Rohit Jyotiba Chougule, 19200240
Assignment No. 7

Question 1:

1. Spatial Dataware house includes spatial dimensions, measures about a specific collection of data. It contains topological or distance information about the data. Various domains where Data Mining can be applied on a Spatial data include geomarketing, environmental studies, remote sensing, geographical information systems (GIS). Spatial Data are highly self co-related. The attributes of a spatial data most often has information related to spatial locations, e.g., longitude, latitude, and elevation.
   A Spatial Dataware house can be designed Motorway traffic data collected for city traffic planning.

   Various aspects & measures to be considered when designing a Spatial Dataware house for traffic data include, which are the dimensions for the data warehouse:

   1. Time:
      Time is the dimension that has data about the various timestamps and span of the day. Like, day of the week, day, month year. Time here is the temporal data in the data warehouse.
   2. Area:
      Area dimension has the details on Highway number/Road number that could be helpful in uniquely identifying the road where the measure needs to be done on. Map coordinates are helpful to in identifying exact details on a particular spot on a road.
   3. Vehicle:
      The vehicle dimension consists details on the type of vehicles, vehicle density, vehicle speeds.
   4. Incident:
      The incident dimension has details on whether any incident like accident or car breakdown has taken place and whether any action has been taken.
   5. Traffic_signal:
      The traffic signal dimension can be helpful to identify if the traffic flow at the time instant and identify the root to the congestion.

   The above dimensions are the key dimensions that I believe are the important when designing a Spatial Dataware house for motorway traffic planning.

   For the Traffic data spatial data warehouse, the measurements are:

   1. Vehicle_Density:
      The vehicle density measures the number of vehicles passing by a particular road or an area during a time span.

The vehicle density is one of the key measures out of the total measures in the data warehouse.

2. Traffic_Flow

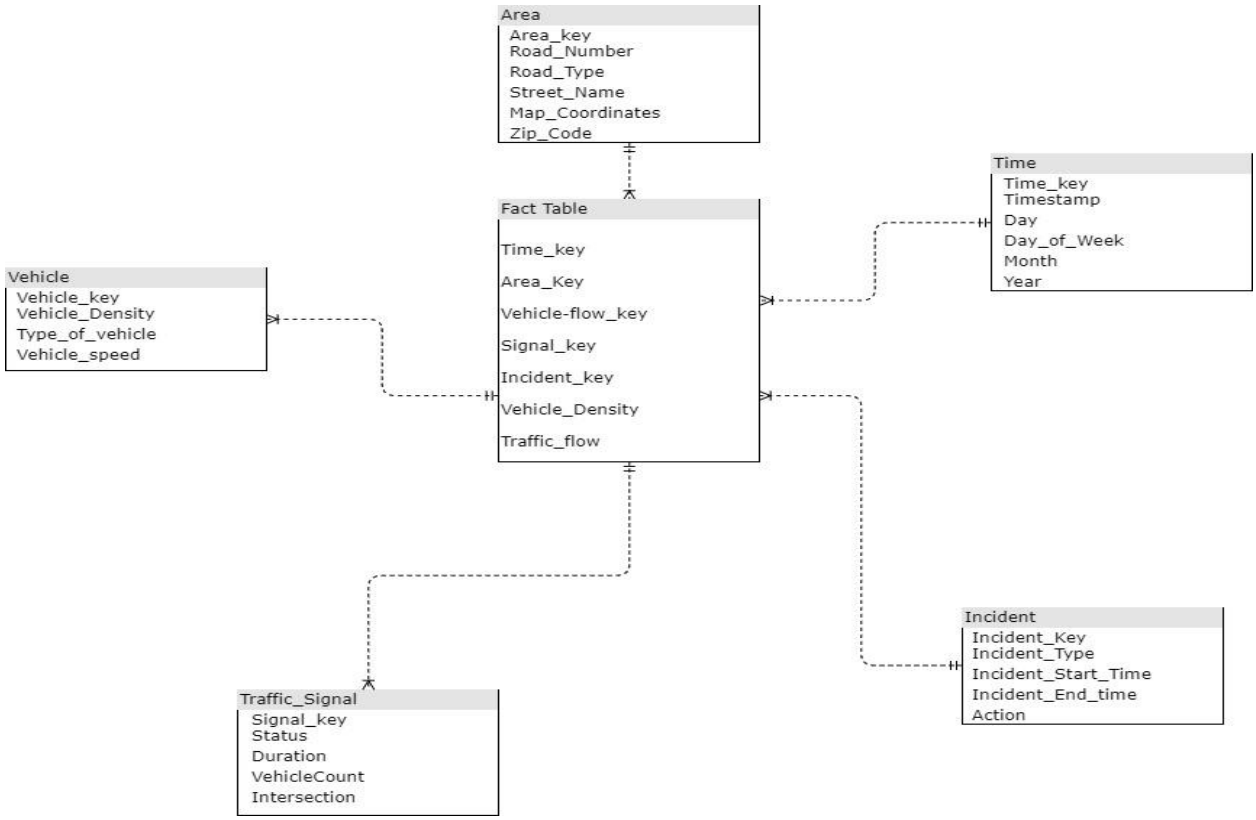
   Traffic_flow is the measure that classifies the data into categories like low, medium and high, depending on the number of vehicles and speed of the vehicles at a point. The traffic flow measure can be obtained from various keys in the dimension tables, it can be at Traffic_signal or can be collected from the GPS details of the vehicles having that provision. The video cameras installed with the traffic signals can also be utilized for measuring this aspect.

