# Data mining and discovery

# SQL Assignment

Student name: Sai Krishna Vavilli

Student ID 23022047

#### Introduction:

This report aims to showcase how a banking database was created using SQLite and Python. The database is designed with several tables to manage customer information, track account balances, and record transactions between accounts. Key features of the database include:

- (a) Ensuring transaction amounts cannot exceed the sender's available balance.
- (b) Using UUIDs to uniquely identify accounts and transactions.
- (c) Following relational database principles, with proper use of foreign keys to maintain data integrity.
- (d) It also automatically updates the bank balance in real-time following each transaction.
- (e) The transaction ID serves as a compound key i.e (unique indentifier) for each transaction. By integrating a UUID into the transaction ID, the system guarantees that each transaction is distinct, thereby effectively preventing the creation of duplicate transactions.

#### Database Structure and Schema:

The database is designed with three tables: customer, balance, and transaction. Below is the schema for each table.

Customer Table (1000 rows)

customer id: (Primary Key) Unique identifier for each customer (UUID).

first\_name : (Nominal Data) Customer's first name (Text).
last name : (Nominal Data) Customer's last name (Text)

email : (Nominal Data) Customer's email address (Text, can be NULL).

Phone : (Nominal Data) Customer's phone number (random 10 digit

integer, can be NULL).

date\_of\_birth : (Interval Data) Customer's date of birth (Date Time, can be NULL)

account\_type: (Ordinal Data) The customer's type of account, ranked as 'silver', 'gold', or 'platinum'.

# Balance Table (1000 rows)

balance\_id :(Primary Key): Unique identifier for each balance record (UUID).

customer\_id :(Foreign Key): Links to the customer table to indicate which customer owns the balance.

Balance\_amount : (Ratio Data) The balance of the customer's account (Real Number, cannot be NULL).

# Transaction Table (10000 rows)

transaction\_id (Compund Key): Unique identifier for each transaction (UUID). sender\_id (Foreign Key): Links to the balance\_id of the sender from the balance table.

receiver\_id (Foreign Key): Links to the balance\_id of the receiver from the balance table.

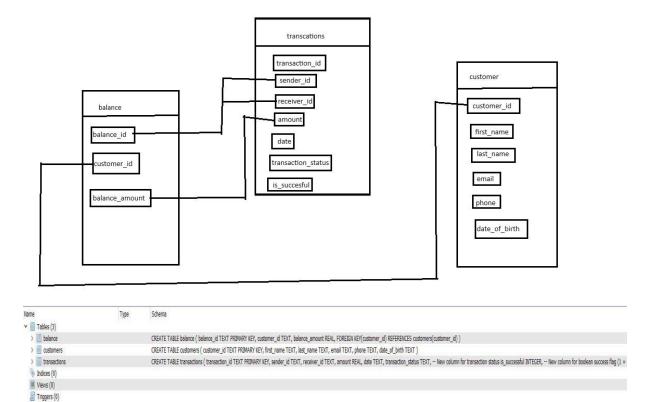
amount: (Ratio Data) The transaction amount (Real Number, cannot exceed sender's balance).

transaction\_date: (Interval Data) The date and time when the transaction occurred (Date Time, stored in YYYY-MM-DD HH:MM:SS format). transcation status: (Nominal Data) This provides a categorical variable

indicating whether the status is 'failed', 'pending', or 'completed'.

is\_successful: (Nominal Data) This column provides a Boolean value indicating whether it is successful or not.

# Graphical representation of the database



## **Transaction Table:**

We now use sender\_id and receiver\_id as a compound key by adding a UNIQUE constraint:

CONSTRAINT sender\_receiver\_unique UNIQUE (sender\_id, receiver\_id). This constraint ensures that no two transactions can occur between the same sender\_id and receiver\_id pair, preventing duplicates.

Each transaction still has a unique transaction\_id (UUID), which is the primary key of the table. This allows each transaction to be uniquely identified, even if multiple transactions occur between the same sender and receiver.

The sender\_ID and receiver\_ID are derived from the balance\_id. The balance table must be updated dynamically after each transaction to reflect the new balances for both the sender and receiver based on the transaction amount.

