Rushi Samirbhai Patel
B00886157
rs397441@dal.ca

Assignment 5

Problem 1

Business Intelligence Reporting using Cognos

Identifying Data fields by dimensions:

The below table describes the measurable fields with a brief description on why I chose the particular facts and dimensions:

Table 1.1: Measurable Fields from the weather dataset

Id	Facts	Dimensions	Reasons
1	Air Temperature	Radiation, Station	Air Temperature can be analyzed by the solar radiation as well as the station (city location) depending on the location. Thus radiation and station are factors to analyze air temperature which would be our fact.
2	Dew Temperature	Radiation, Station	Dew Temperature can be analyzed by the solar radiation as well as the station (city location) depending on the location. Thus radiation and station are factors to analyze dew temperature which would be our fact.
3	Atmospheric Pressure	Station, State, Region, Latitude, Longitude, Height	Atmospheric pressure can be analyzed by the state, region as well as the station (city location) depending on the location. It also depends on the height of the location along with the location's latitude and longitude. Thus station, state, region, latitude, longitude and height are factors to analyze atmospheric pressure which would be our fact.
4	Relative Humidity	Station, Dew Temperature,	Relative Humidity can be analyzed by the dew temperature (since this will

		Latitude, Longitude	have a direct effect towards humidity) as well as the station (city location) depending on the location. Humidity will also depend on the latitude and longitude of the location. Thus station, dew temperature, latitude and longitude are factors to analyze relative humidity which would be our fact.
5	Wind speed	Latitude, Longitude, Height	Wind speed can be analyzed by using latitude, longitude and most importantly the height of the location since it has a direct effect on the wind speed. Thus latitude, longitude and height are factors to analyze wind speed which would be our fact.
6	Max/Min Temperature	Date and Hour	Max/Min Temperature are calculated for the previous hour so it is obvious that date and hour are our key aspects to analyze max/min temperature.
7	Max/Min Dew Temperature	Date and Hour	Max/Min Dew Temperature are calculated for the previous hour so it is obvious that date and hour are our key aspects to analyze max/min dew temperature.
8	Max/Min Relative Humidity	Date and Hour	Max/Min Relative Humidity is calculated for the previous hour so it is obvious that date and hour are our key aspects to analyze max/min relative humidity.

Steps to clean the dataset:

1. **Renamed the columns from portuguese language to its english translation** for better readability : data -> date, hora->hour, precipitacao total,horario (mm) ->total precipitation (mm), pressao atmosferica ao nivel da estacao (mb)->atmospheric pressure at station height (mb), pressao atmosferica max. na hora ant. (aut) (mb)->atmospheric pressure max. in the previous hour (mb), pressao atmosferica min. na hora ant. (aut) (mb)->atmospheric pressure min. in the previous hour (mb),

temperatura do ar - bulbo seco (Â°c)->air temperature - dry bulb (Â°c), temperatura do ponto de orvalho (Â°c)->dew point temperature (Â°c), temperatura maxima na hora ant. (aut) (Â°c)->max. temperature in the previous hour (Â°c),temperatura minima na hora ant. (aut) (Â°c)->min. temperature in the previous hour (Â°c), temperatura orvalho max. na hora ant. (aut) (Â°c)->dew temperature max. in the previous hour (Â°c), temperatura orvalho min. na hora ant. (aut) (Â°c)->dew temperature min. in the previous hour (Â°c),umidade rel. max. na hora ant. (aut) (%)->relative humidity max. in the previous hour (%),umidade rel. min. na hora ant. (aut) (%)->relative humidity min. in the previous hour (%), umidade relativa do ar, horaria (%)->air relative humidity (%), vento direcao horaria (gr) (Â° (gr))->wind direction (Â° (gr)),vento rajada maxima (m/s)->wind gust (m/s), vento velocidade horaria (m/s)-> wind speed (m/s).

