Car Insurance

Data Warehouse design

Business process

The Data Warehouse is designed for the Dealing with Claims process undertaken by *NoLimit* car insurer as described in *Car Insurance: Requirements Process Specification*.

Relational Database schema

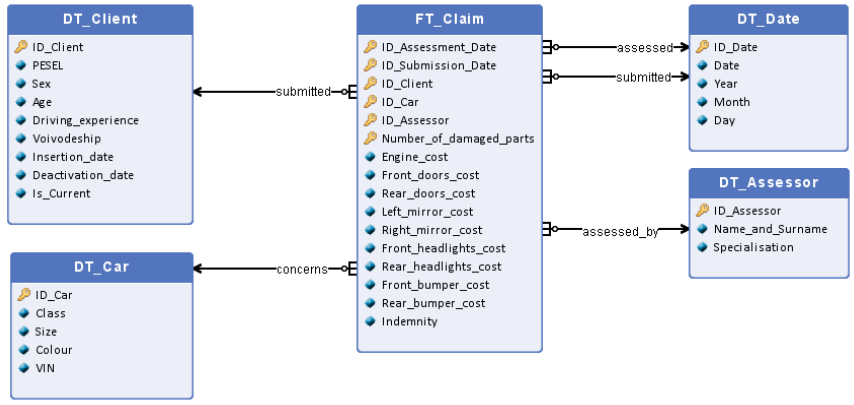

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
FT_Claim (Fact Table)	**Each tuple describes a fact of dealing with a claim.**		
	ID_Assessment_Date	Numeric	FK DT_Date; the day of claim evaluation
	ID_Submission_Date	Numeric	FK DT_Date; the day of claim submission
	ID_Client	Numeric	FK DT_Client; the claimant
	ID_Car	Numeric	FK DT_Car; the damaged car
	ID_Assessor	Numeric	FK DT_Assessor; the claims adjuster who evaluates the claim
	Number_of_damaged_parts	Numeric	The result of claim evaluation - a count of parts requiring repair or replacement. An integer in range 0-9.
	Indemnity	Money	Total money value granted as a result of claim; a nonnegative value

	Engine_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Front_doors_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Rear_doors_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Left_mirror_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Right_mirror_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Front_headlights_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Rear_headlights_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Front_bumper_cost	Money	The cost of a given part determined as a result of the claim evaluation; a nonnegative value
	Rear_bumper_cost	Money	The cost of a given part determined as a

DT_Client (Dimension Table)			result of the claim evaluation; a nonnegative value
	Each tuple describes a *NoLimit's* client.		
	ID_Client	Numeric	PK (surrogate key)
	PESEL	Char(11)	BK
	Sex	VarChar(6)	Values: "female", "male"
	Age	VarChar(13)	Values: „from 18 to 21", "from 22 to 29", "from 30 to 49", "from 50 to 64", "more than 64"
	Driving_experience	VarChar(26)	Values: "up to one year", "between one and five years", "between five and ten years", "more than ten years"
	Voivodeship	VarChar(19)	Client's place of residence – voivodeship; Values: Dolnośląskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie, Łódzkie, Małopolskie, Mazowieckie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Śląskie, Świętokrzyskie, Warmińsko-Mazurskie, Wielkopolskie, Zachodniopomorskie
	Insertion_date	Date	Date (YYYY/MM/DD); The date when the tuple is inserted (SCD2 implementation).

	Deactivation_date	Date	Date (YYYY/MM/DD); The date when the tuple is marked as obsolete (SCD2 implementation).
	Is_Current	Boolean	1 if information is current, otherwise 0 (SCD2 implementation).
DT_Car (Dimension Table)	**Each tuple describes a damaged car the claim is made for.**		
	ID_Car	Numeric	PK (surrogate key)
	VIN	Char(17)	BK
	Class	VarChar(7)	Values: “cheap”, “medium”, “premium”
	Size	VarChar(6)	Values: “small”, “medium”, “large”, “cargo”
	Colour	VarChar(15)	Main colour or theme
DT_Assessor (Dimension Table)	**Each tuple describes a claims adjuster evaluating the claim – a *NoLimit’s* employee.**		
	ID_Assessor	Numeric	PK
	Name_and_Surname	VarChar(50)	Name and surname
	Specialisation	VarChar(7)	The type of car the Assessor specialises in; Values: “cargo”, “premium”, “casual”
DT_Date (Dimension Table)	**One tuple describes one day.**		
	ID_Date	Numeric	PK
	Date	Date	Date (YYYY/MM/DD); BK
	Year	4 digits	Year; Value between 1993 and 2023.
	Month	VarChar(10)	Month; Values: January, February, March, April, May, June, July, August, September, October, November and December.

	Day	1 or 2 digits	Day in month

Dimensional model

Facts definitions

Fact 1 Dealing with Claim fact: Dealing with the specific claim submitted by a specific client on a specific day. Concerned a specific car insured by the client. Evaluated by a specific assessor on a specific day, who assessed which specific parts were damaged.

