Q.1) Assuming normal distribution, mean $\mu = 42$ months, standard deviation $(\sigma) = 8$ months.

we have to find $P(20 < x < 30)$.

Normal distribution, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Normalizing the variables.

$z = \frac{(x-\mu)}{\sigma}$ $\therefore z_1 = (20-42)/8 = -2.75$

$z_2 = (30-42)/8 = -12/8 = -1.5$

$\therefore P(20 < x < 30) = P(-2.75 < z < -1.5)$

$= P(z < -1.5) - P(z < -2.75)$

from normal distribution table,

$P(z < -1.5) = 0.0668.$

$P(z < -2.75) = 0.0030.$

$\therefore P(20 < x < 30) = 0.0668 - 0.0030$

$= 0.0638$

Q.2> since, we have a 100+ data, so the ~~med~~ mean cannot be calculated, but as ~~the d~~ here median will be a useful measure, as data is readily available points.

Here, there are 8 data points therefore ~~mean =~~ median = Mean of 4th and 5th data points = $(63+75)/2 = 69$.

Again here 8 different parts of an electronoric component is given, so, the lowest value 36 will also be a useful measure as failure of a single part will may lead ~~measure~~ ~~in~~ to failure of whole electronic component.

Q.3) First formulation time data points = 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91

ble.

~~So~~ Here $n = 8$.

sample Mean = $\frac{\sum \text{times}}{n} = \mu$.

$$= \frac{(1.75 + 1.92 + 2.62 + 2.35 + 3.09 + 3.15 + 2.53 + 1.91)}{8}$$

$$= 19.32 / 8 = 2.415.$$

sample Variable = $\frac{\sum (x_i - \bar{x})^2}{(n-1)} = s^2$.

= ~~10~~.

$$\approx (1.9976 / 7)$$

$$\approx 0.285371$$

∴ sample standard deviation = $\sqrt{Var(x)} = s$

$$= \sqrt{0.285371} \doteq 0.5342$$

For the first data set,

min = 1.75

$Q_1 = (n+1)/4 = 2.25 \rightarrow$ between 2nd and 3rd.

$$= (1.91 + 1.92)/2 = 1.915.$$

Q_2 = median = between 4th and 5th.

$$= (2.35 + 2.53)/2 = 2.44.$$

similarly $Q_3 = 2.9725$.

Max = 3.15.

similarly for 2nd dataset.

Min = 1.83

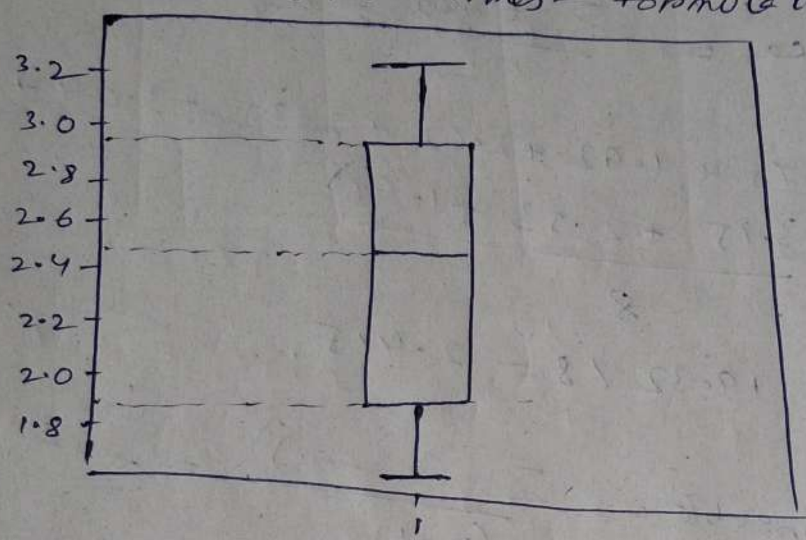
Q_1 = 1.965

Median = 2.76

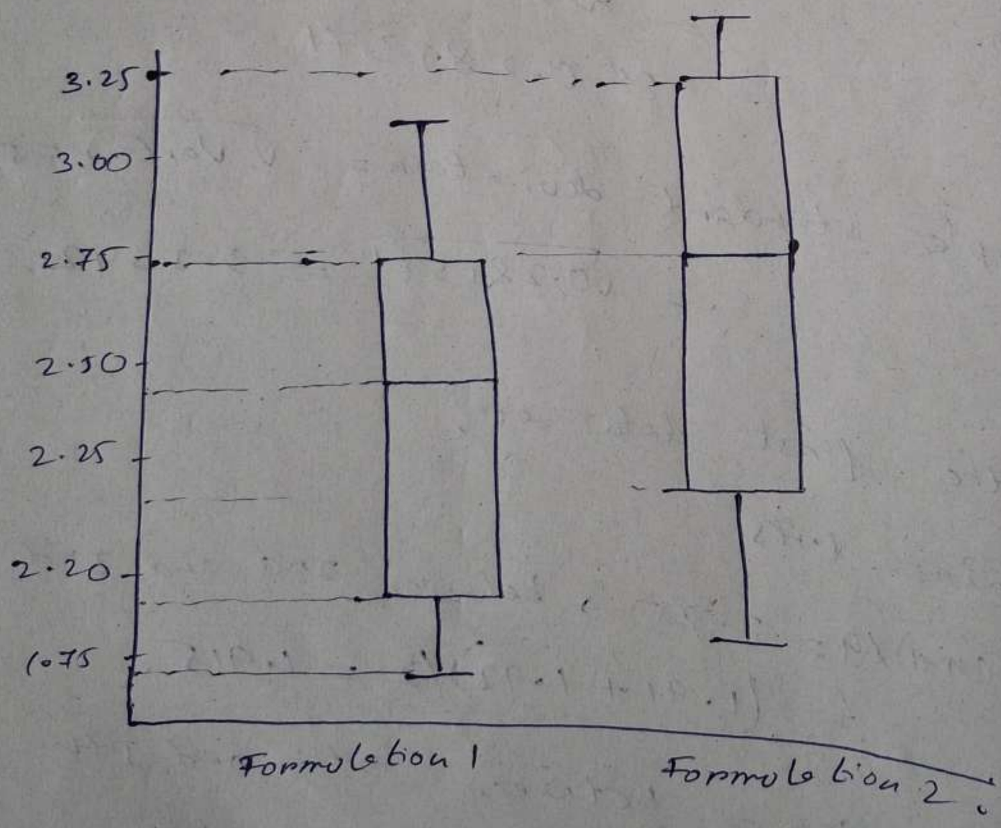
Q_3 = 3.305

Max = 3.40

Cold start Times- Formulation-1.



