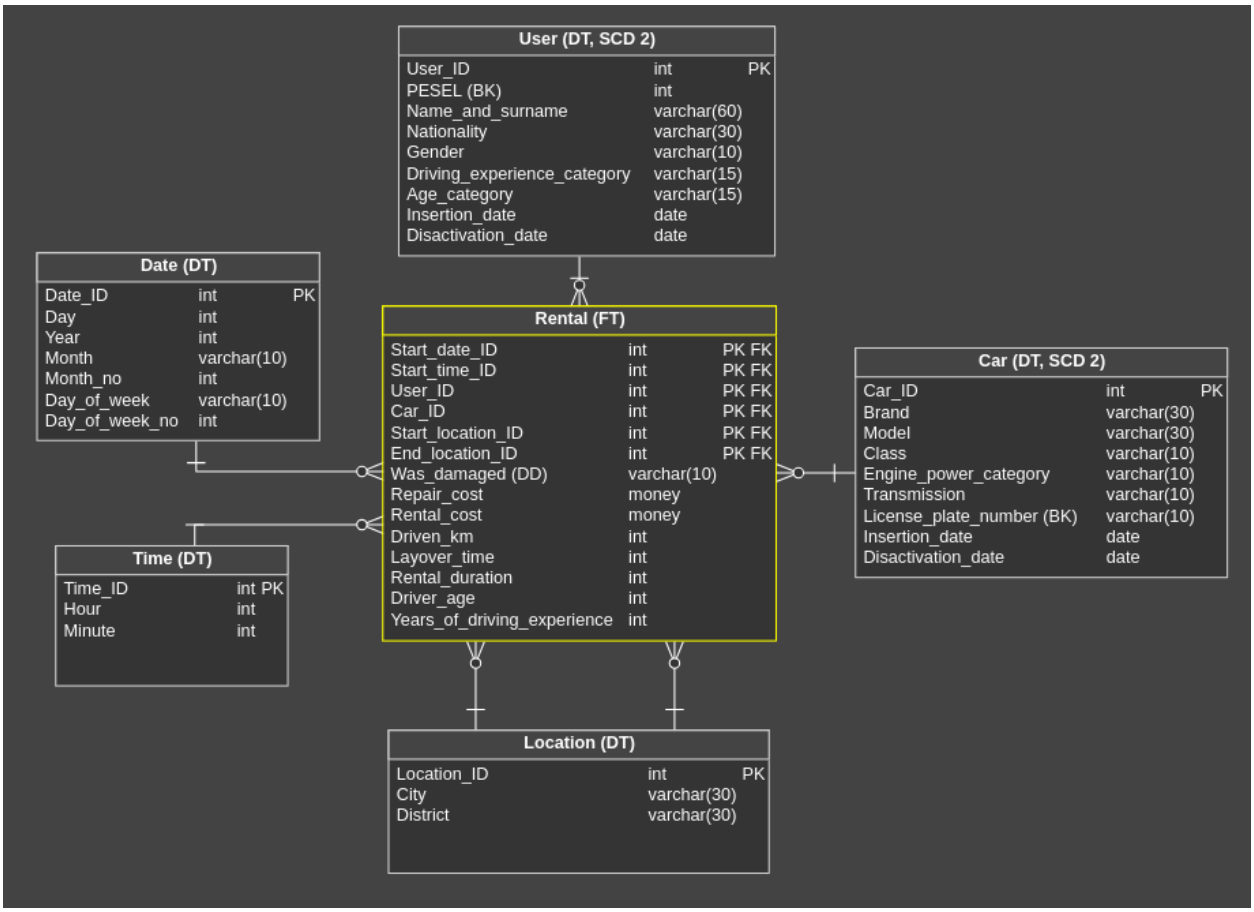
Traficar – Data warehouse design

Relational database schema



2.1. Table Rental (FT)

Description: One tuple describes one fact of rental.

2.1.1. Columns

Column name	Type	Properties	Description
Start_date_ID	int	PK	FK Date Rental start date
Start_time_ID	int	PK	FK Time Rental start time
User_ID	int	PK	FK User User
Car_ID	int	PK	FK Car Car
Start_location_ID	int	PK	FK Location Start location of the rental
End_location_ID	int	PK	FK Location End location of the rental
Was_damaged	varchar(12)	DD	Degenerated dimension determining if there was any damage made during rental. Possible values: Damaged, Undamaged
Repair_cost	money		Cost of damage reparation. 0 if there were no damage during rental
Rental_cost	money		Total cost of the rental (excluding repairing cost)
Driven_km	int		Driven km in the rental
Layover_time	int		Layover time
Rental_duration	int		Rental duration in minutes
Driver_age	int		Age of person driving rented vehicle
Years_of_driving_experience	int		Number of years driving person have their license

2.2. Table Time (DT)

Description: One tuple describes one minute (independently on date).

