

# Homework -1

Farhan Alam

12340740

January 19, 2025

## 1 Theory Questions

Number of points for the question is indicated in square brackets.

### Question 1 [10]

Sample from 1 to 1000 with equal probabilities and let  $x$  be the first digit (e.g.,  $x = 1$  if the number is 15). What are the probabilities  $p_1$  to  $p_9$  (adding to 1) of  $x = 1, \dots, 9$ ? What are the mean and variance of  $x$ ?

### Answer

The probabilities  $p_1$  to  $p_9$  are derived from the frequency of leading digits for each no. from 1 to 9 we have the following series where a certain no  $n$  is the leading no :-

1. The singleton no.  $n$  i.e 1 number
2. Numbers from  $[n*10, n*10+9]$  which has 10 numbers
3. Numbers from  $[n*100, n*100+99]$  which has 100 numbers

So we have a total of  $1+10+100=111$  no.'s for each  $n$  from 1 to 9 .

But for the number 1 we also have the no. 1000 therefore the total being 112 for 1.

$\therefore p_1 = \frac{112}{1000} = 0.112$  and  $p_2$  to  $p_9 = \frac{111}{1000} = 0.111$  The formula for the mean (expected value) is:

$$\mu = E[X] = \sum_{i=1}^n x_i P(x_i)$$

By putting values we get the mean as :

$$\mu = E[X] = 4.995$$

The formula for the variance is:

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2] = \sum_{i=1}^n P(x_i)(x_i - \mu)^2$$

In this case it would be formulated as:

$$\sigma^2 = \text{Var}(X) = E[(X - 4.995)^2] = \sum_{i=1}^n P(x_i)(x_i - 4.995)^2$$

By putting values we get the mean as :

$$\sigma^2 = \text{Var}(X) = 6.675$$

## Question 2 [10]

**Prelude:** Note that the convolution of two univariate Gaussian's  $p_1(x)$   $p_2(x)$  is also Gaussian.  $(p_1 * p_2)(x) = \int_{-\infty}^{\infty} p_1(t)p_2(x-t)dt = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}$

A good proof takes Fourier transforms  $F$  and uses the convolution theorem:

$$F[p_1(x) * p_2(x)] = F(p_1(x))F(p_2(x))$$

This is good because the transforms  $F(p_i(x))$  are multiples of  $\exp -\frac{\sigma_i^2 k^2}{2}$ . Multiplying those transforms gives a product. That product involves  $(\sigma_1^2 + \sigma_2^2)$ . Then the inverse Fourier transform produces  $p_1 * p_2$  as another Gaussian.

**Question:** What is the variance for the convolution of n identical Gaussian's  $N(0, \sigma^2)$  ?

### Answer

Each univariate gaussian distribution is given by :  $N(0, \sigma^2)$  We have been provided with  $n$  such gaussian distribution and have been asked the result of  $(p_1 * p_2 * \dots * p_n)(x)$

Given that for two univariate Gaussian's  $p_1(x) * p_2(x)$  is also Gaussian with distribution  $N(0, \sigma_1^2 + \sigma_2^2)$  where  $\sigma_1^2$  and  $\sigma_2^2$  are variance of  $p_1(x)$  and  $p_2(x)$  respectively.

$\therefore$  For  $(p_1 * p_2 * \dots * p_n)(x)$  we can take two gaussian distributions at a time and then find its distribution and calculate  $(p_1 * p_2 * \dots * p_n)(x)$  in this

manner . By doing so we get  $(p_1 * p_2 * \dots * p_n)(x) = \frac{1}{\sqrt{2\pi(\sum_{i=1}^n \sigma_i^2)}} e^{-\frac{x^2}{2(\sum_{i=1}^n \sigma_i^2)}}$

which resembles

$$N\left(0, \sum_{i=1}^n \sigma_i^2\right)$$

$\therefore$  The variance for the  $(p_1 * p_2 * \dots * p_n)(x)$  convolution distribution is given by  $\sum_{i=1}^n \sigma_i^2$

As each distribution has variance  $\sigma^2$ , therefore, the final answer becomes:

$$\text{Var}(p_1 * p_2 * \dots * p_n)(x) = n\sigma^2$$

### Question 3 [10]

Explain why the probability distribution  $P(x)$  for the sum of two random variables is the convolution  $P = p_1 * p_2$  of their separate probability distributions. An example comes from rolling two dice and adding the results:

Probabilities for sum  $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) * (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) =$   
Probabilities of 2 to 12  $(\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{7}{36}, \frac{8}{36}, \frac{9}{36}, \frac{10}{36}, \frac{11}{36}, \frac{1}{36})$

### Answer

The probability distribution  $P(x)$  for the sum of two random variables  $X_1$  and  $X_2$  is given by the convolution(adding) of their individual probability distributions  $P_1$  and  $P_2$ , denoted as:

$$P(x) = (P_1 * P_2)(x) = \sum_k P_1(k)P_2(x - k).$$

This convolution adds all possible ways the sum  $x$  can be achieved by adding values from  $X_1$  and  $X_2$ .

This can also be explained by rolling of dice as follows :-

When rolling two dice, each die has a uniform probability distribution  $P_1$  and  $P_2$ , where each outcome (1 through 6) has a probability of  $\frac{1}{6}$ . The result of adding the two dice is a random variable  $X = X_1 + X_2$ , where  $X_1$  and  $X_2$  are the results from each die.

The probability distribution for each die roll is:

$$P_1 = P_2 = \left( \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right).$$

To find the probability distribution for the sum  $X$ , we calculate the convolution of the two individual probability distributions.

The probability of each sum  $X = x$  (ranging from 2 to 12) is the sum of the probabilities for all pairs of outcomes  $(X_1, X_2)$  such that their sum equals  $x$ . This can be expressed as:

$$P(X = x) = \sum_k P_1(k)P_2(x - k).$$

For example: - The probability of rolling a sum of 2 comes from the pair (1, 1), so:

$$P(X = 2) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

- The probability of rolling a sum of 3 comes from the pairs (1, 2) and (2, 1), so:

$$P(X = 3) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}.$$

Similarly we can do for other no in the dice. This process continues for all sums from 2 to 12.

Which gives us the resulting distribution:-

$$P(X) = \left( \frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36} \right).$$

This distribution shows that sums near the middle (e.g., 7) are more likely, while sums at the extremes (e.g., 2 and 12) are less likely, compared to the one in middle.

The convolution  $P = P_1 * P_2$  for the sum of two random variables works

because it combines the probabilities of all possible combinations of outcomes that add up to a given sum. In the case of two dice, the resulting probability distribution is symmetric and follows the pattern described above only 7 being an exception.

#### Question 4 [10]

Prove this amazing identity when random samples  $0 < x_1 < x_2 < \dots < x_n$  have possibilities  $p_1$  to  $p_n$ :

#### Answer

$$\text{mean} = E[x] = \sum_1^n p_i x_i = \int_{t=0}^{\infty} (\text{Probability that } x > t) dt.$$

Hint: The probability is  $\sum p_i = 1$  up to  $t = x_1$  and then it is  $1 - p_1$  as far as  $t = x_2$ .