Comparative Cold start Times.



→ Neither box plot points plotted beyond the whiskers, indicating no outliers by 1.5 × IQR rule.

Conclusion

more con-

with a

Formulatio

and gr

If quic

1 appea

Conclusion :- Formulation 1 offers faster and more consistent cold start ignition times, with a lower median and tighter spread. Formulation 2, exhibits slower average and greater variability.

If quick, reliable is priority, Formulation 1 appears superior.

4) For first ~~two~~ entries let's analize

a) Min-max weight normalization

$$\text{formula} = \frac{\text{Weight}(w_i) - \text{Min}(min_i)}{\text{Max}(max) - \text{Min}.}$$

$$= \frac{50-41}{136-41} = 0.094737$$

b) P. lee has weight 50kg which is less than 60kg. Therefore his/her weight falls in low category.

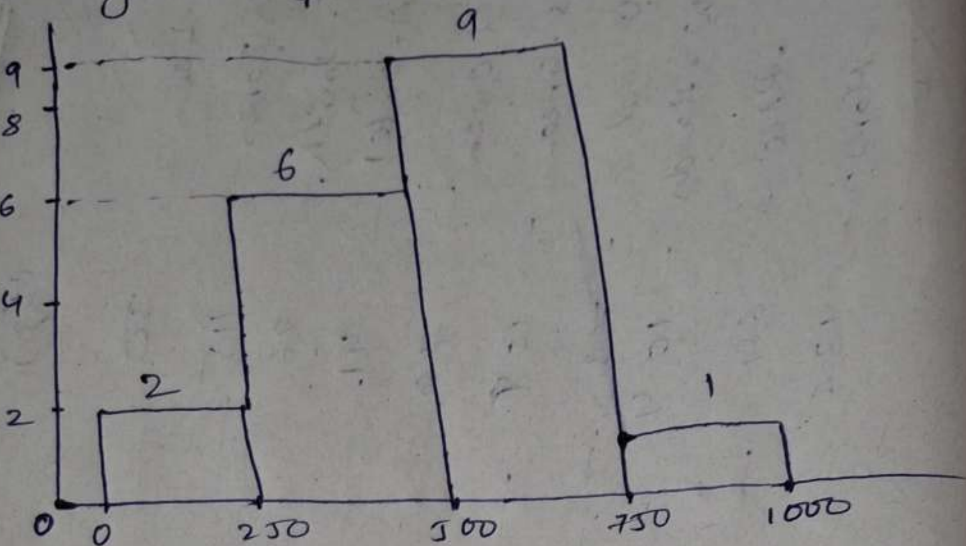
c) $$BMI = \frac{\text{weight}(Kg)}{(\text{Height}(m))^2} = \frac{50\,kg}{(1.52)^2\,m^2}$$

$$= \cancel{2.16}\ 21.641274\ kg/m^2$$

iv) Table with all the newly added columns.

Name	Weight (Kg)	Height (m)	BMI (Kg/m²)	Normalized Weight	Weight Category	Blood Pressure Systolic/Diastolic	Diabetes
1. P. Lee	50	1.52	21.641274	0.094737	Low	68/112	0
2. R. Jones	115	1.77	36.707204	0.778947	High ~~Medium~~	110/152	1
3. J. Smith	96	1.83	28.666129	0.578947	Medium	88/132	0
4. A. Patel	41	1.55	17.065557	0.0	Low	76/125	0
5. M Owen	79	1.82	23.849777	0.40	Medium	65/105	0
6. S. Green	109	1.89	30.514263	0.715789	High	114/159	0
7. N Cook	73	1.76	23.566632	0.336842	Medium	108/136	[illegible]
8. W. Hands	104	1.71	35.566499	0.663158	High	101/132	+
9. P. Rice	64	1.74	21.138856	0.242105	Medium	101/132	0
10. F. Marsh	136	1.78	42.923873	1.0	High	121/165	1

5. a) Histogram of Sale Price.



b) Contingency Table: Store vs. Product Category

Store	Desktop	Laptop	Printer	Scanner
New York, NY	3	1	2	4
Washington, DC	2	2	2	2

c) i) Summary Table - Grouped by Customer.

Customer	Transaction Count	Total Sale Price ($)
B. March	3	1700
E. Sims	1	700
G. Hinton	4	2150
H. Fu	1	450
H. Taylor	1	400
J. Bain	1	500
L. Nye	2	900
P. Judd	2	900
S. Cann	1	600
T. Goss	2	750

