Data Collection and Preprocessing Phase
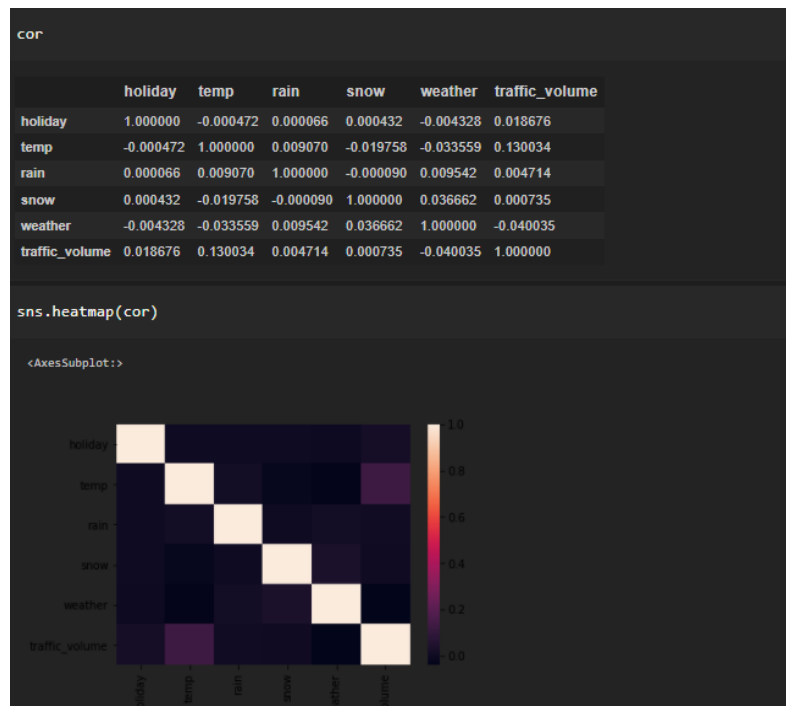
Date	10 July 2024
Team ID	739970
Project Title	**TRAFFICTELLIGANCE-Advanced Traffic Volume Estimation With Machine Learning**
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description

Data Overview	```python
	# used to understand the descriptive analysis of the data
	data.describe()
	``
	
	|
		temp	rain	snow	traffic_volume

---	---	---	---	---
count	48151.000000	48202.000000	48192.000000	48204.000000
mean	281.205351	0.334278	0.000222	3259.818355
std	13.343675	44.790062	0.008169	1986.860670
min	0.000000	0.000000	0.000000	0.000000
25%	272.160000	0.000000	0.000000	1193.000000
50%	282.460000	0.000000	0.000000	3380.000000
75%	291.810000	0.000000	0.000000	4933.000000
max	310.070000	9831.300000	0.510000	7280.000000	

Heat Map	
Pair Plot	

Box Plot	
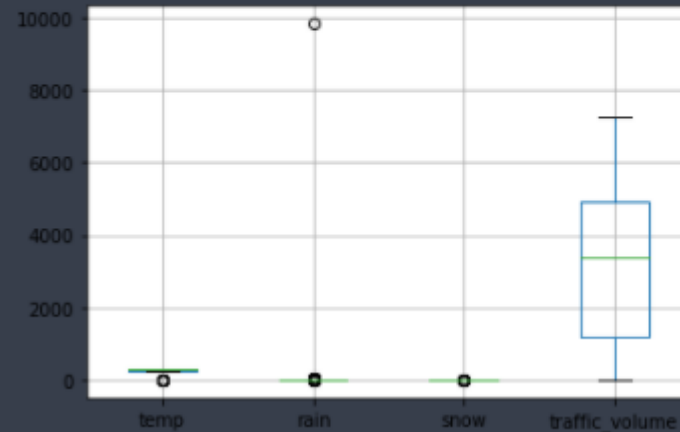
Outliers and Anomalies	-

Data Preprocessing Code Screenshots

Loading Data	

| Handling Missing Data | ```python
data['temp'].fillna(data['temp'].mean(),inplace=True)
data['rain'].fillna(data['rain'].mean(),inplace=True)
data['snow'].fillna(data['snow'].mean(),inplace=True)

print(Counter(data['weather']))

Counter({'Clouds': 15144, 'Clear': 13383, 'Mist': 5942, 'Rain': 5665, 'Snow': 2875, 'Drizzle': 1818, 'H
 'Fog': 912, nan: 49, 'Smoke': 20, 'Squall': 4})

data['weather'].fillna('Clouds',inplace=True)
``` |
| Data Transformation | ```python
spliiting the date column into year,month,day
data[["day", "month", "year"]] = data["date"].str.split("-", expand = True)

spliiting the date column into year,month,day
data[["hours", "minutes", "seconds"]] = data["Time"].str.split(":", expand = True)

data.drop(columns=['date','Time'],axis=1,inplace=True)

data.head()
```

	holiday	temp	rain	snow	weather	traffic_volume	day	month	year	hours	minutes	seconds
0	7	288.28	0.0	0.0	1	5545	02	10	2012	09	00	00
1	7	289.36	0.0	0.0	1	4516	02	10	2012	10	00	00
2	7	289.58	0.0	0.0	1	4767	02	10	2012	11	00	00
3	7	290.13	0.0	0.0	1	5026	02	10	2012	12	00	00
4	7	291.14	0.0	0.0	1	4918	02	10	2012	13	00	00
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-