SecureBank Data Warehouse

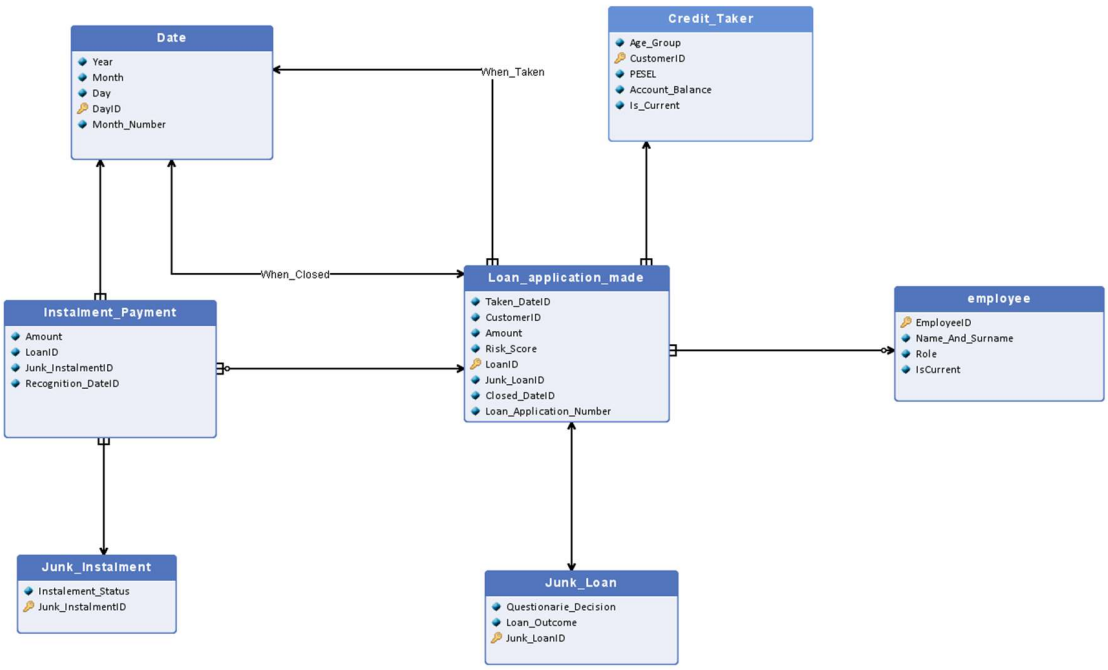

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
LOAN_APPLICATION_MADE	Fact table. One tuple describes one credit application.		
	LoanID	Numeric (PK)	Surrogate key for the loan application
	Loan_Application_Number	Numeric	Loan application identifier (business key)
	Taken_DateID	Numeric (FK)	Application date and in case of acceptance date of taking a loan
	Closed_DateID	Numeric (FK)	Loan closing date
	CustomerID	Numeric (FK)	Link to credit taker
	Amount	Decimal	Requested loan amount
	Risk_Score	Decimal	Calculated credit risk score (0-10)

	Junk_LoanID	Numeric (FK)	Link to Junk containing final loan status and questionnaire decision
	EmployeeID	Numeric (FK)	Advisor who assisted the loan or ID of "no employer"
INSTALLMENT_PAYMENT	Fact table. One tuple describes one repayment toward a loan.		
	LoanID	Numeric (FK)	Refers to loan being repaid
	Recognition_DateID	Numeric (FK)	Payment registration date
	Amount	Decimal	Amount of the installment payment
	Junk_InstalmentID	Numeric (FK)	Link to installment junk
DATE	Dimension table. Describes dates in the warehouse.		
	DayID	Numeric (PK)	Surrogate key for the date
	Year	Numeric	Year part of the date
	Month	Varchar	Month name(January, February, March, April, May, June, July, August, September, October, November, December)
	Day	Numeric	Day of the month
	Month_Number	Numeric	Numerical representation of a month
CREDIT_TAKER	Dimension table. SCD2. Contains information about credit applicants.		
	CustomerID	Numeric (PK)	Surrogate key identifying the credit taker

