

# “Thunder Rental” – Data warehouse design

The Data warehouse is designed for the scooter rental business process. This process is described in the document “Requirements specification for “Rental of the electric scooters” business process”.

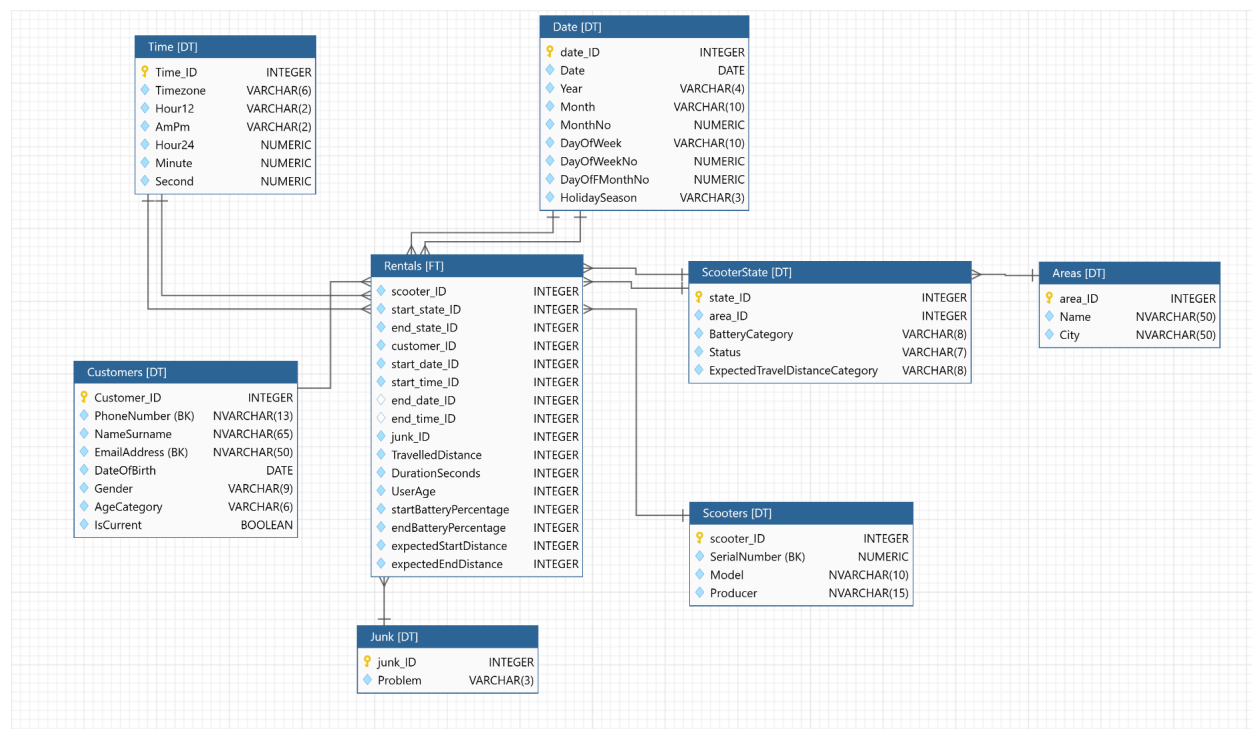


TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
<b>Rentals (FACT TABLE)</b>	<b>One tuple describes one fact of rental</b>		
	scooter_ID	INTEGER	FK Scooters ID of the scooter
	start_state_ID	INTEGER	FK StateEvent ID of the start state of the rental
	end_state_ID	INTEGER	FK EndEvent ID of the end state of the rental
	customer_ID	INTEGER	FK Customers ID of the customer
	start_date_ID	INTEGER	FK Date

