SQL CASE STUDY ON DATA BANK

Table 1: Regions

region_id	region_name
1	Africa
2	America
3	Asia
4	Europe
5	Oceania

Table 2: Customer Nodes

customer_id	region_id	node_id	start_date	end_date
1	3	4	2020-01-02	2020-01-03
2	3	5	2020-01-03	2020-01-17
3	5	4	2020-01-27	2020-02-18
4	5	4	2020-01-07	2020-01-19
5	3	3	2020-01-15	2020-01-23
6	1	1	2020-01-11	2020-02-06
7	2	5	2020-01-20	2020-02-04
8	1	2	2020-01-15	2020-01-28
9	4	5	2020-01-21	2020-01-25
10	3	4	2020-01-13	2020-01-14

Table 3: Customer Transactions

customer_id	txn_date	txn_type	txn_amount
429	2020-01-21	deposit	82
155	2020-01-10	deposit	712
398	2020-01-01	deposit	196
255	2020-01-14	deposit	563
185	2020-01-29	deposit	626
309	2020-01-13	deposit	995
312	2020-01-20	deposit	485
376	2020-01-03	deposit	706
188	2020-01-13	deposit	601
138	2020-01-11	deposit	520

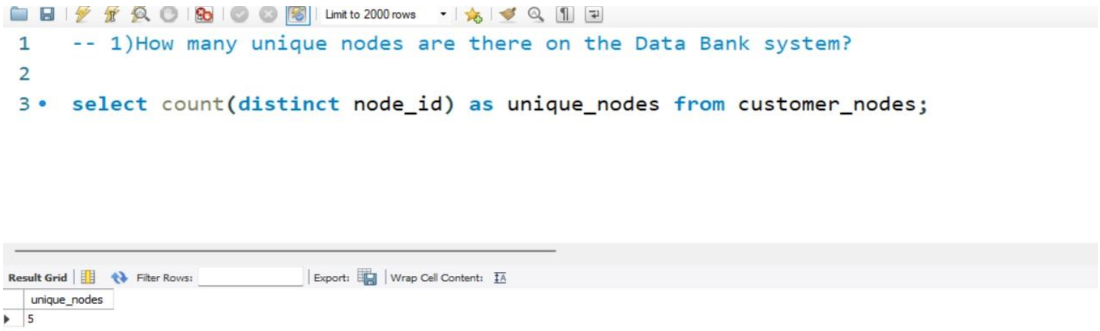
A. Customer Nodes Exploration

1. How many unique nodes are there on the Data Bank system?
2. What is the number of nodes per region?
3. How many customers are allocated to each region?
4. How many days on average are customers reallocated to a different node?
5. What is the median, 80th and 95th percentile for this same reallocation days metric for each region?

B. Customer Transactions

1. What is the unique count and total amount for each transaction type?
2. What is the average total historical deposit counts and amounts for all customers?
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?

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

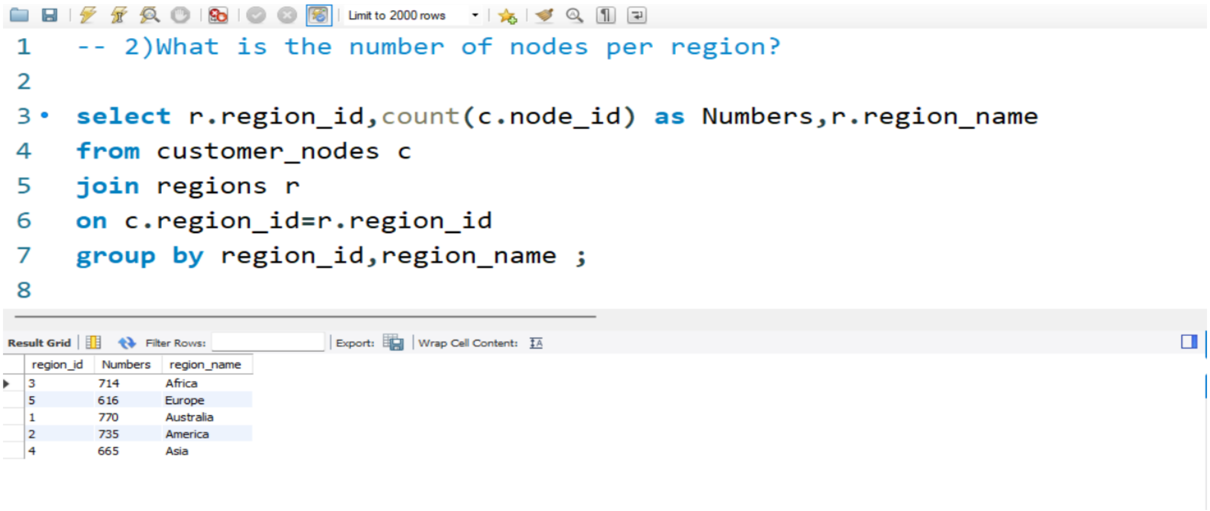


```sql
-- 1)How many unique nodes are there on the Data Bank system?

select count(distinct node_id) as unique_nodes from customer_nodes;
```