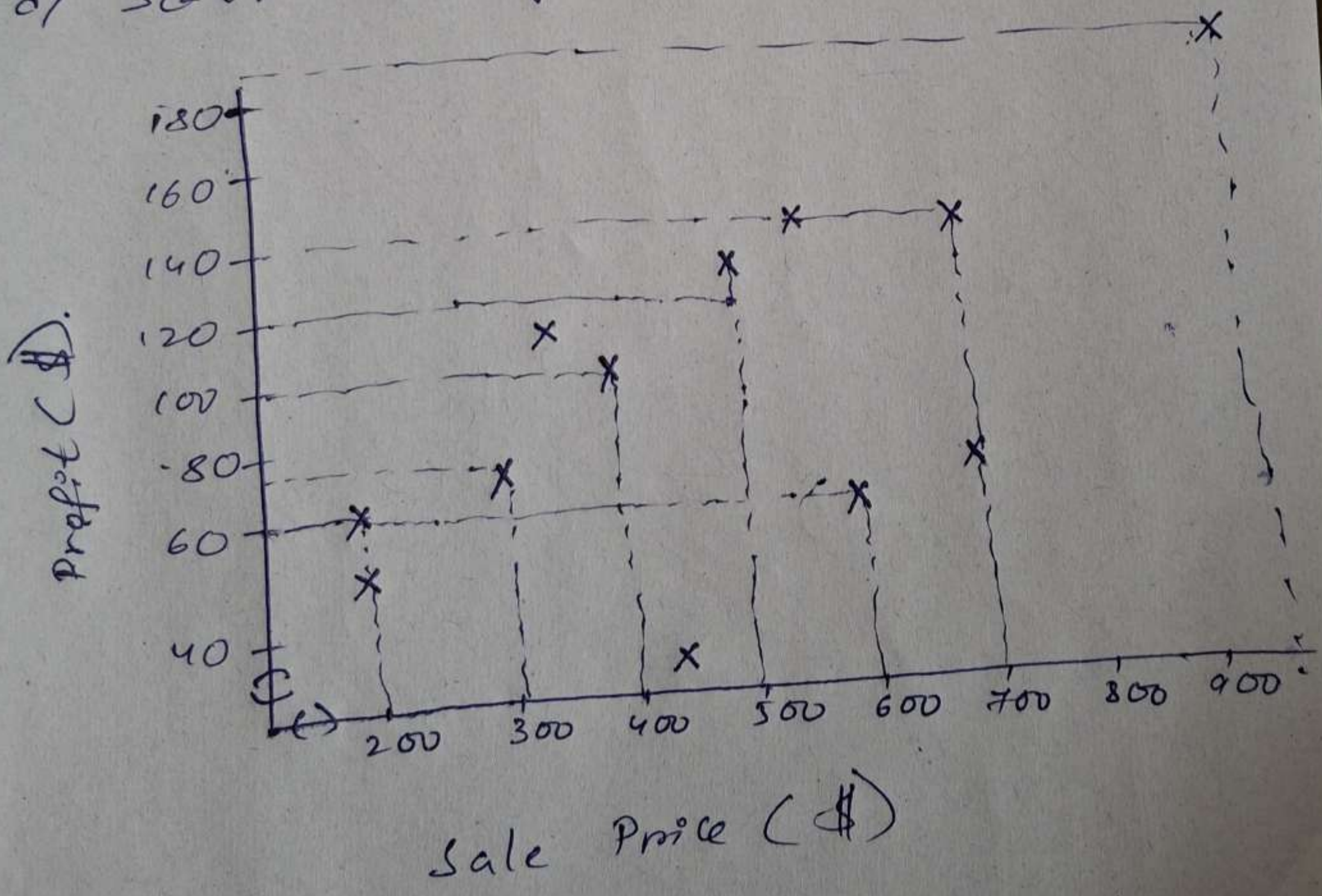
ii) Summary Table ii – Grouped by store.

Store	Transaction Count	Mean Sale Price ($)
New York, NY.	10	485.0
Washington DC.	8	525.0

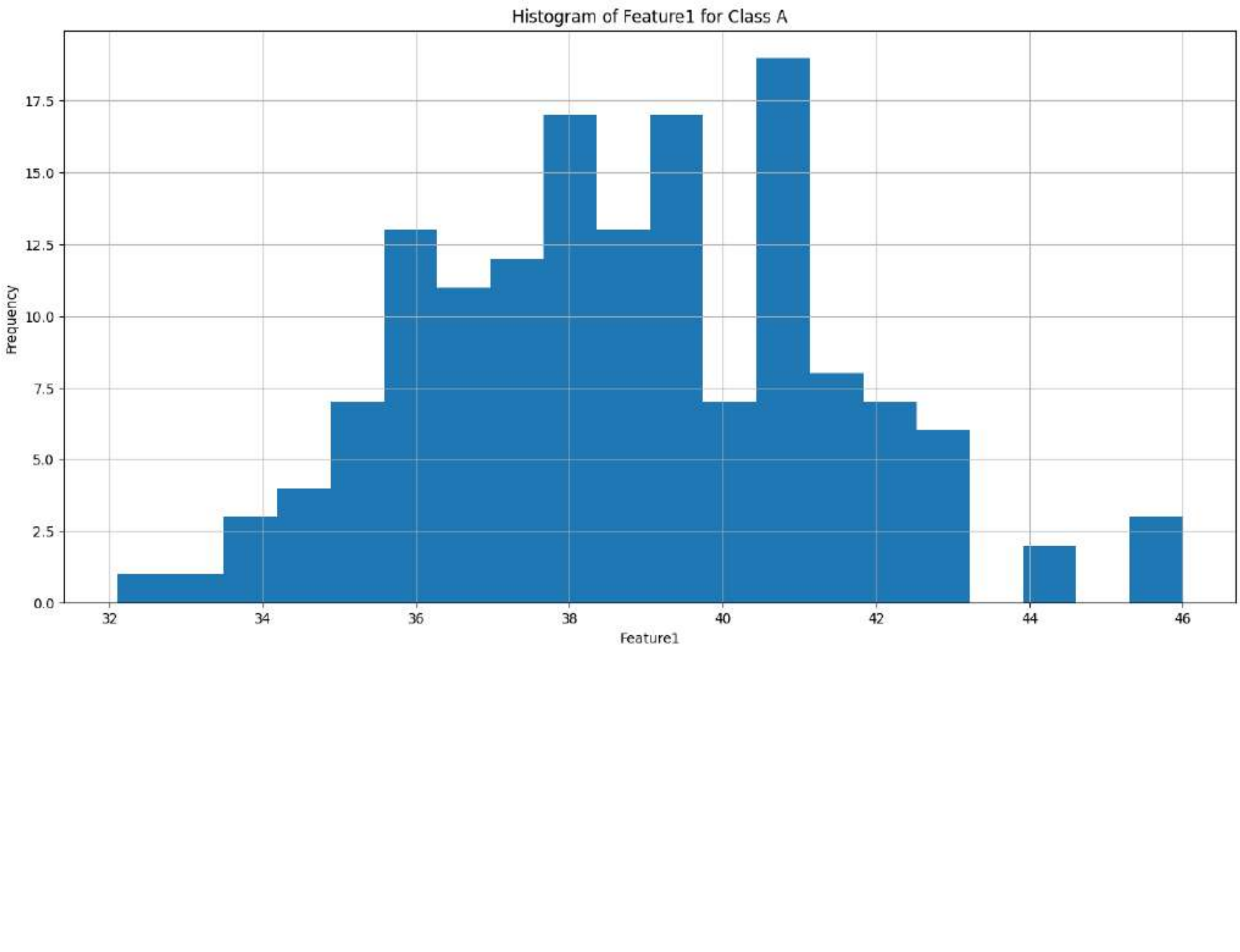
iii) Grouped by Product Category.