Fact table: *Claim*

Granularity:
- a specified claim,
- a specified client (DT_Client),
- a specified car (DT_Car),
- a specified assessor (DT_Assessor),
- a specified date of submission (DT_Date),
- a specified date of assessment (DT_Date).

Measures and aggregation functions:
Cost of engine – SUM (Engine_cost)
Cost of front doors – SUM (Front_doors_cost)
Cost of rear doors – SUM (Rear_doors_cost)
Cost of left mirror – SUM (Left_mirror_cost)
Cost of right mirror – SUM (Right_mirror_cost)
Cost of front headlights – SUM (Front_headlights_cost)
Cost of rear headlights – SUM (Rear_headlights_cost)
Cost of front bumper – SUM (Front_bumper_cost)
Cost of rear bumper – SUM (Rear_bumper_cost)
Number of claim facts – COUNT (1)
Value of claim – SUM (Indemnity)
Percentage of indemnity related to Engine cost – (SUM (Cost of engine) / SUM (Value of claim)) / COUNT (1)

Dimensions definitions

Dimensions for Fact 1 Dealing with Claim fact:

DIMENSION/ DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
NUMBER OF DAMAGED PARTS	**FT_Claim.Number_of_damaged_parts**	**Degenerate dimension**
CLIENT	**DT_Client**	**Dimension**
PESEL	DT_Client.PESEL	Dimension attribute
SEX	DT_Client.Sex	Dimension attribute

AGE	DT_Client.Age	Dimension attribute
DRIVING EXPERIENCE	DT_Client.Driving_experience	Dimension attribute
VOIVODESHIP	DT_Client.Voivodeship	Dimension attribute
CAR	**DT_Car**	**Dimension**
VIN	DT_Car.VIN	Dimension attribute
CLASS	DT_Car.Class	Dimension attribute
SIZE	DT_Car.Size	Dimension attribute
COLOUR	DT_Car.Colour	Dimension attribute
ASSESSOR	**DT_Assessor**	**Dimension**
NAME AND SURNAME	DT_Assessor.Name_and_Surname	Dimension attribute
SPECIALISATION	DT_Assessor.Specialisation	Dimension attribute
SUBMISSION DATE	**DT_Date**	**Dimension**
SUBMISSION YEAR	DT_Date.Year	Dimension attribute
SUBMISSION MONTH	DT_Date.Month	Dimension attribute
SUBMISSION DAY	DT_Date.Day	Dimension attribute
ASSESSMENT DATE	**DT_Date**	**Dimension**
ASSESSMENT YEAR	DT_Date.Year	Dimension attribute
ASSESSMENT MONTH	DT_Date.Month	Dimension attribute
ASSESSMENT DAY	DT_Date.Day	Dimension attribute
SUBMISSION DATE HIERARCHY	• DT_Date.Year •• DT_Date.Month ••• DT_Date.Day	Hierarchical dimension
ASSESSMENT DATE HIERARCHY	• DT_Date.Year •• DT_Date.Month ••• DT_Date.Day	Hierarchical dimension

CAR TYPE HIERARCHY	• DT_Car.Size •• DT_Car.Class ••• DT_Car.Colour	Hierarchical dimension

Checking the feasibility of queries based on the multidimensional model

What is the effect of car properties on the amount and number of claims?

1. Compare the number/amount of claims of different vehicle classes (cheap, medium, premium) in the analysed month relative to previous months.

Measure: Value of claim
Measure: Number of claim facts
Dimension: Car (dimension attribute: Class)
Dimension: Submission date (dimension attribute: Submission month)

2. Compare the number/amount of claims of different vehicle colours in the analysed month relative to previous months.

Measure: Value of claim
Measure: Number of claim facts
Dimension: Car (dimension attribute: Colour)
Dimension: Submission date (dimension attribute: Submission month)

3. Compare the number/amount of claims with respect to the number of damaged parts in the analysed month relative to previous months.

Measure: Value of claim
Measure: Number of claim facts
Dimension (degenerate): Number of damaged parts
Dimension: Assessment date (dimension attribute: Assessment month)

4. How do cars of different sizes differ in terms of percentage of indemnity related to engine?

Measure: Percentage of indemnity related to Engine cost
Dimension: Car (dimension attribute: Size)

5. Does the indemnity differ for cars evaluated by assessors with different specialisations (“cargo”, “premium”, “casual”)?

Measure: Value of claim
Dimension: Assessor (dimension attribute: Specialisation)

What is the effect of driver characteristics on the amount and number of claims?

