Assignment – 4

1. Business Intelligence Reporting using Cognos

The dataset from Kaggle had 31 columns in it. I have observed all the columns and their relevance is noted. One column glord is removed from table because it was loosely coupled with other columns of table. Fact table name is weatherFact table and it contains columns wsid, wsnm, inme and mdct. Primary key for this table is wsid whereas inme and mdct are foreign keys. DimLocation is a location dimension table with primary key inme. It includes other columns lat, lon, elv, city and prov. In location table common locations are identified and entered in table. DimWind table is dimension table, it has primary keys wsid and mdct. Other attributes are wdsp, wdct and gust. DimHumidity is a dimension table with primary keys wsid and mdct. Other attributes of the table are hmdy, hmax, hmin and prcp. DimTime table is also a dimension table with primary key wsid and mdct and other attributes are date, yr, month, hr and da. DimTemp table is a dimension table with primary key wsid and mdct. Temp, dewp, tmax, tmin, dmax and dmin are other columns. DimAirPressure is a dimension table, primary key is wsid and mdct. Other attributes are stp, smax and smin.

For ease, first 100 records of each weather station are extracted from csv file. One column glord is removed as it was not strongly related to any other column of tables.

Star schema is created for model of weather database. In star schema diagram, fact table is related to all the other tables. No other tables are related to each other. There is one many to one relationship between fact and location dimension table where there can be one location which has multiple weather stations. In all the other fact dimension table relationship, there is one to one relationship.