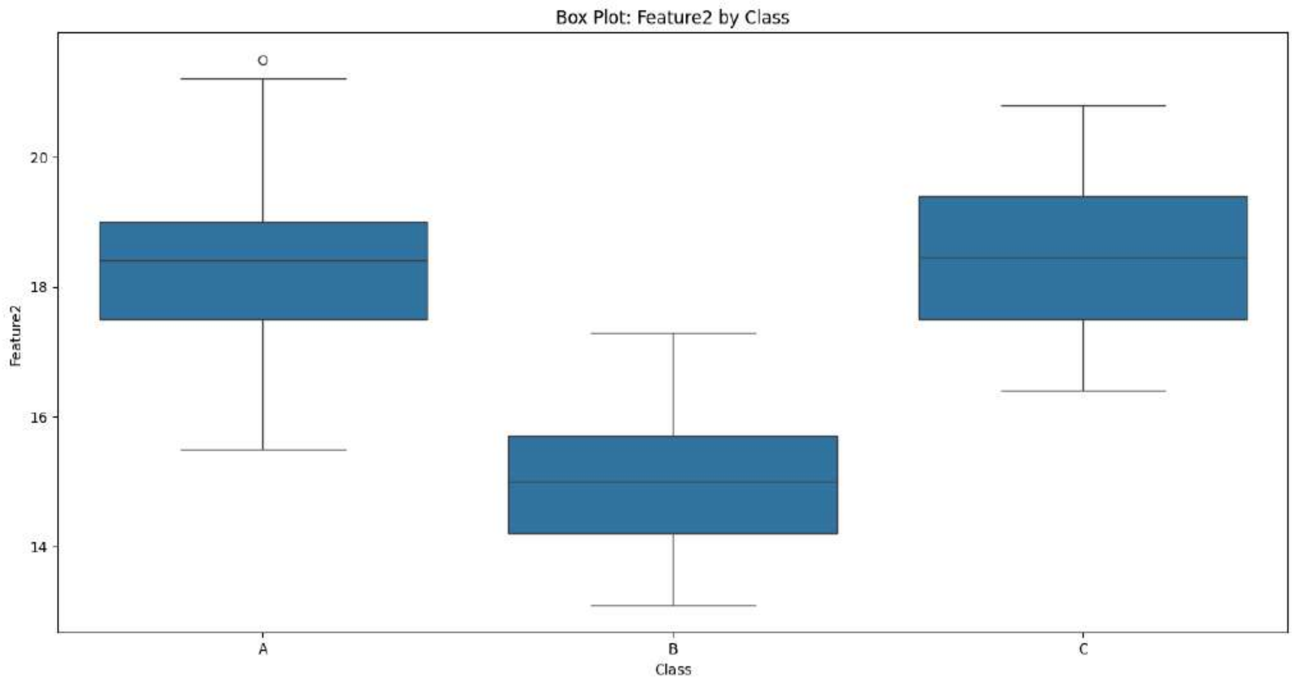
Product Category	Transaction Count	Total Profit ($)
Desktop.	5	295
Laptop.	3	470
Printers.	4	360
Scanners.	6	640.

