FINAL VISUALIZATION REPORT

Name = Muhammad Usman Javed Student ID = 23070968

Title = Fruit Quality Data Analysis (Orange)

Dataset Overview:

About 24 distinct orange varieties, including Cara, Temple, Star Ruby, Navel, and others are included in the study. A range of colours, from deep orange to light orange, are displayed by these kinds. A variety of parameters, including weight, size, brix, pH, softness, harvest time, and maturity, were taken into consideration in order to assess and forecast the quality of oranges. Kaggle is where the orange quality dataset was found. The dataset is shown in Table 1 and was analysed using libraries like pandas, matplotlib, seaborn, and numpy.

	Size (cm)	Weight (g)	Brix (Sweetness)	pH (Acidity)	Softness (1-5)	HarvestTime (days)	Ripeness (1-5)	Color	Variety	Blemishes (Y/N)	Quality (1-5)
0	7.5	180	12.0	3.2	2.0	10	4.0	Orange	Valencia	N	4.0
1	8.2	220	10.5	3.4	3.0	14	4.5	Deep Orange	Navel	N	4.5
2	6.8	150	14.0	3.0	1.0	7	5.0	Light Orange	Cara Cara	N	5.0
3	9.0	250	8.5	3.8	4.0	21	3.5	Orange-Red	Blood Orange	N	3.5
4	8.5	210	11.5	3.3	2.5	12	5.0	Orange	Hamlin	Y (Minor)	4.5
...	...	...	...	...	...	...	...	...	...	...	...
236	8.0	194	10.9	3.6	5.0	13	1.0	Orange-Red	Tangerine	Y (Scars)	5.0
237	7.4	275	8.5	3.5	5.0	20	5.0	Light Orange	Minneola (Hybrid)	N	4.0
238	7.5	196	15.7	3.0	3.0	13	3.0	Deep Orange	Temple	Y (Minor Insect Damage)	5.0
239	7.2	251	9.8	4.3	3.0	23	1.0	Light Orange	Moro (Blood)	Y (Minor Insect Damage)	3.0
240	7.3	282	10.5	2.9	4.0	21	2.0	Orange-Red	Jaffa	Y (Minor)	4.0

Table-1: Quality data set about Oranges

Statistical Study: Significant statistical values are presented in Table 2.

	Size (cm)	Weight (g)	Brix (Sweetness)	pH (Acidity)	Softness (1-5)	HarvestTime (days)	Ripeness (1-5)	Quality (1-5)
count	241.000000	241.000000	241.000000	241.000000	241.000000	241.000000	241.000000	241.000000
mean	7.844813	205.128631	10.907884	3.473900	3.072614	15.344398	3.599585	3.817427
std	1.086002	56.461012	2.760446	0.421007	1.323630	5.323852	1.205214	1.014410
min	6.000000	100.000000	5.500000	2.800000	1.000000	4.000000	1.000000	1.000000
25%	6.900000	155.000000	8.500000	3.200000	2.000000	11.000000	3.000000	3.000000
50%	7.800000	205.000000	11.000000	3.400000	3.000000	15.000000	4.000000	4.000000
75%	8.700000	252.000000	13.400000	3.800000	4.000000	20.000000	4.500000	4.500000
max	10.000000	300.000000	16.000000	4.400000	5.000000	25.000000	5.000000	5.000000

Table-2: Statistical Data

Histogram plots:

Histogram plots provide rapid and efficient insights into data distribution, variability, and patterns, histograms are essential for exploratory data analysis. They provide guidance for making decisions, point out important trends, and provide as a springboard for more complex analysis methods.As shown in Table 2 and Figure 1 plots, the distribution of orange data depending on different parameters and their corresponding ranges is shown by the histogram plots with in a distinct frame.

