The Data Bank team have prepared a data model for this case study as well as a few example rows from the complete dataset below to get you familiar with their tables.

**Entity Relationship Diagram**

**A diagram of a customer

Description automatically generated with medium confidence**

**Table 1: Regions**

Just like popular cryptocurrency platforms - Data Bank is also run off a network of nodes where both money and data is stored across the globe. In a traditional banking sense - you can think of these nodes as bank branches or stores that exist around the world.

This regions table contains the region\_id and their respective region\_name values

|  |  |
| --- | --- |
| **region\_id** | **region\_name** |
| 1 | Africa |
| 2 | America |
| 3 | Asia |
| 4 | Europe |
| 5 | Oceania |

**Table 2: Customer Nodes**

Customers are randomly distributed across the nodes according to their region - this also specifies exactly which node contains both their cash and data.

This random distribution changes frequently to reduce the risk of hackers getting into Data Bank’s system and stealing customer’s money and data!

Below is a sample of the top 10 rows of the data\_bank.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**

This table stores all customer deposits, withdrawals and purchases made using their Data Bank debit card.

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |

Case Study questions:

The following case study questions include some general data exploration analysis for the nodes and transactions before diving right into the core business questions and finishing with a challenging final request.

The case study questions are grouped into the following categories:

A. Customer Nodes Exploration

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

select count(distinct node\_id)

from customer\_nodes

A screenshot of a computer

Description automatically generated

There are 5 unique nodes(branches) in the Data Bank System.

2.  What is the number of nodes per region?

Select count (c.Node\_id ), r.region\_name

from customer\_nodes as c

Inner join regions as r

on c.region\_id = r.region\_id

Group by r.region\_name

A screenshot of a computer

Description automatically generated

Australia had the highest number of nodes occurrences (770), followed by America (735) with Europe having the least number of nodes (616).

3.How many customers are allocated to each region?

Select count (distinct c.customer\_ID), r.region\_Name

From customer\_nodes as c

Inner join regions as r

on c.region\_ID = r.region\_ID

Group By R.region\_Name;

A screenshot of a computer

Description automatically generated

Australia had the highest number of customers allocated to that region, followed by America, while Europe had the least number of customers.

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

SELECT AVG(f) FROM

(select JULIANDAY(end\_date) - JULIANDAY(start\_date) AS f

from customer\_nodes

where end\_date != '99991231') AS T

B. Customer Transactions

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

Select txn\_type, Count (txn\_type), sum (txn\_amount)

From customer\_Transactions

Group by txn\_type;A screenshot of a graph

Description automatically generated

There were more deposits (2671), followed by purchases (1617), and then withdrawals (1580).

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

Select count (c.txn\_type), avg (c.txn\_amount)

From customer\_transactions as c

Inner join customer\_nodes as n

On c.customer\_id = n.customer\_id

inner join regions as r

on r.region\_id = n.region\_id

group by c.txn\_type;A screenshot of a cell phone

Description automatically generated

The average deposit count for a customer is 5 and the average deposit amount for a customer is 2,718.

The above data points and the SQL queries along with their explanations are in my GitHub repo.

Apart from that, I have also created an interactive dashboard on tableau that brings the entire story of the data flow together.