Data Science & Artificial Intelligence

Warehousing

Part- 02

By- Vishvadeep Gothi sir

Topic : Datawarehousing

#Q. In a data cube, what is true about measures and their evaluation?

- Identify the correct statements.

A A data cube measure is a numeric function.

B Measures can be evaluated at each point in the data cube space.

C Measured values are computed by aggregating data corresponding to dimension-value pairs.

D Measures can only be evaluated for a single dimension-value pair.

Topic : Datawarehousing

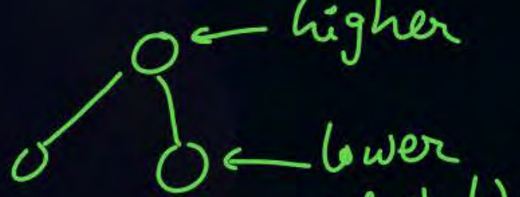
#Q. Consider the following about roll-up operations for data aggregation:

S1 : Roll-up operation aggregates data to a higher level of details.

S2: Roll up operation performed by dimension reduction; dimension are not removed from cube.

↳ reducing no. of columns

Which of the above is/are correct?



A Only S1

B Only S1 and S2

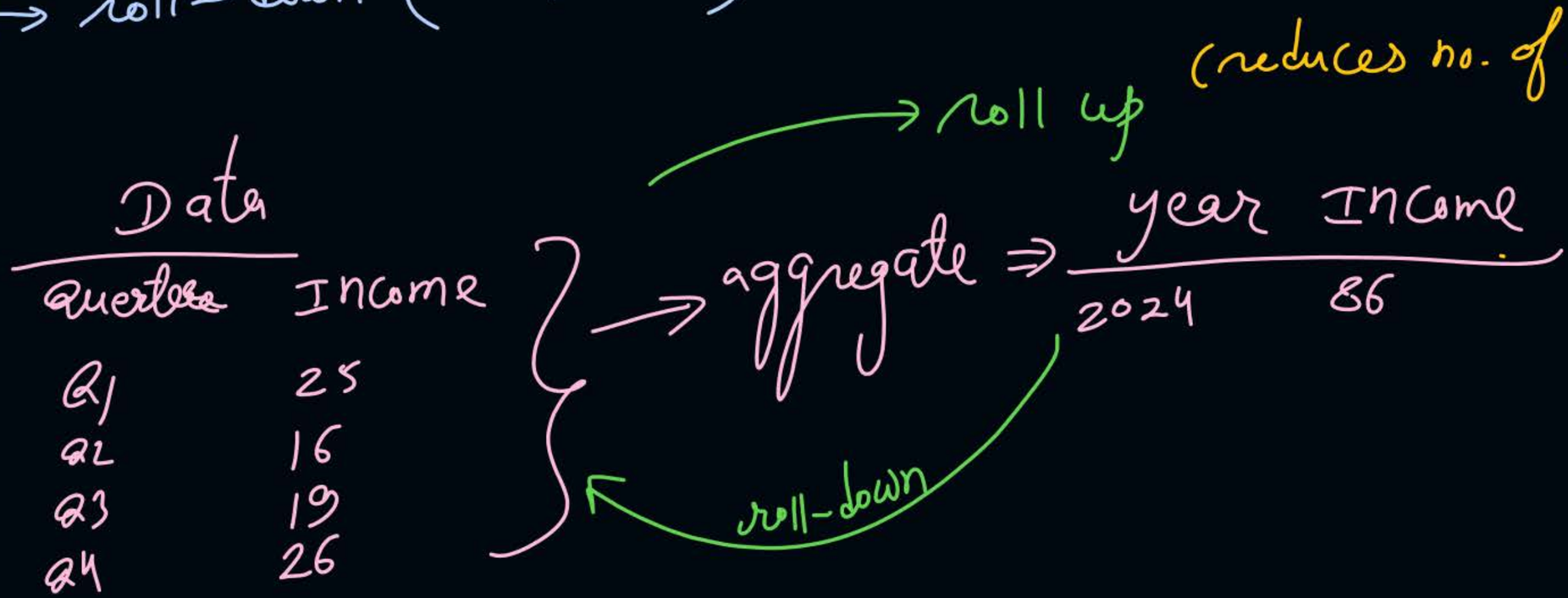
C Only S2

D None of these

aggregation

→ roll-up (drill up)

→ roll-down (drill down)



Topic : Datawarehousing

Ans = 0.218

#Q. Consider a dataset with income value as an attribute in rupee format:
10000, 12000, 14000, 16000, 18000, 20000, 22000 and 24000
What is the Z-score normalized value of 18000 ?

