Session 15 - Advanced Constructs - CTEs & Views

24 September 2024 18:10

Agenda DSML Aug24 Beginner 2

1. Problem Statement
2. Ad-hoc Reporting
3. CTEs & their Advantages
4. Break
5. Views
6. When to use CTE vs View
7. Practice Question

Problem Statement

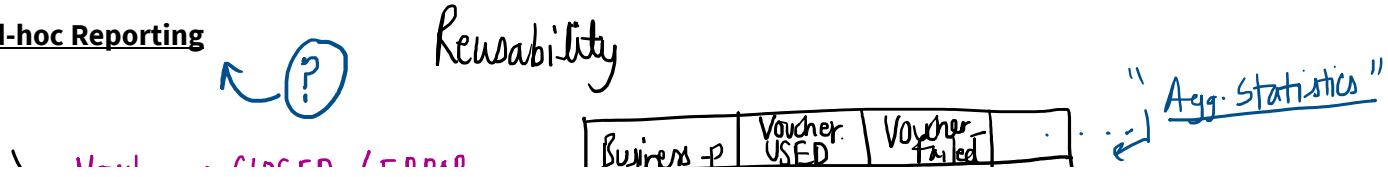
RelianceGuard Insurance is a leading provider of insurance policies, committed to offering exceptional coverage and services to its diverse customer base. However, the company has recently observed a decline in customer retention rates. To address this issue, RelianceGuard Insurance aims to analyze its customer and policy data to gain insights into customer behaviour and identify factors influencing policy renewals. By leveraging this analysis, the company seeks to ultimately improve its overall customer retention rates and maintain a competitive edge in the insurance market.

Data Dictionary:

- **Customer:** Customer ID number
- **Customer_Lifetime_Value:** Customer's total worth to the business over the life of the relationship
- **Response:** True or False response to a renewal offer
- **Coverage:** Type of policy (Basic, Extended, Premium)
- **Monthly_Premium_Auto:** Amount of customer's monthly insurance payments
- **Months_Since_Last_Claim:** Number of months between the customer's last reported insurance claim
- **Months_Since_Policy_Inception:** Number of months since the customer began an insurance policy
- **Number_of_Open_Complaints:** Number of unresolved customer complaints
- **Number_of_Policies:** Number of policies the customer currently owns
- **Policy_Type:** (Corporate Auto, Personal Auto, Special Auto)
- **Policy:** 3 levels (L1, L2, L3) per Policy Type (Corporate, Personal, Special)
- **Renew_Offer_Type:** 4 types of renewal offers (Offer 1, Offer 2, Offer 3, Offer 4)
- **Sales_Channel:** Channels to purchase a policy (Agent, Branch, Call Center, Web)
- **Total_Claim_Amount:** Cumulative amount of claims since policy inception

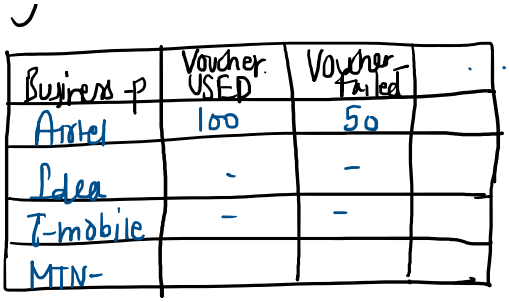
Dataset link: LINK

Ad-hoc Reporting



↘ Voucher ⇒ CLOSED / ERROR

↘ Airtel, Idea, MTN

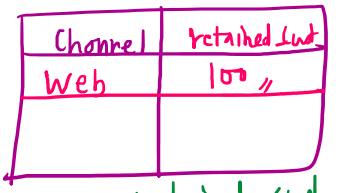


Q. Analyse customer retention by sales channel to identify which channels are most effective at retaining customers.

↑ group by

Step 1: Filter out the data that we need.

```
SELECT Sales_Channel,
    SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) AS
retained_customers
from `rg.relianceguard_insurance`
group by Sales_Channel;
```



retained cust.

