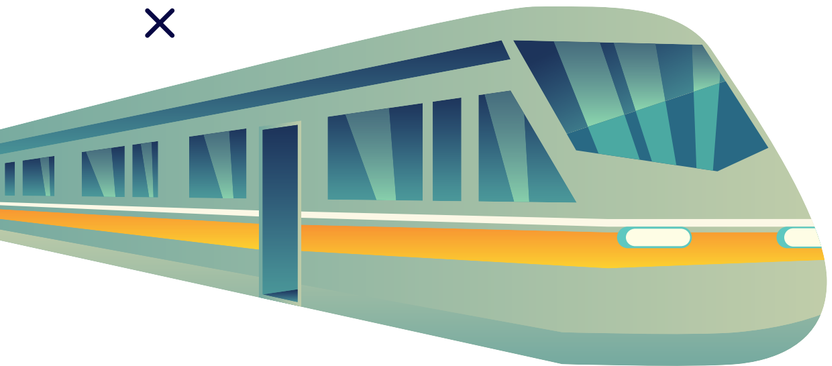
Mini project 2

TIME SERIES ANALYSIS ON TRAFFIC VOLUME

Presented by Maram Alshehri, Shahad Faiz, Fai Aladyani



AGENDA

1. INTRODUCTION
2. PROPOSED FRAMEWORK
3. DATA DESCRIPTION
4. EXPLANATORY DATA ANALYSIS (EDA)
5. DATA PRE-PROCESSING
6. IMPLEMENTED MODELS
7. PERFORMANCE METRICS
8. RESULTS AND DISCUSSION
9. CONCLUSION AND FUTURE WORK

× × × ×

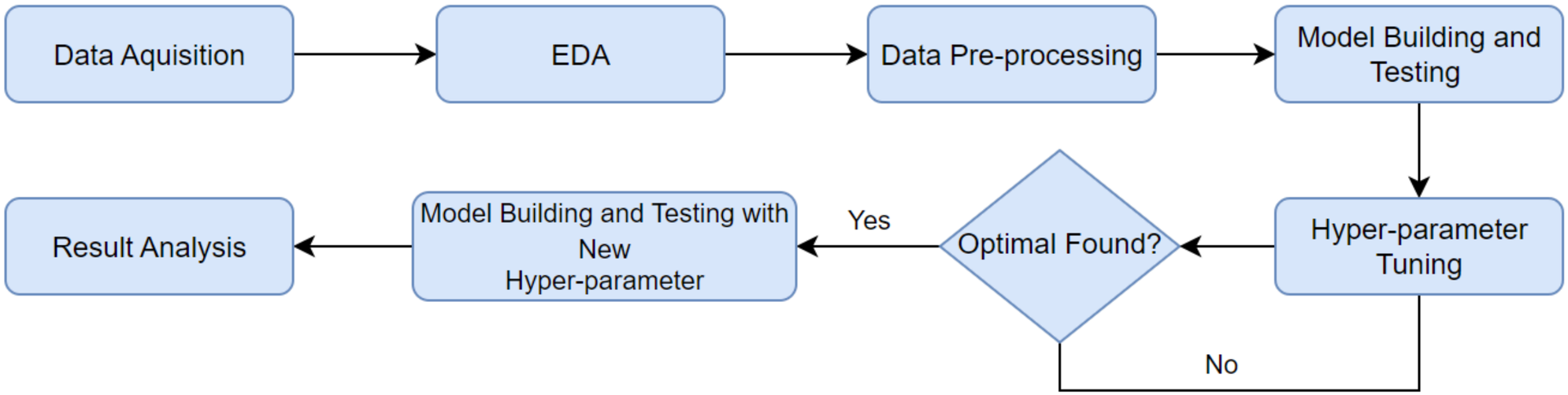
INTRODUCTION

Project Objective

The objective is to develop and evaluate advanced Model in Time Series Forecasting, such as RNN, LSTM and GRU, to accurately forecast traffic volume on the Metro Interstate using the some dataset. Our goal is to gain insights into traffic patterns and trends to inform transportation management strategies and enhance mobility.

PROPOSED FRAMEWORK

The framework consist of seven steps:



1- Data Acquisition

Collecting suitable data from reliable source for the purpose of cleaning and processing the raw data for analysis.

2- Explanatory Data Analysis (EDA)

Analyzing data set to summarize their main characteristics, often using statistical graphics and other data visualization methods

3- Data Pre-processing

Handelling null values, handling categorical data, outlier detection, handling outliers, and data normalization improves the quality of the data.

4- Model Building and Testing

This project employed RNN, LSTM, and GRU model building algorithms.

5- Hyper-parameter Tuning

Different experiments to identify the optimal parameters for high accuracy and performance . In deep learning this include tuning the number of layers and the size of each layer, the learning rate, the activation functions, and the regularization methods.

6- Model Building Testing with New Hyper-parameter

Each model will undergo a second round of training using the optimal hyper-parameters from the previous round.

7- Result Analysis

The RNN, LSTM, and GRU models developed and tested through the proposed framework demonstrated promising performance in accurately predicting traffic volume.

DATA DESCRIPTION

Name of dataset : Metro Interstate Traffic Volume

Dataset Source : Kaggle, **From** 2012-10-02 **To** 2018-09-30 - **NYC**

Programming Language : Python

Framework : TensorFlow and Keras.

