Business Intelligence

Q1. a) What are the important BI reporting practices ?

- Reporting is a report authoring tool that professional and developers can use to build ~~the~~ use to build different types of reports using multiple database

- Types of Reports :-
  1. List
  2. Crosstabs
  3. Map
  4. Charts

① List :-
- List is used for showing detailed information from your database.
- eg - Item list, Customer list
- Data in list is shown in rows and columns.
- Each column shows all the values for a data item in the database.
- Different operations that can be performed on the list
  a) set list properties
  b) Hide columns in list reports
  c) Use repeaters.

② Crosstabs :-
- Crosstab reports are also known as matrix reports.

- They are used to show relationships between three or more query items.
- Data is shown in rows and columns with information summarized at the intersection point.

③ Chart :-
- Charts are used in presenting the data in a way that is useful to the end users
- Many different types of charts are available like pie, bar, line etc.
- More than one chart type can be used within a chart which is known as combination chart.

Q1. b) Discuss the importance of drill up, drill-down, and drill-through capabilities in report preparation.

- While working with dimensionally modelled relation data sources, reports can be created which will allow the reader to view more general information with a dimensional hierarchy.

1. Drill-Down –
- Allows users to move from a high-level summary to more detailed data, helping in root cause analysis.

2. Drill-up –
- Help in viewing aggregated insights by rolling up data for a broader perspective.

3. Drill - Through -
- Enables users to navigate from one report / dashboard to another related reports with more context.
- Using drill through one can move from one report to another within a session while maintaining focus on same data.

Q1. C) Explain with examples the use of Data Grouping and sorting, Filtering is important in BI Reports.

① Data Grouping :-
- Data grouping can be used to group records.
- It creates certain fields on certain criteria to make the report easier to read.
- Grouping allows to separate group of records.
- Groups related data to provide a structured view, making reports easier to analyze.

② Sorting :-
- Sorting data in reports can be done in two ways.
- Firstly sort the data source object itself and add groups to the report and specify how each group should be sorted using the group by and sort properties
- Arrange data in ascending or decending order for better understanding.

③ Filtering :-
- Filtering is useful in simplifying large amount of data and only displaying data what the users really need to see.
- Filters ensure that the reports contains data data only to specific to business query.
- To retrieve the desired data it is important to design the filter correctly.
- Hidden filters can be enables for additional filters.

Q3. a) Need for data pre-processing and two-techniques used.

- Need for Data Pre-Processing :-
- Data pre-processing is an essential step in data analysis and business intelligence because raw data is often incomplete, inconsistent, or noisy.
- Pre-processing ensures that the data used for decision making is clean.
- Need for pre-processing -
① Handling Missing values
② Fixing Inconsistencies.
③ Enhancing Data Qualities
④ Improving Effeciency.

Data Pre-Processing techniques :-

① Data cleaning -
- The process of identifying and correcting errors

in the dataset.
- Filling missing values using mean, median or mode.
- Removing duplicate or unnecessary records.

2. Data Normalization :-
- A technique used to scale numeric data within a specific range to ensure uniformity
- It uses methods like -
  - min - max scaling : Converts data to a 0 - 1 range.
  - Z - score normalization : Transforms data using means and standard deviation.

Q3. b) What is data transformation ? Why it is needed ? Explain at least 3 techniques.

- Data transformation is the process of converting data from one format, structure or value set into another to make it more suitable for analysis.
- Need for data transformation -
  1. Ensures consistency
  2. Data compatability
  3. Organizing

- Techniques for data transformation.

1. Aggregation -
- Summarizes details data into higher - level information.
- eg - Sales records for individual days can be aggregated to show total amount or yearly sales.

② Discretization -

- Converts continuous numeric data into categorical values
- A dataset with customer ages ranging from 18 to 65 can be discretized into categories:
  - 18 - 25 (Young Adults)
  - 26 - 40 (Middle Age)
  - 41 - 65 (Seniors)

③ Encoding -

- Converts categorical data into numerical format for analysis.
- In a dataset containing "Gender" as Male / Female, label encoding can assign Male = 1, Female = 0 making it easier for machine learning models to process.

Q3. c) What is data reduction? Explain Dimensional -ity Reduction and Data compression.

- Data reduction minimizes the volume of data while preserving essential information.
- The helps in improving storage efficiency, reducing compution time, and enhancing analytical performance.

① Dimensionality :-

- Reducing the number of features / attributes in a dataset while keeping the most relevant information.

- Too many variables can lead to overfitting and slow processing.
- Helps in visualizing high-dimensional data by reducing it to a smaller number of key factors.

② Data compression :-
- Reducing the size of data storage while maintaining essential details.
- Large datasets require significant storage and processing power.

Q5. a) What is logistic regression? Discuss the type of logistic regression.

- Logistic Regression is a statistical technique used for classification problems.
- Unlike linear regression, which predicts continuous values, logistic regression estimates the probability that a given input belongs to a specific category.
- The output is mapped values between 0 and 1.
- It is simple and efficient.
- Widely used in business., Helthcare.

Types of Logistic Regression -

① Binary Logistic Regression -

- Used when the target variables has only two possible outcome.

Formula -

$$P(Y=1) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n)}}$$

② Multinomial Logistic Regression :-

- Used when there are three or more unordered categories in the dependent variables.
- Unlike binary logistic regression, this regression calculates the probability of each category independently.

Q5. b) How the classification and clustering are different. Discuss use with example.

Classification	Clustering
1) Supervised learning technique that assigns labels to data	1) Unsupervised learning technique that groups data based on similarities.
2) Classification requires labeled data for training	2) Clustering works with unlabeled data.
3) It categorizes data into predefined data groups.	3) It identifies the hidden patterns and structures.
4) It uses algorithms like - Decision - tree, Neural Network.	4) It uses algorithms such as - K-Mean, Hierarchical clustering.

5) ex :- classifying emails as spam & not spam

5) eg :- Grouping customers based on purchasing behaviour into different segment.

Q5. c) What is decision tree ? Explain with a case study.

- A decision tree is a machine learning algorithm used for classification and regression.
- It splits data into smaller sub-groups using decision rules, forming a tree-like structure.
- Each node represents a decision, and each branch leads to an outcome.

- Key components of decision tree -

1. Root Node
2. Decision Nodes
3. Leaf Nodes.

- Case study :-

Scenario -

A bank wants to automate loan approvals based on income, credit score, and existing loans.