EE708 Assignment 1- Solutions

1. Probability a Z-Phone lasts between 20 and 30 months

Let $X \sim N(\mu = 42, \sigma = 8)$. We want:
$P(20 < X < 30)$
$= P((20 - 42) / 8 < Z < (30 - 42) / 8)$
$= P(-2.75 < Z < -1.50)$

Using the standard normal CDF:
$\Phi(-1.50) \approx 0.0668$
$\Phi(-2.75) \approx 0.0030$

Therefore:
$P(-2.75 < Z < -1.50) = 0.0668 - 0.0030 = 0.0638 \approx 6.38\%$

2. Choosing a Meaningful Measure of Location for Censored Data

We have failure times for eight electronic components, one of which is right-censored (100+).

therefore the actual failure time for this unit is unknown, the mean is not a suitable measure.

Instead, we use the median, a more robust measure of central tendency.

Sorted failure times: 36, 45, 51, 63, 75, 80, 90, 100 (with 100 being censored)

Simple Median (ignoring censoring): (63 + 75)/2 = 69.0 hours

Kaplan-Meier Estimated Median Survival Time: 75.0 hours

```
Data with censoring info:
    Time  Observed
0    75          1
1    63          1
2   100          0
3    36          1
4    51          1
5    45          1
6    80          1
7    90          1

Simple Median (including censored value): 69.0

Kaplan-Meier estimated median survival time: 75.0
```

3. Cold Start Ignition Times Analysis

First gasoline: [1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91]
Second gasoline: [1.83, 1.99, 3.13, 3.29, 2.65, 2.87, 3.40, 2.46, 1.89, 3.35]
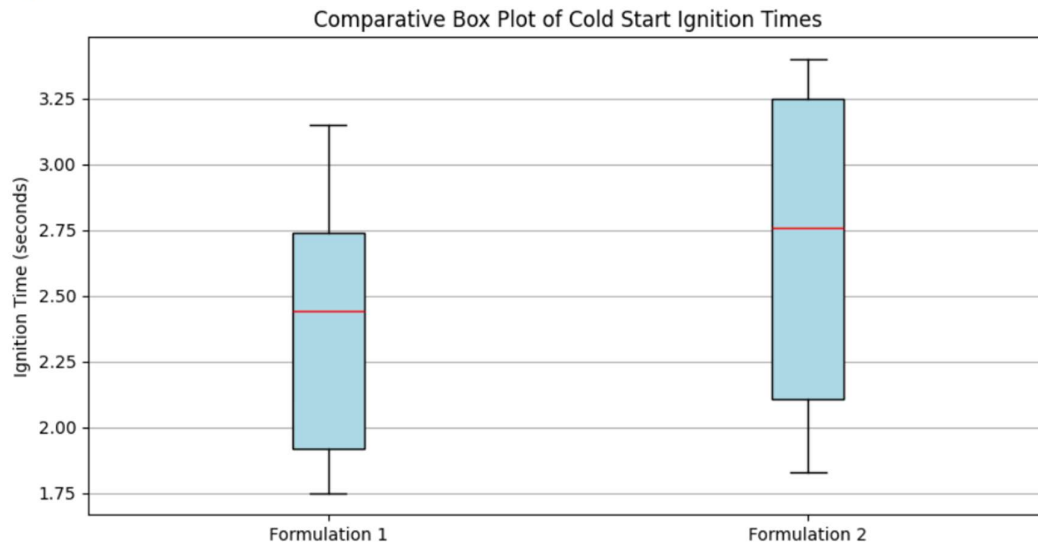
Sample statistics for first formulation:
Mean = 2.42 seconds
Sample Variance = 0.2799
Sample Std Deviation = 0.529

Comparative box plots were generated to visualize both formulations.

```
Mean: 2.42 seconds
Sample Variance: 0.2854
Sample Standard Deviation: 0.5342
```