	PESEL	Varchar	National identifier (business key)
	Age_Group	Varchar	Age interval.(18-35, 36-60,61+)
	Account_Balance_Level	Varchar	Categorized account balance: high (>2mln), medium (2mln, 1mln), low (<1mln). Values in euro
	Is_Current	Boolean	Flag which holds information if this entity is latest appearance of customer
EMPLOYEE	Dimension table. SCD2. Contains basic employee information. There is on entity as "no employer" with nulled fields		
	EmployeeID	Numeric (PK)	Unique identifier for the employee
	Name_And_Surname	Varchar(nullable)	First and last name of the employee
	Role	Varchar(nullable)	Employees' role in the company: intern, junior salesman, salesman, senior salesman
	Is_Current	Boolean	Flag which holds information if this entity is latest appearance of Employee
JUNK_LOAN	Dimension table. Stores additional information about Loan		
	Junk_LoanID	Numeric (PK)	Surrogate key for Junk
	Loan_Outcome	Varchar	The final state of the loan after it is closed. (defaulted, closed, null)
	Questionnaire_Decision	Varchar	Decision about acceptance of the loan questionnaire. Allowed options: rejected, accepted.

JUNK_INSTALMENT	Dimension table. Stores information about instalment status		
	Junk_InstalmentID	Numeric (PK)	Surrogate key for Junk
	Instalment _Status	Varchar	Final state of installment. (on-time, late, defaulted).

Dimensional model

Fact definitions

Fact 1 Loan application fact

Loan application submitted on a specified date by a specified customer with a requested amount and a calculated risk score, assessed with a specified status. The application could be assisted by a specified bank employee or none.

Granularity:

- a specified date of application,
- a specified date of closing,
- a specified customer (with attributes such as age group, income level etc.)
- a specified application status (e.g., rejected, on-time, defaulted)
- a specified advisor (with specified position)assisting the process (or none),

Measures and aggregation functions:

- Number of application facts – COUNT (1)
- Sum from Risk Score – SUM (Risk_Score)
- Sum from loan Amount –SUM(Amount)
- Average loan Amount - Sum from loan Amount /number of application Facts
- Average Risk score - Sum from Risk Score /number of application Facts

Fact 2 Loan instalment recognition fact

Loan instalment with specific amount been recognized as *paid on time/late/defaulted* on a specified date with a connection to the specific loan.

Granularity:

- a specified loan,
- a specified date of recognition
- a specified instalment status (e.g. late, on-time, defaulted)

Measures and aggregation functions:

- Number of instalment recognition facts – COUNT (1)
- Sum from instalment Amount - SUM(Amount)

- Average instalment Amount – Sum from instalment Amount / Number of instalment facts

Dimension definitions

Dimensions for Fact 1 Loan application fact:

DIMENSION/DIMENSION ATTRIBUTE	DIMENSION/DIMENSION ATTRIBUTE	TYPE
APPLICATION DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Day	Hierarchical dimension
APPLICATION DATE	Date	Dimension
APPLICATION YEAR	Date.Year	Dimension attribute
APPLICATION MONTH	Date.Month	Dimension attribute
APPLICATION DAY	Date.Day	Dimension attribute
CLOSURE DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Day	Hierarchical dimension
CLOSURE DATE	Date	Dimension
CLOSURE YEAR	Date.Year	Dimension attribute
CLOSURE MONTH	Date.Month	Dimension attribute
CLOSURE DAY	Date.Day	Dimension attribute
CREDIT TAKER	Credit_Taker	Dimension
CREDIT TAKER PESEL	Credit_Taker.PESEL	Dimension attribute
CREDIT TAKER AGE GROUP	Credit_Taker.Age_Group	Dimension attribute
CREDIT TAKER BALANCE LEVEL	Credit_Taker.Account_Balance_Level	Dimension attribute
LOAN APPLICATION NUMBER	Loan_Application_Number	Degenerated dimension
LOAN JUNK	Junk_Loan	Dimension
LOAN OUTCOME	Junk_Loan.Loan_Outcome	Dimension attribute
LOAN QUESTIONARE DECISION	Junk_Loan. Questionnaire_Decision	Dimension attribute
EMPLOYEE	Employee	Dimension
EMPLOYEE NAME	Employee.Name_And_Surname	Dimension attribute
EMPLOYEE ROLE	Employee.Role	Dimension attribute
EMPLOYEE HIERARCHY	• Employee.Role ••Employee.Name_And_Surname	Hierarchical dimension

Dimensions for Fact 2 Loan instalment recognition fact:

DIMENSION/DIMENSION ATTRIBUTE	DIMENSION/DIMENSION ATTRIBUTE	TYPE
LOAN PAYMENT RECOGNISION DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Day	Hierarchical dimension

RECOGNISION DATE	Date	Dimension
RECOGNISION YEAR	Date.Year	Dimension attribute
RECOGNISION MONTH	Date.Month	Dimension attribute
RECOGNISION DAY	Date.Day	Dimension attribute
CREDIT TAKER	Credit_Taker	Dimension
CREDIT TAKER PESEL	Credit_Taker.PESEL	Dimension attribute
CREDIT TAKER AGE GROUP	Credit_Taker.Age_Group	Dimension attribute
CREDIT TAKER BALANCE LEVEL	Credit_Taker.Account_Balance_Level	Dimension attribute
INSTALLMENT JUNK	Junk_Instalment	Dimension
INSTALLMENT STATUS	Junk_Instalment.Status	Dimension attribute
EMPLOYEE	Employee	employee
EMPLOYEE NAME	Employee.Name_And_Surname	employee Name
EMPLOYEE ROLE	Employee.Role	Employee Role
EMPLOYEE HIERARCHY	• Employee.Role ••Employee.Name_And_Surname	Employee Hierarchy
LOAN	Loan_application_made	Dimension
LOAN APPLICATION NUMBER	Loan_application_made.Loan_Application_Number	Dimension attribute
LOAN JUNK	Junk_Loan	Dimension
LOAN OUTCOME	Junk_Loan.Loan_Outcome	Dimension attribute
LOAN QUESTIONARE DECISION	Junk_Loan. Questionnaire_Decision	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. Calculate the average credit risk score of defaulted loans

Measure: Average Risk_score
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = 'defaulted')

---

2. Determine if the trend of defaults has increased over the last 12 months.

Measure: Number of application facts
Dimension: Date (attributes: Year, Month)
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ' 'defaulted')

---

3. Identify which age group is most likely to default.

Measure: Number of application facts
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ' 'defaulted')
Dimension: Credit_Taker (attribute: Age_Group)

---

4. List 5 most common numbers of late payments of instalments for defaulted loans.

Measure: Number of instalment recognition facts
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘ ‘defaulted’)
Dimension: Junk_Instalment (attribute: Instalment _Status, where Instalment _Status = ‘late’)

---

5. Compare default rates between customers with high and low account balances.

Measure: Number of application facts
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘defaulted’)
Dimension: Credit_Taker (attribute: Account_Balance_Level)
]
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘closed ’))
Dimension: Credit_Taker (attribute: Account_Balance_Level)
]

---

6. Compare the rate of defaulted to successful loans between customers with and without an advisor.

Measure: Number of application facts
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘defaulted’)
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS null [for “no employee”])
]
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘closed ’))
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS NOT null [for “no employee”])
]

---

7. Calculate the mean value of the loans taken without an advisor.

Measure: Average loan Amount
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS null [for “no employee”])

---

8. Determine whether the trend of loans without an advisor is increasing over the last 12 months.

Measure: Number of application facts
Dimension: Date (attributes: Year, Month**)**
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS null [for “no employee”])
Dimension: Junk_Loan (attribute: Questionnaire_Decision where Questionnaire_Decision = ‘accepted ’))

---

9. Identify 5 advisors with the highest rate of defaulted credits to successful ones.

Measure: Number of application facts
Dimension: employee (attribute: EmployeeID)
Dimension: Junk (attribute: Status)

9.1 **Rank the employee roles according to their rate of defaulted credits to successful ones.**

Measure: Number of application facts
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘defaulted’)
Dimension: Employee (attribute: Role)
]
[
Dimension: Junk_Loan (attribute: Loan_Outcome where Loan_Outcome Status = ‘closed ’))
Dimension: Employee (attribute: Role)
]

---

10. Find age groups where the ratio of loans taken with an advisor exceeds 0.8.

Measure: Number of application facts
[

Dimension: Junk_Loan (attribute: Questionnaire_Decision where Questionnaire_Decision = 'accepted '))
Dimension: Credit_Taker (attribute: Age_Group)
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS null)
]
[