2.2.1. Columns

Column name	Type	Properties	Description
Time_ID	int	PK	PK (surrogate key)
Hour	int		Hour. Allowed values from 0 - 23
Minute	int		Minute. Allowed values from 0 - 59

2.3. Table User (DT, SCD 2)

Description: One tuple describes one user.

2.3.1. Columns

Column name	Type	Properties	Description
User_ID	int	PK	PK (surrogate key)
PESEL	int	BK	User's PESEL (11-digit number)
Name_and_surname	varchar(60)		User's first and last name
Nationality	varchar(30)		User's nationality
Gender	varchar(10)		User's gender. Allowed values: male, female
Driving_experience_category	varchar(15)		User driving experience category. Allowed values: Beginner, Intermediate, Experienced, Expert. 0 - 2 years: Beginner 2 – 5 years: Intermediate 5 – 10 years: Experienced 10+ years: Advanced
Age_category	varchar(15)		User age category. Allowed values: Young, Young adult, Adult, Elderly. 18 – 24: Young 25 – 34: Young adult 35 – 60: Adult 60+ Elderly
Insertion_date	date		Insertion date of the user (SCD2 implementation)
Disactivation_date	date		Deactivation date of the user (SCD2 implementation)

2.4. Table Car (DT, SCD 2)

Description: One tuple describes one car.

2.4.1. Columns

Column name	Type	Properties	Description
Car_ID	int	PK	PK (surrogate key)
Brand	varchar(30)		Brand of rental car
Model	varchar(30)		Model of rental car
Class	varchar(10)		Luxury class of car. Possible values: Basic, Comfort, Luxury
Engine_power_category	varchar(10)		Engine power of the car Allowed values: Small, Average, Big. < 110 hp: Small < 180hp: Average <= 180hp: Big
Transmission	varchar(10)		Determines if the car has manual or automatic

			transmission. Allowed values: Manual, Automatic.
License_plate_number (BK)	varchar(10)	BK	License plate number of the car
Insertion_date	date		Insertion date of the car (SCD2 implementation)
Inactive_date	date		Inactive date of the car (SCD2 implementation)

2.5. Table Location (DT)

Description: One tuple describes one location.

2.5.1. Columns

Column name	Type	Properties	Description
Location_ID	int	PK	PK (surrogate key)
City	varchar(30)		City, where the rentals has been made
District	varchar(30)		City's district where rentals has been made

2.6. Table Date (DT)

Description: One tuple describes one day.

2.6.1. Columns

Column name	Type	Properties	Description
Date_ID	int	PK	PK (surrogate key)
Day	int		Day
Year	int		Year
Month	varchar(10)		Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November and December
Month_no	int		Month's numeric value
Day_of_week	varchar(10)		Day of week Allowed values: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday
Day_of_week_no	int		Weekday's numeric value

Dimensional model

Fact definitions

Rental fact:

Rental of specific car by specific user on specific day at a specified time from specific location ended on specific location.

Fact table

Rental

Granularity

- A specified rented car
- A specified renting user
- A specified date of rental
- A specified time of rental
- A specified rental starting and ending locations

Measures and aggregation functions:

- Number of rental facts – COUNT (1)

- MAX Driven kilometers - MAX(Driven_km)
- MIN Driven kilometers - MIN(Driven_km)
- Driven kilometers - SUM(Driven_km)
- AVG Driven kilometers – Driven kilometers / Number of rental facts

- Layover time – SUM(Layover_time)
- AVG Layover Time - Layover time / Number of rental facts,

- MIN Rental Duration - MIN(Rental_duration)
- MAX Rental Duration - MAX(Rental_duration)
- Rental durations - SUM(Rental_duration)
- AVG Rental duration – Rental durations / Number of rental facts

- MIN Repair cost – MIN(Repair_cost)
- MAX Repair cost – MAX(Repair_cost)
- Repair costs - SUM(Repair_cost)
- AVG Repair cost – Repair costs / Number of rental facts

- MIN Rental cost – MIN(Rental_cost)
- MAX Rental cost – MAX(Rental_cost)
- Rental costs – SUM(Rental_cost)
- AVG Rental cost – Rental costs / Number of rental facts

- MIN Driver age – MIN(Driver_age)
- MAX Driver age – MAX(Driver_age)
- Driver ages – SUM(Driver_age)
- AVG Driver age – Driver ages / Number of rental facts

- MIN Years of driving experience – MIN(Years_of_driving_experience)
- MAX Years of driving experience – MAX(Years_of_driving_experience)
- Years of driving experience – SUM(Years_of_driving_experience)
- AVG Years of driving experience – Years of driving experience / Number of rental facts