2. Replaced the value **-9999 to 'null'** suggesting that the data was not registered for that particular column at that particular hour or date.
3. Created multiple csvs for facts and their respective dimensions namely :
    - Air_temperature_dataset
    - Air_temperature_dimensions
    - Atmospheric_dataset_dimensions
    - Atmospheric_pressure_dataset
    - Date_hour_dimensions
    - Dew_temperature_dataset
    - Dew_temperature_dimensions
    - Max_min_dew_temperature_dataset
    - Max_min_relative_humidity_dataset
    - Max_min_temperature_dataset
    - Relative_humidity_dataset
    - Relative_humidity_dimensions
    - Wind_speed_dataset
    - wind_speed_dimensions

Rushi Samirbhai Patel rs397441@dal.ca
B00886157

Star schema using Cognos:

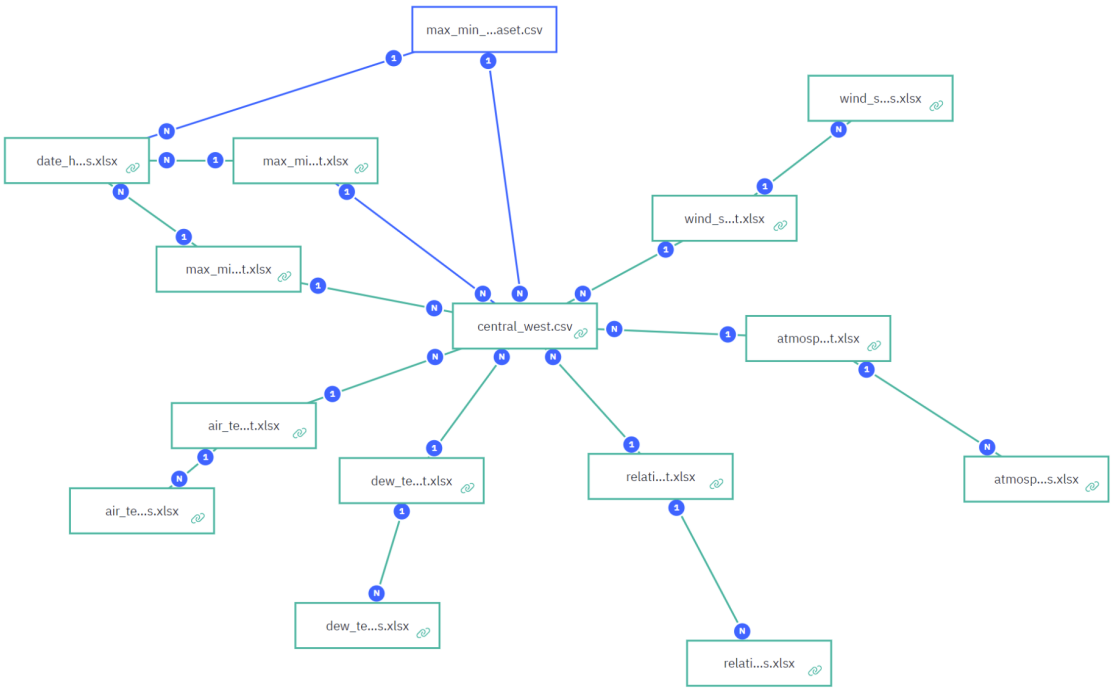

Figure 1.1: Star Schema of the weather dataset

The figure 1.1 demonstrates the star schema of the dataset using cognos. The datasets that I segregated based on the facts and their dimensions are closely related in the star schema. The max_min_temperature_dataset, max_min_relative_humidity_dataset and max_min_dew_temperature_dataset have the same dimensions namely date_hour_dimensions since all these facts would depend on hour and date because they are the max and min of the previous hour. Similarly, air_temperature_dataset is linked with its dimensions set which is linked with station and radiation. The dew_temperature_dataset is linked with its dimensions namely radiation and station too since these are required to determine both dew temperature and air temperature. The relative_humidity_dataset is linked with its dimensions namely Station, Dew Temperature, Latitude, Longitude since the location along with the dew temperature will be affected towards the relative humidity of that location. The atmospheric pressure is directly linked with its dimensions namely Station, State, Region, Latitude, Longitude, Height since height along with the location is required to analyze the atmospheric pressure. Finally, wind speed is linked to its dimensions namely latitude, longitude and height since these factors will be required to analyze the wind speed. All these datasets

are then linked to the central_west dataset since that dataset is considered as the linking dataset of all my fact tables and dimension tables (datasets).

