Case Study #4 - Data Bank

A. Customer Nodes Exploration

1. How many unique nodes are there on the Data Bank system?

```
SELECT
    COUNT(DISTINCT node_id) as unique_nodes_count
FROM customer_nodes
```

2. What is the number of nodes per region?

```
SELECT
    cn.region_id,
    region_name,
    COUNT(DISTINCT node_id)
FROM customer_nodes AS cn
JOIN regions AS r ON cn.region_id = r.region_id
GROUP BY cn.region_id, region_name
```

3. How many customers are allocated to each region?

```
SELECT
    region_id,
    COUNT(*)
FROM customer_nodes
GROUP BY region_id
ORDER BY region_id
```

4. How many days on average are customers reallocated to a different node?

```
SELECT ROUND(AVG(avg_day)) as avg_node_reallocation_days
FROM (
    SELECT
```

```
        SUM(TIMESTAMPDIFF(DAY,start_date,end_date)) as avg_day
    FROM customer_nodes
    WHERE YEAR(end_date) != 9999
    GROUP BY customer_id,node_id) as sub
```

5. What is the median, 80th and 95th percentile for this same reallocation days metric for each region?

```
CREATE VIEW avg_day_reallocation AS (
SELECT
        region_id,
        SUM(TIMESTAMPDIFF(DAY,start_date,end_date)) as avg_day
    FROM customer_nodes
    WHERE YEAR(end_date) != 9999
    GROUP BY region_id,customer_id,node_id
);
SELECT MAX(avg_day) as median
FROM
    (SELECT
        avg_day,
        NTILE(4) OVER(ORDER BY avg_day) AS quartile
    FROM avg_day_reallocation) as sub
WHERE quartile = 2;

SELECT MAX(avg_day) as percentile_80
FROM
    (SELECT
        avg_day,
        NTILE(5) OVER(ORDER BY avg_day) AS quartile
    FROM avg_day_reallocation) as sub
WHERE quartile = 4;

SELECT MAX(avg_day) as percentile_95
FROM
    (SELECT
        avg_day,
        NTILE(20) OVER(ORDER BY avg_day) AS quartile
    FROM avg_day_reallocation) as sub
WHERE quartile = 19;
```

B. Customer Transactions

1. What is the unique count and total amount for each transaction type?

```
SELECT
    txn_type,
    COUNT(*),
    SUM(txn_amount)
FROM customer_transactions
GROUP BY txn_type
```

2. What is the average total historical deposit counts and amounts for all customers?

```
SELECT
    ROUND(AVG(txn_count)) as avg_txn_count,
    ROUND(AVG(avg_amount_each_customer)) as avg_amount
FROM
    (SELECT
        txn_type,
        COUNT(*) as txn_count,
        AVG(txn_amount) as avg_amount_each_customer
    FROM customer_transactions
    WHERE txn_type = 'deposit'
    GROUP BY customer_id) as sub
```

3. For each month - how many Data Bank customers make more than 1 deposit and either 1 purchase or 1 withdrawal in a single month?

```
SELECT
    monthview,
    COUNT(DISTINCT customer_id) as customer_count
FROM
    (SELECT
        MONTH(txn_date) as monthview,
        customer_id,
        SUM(IF(txn_type = 'deposit', 1,0)) as deposit_count,
        SUM(IF(txn_type != 'deposit', 1,0)) as purchase_or_witho
    FROM customer_transactions
    GROUP BY monthview, customer_id) as sub
WHERE deposit_count > 1 and purchase_or_withdrawal_count >=1
GROUP BY monthview
```

4. What is the closing balance for each customer at the end of the month? Also show the change in balance each month in the same table output.

```
SELECT
    *,
    SUM(total_month_change) OVER (
        PARTITION BY customer_id
        ORDER BY last_day_of_month
    ) AS ending_balance
FROM
    (SELECT
        customer_id,
        LAST_DAY(txn_date) as last_day_of_month,
        SUM(CASE
            WHEN txn_type = 'deposit' THEN txn_amount
            ELSE -txn_amount END) as total_month_change
    FROM customer_transactions
    WHERE customer_id = 3
    GROUP BY customer_id, last_day_of_month
    ORDER BY customer_id, last_day_of_month
    ) as sub
```

5. Comparing the closing balance of a customer’s first month and the closing balance from their second nth, what percentage of customers:

- **What percentage of customers have a negative first month balance? What percentage of customers have a positive first month balance?**

```
CREATE VIEW customer_monthly_balances AS
(
    SELECT
    *,
    SUM(total_month_change) OVER (
        PARTITION BY customer_id
        ORDER BY last_day_of_month
    ) AS ending_balance,
    ROW_NUMBER() OVER(PARTITION BY customer_id ORDER BY last_day
    FROM
        (SELECT
            customer_id,
            LAST_DAY(txn_date) as last_day_of_month,
            SUM(CASE
                WHEN txn_type = 'deposit' THEN txn_amount
                ELSE -txn_amount END) as total_month_change
        FROM customer_transactions
        GROUP BY customer_id, last_day_of_month
        ORDER BY customer_id, last_day_of_month
        ) as sub
);


SELECT
    ROUND(COUNT(*)/(SELECT COUNT(DISTINCT customer_id) FROM cust
    100-ROUND(COUNT(*)/(SELECT COUNT(DISTINCT customer_id) FROM
FROM customer_monthly_balances
WHERE ending_balance > 0 AND sequence = 1
```

- **What percentage of customers increase their opening month’s positive closing balance by more than 5% in the following month?**

```
SELECT
    SUM(IF(ROUND((following_balance - ending_balance)/ending_bal
    (SELECT COUNT(DISTINCT customer_id) FROM customer_monthly_ba
FROM
    (SELECT
        *,
        LEAD(ending_balance) OVER(PARTITION BY customer_id) as
    FROM customer_monthly_balances)  as sub
WHERE sequence = 1
```

- **What percentage of customers reduce their opening month's positive closing balance by more than 5% in the following month?**

```
SELECT
    SUM(IF(ROUND((following_balance - ending_balance)/ending_bal
    (SELECT COUNT(DISTINCT customer_id) FROM customer_monthly_ba
FROM
    (SELECT
        *,
        LEAD(ending_balance) OVER(PARTITION BY customer_id) as
    FROM customer_monthly_balances)  as sub
WHERE sequence = 1
```

- **What percentage of customers move from a positive balance in the first month to a negative balance in the second month?**

```
SELECT
    ROUND(COUNT(*) / (SELECT COUNT(DISTINCT customer_id) FROM cu
FROM
    (SELECT
        *,
        LEAD(ending_balance) OVER(PARTITION BY customer_id) AS
    FROM customer_monthly_balances)  AS sub
WHERE sequence = 1 AND ending_balance > 0 AND following_balance
```