Comparative Box Plot of Cold Start Ignition Times

4. Patient Data Preprocessing

a) Min-Max Normalization of Weight (kg):
Each value is transformed to the [0, 1] range using the formula:
normalized_value = (value − min) / (max − min)

b) Weight Binning into categories:
- Low: < 60 kg
- Medium: 60–100 kg
- High: > 100 kg

c) Body Mass Index (BMI) is calculated using:
BMI = Weight (kg) / (Height (m))^2

```
       Name  Weight (kg)  Weight (Normalized) Weight Category  Height (m)  \
0    P. Lee           50             0.094737             Low        1.52
1  R. Jones          115             0.778947            High        1.77
2  J. Smith           96             0.578947          Medium        1.83
3  A. Patel           41             0.000000             Low        1.55
4   M. Owen           79             0.400000          Medium        1.82
5  S. Green          109             0.715789            High        1.89
6   N. Cook           73             0.336842          Medium        1.76
7  W. Hands          104             0.663158            High        1.71
8   P. Rice           64             0.242105          Medium        1.74
9  F. Marsh          136             1.000000            High        1.78

         BMI
0  21.641274
1  36.707204
2  28.666129
3  17.065557
4  23.849777
5  30.514263
6  23.566632
7  35.566499
8  21.138856
9  42.923873
```

5. Retail Transaction Data Analysis

Histogram of Sale Price
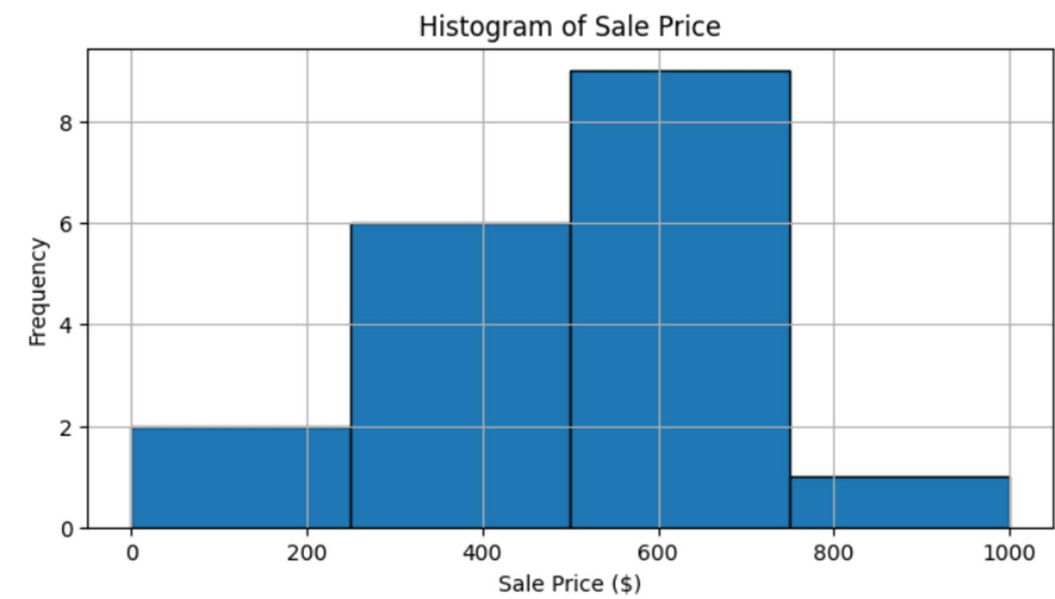
```
Contingency Table:
 Product Category  Desktop  Laptop  Printer  Scanner
Store
New York, NY             3       1        2        4
Washington, DC           2       2        2        2

Grouped by Customer:
             count   sum
Customer
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




Grouped by Store:
                count    mean
Store
New York, NY       10   485.0
Washington, DC      8   525.0

Grouped by Product Category:
                  count   sum
Product Category
Desktop               5   295
Laptop                3   470
Printer               4   360
Scanner               6   640
```
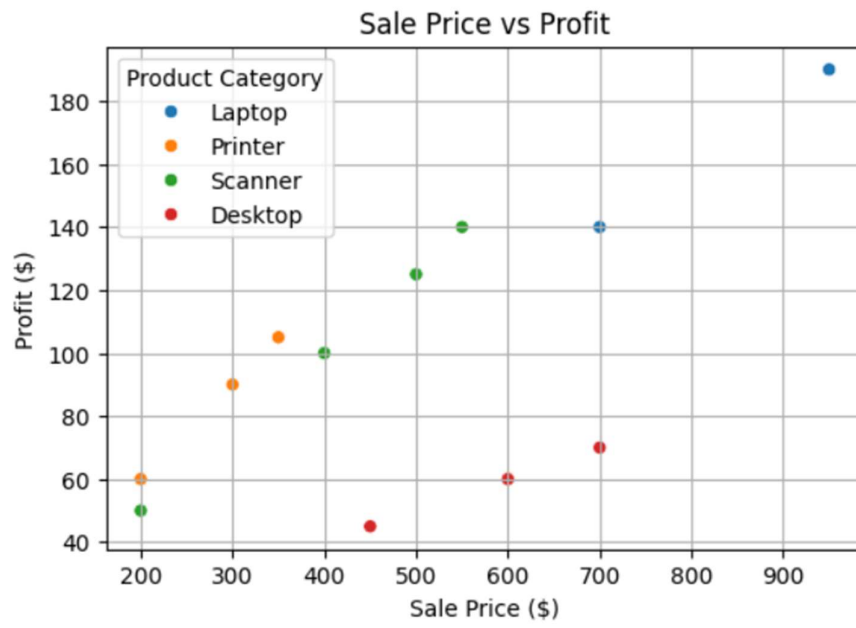
Sale Price vs Profit