Dimension definitions

Dimension/Dimension Attribute	Table / Column	TYPE
WAS DAMAGE	Rental.Was_damaged	Degenerate dimension
USER	User	Dimension
USER PESEL	User.PESEL	Dimension attribute
USERNAME AND SURNAME	User.Name_and_surname	Dimension attribute
USER NATIONALITY	User.Nationality	Dimension attribute
USER DRIVING EXPERIENCE CATEGORY	User.Driving_experience_category	Dimension attribute
USER AGE CATEGORY	User.Age_category	Dimension attribute
USER GENDER	User.Gender	Dimension attribute
USER HIERARCHY	• User.Age_category •• User.Gender ••• User.Nationality	Hierarchical dimension
CAR	Car	Dimension
CAR BRAND	Car.Brand	Dimension attribute
CAR MODEL	Car.Model	Dimension attribute
CAR CLASS	Car.Class	Dimension attribute
CAR ENGINE POWER CATEGORY	Car.Engine_power_category	Dimension attribute

CAR TRANSMISSION	Car.Transmission	Dimension attribute
CAR LICENSE PLATE NUMBER	Car.License_plate_number	Dimension attribute
RENTAL START LOCATION	Location	Dimension
RENTAL START CITY	Location.City	Dimension attribute
RENTAL START DISTRICT	Location.District	Dimension attribute
RENTAL END LOCATION	Location	Dimension
RENTAL END CITY	Location.City	Dimension attribute
RENTAL END DISTRICT	Location.District	Dimension attribute
RENTAL START LOCATION HIERARCHY	• Location.City •• Location.District	Hierarchical dimension
RENTAL END LOCATION HIERARCHY	• Location.City •• Location.District	Hierarchical dimension
RENTAL START DATE	Date	Dimension
RENTAL START YEAR	Date.Year	Dimension attribute
RENTAL START MONTH	Date.Month	Dimension attribute
RENTAL MONTH NO	Date.Month_no	Dimension attribute
RENTAL START DAY	Date.Day	Dimension attribute
RENTAL START DAY OF WEEK	Date.Day_of_week	Dimension attribute
RENTAL DAY OF WEEK NO	Date.Day_of_week_no	Dimension attribute
START DAY OF WEEK HIERARCHY	• Date.Year •• Date.Month ••• Date.Day_of_week	Hierarchical dimension
START DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Day	Hierarchical dimension
RENTAL START TIME	Time	Dimension
RENTAL START HOUR	Time.Hour	Dimension attribute
RENTAL START MINUTE	Time.Minute	Dimension attribute
START TIME HIERARCHY	• Time.Hour •• Time.Minute	Hierarchical dimension

Checking the feasibility of queries based on the multidimensional model

1. Compare the popularity of available cars in terms of the number of rentals over the past year.

Measure – Number of rental facts

Dimension – Car (dimension attributes: Car model)

Dimension – Rental start date (dimension attributes: Rental start year)

2. Which car models generate the highest revenue per vehicle?

Measure – Rental costs

Dimension – Car (dimension attributes: Car Model)

3. Compare the earnings of cars by grouping them according to engine power and type of transmission.

Measure – Rental costs

Dimension – Car (dimension attributes: Car transmission, Car engine power category)

4. Compare which cars and which class of cars were most frequently used in different city regions.

Measure – Number of rental facts

Dimension – Car (dimension attributes: Car model, Car class)

Dimension – Rental start location (dimension attributes: Rental start district)

Dimension – Rental end location (dimension attributes: Rental end district)

5. Create a ranking of the 5 car models with the highest total mileage driven by customers in the last year.

Measure – Driven kilometers

Dimension – Car (dimension attributes: Car model)

Dimension – Rental start date (dimension attributes: Rental start year)

6. Create a ranking of age groups responsible for causing the most damage over the past year.

Measure – Number of rental facts

Degenerated Dimension – Was damaged

Dimension – User (dimension attributes: User age category)

Dimension – Rental start date (dimension attributes: Rental start year)

7. Create a ranking of driving experience levels that resulted in the highest number of accidents over the past year.

Measure – Number of rental facts

Degenerated Dimension – Was damaged

Dimension – User (dimension attributes: User driving experience category)

Dimension – Rental start date (dimension attributes: Rental start year)

8. What is the average total number of rides for users who caused damage compared to those who never caused damage?

Measure – Number of rental facts

Degenerated Dimension – Was damaged

Dimension – User (dimension attribute: User ID)

9. Considering the number of rides taken and the number of reported damages, what is the theoretical probability of causing damage within each age group of users?