100 / (500) = 20%

↖ total Cust.

Step 2: Calculate retention rates.

retained_cust. total Count.

```
SELECT Sales_Channel,
    (SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) * 1.0 / count(*)) AS
retained_ratio
from `rg.relianceguard_insurance`
group by Sales_Channel;
```

	2023	2024
Bajaj	900	300
Reliance	500 ↑	800 ↑

Step 3: Extract the top Sales channel.

```
SELECT Sales_Channel,
    (SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) * 1.0 / count(*)) AS
retained_ratio
from `rg.relianceguard_insurance`
group by Sales_Channel
order by retained_ratio desc
limit 1;
```

Common Table Expressions (CTEs)

Goal : store queries (and the results of queries) for reuse in reports and other analyses.

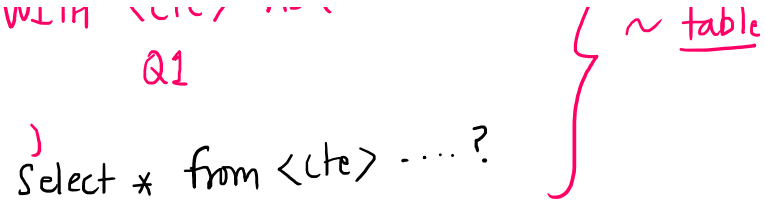
↘ 1. CTE (Commn table Expression) ✓

2. Views ✓

WITH <cte> AS (Alias

Q1

} ~ table



QUIZ:

CTEs can only be referenced once in a single SQL query. "False"

CTE Usecase & Advantages

Question 1: Among RelianceGuard Insurance's various sales channels, which channel boasts the highest customer retention rate?

```
WITH sales_by_channel AS (
    SELECT  Sales_Channel,
            count(*) AS total_customers,
            SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) AS
retained_customers
    from `rg.relianceguard_insurance`
    group by Sales_Channel
),
retention_rates AS (
    SELECT Sales_Channel,
           (retained_customers * 1.0 / total_customers) * 100 AS retention_rates
    from sales_by_channel
)
select * from retention_rates
order by retention_rates desc
limit 1;
```

Recursive CTE ?

Break till : 22:28

Homework: - Provided at the end.

Question 2: Write a query to analyze customer profiles and renewal offer types to identify the most receptive customer segments for each offer.

Step 1: Calculate response rates for each combination of renewal offer type and customer profile attributes

Renewal_Offer_type
State,
Gender,
Education,
Employment_Status,
Marital_Status,
Vehicle_Class,
Vehicle_Size,

Step 2: Rank these combinations by response rate.

Step 3: Select the top combinations with the highest response rates.

SOLUTION:

Views = Another approach to CTE. ✓

Query → View ≈ "table"

Views ~ A bit slower in execution time.

QUIZ:

Queries that reference views typically execute faster than queries that directly access tables

View ≈ table
(re-usable Query)
)

false

Question: RelianceGuard Insurance is concerned about potential customer churn and wants to identify customers who might be at risk of not renewing their policies. To achieve this, write a query to analyze customer data from the past month (as of March 1st, 2024).

CUSTOMER
Low Engagement - Months_Since_Last_Claim
Complaints - Number_of_Open_Complaints

