

Data Collection and Preprocessing Phase

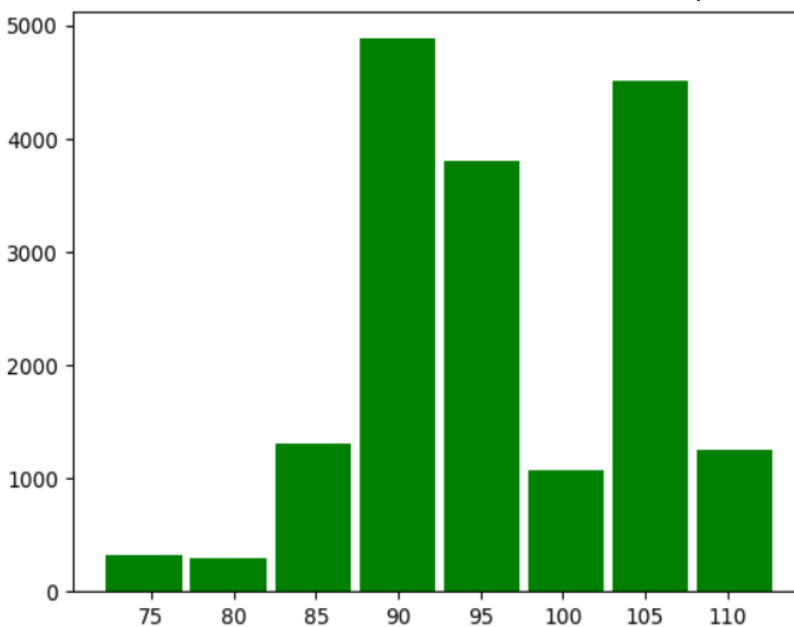
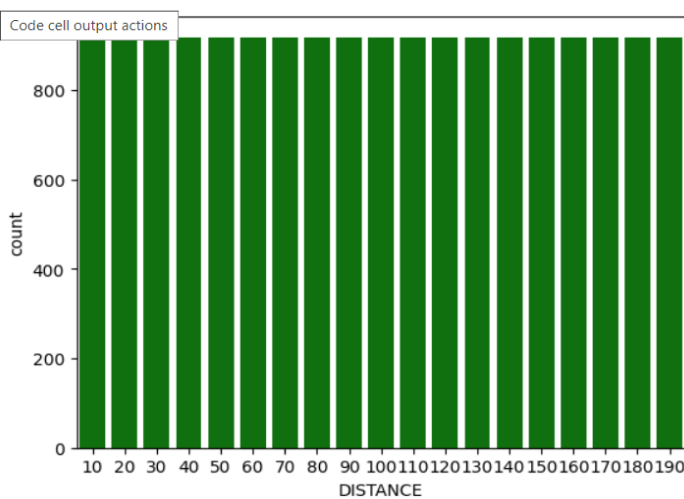
Date	15 March 2024
Team ID	740139
Project Title	Acoustic Fire Extinguishing Prediction
Maximum Marks	6 Marks