#### Answer

For a discrete random variable with ordered values  $x_1, x_2, \dots, x_n$  and corresponding probabilities  $p_1, p_2, \dots, p_n$ , the cumulative distribution function (CDF) is defined as:

$$F(t) = P(x \leq t)$$

Thus, the probability that  $x > t$  is:

$$P(x > t) = 1 - F(t)$$

Substituting  $P(x > t) = 1 - F(t)$  into the integral we get :

$$\int_0^{\infty} (1 - F(t)) dt$$

For the given distribution  $x_1 < x_2 < \dots < x_n$ , the CDF  $F(t)$  can be written as:

$$F(t) = \begin{cases} 0, & \text{if } t < x_1 \\ \sum_{i=1}^k p_i, & \text{if } x_k \leq t < x_{k+1}, \text{ for some } k \\ 1, & \text{if } t \geq x_n \end{cases}$$

The integral can be then separated into intervals based on the ordered values  $x_1, x_2, \dots, x_n$ :

$$\int_0^\infty (1-F(t)) dt = \int_0^{x_1} (1-F(t)) dt + \int_{x_1}^{x_2} (1-F(t)) dt + \dots + \int_{x_{n-1}}^{x_n} (1-F(t)) dt$$

For  $t < x_1$ ,  $F(t) = 0$ , so  $1 - F(t) = 1$ . The integral becomes:

$$\int_0^{x_1} 1 dt = x_1$$

Similarly for  $t \in [x_k, x_{k+1})$ ,  $F(t) = \sum_{i=1}^k p_i$ , so  $1 - F(t) = 1 - \sum_{i=1}^k p_i$ . The integral becomes:

$$\int_{x_k}^{x_{k+1}} \left(1 - \sum_{i=1}^k p_i\right) dt = (x_{k+1} - x_k) \left(1 - \sum_{i=1}^k p_i\right)$$

Summing over all intervals, we get:

$$\mathbb{E}[x] = x_1 + (x_2 - x_1)(1 - p_1) + (x_3 - x_2)(1 - (p_1 + p_2)) + \dots + (x_n - x_{n-1}) \left(1 - \sum_{i=1}^{n-1} p_i\right)$$

Which upon adding we get is equal to:

$$\sum_{i=1}^n p_i x_i$$

which in turn is equal to:

$$E[X]$$

Hence proved,

## MML 6.2, Question 5 [10]

Consider a matrix of two Gaussian distributions

$$0.4\mathcal{N}\left(\begin{bmatrix} 10 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 0.6\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.4 & 2.0 \\ 2.0 & 1.7 \end{bmatrix}\right).$$

- Compute the marginal distributions for each dimension.
- Compute the mean, mode and median for each marginal distribution.
- Compute the mean and mode for the two-dimensional distribution.

## Answer

We are given the mixture of two Gaussian distributions:

$$P(x) = 0.4 \cdot N\left(\begin{bmatrix} 10 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 0.6 \cdot N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.4 & 2.0 \\ 2.0 & 1.7 \end{bmatrix}\right).$$

The marginal distributions for each component are as follows:

1. For the first Gaussian we have:

$$X_1^{(1)} \sim N(10, 1), \quad X_2^{(1)} \sim N(2, 1).$$

2. For the second Gaussian we have:

$$X_1^{(2)} \sim N(0, 8.4), \quad X_2^{(2)} \sim N(0, 1.7).$$

Overall we have:

$$P_{X_1}(x_1) = 0.4 \cdot N(10, 1) + 0.6 \cdot N(0, 8.4),$$

$$P_{X_2}(x_2) = 0.4 \cdot N(2, 1) + 0.6 \cdot N(0, 1.7).$$

Now we will calculate the mean mode and meadian

1. Mean:

$$\mu_{X_1} = 0.4 \cdot 10 + 0.6 \cdot 0 = 4, \quad \mu_{X_2} = 0.4 \cdot 2 + 0.6 \cdot 0 = 0.8.$$

2. Mode: The mode of a mixture distribution is numerically estimated.

For simplicity:

$$\text{Mode}X_1 \approx 0, \quad \text{Mode}X_2 \approx 0.$$

1. Mean:

$$\mu = 0.4 \cdot \begin{bmatrix} 10 \\ 2 \end{bmatrix} + 0.6 \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 0.8 \end{bmatrix}.$$

2. Mode: The mode must be numerically determined from the combined density function.

## MML 6.3, Question 6 [10]

You have written a computer program that sometimes compiles and sometimes not (code does not change). You decide to model the apparent stochasticity (success vs no success)  $x$  of the compiler using a Bernoulli distribution with parameter  $\mu$ :

$$p(x | \mu) = \mu^x(1 - \mu)^{1-x}(1 - x), \quad x \in \{0, 1\}$$

## Answer

We are given that the success (0 or 1) of the compiler follows a Bernoulli distribution, which is expressed as:

$$p(x|\mu) = \mu^x(1 - \mu)^{1-x}, \quad x \in \{0, 1\}.$$

## Conjugate Prior

The conjugate prior corresponding to the Bernoulli likelihood is the Beta distribution, given by:

$$p(\mu|\alpha, \beta) = \frac{\mu^{\alpha-1}(1 - \mu)^{\beta-1}}{B(\alpha, \beta)},$$

where  $B(\alpha, \beta)$  is the Beta function, defined as:

$$B(\alpha, \beta) = \int_0^1 \mu^{\alpha-1}(1 - \mu)^{\beta-1} d\mu.$$

## Likelihood Function

The likelihood function for observing the data points  $x_1, x_2, \dots, x_N$  is:

$$p(x_1, x_2, \dots, x_N|\mu) = \prod_{i=1}^N p(x_i|\mu) = \prod_{i=1}^N \mu^{x_i}(1 - \mu)^{1-x_i}.$$

Let:

$$S = \sum_{i=1}^N x_i,$$

which represents the total number of successes. The likelihood function can then be simplified as:

$$p(x_1, x_2, \dots, x_N|\mu) = \mu^S(1 - \mu)^{N-S}.$$

## Posterior Distribution

By applying Bayes' theorem, we can find the posterior distribution:

$$p(\mu|x_1, x_2, \dots, x_N) \propto p(x_1, x_2, \dots, x_N|\mu) \cdot p(\mu|\alpha, \beta).$$



Substituting in the likelihood and prior distributions:

$$p(\mu|x_1, x_2, \dots, x_N) \propto \mu^S(1 - \mu)^{N-S} \cdot \mu^{\alpha-1}(1 - \mu)^{\beta-1}.$$

Now, we combine the exponents:

$$p(\mu|x_1, x_2, \dots, x_N) \propto \mu^{S+\alpha-1}(1 - \mu)^{N-S+\beta-1}.$$

This expression is the kernel of a Beta distribution, so the posterior distribution is:

$$p(\mu|x_1, x_2, \dots, x_N) = \text{Beta}(\mu|S + \alpha, N - S + \beta).$$

## Posterior Parameters

The parameters of the posterior Beta distribution are:

$$\alpha_{\text{posterior}} = S + \alpha, \quad \beta_{\text{posterior}} = N - S + \beta.$$

This describes the posterior distribution for the success probability  $\mu$ , given the data and the prior Beta distribution.

## MML 6.4, Question 7 [10]

There are two bags. The first bag contains four mangoes and two apples; the second bag contains four mangoes and four apples.