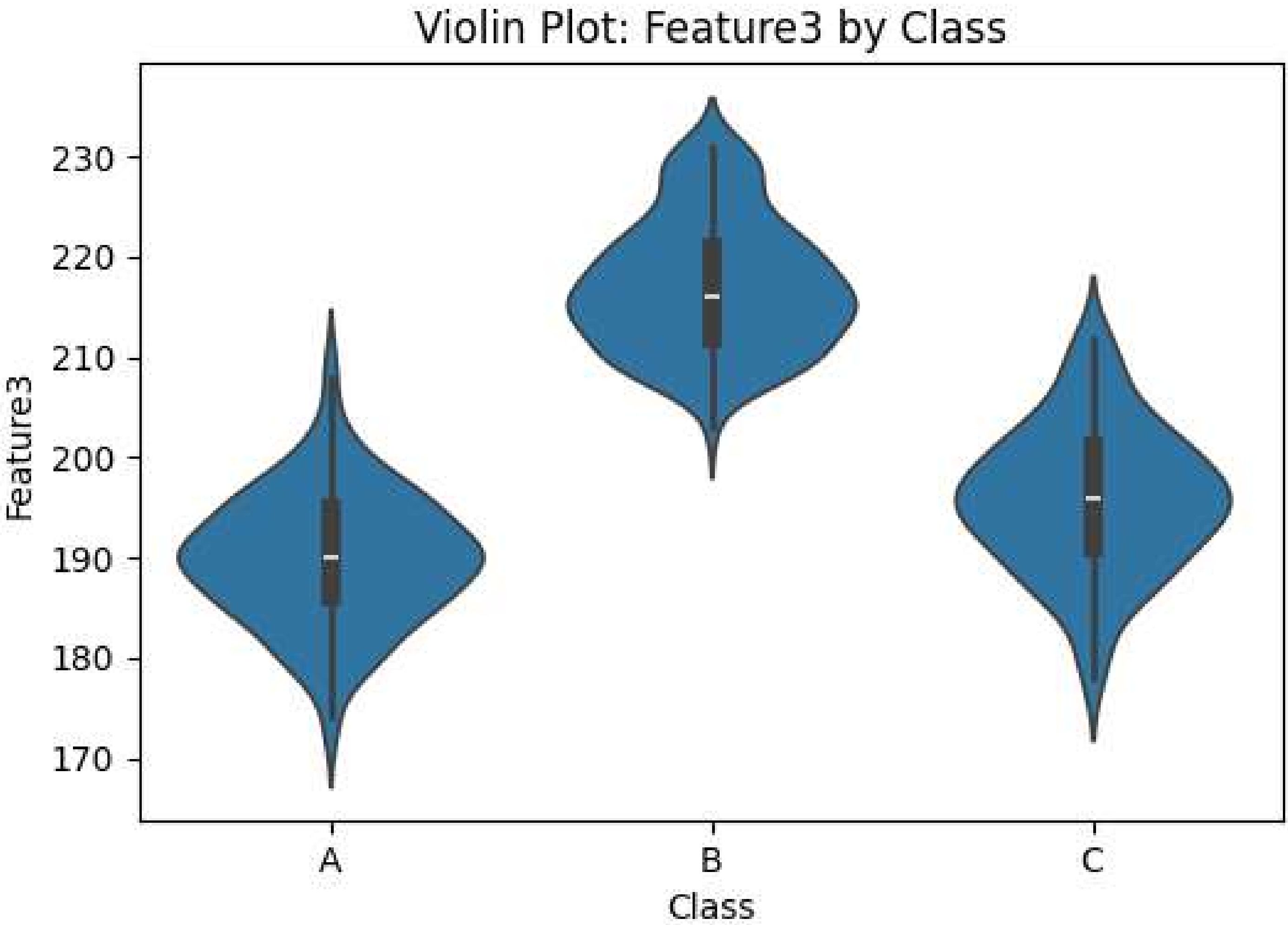
d) Scatter Plot of Sale Price vs Profit.

