M T W T F S S
Page No. 1
Date:
YOUVA

Assignment - 1

1) What is Business Report? What are the main characteristics of a Business Report?

A business report is a concise, structured document that presents data, analysis, and recommendations regarding a specific business issue or project, aimed at informing decision-making and facilitating organizational performance.

Main characteristics of Business Report →

1.) Objective-driven
2.) Structured format
3.) Concise and focused
4.) Data-driven
5.) Audience-oriented
6.) Professionally presented
7.) Actionable recommendations
8.) Timeliness.

2.) What is Business Analytics? What is the need of BA? How it is helpful in any Organization elaborate with an example.

Business data Analytics takes that idea, but puts it in the content of business insight, often with prebuilt business content and tools that expedite the analysis process. In other words, data analytics is more of a general description of the mordern analytics process.

Need of BA →

1) Enhance Customer experience
2) Make informed decisions
3) Reduce employee turnover
4) Improving Efficiency
5) Identify frauds
6) Improved Advertising
7) Make the most of Your Investment
8) Better Product Management

3) How many data Structure R has?

The most essential data stuructars used in R include →

1) Vector
2) List
3) Dataframe
4) Array
5) Matrix
6) Factor

4) What are R packages?

R packages are extensions to the R Statistical programming language. R packages contain code, data, and documentation in a stadardised collection format that can be installed by users of R, typically via a centralised software Repository such as →

CRAN (Comprehensive R Archive Network)

5) How can you load a CSV file in R?

5) Elaborate the following R objects. a) Vector b) data-frame c) Matrix d) list

a) Vector → One-dimensional Array holding elements of the same data type, created using 'c()'. They are commonly used for storing and manipulating data elements, performing mathematical operations, and indexing subset of data.

b) Data-frame → Two-dimensional structure resembling a Table, with rows and columns for storing datasets. They can be created using the 'data.frame()' function.

c) Matrix → Two-dimensional array containing elements of the same data type, used for mathematical operation. Matrix can be created using matrix() function.

d) list → Versatile data structure holding elements of different type, useful for organizing heterogeneous data of object. They are created using the 'list()' function.

6) Explain different Matrix operations function in R?

- %*% operators for matrix multiplication
- Solve() function compute the inverse of a square matrix.
- t() function is used to transpose a matrix, swapping the rows and columns.

- dim() and dimnames() functions
dim() functions return the dimensions of a matrix as a numeric vector, where first element represent the number of ~~column~~ rows and second tells number of columns.
dimnames() function return the names of the rows and columns of a matrix as a list.
- cbind() and rbind() function - cbind is used to add the column in the matrix and rbind() is used to add the rows in the matrix.
- diag() function - To create the diagonal elements of a matrix.
- det() function - Use to calculate the determinant of a square matrix.
- sum(), colSums(), and rowSums() function.
- Mean() colMeans, and rowMeans() function.

7.) What are the data structures in R that is used to perform statistical analyses and create graph?
1.) Vectors → One-dimensional array used for storing numeric, characters, logical, or factor data.
2.) Dataframe → Two-dimensional structures similar to tables, where rows represent observations and columns represent variable.
3.) List → Versatile data structures capable of holding elements of different types including vectors matrices, data frames, or even other lists.
4.) Matrices → Two-dimensional Array.
5.) factors → special type of vector used to represent categorial data.
6.) Array - Multidimensional extensions of vectors useful for handling data with more than two dimensions.

For graph creation, then primary data structure include-

1) Vector
2) Dataframe
3) Matrices
4) List

8.) Define mean, median, mode and standard deviation.

Meadian → The middle value of a sorted dataset, separating the higher half from the lower half. Obtained using the 'median()' function.

Mean → The arithmetic average of a set of values. Calculated using the 'mean()' function.

Mode → The value that appears most frequently in a dataset. Not directly available in base R, but can be computed using functions from external packages or custom functions.

Standardard Deviation → A measure of the dispersion or spread of a dataset around its mean. Computed using the 'sd()' function.

9.) Briefly explain the below term structured Data?

Structured data refers to data that is organised and stored in a predefined format or schema, making it easily searchable, accessible, and analyzable. It follows a consistent and predefined structure, typically

In tables with rows and columns. Structured data is highly organized and has a fixed schema, which allows for efficient storage, retrieval, and analysis using relational database or spreadsheet applications.

