

## Data Collection and Preprocessing Phase

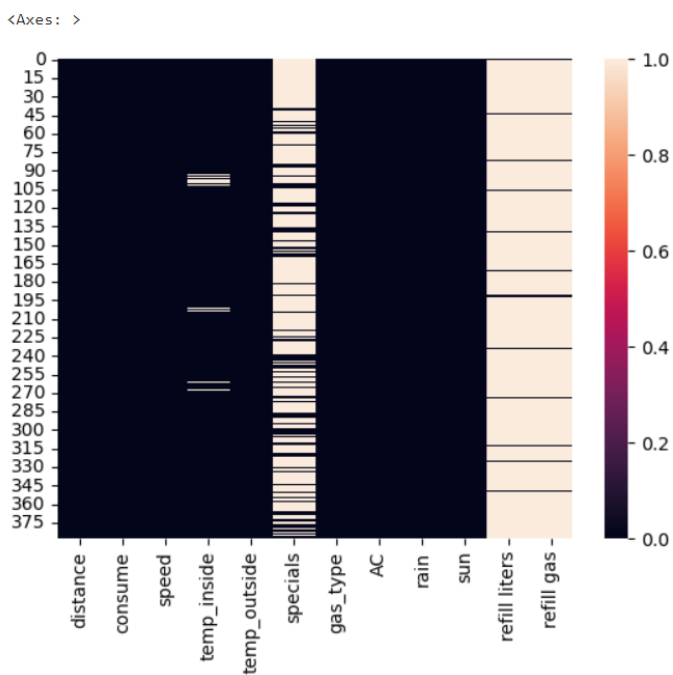
Date	8 July 2024
Team ID	SWTID1720195303
Project Title	Predictive Modeling For Fleet Fuel Management Using Machine Learning
Maximum Marks	6 Marks

## Data Exploration and Preprocessing Template

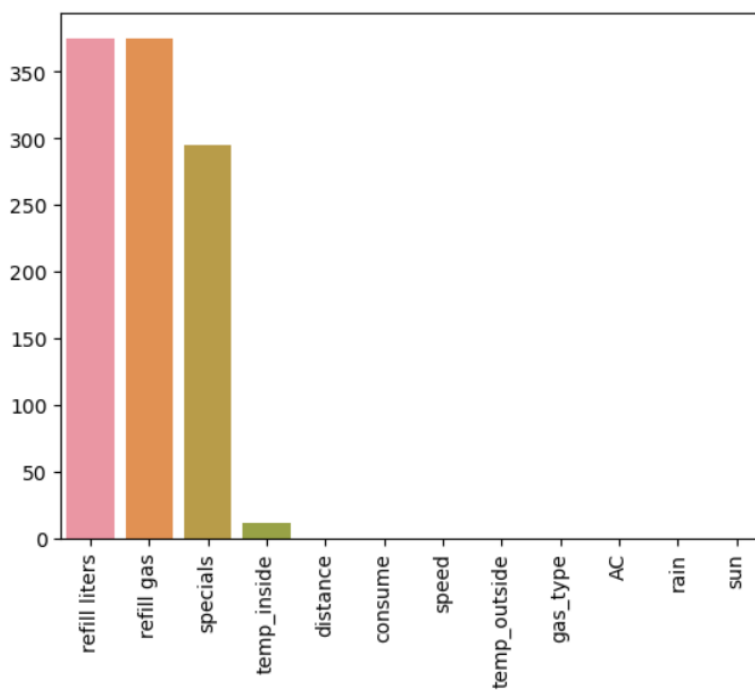
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	<pre> distance  consume  speed  temp_inside  temp_outside  specials  gas_type  AC  \ 0      28.0      5.0    26      21.5           12      NaN      E10    0 1      12.0      4.2    30      21.5           13      NaN      E10    0 2      11.2      5.5    38      21.5           15      NaN      E10    0 3      12.9      3.9    36      21.5           14      NaN      E10    0 4      18.5      4.5    46      21.5           15      NaN      E10    0  rain  sun  refill  liters  refill  gas 0     0    0      45.0    E10 1     0    0      NaN     NaN 2     0    0      NaN     NaN 3     0    0      NaN     NaN 4     0    0      NaN     NaN </pre>