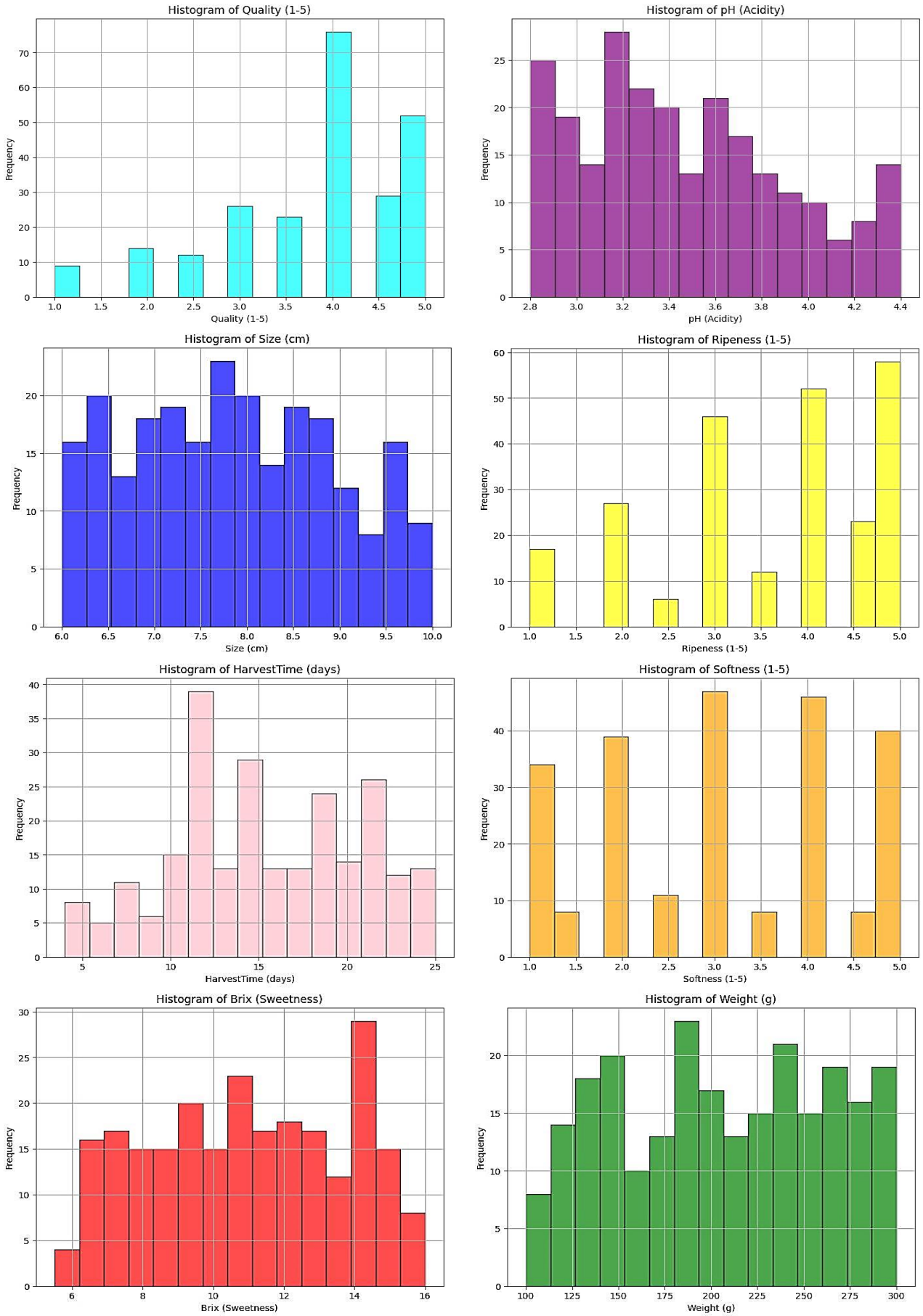

Figure 1: Histogram

Pie Chart:

Pie charts work best in situations where there are few categories and notable distinctions between them. Figure 2 displays a pie chart that illustrates the weighting of five distinct orange colours in the data**.**