#### **Process Flow:**

#### Transaction Verification:

Before proceeding with the transaction, verify if the sender has sufficient funds. If the sender's balance is greater than or equal to the transaction amount, the transaction can continue. Otherwise, it should be flagged as unsuccessful.

## Balance Update:

If the transaction is successful, update the sender's balance by subtracting the transaction amount and update the receiver's balance by adding the same amount.

### **Commit Changes:**

After the balances have been updated, commit the changes to the database to ensure that the transaction is permanently recorded and the balances are accurate.

This ensures that the transaction is handled such a way that no money is created nor destroyed which is important for banking.

### **DATA GENERATION:**

Here is how I generated the database

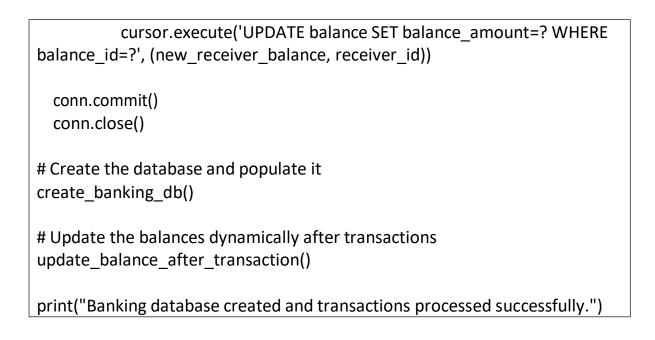
```
import sqlite3
import uuid
import random
from datetime import datetime, timedelta
# Function to create and populate the database
def create_banking_db():
  # Connect to SQLite database (or create it if it doesn't exist)
  conn = sqlite3.connect('banking database.db')
  cursor = conn.cursor()
  # Create Account Types Table (Ranking System for Account Types)
  cursor.execute(""
  CREATE TABLE IF NOT EXISTS account types (
    account type TEXT PRIMARY KEY, -- Silver, Gold, Platinum
    rank INTEGER -- Rank: Silver = 1, Gold = 2, Platinum = 3
  "")
  # Insert account types into the account types table (Ordinal Data)
  cursor.executemany("
  INSERT OR IGNORE INTO account types (account type, rank) VALUES (?, ?)
  "',[
    ('Silver', 1),
    ('Gold', 2),
    ('Platinum', 3)
  1)
  # Create Customer Table with account_type (Foreign Key)
  cursor.execute(""
  CREATE TABLE IF NOT EXISTS customers (
    customer id TEXT PRIMARY KEY,
    first name TEXT,
    last name TEXT,
    email TEXT,
    phone TEXT,
    date of birth TEXT,
```

```
account_type TEXT, -- Foreign Key to account_types table
    FOREIGN KEY(account type) REFERENCES account types(account type)
 ·''')
  # Create Balance Table
  cursor.execute(""
  CREATE TABLE IF NOT EXISTS balance (
    balance id TEXT PRIMARY KEY,
    customer id TEXT,
    balance amount REAL,
    FOREIGN KEY(customer id) REFERENCES customers(customer id)
 "")
  # Create Transaction Table with compound key (transaction id, sender id,
receiver id)
  cursor.execute(""
  CREATE TABLE IF NOT EXISTS transactions (
    transaction id TEXT,
    sender id TEXT,
    receiver id TEXT,
    amount REAL,
    date TEXT,
    transaction_status TEXT,
    is successful INTEGER,
    PRIMARY KEY (transaction id, sender id, receiver id),
    FOREIGN KEY(sender id) REFERENCES balance(balance id),
    FOREIGN KEY(receiver id) REFERENCES balance(balance id)
 "")
  # Generate random data for customers
  def generate random customer():
    first names = ['John', 'Jane', 'Alice', 'Bob', 'Charlie', 'David', 'Emily']
    last_names = ['Smith', 'Doe', 'Johnson', 'Brown', 'Taylor', 'Anderson']
    emails = ['@gmail.com', '@yahoo.com', '@outlook.com']
    phone numbers = ['+1234567890', '+1987654321', '+1112233445']
    date of birth = (datetime.today() -
timedelta(days=random.randint(18*365, 60*365))).strftime('%Y-%m-%d')
```

```
# Assigning account_type from 'Silver', 'Gold', 'Platinum'
    account types = ['Silver', 'Gold', 'Platinum']
    account_type = random.choice(account_types)
    customer id = str(uuid.uuid4())
    first name = random.choice(first names)
    last name = random.choice(last names)
    email = first name.lower() + last name.lower() + random.choice(emails)
    phone = random.choice(phone numbers)
    return customer id, first name, last name, email, phone,
date of birth, account type
  # Insert 1000 customers into the customer table
  for in range(1000):
    customer = generate_random_customer()
    cursor.execute(""
    INSERT INTO customers (customer id, first name, last name, email,
