Analysis Report

1. Summary Statistics

	a	b	c	d	e	f	g	h	i
count	214.000000	214.000000	214.000000	214.000000	214.000000	214.000000	214.000000	214.000000	214.000000
mean	1.518365	13.407850	2.684533	1.444907	72.650935	0.497056	8.956963	0.175047	0.057009
std	0.003037	0.816604	1.442408	0.499270	0.774546	0.652192	1.423153	0.497219	0.097439
min	1.511150	10.730000	0.000000	0.290000	69.810000	0.000000	5.430000	0.000000	0.000000
25%	1.516522	12.907500	2.115000	1.190000	72.280000	0.122500	8.240000	0.000000	0.000000
50%	1.517680	13.300000	3.480000	1.360000	72.790000	0.555000	8.600000	0.000000	0.000000
75%	1.519157	13.825000	3.600000	1.630000	73.087500	0.610000	9.172500	0.000000	0.100000
max	1.533930	17.380000	4.490000	3.500000	75.410000	6.210000	16.190000	3.150000	0.510000
range	0.022780	6.650000	4.490000	3.210000	5.600000	6.210000	10.760000	3.150000	0.510000
IQR	0.002635	0.917500	1.485000	0.440000	0.807500	0.487500	0.932500	0.000000	0.100000

The levels of additives from 'b' to 'i' show greater variability compared to additive 'a'. This is because of the large standard deviation and range. Additive 'a' has a lower mean and narrower variability. Its standard deviation is 0.003037 and its range is 0.022780.

These summary statistics show that the petrol formulations have different levels of additives. Due to the significant differences in mean values, standard deviations, ranges and IQRs, it is likely that the formulations have distinct characteristics in terms of burning pattern and additive compositions.

2. Correlation Results

	a	b	c	d	e	f	g	h	i
a	1.000000	-0.191885	-0.122274	-0.407326	-0.542052	-0.289833	0.810403	-0.000386	0.143010
b	-0.191885	1.000000	-0.273732	0.156794	-0.069809	-0.266087	-0.275442	0.326603	-0.241346
c	-0.122274	-0.273732	1.000000	-0.481799	-0.165927	0.005396	-0.443750	-0.492262	0.083060
d	-0.407326	0.156794	-0.481799	1.000000	-0.005524	0.325958	-0.259592	0.479404	-0.074402
e	-0.542052	-0.069809	-0.165927	-0.005524	1.000000	-0.193331	-0.208732	-0.102151	-0.094201
f	-0.289833	-0.266087	0.005396	0.325958	-0.193331	1.000000	-0.317836	-0.042618	-0.007719
g	0.810403	-0.275442	-0.443750	-0.259592	-0.208732	-0.317836	1.000000	-0.112841	0.124968
h	-0.000386	0.326603	-0.492262	0.479404	-0.102151	-0.042618	-0.112841	1.000000	-0.058692
i	0.143010	-0.241346	0.083060	-0.074402	-0.094201	-0.007719	0.124968	-0.058692	1.000000

- Additive 'a' tends to have a strong negative relationship with additives 'e' with coefficient of -0.54. This means that when the levels of additive 'a' are high, the levels of additive 'e' is usually low and vice versa.
- Additive 'b' has a weak negative relationship with additives 'c' and 'g' for coefficient of -0.27 and -0.28 respectively. This suggests that when the levels of additive 'b' change, the levels of additives 'c' and 'g' may change slightly in the opposite direction.
- Moderate negative relationship occurs between additives 'c' and 'd' of coefficient -0.48. This suggests that when the levels of additive 'c' increase, the levels of additive 'd' tend to decrease and vice versa.
- Additive 'd' shows a moderate positive relationship with additive 'h' of 0.48. This suggests that as the levels of additive 'd' change, the levels of additive 'h' may change too in the same direction.
- Additive 'h' has weak negative relationships with additive 'i' of coefficient -0.06 . This suggests that there is only slight changes in the opposite directions.
- A weak positive relationship occurs between additives 'i' and 'g' for coefficient 0.13. This suggests that when the levels of additive 'i' increase, the levels of additive 'g' may also increase slightly.