We also have a biased coin, which shows "heads" with a probability 0.6 and "tails" with a probability 0.4. If the coin shows "heads" we pick a fruit at random from bag1; otherwise, we pick a fruit at random from bag 2.

Your friend flips the coin (you cannot see the result), picks a fruit randomly from the corresponding bag, and presents you as a mango.

What is the probability that the mango was picked from bag 2?

*Hint:* Use Bayes' theorem.

## Answer

Given that:

$$p(\text{Head})=0.6$$

$$p(\text{Tail})=0.4$$

No.of fruits in bag 1=6

No.of mangoes in bag 1=4

No.of fruits in bag 2=8

No.of mangoes in bag 2=4

$$p(\text{getting mango from bag 1}) = \frac{4}{6} = 0.66$$

$$p(\text{getting mango from bag 2}) = \frac{4}{8} = 0.50$$

But for choosing these bags we are flipping the coin which has heads and tails probability as mentioned above and as flipping is independent of getting mango from the bag  $\therefore$

$$p(\text{getting mango from bag 1}) = p(\text{Head}) * \frac{4}{6} = 0.6 * \frac{4}{6} = 0.396$$

$$p(\text{getting mango from bag 2}) = p(\text{Tail}) * \frac{4}{8} = 0.4 * \frac{4}{8} = 0.2$$

$\therefore$

$$p(\text{mango}) = 0.396 + 0.2 = 0.596$$

We have been asked to find the probability of a being from bag 2 given that we have received a mango; this can be done by Bayes theorem as follows:

Let the probability of getting a mango from bag 1 be specified as: B2

Let the probability of getting a mango be specified as : M

So, we have been technically asked the value of :

$p(B_2 | M)$ , which by Bayes' law, can be written as:

$$p(B_1 | M) = \frac{p(B_2 \cap M)}{p(M)} = \frac{0.2}{0.596} = 0.33$$

## MML 6.5, Question 8 [30]

Consider the time series model:

$$x_{t+1} = \mathbf{A}x_t + \omega, \quad \omega \sim \mathcal{N}(0, Q)$$

$$y_t = \mathbf{C}x_t + v, \quad v \sim \mathcal{N}(0, R)$$

where  $\omega, v$  are i.i.d. Gaussian noise variables.

Further, assume that  $p(x_0) = N(\mu_0, \Sigma_0)$ .

- a. What is the form of  $p(x_0, x_1, \dots, x_T)$  ? Justify your answer ( you do not have to explicitly compute the joint distribution).
- b. Assume that  $p(x_t|y_1, \dots, y_T) = N(\mu_t, \Sigma_t)$ .
  1. Compute  $p(x_{t+1}|y_1, \dots, y_T)$ .
  2. Compute  $p(x_{t+1}, |y_{t+1}|y_1, \dots, y_T)$ .
  3. At time  $t+1$ , we observe that the value  $y_{t+1} = \hat{y}$ . Compute the conditional distribution  $p(x_{t+1}|y_1, \dots, y_{t+1})$ .

## Model Setup

We define the state-space model as follows:

$$x_{t+1} = Ax_t + \omega, \quad \omega \sim \mathcal{N}(0, Q),$$

$$y_t = Cx_t + v, \quad v \sim \mathcal{N}(0, R),$$

with the prior distribution:

$$p(x_0) = \mathcal{N}(\mu_0, P_0).$$

—

### Part (a): Form of $p(x_0, x_1, \dots, x_T)$

The joint distribution  $p(x_0, x_1, \dots, x_T)$  is a multivariate Gaussian distribution due to the linear nature of the model and the Gaussian noise terms. Here's the reasoning: 1. The prior  $p(x_0)$  is Gaussian:  $\mathcal{N}(\mu_0, P_0)$ . 2. The state transition equation  $x_{t+1} = Ax_t + \omega$  preserves the Gaussian distribution, as linear transformations of Gaussian variables plus Gaussian noise result in a Gaussian distribution. 3. Hence, the joint distribution is:

$$p(x_0, x_1, \dots, x_T) = \mathcal{N}(\mu, \Sigma),$$

where  $\mu$  and  $\Sigma$  are the mean vector and covariance matrix, which depend on the system dynamics  $A$ , noise covariance  $Q$ , and the initial conditions.

—

## Part (b): Recursive Computations

Let the posterior distribution at time  $t$  be:

$$p(x_t|y_1, \dots, y_T) = \mathcal{N}(\mu_t, P_t).$$

### 1. Compute $p(x_{t+1}|y_1, \dots, y_T)$

Using the state equation  $x_{t+1} = Ax_t + \omega$ , the conditional distribution is:

$$p(x_{t+1}|x_t) = \mathcal{N}(Ax_t, Q).$$

Taking the expectation over the posterior  $p(x_t|y_1, \dots, y_T)$ , we get the predicted distribution:

$$p(x_{t+1}|y_1, \dots, y_T) = \mathcal{N}(A\mu_t, AP_tA^\top + Q).$$

### 2. Compute $p(x_{t+1}, y_{t+1}|y_1, \dots, y_T)$

Using the observation equation  $y_{t+1} = Cx_{t+1} + v$ , the joint distribution of  $(x_{t+1}, y_{t+1})$  is Gaussian:

$$p(x_{t+1}, y_{t+1}|y_1, \dots, y_T) = \mathcal{N}\left(\begin{bmatrix} A\mu_t \\ CA\mu_t \end{bmatrix}, \begin{bmatrix} AP_tA^\top + Q & (AP_tA^\top + Q)C^\top \\ C(AP_tA^\top + Q) & C(AP_tA^\top + Q)C^\top + R \end{bmatrix}\right).$$

Let us define temporary variables for simplicity:

$$\mu_{t+1} = A\mu_t, \quad \Sigma_{t+1} = A\Sigma_t A^\top + Q, \quad p(x_{t+1}|y_1, \dots, y_t) = \mathcal{N}(\mu_{t+1}, \Sigma_{t+1})$$

Then,  $y_{t+1}$  is derived from the parameters of the distribution  $p(x_{t+1}|y_1, \dots, y_t)$  using similar steps as in the first question, resulting in:

$$p(y_{t+1}|y_1, \dots, y_t) = \mathcal{N}(y_{t+1}|CA\mu_t, C(AP_tA^\top + Q)C^\top + R)$$

The conditional distribution of  $x_{t+1}$  given the previous observations and the new observation  $y_{t+1}$  is obtained as:

$$p(x_{t+1}|y_1, \dots, y_t, y_{t+1}) = \frac{p(y_{t+1}, x_{t+1}|y_1, \dots, y_t)}{p(y_{t+1}|y_1, \dots, y_t)}$$

This is equivalent to:

$$= \frac{\mathcal{N}(y_{t+1}|CA\mu_t, R)\mathcal{N}(x_{t+1}|A\mu_t, A\Sigma_t A^\top + Q)}{\mathcal{N}(y_{t+1}|CA\mu_t, C(AP_tA^\top + Q)C^\top + R)}$$

## MML 6.12, Question 9 [40]

**Manipulation of Gaussian Random Variable** Consider a Gaussian random variable  $x \sim \mathcal{N}(x|\mu_x, \Sigma_x)$ , where  $x \in \mathbb{R}^D$ . Furthermore, we have

$$y = Ax + b + \omega,$$

where  $y \in \mathbb{R}^D$ ,  $A \in \mathbb{R}^{E \times D}$ ,  $b \in \mathbb{R}^E$  and  $\omega \sim \mathcal{N}(\omega|0, Q)$  is independent Gaussian noise. “Independent” implies that  $x$  and  $\omega$  are independent random variables and that  $Q$  is diagonal. a. Write down the likelihood  $p(y|x)$ .

b. The distribution  $p(y) = \int p(y|x)p(x)dx$  is Gaussian. Compute the mean  $\mu_y$  and covariance  $\Sigma_y$ . Derive your result in detail.

c. The random variable  $y$  is being transformed according to the measurement mapping

$$z = Cy + v,$$

where  $z \in \mathbb{R}^F$ ,  $C \in \mathbb{R}^{F \times E}$ ,  $v \sim \mathcal{N}(v|0, R)$  is independent Gaussian (measurement) noise.