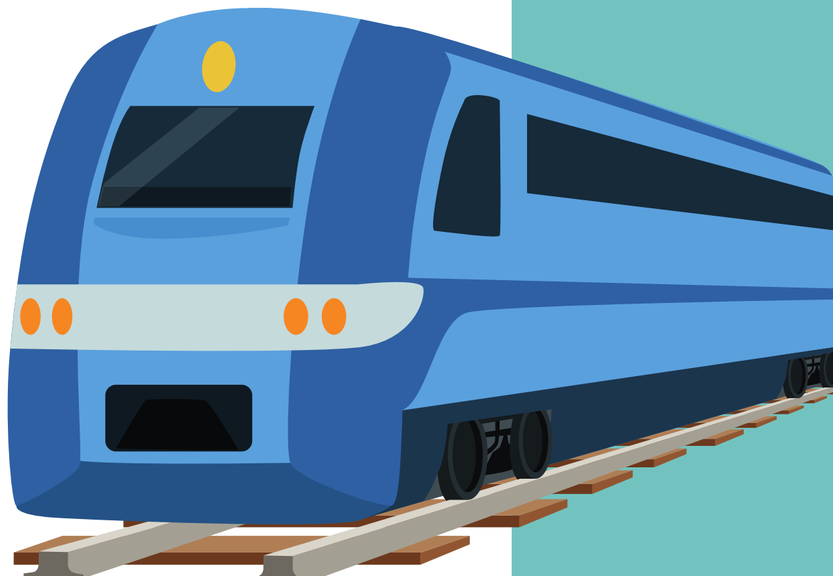
kaggle

python™

statsmodels

TensorFlow

Keras



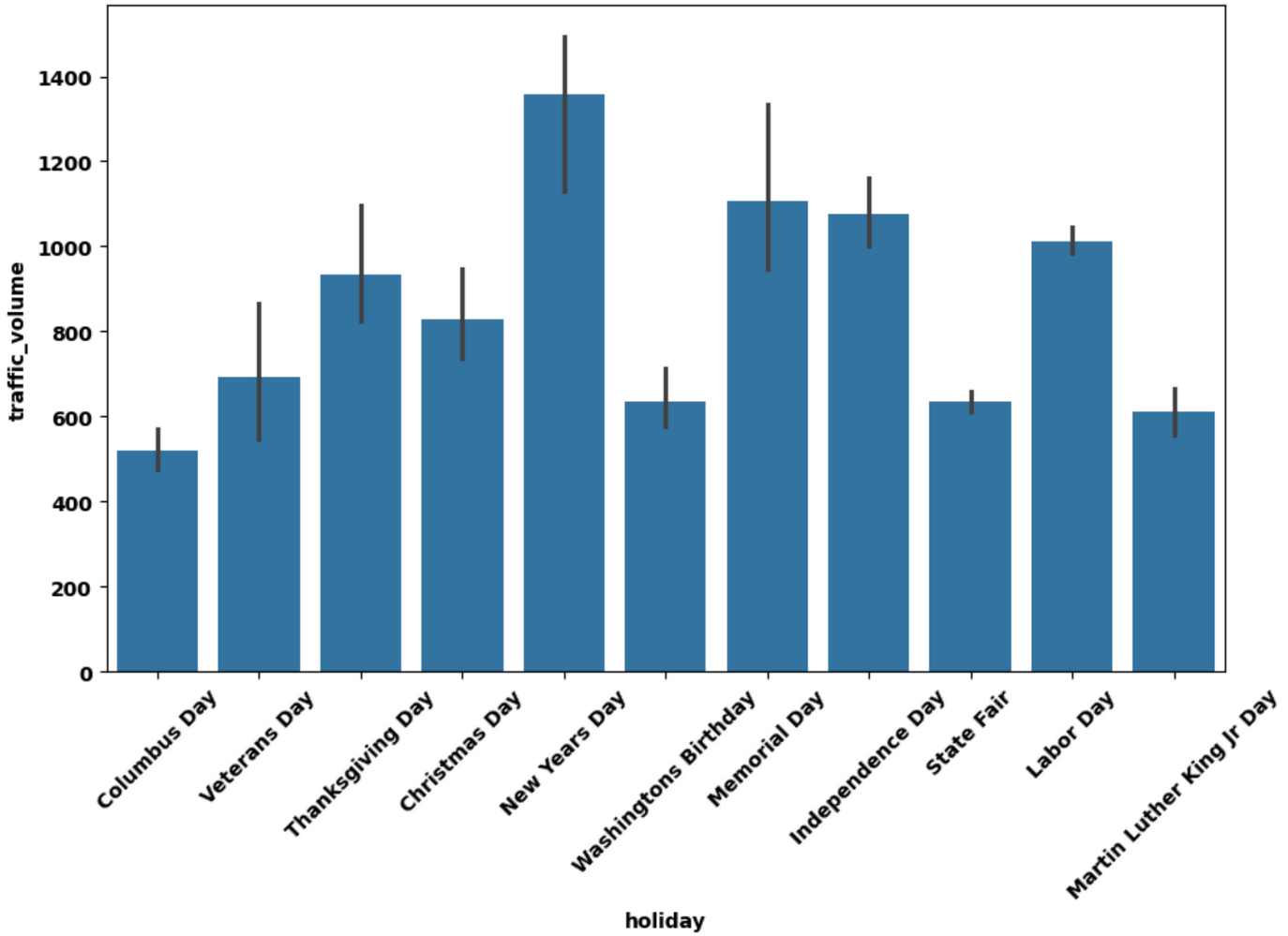
	holiday	temp	rain_1h	snow_1h	clouds_all	weather_main	weather_description	date_time	traffic_volume
0	NaN	288.28	0.0	0.0	40	Clouds	scattered clouds	2012-10-02 09:00:00	5545
1	NaN	289.36	0.0	0.0	75	Clouds	broken clouds	2012-10-02 10:00:00	4516
2	NaN	289.58	0.0	0.0	90	Clouds	overcast clouds	2012-10-02 11:00:00	4767
3	NaN	290.13	0.0	0.0	90	Clouds	overcast clouds	2012-10-02 12:00:00	5026
4	NaN	291.14	0.0	0.0	75	Clouds	broken clouds	2012-10-02 13:00:00	4918
...	...	...	...	...	...	...	...	...	...
48199	NaN	283.45	0.0	0.0	75	Clouds	broken clouds	2018-09-30 19:00:00	3543
48200	NaN	282.76	0.0	0.0	90	Clouds	overcast clouds	2018-09-30 20:00:00	2781
48201	NaN	282.73	0.0	0.0	90	Thunderstorm	proximity thunderstorm	2018-09-30 21:00:00	2159
48202	NaN	282.09	0.0	0.0	90	Clouds	overcast clouds	2018-09-30 22:00:00	1450
48203	NaN	282.12	0.0	0.0	90	Clouds	overcast clouds	2018-09-30 23:00:00	954

48204 rows × 9 columns

Size of dataset :

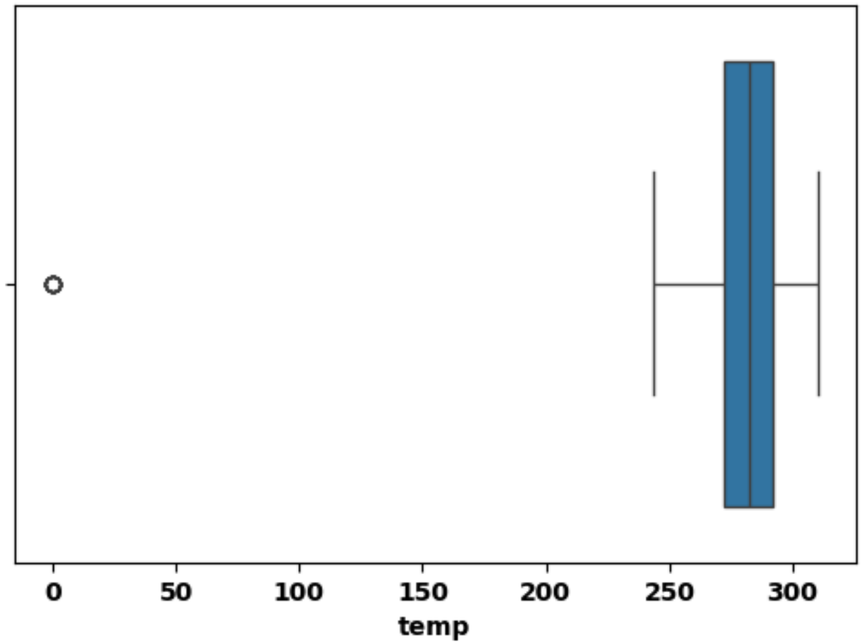
- **8 Features , 1 Target Traffic_Volume**
- **48204 instances**

