ALY6030 Assignment 3 — Public Housing Inspections Star Schema

Jiaqiao Feng

Question 1: How many facts are there in this dataset?

1.1. Which facts do you identify?

There are **two facts** in this dataset:

- INSPECTION_SCORE – a numerical score ranging from 0 to 100.
- COST_OF_INSPECTION_IN_DOLLARS – the dollar amount spent on the inspection.

1.2. For the facts that you identify, what type of facts are they?

- INSPECTION_SCORE is a **non-additive fact** because you cannot meaningfully sum inspection scores across different rows.
- COST_OF_INSPECTION_IN_DOLLARS is an **additive fact**, as you can sum costs across all inspections for analysis.

Question 2: How many dimensions are there in this dataset?

2.1. Which dimensions do you identify?

There are **four dimensions** in this dataset:

1. **Public Housing Agency** – PUBLIC_HOUSING_AGENCY_NAME
2. **Development** – INSPECTED_DEVELOPMENT_NAME

3. **Location** – includes INSPECTED_DEVELOPMENT_ADDRESS, INSPECTED_DEVELOPMENT_CITY, and INSPECTED_DEVELOPMENT_STATE
4. **Date** – INSPECTION_DATE

Question 3: which type (or types) of fact tables would you use and why?

To show both each inspection and a monthly summary of costs, I would use two types of fact tables. First, I would use a **transaction fact table** to store each inspection with its score and cost. This helps when we need to look at every inspection one by one.

Second, I would create a **periodic snapshot fact table** to keep the total inspection cost for each month. This is useful when managers want to see reports by month without doing the calculation every time. Using both tables helps answer different business questions more easily.

Question 4: how should we handle this slowly changing dimension? Select from types 0, 1, 2, or 3 from the Kimball reading

I would use Type 2 “slowly changing dimension” to handle changes in public housing agency names and addresses. This is because Type 2 keeps the old records and only adds a new row when there is a change. It helps us track the history of each agency, which is important if agencies merge or change names. By keeping both current and past versions, we can still know which agency was responsible at the time of each inspection. This gives more accurate reporting over time.

Question 5: Execute SQL Query

Please find out the query in the .sql file as well as the csv spreadsheet. The result contains all the PHAs that saw an *increase* in the cost of performing an inspection in their jurisdiction.