- Write down  $p(z|y)$
- Compute  $p(z)$ , i.e. the mean  $\mu_z$  and the covariance  $\Sigma_z$ . Derive your result in detail.
- d. Now, a value  $\hat{y}$  is measured. Compute the posterior distribution  $p(x|\hat{y})$ .  
*Hint for Solution:* This posterior is also Gaussian, i.e. we need to determine only its mean and covariance matrix. Start by explicitly computing the joint Gaussian  $p(x, y)$ . This also requires us to compute the cross-covariances  $Cov_{x,y}[x, y]$  and  $Cov_{y,x}[y, x]$ . Then apply the rules for Gaussian conditioning.

## Answer

We are given:

- $x \sim \mathcal{N}(\mu_x, P_x)$ , where  $x \in \mathbb{R}^D$ .
- $y = Ax + b + \omega$ , where  $y \in \mathbb{R}^E$ ,  $A \in \mathbb{R}^{E \times D}$ ,  $b \in \mathbb{R}^E$ , and  $\omega \sim \mathcal{N}(0, Q)$  with  $Q$  being diagonal. The noise  $\omega$  is independent of  $x$ .

**(a): Likelihood  $p(y|x)$**

The likelihood  $p(y|x)$  follows directly from the relationship:

$$y = Ax + b + \omega,$$

where  $\omega \sim \mathcal{N}(0, Q)$ . Given  $x$ ,  $y$  is Gaussian with:

$$p(y|x) = \mathcal{N}(y|Ax + b, Q).$$

---

**(b): Distribution  $p(y)$**

The marginal distribution  $p(y)$  is obtained by integrating out  $x$ :

$$p(y) = \int p(y|x)p(x) dx.$$

Since both  $p(y|x)$  and  $p(x)$  are Gaussian,  $p(y)$  is also Gaussian. To compute its parameters:

1. Mean  $\mu_y$ :

$$\mu_y = \mathbb{E}[y] = \mathbb{E}[Ax + b + \omega] = A\mu_x + b.$$

2. Covariance  $P_y$ :

$$P_y = \text{Var}(y) = \mathbb{E}[(y - \mu_y)(y - \mu_y)^\top].$$

Substituting  $y = Ax + b + \omega$ :

$$P_y = AP_xA^\top + Q.$$

Thus:

$$p(y) = \mathcal{N}(y|\mu_y, P_y), \quad \text{where } \mu_y = A\mu_x + b, \quad P_y = AP_xA^\top + Q.$$

---

**(c): Transformation  $z = Cy + v$**

We are given:

$$z = Cy + v, \quad v \sim \mathcal{N}(0, R),$$

where  $C \in \mathbb{R}^{F \times E}$  and  $v$  is independent of  $y$ .

**Likelihood  $p(z|y)$ :**

The likelihood  $p(z|y)$  follows from the relationship:

$$z = Cy + v, \quad v \sim \mathcal{N}(0, R).$$

Thus:

$$p(z|y) = \mathcal{N}(z|Cy, R).$$

**Distribution  $p(z)$ :**

The marginal distribution  $p(z)$  is obtained by integrating out  $y$ :

$$p(z) = \int p(z|y)p(y) dy.$$

Since  $p(z|y)$  and  $p(y)$  are Gaussian,  $p(z)$  is also Gaussian. To compute its parameters:

1. Mean  $\mu_z$ :

$$\mu_z = \mathbb{E}[z] = \mathbb{E}[Cy + v] = C\mu_y.$$

2. Covariance  $P_z$ :

$$P_z = \text{Var}(z) = CP_yC^\top + R.$$

Thus:

$$p(z) = \mathcal{N}(z|\mu_z, P_z), \quad \text{where } \mu_z = C\mu_y, \quad P_z = CP_yC^\top + R.$$

—

**(d): Posterior  $p(x|\hat{y})$**

Given a measurement  $\hat{y}$ , the posterior distribution  $p(x|\hat{y})$  is Gaussian. To compute its mean and covariance, we start with the joint distribution  $p(x, y)$ .

**Joint Distribution  $p(x, y)$ :**

From the model:

$$p(x) = \mathcal{N}(x|\mu_x, P_x), \quad p(y|x) = \mathcal{N}(y|Ax + b, Q).$$

The joint distribution  $p(x, y)$  is Gaussian:

$$p(x, y) = \mathcal{N} \left( \begin{bmatrix} x \\ y \end{bmatrix} \middle| \begin{bmatrix} \mu_x \\ A\mu_x + b \end{bmatrix}, \begin{bmatrix} P_x & P_xA^\top \\ AP_x & AP_xA^\top + Q \end{bmatrix} \right).$$

**Posterior**  $p(x|\hat{y})$ :

Using Gaussian conditioning, the posterior  $p(x|\hat{y})$  is:

$$p(x|\hat{y}) = \mathcal{N}(x|\mu_{x|\hat{y}}, P_{x|\hat{y}}),$$

where:

$$\begin{aligned}\mu_{x|\hat{y}} &= \mu_x + P_x A^\top (AP_x A^\top + Q)^{-1} (\hat{y} - (A\mu_x + b)), \\ P_{x|\hat{y}} &= P_x - P_x A^\top (AP_x A^\top + Q)^{-1} AP_x.\end{aligned}$$

## 2 Computer Based

### Question 10 [10]

Computer Experiment: Find the average  $A_{1000000}$  of a million random 0-1 samples! What is the value of the standardized variable:  $X = \frac{A_N - \frac{1}{2}}{\frac{1}{2\sqrt{N}}}$ ?

#### 2.1 Answer

Here is the Python code:

```
import random
import statistics as st
a = []
for i in range(0, 1000000):
    a.append(random.randrange(2))
print(st.mean(a)) #prints the average of the million random 0-1 sample
print((st.mean(a) - 1/2) / (1/(2*1000))) #prints standardized value
```

The resultant mean is:0.500317

The resultant standardized value is:0.158000

This is very close to the true mean value (0.500) which is calculated theoretically .Hence proving the Central limit theorem to be true. The standardized value wpuld tend to  $N(0,1)$  hencen giving us a result which is very close to 0.



## Question 11 [30]

Dataset: “<https://github.com/mwaskom/seaborn-data/blob/master/penguins.csv>”

Problem:

1. Download the dataset. Understand and summarize the dataset
2. Perform EDA
3. Identify clusters wherever possible.
4. Think of different ways you can predict the species a given penguin belongs to...

