BUSINESS INTELLIGENCE REPORTING USING COGNOS

Problem 1

PROBLEM 1

3) **Measurable fields and Dimensions**:

The dataset contains the temperature parameters provided by the weather stations in Brazil. The weather parameters are used by meteorological department to provide weather forecasting[1].

The measured fields are Weather Station Identifier, Humidity, Air temperature, Air Pressure, Wind Speed, Wind Direction solar radiation, wind gust and due point temperature. I select them as the measurable fields as they are recorded using corresponding instruments. Like we need a thermometer or some kind of device to collect the readings of the temperature. Barometer for pressure and so on.

While details like location of the weather station, its coordinates and province, state are dimensions with respect to the Weather station. Also, the time of recording is further modularized to date, day, year, month and hour. So single timestamp is measurable field while the modularized fields are dimensions with respect to the recorded time intervals.

These various parameters add dimension to the data. Fetching the temperature parameters based on a location. Fetching the readings based on the time interval. Fetching details based on time interval in a location etc. Since different instruments are used to read various parameters of the weather station, it is good to normalize it into separate table. Fields can be clubbed with a generated id to fetch details based on specific temperature or pressure in the future.

As Business Intelligence focus more on answering questions to the data, the data needs to be clean, organized and accurate.

4) Cleaning dataset

- mdct comprises of the fields like date, hour, month, year and day. Hence only the mdct is kept in the schema and other fields are moved to separate schema for dimensional purpose.
- Deleting rows where the readings are 0. As those are inefficient and provides no information.(51,059 entries)
- We can replace the empty string columns to -1 as it is an invalid reading and can be ignored in future data analytics. However, places where there is no due point temperature , it can be calculated with the help of air temperature and relative humidity.[2]
- Since separate instruments are needed to measure the various attributes at the weather station in a given time,
    - a new schema for air pressure which clubs stp, stpmin, stpmax, mdct wsid.
    - a new schema for air temperature which club temp, tmax, tmin, mdct, wsid.
    - a new schema for relative humidity which club hmd, hmax, hmin, wsid, mdct.
    - a new schema for wind parameters which club wdsp, wdct, gusp, wsid, mdct.
    - a new schema for dew point temperature which club dewp, dmax, dmin.

6) **Star Schema:**

The model clearly makes use of the meaningful schema created in the cleaning process. The various instruments readings at a particular time in weather station is mapped to a uniquely generated id. This generated id will be the primary key of those schema. This in turns become the foreign key of the fact table which resides at the centre. The fact table consists of data that are measured. So, visualizing the fact data in various aspects is isolated and stored in the dimensional table.

By having the model this way, the comparison on a particular time frame at a particular location based on air temperature, wind, air pressure, humidity, due point can be analysed using the slice and dice method.

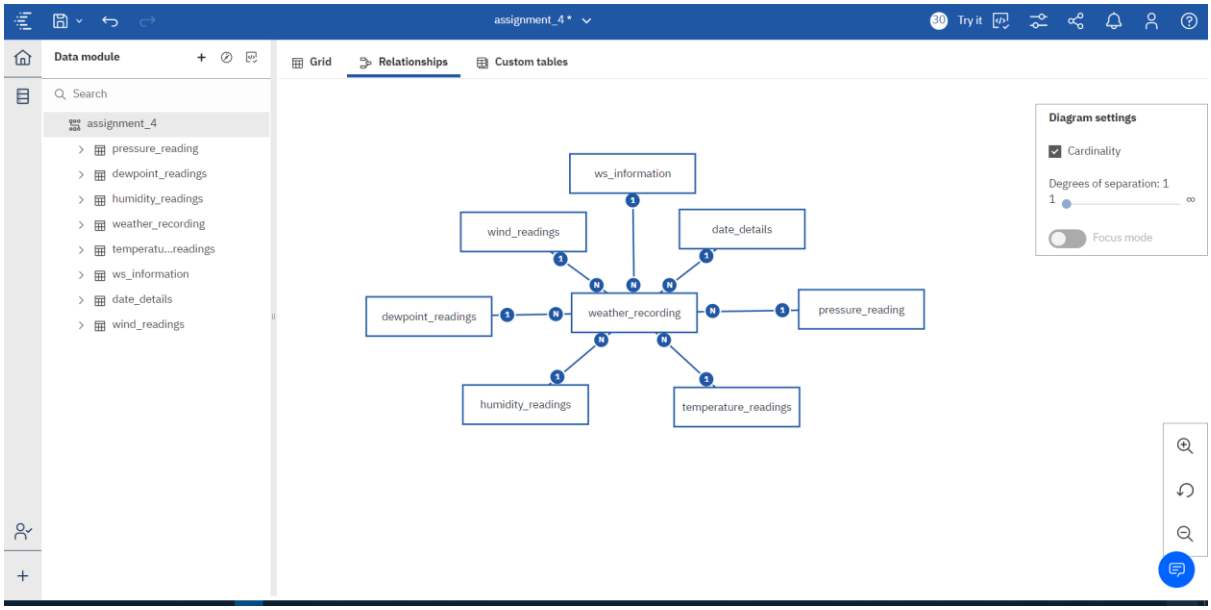

Figure 1 Star schema

This model can be further breaking down into snowflakes schema by isolating the location details of the weather station and temperature readings etc. Since the data is more clearly organized in this model, it can answer questions that can help the meteorological department to make good decisions.