Dimension: Junk_Loan (attribute: Questionnaire_Decision where Questionnaire_Decision = 'accepted '))
Dimension: Credit_Taker (attribute: Age_Group)
Dimension: Employee (attribute: Name_And_Surname where Name_And_Surname IS NOT null)
]

Date in the Date sources needed to fill the Date warehouse

TABLE_NAME	COLUMN	SOURCE AND DESCRIPTION
LOAN_APPLICATION_MADE	One tuple describes one fact of loan application made.	
	LoanID	Loan ID. Surrogate key - generated by database.
	Loan_Application_Number	Loan ID. Taken from CREDIT questionnaire, Column Q_ID – Unique application ID.
	Taken_DateID	Date ID of taking a loan. Foreign key from dimension table. Based on Column Q_DATE from CREDIT questionnaire.
	Closed_DateID	Date ID of closing a loan. Foreign key from dimension table. Based on Bankoteka Loan.Date_Of_Credit_Closing (for rejected loans equal to '0')
	CustomerID	Customer ID. Foreign key from dimension table. Based on Column CL_ID from CREDIT questionnaire, – Bank customer ID.
	Amount	The quantity of money loaned to customer. Taken from Column DES_CRED_VAL from CREDIT questionnaire.
	Risk_Score	Risk score assigned to loan. Taken from Column RISK_CALC from CREDIT questionnaire.
	Junk_LoanID	Junk loan ID. Foreign key from dimension table.

	EmployeeID	Employee ID. Foreign key from dimension table. Based on CREDIT questionnaire, Column E_ASSIST – Employee who assisted (or Default = no advisor) in a process of loan taking.
INSTALLMENT_PAYMENT	One tuple describes one fact of installment payment.	
	LoanID	Loan ID. Foreign key from dimension table. Based on Column Q_ID from CREDIT questionnaire. Unique ID of a loan which the instalment is repaying.
	Recognition_DateID	Date ID. Foreign key from dimension table. Based on Loan_Payment. Date from BANKOTEKA.
	Amount	The quantity of money paid as a partial repayment of a loan. Based on Loan_Payment.Amount from BANKOTEKA.
	Junk_InstalmentID	Junk instalment ID. Foreign key from dimension table.
CREDIT_TAKER	One tuple describes one credit taker in the specified age group, with specified account balance at the moment of taking a loan	
	CustomerID	Customer ID. Surrogate key - generated by database.
	PESEL	Customer's PESEL number. Business taken from Customer.PESEL from BANKOTEKA.
	Age_Group	Based on BANKOTEKA, Customer.Date_Of_Birth and Loan.Approval_Date – Age group derived from those dates. Allowed values: 18-35, 36-60,61+.
	Account_Balance_Level	State of account at the time of loan submission. Based on BANKOTEKA, Account.Balance. Allowed values: : high (>2mln), medium (2mln, 1mln), low (<1mln).

	Is_Current	"1" if information is current, otherwise "0" (SCD2 implementation).
EMPLOYEE	One tuple describes one employee with specified role.	
	EmployeeID	Employee ID. Based on Employee.EmployeeID from BANKOTEKA – Unique employee ID.
	Name_And_Surname	Name and surname of an employee. Based on Employee.Name and Employee.Surname from BANKOTEKA.
	Role	Role of an employee. Based on Employee.Role from BANKOTEKA. Allowed values: intern, junior, salesman, senior salesman.
JUNK_INSTALMENT	The tuples correspond to "all" possible values for Instalment_Status and are generated before ETL process.	
	Junk_ InstalmentID	Junk instalment ID. Surrogate key - generated by database.
	Instalement_Status	Based on Loan_Payment.Status from Bankoteka. Allowed values: on-time, late, defaulted.
JUNK_LOAN	The tuples correspond to three possible combinations of values for Questionnaire_Decision and Loan_Outcome: [rejected, null], [accepted, defaulted], [accepted, closed]. They are generated before ETL process.	
	Junk_LoanID	Junk loan ID. Surrogate key - generated by database.
	Questionnaire_Decision	Decision about acceptance of the loan. Based on Column N - CR_DECISION from CREDIT questionnaire. Allowed options: rejected, accepted.
	Loan_Outcome	The final state of the loan after it is closed. Based on Loan.Status from Bankoteka. Allowed options: defaulted, closed, null
DATE	**One tuple describes one day.** All the data in this table are generated tuple by tuple based on any calendar, before ETL process.	