## Answer

Here is the python code :-

```
import pandas as pd
import seaborn as sns
url="https://raw.githubusercontent.com/mwaskom/seaborn-data/refs/heads/master/penguins.csv"
df=pd.read_csv(url)
print(df.shape)
df.dropna(inplace=True)
print(df['species'].value_counts())
print(df['island'].value_counts())
print(sns.scatterplot(data=df,x='flipper_length_mm',y='body_mass_g',hue='species'))
import matplotlib.pyplot as plt
plt.figure()
print(sns.scatterplot(data=df,x='flipper_length_mm',y='body_mass_g',hue='species'))
plt.figure()
sns.countplot(data=df, x='species', hue='island')
plt.figure()
sns.scatterplot(data=df,x='bill_length_mm',y='bill_depth_mm',hue='species')
plt.figure()
sns.scatterplot(data=df,x='flipper_length_mm',y='body_mass_g',hue='sex')
plt.figure()
print(sns.countplot(data=df,x='species',hue='sex'))
```

1.First, the data is loaded as a data frame and then the rows which were having nan values were dropped thus reducing row length from 344 to 333 as we still have 333 rows we have enough data to perform EDA. 2.Then we calculated the no. of penguin per species which gave us the result as follows :-

<b>Species</b>	<b>Count</b>
Adelie	146
Gentoo	119
Chinstrap	68

Table 1: Distribution of penguin species in the dataset.

Which shows Adeilie has the highest presence followed by gentoo then followed by Chinstrap.

3.Then we calculated the no. of penguin residing on each island which gave us the result as follows:-

From this we can observe and say that Biscoe and Dream island have favor-

<b>Island</b>	<b>Number of Penguins</b>
Biscoe	163
Dream	123
Torgersen	47

Table 2: Penguin counts on different islands.

able condtion for penguins to survive where as Torgensen island is not that favourable for penguins to survive.

4.Then we plotted a scatterplot between flipper length and body mass with species being the hue. and got the result as follows :-

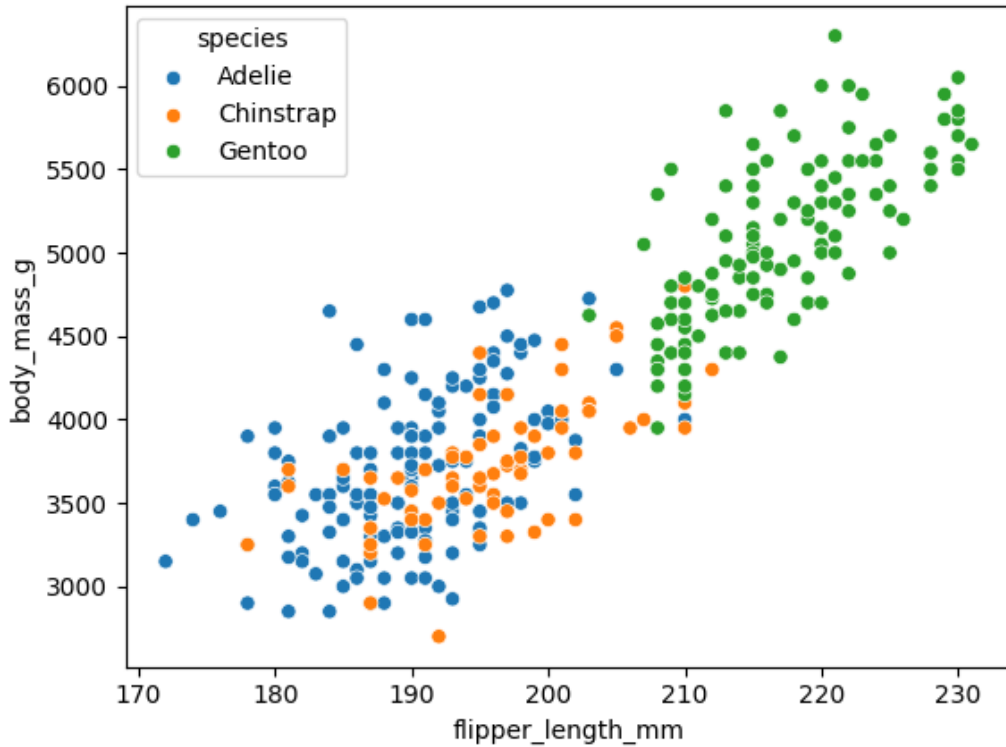


Figure 1: Flipper length vs body mass.

From this we can easily conclude that as flipper length increase body mass also increases for the penguins and the 'Gentoo' species have a higher body mass compared to the 'Chinstrap' and 'Adelie' species. Adelie and Chinstrap species have equivalent mass but the mass of Adelie might be slightly higher on average. Thus we can build the following relation :-

$$\text{Mass of Gentoo} > \text{Mass of Adelie} \geq \text{Mass of Chinstrap}.$$

5. Next we plotted flipper length vs body mass with hue as the island on which the penguins are residing and got the result as follows :

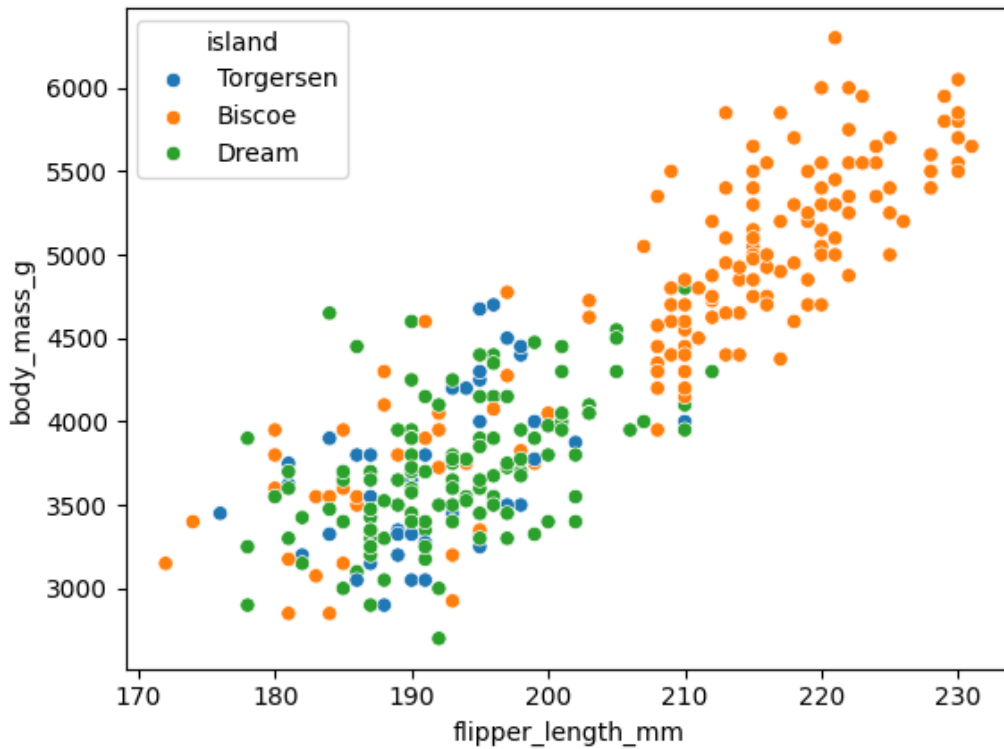


Figure 2: flipper length vs body mass.