prakash, -, - 0
Saurabh 1
Sanket 0
Kiran 1
→ Sum(Response) =

Table
↳ Data

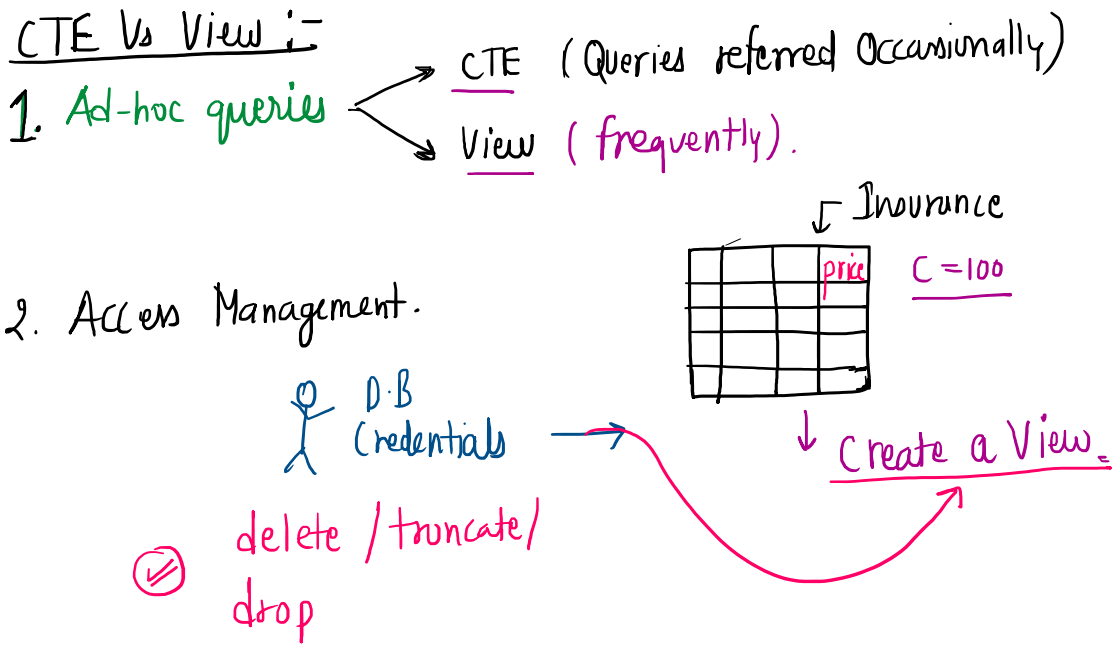
View
↳ Query
=

QUIZ:

What is the primary difference between a database view and a table in SQL?

Query ↑ view
↑ data

CTE Vs View :-



Homework:

Question 2: Write a query to analyze customer profiles and renewal offer types to identify the most receptive customer segments for each offer.

Step 1: Calculate response rates for each combination of renewal offer type and customer profile attributes

Renew_Offer_type
State,
Gender,
Education,
Employment_Status,
Marital_Status,
Vehicle_Class,
Vehicle_Size,

Step 2: Rank these combinations by response rate.

Step 3: Select the top combinations with the highest response rates.

SOLUTION:

```
WITH Response_Rates_CTE AS (
    SELECT
        Renew_Offer_Type,
        State,
        Gender,
        Education,
        Employment_Status,
        Marital_Status,
        Vehicle_Class,
        Vehicle_Size,
        COUNT(*) AS Total_Customers,
        SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) AS Total_Accepted,
        (SUM(CASE WHEN Response = true THEN 1 ELSE 0 END) * 1.0 / COUNT(*)) AS Response_Rate
    FROM `rg.relianceguard_insurance`
    GROUP BY
        Renew_Offer_Type,
        State,
        Gender,
        Education,
        Employment_Status,
```

```
        Marital_Status,
        Vehicle_Class,
        Vehicle_Size
),
Ranked_Response_Rates_CTE AS (
    SELECT
        Renew_Offer_Type,
        State,
        Gender,
        Education,
        Employment_Status,
        Marital_Status,
        Vehicle_Class,
        Vehicle_Size,
        Response_Rate,
        DENSE_RANK() OVER (PARTITION BY Renew_Offer_Type ORDER BY Response_Rate DESC) AS Response_Rank
    FROM Response_Rates_CTE
)
SELECT  Renew_Offer_Type,
        State,
        Gender,
        Education,
        Employment_Status,
        Marital_Status,
        Vehicle_Class,
        Vehicle_Size,
        Response_Rate
  FROM Ranked_Response_Rates_CTE
  WHERE Response_Rank = 1;
```