Team 17

Dimensional Model

Business Process: Based on the requirements of the Chicago and Dallas food inspection dataset, the business process selected for the dimensional model is food safety inspections. The goal of this process is to ensure that food establishments are following health and safety guidelines to prevent any health hazards for consumers

Grain: Each inspection per row. There could be multiple inspections happening across the city per day.

Dimensions:

1. DimBusiness
2. DimAddress
3. DimFoodRiskCategory
4. DimFoodinspectionType
5. DimViolationCode
6. DimFacilityType
7. DimLicense

Facts:

1. FactFoodInspections
2. FactInspectionViolations

Data discrepancies

(CH): Chicago Dataset

(D): Dallas Dataset

(BL): Business Licences

1. (CH) Two records with the exact details except for License No
   DBA_Name : Thunderbird catering co
   Inspection date: 03/22/2010
2. (CH) No standard address format is being followed in the restaurant's street address for Chicago Dataset. So, regex cannot be applied directly on this column to split into street number, type, direction, etc.
3. (D) One Restaurant Name with multiple addresses. Eg: WINGSTOP
4. Same Restaurants with discrepancies in name, such as:
   (CH) SPORTSERVICE SOLDIER FIELD
   SPORT SERVICE SOLDIER FIELD
   (D) BRYAN STREET'S TAVERN
   BRYAN STREET TAVERN
   In this case, SQL considers this as different records.
5. (D) One StreetAddress has multiple restaurants linked to it. Ques: how can we make it type 2 scd? We do not make address as type 2 SCD. Since the combination of name, street address is not unique.
   Eg, address: '650 S GRIFFIN ST'
6. (D) 11 records where Restaurant Name = Null
7. (D) Multiple formats for Inspection Date
8. (CH) 634 Restaurants with License # = 0 (License # is not unique on its own. It has to be combined with other attributes to make a record unique)
9. (CH) Many instances exist where a single license # is associated with many restaurants.
10. (D) Violation Code: 47; not differentiating between 'OTHER VIOLATIONS' and 'Other Violations'; count is increasing by 1

11. (CH) Three restaurants where the street address is blank.
12. (CH) 36626 records having combination of DBA Name, AKA Name, and Street Address as unique.
13. Inspection Date cannot be used to create SCD on ViolationCode.
    Let's say you have a violation code 55 with an ABC description. If you associate it with Inspection Date, it will change as that particular violation is encountered.
    So, on inspection date 03/20/2023 → Violation Code 55 → Violation Description ABC
    03/22/2023 → Violation Code 55 → Violation Description DEF
    03/22/2023 → Violation Code 55 → Violation Description ABC
    03/24/2023 → Violation Code 55 → Violation Description DEF
    Now, the active flag will be 'Y' with 03/24/2023. But if you perform any visualization or analysis for Violation Code 55, it will show DEF in the Violation Description. The violation Code keeps on changing from ABC to DEF and vice-versa. So, it is not advisable to make this dimension an SCD.

Loading Rules:

Pre-stage table: stores 'as-is' data

Stage table: stores data with transformations applied.

1. DimAddress: Unique StreetAddress
2. (D) There are 3 restaurants with DBA Name as null. For those values, we are going to populate the attribute value as 'Restaurant Name Not Available'
3. (CH) We have removed licenses which are not related/relevant to food inspection licenses. There are licenses which belong to Indiana, Colorado.
4.

-1 : FK Not Available

-99 : Violation Code Not Available

Columns not in the dimensional model:

After thoughtful consideration and n number of brainstorming sessions, we have decided to omit certain columns from our dimensional model. We believe those columns are not going to be used for any visualizations as of now. But we have maintained them in the staging table, in case we need them in the future.

(D) Violation Points (1-25), Violation Detail (1-25), Inspection Month, Inspection Year

(CH) Location

(BL) Site Number, Legal Name, City, State, Zip code, Ward, Precinct, Ward Precinct, Police District, Business Activity ID, Business Activity, SSA, Latitude, Longitude, Location