Data Collection and Preprocessing Phase

Date	20 June 2025
Team ID	LTVIP2025TMID44004
Project Title	Traffictelligence-Advanced-Traffic-Volume Es ma on-With-Machine-Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

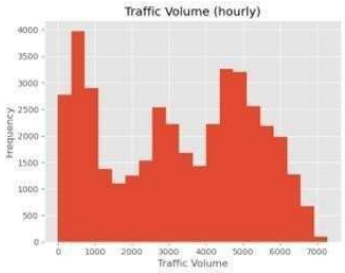
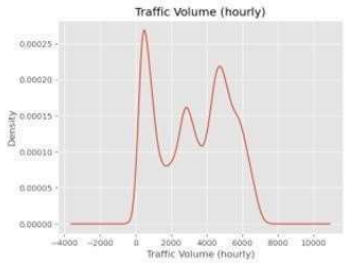
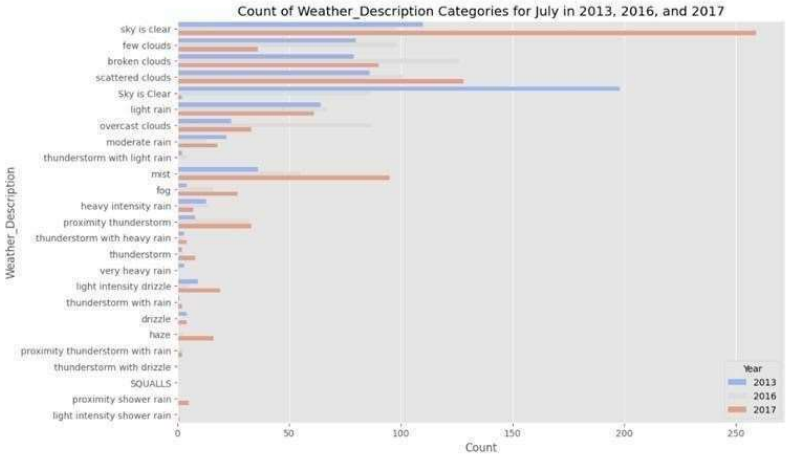
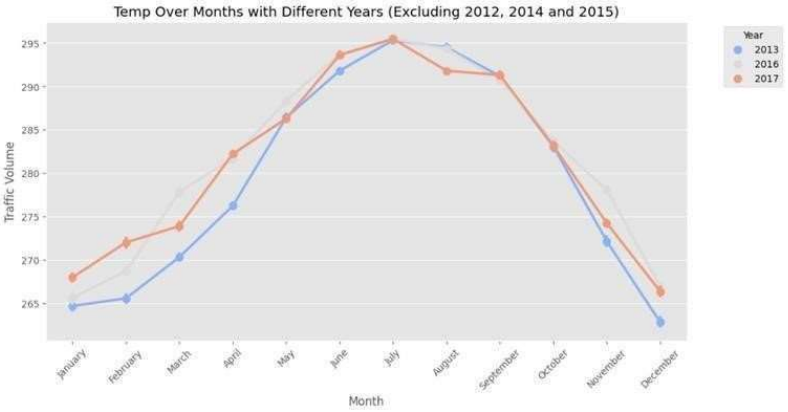
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