			ID of the date at the start of the rental
	start_time_ID	INTEGER	FK Time ID of the time at the start of the rental
	end_date_ID	INTEGER	FK Date ID of the date at the end of the rental
	end_time_ID	INTEGER	FK Time ID of the time at the end of the rental
	junk_ID	INTEGER	FK Junk ID of the junk
	TravelledDistance	INTEGER	Distance traveled on a scooter by customer during the rental shown in meters
	DurationSeconds	INTEGER	Duration of the rental (in seconds)
	UserAge	INTEGER	The age of the customer that made the rental
	StartBatteryPercentage	INTEGER	Remaining battery percentage of the scooter at the start of the rental
	EndBatteryPercentage	INTEGER	Remaining battery percentage of the scooter at the end of the rental
	ExpectedStartDistance	INTEGER	Expected travel distance remaining on the scooter at the start of the rental
	ExpectedEndDistance	INTEGER	Expected travel distance remaining on the scooter at the end of the rental
<b>Junk (DIMENSION TABLE)</b>	<b>The tuples correspond to "all" possible values for Problem</b>		
	junk_ID	INTEGER	PK ( <i>surrogate key</i> )
	Problem	VARCHAR(3)	"YES" if there was a problem during the rental and it ended because of that problem, "NO" if not
<b>Customers (DIMENSION TABLE)</b>	<b>One tuple describes one customer, in the specified age category, with the specified phone number, email address, date of birth and gender</b>		
	customer_ID	INTEGER	PK ( <i>surrogate key</i> )
	PhoneNumber	NVARCHAR(13)	Phone number of the customer, unique, only one account can be created with a specific phone number. Format : at the beginning 2

			or 3 digits associating number with a country, then 9 or 10 digit numbers depending on a country ( <i>business key</i> )
	NameSurname	NVARCHAR(65)	Name and surname of the customer
	EmailAddress	NVARCHAR(50)	Email address of the customer, unique, only one account can be created with a specific email address ( <i>business key</i> )
	DateOfBirth	DATE	Date of birth of the customer
	Gender	VARCHAR(9)	Gender of customer. Allowed values: "Male", "Female" or "Different"
	AgeCategory	VARCHAR(6)	Age category of the customer Allowed values: 18-25, 26-35, 36-45, 46-55, 56+
	IsCurrent	BOOLEAN	1 if information is current, otherwise 0 (SCD2 implementation)
<b>Scooters (DIMENSION TABLE)</b>	<b>One tuple describes one scooter with its serial number, model and producer</b>		
	scooter_ID	INTEGER	PK ( <i>surrogate key</i> )
	SerialNumber	NUMERIC	Serial number of the vehicle ( <i>business key</i> )
	Model	NVARCHAR(10)	The specific model of the scooter
	Producer	NVARCHAR(15)	The producer of the specific scooter
<b>ScooterState (DIMENSION TABLE)</b>	<b>One tuple describes one scooter's state with its status, battery percentage category, expected travel distance category, and area. New ScooterState is generated when the status is changed</b>		
	state_ID	INTEGER	PK ( <i>surrogate key</i> )
	area_ID	INTEGER	FK Areas ID of the area
	BatteryCategory	VARCHAR(8)	Category of the remaining battery percentage of the scooter at the start of the state Allowed values: if 0%-20% "VeryLow", if 21%-40% "Low", if 41%-60% "Medium", if 61%-80% "High", if 81%-100% "VeryHigh"

	Status	VARCHAR(7)	Status of the scooter at the start of the state Allowed values: Free, Rented and Service
	ExpectedTravelDistanceCategory	VARCHAR(8)	Category of expected travel distance remaining on the scooter at the start of the state Allowed values: if 0m-5000m "VeryLow", if 5001m-10000m "Low", if 10001m-15000m "Medium", if 15001m-20000m "High", if 20001m-25000m "VeryHigh"
<b>Areas (DIMENSION TABLE)</b>	<b>One tuple describes one area with the city in which the area is located</b>		
	area_ID	INTEGER	PK ( <i>surrogate key</i> )
	Name	NVARCHAR(50)	Name of the area
	City	NVARCHAR(50)	City in which the area is located
<b>Date (DIMENSION TABLE)</b>	<b>One tuple describes one day</b>		
	date_ID	INTEGER	PK ( <i>surrogate key</i> )
	Date	DATE	Date
	Year	VARCHAR(4)	Year
	Month	VARCHAR(10)	Month Allowed values: January, February, March, April, May, June, July, August, September, October, November and December
	MonthNo	NUMERIC	Month's numeric value
	DayOfWeek	VARCHAR(10)	Day of week Allowed values: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday
	DayOfWeekNo	NUMERIC	Weekday's numeric value
	DayOfMonthNo	NUMERIC	Numerical value of the day of the month
	HolidaySeason	VARCHAR(3)	"YES" if this is holiday season, "NO" if not