phone, date_of_birth, account_type)
    VALUES (?, ?, ?, ?, ?, ?, ?)
    ''', customer)
  # Generate random balances for customers
  def generate random balance(customer id):
    balance amount = random.uniform(100, 5000) # Random balance
between 100 and 5000
    balance id = str(uuid.uuid4())
    return balance id, customer id, balance amount
  # Insert balances for customers
  for in range(1000):
    customer id = str(uuid.uuid4()) # Random customer id for balance
    balance = generate_random_balance(customer id)
    cursor.execute(""
    INSERT INTO balance (balance_id, customer_id, balance_amount)
    VALUES (?, ?, ?)
    ", balance)
  # Generate random transactions
```

```
def generate_random_transaction(sender_id, receiver_id, amount, date):
    transaction id = str(uuid.uuid4())
    # Randomly assign status and success flag
    status = random.choice(['Pending', 'Completed', 'Failed']) # Transaction
status
    is successful = 1 if status == 'Completed' else 0 # Success flag (1 for
success, 0 for failure)
    return transaction id, sender id, receiver id, amount, date, status,
is successful
  # Insert 10,000 transactions (with realistic constraints)
  for in range(10000): # Insert 10,000 transactions
    # Select a random sender and receiver from the balance table
    sender balance = cursor.execute('SELECT balance id FROM balance
ORDER BY RANDOM() LIMIT 1').fetchone()
    receiver balance = cursor.execute('SELECT balance id FROM balance
ORDER BY RANDOM() LIMIT 1').fetchone()
    if sender balance and receiver balance:
      sender id = sender balance[0]
      receiver id = receiver balance[0]
      # Ensure sender's balance is enough for the transaction
      sender balance amount = cursor.execute('SELECT balance amount
FROM balance WHERE balance id=?', (sender id,)).fetchone()[0]
      amount = random.uniform(10, sender balance amount) # Ensure
transaction amount is less than or equal to balance
      transaction date = datetime.today().strftime('%Y-%m-%d %H:%M:%S')
      # Generate the transaction
      transaction = generate random transaction(sender id, receiver id,
amount, transaction date)
      # Insert the transaction into the table
      cursor.execute(""
      INSERT INTO transactions (transaction_id, sender_id, receiver_id,
amount, date, transaction status, is successful)
      VALUES (?, ?, ?, ?, ?, ?, ?)
      ", transaction)
```

```
# Commit the changes
  conn.commit()
  # Close the connection
  conn.close()
# Function to update the balance dynamically after each transaction
def update balance after transaction():
  conn = sqlite3.connect('banking database.db')
  cursor = conn.cursor()
  # Get all transactions
  cursor.execute('SELECT * FROM transactions')
  transactions = cursor.fetchall()
 for transaction in transactions:
    sender id = transaction[1]
    receiver id = transaction[2]
    amount = transaction[3]
    transaction status = transaction[5] # Get the transaction status
    is successful = transaction[6] # Get the success flag
    # If transaction is successful, update balances
    if is successful == 1:
      # Update sender's balance
      cursor.execute('SELECT balance amount FROM balance WHERE
balance id=?', (sender id,))
      sender balance = cursor.fetchone()
      if sender_balance and sender_balance[0] >= amount:
        new sender balance = sender balance[0] - amount
        cursor.execute('UPDATE balance SET balance amount=? WHERE
balance id=?', (new sender balance, sender id))
        # Update receiver's balance
        cursor.execute('SELECT balance amount FROM balance WHERE
balance id=?', (receiver id,))
        receiver balance = cursor.fetchone()
        if receiver balance:
          new_receiver_balance = receiver_balance[0] + amount
```



Displaying null values in customer table:

```
SQL 1 SELECT
COUNT(*) AS null_email_phone_count
FROM
customer
WHERE
email IS NULL OR phone IS NULL;

null_email_phone_count
1 752

Execution finished without errors.
```

