**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\_withdrawal\_count

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\_of\_month) AS sequence

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 customer\_monthly\_balances) \* 100,1) AS positive\_percentage,

100-ROUND(COUNT(\*)/(SELECT COUNT(DISTINCT customer\_id) FROM customer\_monthly\_balances) \* 100,1) AS negative\_percentage

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\_balance \* 100) > 5.0,1,0))/

(SELECT COUNT(DISTINCT customer\_id) FROM customer\_monthly\_balances) \* 100 AS percentage

FROM

(SELECT

\*,

LEAD(ending\_balance) OVER(PARTITION BY customer\_id) as following\_balance

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\_balance \* 100) < 5.0,1,0))/

(SELECT COUNT(DISTINCT customer\_id) FROM customer\_monthly\_balances) \* 100 AS percentage

FROM

(SELECT

\*,

LEAD(ending\_balance) OVER(PARTITION BY customer\_id) as following\_balance

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 customer\_monthly\_balances) \* 100,1) AS percentage

FROM

(SELECT

\*,

LEAD(ending\_balance) OVER(PARTITION BY customer\_id) AS following\_balance

FROM customer\_monthly\_balances) AS sub

WHERE sequence = 1 AND ending\_balance > 0 AND following\_balance < 0