Time (DIMENSION TABLE)	One tuple describes one unit of time with accuracy to seconds (independently on date)		
	time_ID	INTEGER	PK ( <i>surrogate key</i> )
	Timezone	VARCHAR(7)	Determines the time zone of described unit of time in UTC+ or UTC- format
	Hour12	VARCHAR(3)	Hour Allowed values from 1 – 12
	AmPm	VARCHAR(3)	“AM” if the hour is before noon and “PM” if after noon
	Hour24	NUMERIC	Hour Allowed values from 0 – 23
	Minute	NUMERIC	Minute Allowed values from 0 – 59
	Second	NUMERIC	Second Allowed values from 0 – 59

## Dimensional model

### Fact 1

Rental fact: Rental of the scooter (particular scooter producer, particular model, particular vehicle). Rental performed by the user with the specified phone number, email address, age, gender, name and surname. Rented using the mobile app. Rental started at the specified day and time, a particular state of the vehicle (specific battery status, specific location, specific status, specific expected travel distance). Rental ended with a particular state of the vehicle (specific battery status, specific location, specific status, specific expected travel distance). Each rental resulted in a specified travel distance.

### Fact table: Rentals

Granularity:

- a specified rental,
- a specified scooter,
- a specified start vehicle state,
- a specified end vehicle state,
- a specified date and time of the start of the rental,
- a specified date and time of the end of the rental,
- a specified traveled distance during the rental,
- a specified customer in the specified age, with the specified gender.

### Measures and aggregation functions:

Number of rental facts – COUNT (1)

Number of users - COUNT DISTINCT (PhoneNumber)

Number of scooters - COUNT DISTINCT (scooter\_ID)

Total travel distance - SUM (TravelledDistance)

Total duration of rentals - SUM (Duration)

Sum of battery percentage - SUM (batteryPercentage)  
Sum of expected travel distance - SUM (expectedDistance)  
Average battery percentage - "Sum of battery percentage"/"Number of rental facts"  
Average number of rentals per user - "Number of rental facts"/"Number of users"  
Average expected travel distance - "Sum of expected travel distance"/"Number of rentals facts"  
Average travel distance - "Total travel distance"/"Number of rental facts"  
Average rental duration - "Total duration of rentals"/"Number of rental facts"  
Percentage of rentals - "Number of rental facts"/"Number of rental facts"

DIMENSION / DIMENSION ATTRIBUTE	TABLE / COLUMN	TYPE
<b>Scooter start state</b>	<b>ScooterState</b>	<b>Dimension</b>
Battery level category	ScooterState.BatteryCategory	Dimension attribute
Vehicle status	ScooterState.Status	Dimension attribute
Expected travel distance left category	ScooterState.ExpectedTravelDistanceCategory	Dimension attribute
<b>Scooter end state</b>	<b>ScooterState</b>	<b>Dimension</b>
Battery level category	ScooterState.BatteryCategory	Dimension attribute
Vehicle status	ScooterState.Status	Dimension attribute
Expected travel distance left category	ScooterState.ExpectedTravelDistanceCategory	Dimension attribute
<b>Scooter</b>	<b>Scooters</b>	<b>Dimension</b>
Scooter identifier	Scooters.SerialNumber	Dimension attribute
Scooter producer	Scooters.Producer	Dimension attribute
Scooter model	Scooters.Model	Dimension attribute
<b>Start area</b>	<b>Areas</b>	<b>Dimension</b>
Area name	Areas.Name	Dimension attribute
City name	Areas.City	Dimension attribute
Start area hierarchy	<ul style="list-style-type: none"> <li>• Areas.City</li> <li>•• Areas.Name</li> </ul>	Hierarchical Dimension
<b>End area</b>	<b>Areas</b>	<b>Dimension</b>
Area name	Areas.Name	Dimension attribute
City name	Areas.City	Dimension attribute
End area hierarchy	<ul style="list-style-type: none"> <li>• Areas.City</li> </ul>	Hierarchical