If proper analysis is made using a machine learning algorithm in the data, a perfect weather forecast can be made. That comes under Data Analytics stage.

8) Visualizations:

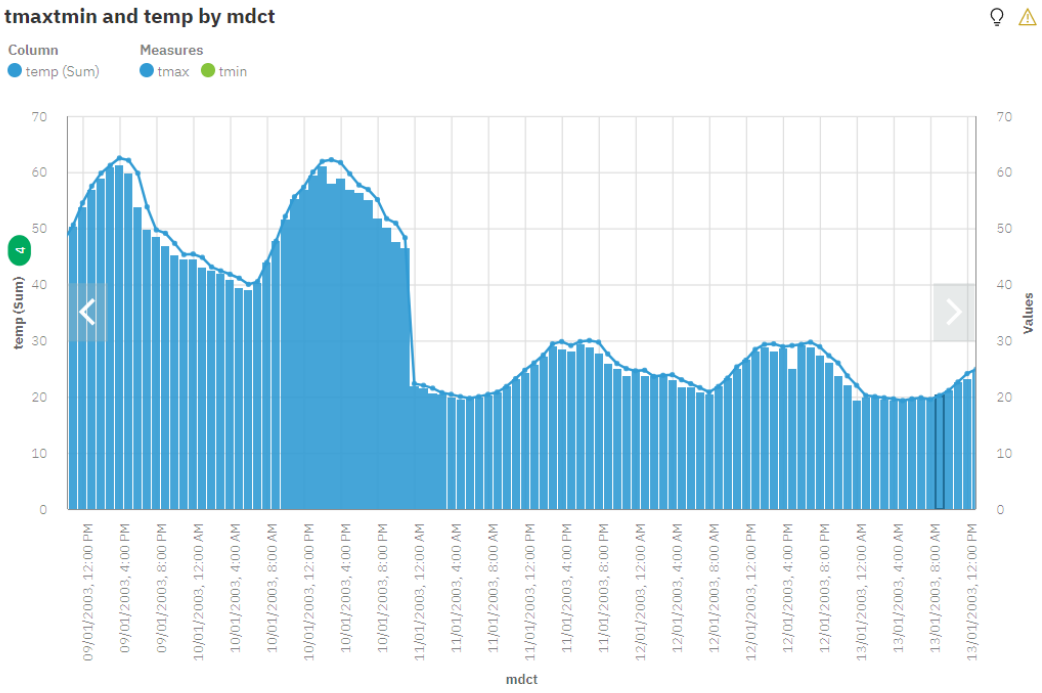

Figure 2 Above is a visualization based on air temperature in various intervals of time.

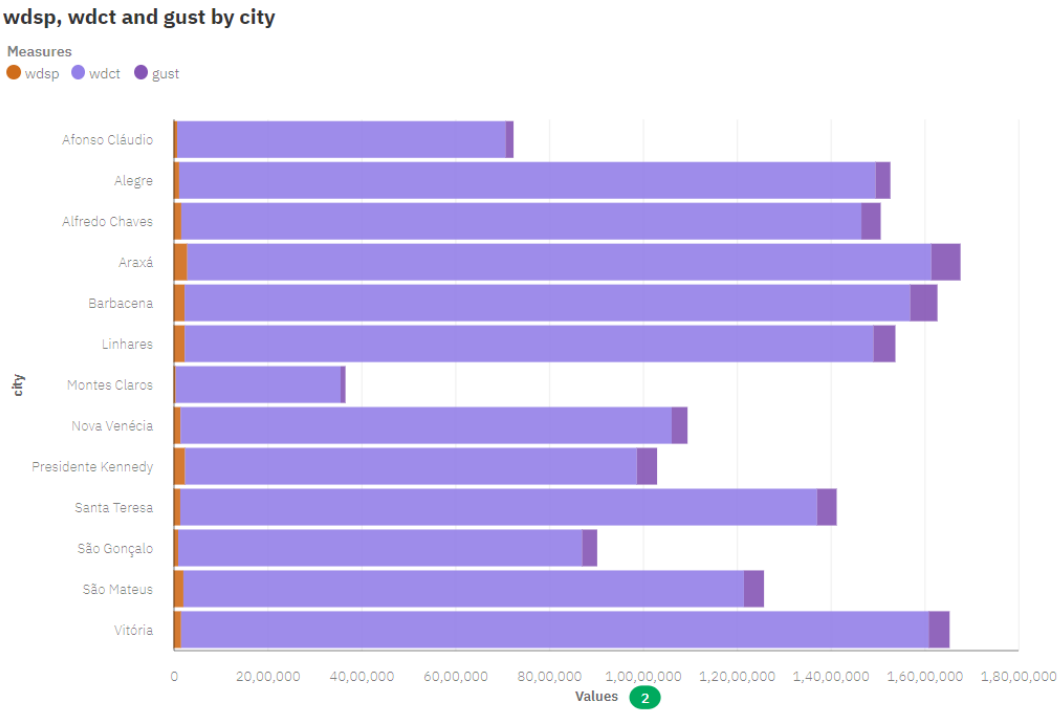

Figure 3 Above is a data visualization based on the wind readings in a particular city.

