Assignment 3 — Public Housing Inspections Star Schema

By Yue Zeng

ALY 6030 Data Warehousing and SQL

Date: Nov 30. 2024

Professor: Adam Jones

1.

- **How many facts are there?**
  There are 2 main facts in the dataset.
- **Which facts do you identify?**
  COST_OF_INSPECTION_IN_DOLLARS: Represents the cost incurred for each public housing inspection.
  INSPECTION_SCORE: Reflects the quality or performance score assigned to a housing development after inspection.
- **What type of facts are they?**
  COST_OF_INSPECTION_IN_DOLLARS is fully additive and can be summed across agencies, locations, and time periods.
  INSPECTION_SCORE is semi-additive, where it can be averaged over time or across developments but does not make sense to sum.

2.

- **How many dimensions are there?**
  There are 6 dimensions in the dataset.
- **Which dimensions do you identify?**
  PUBLIC_HOUSING_AGENCY_NAME
  INSPECTED_DEVELOPMENT_NAME
  INSPECTED_DEVELOPMENT_ADDRESS
  INSPECTED_DEVELOPMENT_CITY
  INSPECTED_DEVELOPMENT_STATE
  INSPECTION_DATE

3.

- **Which types of fact tables would I use and why?**
  I would use two types of fact tables: Transation fact table and Periodic summary fact table, which provides flexibility for granular analysis and summary reporting. For individual inspection records, allowing granular-level analysis. Transaction fact tables are typically used to store detailed event-level data (Kimball & Ross, 2013).
  For summarizing inspection costs by month, which is useful for high-level trend analysis over time.

4.

- **How should we handle this slowly changing dimension**
  I recommend using Type 2 Slowly Changing Dimension (SCD), where historical changes are tracked by adding new rows when changes occur. This allows querying both current and historical data, ensuring data integrity and maintaining a record of PHA name/address changes over time.

References

Northeastern University. (2024). *ALY 6030 Datawarehousing and SQL: Assignment 3 — Public Housing Inspections Star Schema [Data set]*. Course materials.

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: *The Definitive Guide to Dimensional Modeling* (3rd Edition). Wiley.