From this we can observe that the mass of penguins residing on the Biscoe island on an average is the highest then the other two have similar attributes when it comes to weights this signifies high presence of 'Gentoo' species on the Biscoe island. 6. Next we plotted the count of each species on each island and the plot was as follows:-

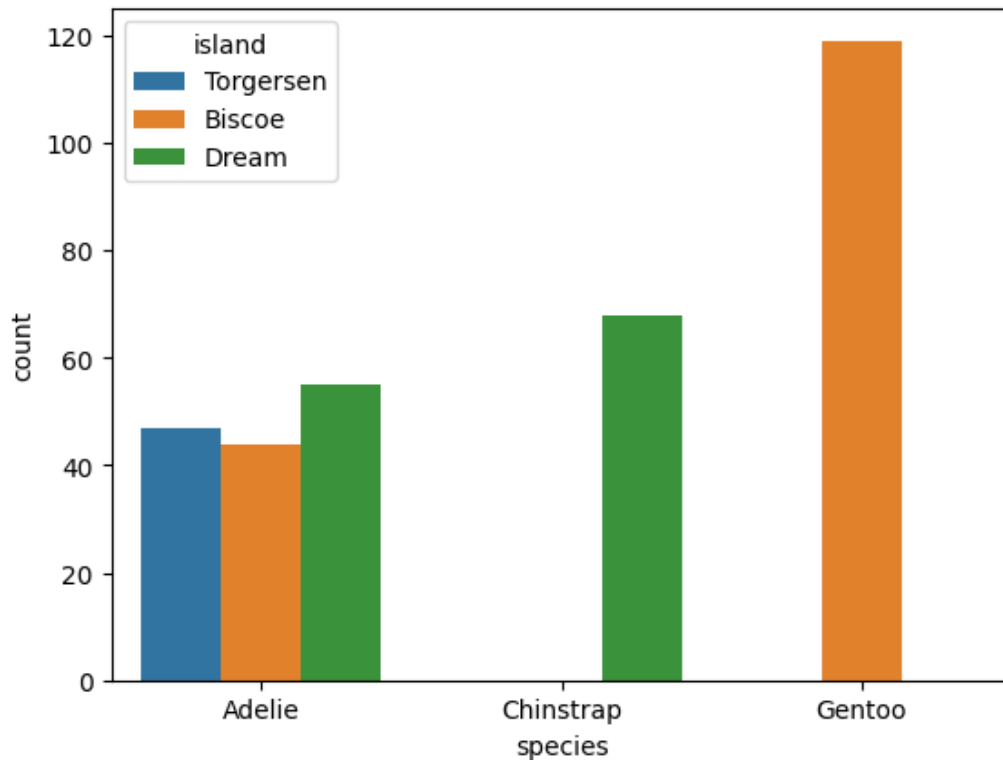


Figure 3: flipper length vs body mass.

From this we can see that the gentoo species are only found on the Biscoe island and the Chinstrap species are found only on the Dream island where as the Adeile species are found on all the islands. From this we can confer that the Adeile species is capabale of surviving on all three islands but the other two are island specific. From this we can say that the Adeile species can easily adapt to their surrounding whereas the other species cannot.

7. Next we plotted the graph between bill length and bill depth and took hue as species and got the following graph as follows :-

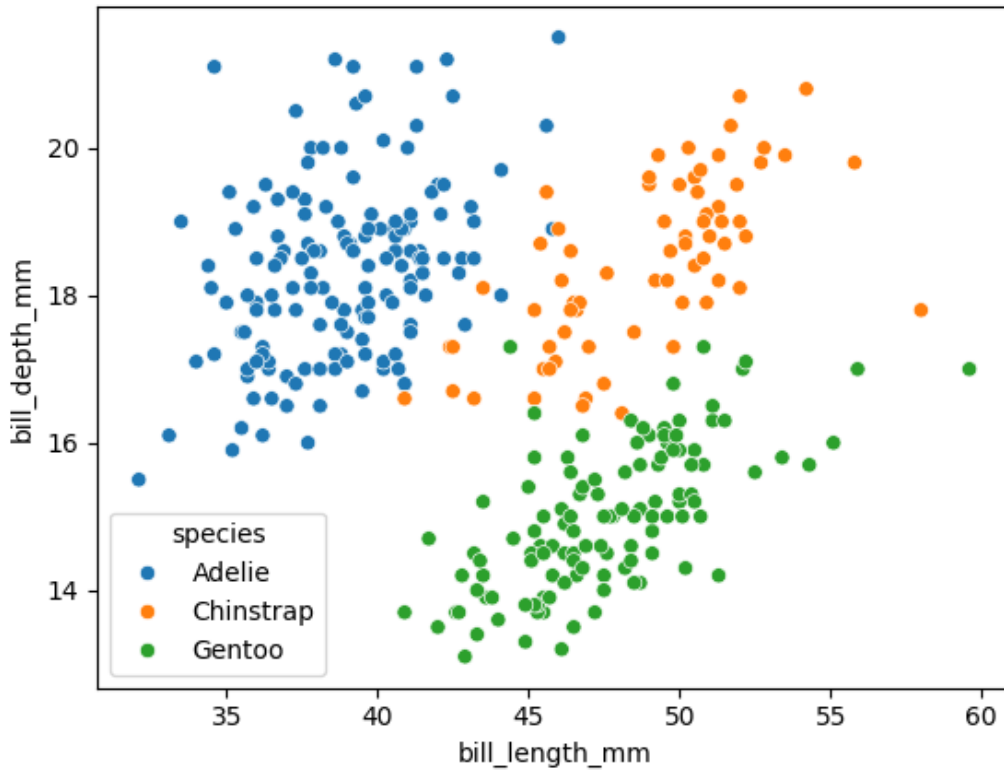


Figure 4: flipper length vs body mass.

From this we can infer the following -

1. Adelie species have more bill depth compared to bill length.
2. Gentoo species have more bill length compared to bill depth.
3. Chinstrap species have similar bill depth and bill length.

8. Last but not least, a countgraph of each species was plotted with hue as sex (male, female) and the results were as follows :-