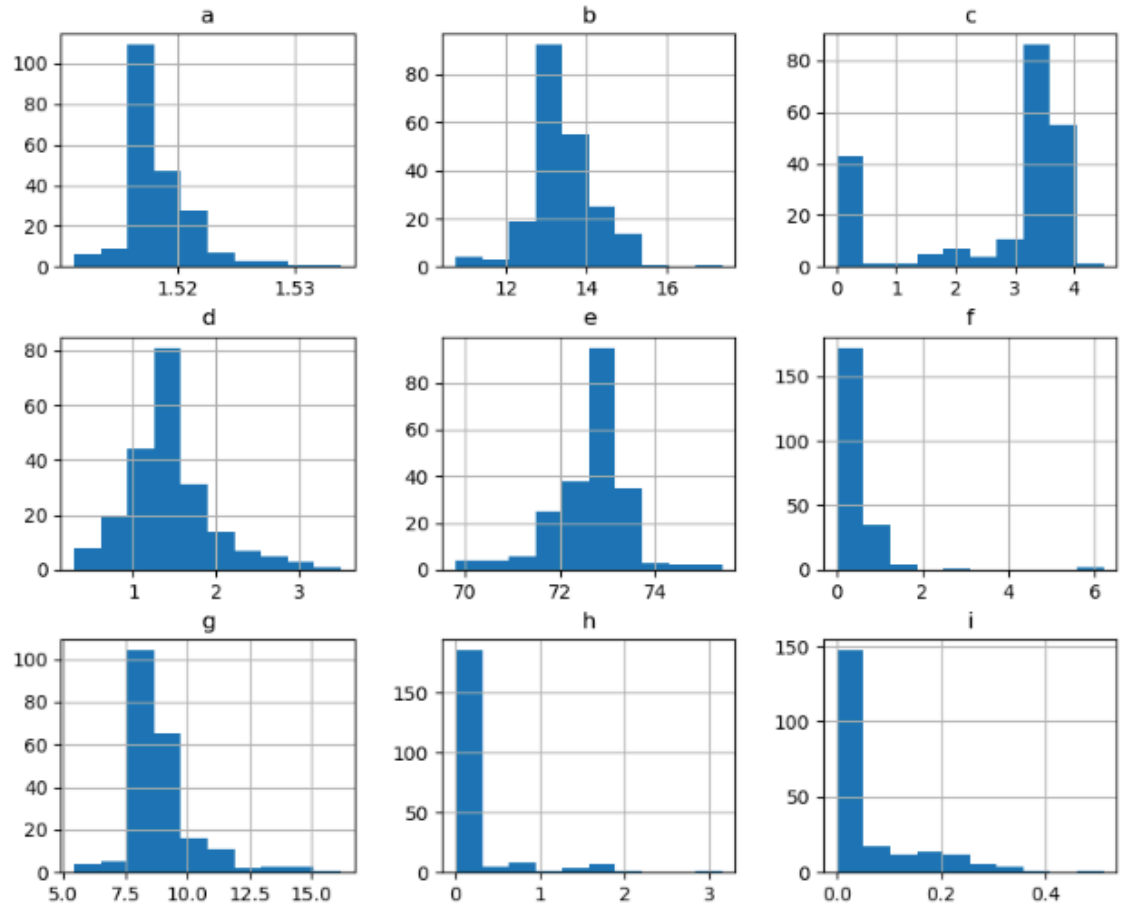
3. ANOVA Results

statistic=168331.96885835085, pvalue=0.0

Based on the analysis, a large F-statistic value and a low p-value indicate that there are differences among the petrol formulations. This suggests strong evidence against the null hypothesis. However, a large F-statistic value is not common and may suggest issues with the data or analysis. It's possible that outliers (Boxplot) in the data are increasing the F-statistic value.

