Task: design a data warehouse snowflake schema for analyzing air pollution

Problem description

Air pollution is monitored in multiple cities in a country. The monitoring is performed by *stations*. A station is characterized by a unique identifier and name. A station is located in a district - in a given district only one station exists. A city includes multiple districts. Every district is characterized by its name, area (in m^2), and a number of inhabitants. A city is characterized by a name, a number of inhabitants, and a binary information whether a heavy industry is located in the city.

A station measures the following types of pollution: PM2.5, PM10, ozone, dioxide. Each measurement is provided by a dedicated sensor (uniquely identified by its 10-digits number). A measurement is taken every 60sec by every sensor.

Additionally, whether stations provide information on:

- air humidity (in %),
- wind speed (in m/s),
- air temperature,
- sky condition (cloudy, partly cloudy, sunny),
- falls (rain, snow).

The data are measured every 30sec.

Expected functionality

The designed data warehouse has to provide means for the following analyses.

- Drawing a chart with average daily pollution within a given time period for a given district and for a given pollution type.
- For a given district finding a time moment with maximum pollution of a given type.
- For a given city finding a district with the highest daily average pollution of a PM10.
- Finding a district with maximum pollution of dioxide within the last year in sunny days.