3. Area or Region is another spatial measure that can be used to identify the regions that are highly affected by motorway traffic.

   We have a hierarchy for the dimension Time in our data warehouse as –

   Time <- Day of week <- Month <- Year

Designed below is a Star schema for the Traffic data warehouse:



2. What information can we mine from such a spatial data warehouse to help city planners?
- Information regarding the vehicle density at different hours can be identified to suggest the city planners to add/remove or modify the traffic signal on the route.
- If possible, finding alternative routes and diverting traffic during peak hours depending on the vehicle density identified.
- Identifying the type/segment of vehicles that are frequent on any of the roads and depending on which the road lanes can be divided and reserved.
- Traffic density at different hours can be identified to understand the peak hours and route the traffic accordingly. If required new roads or alternative roads can be built depending on the need and intensity of the traffic.

3. This data warehouse contains both spatial and temporal data. Propose one mining technique that can efficiently mine interesting patterns from such a spatiotemporal data warehouse.

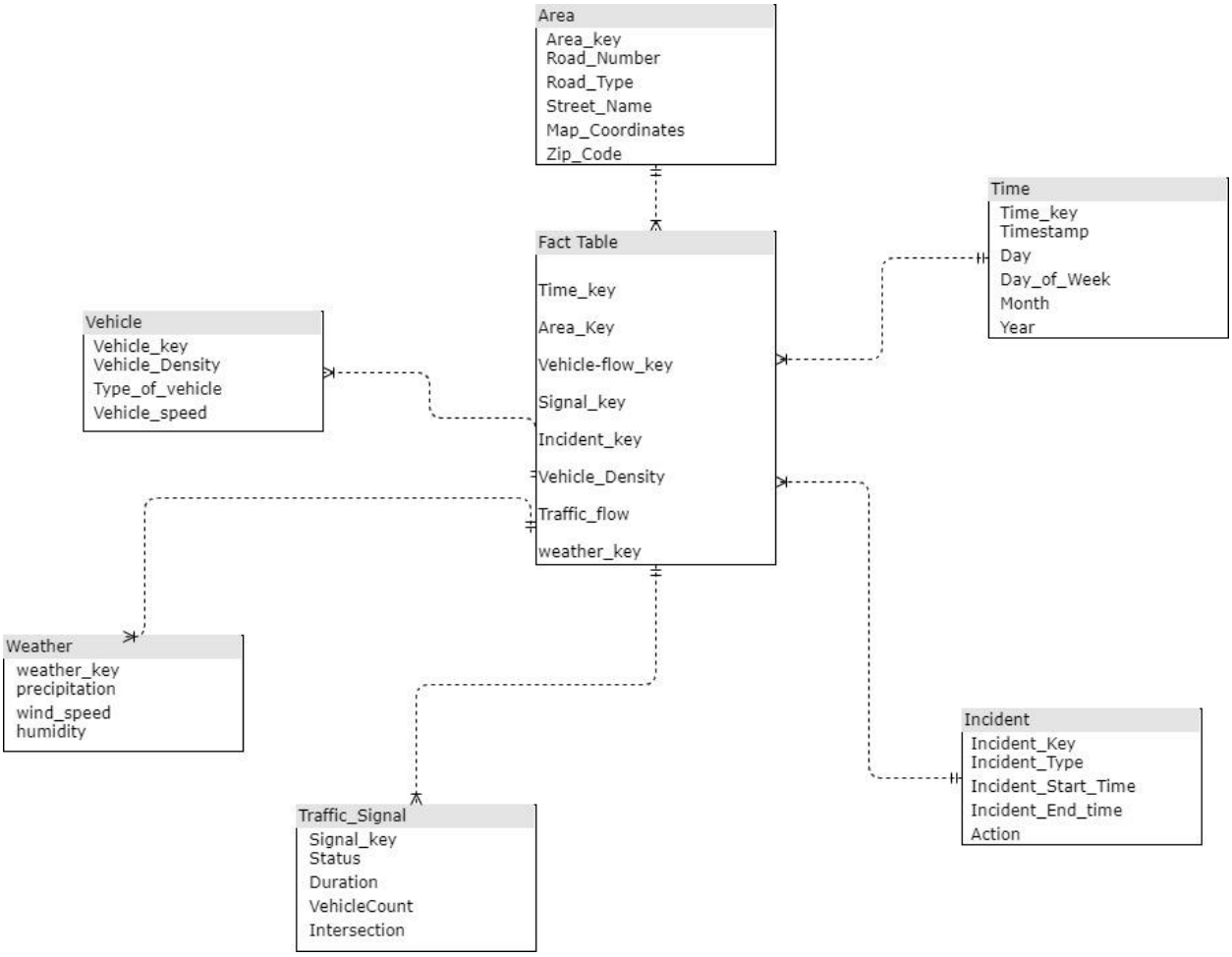
- The data warehouse that has both, spatial and time series trend data is known as a spatiotemporal data warehouse. Information gained from mining the data from a spatiotemporal data warehouse can be very fruitful as it contains the data over a period and includes the spatial parameter for the data as well.