10.) What about Arithmetic and Boolean operators in R programming?

Arithmetic Operators →

1) Addition (+)
2) Substraction (-)
3) Multiplication (*)
4) Division (/)
5) Exponentiation (^)
6.) Modulo (%%)
7.) Integer Division (%/%)

Boolean Operators →

1) Logical AND (&&)
2) Logical OR (||)
3) Logical Not (!)
4.) Equal to (==)
5.) Not Equal to (!=)
6.) Greater than (>)
7.) Less than (<)
8.) Greater than or equal to (>=)
9.) Less than or equal to (<=).

11.) find median and mode of the following numbers : 12, 13, 11, 10, 9, 11, 7, 10, 15, 16, 11.

Code →

```
numbers → c (12,13,11,10,9,11,7,10,15,16,11)
median_value ← median (numbers)
mode_value ← names (table (numbers)) [table (
                numbers) == max (table [numbers])]
print (median_value)
print (mode_value)
```

12.) Explain the importance of dataframe ?

Importance of dataframe as follow as-

1.) Tabular Representation
2.) Data management
3.) Statisatical Analysis
4.) Graphical Visualization
5.) Data Integration
6.) Interoperability

13.) Write the syntax of if-else in R ?

Syntax →

```
if (condition) {
    # code to be executed if the condition is
      True
} else {
    # code to be executed if the condition is
      False
```

14.) Explain how to save the data in R.

1.) CSV (Comma-Separated Values):
write.csv(data_frame, "filename.csv", row.names=FALSE)

2.) Excel (require 'writexl' or 'xlsx' package):
writexl::write_xlsx(data_frame, "filename.xlsx")
xlsx::write.xlsx(data_frame, "filename.xlsx")

3.) RData (binary format to save R objects):
save(data_frame, file = "filename.RData")

4.) RDS (R Data Serialization) (binary format to save a single R object):
saveRDS(data_frame, file = "filename.rds")

15.) Explain types of business analytics with necessary examples. ?

1.) Descriptive Analytics → Descriptive analytics focuses on summarizing historical data to gain insights into past performance and understand trends and patterns.

2.) Predictive Analytics → Predictive analytics involves forecasting future outcomes or trends based on historical data and statistical models.

3.) Prescriptive Analytics → Prescriptive analytics aims to provide recommendations or

actionable insights to optimize decision-making and improve business outcomes.

16.) What is Social Media Analytics? Explain the role of analytics in facebook?

It refers to the process of collecting and analyzing data from social media platforms to gain insights into user behaviour, preferences and trends.

Here are some role analytics in facebook

1.) Audience Insights
2.) Content Performance
3.) Ad. Campaign Optimization
4.) Social listening and sentimental Analysis
5.) CRM (Custom Relationship Management).

17.) Define Various types of digital data and also discuss the sources of these digital data.

Various types of digital data include:

1.) Text-Data - Textual information such as social media posts, emails, website content, and customer reviews.
2.) Numerical Data - Numeric information such as sales figures, financial data, and numerical measurements from sensors or devices.
3.) Audio Data - Sound recordings, podcasts, voice notes, and other audio files.
4.) Image Data - Visual content such as photographs, graphics, charts, and diagrams.
5.) Video Data - Moving visual content including videos, animations, and live streams.

6.) Geospatial Data- Location-based information obtained from GPS, mapping services, and geotagged content.

7.) Time Series Data- Data collected over time, often used for trend analysis, forecasting and monitoring changes.

8.) Web Data- Information gathered from web sources including webpages.

Source of digital data include:

1.) Websites and Blogs- Content published on websites, blogs, and online forums.

18.) Discuss the importance of business analytics?

Business analytics is crucial for-

1.) Understanding customer behaviour
2.) Optimizing operations
3.) Enhancing Marketing effectiveness
4.) Mitigating risks
5.) Improving financial performances
6.) Driving innovation
7.) Enhancing decision-making with data-driven insights.

19.) Display 1 to 30 number in R.

```
number <- 1:30
print (number)
```

20.) Explain different data structures in R. Display an 5 cities names in R using objects.

Different data structures in R include:

1.) Vectors: A one-dimensional array that can hold elements of the same data types.

2.) Matrices: A two-dimensional array with rows and columns, where all elements are of the same data type.