6. Exploratory Data Analysis (EDA) using A1.csv
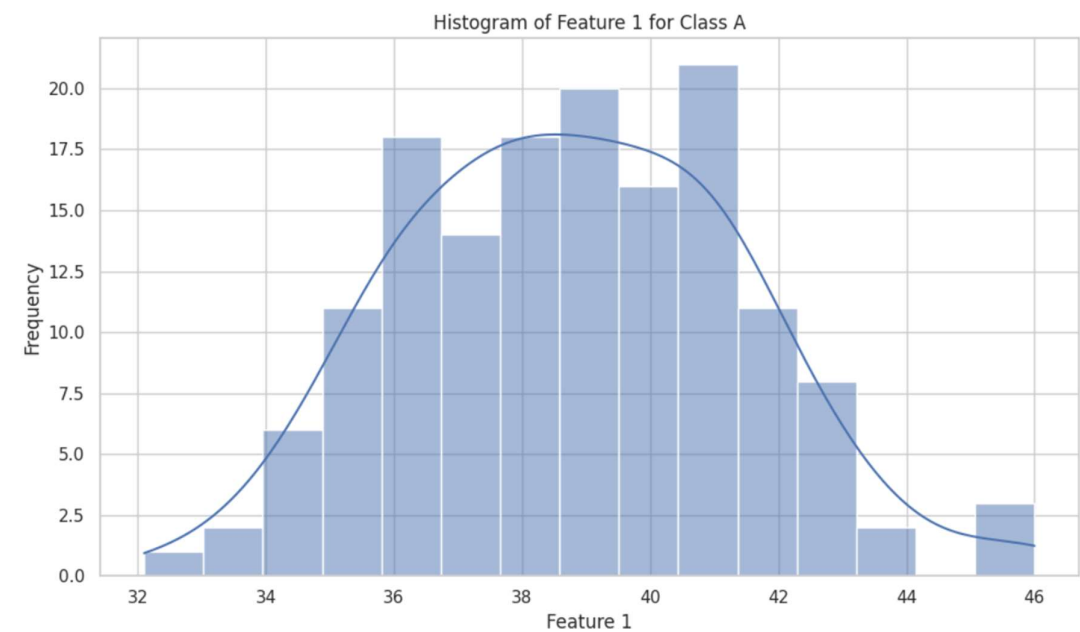
6(a)

```
Frequency of samples per class:
 Classes
A    151
B    123
C     68
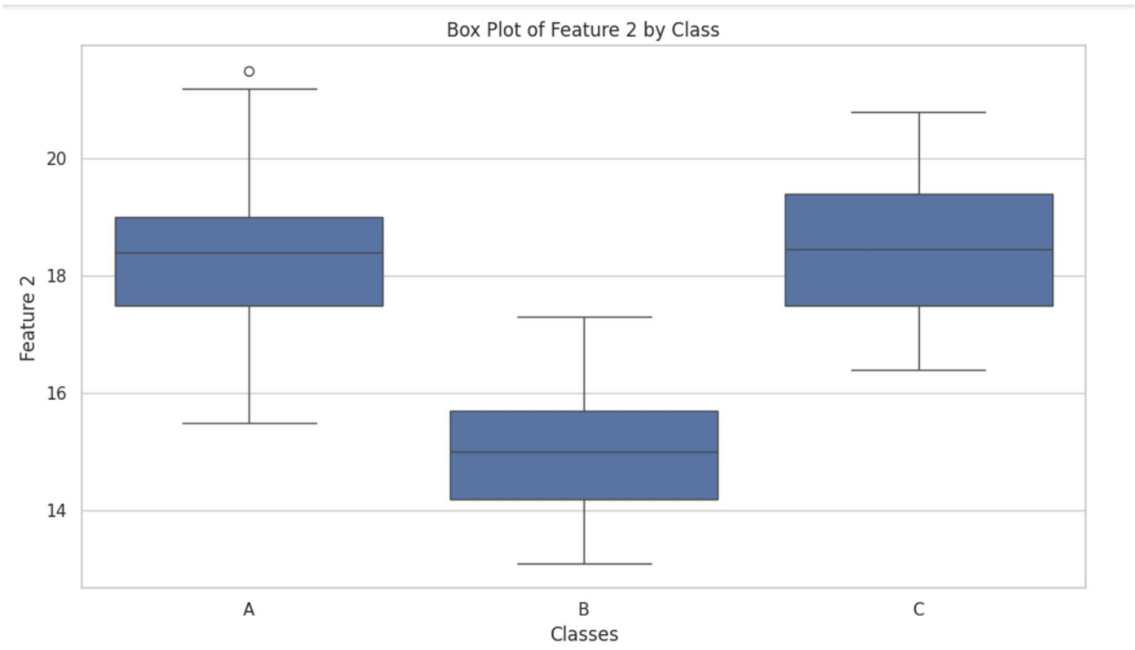```

### 6(b)