EXPLANATORY DATA ANALYSIS (EDA)

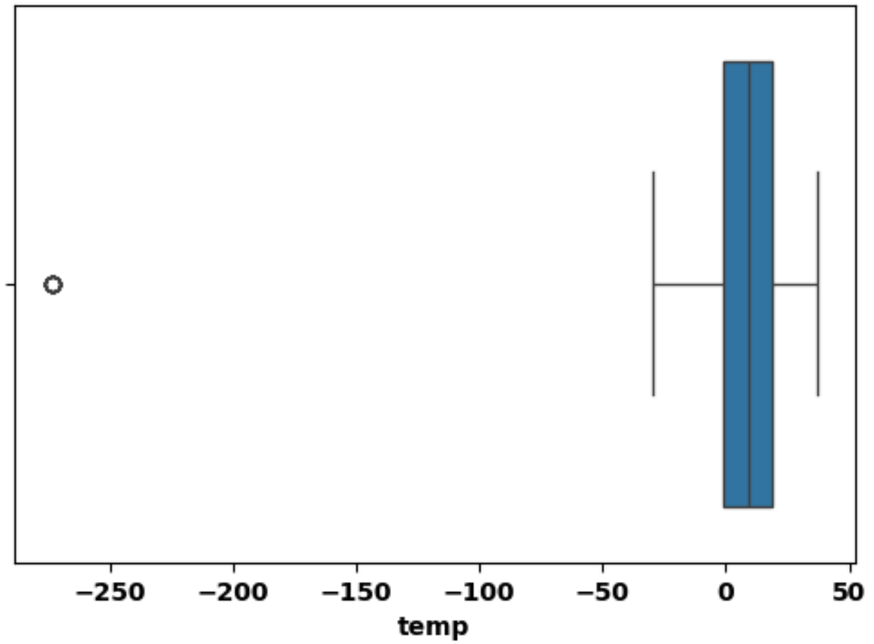


- Bar plot of traffic volume on holidays

EXPLANATORY DATA ANALYSIS (EDA)