3.) Lists: A collection of objects of different lengths and data types.

4.) Data Frames: A two-dimensional tabular data structure with rows and columns.

5.) Factors: A special type of vector used to represent categorical data with predefined levels.

6.) Arrays: Multidimensional generalization of vectors and matrices.

7.) Data Tables: An extension of data frames, offering additional functionality and performance improvements for large datasets.

8.) Tibbles - Modern data frames with enhanced features for easier data manipulation and printing.

To display 5 cities names

```
cities <- c("New York", "London", "Tokyo", "Paris", "Sydney")
print(cities)
```

21.) How you Apply Vector? Use dataframes in R. Also discuss the process to use data from an external file in R.

Applying Vector -

Creating Vectors:

You can create vectors using the 'c()' function or by using sequences like 'seq()' or 'rep()'.

```
numeric_vector <- c(1,2,3,4,5)
character_vector <- c("a", "b", "c", "d", "e")
result_vector <- numeric_vector + 2
```

Using Data frame -

Creating Data Frames:

You can create data frames using the 'data.frame()' function.

Code -
```
data <- data.frame(
    ID = c(1,2,3,4,5),
    name = c("John", "Jane", "Doe", "Alice",
             "Bob"),
    Age = c(25, 30, 35, 40, 45) )
```

Accessing Data in Data frame:

```
names <- data$Name
ages <- data[, "Age"]
```

Using Data in from an External file:

Reading Data from a csv file

You can use 'read.csv()' function

```
data <- read.csv("file.csv")
```

Writing Data to External files:
You can write data frames to external files using functions like 'write.csv()', write_excel_csv()'

Code -

```
write.csv (data, "file.csv")
library (writexl)
write_xlsx (data, "file.xlsx")
```

22.) Explain different types of operations in R.

1.) Arithmetic Operators: Arithmetic operators in R are used to perform mathematical calculations on numeric values R provide several arithmetic operators that allow you to add, subtract, multiply, divide, raise to a power, and find the remainder.

2.) Assignment Operators: In R assignment operators is used to assign values to variables equal sign (=) or less than and minus signs (<-).

3.) ~~Relational~~ Logical Operators: ~~Relational~~ Logical operators in R are used to combine or manipulate logical values (True or False).

4.) ~~Logical~~ Relational Operators: ~~Logical~~ Relational operators in R are used to ~~combine~~ compare values and determine relationship b/w them.

5) Membership Operators: In R, membership operators are used to check if elements are present in a vector or not R provide two membership operators %in% and %!in%

6) Bitwise Operators: Bitwise operators in R are used to perform bitwise operations on binary representations of integers.

23) Write about Data frame? Write about operations on data frame.

A dataframe is a fundamental data structure in R that represents a two dimensional tabular data object. A data frame can contain different types of data, such as numeric, character, logical, or factors, and it allows for easy manipulations and analysis of data.

Operations on a data frame involve accessing and manipulating it is contents some common operations on a data frame include:

1) Accessing data
2) filtering data
3) Adding Columns
4) Aggregating data
5) Sorting data.

24) What is R? State the features of R language.

R is a programming language and software environment designed for statistical computing and graphics. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland Newzealand, in the early 1990s. It's extensive functionality, rich set of packages, and active community make it a popular choice for anyone working with data, from statisticians and researchers to data scientists and analysts.

features of R →

- Open-source nature
- Extensive repository of packages
- Rich statistical capabilities.
- Powerful graphics and visualization tools.
- Robust data handling and manipulating functions.
- Support for reproducible research through scripting.

25) What is the difference b/w Matrix and Dataframes.

	Matrix	Dataframes
1) Structures	Homogeneous	Heterogeneous
2) Dimension	fixed	flexible
3) Column name	No	Yes
4) Data types	Same for all elements	Can vary by Columns
5) Usage	Mathematical operation Linear Algebra	Can vary by Columns.

6.) functionality	Limited	Data Analysis, Manipulations, Modelling Extensive (data manipulation, subsetting, merging.

26.) How will you merge two dataframes in R programming language?