unique_nodes
5

2)What is the number of nodes per region?

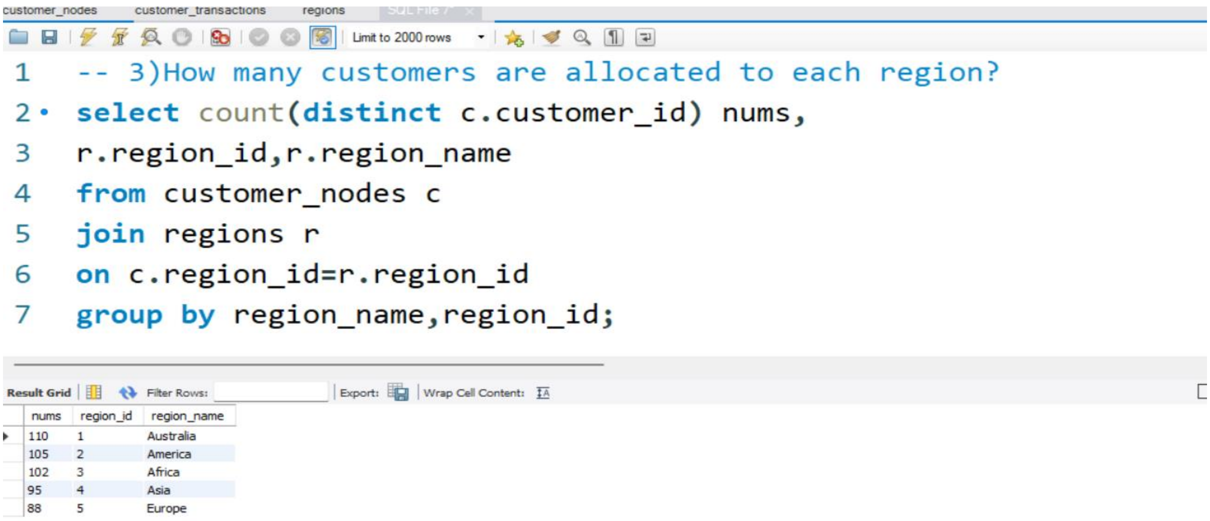


```sql
-- 2)What is the number of nodes per region?

select r.region_id,count(c.node_id) as Numbers,r.region_name
from customer_nodes c
join regions r
on c.region_id=r.region_id
group by region_id,region_name ;
```

region_id	Numbers	region_name
3	714	Africa
5	616	Europe
1	770	Australia
2	735	America
4	665	Asia