In conclusion, the results show that at least one group of petrol formulation's mean is significantly different from the others. Therefore, it can be concluded that there are significant differences among the petrol formulations. However, it's important to consider conducting physical tests or experiments to complement the statistical analysis.

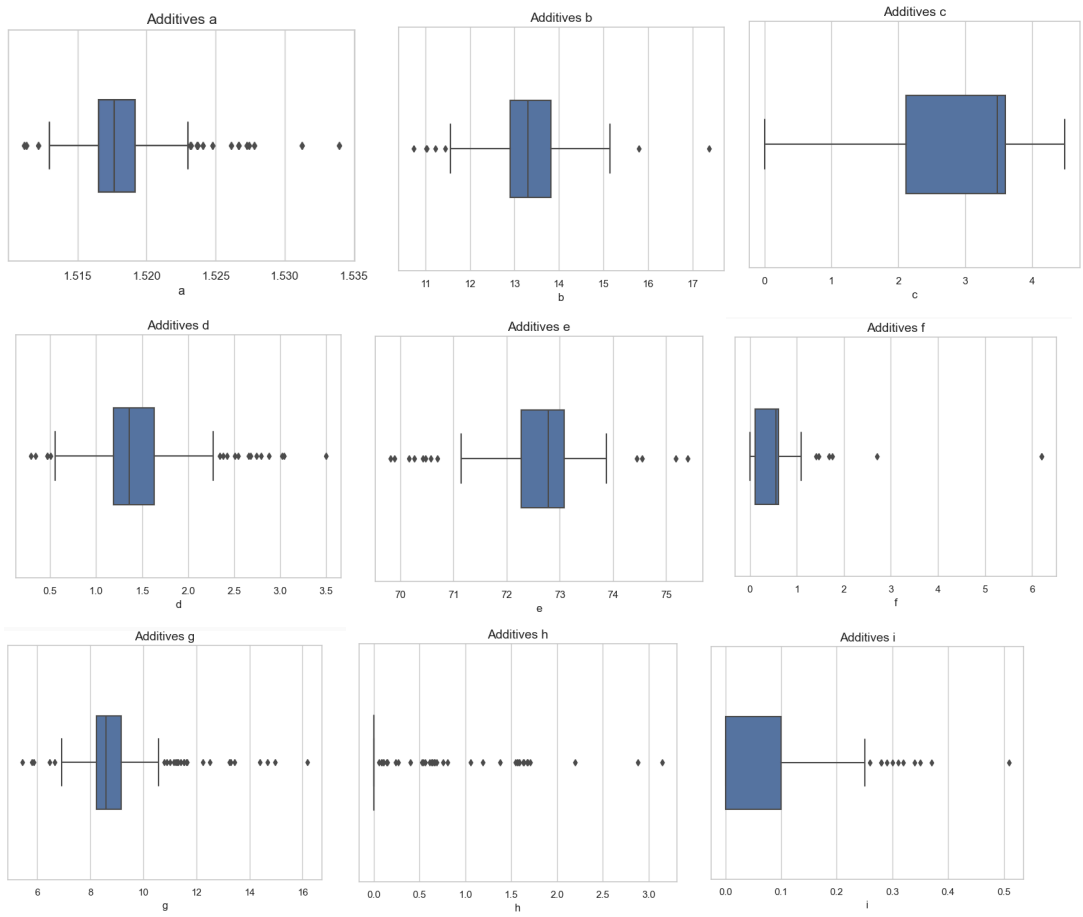
Histograms



According to the histogram, the additives 'a', 'b', 'd', 'f', 'g', 'h', and 'i' have a positive skewness, indicating that these additives have more occurrences of lower values and fewer occurrences of higher values. This skewed distribution means that these additives tend to have higher values, which may suggest that some formulations are more effective at preventing engine knocking, gum formation or ensuring stability in storage at lower concentrations. In other words, positive skewness may indicate that certain additives work better at lower concentrations.