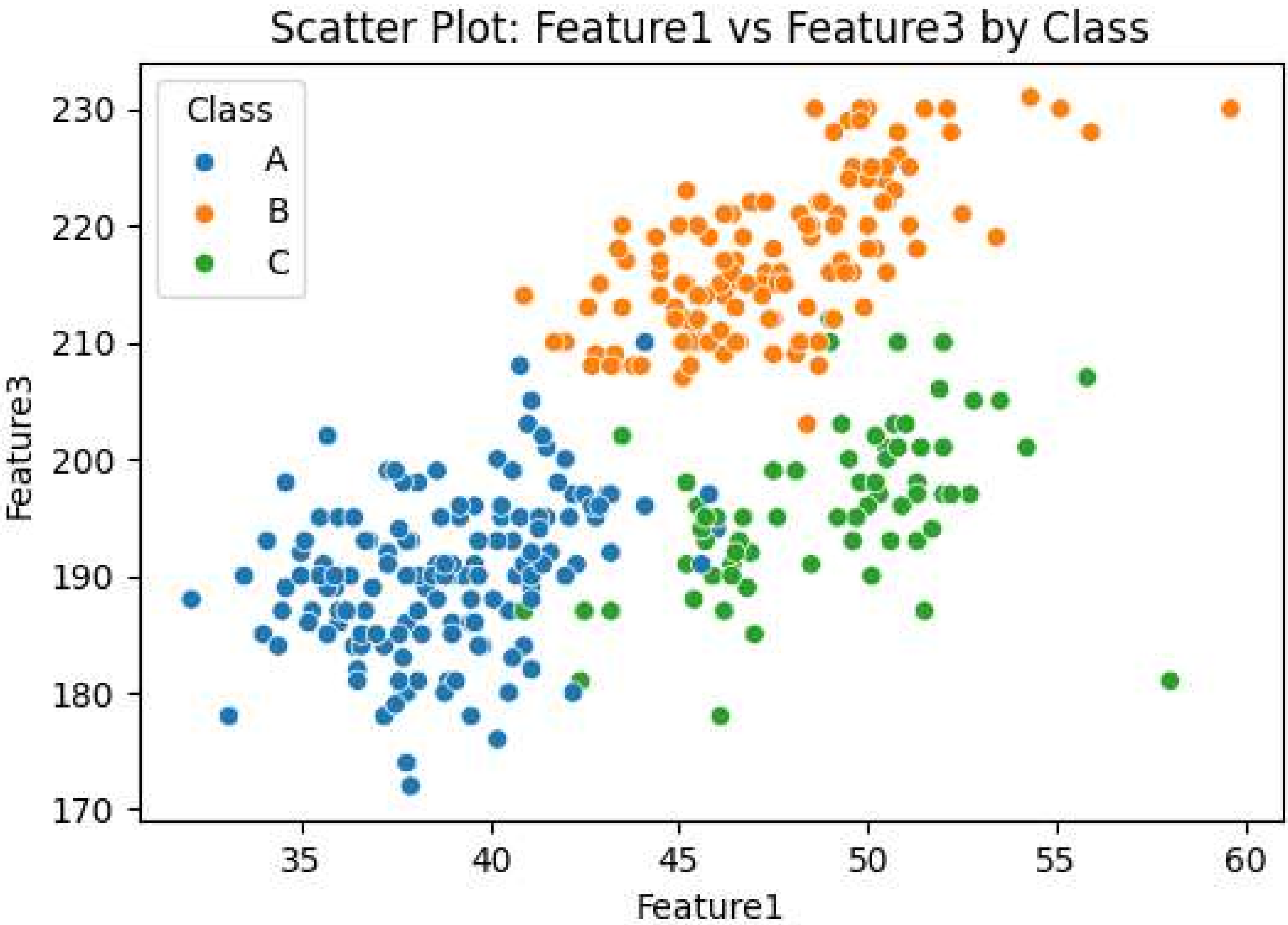


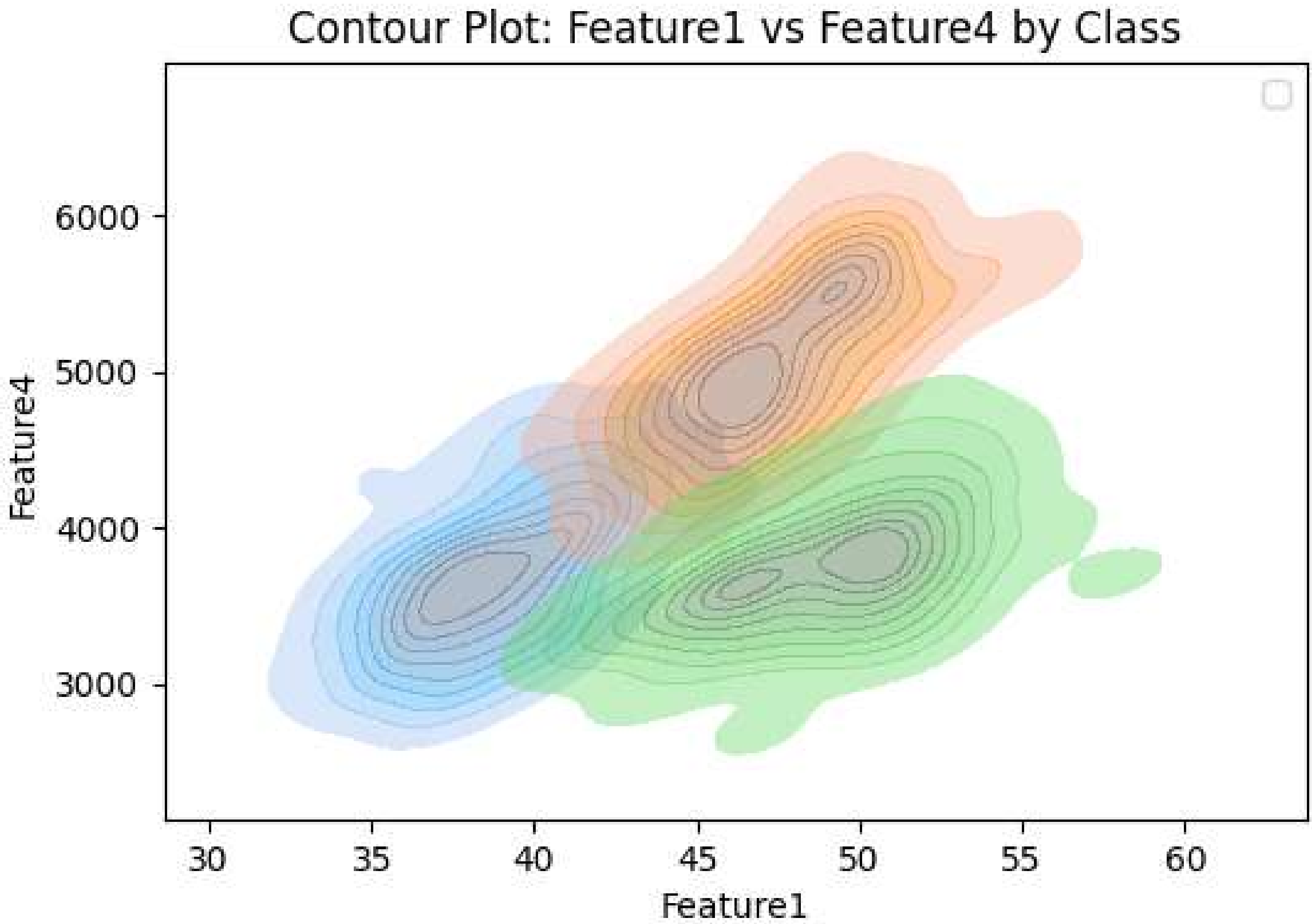
```
Frequency of samples for each class:
Class
A    151
B    123
C     68
Name: count, dtype: int64

Data Description:
          Feature1    Feature2    Feature3     Feature4
count  342.000000  342.000000  342.000000   342.000000
mean    43.921930   17.151170  200.915205  4201.754386
std      5.459584    1.974793   14.061714   801.954536
min     32.100000   13.100000  172.000000  2700.000000
25%     39.225000   15.600000  190.000000  3550.000000
50%     44.450000   17.300000  197.000000  4050.000000
75%     48.500000   18.700000  213.000000  4750.000000
max     59.600000   21.500000  231.000000  6300.000000

Interquartile Range (IQR):
 Feature1       9.275
Feature2       3.100
Feature3      23.000
Feature4    1200.000
dtype: float64
```