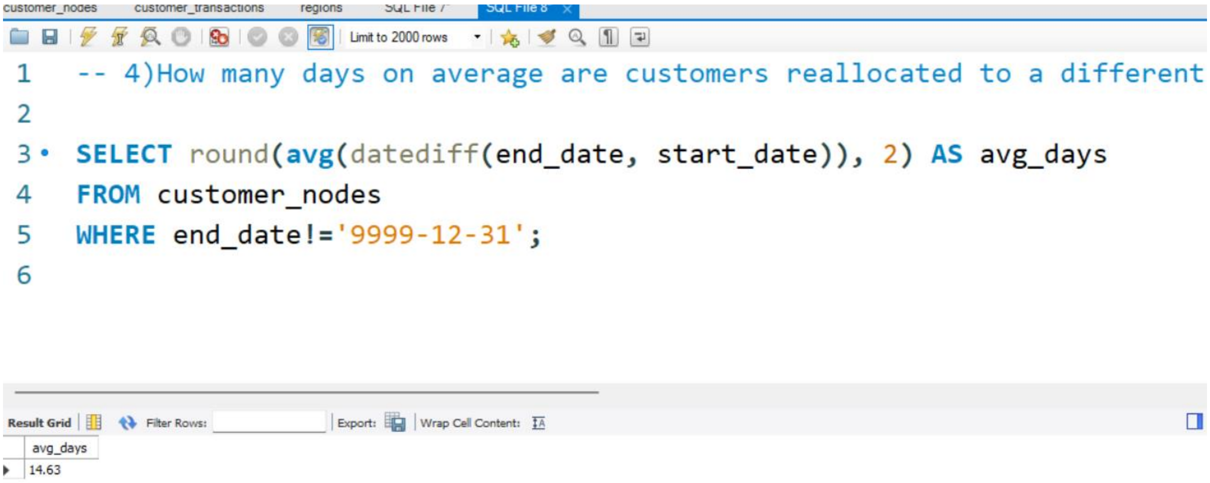
3)How many customers are allocated to each region?



```sql
-- 3)How many customers are allocated to each region?
select count(distinct c.customer_id) nums,
r.region_id,r.region_name
from customer_nodes c
join regions r
on c.region_id=r.region_id
group by region_name,region_id;
```

nums	region_id	region_name
110	1	Australia
105	2	America
102	3	Africa
95	4	Asia
88	5	Europe

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



```
-- 4)How many days on average are customers reallocated to a different

SELECT round(avg(datediff(end_date, start_date)), 2) AS avg_days
FROM customer_nodes
WHERE end_date!='9999-12-31';