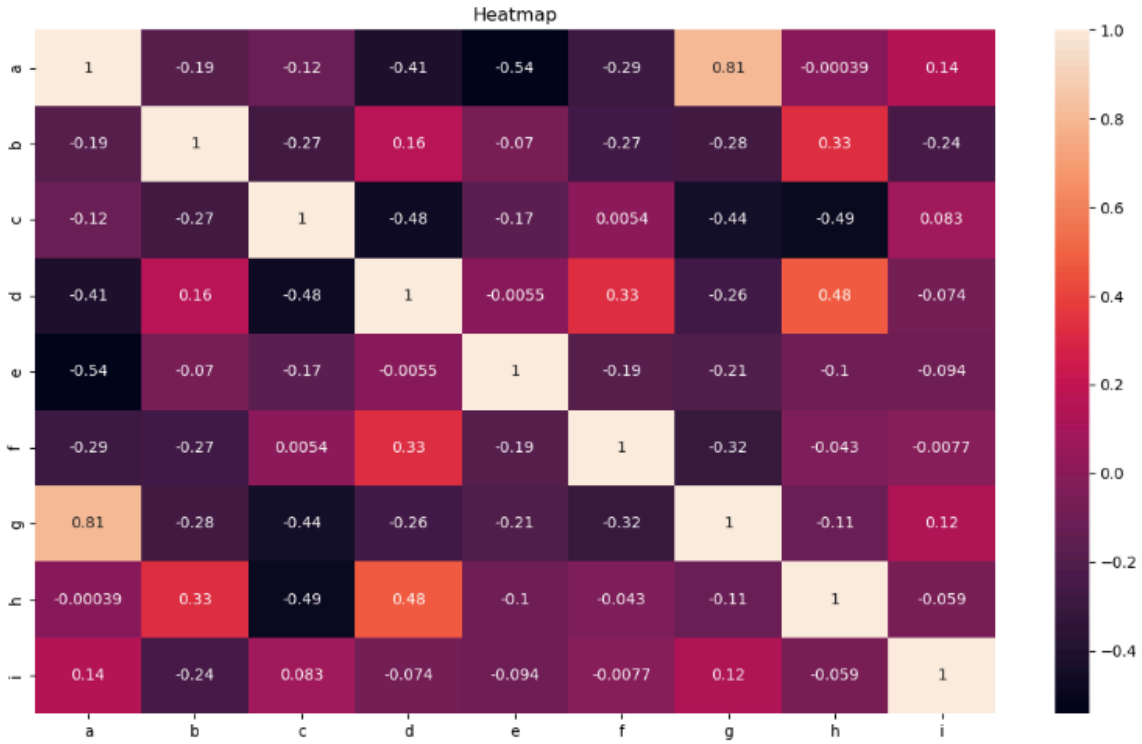
On the other hand, a negative skewness for additive 'c' and 'e' suggests that these additives have more occurrences of higher values and fewer occurrences of lower values, resulting in a distribution that is skewed towards lower values. A negative skewness may indicate formulations with higher additive concentrations are needed to achieve the desired effects.

Boxplot



Based on the boxplot, we can further clarify the presence of outliers. We can see that there are more than one additives have outliers, specifically additives 'a', 'd', 'e', 'g' and ‘h’ have a relatively more outliers. The identification of outliers suggests that there are potential discrepancies in the data associated with these additives and variations in additive concentrations that may give impact on the consistency of petrol formulations.