In R, You can merge two data frames using various functions and methods, depending on the type of merge options you want to perform. Here are some commonly used method for merging dataframes in R

1) 'merge()' function → The merge function is a versatile method for merging data frames based on common columns

ii) 'rbind()' function → The rbind() function can be used to merge two data frames vertically by appending rows from one data frame to another.

iii) 'bind_rows()' function (from 'dplyr' package): The bind_rows() function from the dplyr package is another option for vertically merging dataframes. It can handle dataframes

with different column names and order.

iv) 'cbind()' function: The cbind() functions can be used to merge data frames horizontally by combining columns from different dataframes.

27.) Explain the importance of dataframes in R?

Dataframes are fundamental data structure in R that play a crucial role in data Analysis and manipulation. Here are some key reasons why data frames are important in R.

1) Tabular Representation
2) Heterogenous data
3) Data cleaning or transformation
4) Data analysis and statistics.
5) Data manipulation and subsetting
6) Integration with Graphical libraries
7) Compatibility with other R packages.

28.) Write a R program to create a data frame from two given vectors like Names:

Radha	Kavita	Golu	Ravi
24	14	19.2	54.5

Code ⇒

```
#Vectors
vec1 <- c ("Radha", "kavita", "Golu",
           "Ravi")
vec2 <- c (24, 14, 19.2, 54.5)

#Creating Data frame
data <- data.frame (Name = vec1, Age =
                    vec2)

data
```

Output ⇒

	Name	Age
1	Radha	24
2	kavita	14
3	Golu	19.2
4	Ravi	54.5

8/9/2024

29) How can you load a .csv file in R?

In R, you can load a csv (comma-separated values) file using the read.csv() function. Here's an example of how to do it:

```
# Set the file path to your csv file
setwd("path of the file")

# Load the csv file
data <- read.csv("file name")

# Print the csv file
print(data)
```

30) How will you identify and treat the missing value and outlier data in R?

Identifying Missing Values: i) The is.na() function can be used to identify missing values in a dataset. It returns a logical vector indicating which values are missing (TRUE) and which are not (FALSE).

ii) You can also use functions like complete.cases() or na.omit() to identify and remove rows with missing values.

Treating Missing Values: i) Deleting Missing Values
ii) Imputing Missing Values

Identifying Outliers: i) Boxplots
ii) Summary Statistics

Treating Outliers : i) Removing Outliers
ii) Winsorization
iii) Transformation
iv) Robust Statistical Methods

31.) What are the different ways to read the dataset? How to create and rename a variable in R? What are the read write method available in R and explain?

1.) Reading Data:
- Use 'read.csv()' for csv file.
- Use 'read.table()' for general tabular data with customizable delimiters.
- read-delim()' is specifically for tab-delimited files.
- The readers package offers faster alternatives like 'read_csv()' for csv files and similar functions for delimited files.

2.) Creating and Renaming Variables:
- Create variables by assigning values with '<-' or '='.
- Rename variables in data frame using 'names()'.

3.) Reading / Write Method.
- read.csv()' / write.csv()
- read.table()/ write.table().
- read.delim'()/ write.delim()

32.) How to create a Bar Chart & Histogram?

```
#Creating a Bar Chart
data <- c(10, 20, 30, 40)
names <- c("A", "B", "C", "D")

# Using Barplot() function

barplot(data, names.arg = names, xlab = "Categories",
        ylab = "Values", main = "Bar Chart")
```

```
#Creating Histogram
data <- rnorm(100, mean = 0, sd = 1)
hist(data, xlab = "Values", ylab = "Frequency", main =
     "Histogram")
```

33.) How to create a Scatter plot & a Box plot?

```
#Creating a Scatter plot
x <- c(1, 2, 3, 4, 5)
y <- c(3, 5, 7, 9, 11)
plot(x, y, xlab = "X-axis", ylab = "Y-axis",
     main = "Scatter plot")
```

```
#Creating Box plot

x<- rnorm (100, mean=0, sd=1)

boxplot(x, main ="Boxplot")
```

34.) Generalize the graphical analysis in data analysis? List the various plots in R and explain in detail.

Graphical analysis is a fundamental aspect of data analysis, as it allows us to visually explore and understand the patterns, relationships and distribution within our data.