```
Data Description:
         Feature 1   Feature 2   Feature 3   Feature 4
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):
 Feature 1   9.275
Feature 2   3.100
Feature 3   23.000
Feature 4   1200.000
dtype: float64
```
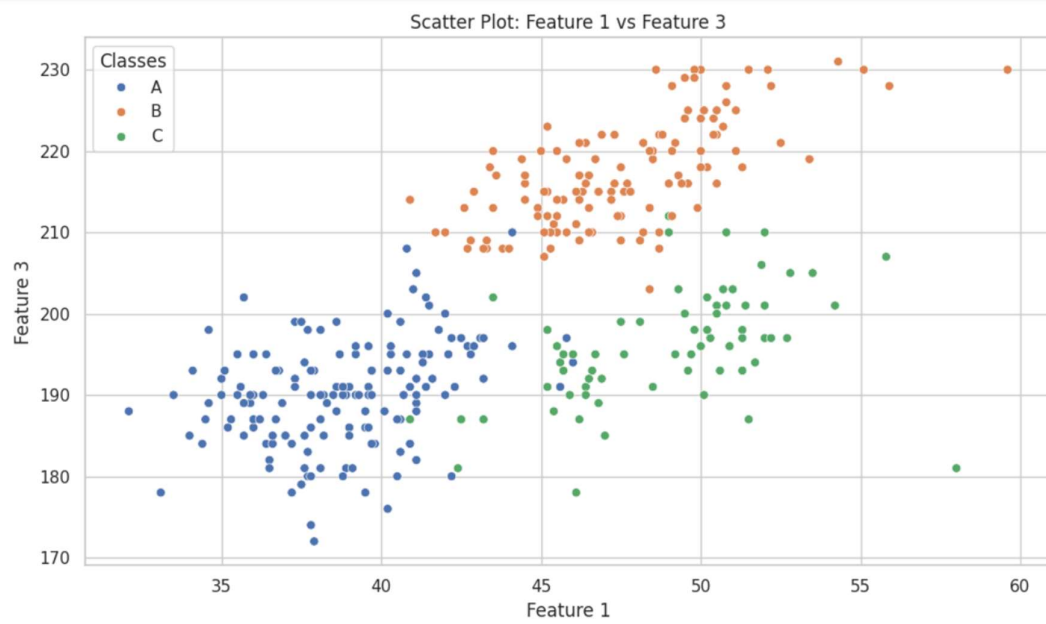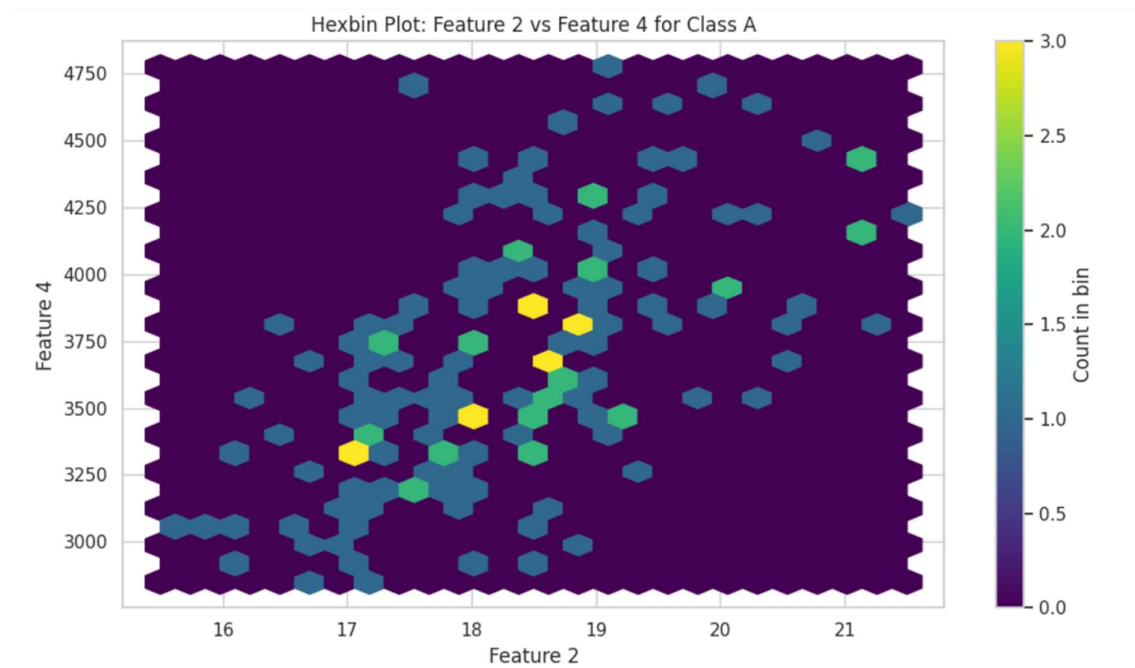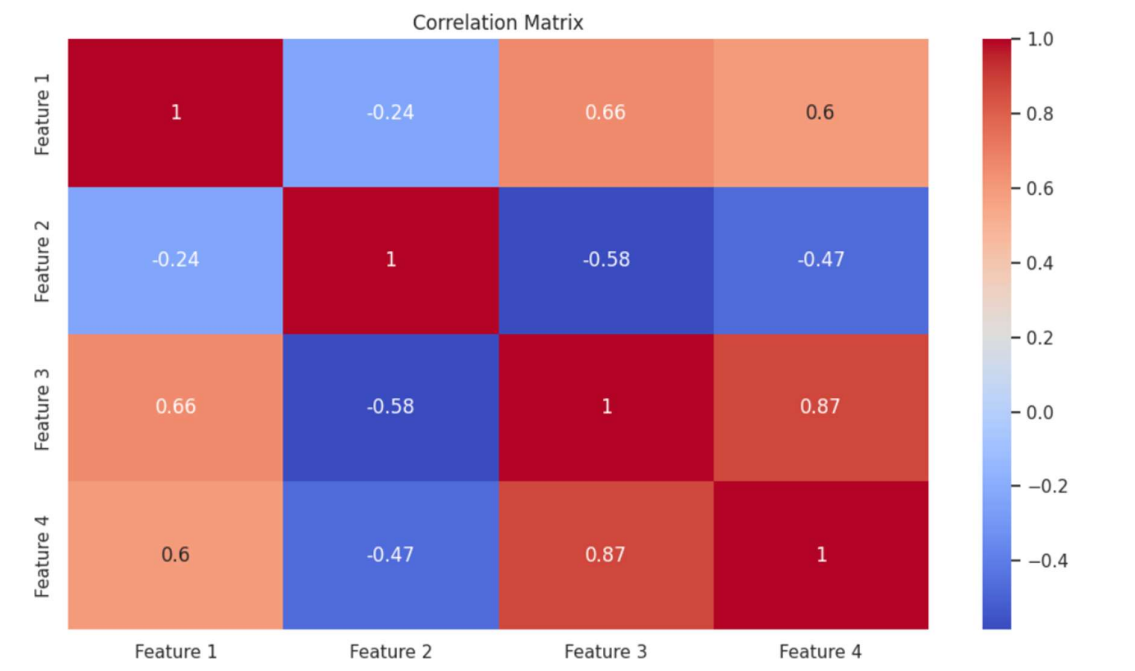
### 6(c)



Histogram of Feature 1 for Class A

6(d)


Box Plot of Feature 2 by Class

6(e)


Violin Plot of Feature 3 by Class

6(f)

Scatter Plot: Feature 1 vs Feature 3

6(g)


Contour Plot: Feature 1 vs Feature 4 by Class

6(h)

Hexbin Plot: Feature 2 vs Feature 4 for Class A

6(i)



Correlation Matrix

6(j)

**j)**



Pair Plot of Features by Class