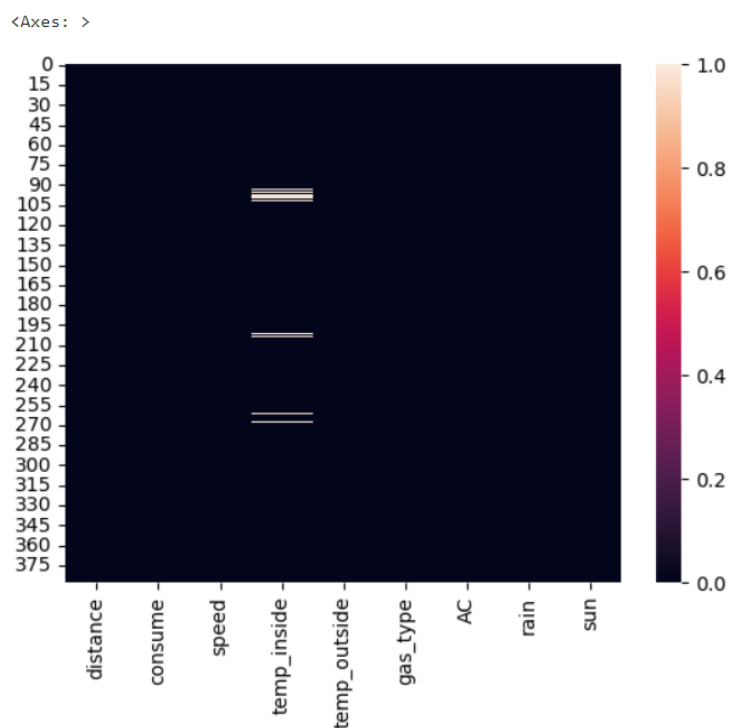
## Univariate Analysis



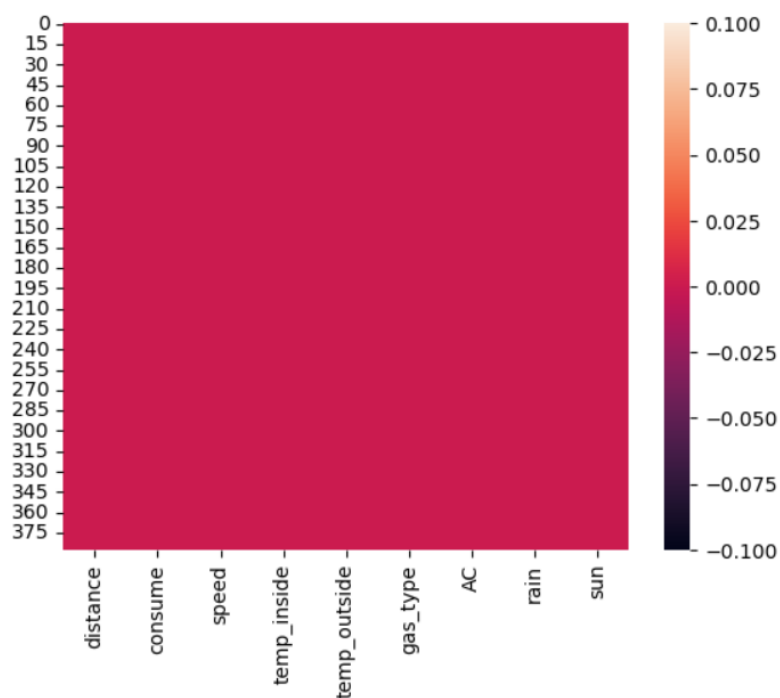
## Bivariate Analysis



## Multivariate Analysis



## Outliers and Anomalies



## Data Preprocessing Code Screenshots

## Loading Data

```
data2 = pd.read_excel('measurements2.xlsx')
```

```
print(data2.head())
```

	distance	consume	speed	temp_inside	temp_outside	specials	gas_type	AC	\
0	28.0	5.0	26	21.5	12	NaN	E10	0	
1	12.0	4.2	30	21.5	13	NaN	E10	0	
2	11.2	5.5	38	21.5	15	NaN	E10	0	
3	12.9	3.9	36	21.5	14	NaN	E10	0	
4	18.5	4.5	46	21.5	15	NaN	E10	0	

	rain	sun	refill liters	refill gas
0	0	0	45.0	E10
1	0	0	NaN	NaN
2	0	0	NaN	NaN
3	0	0	NaN	NaN
4	0	0	NaN	NaN

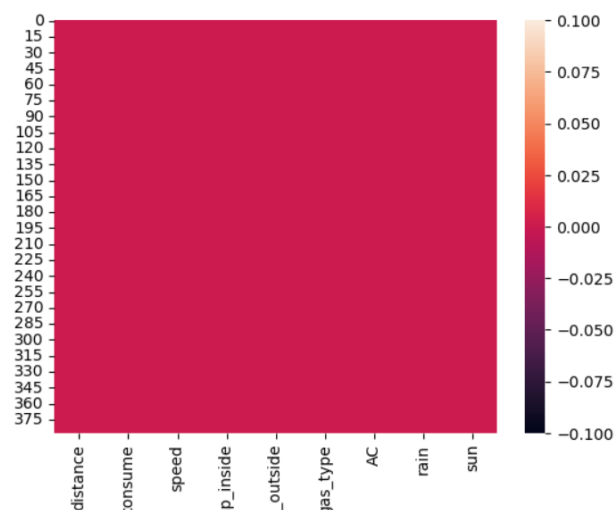
## Handling Missing Data

```
[8]: temp_inside_mean=np.mean(data2['temp_inside'])
print(temp_inside_mean)
```

```
21.929521276595743
```

```
[9]: data2['temp_inside'].fillna(temp_inside_mean,inplace=True)
sns.heatmap(data2.isnull())
```

```
[9]: <Axes: >
```



Data Transformation	<pre>from sklearn import metrics print(metrics.mean_squared_error(y_test,y_pred)) print(metrics.mean_absolute_error(y_test,y_pred)) print(np.sqrt(metrics.mean_squared_error(y_test,y_pred)))</pre> <p>0.7424532609047074 0.6635761182069616 0.861657275780056</p> <pre>dum1=pd.get_dummies(data2['gas_type']) print(dum1)</pre> <pre>      E10  SP98 0    True  False 1    True  False 2    True  False 3    True  False 4    True  False ..     ...   ... 383  False   True 384  False   True 385  False   True 386  False   True 387  False   True  [388 rows x 2 columns]</pre> <pre>data2=pd.concat([data2,dum1],axis=1)</pre>
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-