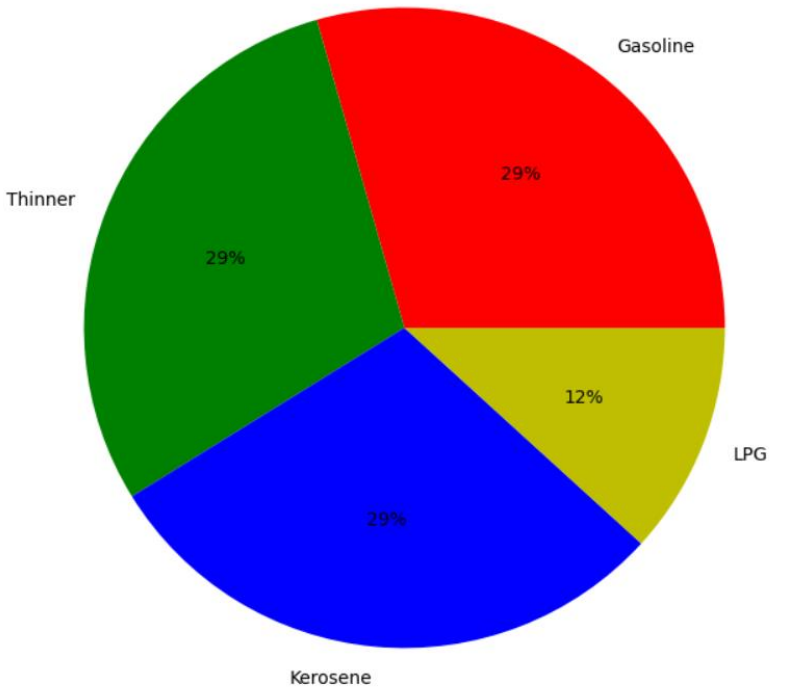
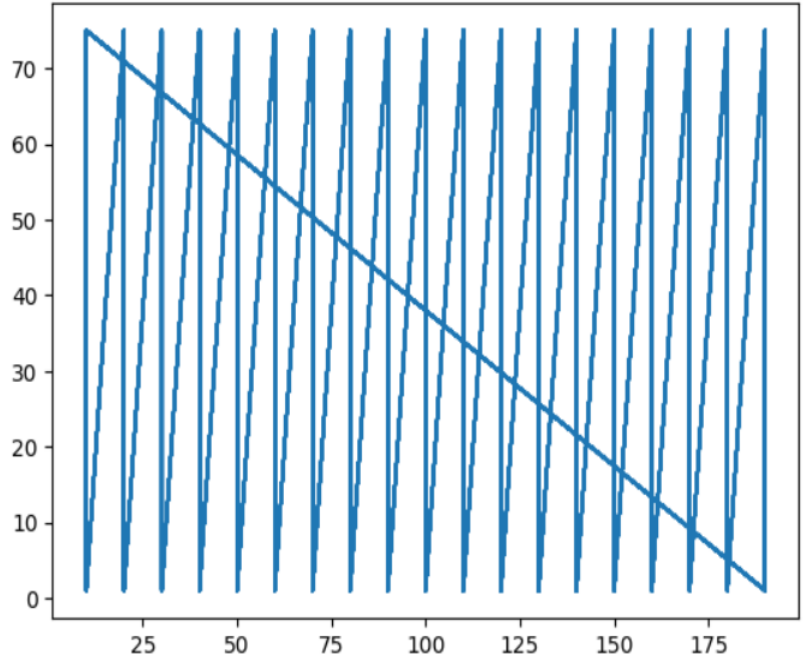
Data Exploration and Preprocessing Template