	●● Areas.Name	Dimension
<b>Customer</b>	<b>Customers</b>	<b>Dimension</b>
Phone number	Customer.PhoneNumber	Dimension attribute
Name and surname	Customer.NameSurname	Dimension attribute
Email address	Customer.EmailAddress	Dimension attribute
Date of birth	Customer.DateOfBirth	Dimension attribute
Gender	Customer.Gender	Dimension attribute
Age category	Customer.AgeCategory	Dimension attribute
<b>Rental start date</b>	<b>Date</b>	<b>Dimension</b>
Year	Date.Year	Dimension attribute
Month	Date.Month	Dimension attribute
Month number	Date.MonthNo	Dimension attribute
Day of week	Date.DayOfWeek	Dimension attribute
Day of week number	Date.DayOfWeekNo	Dimension attribute
Day of month number	Date.DayOfMonthNo	Dimension attribute
Holiday season day	Date.HolidaySeason	Dimension attribute
<b>Rental end date</b>	<b>Date</b>	<b>Dimension</b>
Year	Date.Year	Dimension attribute
Month	Date.Month	Dimension attribute
Month number	Date.MonthNo	Dimension attribute
Day of week	Date.DayOfWeek	Dimension attribute
Day of week number	Date.DayOfWeekNo	Dimension attribute
Day of month number	Date.DayOfMonthNo	Dimension attribute
Holiday season day	Date.HolidaySeason	Dimension attribute
Rental start date hierarchy	<ul style="list-style-type: none"> <li>● Date.Year</li> <li>●● Date.Month</li> <li>●●● Date.DayOfMonthNo</li> </ul>	Hierarchical dimension
Rental end date hierarchy	<ul style="list-style-type: none"> <li>● Date.Year</li> <li>●● Date.Month</li> <li>●●● Date.DayOfMonthNo</li> </ul>	Hierarchical dimension

<b>Rental Start Time</b>	<b>Time</b>	<b>Dimension</b>
Timezone	Time.Timezone	Dimension attribute
Hour in 12h format	Time.Hour12	Dimension attribute
AM/PM signature for 12h time format	Time.AmPm	Dimension attribute
Hour in 24h format	Time.Hour24	Dimension attribute
Minute	Time.Minute	Dimension attribute
Second	Time.Second	Dimension attribute
<b>Rental End Time</b>	<b>Time</b>	<b>Dimension</b>
Timezone	Time.Timezone	Dimension attribute
Hour in 12h format	Time.Hour12	Dimension attribute
AM/PM signature for 12h time format	Time.AmPm	Dimension attribute
Hour in 24h format	Time.Hour24	Dimension attribute
Minute	Time.Minute	Dimension attribute
Second	Time.Second	Dimension attribute
Rental start time hierarchy	<ul style="list-style-type: none"> <li>● Time.Timezone</li> <li>●● Time.Hour24</li> <li>●●● Time.Minute</li> <li>●●●● Time.Second</li> </ul>	Hierarchical Dimension
Rental end time hierarchy	<ul style="list-style-type: none"> <li>● Time.Timezone</li> <li>●● Time.Hour24</li> <li>●●● Time.Minute</li> <li>●●●● Time.Second</li> </ul>	Hierarchical Dimension
<b>Junk</b>	<b>Junk</b>	<b>Dimension</b>
Problem	Junk.Problem	Dimension attribute