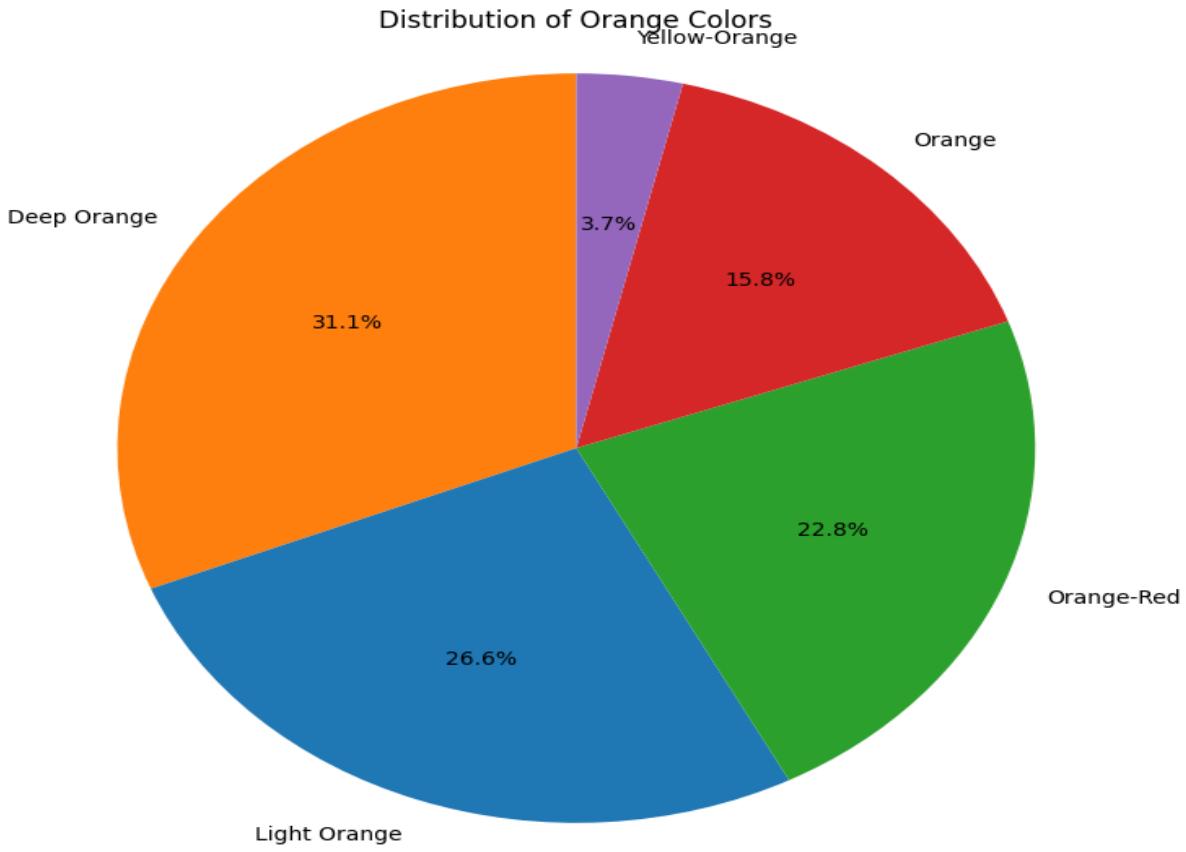

Figure 2: Pie-Chart

Kernel Density Estimation (KDE) Plot:

A crucial non-parametric technique in data analysis for determining the probability density function (PDF) of a continuous variable is Kernel Density Estimation (KDE).Higher-quality oranges have a higher degree of sweetness, which is more densely concentrated within a smaller range, as shown by the graph in Figure 3. The clear distinction between density peaks across different quality grades suggests that orange quality assessment is capable of accurately identifying differences in sweetness levels.