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

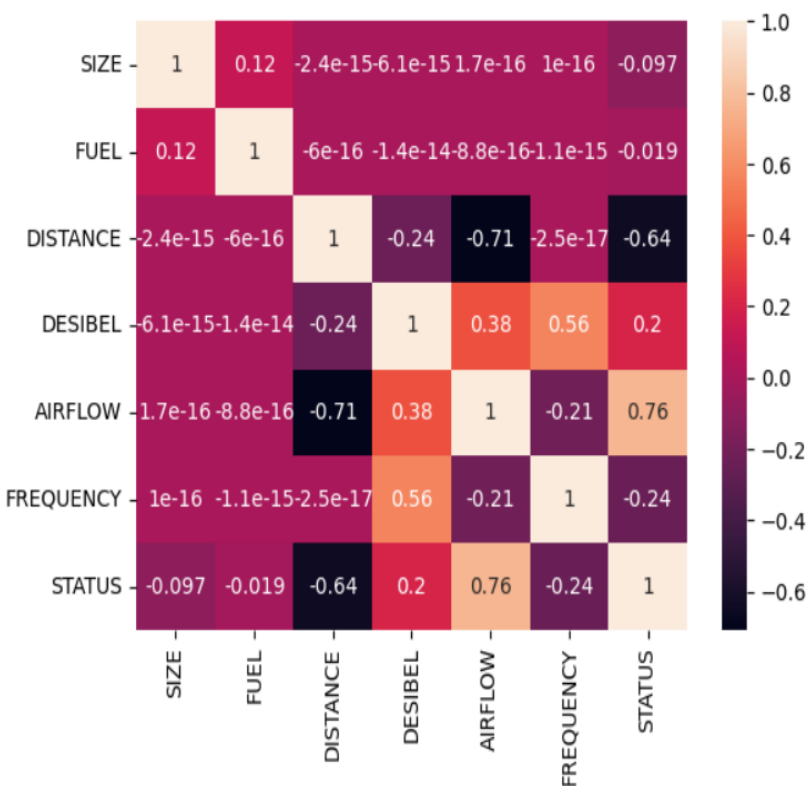
Section	Description																																																																								
Data Overview	Dimensions: 602rows*9columns Descriptive statistics:																																																																								
	<table><tr><th></th><th>SIZE</th><th>FUEL</th><th>DISTANCE</th><th>DESTBEL</th><th>AIRFLOW</th><th>FREQUENCY</th><th>STATUS</th></tr><tr><td>count</td><td>17442.000000</td><td>17442.000000</td><td>17442.000000</td><td>17442.000000</td><td>17442.000000</td><td>17442.000000</td><td>17442.000000</td></tr><tr><td>mean</td><td>3.411765</td><td>1.411765</td><td>100.000000</td><td>96.379142</td><td>6.975634</td><td>31.611111</td><td>0.497821</td></tr><tr><td>std</td><td>1.750977</td><td>1.191120</td><td>54.773826</td><td>8.164096</td><td>4.736169</td><td>20.939149</td><td>0.500010</td></tr><tr><td>min</td><td>1.000000</td><td>0.000000</td><td>10.000000</td><td>72.000000</td><td>0.000000</td><td>1.000000</td><td>0.000000</td></tr><tr><td>25%</td><td>2.000000</td><td>0.000000</td><td>50.000000</td><td>90.000000</td><td>3.200000</td><td>14.000000</td><td>0.000000</td></tr><tr><td>50%</td><td>3.000000</td><td>1.000000</td><td>100.000000</td><td>95.000000</td><td>5.800000</td><td>27.500000</td><td>0.000000</td></tr><tr><td>75%</td><td>5.000000</td><td>3.000000</td><td>150.000000</td><td>104.000000</td><td>11.200000</td><td>47.000000</td><td>1.000000</td></tr><tr><td>max</td><td>7.000000</td><td>3.000000</td><td>190.000000</td><td>113.000000</td><td>17.000000</td><td>75.000000</td><td>1.000000</td></tr></table>		SIZE	FUEL	DISTANCE	DESTBEL	AIRFLOW	FREQUENCY	STATUS	count	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000	mean	3.411765	1.411765	100.000000	96.379142	6.975634	31.611111	0.497821	std	1.750977	1.191120	54.773826	8.164096	4.736169	20.939149	0.500010	min	1.000000	0.000000	10.000000	72.000000	0.000000	1.000000	0.000000	25%	2.000000	0.000000	50.000000	90.000000	3.200000	14.000000	0.000000	50%	3.000000	1.000000	100.000000	95.000000	5.800000	27.500000	0.000000	75%	5.000000	3.000000	150.000000	104.000000	11.200000	47.000000	1.000000	max	7.000000	3.000000	190.000000	113.000000	17.000000	75.000000	1.000000
		SIZE	FUEL	DISTANCE	DESTBEL	AIRFLOW	FREQUENCY	STATUS																																																																	
	count	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000	17442.000000																																																																	
	mean	3.411765	1.411765	100.000000	96.379142	6.975634	31.611111	0.497821																																																																	
	std	1.750977	1.191120	54.773826	8.164096	4.736169	20.939149	0.500010																																																																	
	min	1.000000	0.000000	10.000000	72.000000	0.000000	1.000000	0.000000																																																																	
	25%	2.000000	0.000000	50.000000	90.000000	3.200000	14.000000	0.000000																																																																	
	50%	3.000000	1.000000	100.000000	95.000000	5.800000	27.500000	0.000000																																																																	
	75%	5.000000	3.000000	150.000000	104.000000	11.200000	47.000000	1.000000																																																																	
max	7.000000	3.000000	190.000000	113.000000	17.000000	75.000000	1.000000																																																																		