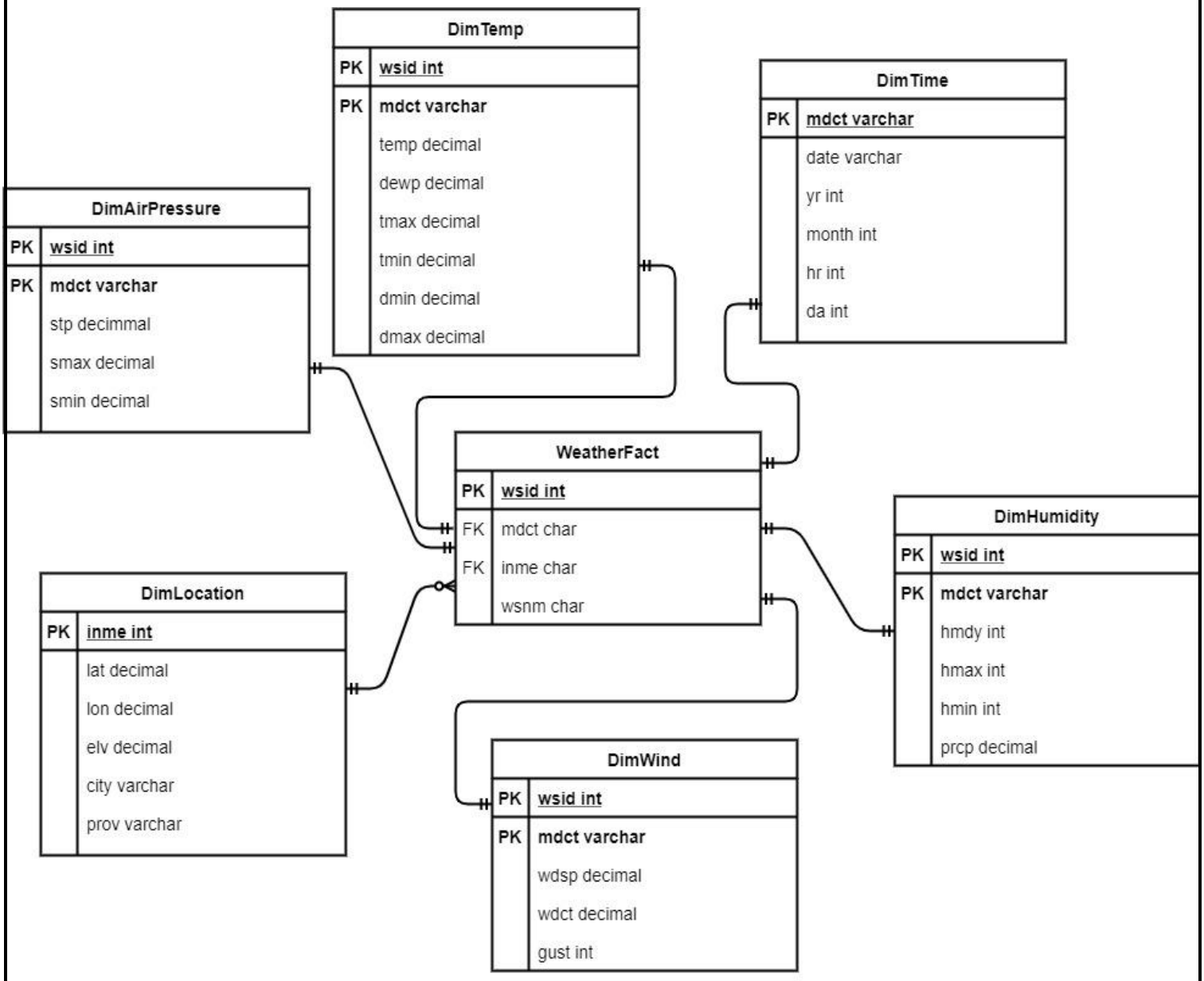

Figure 1 – Star schema diagram

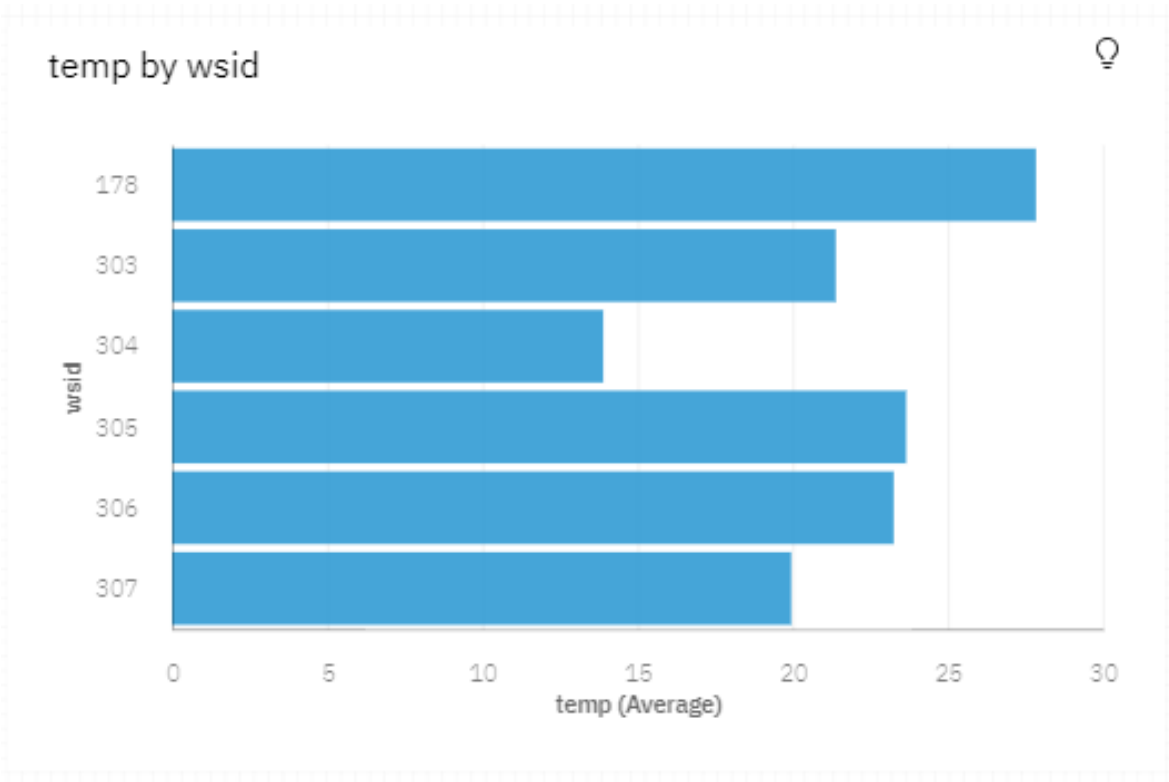

Figure 2 – Temperature by weather station bar chart in Cognos

GitLab link: https://git.cs.dal.ca/mihir/assignment-4-csci5408.git

References

[1] "Deadlock in DBMS," [Online]. Available: https://www.javatpoint.com/deadlock-in-dbms. [Accessed October 2020].

[2] "Visualize Clustering Using ggplot2," [Online]. Available: https://rpubs.com/aephidayatuloh/clustervisual.

[3] "The Snowflake Schema," [Online]. Available: https://www.vertabelo.com/blog/data-warehouse-modeling-the-snowflake-schema/.

[4] "Difference between Star Schema and Snowflake Schema," [Online]. Available: https://www.geeksforgeeks.org/difference-between-star-schema-and-snowflake-schema/?ref=rp.