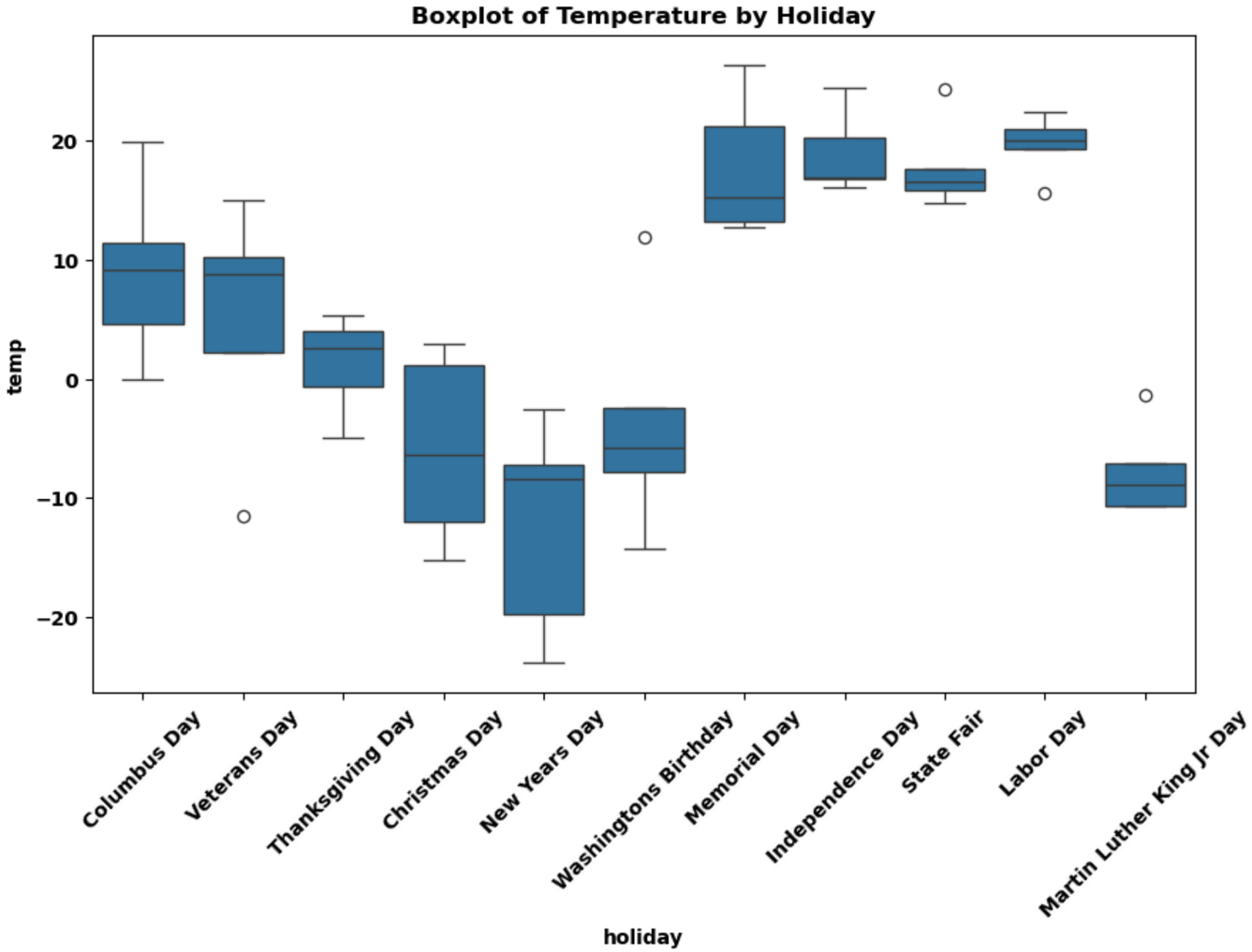


- Box plot of Temperature column in Kelvin



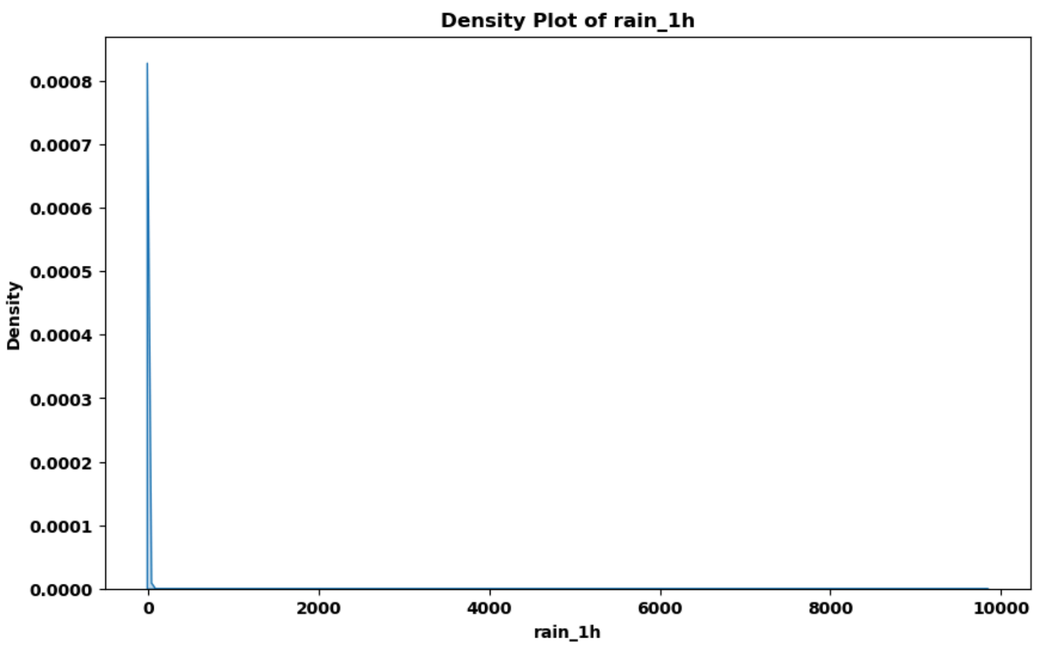
- Box plot of Temperature column in Celsius

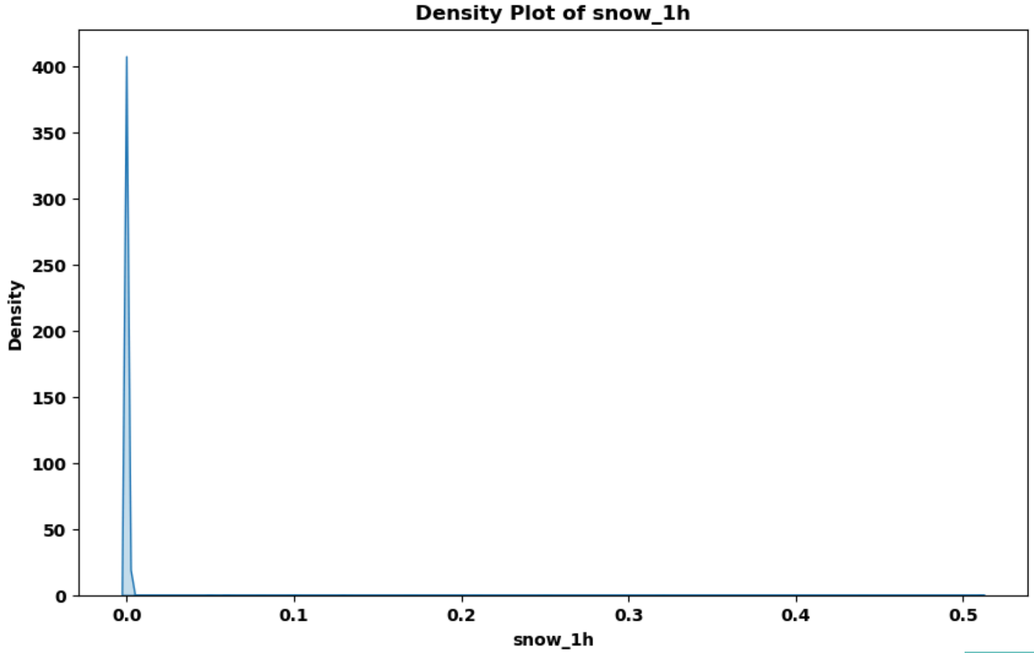
EXPLANATORY DATA ANALYSIS (EDA)



- Box plot of temperature on holidays

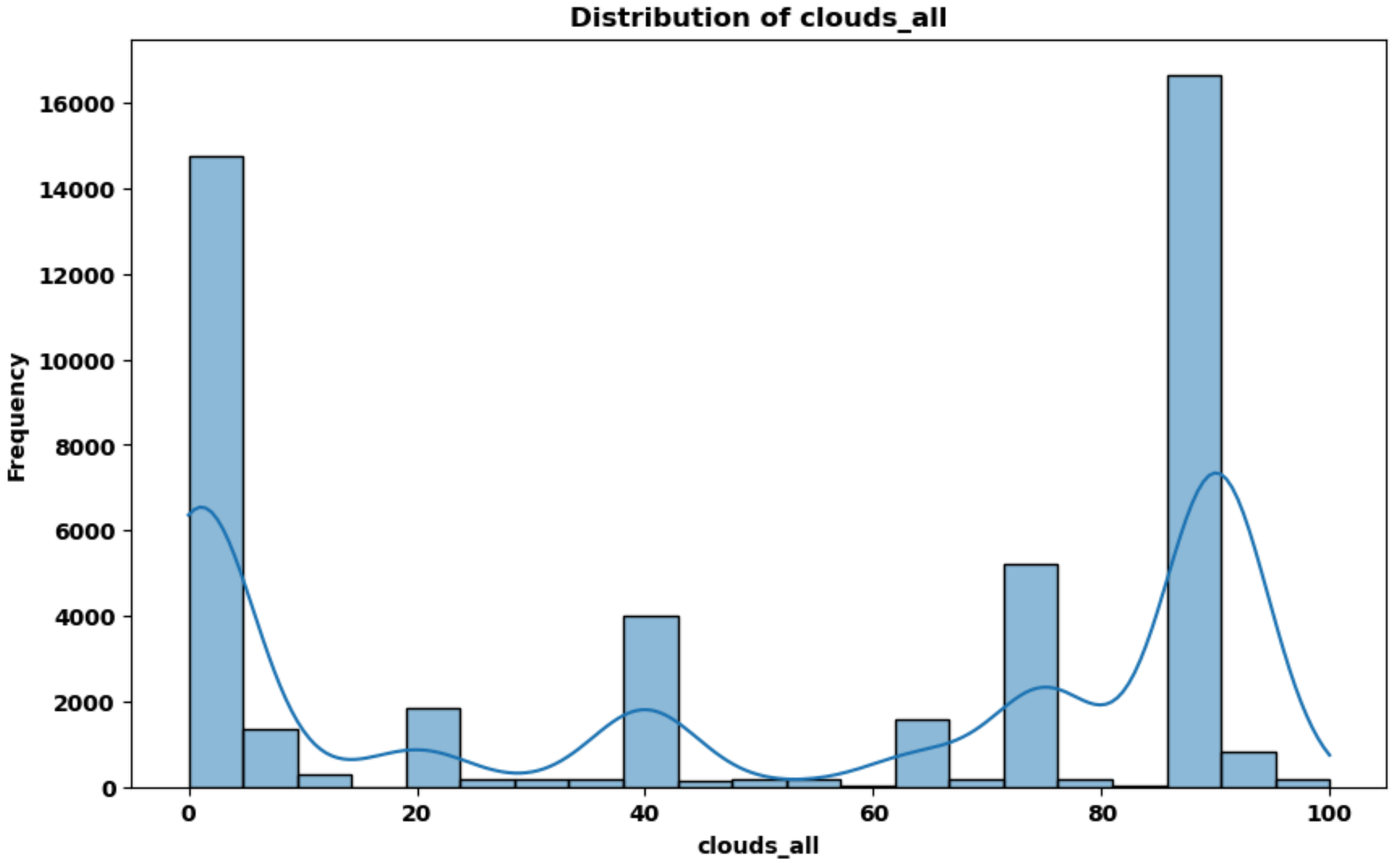
EXPLANATORY DATA ANALYSIS (EDA)