Here are some commonly used plots in R and their explanations.

i.) Scatter plot :- A scatter plot is used to visualize the relationship between two continuous variables. Each data point is represented as a dot on the plot, with one variables on the axis and the other on the y-axis.

ii.) Line plot :- A line plot is used to display the trend or progression of a variable over time or any ordered category. It is useful for visualizing temporal data or ordered data point.

iii) Bar plot :- A bar plot is used to compare and display categorical data.

iv.) Histogram:- A histogram is used to visualize the distribution of a continous variables. It displays the frequency or proportion of data falling within specified intervals or bins.

v.) Box-plot:- A box-plot, also known as a box-and whisker plot, provides a summary of the distribution of a continous variable.

35.) Write a R script to create a line graph.

Here's an example of an R script to create a line graph using the built-in dataset "Air-Passangers":

```
#Load the AirPassangers dataset
data(AirPassangers)
ts data <- ts(AirPassangers, start = c(1949, 1), frequency = 12)

plot(tsdata, type = "l", xlab = "Year", ylab = "Passengers",
     main = "Airpassangers")

Grid()
```

36.) Explain the importance of data Visulization techniques?

Importance of Data Visulization techniques-

i) Rich Visulization Capabilities
ii) Seamless Integration with Data Analysis

iii) EDA (Exploratory Data Analysis
iv) Effective Communication and presentation
v) Interactivity and Dynamic Visualization
vi) Reproductivity and Documentation
vii) Integration with statistical Modelling and Machine learning.

37.) How can business managers use the Descriptive, predictive, and prescriptive Analytics for making better business decision? Discuss with examples.

Descriptive Analytics helps managers understand what happened in the past, like sales trends, Predictive analytics forecast future outcomes, like predicting future sales based on past data. Prescriptive analytics suggests actions to take, like adjusting marketing strategies based on predicted sales. For examples, a retail manager might use descriptive analytics to analyze past sales, predictive analytics to analyze past sales, predictive analytics to forecast future sales, and prescriptive analytics to decide on promotional strategies to optimize sales based on those forecast.

38) Briefly Explain the term Marketing Analysis.

Marketing analysis is the process of assessing market trends, customer preferences, competitor decisions. It involves gathering and interpreting data to understand consumer behaviour and market dynamics, helping businesses develop effective marketing strategies.

39.) Briefly explain the term Structured Data.

Structured data refers to organized information that is typically stored in databases or spreadsheets with a clear and defined format. It is easily searchable, sortable, and can be analyzed using algorithms or software. Examples include tables of numerical data, such as sales figures, customer demographics or inventory levels.

40.) Write navigation to apply scatter plot, boxplot, and histogram in R.

* Load the require packages
* prepare the data

Scatter plot:-
- Use the plot() function
- Specify the variables
- Customize the plot
- Display the plot

Box plot:-
- Use the ~~list~~ → funct boxplot() function
- Customize the plot by adding lables
- Use the plot() function to display the note.

Histogram:-
- Use the hist() function
- Customize the plot by adding lables
- Use the plot() function to display the note.

41.) Explain Data Visualization: which Different types of charts are used for qualitative data, quantitative data as well as data visualization for text data?

Data Visualization is the process of representing data and information visually using charts, graphs, maps and other visual elements.

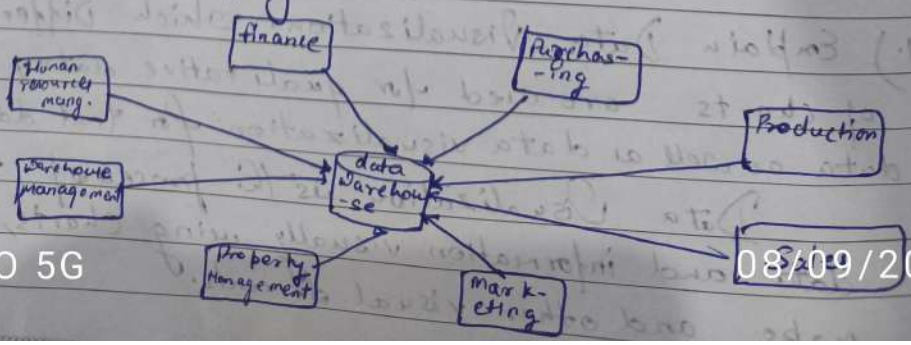
Qualitive Data Visualization:
i) Bar chart
ii) Pie chart.
iii) Staked Bar chart

Quantitative Data Visualization:
i) Line chart
ii) Scatter plot
iii) Histogram
iv) Box plot