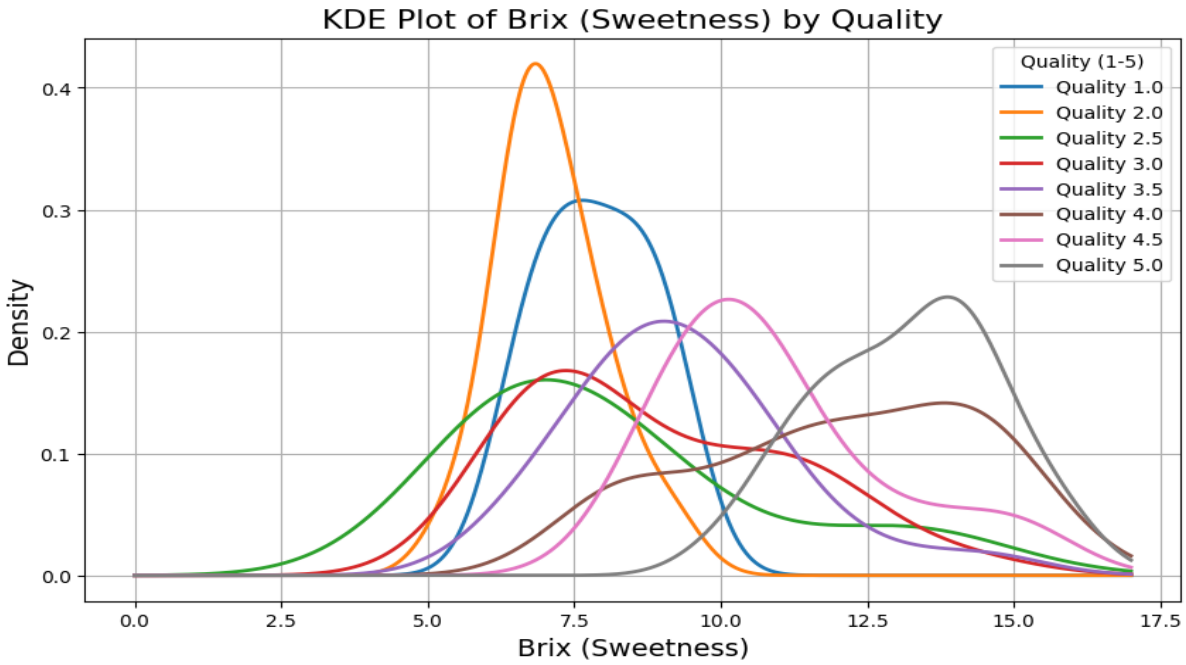

Figure 3: KDE plot

Correlation Heat Map Plot:

Using colour gradients to depict data values, a heat map is an effective data visualisation tool for data analysis. Its significance stems from its capacity to communicate intricate data linkages in an intuitive and efficient manner. The correlation matrix reveals some noteworthy conclusions and revelations, including:

1.With correlation coefficients of 0.305 and 0.330, respectively, there is a weakly positive relationship between their size, weight, and pH. This implies that larger oranges are typically heavier and more acidic. The size of oranges and their Brix levels, however, have a weakly negative connection of 0.301. According to these results, larger oranges often contain slightly higher levels of an acidity and relatively lower levels of sweetness.

2. Harvest Time (days), softness, and quality rating have a moderately negative association with pH (acidity). Accordingly, oranges with higher freshness, lower acidity (higher pH), and less softness are often rated as being of higher grade.