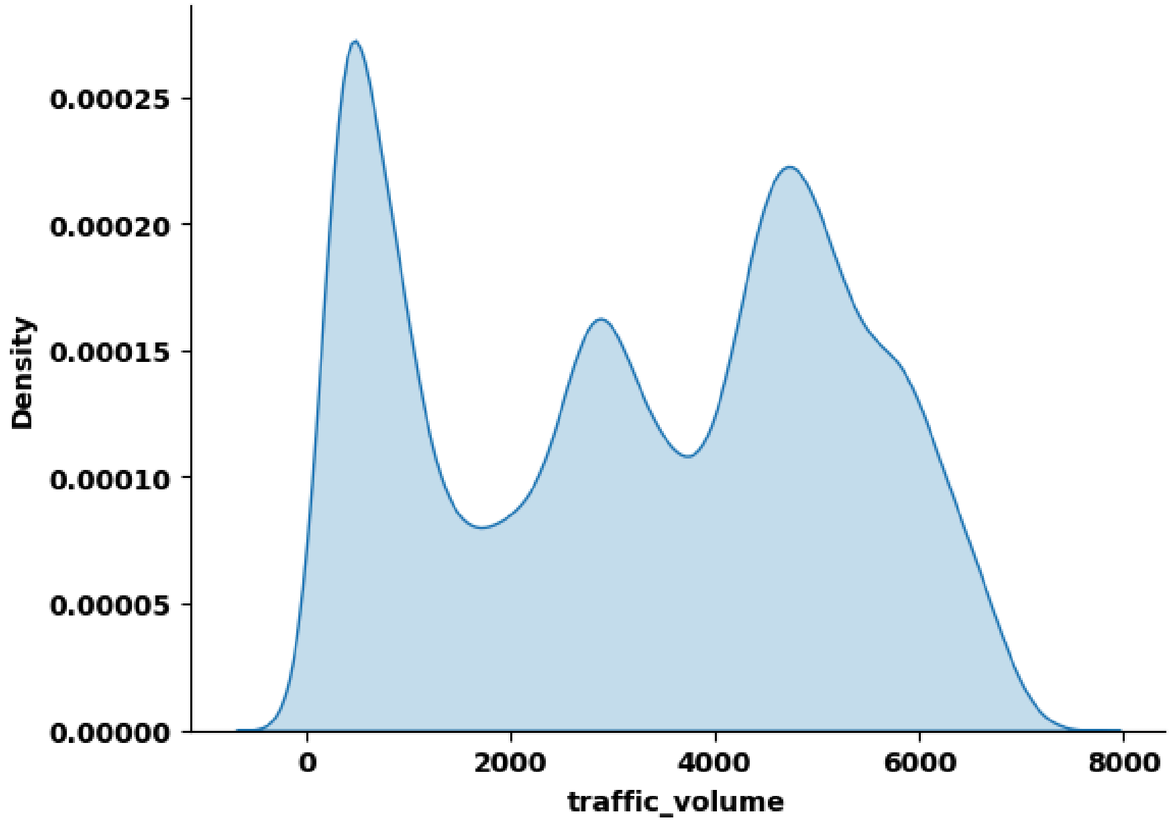
- Density distribution of rain_h1 and snow_1h in mm

EXPLANATORY DATA ANALYSIS (EDA)



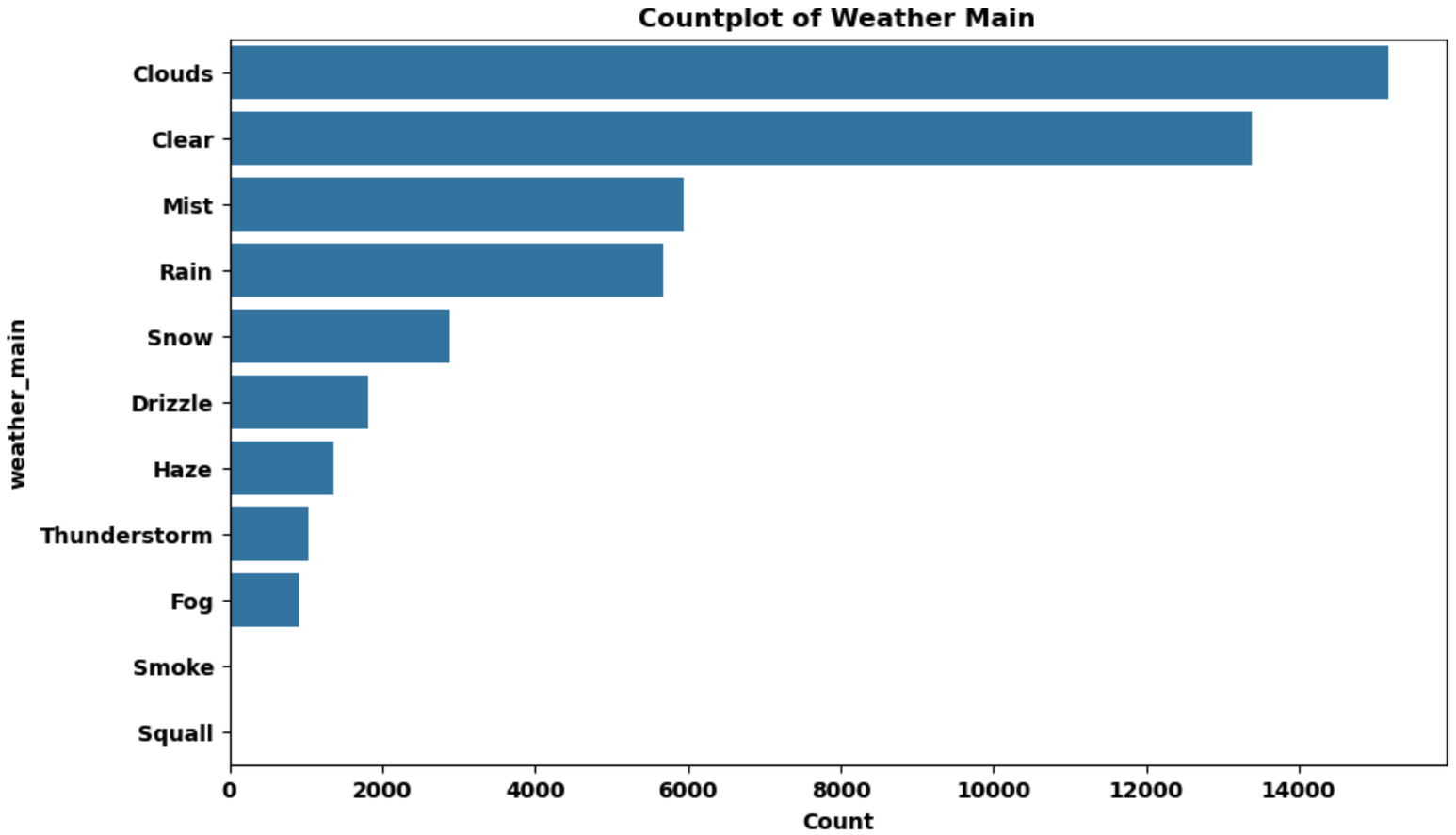
- Histogram distribution of clouds_all in percentage