## Checking the feasibility of queries based on the multidimensional model

1. Which cities see the highest demand for scooter rentals for the analyzed month compared to the previous month?

*Measure: Number of rental facts*

*Dimension: Area (dimension attribute: City name)*

*Dimension: Date (dimension attribute: Month, Month number)*



2. What is the number of rentals per age group for the analyzed month and for the previous month?  
*Measure: Number of rental facts*  
*Dimension: Customer (dimension attribute: Age group)*  
*Dimension: Date (dimension attribute: Month, Month number)*
3. Give the average number of rentals per user, per month.  
*Measure: Average number of rentals per user*  
*Dimension: Date (dimension attribute: Month, Month number)*
4. Give the average battery life remaining on a scooter at the start of each rental in the current and previous month.  
*Measure: Average battery percentage*  
*Dimension: Date (dimension attribute: Month, Month number)*
5. Compare the percentage distribution of rentals depending on gender for the current month with the percentage distribution for the previous month.  
*Measure: Percentage of rentals*  
*Dimension: Date (dimension attribute: Month, Month number)*  
*Dimension: Customer (dimension attribute: Gender)*
6. Compare the average expected travel distance at the end of the rental between cities.  
*Measure: Average expected travel distance*  
*Dimension: Area (dimension attribute: City name)*  
*Dimension: Date (dimension attribute: Month, Month number)*
7. What is the average rental duration this month compared to the previous month?  
*Measure: Average rental duration*  
*Dimension: Date (dimension attribute: Month, Month number)*
8. Give the average battery life remaining on scooters at the end of each rental.  
*Measure: Average battery percentage*  
*Dimension: ScooterState (Status)*
9. What is the average travel distance per rental from each month for the last year?  
*Measure: Average travel distance*  
*Dimension: Date (dimension attribute: Month, Month number, Year)*
10. Which day of the week has the highest scooter usage for the last month?  
*Measure: Number of rentals*  
*Dimension: Date (dimension attribute: Month, Month number, DayOfWeek, DayOfWeekNo)*

11. How many rentals does a scooter have on average per day in the current and previous month?  
*Measure: Number of rentals*  
*Dimension: Date (dimension attribute: Month, Month number, DayOfMonthNo)*
12. Compare the share of rentals between various models of scooters for the last year.  
*Measure: Percentage of rentals*  
*Dimension: Scooter (dimension attribute: Scooter producer, Scooter model)*
13. Compare the number of rentals between various areas in cities.  
*Measure: Number of rental facts*  
*Dimension: Area (dimension attribute: City name, Area name)*

## Checking if there are data in data sources needed to fill the data warehouse

TABLE NAME	COLUMN	SOURCE
<b>Rentals (FACT TABLE)</b>	<b>One tuple describes one fact of rental</b>	
	scooter_ID	Scooter ID. Foreign key from dimension table. Based on ID, model and producer stored in the Scooters table in THE MOBILE APP DATABASE source.
	start_state_ID	Start state ID. Foreign key from dimension table. Based on Scooter ID, timestamp and status (columns A, B and D) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	end_state_ID	End state ID. Foreign key from dimension table. Based on Scooter ID, timestamp and status (columns A, B and D) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	customer_ID	Customer ID. Foreign key from dimension table. Based on Phone number stored in the Customers table in THE MOBILE APP DATABASE source.
	start_date_ID	Start date ID. Foreign key from dimension table. Based on timestamp (column B) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	start_time_ID	Start time ID. Foreign key from dimension table. Based on timestamp (column B) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.

