# Paper Solution Midterm

## Question No 1:

1. Data science helps organizations make data-driven decisions rather than relying on guesswork. It involves analyzing large datasets to uncover hidden patterns, trends, and insights that improve efficiency and accuracy.

**Healthcare:** Data science aids in disease prediction, personalized medicine, and medical image analysis. For example, hospitals use machine learning to predict patient readmissions and optimize treatment plans.

**Finance:** Banks use data science for fraud detection, credit scoring, and risk management. For example, algorithms analyze transaction data to detect suspicious activities in real time.

1. Typical Steps in the Data Science Process

**Data Collection**: Gathering relevant data from sources like databases, sensors, or web APIs. Ensures enough quality information for analysis.

**Data Cleaning:** Removing errors, duplicates, or missing values. Important for reliability and accuracy.

**Data Exploration (EDA):** Visualizing and summarizing data to understand patterns and relationships. Helps in selecting the right features and models.

**Modeling:** Applying machine learning or statistical techniques to make predictions or classifications. Transforms insights into actionable results.

**Evaluation & Deployment:** Testing model performance and implementing it in real systems. Ensures the solution works effectively in real-world conditions.

## Question No 2:

1. salary <- dataset$Salary

avg\_salary <- mean(salary, na.rm = TRUE)

print(paste("Average Salary:", avg\_salary))

1. ages <- dataset$Age

min\_age <- min(ages, na.rm = TRUE)

max\_age <- max(ages, na.rm = TRUE)

print(paste("Minimum Age:", min\_age))

print(paste("Maximum Age:", max\_age))

1. employee <- list(

Name = "Ali Khan",

Department = "Finance",

Age = 30,

Salary = 85000

)

print(employee$Name)

print(employee$Department)

print(employee$Age)

print(employee$Salary)

1. Lists can hold multiple data types (numeric, character, logical, etc.), while vectors can only contain elements of the same type.

Lists can store complex data structures, such as data frames or other lists.

Therefore, lists are more flexible and suitable for storing detailed records, while vectors are used for simple, uniform data.

1. mean\_salary <- mean(dataset$Salary, na.rm = TRUE)

sd\_salary <- sd(dataset$Salary, na.rm = TRUE)

correlation <- cor(dataset$Salary, dataset$Age, use = "complete.obs")

print(paste("Mean Salary:", mean\_salary))

print(paste("Standard Deviation of Salary:", sd\_salary))

print(paste("Correlation between Salary and Age:", correlation))

## Question No 3:





