

Data Collection and Preprocessing Phase

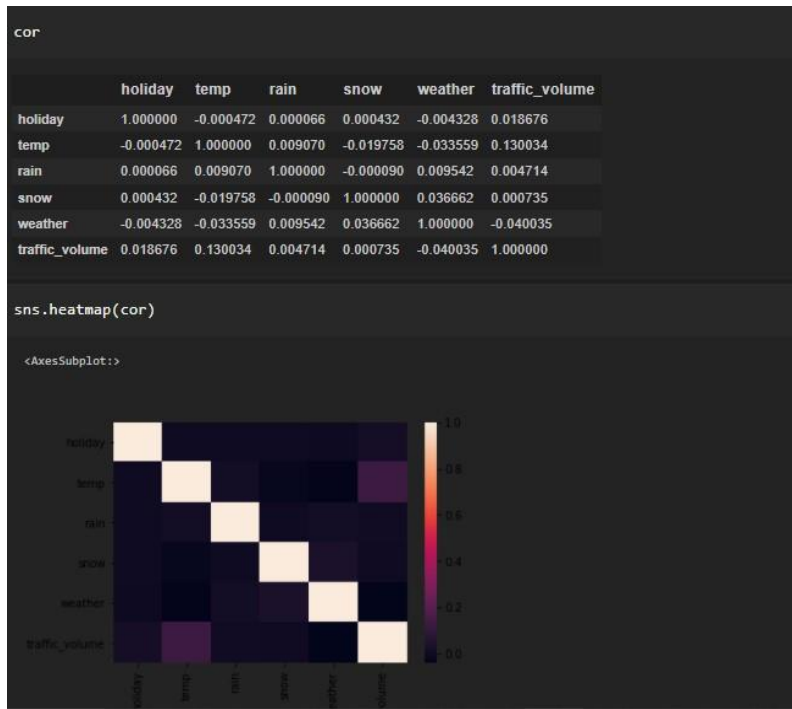
Date	10 July 2024
Team ID	739797
Project Title	TRAFFICTELLIGENCE-Advanced Traffic Volume Estimation with Machine Learning
Maximum Marks	6 Marks

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	<pre># used to understand the descriptive analysis of the data data.describe()</pre>

Heat Map



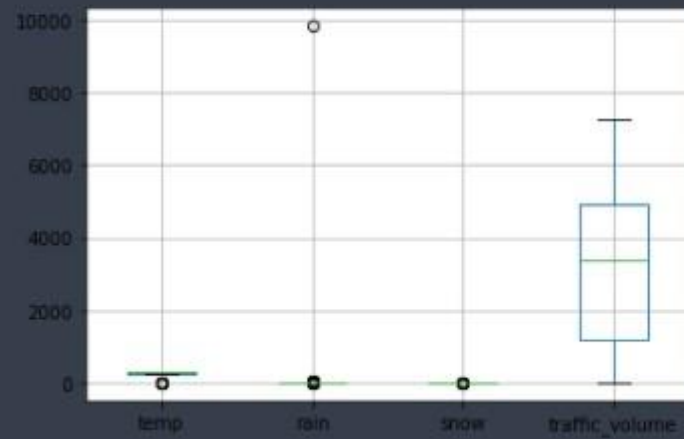
Pair Plot



Box Plot

```
data.boxplot()
```

```
<AxesSubplot:>
```



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
# displaying first 5 columns of the data
data.head()
```

	holiday	temp	rain	snow	weather	date	Time	traffic_volume
0	None	288.28	0.0	0.0	Clouds	02-10-2012	09:00:00	5545
1	None	289.36	0.0	0.0	Clouds	02-10-2012	10:00:00	4516
2	None	289.58	0.0	0.0	Clouds	02-10-2012	11:00:00	4767
3	None	290.13	0.0	0.0	Clouds	02-10-2012	12:00:00	5026
4	None	291.14	0.0	0.0	Clouds	02-10-2012	13:00:00	4918

Handling Missing Data	<pre>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)</pre>																																																																														
Data Transformation	<pre># splitting the date column into year,month,day data[["day", "month", "year"]] = data["date"].str.split("-", expand = True) # splitting 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()</pre> <table><thead><tr><th></th><th>holiday</th><th>temp</th><th>rain</th><th>snow</th><th>weather</th><th>traffic_volume</th><th>day</th><th>month</th><th>year</th><th>hours</th><th>minutes</th><th>seconds</th></tr></thead><tbody><tr><td>0</td><td>7</td><td>288.28</td><td>0.0</td><td>0.0</td><td>1</td><td>5545</td><td>02</td><td>10</td><td>2012</td><td>09</td><td>00</td><td>00</td></tr><tr><td>1</td><td>7</td><td>289.36</td><td>0.0</td><td>0.0</td><td>1</td><td>4516</td><td>02</td><td>10</td><td>2012</td><td>10</td><td>00</td><td>00</td></tr><tr><td>2</td><td>7</td><td>289.58</td><td>0.0</td><td>0.0</td><td>1</td><td>4767</td><td>02</td><td>10</td><td>2012</td><td>11</td><td>00</td><td>00</td></tr><tr><td>3</td><td>7</td><td>290.13</td><td>0.0</td><td>0.0</td><td>1</td><td>5026</td><td>02</td><td>10</td><td>2012</td><td>12</td><td>00</td><td>00</td></tr><tr><td>4</td><td>7</td><td>291.14</td><td>0.0</td><td>0.0</td><td>1</td><td>4918</td><td>02</td><td>10</td><td>2012</td><td>13</td><td>00</td><td>00</td></tr></tbody></table>		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
	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	-																																																																														