	end_date_ID	End date ID. Foreign key from dimension table. Based on timestamp (column B) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	end_time_ID	End time ID. Foreign key from dimension table. Based on timestamp (column B) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	junk_ID	Junk ID. Foreign key from dimension table. Based on Problem stored in the Rentals table in THE MOBILE APP DATABASE source.
	TravelledDistance	Travelled distance taken from Travelled distance column in Rentals table in THE MOBILE APP DATABASE source.
	DurationSeconds	Duration of the rental taken from End timestamp and Start timestamp columns stored in the Rentals table in THE MOBILE APP DATABASE source.
	UserAge	User age in years taken from Birth date column in Customers table in THE MOBILE APP DATABASE source.
	StartBatteryPercentage	Start battery percentage of the scooter taken from Battery percentage (column C) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	EndBatteryPercentage	End battery percentage of the scooter taken from Battery percentage (column C) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	ExpectedStartDistance	Start expected travel distance of the scooter taken from Expected travel distance left (column E) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	ExpectedEndDistance	End expected travel distance of the scooter taken from Expected travel distance left (column E) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
<b>Junk (DIMENSION TABLE)</b>	<b>The tuples correspond to "all" possible values for Problem</b>  The tuples correspond to "all" possible values for Problem and are generated before ETL process.	
	junk_ID	Junk ID. Surrogate key - generated by database.
<b>Customers (DIMENSION TABLE)</b>	<b>One tuple describes one customer, in the specified age category, with the specified phone number, email address, date of birth and gender. (Implementation of SCD 2)</b>	
	customer_ID	Customer ID. Surrogate key - generated by the database.

	PhoneNumber	Phone number of the customer. Business key taken from Phone number column in the Customers table in THE MOBILE APP DATABASE source.
	NameSurname	Name and surname of the customer. Taken from Name and Surname columns in the Customers table in THE MOBILE APP DATABASE source.
	EmailAddress	Email address of the customer. Business key taken from Email address column in the Customers table in THE MOBILE APP DATABASE source.
	DateOfBirth	Birthdate of the customer. Taken from Birth date column in the Customers table in THE MOBILE APP DATABASE source.
	Gender	Gender of the customer. Taken from Gender column in the Customers table in THE MOBILE APP DATABASE source.
	AgeCategory	Age category that customer falls in. Based on the Birth date column stored in the Customers table in THE MOBILE APP DATABASE source.
	IsCurrent	1 if information is current, otherwise 0 (SCD2 implementation)
<b>Scooters (DIMENSION TABLE)</b>	<b>One tuple describes one scooter with its serial number, model and producer</b>	
	scooter_ID	Scooter ID. Surrogate key - generated by database.
	SerialNumber	Serial number of the vehicle. Business key taken from ID column in the Scooters table in THE MOBILE APP DATABASE source.
	Model	The specific model of the scooter. Taken from Model column in the Scooters table in THE MOBILE APP DATABASE source.
	Producer	The producer of the specific scooter. Taken from Producer column in the Scooters table in THE MOBILE APP DATABASE source.
<b>ScooterState (DIMENSION TABLE)</b>	<b>One tuple describes one scooter's state with its status, battery percentage category, expected travel distance category, and area. New ScooterState is generated when the status is changed</b>	
	state_ID	State ID. Surrogate key - generated by database.
	area_ID	Area ID. Foreign key from dimension table. Based on location latitude, location longitude (columns F and G) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source and city taken from City column in

		the Rentals table in THE MOBILE APP DATABASE source.
	BatteryCategory	Category of the battery level of the scooter taken from Battery percentage (column C) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	Status	Status of the scooter taken from Status (column D) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	ExpectedTravelDistanceCategory	Category of the expected travel distance remaining of the scooter taken from Expected travel distance left (column E) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
<b>Areas (DIMENSION TABLE)</b>	<b>One tuple describes one area with the city in which the area is located</b>	
	area_ID	Area ID. Surrogate key - generated by the database.
	Name	Name of the area. Based on Location latitude and Location longitude (columns F and G) stored in SCOOTER MANAGEMENT SYSTEM REPORTS source.
	City	City in which the area is located. Taken from City column in the Rentals table in THE MOBILE APP DATABASE source.
<b>Date (DIMENSION TABLE)</b>	<b>One tuple describes one day</b>  All the data in this table is generated tuple by tuple based on the timestamp, before the ETL process.	
<b>Time (DIMENSION TABLE)</b>	<b>One tuple describes one unit of time with accuracy to seconds (independently on date).</b>  All the data in this table is generated tuple by tuple based on the timestamp, before the ETL process.	