Result: 1 rows returned in 2ms



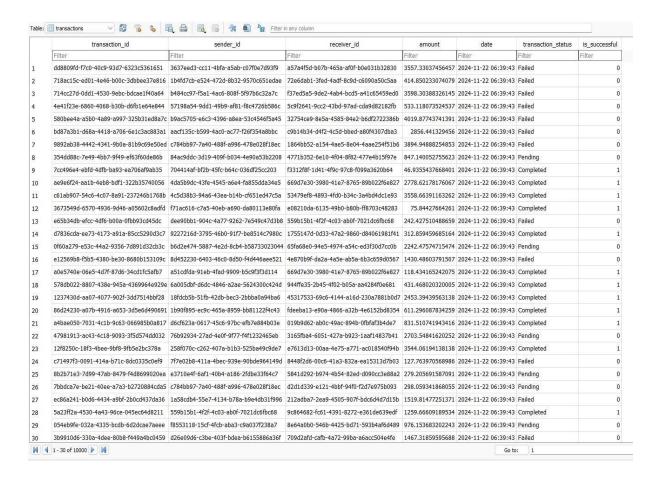
# Customer table:

	customer_id	first_name	last_name	email	phone	date_of_birth
	Filter	Filter	Filter	Filter	Filter	Filter
1	b590fdd2-889f-474c-865d-f657927e429e	Jane	Anderson	janeanderson@yahoo.com	+1987654321	1999-01-03
2	17b5df9e-857f-4f50-9244-7f8ef47d7c76	Emily	Taylor	emilytaylor@yahoo.com	+1987654321	1991-07-12
3	1a058b84-29a0-4c89-b3d4-8f088444c75e	Emily	Smith	emilysmith@yahoo.com	+1987654321	1971-04-05
4	f33b9eb3-dfab-4326-b283-c7d4e040d7ff	David	Brown	davidbrown@gmail.com	+1234567890	1987-12-15
5	4710eff6-9234-404b-9177-47fe66ed2c14	Jane	Taylor	janetaylor@outlook.com	+1112233445	1976-03-23
5	8f4f8f66-d1f8-40af-9e1c-3ed90b25ab7d	Bob	Taylor	bobtaylor@gmail.com	+1987654321	1977-01-19
7	561e588e-33eb-4e73-bd8a-0d76f3af741c	Charlie	Johnson	charliejohnson@gmail.com	+1112233445	2004-05-15
8	e731cea3-f489-43bf-8545-5eba9d16ce89	Emily	Doe	emilydoe@yahoo.com	+1112233445	1967-09-24
9	127d83cb-0e2b-4141-909d-c004247a0839	Bob	Doe	bobdoe@outlook.com	+1987654321	1987-09-28
10	b9ffef06-3080-4058-9f99-88651996a96d	John	Doe	johndoe@outlook.com	+1987654321	1987-10-16
11	395328e7-7fbe-4ec9-b313-fd401ba9c2c6	Bob	Anderson	bobanderson@gmail.com	+1112233445	1988-02-19
12	3ff54453-34a4-4252-8c76-ccf9e2efed62	John	Brown	johnbrown@outlook.com	+1987654321	1992-08-30
13	64d30d2f-d342-4883-a4cc-ebdc1e64427a	Jane	Brown	janebrown@gmail.com	+1112233445	1969-12-28
14	67140c4d-b4ac-49c2-ae31-d1073e9272e2	David	Anderson	davidanderson@gmail.com	+1987654321	1988-01-01
15	b85ce862-74c8-494a-b55c-238009521502	David	Brown	davidbrown@yahoo.com	+1987654321	1979-03-23
16	f3aa0c1f-7afe-48bb-9c57-469e0066d935	John	Smith	johnsmith@yahoo.com	+1234567890	1986-07-01
17	b69b1851-638b-45ad-9765-e58f1b02d854	Emily	Taylor	emilytaylor@gmail.com	+1234567890	1974-03-12
18	361a79bb-cad1-47f8-9002-09f6c9813dcf	Alice	Anderson	aliceanderson@yahoo.com	+1112233445	1976-01-13
19	c30c01fb-92a5-4993-9459-8fba0b262f01	Emily	Johnson	emilyjohnson@gmail.com	+1234567890	2005-12-31
20	f9380dc9-8c2b-47be-bf29-e55e10ebc043	Alice	Taylor	alicetaylor@yahoo.com	+1987654321	1975-04-14
21	e460ffd4-4773-4813-b515-c9571b46b730	Charlie	Anderson	charlieanderson@outlook.com	+1987654321	1998-02-04
22	d220cee5-2799-4fc3-991b-29ad2b1d85af	Alice	Smith	alicesmith@outlook.com	+1234567890	1981-06-18
23	34cdcd96-d7fe-4e91-9644-b3e632777d2f	Bob	Smith	bobsmith@gmail.com	+1987654321	1978-03-13
24	fd036278-a8ef-4d6e-a97d-eff39cec4397	Alice	Brown	alicebrown@yahoo.com	+1112233445	1985-04-01
25	87a5bea6-f8b6-4503-8867-615cce643f80	Emily	Taylor	emilytaylor@outlook.com	+1234567890	1966-02-01
26	3b61411e-9f4e-49c6-a482-600455c95d59	John	Anderson	johnanderson@outlook.com	+1234567890	2004-02-29
27	d5f33c20-cf6a-4b20-ac45-e0a90ced5a37	Alice	Smith	alicesmith@outlook.com	+1234567890	1972-05-28
28	4d8c0c49-3a03-4f6e-bec5-0c097b6217bd	Alice	Smith	alicesmith@outlook.com	+1234567890	1978-08-20
29	2e1842db-d141-4ac7-8323-55a8eb90dcf8	John	Doe	johndoe@yahoo.com	+1234567890	1968-06-10
30	90567950-7273-4de7-a60f-aad0fbef86dc	Jane	Doe	janedoe@gmail.com	+1987654321	1982-07-31