```

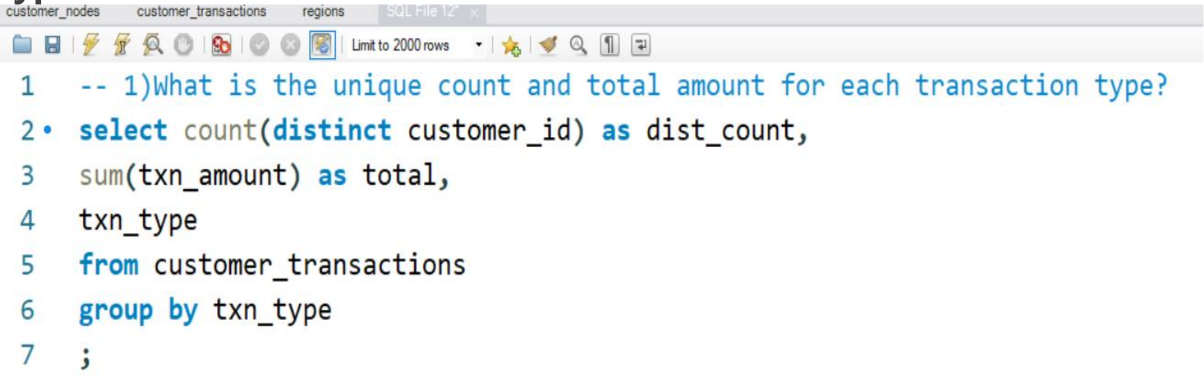
```
avg_days
14.63
```

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

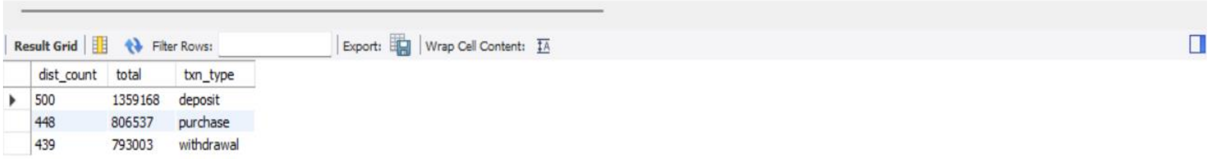
```
WITH reallocation_days_cte AS (
  SELECT *,
        (datediff(end_date, start_date)) AS reallocation_days
  FROM customer_nodes
  INNER JOIN regions USING (region_id)
  WHERE end_date != '9999-12-31'
),
percentile_cte AS (
  SELECT *,
        percent_rank() OVER (PARTITION BY region_id ORDER BY reallocation_days) * 100 AS p
  FROM reallocation_days_cte
)
SELECT region_id,
      region_name,
      reallocation_days
FROM percentile_cte
WHERE p > 80
group by region_id,region_name,reallocation_days
;
```

B. Customer Transactions

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

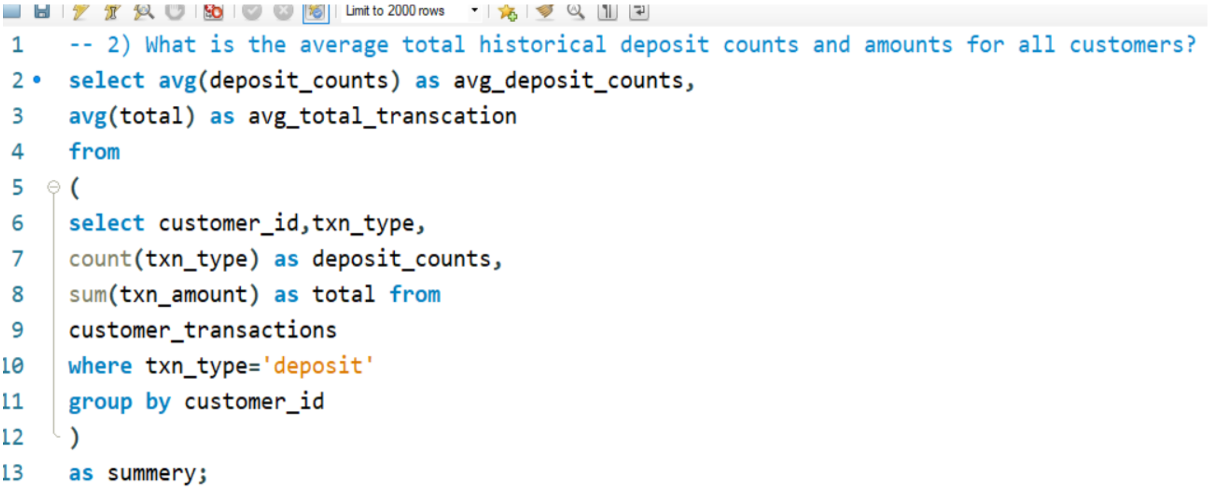


```sql
-- 1)What is the unique count and total amount for each transaction type?
select count(distinct customer_id) as dist_count,
sum(txn_amount) as total,
txn_type
from customer_transactions
group by txn_type
;
```

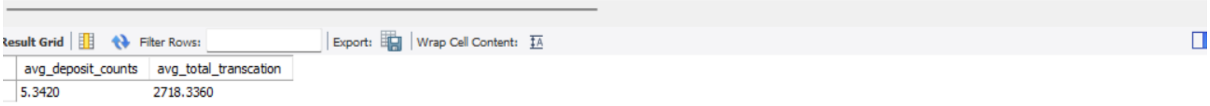


dist_count	total	txn_type
500	1359168	deposit
448	806537	purchase
439	793003	withdrawal