$$\mu = \frac{10000 + 12000 + 14000 + 16000 + 18000 + 20000 + 22000 + 24000}{8}$$

$$= 17000$$

Topic : Data warehousing

μ = mean

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

S.D.

$$z_i = \frac{x_i - \mu}{\sigma}$$

Values	$(x_i - \mu)$	$(x_i - \mu)^2$	Z
10000	−7000	49000000	
12000	−5000	25000000	
14000	−3000	9000000	
16000	−1000	1000000	
18000	+1000	1000000	
20000	3000	9000000	
22000	5000	25000000	
24000	7000	49000000	

σ = 4582.5757

$$Z_{18000} = \frac{18000 - 17000}{4582.5757} = 0.218$$

Topic : Datawarehousing

#Q. Which of the following statements is/are correct?

S1 : Some OLAP multi-dimensional database modeling tools are optimized for snowflake schema.

S2: In snowflake schema data is not redundant as dimensions are normalized.

A Only S1

B Only S2

C Both S1 and S2

D None of the above

Topic : Datawarehousing

warehouse

#Q. Consider the following statements regarding OLAP (Online Analytical Processing) and OLTP (Online Transaction Processing) in data warehousing:

OLTP → on dbms

S1 : OLAP is used for decision support and data analysis based on historical data.

S2: OLTP ensures data integrity and consistency across operational systems.

S3: OLAP databases are optimized for fast transaction processing and real-time updates.

S4: OLTP databases are typically denormalized to improve query performance.

Which of the following is/are true?

A ✓ S1 and S2

B S1 and S4

C S2 and S3

D S1,S2, and S3

Topic : Datawarehousing

Ans = 0.625

#Q. Consider a data set with "income" values as an attribute 50000, 60000, 70000, 80000, 90000, and 98000. To normalize the values the min-max normalization is used. Value normalized value of the "income" 80000 is ___?

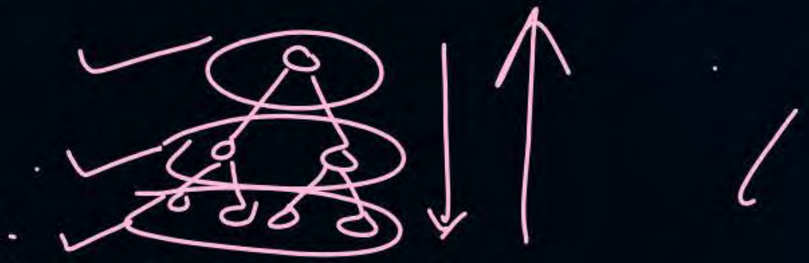
$$x' = \frac{x - min}{max - min}$$

$$x' = \frac{80000 - 50000}{98000 - 50000}$$

$$= 0.625$$

Topic : Datawarehousing

#Q. Consider the following statement about hierarchies in data warehouse: Which of the above is correct?



A A roll-up operation moves from a higher level of to a lower level.

B Hierarchies facilitate efficient data navigation and analysis at different aggregation levels.

C Star schema do not support hierarchical data structure.

D Hierarchical structures can improve query performance in data warehouse.

Topic : Datawarehousing

#Q. Which of the following statements is/are false?

S1: A snowflake schema generally provides better performance for complex queries due to its normalized structure.

S2: In a star schema, dimension tables are often highly normalized.

A. Only S1

B. Only S2

C. Both S1 and S2

D. None of the above

Topic : Datawarehousing

#Q. Which of the following statements is/are correct?

S1: Roll-up is an OLAP operation that reduces the dimensionality of data.

S2: Drill-down operation can be used to decrease the number of dimensions in an OLAP cube.

false

false

A Only S1

B Only S2

C Both S1 and S2

D Neither S1 or S2

Topic : Datawarehousing

#Q. In binning, which of the following methods replaces each value in a bin with the bin's mean value?

A Mean binning

B Median binning

C Equal-width binning

D Frequency binning

Topic : Datawarehousing

#Q. Which of the following is an accurate description of the equal-width binning method in discretization?