# Balance table:



### Transcation table:



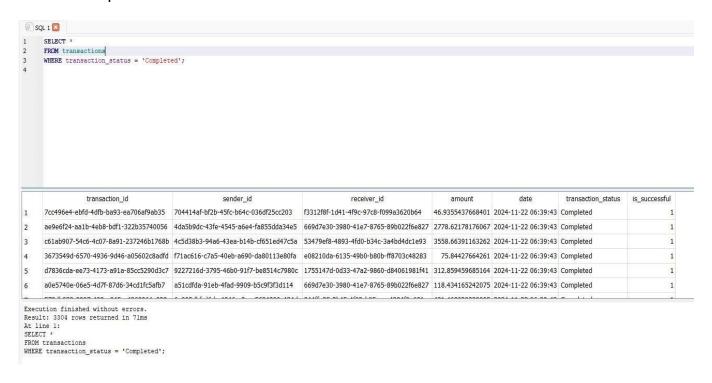
## **Example Queries:**

Get a Customer's Balance

```
SELECT balance_amount
2
     FROM balance
     WHERE customer_id = '2af8d389-4f53-4fa2-809e-672fbcb7eef0';
3
   balance_amount
1 2413.91384395334
Execution finished without errors.
Result: 1 rows returned in 2ms
At line 1:
SELECT balance_amount
FROM balance
WHERE customer_id = '2af8d389-4f53-4fa2-809e-672fbcb7eef0';
```

Get All Transactions of a Specific Customer

## **Get All Completed Transactions**



# Justification for Separate Tables:

By separating the data into different tables — customer, balance, and transaction — we achieve the following:

Reduced Redundancy: Avoid storing the same customer information repeatedly in each transaction or balance record.

Efficient Data Management: Update customer details, balances, or transactions independently without disrupting the structure of other data.

Maintaining Data Integrity: Foreign keys enforce consistency, ensuring that transactions only reference valid customer and balance data.

Scalability: The database structure is scalable, allowing for future expansion (e.g., adding new account types or handling additional transaction attributes).

Security: Sensitive information can be isolated, allowing for better access control and data protection.

## Conclusion

this database design provides a strong foundation for a real-world banking system, balancing efficiency, data integrity, and security while ensuring scalability for future needs. The clear separation of customer information, balances, and transactions allows for effective data management and better system performance. This structure is ideal for supporting the core operations of a banking system, ensuring it can grow and evolve.