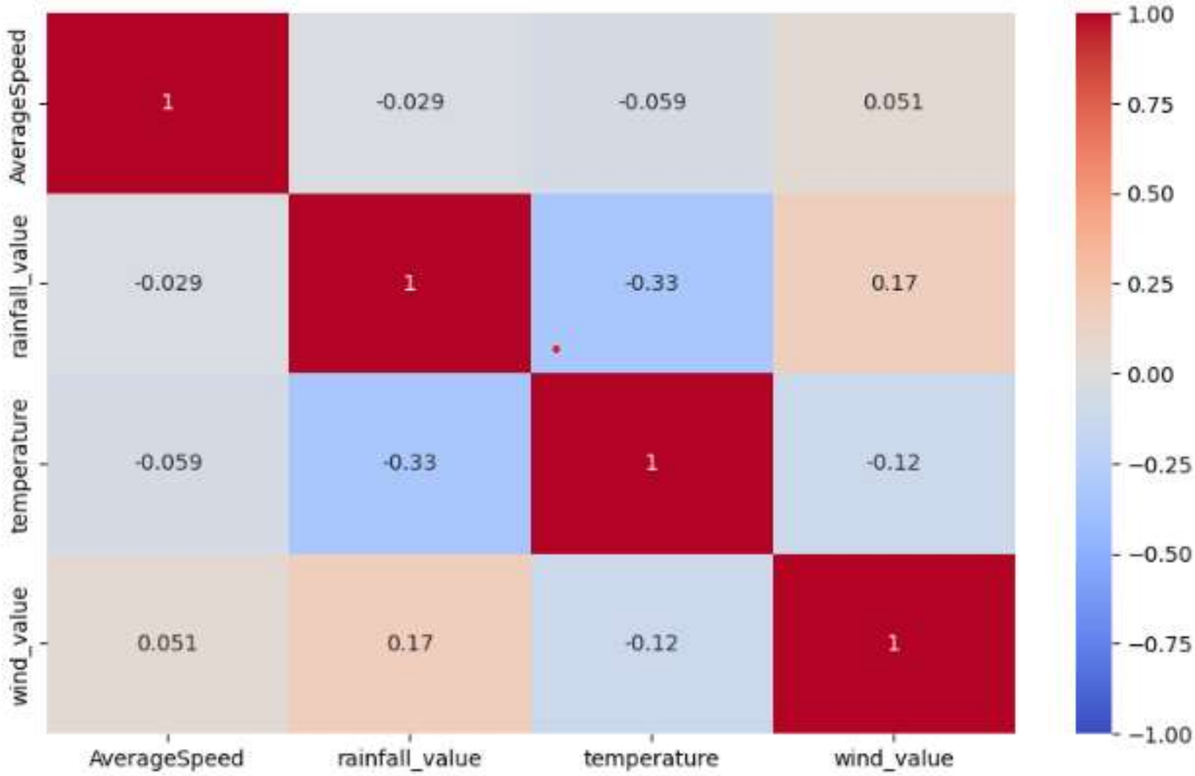

Figure 9.53: Correlation Matrix – September Analysis (F)

The correlation matrix displays the correlation coefficients between each pair of variables, which provides insight into the strength and direction of the linear relationships between them. A positive correlation coefficient indicates a positive relationship between the variables, while a negative correlation coefficient indicates a negative relationship. The correlation coefficients range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

The correlation coefficient between average speed and rainfall value is -0.029291, indicating a weak negative relationship. This suggests that as rainfall increases, the average speed may decrease slightly. The correlation between average speed and temperature is -0.059072, suggesting a weak negative relationship, while the correlation between average speed and wind value is 0.050621, indicating a weak positive relationship. The correlation coefficients between the weather variables (rainfall value, temperature, and wind value) are also provided, showing weak correlations among them.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           AverageSpeed   R-squared:                       0.009
Model:                            OLS   Adj. R-squared:                  0.009
Method:                 Least Squares   F-statistic:                     375.0
Date:                Fri, 07 Apr 2023   Prob (F-statistic):          1.52e-242
Time:                        10:23:00   Log-Likelihood:            -5.3475e+05
No. Observations:              126995   AIC:                         1.069e+06
Df Residuals:                  126991   BIC:                         1.070e+06
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             87.5908      1.309     66.891      0.000      85.024      90.157
rainfall_value    -2.2469      0.107    -20.952      0.000      -2.457      -2.037
temperature       -1.1997      0.049    -24.611      0.000      -1.295      -1.104
wind_value         0.3460      0.019     18.294      0.000       0.309       0.383
==============================================================================
Omnibus:                     2588.035   Durbin-Watson:                   1.505
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1902.685
Skew:                           0.204   Prob(JB):                         0.00
Kurtosis:                       2.560   Cond. No.                         771.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9.54: OLS Regression Results – September Analysis (F)

The output presents the results of the linear multiple regression analysis, where the dependent variable is average speed, and the independent variables are rainfall value, temperature, and wind value.