EXPLANATORY DATA ANALYSIS (EDA)



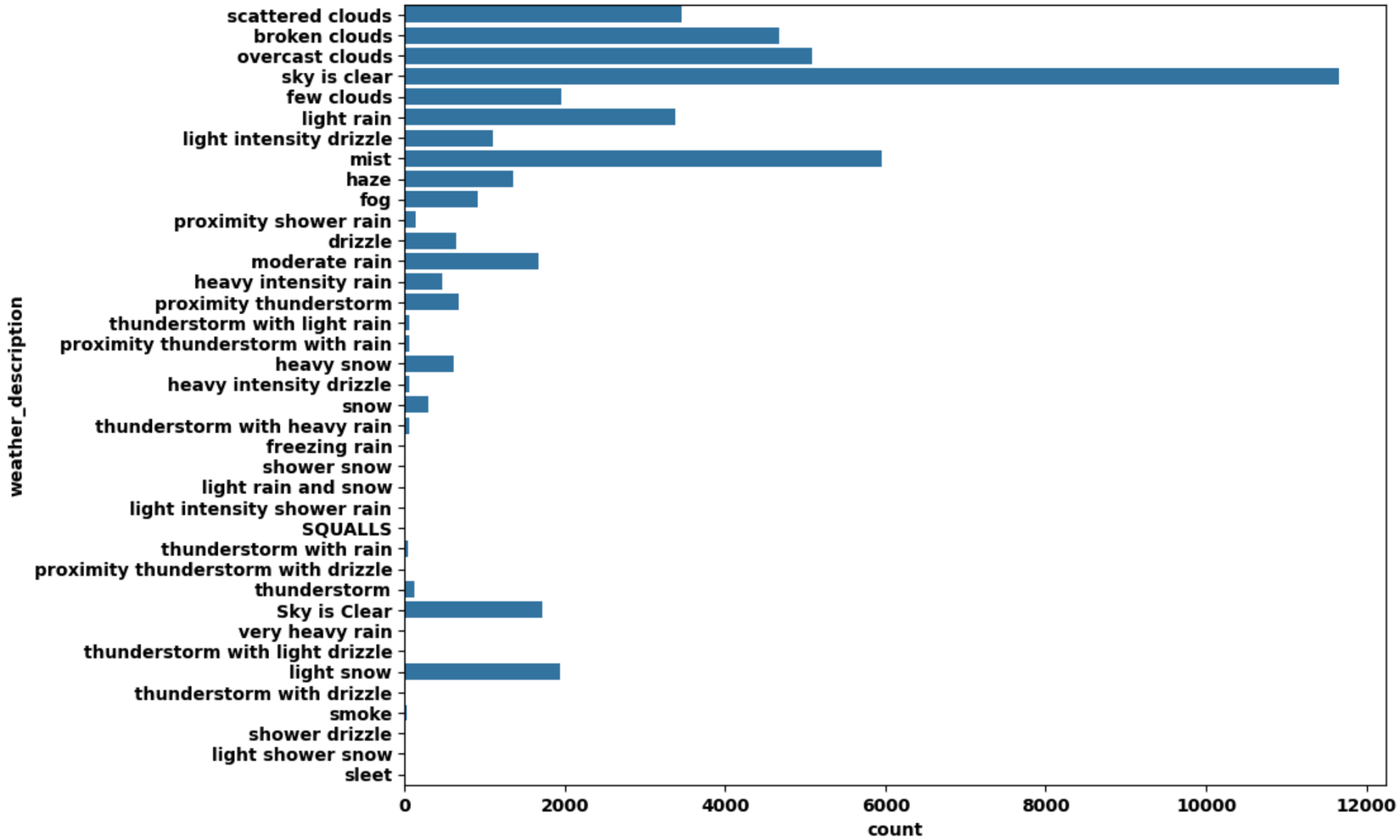
- Density distribution of traffic volume

EXPLANATORY DATA ANALYSIS (EDA)