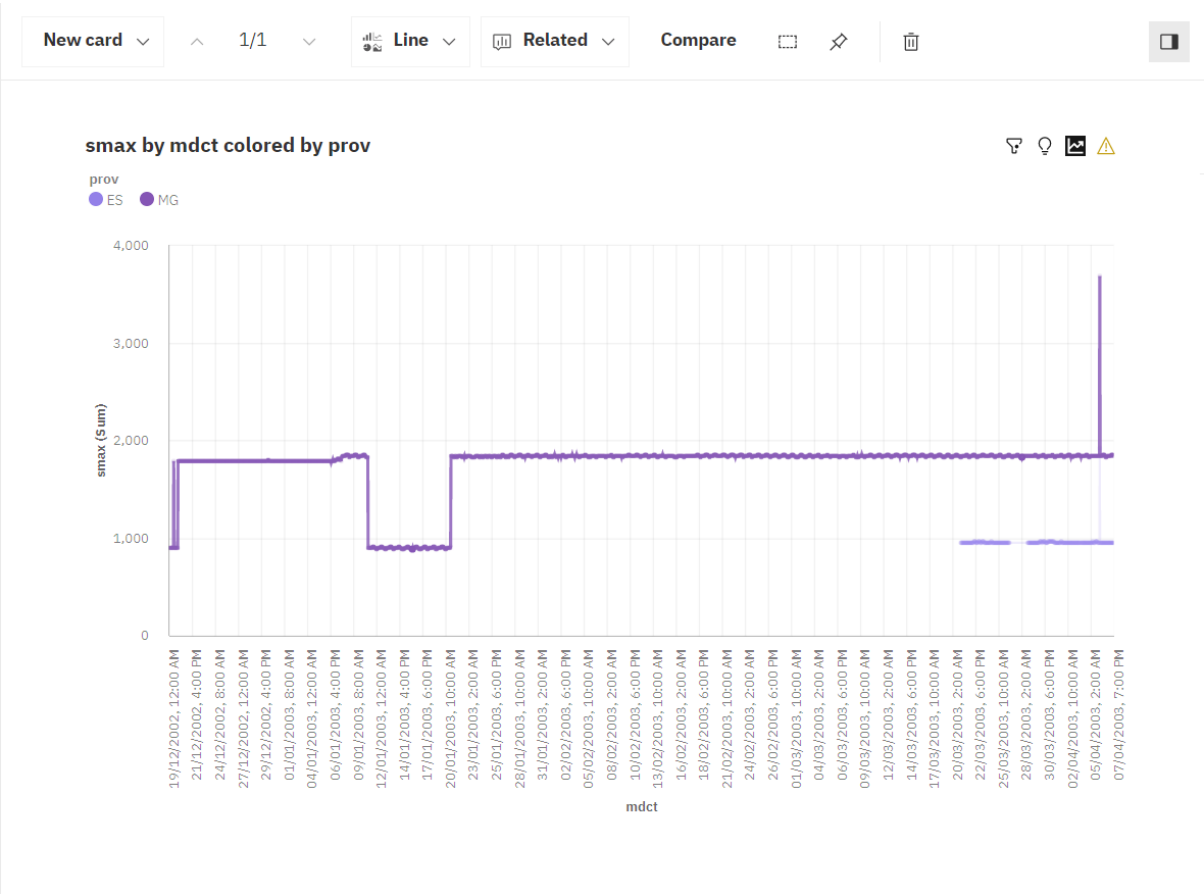

Figure 4 Visualization based on air pressure by time recording by province

PROBLEM 2 and PROBLEM 3

Git repository link:

https://git.cs.dal.ca/kasi/csci-5408-w2021-b00881083-srikrishnan/-/tree/master/Assignment_4. [6][7]

References:

1. https://www.fondriest.com/environmental-measurements/parameters/weather/#:~:text=Weather%20is%20made%20up%20of,quality%20of%20local%20atmospheric%20conditions.
2. https://www.calculator.net/dew-point-calculator.html
3. https://www.kaggle.com/PROPPG-PPG/hourly-weather-surface-brazil-southeast-region
4. https://www.yourarticlelibrary.com/climatology/7-forecasts-of-weather-parameters/88939
5. https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_fm.doc/t_ug_fm_steps_create_star_schema.html#ug_fm_steps_create_star_schema
6. https://github.com/shekhargulati/sentiment-analysis-python/tree/master/opinion-lexicon-English
7. https://git.cs.dal.ca/kasi/csci-5408-w2021-b00881083-srikrishnan/-/tree/master/Assignment_3