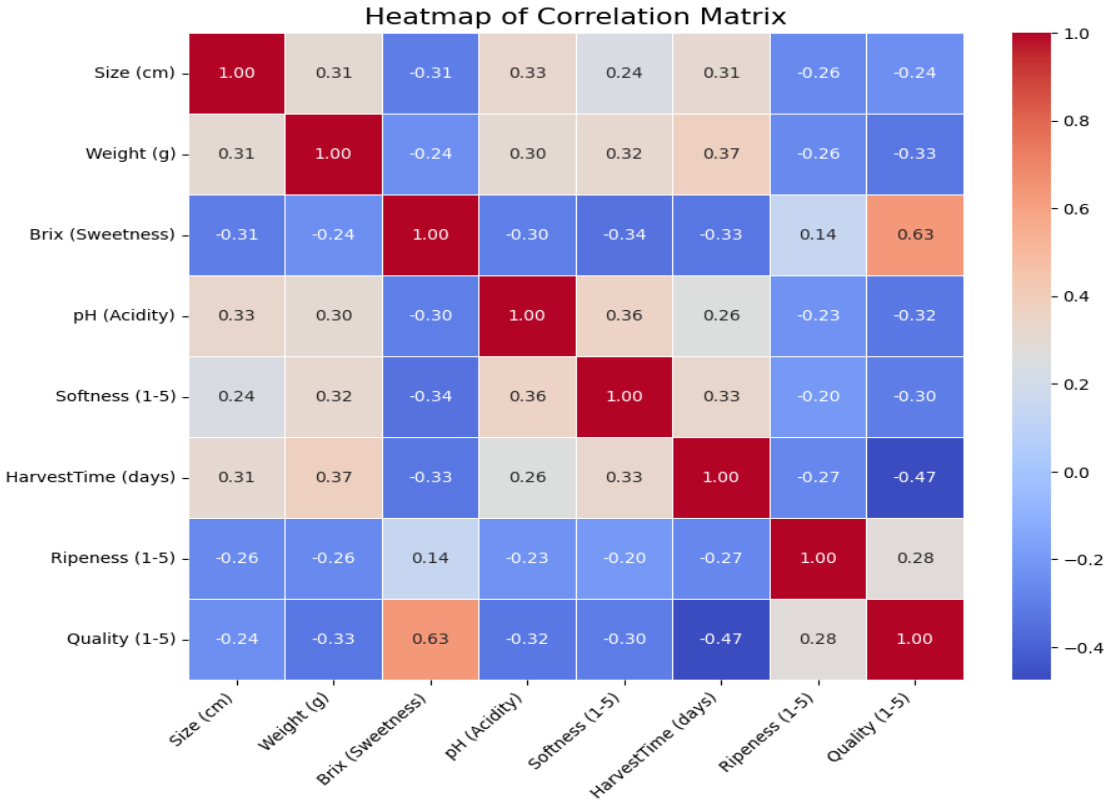

Figure 4: Correlation Heatmap

Blemishes (Y/N)	Brix (Sweetness)		pH (Acidity)		Softness (1-5)		Ripeness (1-5)		Quality (1-5)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
N	11.02	11.00	3.38	3.30	2.99	3.00	3.91	4.00	3.99	4.00
Y (Minor)	11.36	11.45	3.57	3.45	2.18	2.00	3.36	3.00	3.46	4.00
Y (Mold Spot)	9.93	9.20	3.91	3.89	4.00	4.25	3.15	2.75	3.20	3.25
Y (Scars)	10.44	10.00	3.59	3.60	3.38	3.50	3.23	3.00	3.74	3.50
Y (Sunburn Patch)	10.07	9.30	3.51	3.40	3.09	3.00	2.91	3.00	3.52	4.00

Table 3 : Blemish's effect on quality in relation to other factors.

Box Plot:

Some significant findings have been drawn from the information in Table 3 and the illustrations in Figure 5, which explain how flaws affect quality along with other factors.

1. Compared to other blemish types, oranges with 'Y (Minor)' flaws are generally sweeter, firmer, and less mature.
2. 'Y (Mould Spot)' flaws can significantly affect the quality of oranges by making them less sweet, low acidic, riper, and generally softer.
3. Blemish-free oranges ('N') are considered to be of superior quality and may have a slightly greater acidity.