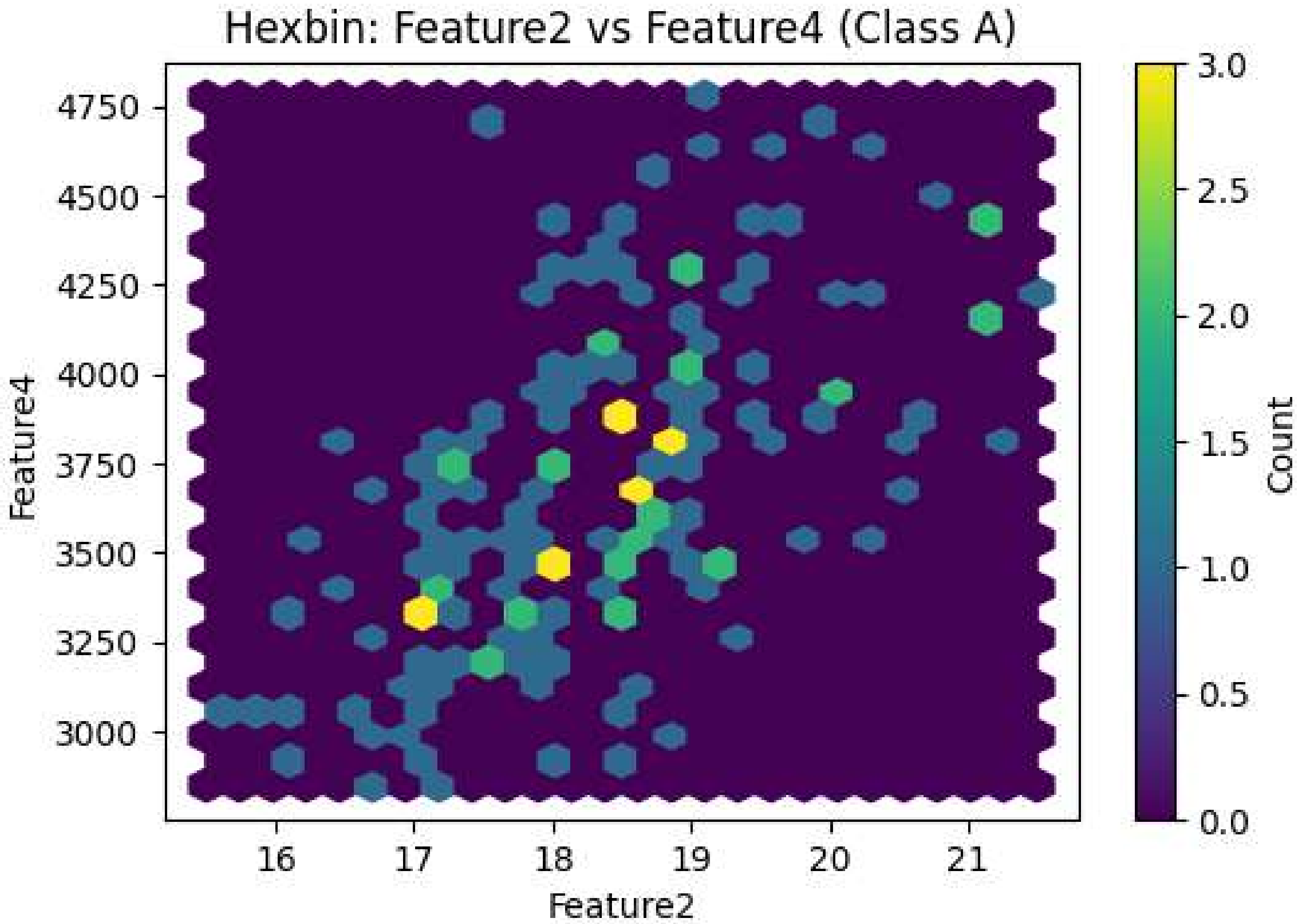

Histogram of Feature1 for Class A
Frequency
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
32
34
36
38
40
42
44
46
Feature1