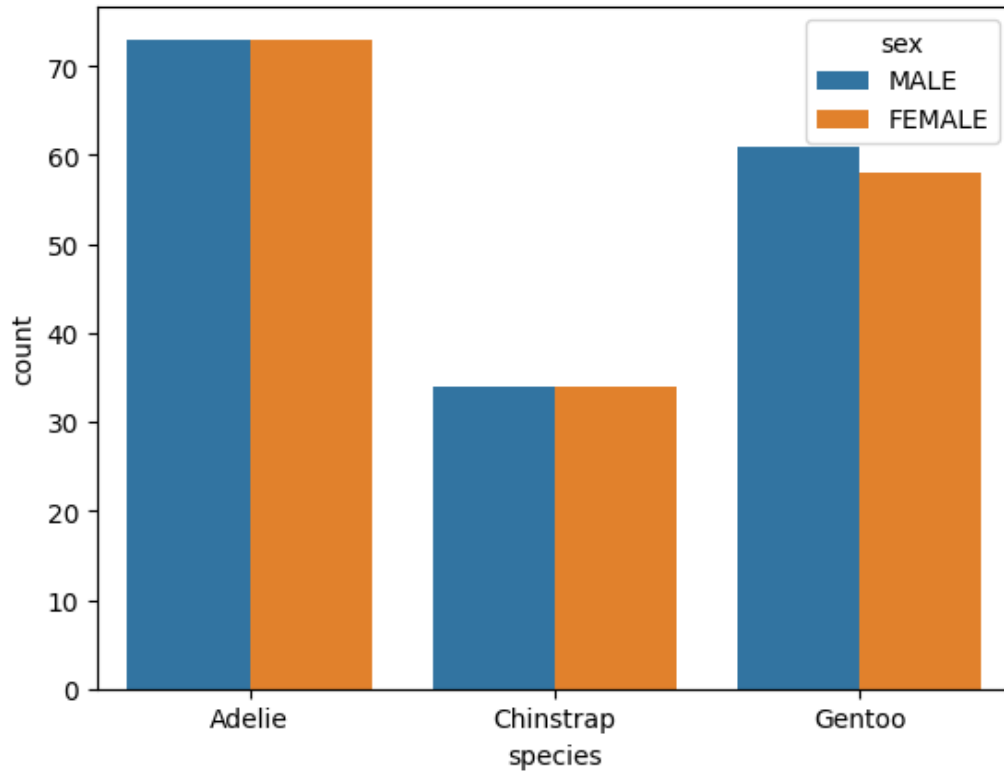


Figure 5: flipper length vs body mass.

Here we can see that there are nearly or equal population of both sex in all three species thus having a thriving species.

## 2.2 Key Takeaways:

So our key takeaways from the EDA of the file are :-

1. Adelie species are higher in population followed by gentoo then followed by Chinstrap.
2. Biscoe and Dream Island are quite favorable for penguin survival compared to Torgensen Island.
- 3.

Mass of Gentoo  $>$  Mass of Adelie  $\geq$  Mass of Chinstrap.

4.The Gentoo species is found only on Biscoe Island, and the Chinstrap species is exclusive to Dream Island, while the Adelie species is present on all three islands. This suggests that the Adelie species is more adaptable, capable of surviving on all islands, while the Gentoo and Chinstrap species are more island-specific.

5.The Adelie species has a greater bill depth than bill length, the Gentoo species exhibits a longer bill length than bill depth, while the Chinstrap species has similar measurements for both bill depth and length.

To predict the species of a penguin, you can use the following characteristics:

Population Size: Adelie  $\downarrow$  Gentoo  $\downarrow$  Chinstrap.

Island: Gentoo is found only on Biscoe, Chinstrap on Dream, and Adelie on all three.

Mass: Gentoo  $\downarrow$  Adelie  $\downarrow$  Chinstrap.

Bill Dimensions: Adelie has more bill depth than length, Gentoo has a longer bill length, Chinstrap has similar bill depth and length.

Prediction Approach:

Mass: Larger mass  $\rightarrow$  Gentoo, smaller  $\rightarrow$  Chinstrap or Adelie.

Island: If on Torgensen  $\rightarrow$  likely Adelie, if on Biscoe  $\rightarrow$  Gentoo, if on Dream  $\rightarrow$  Chinstrap.

Bill Dimensions: Measure the ratio of bill depth to length for classification.

We can use a classifier like Decision Trees or Logistic Regression to predict species based on these features.

The colab link for the computer based assignment is as follows :-

<https://colab.research.google.com/drive/1oAkvXivgzg5T6AvLnm3l9Gn1F3nZMiwQ?usp=sharing>

## 2.3 Data Collection Question

## 3 Introduction

This report presents an exploratory data analysis (EDA) on a dataset containing smartphone usage information with age ,no.of years of usage and price of smartphone. The analysis includes data cleaning, anomaly removal, statistical summary, visualizations, and predictive modeling.



## 4 Data Cleaning and Preprocessing

- **Converted data types:** The 'Price of smartphone' column was converted to a numeric type for proper analysis.
- **Handled missing values:** Missing values in 'Price of smartphone' were replaced with the median price.
- **Removed anomalies:**
  - Entries with unrealistic smartphone ages ( $>10$  years) were removed.
  - Extreme price values ( $>100,000$ ) were also excluded.

## 5 Summary Statistics

After cleaning and augmentation, the dataset contains **69 entries**. Key statistics include:

- **Average smartphone age:** 2.1 years
- **Median smartphone price:** 20,000
- **Price range:** 10,000 - 70,000
- **Age range of respondents:** 17 to 28 years

## 6 Visualizations and Insights

### 6.1 Age Distribution

The majority of respondents are between **18-21 years old**, indicating a younger demographic.

### 6.2 Smartphone Age Distribution

Most users have smartphones aged between **1-3 years**, suggesting frequent upgrades.

## 6.3 Price Distribution

The price distribution is skewed, with **most smartphones priced between 15,000 - 25,000**. A few premium smartphones ( > 70,000) are owned by a small fraction of users.

# 7 Predictive Modeling

A regression model was developed to predict smartphone prices based on user age and smartphone age.

## 7.1 Machine Learning Model: Linear Regression

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
```

Feature selection

```
X = df_extended[['Age', 'How old is your smartphone (in years)?']]
y = df_extended['Price of smartphone']
```

Splitting data into training and testing sets

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Training the model

```
model = LinearRegression()
model.fit(X_train, y_train)
```

Predictions

```
y_pred = model.predict(X_test)
```

Model Evaluation

```
mae = mean_absolute_error(y_test, y_pred)
```

```

mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

print(f"Mean Absolute Error: {mae}")
print(f"Root Mean Squared Error: {rmse}")

```

## 7.2 Clustering with K-Means

To group similar smartphone users based on their age and smartphone price, we apply the K-Means clustering algorithm.

```

from sklearn.cluster import KMeans
import seaborn as sns

```

Selecting features for clustering

```

X_cluster = df_extended[['Age', 'Price of smartphone']]

```

Applying K-Means

```

kmeans = KMeans(n_clusters=3, random_state=42)
df_extended['Cluster'] = kmeans.fit_predict(X_cluster)

```

Visualization