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

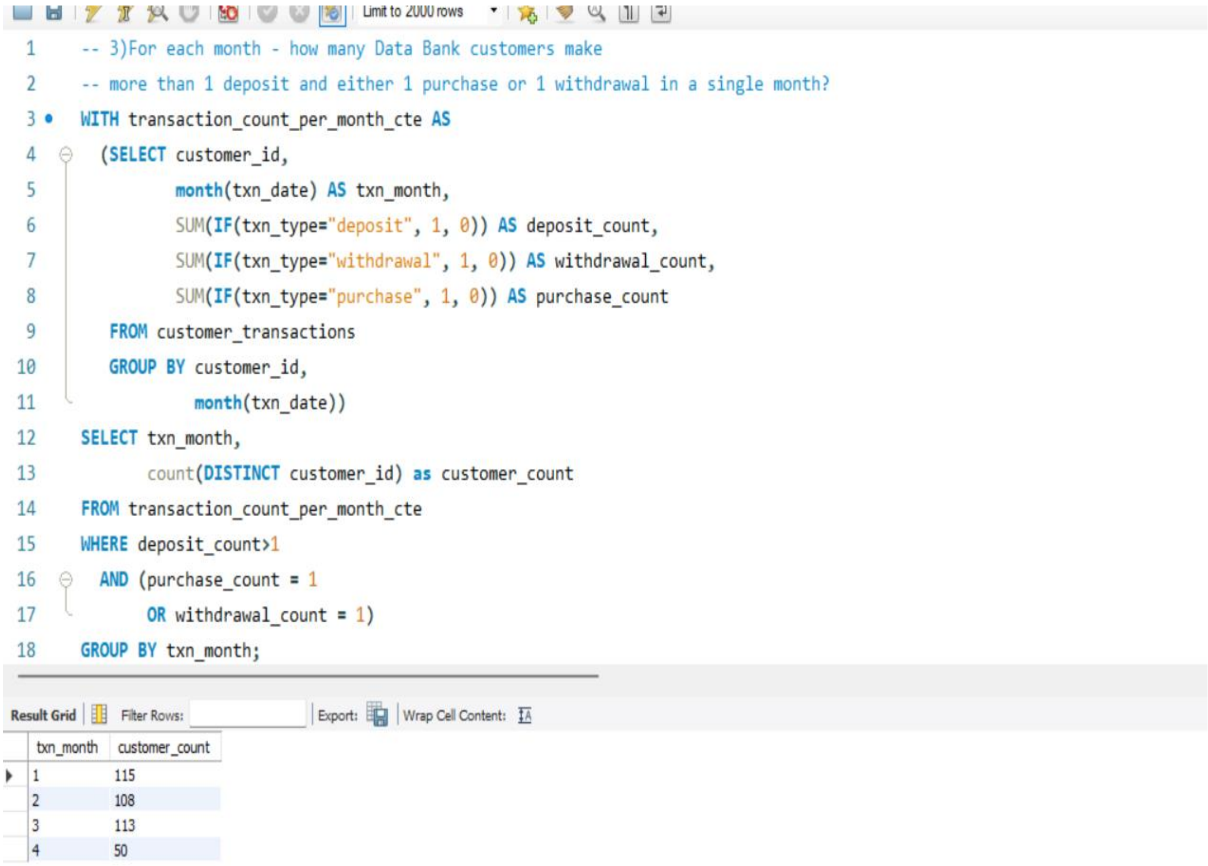


```sql
-- 2) What is the average total historical deposit counts and amounts for all customers?
select avg(deposit_counts) as avg_deposit_counts,
avg(total) as avg_total_transcation
from
(
select customer_id,txn_type,
count(txn_type) as deposit_counts,
sum(txn_amount) as total from
customer_transactions
where txn_type='deposit'
group by customer_id
)
as summery;
```



avg_deposit_counts	avg_total_transcation
5.3420	2718.3360

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?



```sql
-- 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?
WITH transaction_count_per_month_cte AS
  (SELECT customer_id,
          month(txn_date) AS txn_month,
          SUM(IF(txn_type="deposit", 1, 0)) AS deposit_count,
          SUM(IF(txn_type="withdrawal", 1, 0)) AS withdrawal_count,
          SUM(IF(txn_type="purchase", 1, 0)) AS purchase_count
   FROM customer_transactions
   GROUP BY customer_id,
            month(txn_date))
SELECT txn_month,
       count(DISTINCT customer_id) as customer_count
FROM transaction_count_per_month_cte
WHERE deposit_count>1
  AND (purchase_count = 1
       OR withdrawal_count = 1)
GROUP BY txn_month;
```

txn_month	customer_count
1	115
2	108
3	113
4	50