Measure – Number of rental facts

Degenerated Dimension – Was damaged

Dimension – User (dimension attribute: User age category)

10. What is the average cost of repairs caused by users within different levels of driving experience?

Measure – AVG Repair cost

Degenerated Dimension – Was damaged

Dimension – User (dimension attributes: User driving experience category)

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

Rental	**One tuple describes one fact of car rental.**	
	Start_date_ID	Rental start date Id. Foreign key from dimension table. Based on date from Rental_date_start stored in Rentals table in Traficar database source.
	Start_time_ID	Rental start time Id. Foreign key from dimension table. Based on time from Rental_date_start stored in Rentals table in Traficar database source.
	User_ID	User renting the car Id. Foreign key from dimension table. Based on User_ID FK stored in Rentals table in Traficar database source.
	Car_ID	Rented car Id. Foreign key from dimension table. Based on Car_ID FK stored in CarStates table, to which Rentals table have CarStates_ID FK in Traficar database source.
	Start_location_ID	Rental start location Id. Foreign key from dimension table. Based on Start_location stored in

		Rentals table in Traficar database source.
	End_location_ID	Rental end location Id. Foreign key from dimension table. Based on End_location stored in Rentals table in Traficar database source.
	Was_damaged (DD)	Determining if there was any accident during rental. We check if there was any damage during rental by checking if there is Rental_ID in Damage excel sheet.
	Repair_cost	Cost of repairs. If no damage was done during rental, the value is 0. If any damage was made, Repair_cost is taken from Damage repair cost column from Damage excel sheet. We take Repair_cost based on ID of rental on which damage was made column from Damage excel sheet.
	Rental_cost	Total cost of rental (excluding potential repair costs). Taken from Total_cost in Rentals table in Traficar database.
	Driven_km	Kilometers driven during rental. Taken from Driven_km in Rentals table in Traficar database.
	Layover_time	Layover time during rental. Taken from Layover_time in Rentals table in Traficar database.
	Rental_duration	Rental duration. Calculated using Rental_date_start and Rental_date_end from Rentals table in Traficar database.

	Driver_age	User's age calculated from PESEL in User table in Traficar database.
	Years_of_driving_experience	User's driving experience calculated from License_receiving_date in User table in Traficar database.
User	**One tuple describes one user.**	
	User_ID	User ID based on User_ID in User table in Traficar database.
	PESEL	User's PESEL taken from PESEL in Users table in Traficar database.
	Name_and_surname	User's first name and last_name taken from First_name and Last_name in User table in Traficar database.
	Nationality	User's nationality taken from Nationality in User table in Traficar database.
	Gender	User's gender calculated from User's PESEL from PESEL in User table in Traficar database.
	Driving_experience_category	Driving experience category. Allowed values: 0 - 2 years: Beginner 2– 10 years: Experienced 10+ years: Advanced Years of having license calculated from License_reveiving_date in Users table in Traficar database.
	Age_category	User age category. Allowed values: 18 – 24: Young 25 – 34: Young adult 35 – 60: Adult

		60+ Elderly User age calculated from User's PESEL from PESEL in User table in Traficar database.
	Insertion_date	Date of insertion of User (SCD2 implementation).
	Disactivaton_date	Date of insertion same user but with changed Driving_experience_category and Age_category. Previous user inactive_date = new user insertion_date (SCD2 implementation).
Car	**One tuple describes one car.**	
	Car_ID	Car ID based on Car_ID in Cars table in Traficar database.
	Brand	Car brand taken from Brand in Cars table in Traficar database.
	Model	Car model taken from Model in Cars table in Traficar database.
	Class	Luxury class of car. Possible values: Basic, Comfort, Luxury. Car class taken from Traficar database based on pricelists the car had.
	Engine_power_category	Engine power category. Allowed values: Power < 110: Small, Power < 180: Average, Power >= 180: Big. Car power taken from Power in Cars table in Traficar database.
	Transmission	Car transmission type. Allowed values: Automatic, Manual. Transmission type based on

		Manual in Cars table in Traficar database.
	License_plate_number	Car license plate number taken from License_plate_number in Cars table in Traficar database.
	Insertion_date	Date of insertion of Car (SCD2 implementation).
	Disactivation_date	Date of insertion same car but with changed license plate. Previous car inactive_date = new car insertion_date (SCD2 implementation).
Location	**One tuple describes one location.**	
	Location_ID	Location ID. Surrogate key created automatically by database.
	City	Calculated from coordinates taken from Rental_date_start and Rental_date_end in Rentals table in Traficar database.
	District	Calculated from coordinates taken from Rental_date_start and Rental_date_end in Rentals table in Traficar database.
Date	**One tuple describes one day.** All the data in this table is generated tuple by tuple based on any calendar, before ETL process.	
Time	**One tuple describes one minute (independently of date).** All the data in this table are generated tuple by tuple based on clock, before ETL process.	