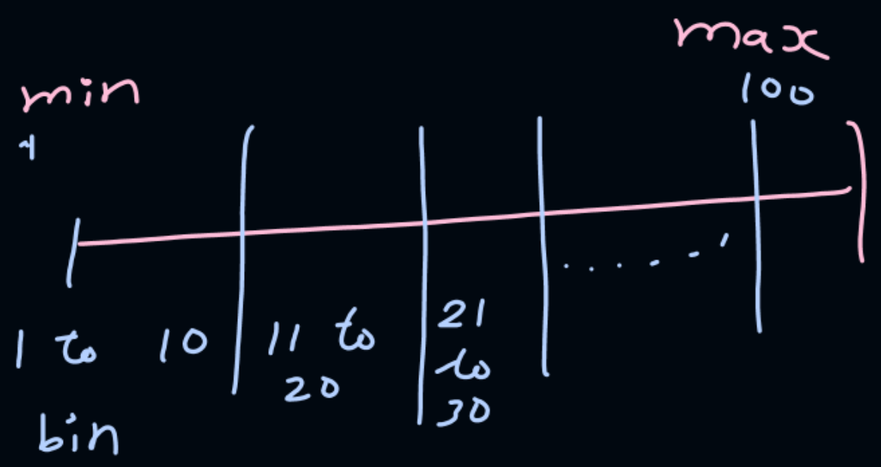
A. The range of data is divided into intervals of equal size, and each interval is replaced by it's representative

B. Data is grouped based on frequency, ensuring each bin contains an equal number of data points.

C. Data points are assigned to bins based on their proximity to the bin's mean value.

D. The data range is split into bins based on predefined thresholds that vary in width.

equal width binning



equal frequency binning

each bin contains equal no. of values

Topic : Datawarehousing

#Q. Which of the following statements is/are correct?

S1 : Roll-up operation reduces data detail by collapsing lower-level attributes into higher-level summaries.

S2 : Drill-down decreases data granularity, providing less detailed views of the data.

A Only S1

B Only S2

C Both S1 and S2

D None of these

Topic : Datawarehousing

#Q. Data warehousing has various characteristics including:

(A) Focuses on modelling and analysis of data relating to a specific area

(B) Data warehouse is an integration of data from various systems like CRM system, SCM system, etc

(C) Time-variant when it records the timestamp of data, allowing it to capture historical changes over time

(D) It is stored permanently i.e data once stored can not be updated

(E) It is stored temporarily i.e data once stored can be updated

Choose the most appropriate answer from the options given below:

A (B), (D), (E) only

B (A), (B), (C), (D) only

C (A), (C), (D) only

D None

Topic : Datawarehousing

#Q. Which of the following is NOT a part of data warehouse architecture?

A Data Staging

B Data Presentation

C Data Collection

D Data Integration

Topic : Datawarehousing

#Q. A fact table typically contains:

A Descriptive attributes

B Measures for analysis

C Hierarchical data

D Metadata

Topic : Datawarehousing

#Q. The main difference between a data warehouse and a database is:

A Data warehouses are used for transaction processing.

B Databases use OLAP for queries.

C Data warehouses are optimized for read-heavy operations.

D Data warehouses are normalized.

Topic : Datawarehousing

#Q. Which of the following OLAP operations allows you to view data from different perspectives?

A Slice

B Dice

C Roll-up

D Pivot

Topic : Datawarehousing

#Q. Which of the following is a key challenge in data integration?

A Increased storage capacity

B Different data formats

C High-speed processors

D Centralized management

Topic : Datawarehousing

#Q. What is metadata?

A Data about data

B Raw data

C Processed data

D Analytical data

Topic : Datawarehousing

#Q. Normalization is typically used to:

A ✓ Reduce redundancy in data

B Increase data retrieval speed

C Simplify the data model

D Enhance data integrity

Topic : Datawarehousing

#Q. Which of the following is NOT a benefit of data warehousing?

A Improved data quality

B Better decision-making

C Increased operational cost

D Historical analysis

Topic : Datawarehousing

#Q. In a data warehouse, the process of validating and cleansing data is important because:

A. It increases storage requirements

B. It reduces data redundancy

C. It ensures data quality ✓

D. It simplifies data modelling

Topic : Datawarehousing

#Q. What is the primary goal of a Data Warehouse?

A Transaction processing

B Analytical reporting

C Data entry

D Data storage

Topic : Datawarehousing

#Q. Which of the following describes OLAP (Online Analytical Processing)?

A. Supports high-volume transaction processing → OLTP

B. Optimized for data retrieval and analysis

C. Focuses on data entry and modification

D. Uses a flat file structure

Topic : Datawarehousing

#Q. What does the "Extract" step in ETL refer to?

A. Transforming the data into a usable format → Transform

B. Loading data into the data warehouse → Load

C. Collecting data from various sources

D. Cleansing the data

Topic : Datawarehousing

#Q. In which schema is data stored in a highly normalized form?

A Star Schema

B Snowflake Schema

C Galaxy Schema

D Fact Constellation

THANK - YOU