- Different data mining techniques are applied for the Spatial data warehouses to retrieve fruitful information. A trend analysis of the traffic can be generated from the traffic data collected to identify – 1. Peak hours during weekdays, 2. Segment of vehicles in traffic (cars, buses, trucks), 3. Trends of traffic during weekends. 4. Efficiency of traffic signals in managing the traffic, 5. Regions affected mostly by traffic.

Association rules mining algorithm can be implemented here to get the trends of traffic data.
In mining Spatiotemporal data warehouse, we retrieve meaningful information in both spatial (Traffic affected areas) and temporal (Time during which traffic affects) measures.

Question 2

- Another dimension for weather can be added to the existing Spatiotemporal data warehouse design.
- The weather dimension will contain data about the rain amount, which is the precipitation, that can be 0 as well in case there is no rain expected on a day. The wind speed may also at times affect the traffic for bikers and cyclists.
- Congestions at intersection on one-way roads can be observed or analyzed with consideration of the traffic signal dimension as well.
- A star schema for data warehouse design has been implemented in this design as well, however a snowflake model can also be implemented, where we can consider each

record from the traffic and weather dimensions and apply mining algorithms to find patterns and predict traffic situations.



- Association rule mining algorithms can be applied here on this spatiotemporal data warehouse to identify patterns and predict the traffic conditions at a location or street on a given day, weather and time.
- Some of the example for association rules after finding frequent traffic congestion data, can be applied are in combination of:
  Time, Road, weather, vehicle-flow, incident, area
- Different confidence values could be tried to create association rules, which could then be helpful in predicting traffic situation.