Assignment 3
Questions 1-4

1. **How many facts are there in this dataset?**

There are two facts in this dataset.

- **Which facts do you identify?**

Cost of Inspection in Dollars (numeric value)

Inspection Score (numeric value)

- **For the facts that you identify, what type of facts are they?**

The facts can be categorized into two types:

Cost of Inspection in Dollars: Additive fact (since you can sum costs)

Inspection Score: Non-additive fact (it's usually average)

2. **How many dimensions are there in this dataset?**

There are 7 dimensions:

- Public Housing Agency Name
- Inspected Development Name
- Inspected Development Address
- Inspected Development City
- Inspected Development State
- Inspection Date
- Inspection ID (used for identifying inspections)

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?**

I think I would need two types of fact tables:

- Transaction Fact Table: Stores inspection-level data with facts like cost and score for each inspection.
- Periodic Snapshot Fact Table: Stores aggregated data on a monthly basis, such as total costs per month for each housing agency.

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.**

A slowly changing dimension (SCD) occurs when the values of a dimension change over time.

SCD types are:

- SCD Type 0: No change (static).
- SCD Type 1: Overwrite old values with new values.
- SCD Type 2: Keep history by adding a new row each time the value changes.
- SCD Type 3: Keep limited history (old and new values in the same row).

SCD Type 2: This is the most appropriate choice because public housing agency names may change due to mergers, and you'll need to keep historical records of these changes. For example, if a housing agency merges, you would create a new record for the merged agency while retaining old records for analysis.