1. Are less experienced drivers responsible for more claims?

Measure: Number of claim facts
Dimension: Client (dimension attribute: Driving experience)

2. Are there differences in number and value of claims between men and women from the same age/experience groups?

Measure: Value of claim
Measure: Number of claim facts
Dimension: Client (dimension attributes: Sex, Driving experience, Age)

3. Compare the number of claims in different voivodeships in the analysed month relative to previous months.

Measure: Number of claim facts

Dimension: Client (dimension attribute: Voivodeship)

Dimension: Submission date (dimension attribute: Submission month)

4. How do women and men differ in terms of percentage of indemnity related to engine?

Measure: Percentage of indemnity related to Engine cost

Dimension: Client (dimension attribute: Sex)

5. Which age groups are responsible for most and least expensive claims?

Measure: Value of claim

Dimension: Client (dimension attribute: Age)

Checking if there are Data in the Data sources needed to fill the Data warehouse

TABLE NAME	COLUMN	SOURCE
FT_Claim	**Each tuple describes a fact of dealing with a claim.**	
	ID_Assessment_Date	Assessment date ID. A foreign key from dimension table. Based on *Evaluation_date* in *Claims* table from *SureSale* source.
	ID_Submission_Date	Submission date ID. A foreign key from dimension table. Based on *Submission_date* in *Claims* table from *SureSale* source.
	ID_Client	Client ID. A foreign key from dimension table. Based on *ID* in *Clients* table from *SureSale* source.
	ID_Car	Car ID. A foreign key from dimension table. Based on *ID* in *Cars* table from *SureSale* source.
	ID_Assessor	Assessor ID. A foreign key from dimension table. Based on *ID* in *Assessors* table from *SureSale* source.
	Number_of_damaged_parts	The number of damaged parts is calculated as a count of nonzero (values: 1 or 2) columns (*Engine-Rear_bumper*) in *Claim* table from*SureSale* system.

	Indemnity	Total money value granted as a result of claim. A sum of columns *Engine_cost-Rear_bumper_cost* from the Data Warehouse.
	Engine_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Engine* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Front_doors_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Front_doors* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Rear_doors_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Rear_doors* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.

	Left_mirror_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Left_mirror* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Right_mirror_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Right_mirror* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Front_headlights_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Front_headlights* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Rear_headlights_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Rear_headlights* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) *

		Value (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Front_bumper_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Front_bumper* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
	Rear_bumper_cost	The cost of a given part determined as a result of the claim evaluation. A product of *Rear_bumper* (*Claims* table from *SureSale* system; possible values: 0, 1, 2) * *Value* (column *D* from *PartsCatalogueCSV*) * 0.5 (a scaling factor). Appropriate *Value* is chosen based on *Car_type_ID* in the ETL process.
DT_Client	**Each tuple describes a *NoLimit's* client (Implementation of SCD2).**	
	ID_Client	Client ID. Surrogate key - generated by database.
	PESEL	Business key taken from *PESEL* in *Clients* from *SureSale* system.
	Sex	Client's sex. Values: "female" (if 'F'), "male" (if 'M'). Based on *Sex* in *Clients* from *SureSale* system.

	Age	Client's age category. Values: „from 18 to 21", "from 22 to 29", "from 30 to 49", "from 50 to 64", "more than 64". Based on *Date_of_birth* in *Clients* from *SureSale* system.
	Driving_experience	Client's driving experience category. Values: "up to one year", "between one and five years", "between five and ten years", "more than ten years". Based on *License_issuing_date* in *Clients* from *SureSale* system.
	Voivodeship	Client's place of residence – voivodeship; Values: Dolnośląskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie, Łódzkie, Małopolskie, Mazowieckie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Śląskie, Świętokrzyskie, Warmińsko-Mazurskie, Wielkopolskie, Zachodniopomorskie. Taken from *Voivodeship* in *Clients* from *SureSale* system.
	Insertion_date	YYYY/MM/DD date. The date when the tuple is inserted (SCD2 implementation).
	Deactivation_date	YYYY/MM/DD date. The date when the tuple is marked as obsolete (SCD2 implementation).

DT_Car	Is_Current	"1" if information is current, otherwise "0" (SCD2 implementation).
	Each tuple describes a damaged car the claim is made for.	
	ID_Car	Car ID. Surrogate key - generated by database.
	VIN	Vehicle Identification Number business key. Taken from *VIN* in *Cars* from *SureSale* system.
	Class	Class of car. Values: "cheap", "medium", "premium". Taken from *Class* in *Car_Types* and *Car_type_ID* in *Cars* from *SureSale* system.
	Size	Size of car. Values: "small", "medium", "large", "cargo". Taken from *Size* in *Car_Types* and *Car_type_ID* in *Cars* from *SureSale* system.
	Colour	Main colour or theme. Taken from *Colour* in *Cars* from *SureSale* system.
DT_Assessor	**Each tuple describes a claims adjuster evaluating the claim – a *NoLimit's* employee.**	
	ID_Assessor	Assessor ID. Surrogate key - generated by database.
	Name_and_Surname	Name and surname of the Assessor.Taken from *Name* and *Surname* in *Assessors* from *SureSale* system.
	Specialisation	The type of car the Assessor specialises in; Values: "cargo", "premium", "casual". Taken from *Specialisation* in *Assessors* from *SureSale* system.

DT_Date	**One tuple describes one day.** All the data in this table are generated tuple by tuple based on a calendar, before the ETL process. Range 1993/01/01, 2023/01/01.