Assignment 3

Question 1a: How many facts are there in this dataset?
Answer 1a: There are **two facts** in this dataset.

- INSPECTION_SCORE
- COST_OF_INSPECTION_IN_DOLLARS

Question 1b: Which facts do you identify?
Answer 1b:

- Inspection Score : This represents the score assigned to an inspection, measuring the quality or compliance of the development.
- Cost of Inspection in Dollars: This represents the monetary cost incurred to perform the inspection.

Question 1c: For the facts that you identify, what type of facts are they?
Answer 1c

- In dimensional modeling, facts are typically categorized as additive, semi-additive, or non-additive facts.
- Inspection Score's type is non-additive, becase we cannot meaningfully sum inspection scores across different inspections. They are better used as averages or ratios in analysis.
- As for the Cost of Inspection in Dollars, the type is additive, becase this cost can be summed across multiple inspections, agencies, developments, time periods,etc.

Question 2a: How many dimensions are there in this dataset?
Answer 2a: There are **five dimensions** in this dataset.

Question 2b: Which dimensions do you identify?
Answer 2b:

- The first dimension is the **Public Housing Agency**. This dimension includes the name of the agency responsible for managing public housing and conducting inspections. It helps analyze inspection activities and performance by different agencies.
- The second dimension is the **Development**. It contains information about the housing development being inspected, such as the development name, address, city, and state. This allows analysis by specific properties or locations.
- The third dimension is **Time (Inspection Date)**. This dimension includes the date when the inspection took place. It is useful for analyzing inspections over time, such as by year, quarter, or month.
- The fourth dimension is **Geography**. While related to the development information, geography can be treated as its own dimension, focusing on the city and state. This enables analysis by region or location.
- The fifth dimension is the **Inspection** itself. This includes unique inspection events, such as the inspection ID, and can provide a detailed view of each inspection performed.

Question 3: Senior management is interested in viewing the facts identified above, at both the inspection level, as well as a periodic summary of inspection costs for each month. Based on this context, if you were to store these data in a set of fact tables, which type (or types) of fact tables would you use and why?

Answer 3: To meet senior management's need for both detailed and summarized views, I would use two types of fact tables: a **transactional fact table** and a **periodic snapshot fact table**. The transactional fact table captures data at the inspection level, including facts such as the inspection score and the cost of each inspection. This table allows detailed analysis of individual inspections and can answer specific questions about performance and cost at a granular level.

The periodic snapshot fact table provides a summarized view of inspection activities over a set time period, such as monthly totals or averages. It is useful for reporting overall trends, such as total inspection costs per month or average scores over time. Using both tables allows flexibility in analysis—supporting both detailed operational reporting and higher-level strategic summaries for management decision-making.

Question 4: Senior Management is also concerned with changes in the names and addresses of the public housing agency names since they tend to get merged with other agencies on a frequent basis. Based on this, how should we handle this slowly changing dimension? Select from types 0, 1, 2, or 3 from the Kimball reading. Justify your answer.

Answer 4: To handle changes in the names and addresses of public housing agencies, I recommend using a **Slowly Changing Dimension (SCD) Type 2** approach. Type 2 tracks historical changes by creating a new record whenever a change occurs, such as when agencies merge or update their names and addresses. This method preserves the full history of the dimension, allowing accurate reporting based on how data looked at any point in time. For example, if an agency merged and changed its name, we would keep both the old and new records with effective dates.

This approach is justified because senior management is specifically concerned with tracking changes over time. Using Type 2 ensures that historical reports remain accurate and consistent with the data that was valid at the time of each inspection. It allows analysis both before and after an agency's name or structure changed, which would not be possible with Type 1 (which overwrites data) or Type 0 (which ignores changes). Type 3 could track limited changes but typically only keeps the current and previous values, which is not sufficient for this scenario.