Dimension Modeling of Weather Dataset:

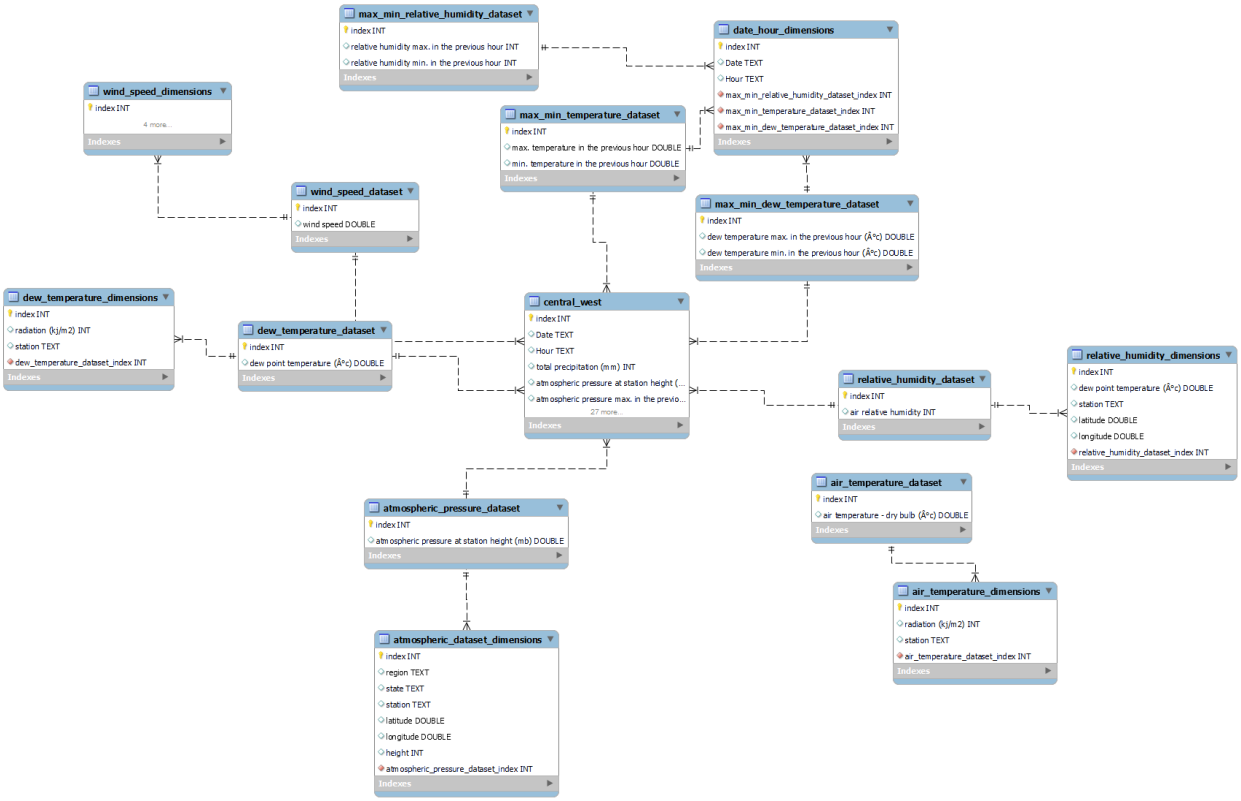

Figure 1.2: Dimension modeling of weather dataset

Figure 1.2 illustrates the model for the weather dataset which I have created using reverse engineering of mysql by importing the created fact datasets along with the dimensions table in mysql and linking them to generate this model.

Rushi Samirbhai Patel
B00886157
rs397441@dal.ca

Visual Analysis using Cognos:

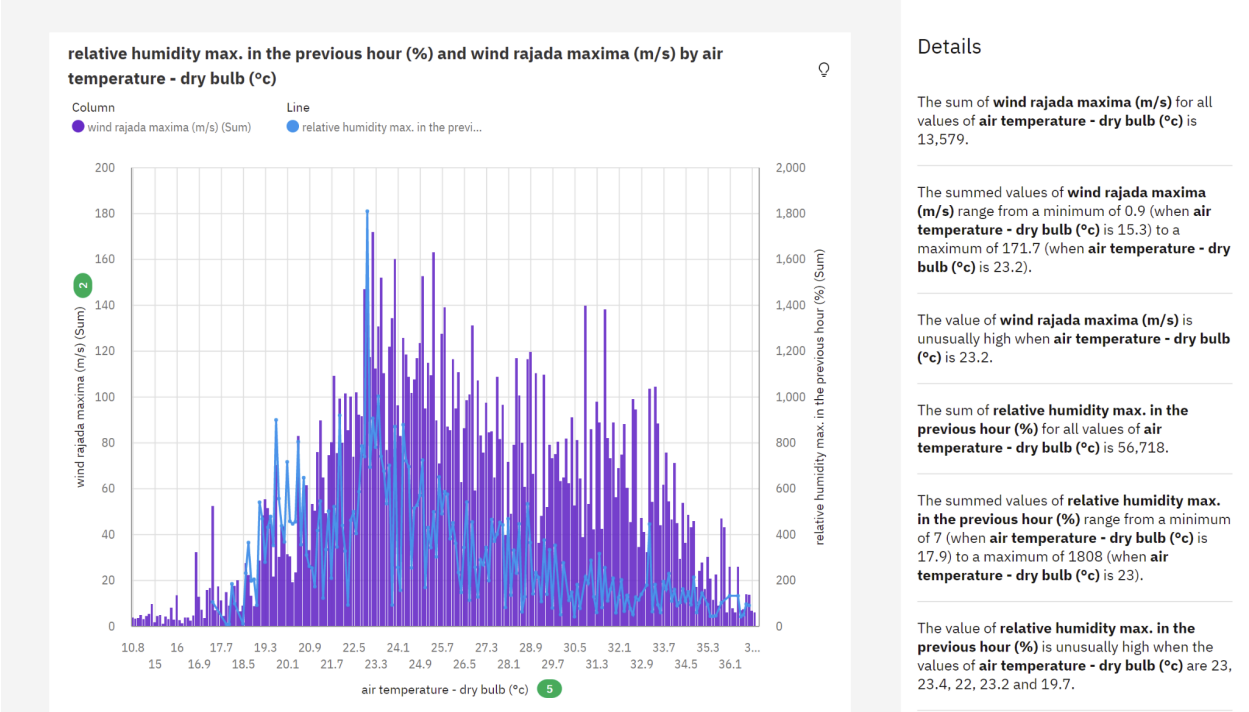

Figure 1.3: Visual analysis of max of Max Relative Humidity in previous hour and max wind speed by air temperature

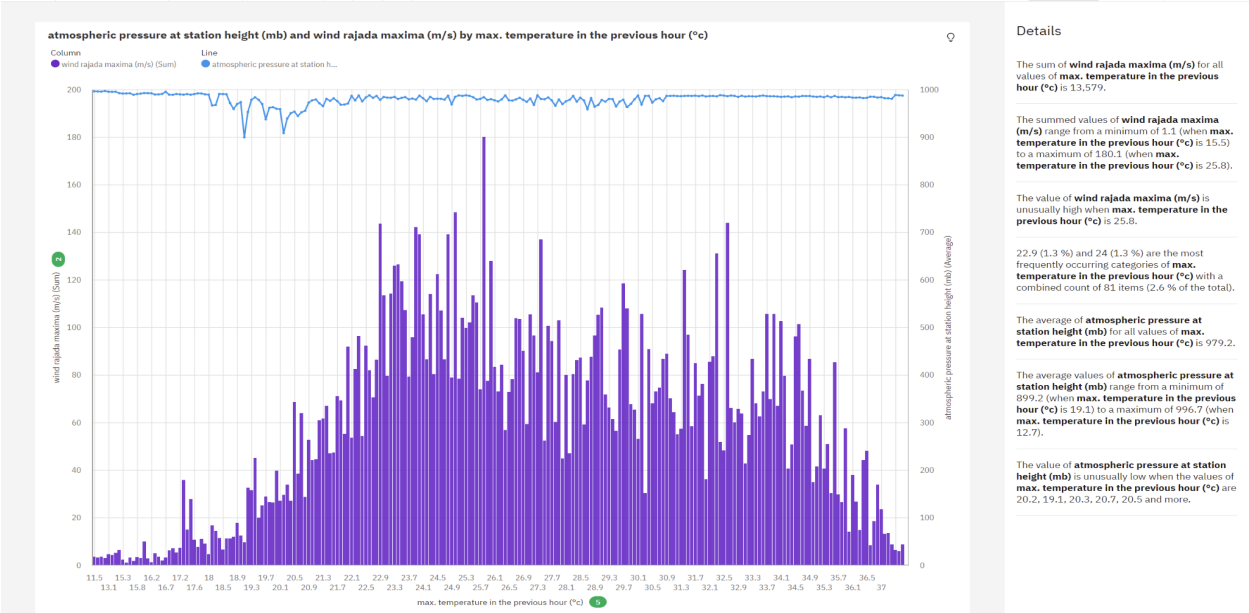

Figure 1.4: Visual analysis of atmospheric pressure at station height and wind speed by max temperature in the previous hour