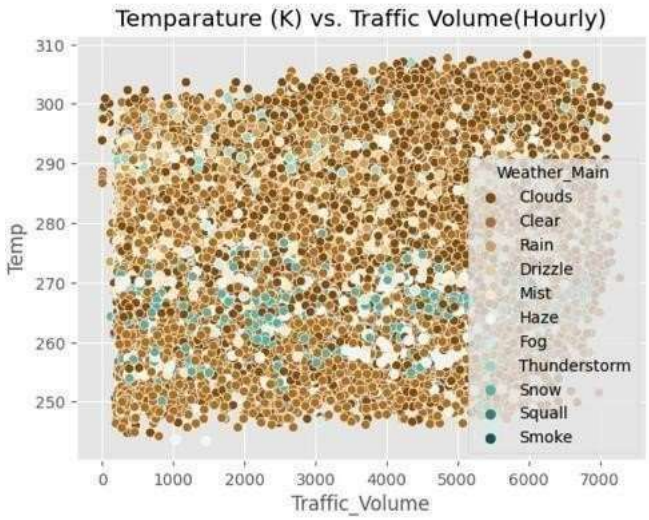
Section	Description

Data Overview	Dimension:- 40632 rows*12 columns

	Holiday	Temp	Rain_1h	Snow_1h	Clouds_All	Weather_Main	Weather_Description	Date_Time	Year	Month	Day	Traffic_Volume
580	NaN	289.06	0.0	0.0	90	Mist	mist	2012-10-24 19:00:00	2012	10	24	3118
6421	NaN	289.06	0.0	0.0	90	Clouds	overcast clouds	2013-05-26 15:00:00	2013	5	26	3588
6605	NaN	289.06	0.0	0.0	1	Clear	sky is clear	2013-06-02 01:00:00	2013	6	2	787
6870	NaN	289.06	0.0	0.0	92	Mist	mist	2013-06-11 00:00:00	2013	6	11	576
6902	NaN	289.06	0.0	0.0	8	Mist	mist	2013-06-12 01:00:00	2013	6	12	377
17564	NaN	289.06	0.0	0.0	75	Clouds	broken clouds	2015-08-19 19:00:00	2015	8	19	3318
17677	NaN	289.06	0.0	0.0	90	Clouds	overcast clouds	2015-08-23 23:00:00	2015	8	23	1041
17747	NaN	289.06	0.0	0.0	40	Clouds	scattered clouds	2015-08-26 21:00:00	2015	8	26	2812
23850	NaN	289.06	0.0	0.0	90	Clouds	overcast clouds	2016-06-01 10:00:00	2016	6	1	4831
23851	NaN	289.06	0.0	0.0	90	Clouds	overcast clouds	2016-06-01 10:00:00	2016	6	1	4831
26108	NaN	289.06	0.0	0.0	90	Fog	fog	2016-08-28 07:00:00	2016	8	28	1228
26109	NaN	289.06	0.0	0.0	90	Mist	mist	2016-08-28 07:00:00	2016	8	28	1228
26110	NaN	289.06	0.0	0.0	90	Rain	light rain	2016-08-28 07:00:00	2016	8	28	1228
26297	NaN	289.06	0.0	0.0	1	Clear	sky is clear	2016-09-04 04:00:00	2016	9	4	360
26972	NaN	289.06	0.0	0.0	12	Clouds	few clouds	2016-09-29 12:00:00	2016	9	29	4484

Descrip ve Sta s cs:-

Univariate Analysis	 
Bivariate Analysis	 

Multivariate Analysis	
	
Outliers and Anomalies	extreme weather, special events, accidents, data errors, or unusual traffic pa erns
Data Preprocessing Code Sc reenshots	

Loading Data

```
df = pd.read_csv(r'C:\Users\bhart\OneDrive\Desktop\Model Deployment\Metro_Interstate_Traffic_Volume_test (2).csv')
df = pd.read_csv(r'C:\Users\bhart\OneDrive\Desktop\Model Deployment\Metro_Interstate_Traffic_Volume_train.csv')
```

1.UNDERSTANDING THE DATA

```
df.shape
```

```
(40255, 14)
```

```
df.head(5)
```

	Unnamed: 0	holiday	temp	rain_1h	snow_1h	clouds_all	weather_main	weather_description	date_time	year	month	day	hour	traffic_volume
0	0	NaN	288.28	0.0	0.0	40	Clouds	scattered clouds	2012-10-02 09:00:00	2012	10	2	09:00	5545
1	1	NaN	289.36	0.0	0.0	75	Clouds	broken clouds	2012-10-02 10:00:00	2012	10	2	10:00	4516
2	2	NaN	289.58	0.0	0.0	90	Clouds	overcast clouds	2012-10-02 11:00:00	2012	10	2	11:00	4767
3	3	NaN	290.13	0.0	0.0	90	Clouds	overcast clouds	2012-10-02 12:00:00	2012	10	2	12:00	5026
4	4	NaN	291.14	0.0	0.0	75	Clouds	broken clouds	2012-10-02 13:00:00	2012	10	2	13:00	4918

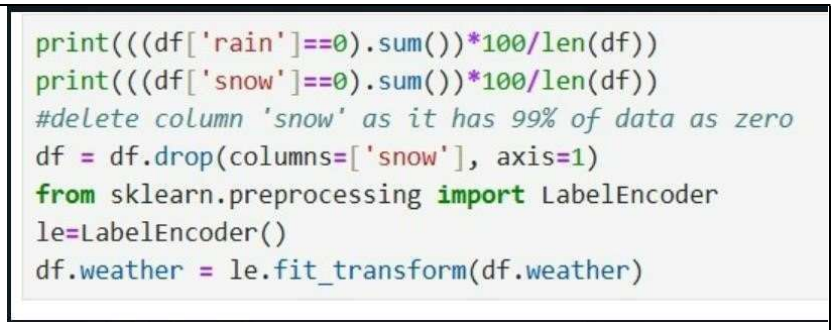
Handling Missing Data

```
df = pd.read_csv('traffic_volume.csv')
print(df.head())
print(df.shape)
print(((df.isnull().sum())*100)/len(df))
```

```
  holiday    temp  rain  snow weather        date      Time  traffic_volume
0     NaN  288.28   0.0   0.0  Clouds  02-10-2012  09:00:00            5545
1     NaN  289.36   0.0   0.0  Clouds  02-10-2012  10:00:00            4516
2     NaN  289.58   0.0   0.0  Clouds  02-10-2012  11:00:00            4767
3     NaN  290.13   0.0   0.0  Clouds  02-10-2012  12:00:00            5026
4     NaN  291.14   0.0   0.0  Clouds  02-10-2012  13:00:00            4918
(48204, 8)
holiday           99.873454
temp               0.109949
rain               0.004149
snow               0.024894
weather            0.101651
date               0.000000
Time               0.000000
traffic_volume     0.000000
dtype: float64
```

```
# Delete column 'holiday'
# delete the rows wit null values in 'temp', 'rain', 'snow', 'weather'
#--------------Handling NUll values-------------
```

```
df=df.drop(columns=['holiday'], axis=1)
df.dropna(inplace=True)
print(df.shape)
print(df.isnull().sum())
```

Data Transformation	
Feature Engineering	Attached the codes in final Submission
Save Processed Data