The R-squared value is 0.009, indicating that only 0.9% of the variation in average speed can be explained by the model. The adjusted R-squared, which accounts for the number of independent variables and sample size, is also 0.009. These R-squared values suggest that the model does not explain a substantial amount of the variation in average speed.

The F-statistic is 375.0, with a p-value (Prob (F-statistic)) of 1.52e-242. This implies that the model is statistically significant, and at least one of the independent variables has a significant effect on the dependent variable.

The coefficients, standard errors, t-statistics, p-values, and 95% confidence intervals for each independent variable are presented in the table. All independent variables have statistically significant effects on average speed, as their p-values are less than 0.05.

The coefficient for rainfall value is -2.2469, suggesting that for each unit increase in rainfall value, the average speed decreases by approximately 2.25 units, holding the other variables constant. The coefficient

for temperature is -1.1997, indicating that for each unit increase in temperature, the average speed decreases by approximately 1.20 units, holding the other variables constant. The coefficient for wind value is 0.3460, implying that for each unit increase in wind value, the average speed increases by approximately 0.35 units, holding the other variables constant.

The Omnibus test has a p-value of 0.000, suggesting that the residuals are not normally distributed. The Durbin-Watson statistic is 1.505, which indicates some positive autocorrelation in the residuals. The Jarque-Bera test has a p-value of 0.00, also suggesting non-normality in the residuals. The skewness and kurtosis values provide further evidence of non-normality.

The condition number is 771, which is not extremely high, suggesting that multicollinearity is not a major issue in the model.

In summary, the linear multiple regression analysis indicates that rainfall value, temperature, and wind value are statistically significant predictors of average speed, but the model explains only a small portion of the variation in average speed. The residuals show signs of non-normality and autocorrelation, which may indicate potential issues with the model's assumptions.

10. Conclusion

In conclusion, the correlation matrices and R-squared values presented for Road Categories A, B, and F in combined data, December, and September provide insights into the relationships between weather factors (rainfall, temperature, and wind) and the average speed of vehicles in different road categories.

For Road Category A, the correlations with rainfall are weak and negative, indicating a slight decrease in average speed as rainfall increases. In contrast, temperature and wind have weak positive correlations. The R-squared values are very low, ranging from 0.6% to 3%, suggesting that weather factors only explain a small portion of the variation in average speed on these roads.

In Road Category B, the correlations between average speed and weather factors are generally weak. Rainfall has a weak positive correlation in the combined data but weak negative correlations in December and September. Temperature and wind display weak negative correlations, with wind showing a stronger negative relationship in December. The R-squared values range from 1.3% to 14.5%, indicating that weather factors have a slightly higher explanatory power for average speed in Road Category B, particularly in December.

For Road Category F, the correlations with rainfall are weak and negative, while the correlations with temperature are weak and negative, and the correlations with wind are weak and positive. The R-squared values range from 0.5% to 1.6%, demonstrating that weather factors explain a limited amount of variation in average speed on these roads.

Overall, the findings suggest that weather factors have limited explanatory power for average speed variations across different road categories. Although there are weak correlations between average speed and weather factors, the R-squared values indicate that other factors not considered in this analysis might have a more significant impact on average speed. Further research is required to identify these factors and develop a more comprehensive understanding of the determinants of average speed on various road categories. This may involve exploring additional variables or employing alternative modeling approaches to better capture the complex relationships between weather factors and average speed.

11. References

Abdel-Aty, M., Ekram, A.-A., Huang, H., & Choi, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis & Prevention*, *43*(5), 1730–1737. https://doi.org/10.1016/j.aap.2011.04.003

Agarwal, M. (n.d.). *Impacts of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity*.

Andrey, D. J. (n.d.). *Weather Information and Road Safety*.

Andrey, J., Mills, B., Leahy, M., & Suggett, J. (n.d.). *Weather as a Chronic Hazard for Road Transportation in Canadian Cities*.