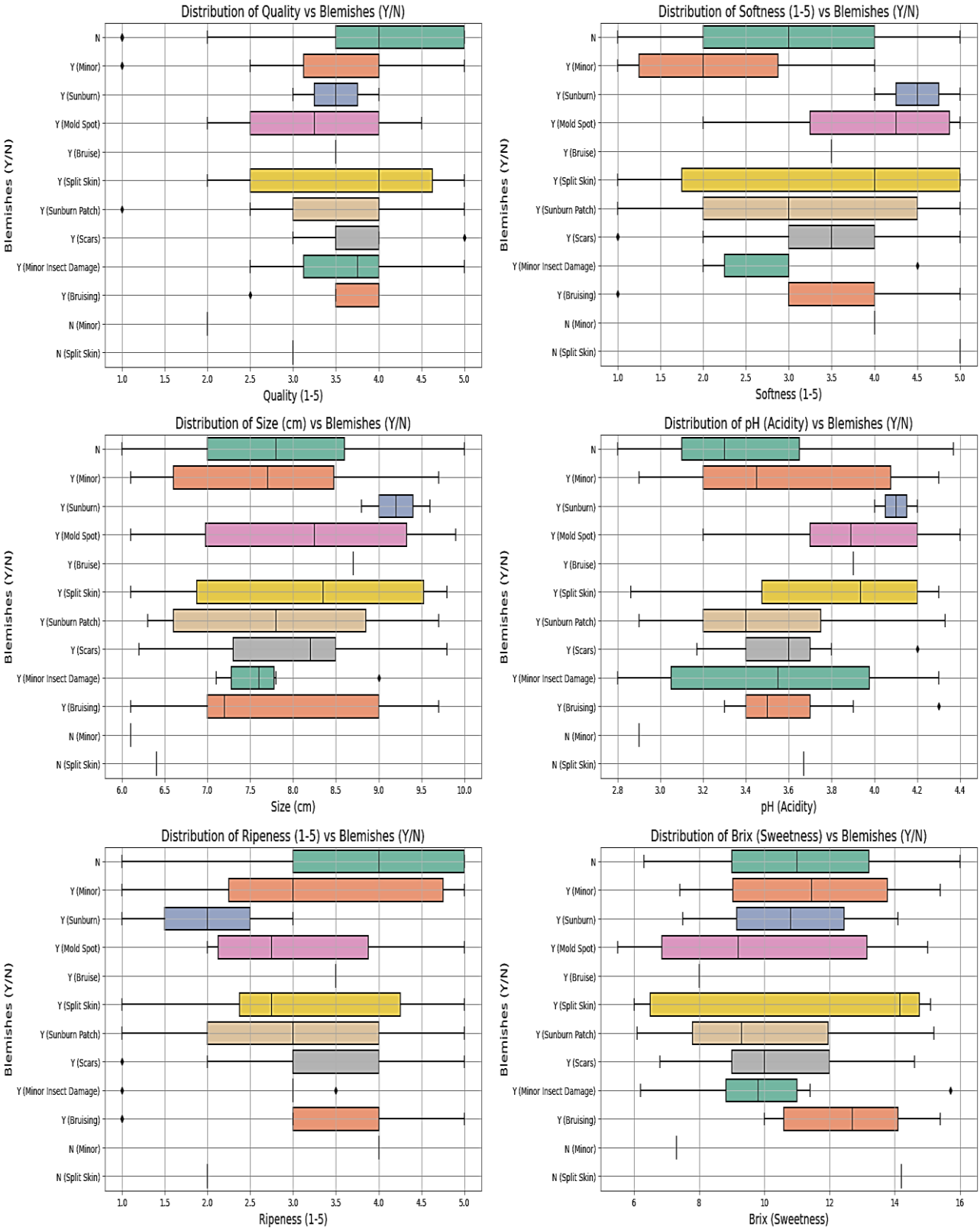

Figure 5: Box plots

Violin Plot:

According to the violin plot, oranges with a higher sweetness level are linked to a higher quality ranking.

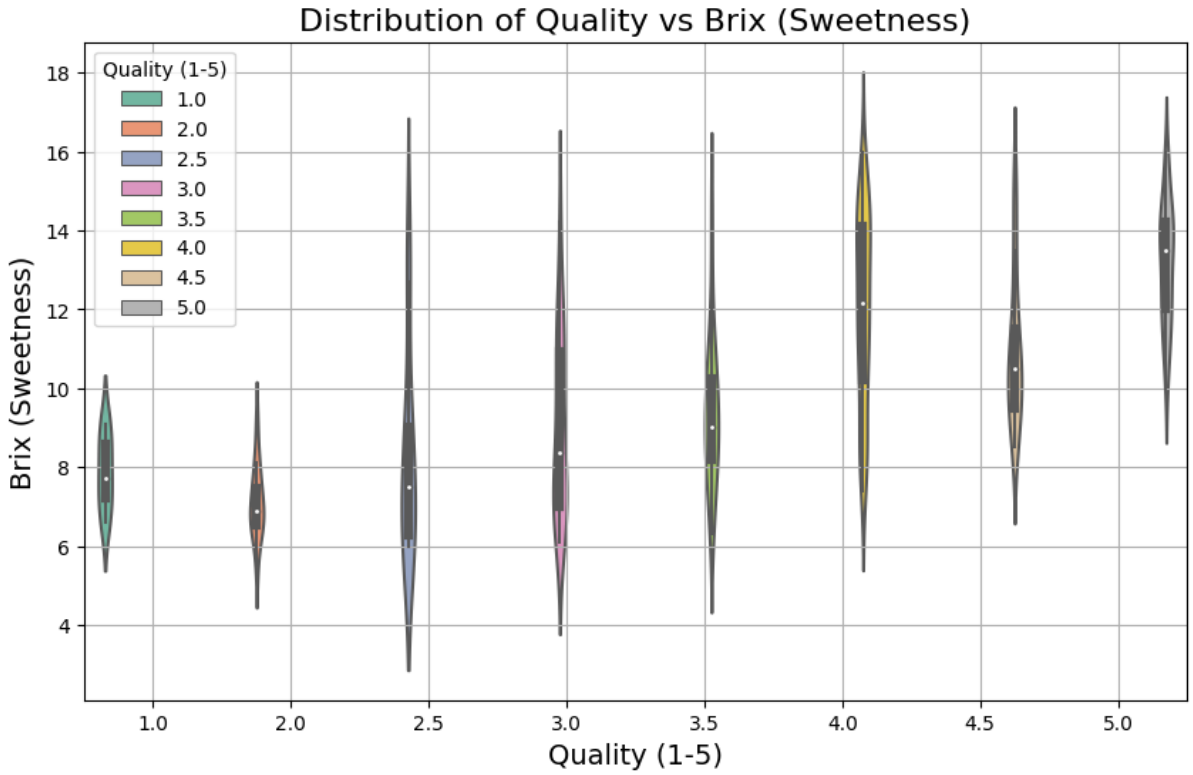

Figure 6: Violin Plot

Conclusion:

The data analysis report yielded the following result, which is provided below.

Higher quality ratings are assigned to oranges with no flaws, greater sweetness, lower acidity, greater freshness, and less softness. In general, larger oranges have slightly higher levels of acidity and relatively less sweetness.