Univariate Analysis



	<p>Pie chart showing the distribution of Fuel of the Fire</p>  <table border="1"> <thead> <tr> <th>Fuel Type</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Gasoline</td> <td>29%</td> </tr> <tr> <td>Thinner</td> <td>29%</td> </tr> <tr> <td>Kerosene</td> <td>29%</td> </tr> <tr> <td>LPG</td> <td>12%</td> </tr> </tbody> </table>	Fuel Type	Percentage	Gasoline	29%	Thinner	29%	Kerosene	29%	LPG	12%								
Fuel Type	Percentage																		
Gasoline	29%																		
Thinner	29%																		
Kerosene	29%																		
LPG	12%																		
<p>Bivariate Analysis</p>	 <table border="1"> <thead> <tr> <th>X-axis Value</th> <th>Y-axis Value (approx.)</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>75</td> </tr> <tr> <td>25</td> <td>65</td> </tr> <tr> <td>50</td> <td>55</td> </tr> <tr> <td>75</td> <td>45</td> </tr> <tr> <td>100</td> <td>35</td> </tr> <tr> <td>125</td> <td>25</td> </tr> <tr> <td>150</td> <td>15</td> </tr> <tr> <td>175</td> <td>5</td> </tr> </tbody> </table>	X-axis Value	Y-axis Value (approx.)	0	75	25	65	50	55	75	45	100	35	125	25	150	15	175	5
X-axis Value	Y-axis Value (approx.)																		
0	75																		
25	65																		
50	55																		
75	45																		
100	35																		
125	25																		
150	15																		
175	5																		

Multivariate Analysis



Outliers and Anomalies

—

Data Preprocessing Code Screenshots

Loading Data

	SIZE	FUEL	DISTANCE	DESIBEL	AIRFLOW	FREQUENCY	STATUS
0	1	gasoline	10	96	0.0	75	0
1	1	gasoline	10	96	0.0	72	1
2	1	gasoline	10	96	2.6	70	1
3	1	gasoline	10	96	3.2	68	1
4	1	gasoline	10	109	4.5	67	1

Handling Missing Data	<pre>[] data.isnull().sum()</pre> <div><div></div><div>SIZE0 FUEL0 DISTANCE0 DESIBEL0 AIRFLOW0 FREQUENCY0 STATUS0 dtype: int64</div></div>																																																
Data Transformation	<pre>[] lb = LabelEncoder()</pre> <pre>[] data["FUEL"] = lb.fit_transform(data["FUEL"])</pre> <div>New Section</div> <div>▼ New Section</div> <pre>[] data.head()</pre> <div><div></div><div><table><thead><tr><th></th><th>SIZE</th><th>FUEL</th><th>DISTANCE</th><th>DESIBEL</th><th>AIRFLOW</th><th>FREQUENCY</th><th>STATUS</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>0</td><td>10</td><td>96</td><td>0.0</td><td>75</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>10</td><td>96</td><td>0.0</td><td>72</td><td>1</td></tr><tr><td>2</td><td>1</td><td>0</td><td>10</td><td>96</td><td>2.6</td><td>70</td><td>1</td></tr><tr><td>3</td><td>1</td><td>0</td><td>10</td><td>96</td><td>3.2</td><td>68</td><td>1</td></tr><tr><td>4</td><td>1</td><td>0</td><td>10</td><td>109</td><td>4.5</td><td>67</td><td>1</td></tr></tbody></table></div></div>		SIZE	FUEL	DISTANCE	DESIBEL	AIRFLOW	FREQUENCY	STATUS	0	1	0	10	96	0.0	75	0	1	1	0	10	96	0.0	72	1	2	1	0	10	96	2.6	70	1	3	1	0	10	96	3.2	68	1	4	1	0	10	109	4.5	67	1
	SIZE	FUEL	DISTANCE	DESIBEL	AIRFLOW	FREQUENCY	STATUS																																										
0	1	0	10	96	0.0	75	0																																										
1	1	0	10	96	0.0	72	1																																										
2	1	0	10	96	2.6	70	1																																										
3	1	0	10	96	3.2	68	1																																										
4	1	0	10	109	4.5	67	1																																										
Feature Engineering	Attached the codes in final submission																																																
Save Processed Data	—																																																