```

plt.figure(figsize=(8, 6))
sns.scatterplot(x=df_extended['Age'], y=df_extended['Price of smartphone'], hue=df_extended['Cluster'])
plt.title("K-Means Clustering of Smartphone Users")
plt.xlabel("Age")
plt.ylabel("Price of Smartphone")
plt.show()

```

Below are certain visualizations which we achieved thorough analysis.

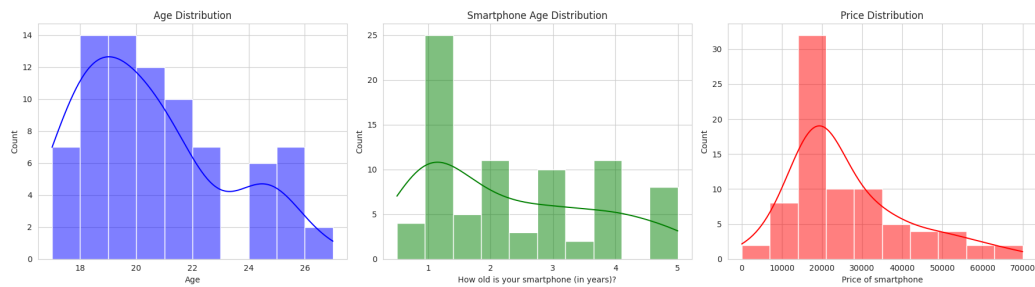


Figure 6: various comparison

The below diagram tells us about popularity of smartphone brand among the users:-

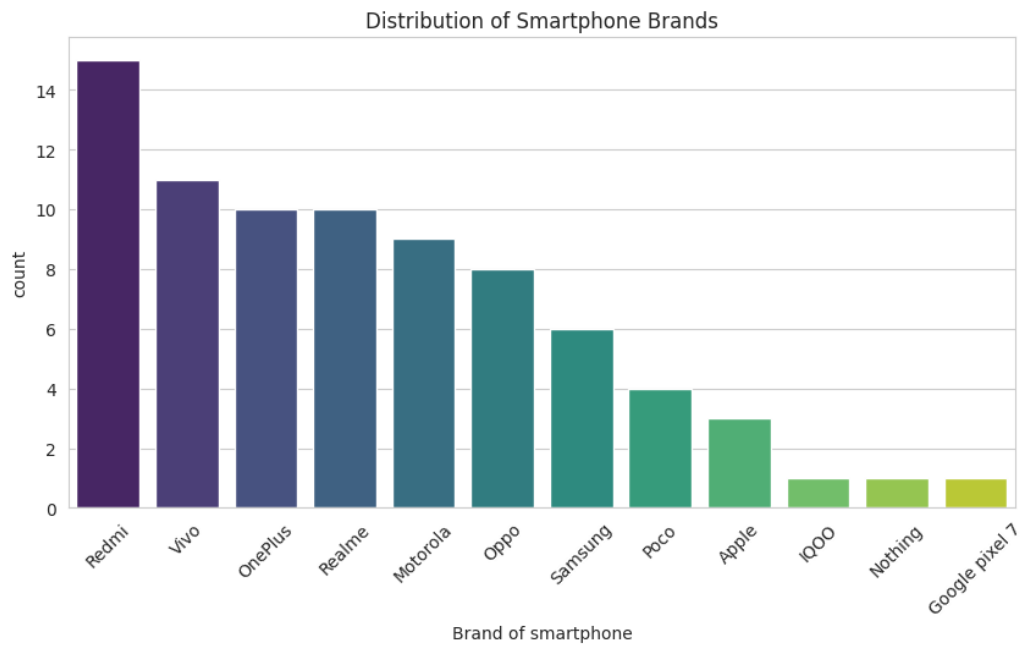


Figure 7:

We can see redmi is the most popular while google pixel 7 being the least. Below is boxplot of smratphone price by brand

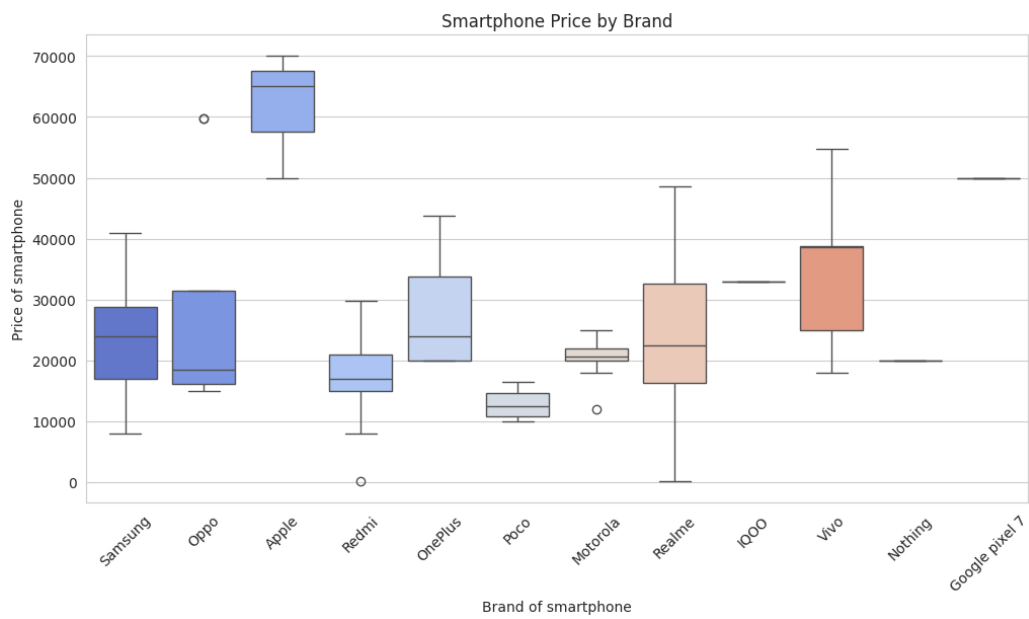


Figure 8:

We can see apple devices costs the most.  
Below is the k means prediction.

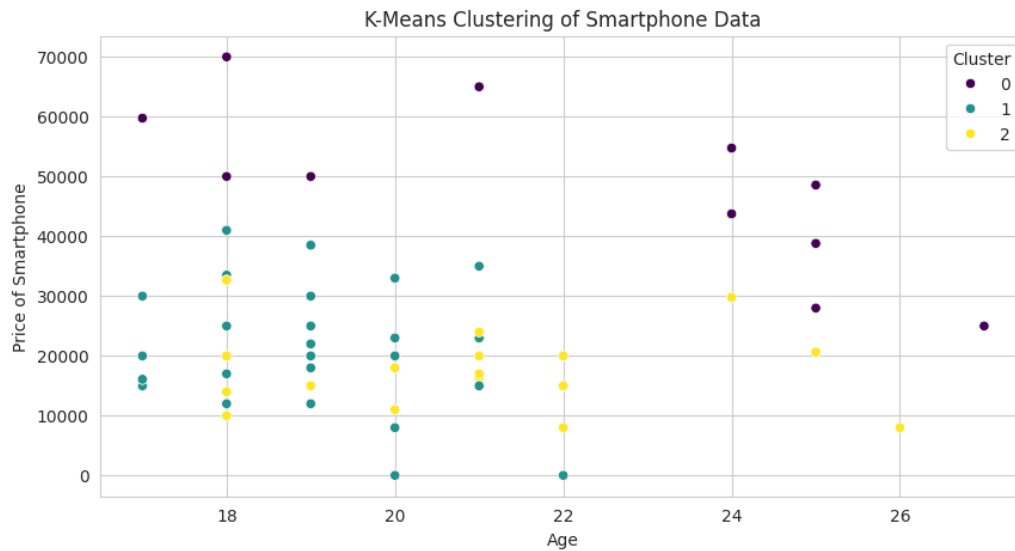


Figure 9:

## 8 Conclusion

This analysis provides insights into smartphone ownership trends and includes predictive models for estimating smartphone prices and clustering users. The cleaned dataset is now more reliable for further analysis, such as segmentation and recommendation systems.

The colab link is as follows:-[https://colab.research.google.com/drive/1znPiNzke3HE6qmIK\\_59K\\_](https://colab.research.google.com/drive/1znPiNzke3HE6qmIK_59K_)  
*sharing*