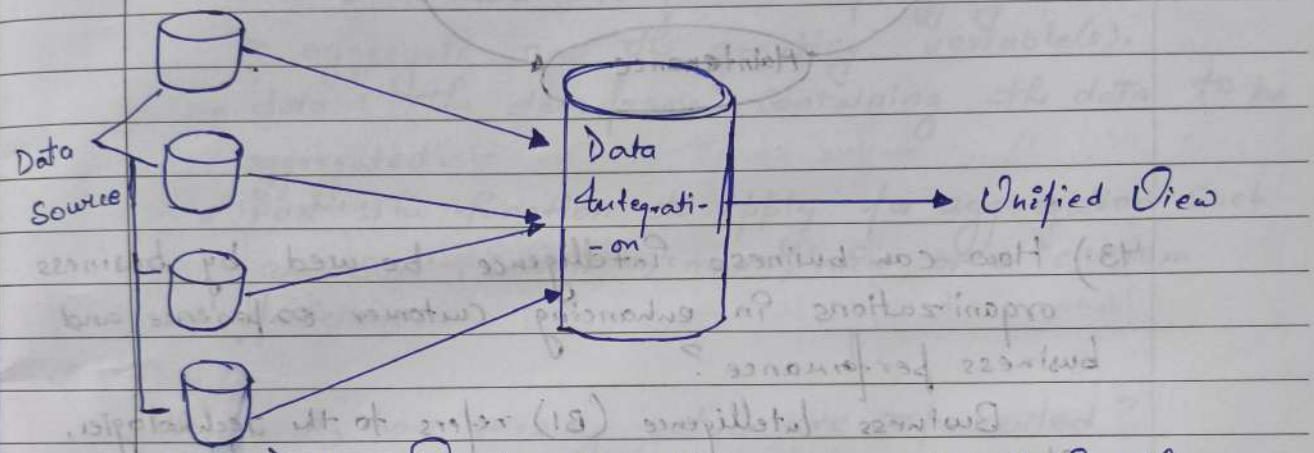
Text Data Visualization:
i) Word cloud
ii) Heatmap
iii) Network Diagram

42.) Describe in detail the different components of business analytics with its conceptual diagram.

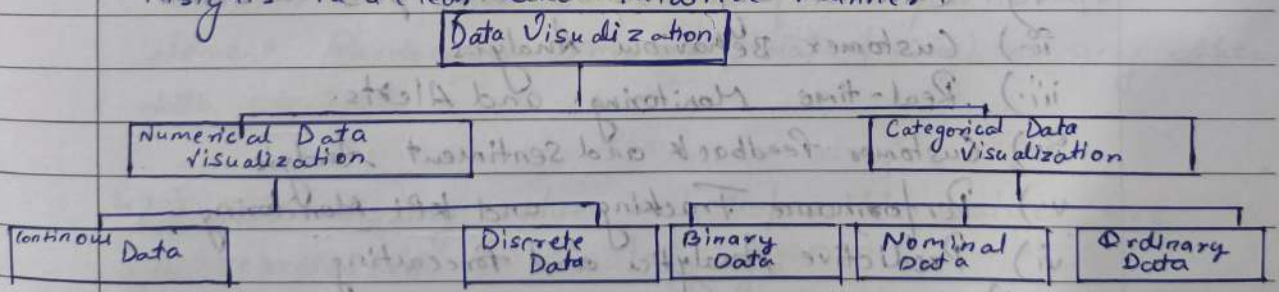
i) Data Sources:- The foundation of business analytics lies in the data. Data sources can include internal organizational data, such as customer records, sales transactions, financial data, operational data, and employee data.



ii.) Data Integration: In the data integration stage, data from various sources is collected, combined, and transformed into a unified format suitable for analysis. This involves task like data cleaning, data transformation, and data enrichment to ensure, to ensure data quality and consistency.

