**Integrated Case Study**

**Business Problem:**

You are a data analyst working in the Loans Division of the Personal Financial Services department of a major US bank called West Banking Group (WBG). WBG has been issuing mortgages for home owners across more than 200 cities in the US.  
  
The head of the loans division is keen to take stock of the set of loans issued over the most recent month (Aug 2022) before the upcoming meeting with the CEO. He has a set of questions he would like you to answer as part of the business intelligence reporting and trend analysis ahead of the meeting.

In order to do the same, required to perform data analysis to gather insights for reporting.

**About the dataset:**

* **Banker\_Data.csv:**Employee details of the bankers in the loans division  
    
  **banker\_id**: ID of bank employee (banker) **- text string**   
  **first\_name**: First name of banker **- text string**   
  **last\_name**: Last name of banker **- text string**   
  **gender**: Gender of banker **- text string**   
  **phone**: Phone number of banker **- text string**   
  **dob**: Date of Birth of banker **- date**   
  **date\_joined**: Date joined as employee **- date**
* **Customer\_Data.csv:**Details of the home owners issued the home loan from WBG

**customer\_id**: ID of bank customer (loanee) - **integer**   
           **first\_name**: First name of customer - **text string**   
           **last\_name**: Last name of customer - **text string**   
           **email**: Email address of customer - **text string**   
           **gender**: Gender of customer - **text string**   
           **phone**: Phone number of customer - **text string**   
           **dob**: Date of Birth of customer - **date**  
           **customer\_since:**Date joined as bank customer - **date**  
           **nationality**: Nationality of customer -**text string**

* **Home\_Loan\_Data.csv:**Details of the property for which the home loan is issued

**loan\_id**: ID of home loan from bank-**text string**   
           **property\_type**: Type of property (categorical)-**text string**   
           **country**: Country of property (default USA)-**text string**   
           **city**: City of property -**text string**   
           **property\_value**: Value of property (US dollars)-**integer**   
           **loan\_percent**: Percent of property value approved for loan-**integer**   
           **loan\_term**: Loan term (number of years)-**integer**   
           **postal\_code**: Postal code of property-**integer**   
           **joint\_loan**: Whether loan is shared with another person - **text\_string**

* **Loan\_Records\_Data.csv:**Transaction records of the home loans (Aug 2022)

**record\_id**: ID of the home loan transaction record -**integer**   
           **customer\_id**: ID of bank customer (loanee) -**integer**   
           **loan\_id**: ID of home loan from bank **-text string**   
           **banker\_id**: ID of bank employee (banker)-**text string**   
           **transaction\_date**: Date of loan approval -**date**

**Setup Database and Tables**

* Create a new database (named Loans) in SQL Server **- 2Marks**
* Import the 4 CSV files and generate 4 tables in the database. Ensure that the tables are named in a manner that is easy to understand, and that the data type for each column is correctly specified (Refer to above section for more info). In particular, ensure that the date format for the date columns are accurately defined (i.e., %Y-%m-%d). **- 10 Marks**
* Write a query to print all the databases available in the SQL Server. **- 1Marks**
* Write a query to print the names of the tables from the Loans database**. - 1Marks**
* Write a query to print 5 records in each table **- 1Marks**
* **Q. Find the average age of male bankers (years, rounded to 1 decimal place) based on the date they joined WBG** **(2 Marks)**
* 
* Your answer
* **Q. Find the customer ID, first name, last name, and email of customers whose email address contains the term 'amazon'.** **(1 Marks)**
* 
* Your answer
* **Q. Find the average age (at the point of loan transaction, in years and nearest integer) of female customers who took a non-joint loan for townhomes.** **(2 Marks)**
* 
* Your answer
* **Q. Find the number of home loans issued in San Francisco.** **(1 Marks)**
* 
* Your answer
* **Q. Find the names of the top 3 cities (based on descending alphabetical order) and corresponding loan percent (in ascending order) with the lowest average loan percent.** **(2 Marks)**
* 
* Your answer
* **Q. Find the maximum property value (using appropriate alias) of each property type, ordered by the maximum property value in descending order.** **(1 Marks)**
* 
* Your answer
* **Q. Find the ID, first name, and last name of the top 2 bankers (and corresponding transaction count) involved in the highest number of distinct loan records.** **(2 Marks)**
* 
* Your answer
* **Q. Find the average loan term for loans not for semi-detached and townhome property types, and are in the following list of cities: Sparks, Biloxi, Waco, Las Vegas, and Lansing.** **(2 Marks)**
* 
* Your answer
* **Q. Find the city name and the corresponding average property value (using appropriate alias) for cities where the average property value is greater than $3,000,000.** **(1 Marks)**
* 
* Your answer
* **Q. Find the total number of different cities for which home loans have been issued.** **(1 Marks)**
* 
* Your answer
* **Q. Create a stored procedure called `city\_and\_above\_loan\_amt` that takes in two parameters (city\_name, loan\_amt\_cutoff) that returns the full details of customers with loans for properties in the input city and with loan amount greater than or equal to the input loan amount cutoff.**
* **Call the stored procedure `city\_and\_above\_loan\_amt` you created above, based on the city San Francisco and loan amount cutoff of $1.5 million**    **(5 Marks)**
* 
* Your answer
* **Q. Find the number of Chinese customers with joint loans with property values less than $2.1 million, and served by female bankers. (3 Marks)**
* 
* Your answer
* **Q. Find the top 3 transaction dates (and corresponding loan amount sum) for which the sum of loan amount issued on that date is the highest.** **(3 Marks)**
* 
* Your answer
* **Q. Find the number of bankers involved in loans where the loan amount is greater than the average loan amount.** **(3 Marks)**
* 
* Your answer
* **Q. Find the sum of the loan amounts ((i.e., property value x loan percent / 100) for each banker ID, excluding properties based in the cities of Dallas and Waco. The sum values should be rounded to nearest integer.** **(3 Marks)**
* 
* Your answer
* **Q. Find the ID, first name and last name of customers with properties of value between $1.5 and $1.9 million, along with a new column 'tenure' that categorizes how long the customer has been with WBG.**
* **The 'tenure' column is based on the following logic:  
  Long: Joined before 1 Jan 2015  
  Mid: Joined on or after 1 Jan 2015, but before 1 Jan 2019  
  Short: Joined on or after 1 Jan 2019**
* **(2 Marks)**
* 
* Your answer
* **Q. Find the ID and full name (first name concatenated with last name) of customers who were served by bankers aged below 30 (as of 1 Aug 2022).** **(3 Marks)**
* 
* Your answer
* **Q. Create a stored procedure called `recent\_joiners` that returns the ID, concatenated full name, date of birth, and join date of bankers who joined within the recent 2 years (as of 1 Sep 2022)**
* **Call the stored procedure `recent\_joiners` you created above (5 Marks)**
* 
* Your answer
* **Q. Create a view called `dallas\_townhomes\_gte\_1m` which returns all the details of loans involving properties of townhome type, located in Dallas, and have loan amount of >$1 million. (3 Marks)**