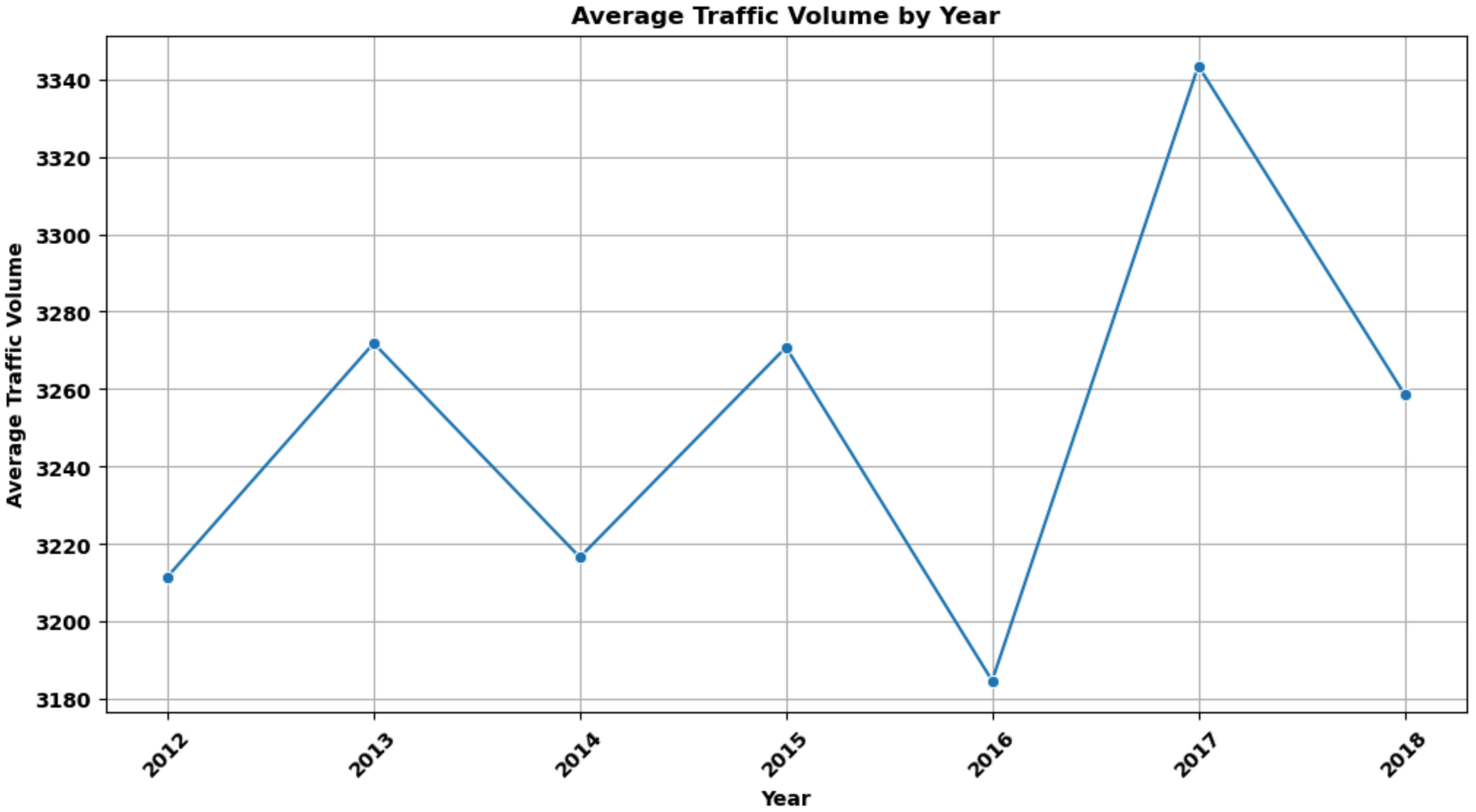


- Count plot of weather_main

EXPLANATORY DATA ANALYSIS (EDA)



- Count plot of weather_description


Average Traffic Volume by Year
Average Traffic Volume
3340
3320
3300
3280
3260
3240
3220
3200
3180
2012
2013
2014
2015
2016
2017
2018
Year

DATA PRE-PROCESSING

The pre-processing consist of five steps:

1- Handle missing values

2- Remove duplicates

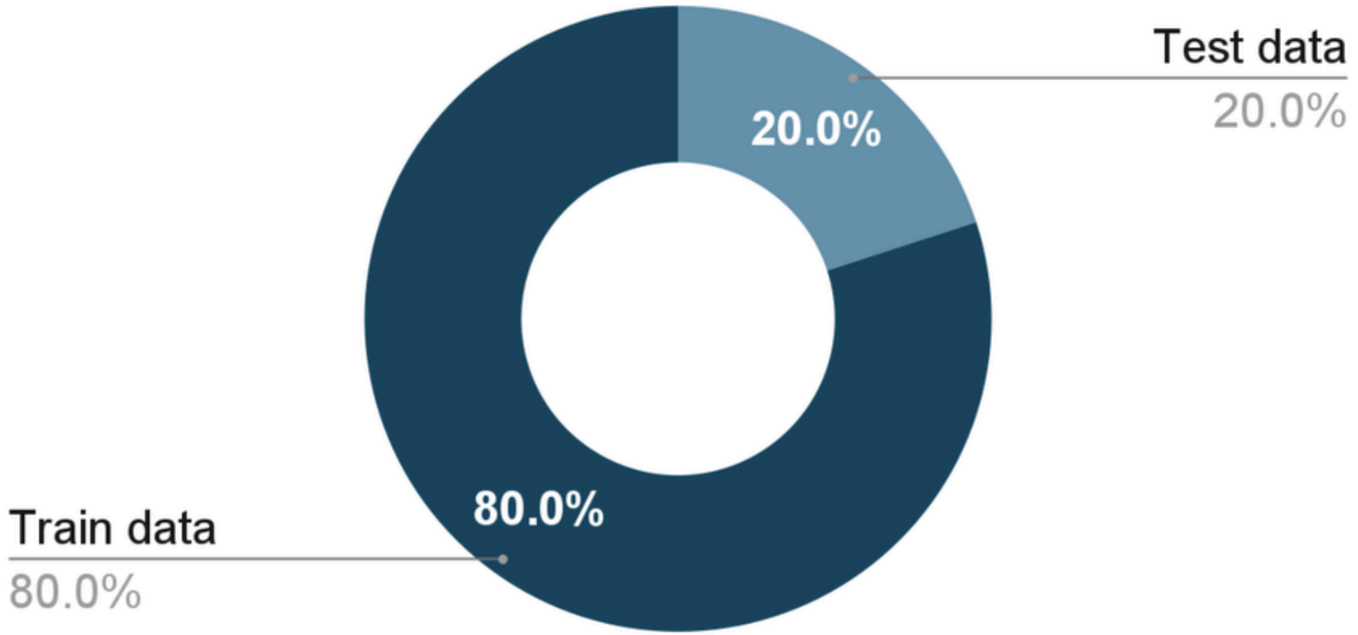
3- Remove outliers

4- Convert to appropriate data types

5- Scale data

IMPLEMENTED MODELS

SPLITTING THE DATASET



- Training set size: 38552, Test set size : 9630

1. RNN

RNN (Recurrent Neural Network) Model for Time Series Forecasting

Model Structure:

- **RNN Layers:**

Two RNN layers with 64 units and return_sequences, and processing the output of the first layer.
Dense Layer:
Single Dense layer with 1 unit, producing the final output value.

Layer (type)	Output Shape	Param #
simple_rnn_4 (SimpleRNN)	(None, 10, 64)	4,800
simple_rnn_5 (SimpleRNN)	(None, 64)	8,256
dense_8 (Dense)	(None, 1)	65

- **Compilation:**

Optimizer: Adam (efficient for large datasets).
Loss Function: Mean Squared Error (suitable for regression tasks).