Box Plot: Feature2 by Class
20
18
16
14
Feature2
A
B
C
Class

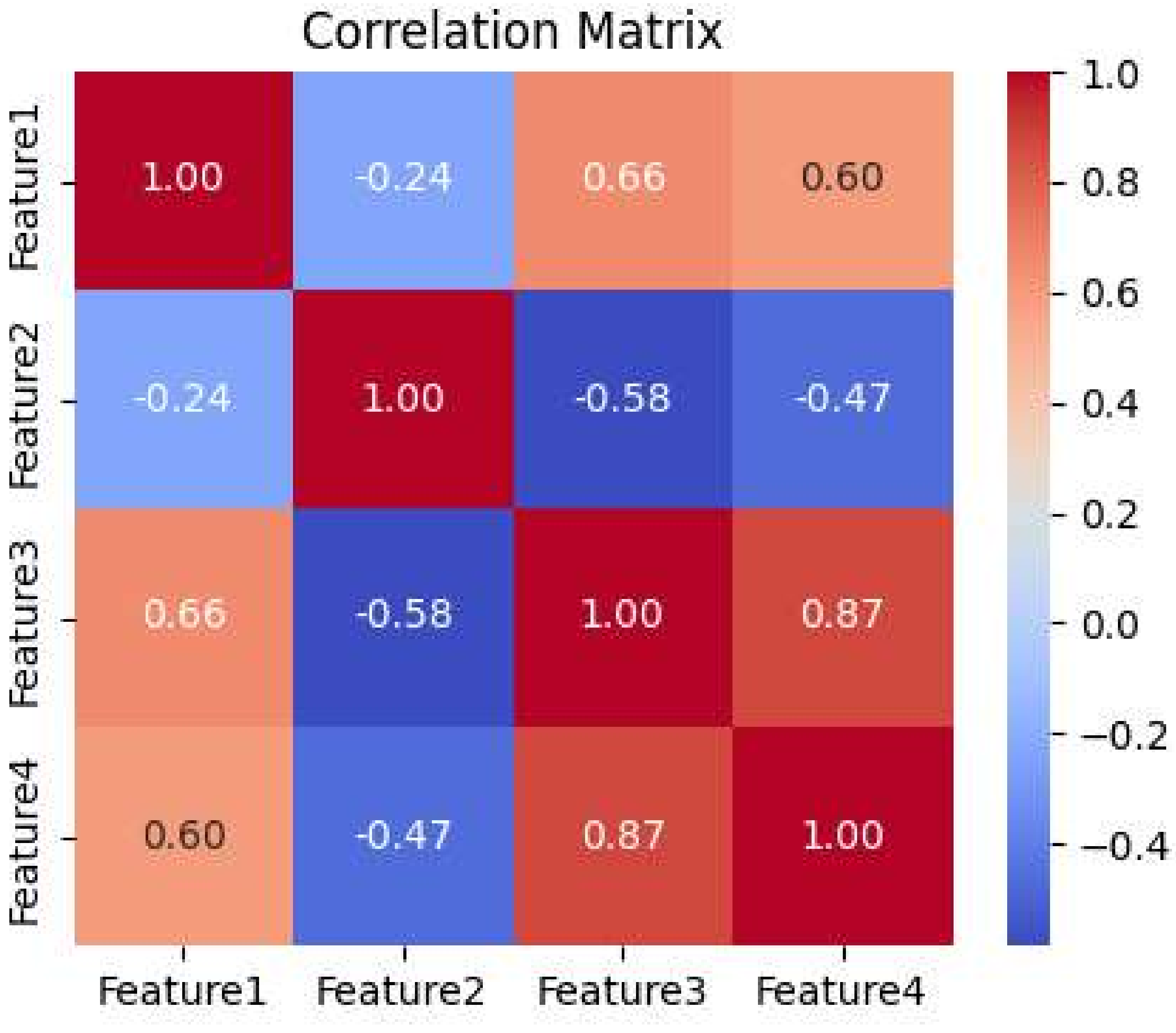

Violin Plot: Feature3 by Class
230
220
210
200
190
180
170
Feature3
A
B
C
Class


Scatter Plot: Feature1 vs Feature3 by Class
Class
A
B
C
Feature3
Feature1
170
180
190
200
210
220
230
35
40
45
50
55
60

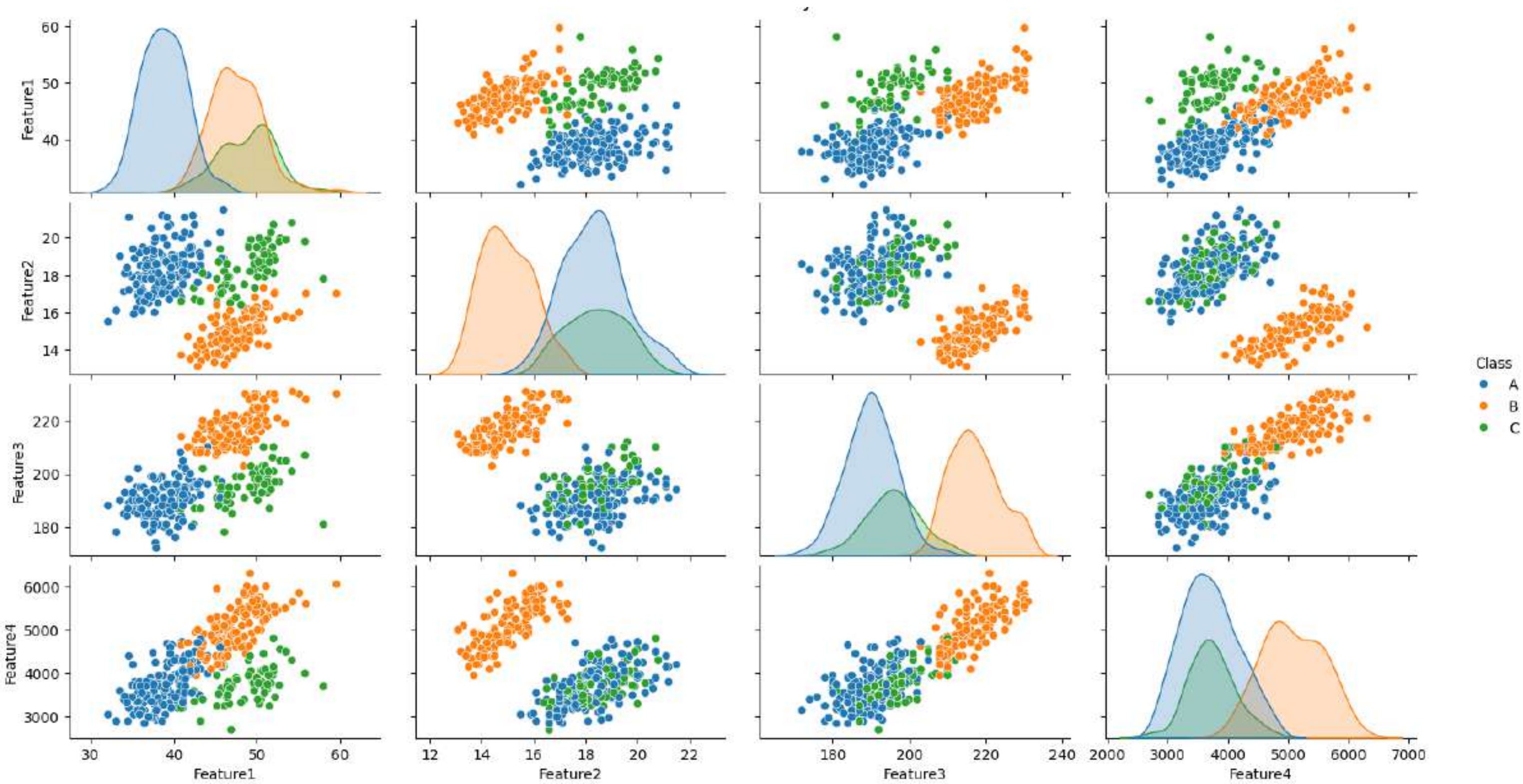

Contour Plot: Feature1 vs Feature4 by Class
Feature4
6000
5000
4000
3000
30
35
40
45
50
55
60
Feature1



Hexbin: Feature2 vs Feature4 (Class A)
4750
4500
4250
4000
3750
3500
3250
3000
Feature4
16
17
18
19
20
21
Feature2
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count



Correlation Matrix
Feature1
Feature2
Feature3
Feature4
1.00
-0.24
0.66
0.60
-0.24
1.00
-0.58
-0.47
0.66
-0.58
1.00
0.87
0.60
-0.47
0.87
1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Feature1
Feature2
Feature3
Feature4
Class
A
B
C