Rushi Samirbhai Patel
B00886157
rs397441@dal.ca

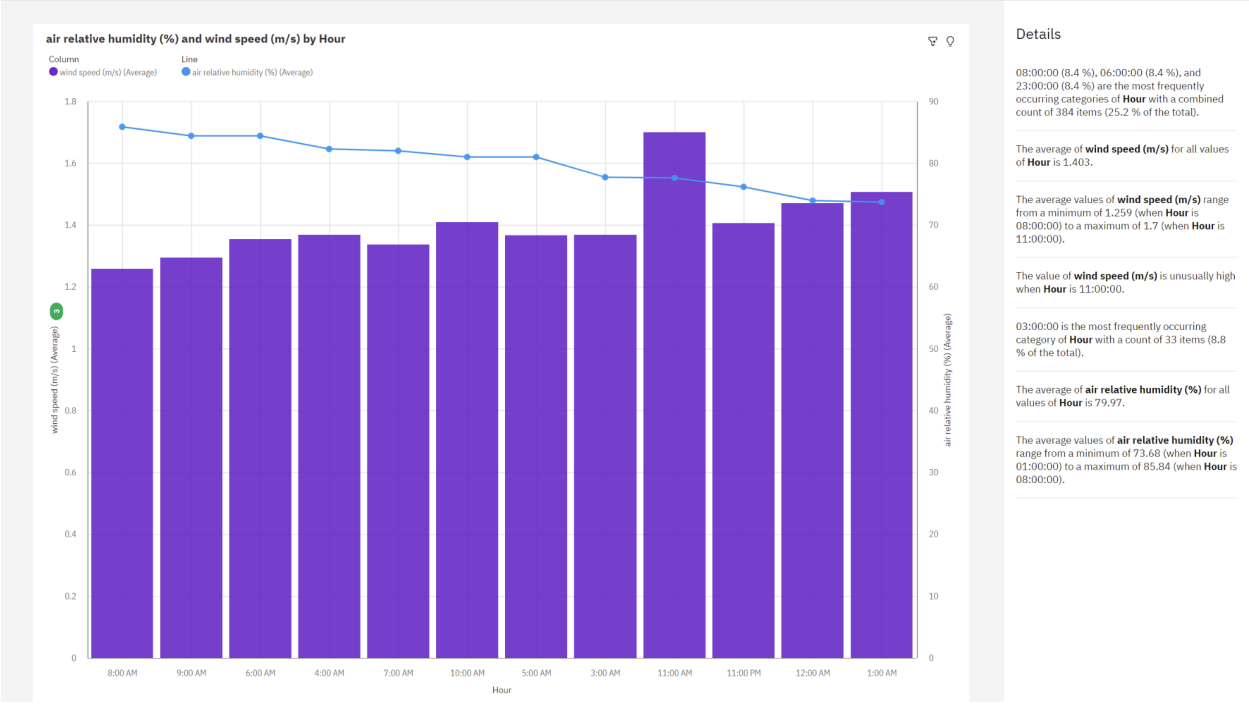

Figure 1.5: Visual Analysis of air relative humidity and wind speed by hour

Problem 2

Link to Gitlab code:https://git.cs.dal.ca/rspatel/csci-5408-w2022-b00886157-rushi_patel/-/tree/main/Assignment-5/Assignment5Problem2

I have removed my MongoDB credentials from the code in the connection string. Please replace the credentials in the code to make the code run.

Problem 3

Link to Gitlab code:https://git.cs.dal.ca/rspatel/csci-5408-w2022-b00886157-rushi_patel/-/tree/main/Assignment-5/Assignment5Problem3

I have removed my MongoDB credentials from the code in the connection string. Please replace the credentials in the code to make the code run.

Rushi Samirbhai Patel
B00886157
rs397441@dal.ca

References:

[1]	"Climate Weather Surface of Brazil - Hourly", *Kaggle.com*, 2022. [Online]. Available: https://www.kaggle.com/PROPPG-PPG/hourly-weather-surface-brazil-southeast-region?select=sudeste.csv. [Accessed: 29- Mar- 2022].