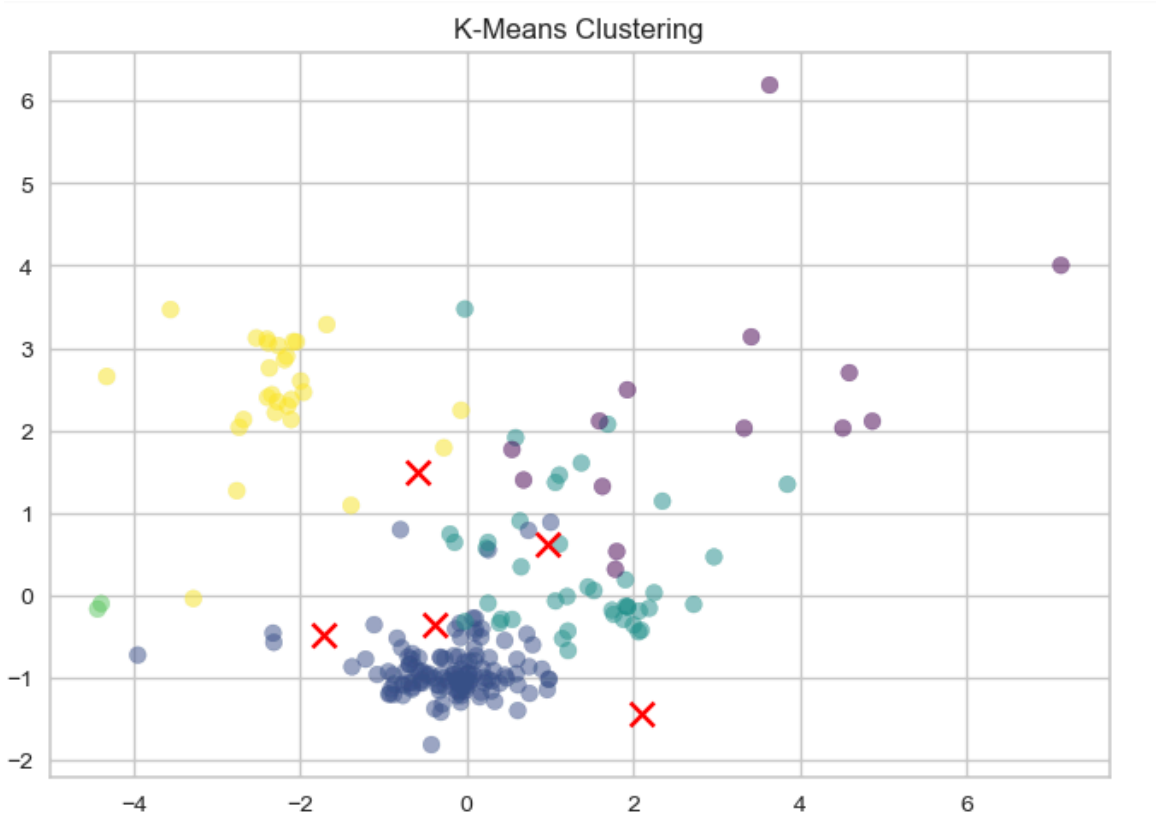
Heatmap



Based on the heatmap, we can see that the lighter color indicates it has stronger positive correlation whereas the darker color indicates it has stronger negative correlation.

- Additive 'a' tends to have a stronger negative relationship with ‘e’ for correlation of -0.54 which can be represented with the black color cell.
- Additive 'b' has a dim purple color that shows a weak negative relationship of -0.27 and -0.28 with additives 'c' and 'g' respectively where there are slightly change in color in an opposite way.
- Both additive ‘c’ and ‘g’ exhibit a moderate negative correlation of -0.44 which indicated by the black colors.
- Additive 'd' shows orange color in heatmap which indicate there is moderate positive relationship with additive 'h' of coefficient 0.48 which both of them show changes in the color in the same direction.

K-Means Clustering



Upon observing the scatter plot and the distribution of clusters, it appears that the chosen number of clusters of 5 effectively captures the variations and patterns in the data to some extent. Each cluster exhibits unique characteristics as evidenced by the varying densities and patterns within clusters. However, there are a few aspects to consider:

- There are instances where clusters appear to overlap, especially light green and purple color. This suggest that the separation between clusters may not be optimal.
- The centroid is positioned relatively close to the blue clusters at bottom of the scatter plot which suggests that a closer relationship or similarity in their data points.
- The potential of the presence of outliers within the yellow cluster at the left side and purple cluster at right side of the scatter plot. These anomalies might signify noise in the data where the K-Means algorithm's partitioning could be refined.

In summary, the scatter plot has provided valuable insights that reveal distinct groupings of data points into the formulations of petrol by looking at the pattern and distribution around centroids.