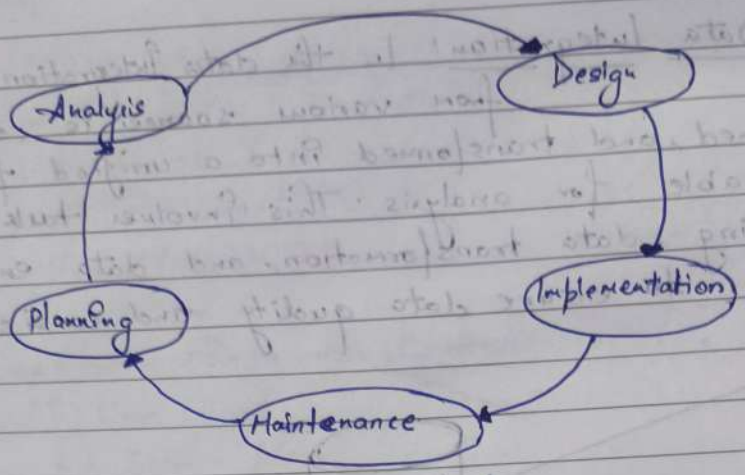


iii) Data Visualization play a crucial role in this stages, where charts, graphs, and visual representation are used to communicate patterns, trends, and insights in a clear and intuitive manner.



iv) Data Modeling – It is the process of analyzing and defining all the different data your business collects and produces, as well as the relationship b/w those bits of data.

Diagram



43.) How can business intelligence be used by business organizations in enhancing customer experience and business performance?

Business Intelligence (BI) refers to the technologies, strategies, and practices used by organizations to analyze business information and gain insights for making informed decisions.

i.) Customer Segmentation and personalization
ii.) Customer Behaviour Analysis
iii.) Real-time Monitoring and Alerts.
iv.) Customer feedback and sentiment Analysis
v.) Performance Tracking and KPI Monitoring
vi) Predictive Analytics and forecasting
vii) Competitive Analysis.

44.) What is R?

R is a programming language and and Software environment designed for Statistical computing and graphics. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, in the early 1990s.

45.) Explain how data is Aggregated in R?

The aggregate() function is a base R function that allows you to aggregate data based on grouping variables. It takes the following general form:

aggregate(formula, data, FUN)

- 'formula': A formula specifying the variables to aggregate and the grouping variable(s).
- 'data': The data frame containing the data to be aggregated.
- 'FUN': The function to apply for aggregation such as 'sum', 'mean', 'max', 'min' or a custom function.

46.) In R, how missing values are represented?

In R, missing values are represented by the special value NA, which stands for "Not Available". The NA value is used to indicate the absence or unknown value for a particular element in a vector, matrix, data frame, or any other data structure.

47.) How missing values and impossible values are represented in R language?

In R, missing values are represented by the special value NA, as mentioned in the previous response. NA is used to indicate the absence or unknown value for a particular element.

48.) State the features of R language.

Some key features of R language

i) Extensive Statistical Capabilities
ii) Data Manipulation and Transformation
iii) Graphics and Data Visualization
iv) Reproducibility and Documentation
v) Open Source and Community-driven

49.) Draw a pie chart for the following data.

Sections	I	II	III	IV	V
No. of Workers	210	390	180	75	145

```
# define the data
Section <- c("I", "II", "III", "IV", "V")
workers <- c(210, 390, 180, 75, 145)
Colors <- c("red", "blue", "green", "yellow", "orange")

pie (worker, label = section, col = colors)

title ("Number of Workers by Section")

legend ("topright", section, fill = Color, title = "Sections")
```

50.) Create a data frame with a = c(1,2,2), b = c(4,5,6), c(7,8,9) and find the value of the following.

I How do I select the c(4,5,6)?
II How do I select the 1?
III How do I select the 5?
IV what is df[,3]?

v. What is d[3,3]?
vi. What is d[2,2]?

```
# Create a data frame
d1 <- data.frame(a = c(1,2,3), b = c(4,5,6), c = c(7,8,9))
# I. Select c(4,5,6)
   selected_value1 <- d1[,"b"]

# II Select 1

   selected_value2 <- d1$a[1]

# III  Select 5

   Selected_value3 <- d1$b[2]

# IV d1[,3] ?
   Column_3 <- d1[,3]

# V. d1[1,]
  row_1 <- d1[1,]

# VI  d1[2,2]
  value_2_2 <- d1[2,2]
```

I. To select the value c(4,5,6).
II. To select the value 1 from a
III. To select the value 5 from b
IV. Select the third column of the data frame. It return vector of value (7,8,9)
V. Select the first row of the data frame. It return data. 1,4,7
VI. Select value in second row and II column. It return value 5.