Andrey, J., & Yagar, S. (1993). A temporal analysis of rain-related crash risk. *Accident Analysis & Prevention*, *25*(4), 465–472. https://doi.org/10.1016/0001-4575(93)90076-9

Antoniou, C., Yannis, G., & Katsochis, D. (2013). *Impact of meteorological factors on the number of injury accidents*.

Bergel-Hayat, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the road accident risk: Weather effects. *Accident Analysis & Prevention*, *60*, 456–465. https://doi.org/10.1016/j.aap.2013.03.006

Brijs, T., Karlis, D., & Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis & Prevention*, *40*(3), 1180–1190. https://doi.org/10.1016/j.aap.2008.01.001

Chen, C., Gong, H., & Paaswell, R. (2008). Role of the built environment on mode choice decisions: Additional evidence on the impact of density. *Transportation*, *35*(3), 285–299. https://doi.org/10.1007/s11116-007-9153-5

Cools, M., Moons, E., & Wets, G. (2010). Assessing the Impact of Weather on Traffic Intensity. *Weather, Climate, and Society*, *2*(1), 60–68. https://doi.org/10.1175/2009WCAS1014.1

Deb, B., Khan, S. R., Tanvir Hasan, K., Khan, A. H., & Alam, Md. A. (2019). Travel Time Prediction using Machine Learning and Weather Impact on Traffic Conditions. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 1–8. https://doi.org/10.1109/I2CT45611.2019.9033922

Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. *Accident Analysis & Prevention*, *36*(4), 637–647. https://doi.org/10.1016/S0001-4575(03)00085-X

Fior, J., & Cagliero, L. (2022). Correlating Extreme Weather Conditions With Road Traffic Safety: A Unified Latent Space Model. *IEEE Access*, *10*, 73005–73018. https://doi.org/10.1109/ACCESS.2022.3190399

Hammad, H. M., Ashraf, M., Abbas, F., Bakhat, H. F., Qaisrani, S. A., Mubeen, M., Fahad, S., & Awais, M. (2019). Environmental factors affecting the frequency of road traffic accidents: A case study of sub-urban area of Pakistan. *Environmental*

Science and Pollution Research, *26*(12), 11674–11685. https://doi.org/10.1007/s11356-019-04752-8

Jung, S., Qin, X., & Noyce, D. A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis & Prevention*, *42*(1), 213–224. https://doi.org/10.1016/j.aap.2009.07.020

Keay, K., & Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident Analysis & Prevention*, *37*(1), 109–124. https://doi.org/10.1016/j.aap.2004.07.005

Koetse, M. J., & Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D: Transport and Environment*, *14*(3), 205–221. https://doi.org/10.1016/j.trd.2008.12.004

Lee, J., Hong, B., Lee, K., & Jang, Y.-J. (2015). A Prediction Model of Traffic Congestion Using Weather Data. *2015 IEEE International Conference on Data Science and Data Intensive Systems*, 81–88. https://doi.org/10.1109/DSDIS.2015.96

Malyshkina, N. V., & Mannering, F. L. (2009). Markov switching multinomial logit model: An application to accident-injury severities. *Accident; Analysis and Prevention*, *41*(4), 829–838. https://doi.org/10.1016/j.aap.2009.04.006

Shang, R., Zhang, Y., Max Shen, Z.-J., & Zhang, Y. (2021). Analyzing the Effects of Road Type and Rainy Weather on Fuel Consumption and Emissions: A Mesoscopic

Model Based on Big Traffic Data. *IEEE Access*, *9*, 62298–62315. https://doi.org/10.1109/ACCESS.2021.3074303

Shin, H., & Lee, J. (2020). Temporal Impulse of Traffic Accidents in South Korea. *IEEE Access*, *8*, 38380–38390. https://doi.org/10.1109/ACCESS.2020.2975529

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, *72*, 244–256. https://doi.org/10.1016/j.aap.2014.06.017

Usman, T., Fu, L., & Miranda-Moreno, L. F. (2010). Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accident Analysis & Prevention*, *42*(6), 1878–1887. https://doi.org/10.1016/j.aap.2010.05.008

Weather Effects on Daily Traffic Accidents and Fatalities.pdf. (n.d.).

Yu, T., Kuang, Q., Hu, J., Zheng, J., & Li, X. (2021). Global-Similarity Local-Salience Network for Traffic Weather Recognition. *IEEE Access*, *9*, 4607–4615. https://doi.org/10.1109/ACCESS.2020.3048116