2. LSTM

LSTM (Long Short-Term Memory) Model for Time Series Forecasting

Model Structure:

- **LSTM Layers:**

Two LSTM layers with 64 units and return_sequences, and processing the output of the first layer.
Dense Layer:
Single Dense layer with 1 unit, producing the final output value.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 10, 64)	18,432
lstm_1 (LSTM)	(None, 64)	33,024
dense (Dense)	(None, 1)	65

- **Compilation:**

Optimizer: Adam (efficient for large datasets).
Loss Function: Mean Squared Error (suitable for regression tasks).

3. GRU

GRU (Gated Recurrent Unit) Model for Time Series Forecasting

Model Structure:

- **GRU Layers:**

Two GRU layers with 64 units and return_sequences, and processing the output of the first layer.
Dense Layer:
Single Dense layer with 1 unit, producing the final output value.

Layer (type)	Output Shape	Param #
gru_4 (GRU)	(None, 10, 64)	14,592
gru_5 (GRU)	(None, 64)	24,960
dense_10 (Dense)	(None, 1)	65

- **Compilation:**

Optimizer: Adam (efficient for large datasets).
Loss Function: Mean Squared Error (suitable for regression tasks).

PERFORMANCE METRICS

Performance Metrics Comparison

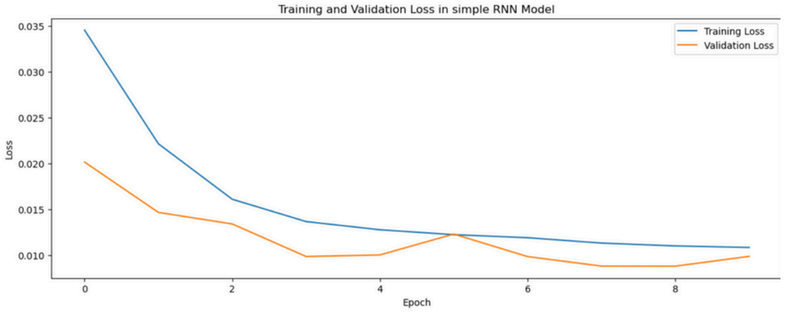
RNN achieved **MAE** of 0.07 , 0.10 and the same **RMSE** of 0.06, 0.09 indicating they have very similar predictive accuracy on the test dataset.

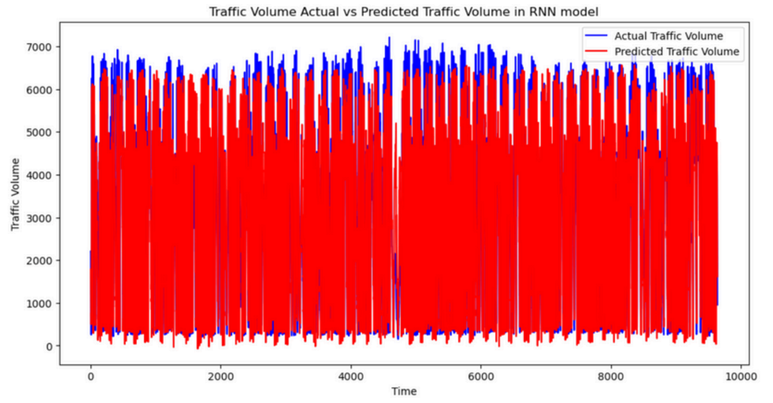
on the other hand, GRU and LSTM achieved similar **MAE** of 0.086 and a **RMSE** of 0.09, suggesting it is performing similarly.

Models	MAE	RMSE
Simple RNN	0.07	0.10
LSTM	0.06	0.09
GRU	0.06	0.09

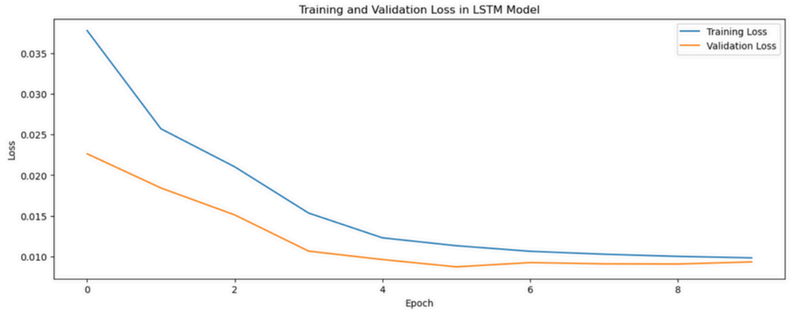
RESULTS AND DISCUSSION

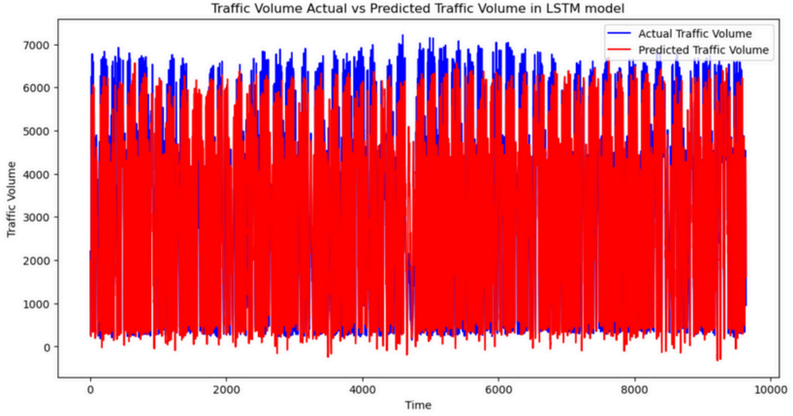
RNN



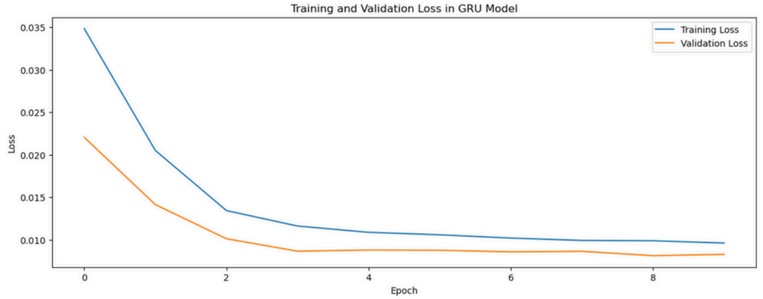


LSTM





GRU



CONCLUSION AND FUTURE WORK

- **The RNN and LSTM, GRU models demonstrate promising performance in predicting traffic volume based on the provided data.**

- **Future work could explore ensemble methods such as XGBoost, or the incorporation of additional relevant features to enhance the overall predictive accuracy across the models.**

THANK YOU FOR LISTENING!