College of Professional Studies

Northeastern University San Jose

MPS Analytics

Course: ALY6030

Assignment:

Assignment 3 Retail Sales Star Schema

Submitted on:

Nov 31, 2024

Submitted to:

Instructor: Adam Jones

Submitted by:

Luyun Qi

Question 1

a. How many facts are there in this dataset?

   Two facts

b. Which facts do you identify?

   Two facts' are: 'COST_OF_INSPECTION_IN_DOLLARS' and 'INSPECTION_SCORE'

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

   'COST_OF INSPECTION_IN_DOLLARS': Additive fact. It can be added up.

   'INSPECTION_SCORE": Non-Additive fact. It can no be added up through any dimension

Question 2

a. How many dimensions are there in this dataset?

   Total Seven dimensions

b. Which dimensions do you identify?

   1. INSPECTION_ID
   2. PUBLIC_HOUSING_AGENCY_NAME
   3. INSPECTED_DEVELOPMENT_NAME
   4. INSPECTED_DEVELOPMENT_ADDRESS
   5. INSPECTED_DEVELOPMENT_CITY
   6. INSPECTED_DEVELOPMENT_STATE
   7. INSPECTION_DATE

Question 3: 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: For Storing these dataset, I'd recommending to use a periodic snapshot fact table. It sum up inspection data on a monthly basis, summarizing the total costs for each PHA or development. This is perfect for senior management, allowing them to easily track monthly trends, analyze costs, and make strategic decisions without getting lost in individual inspection details. The periodic snapshot fact table would include attributes like:

(1). PUBLIC_HOUSING_AGENCY_NAME (dimension)

(2). INSPECTION_DATE (time dimension)

(3). Total Inspection Cost for the Month (measure)

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

Answer: For handling the frequent changes in the names and addresses of the PHA, I would recommend using Type 2. Because Type 2 captures all changes by adding a new row with each update and we can also track history of each change, which is very important when PHA changes names and addresses too often.

Type 1 simply overwritten data, which will lose all historical data.

Type 3 offers historical data, but it is not sufficient for frequent and multiple changes.

Overall, Type 2 is the right fit.

Question 5:

Answer:

public_housing_inspection_data | SQL Query

```sql
SELECT
    PHA_NAME,
    INSPECTION_DATE AS MR_INSPECTION_DATE,
    INSPECTION_COST AS MR_INSPECTION_COST,
    SECOND_MR_INSPECTION_DATE,
    SECOND_MR_INSPECTION_COST,
    (INSPECTION_COST - SECOND_MR_INSPECTION_COST) AS CHANGE_IN_COST,
    ROUND(((INSPECTION_COST - SECOND_MR_INSPECTION_COST) / SECOND_MR_INSPECTION_COST) * 100, 2) AS PERCENT_CHANGE_IN_COST
FROM (
    SELECT
        PUBLIC_HOUSING_AGENCY_NAME as PHA_NAME,
        INSPECTION_DATE,
        COST_OF_INSPECTION_IN_DOLLARS as INSPECTION_COST,
        LAG(INSPECTION_DATE) OVER (PARTITION BY PUBLIC_HOUSING_AGENCY_NAME ORDER BY INSPECTION_DATE DESC) AS SECOND_MR_INSPECTION_DATE,
        LAG(COST_OF_INSPECTION_IN_DOLLARS) OVER (PARTITION BY PUBLIC_HOUSING_AGENCY_NAME ORDER BY INSPECTION_DATE DESC) AS
 SECOND_MR_INSPECTION_COST
    FROM 6030A3.public_housing_inspection_data
) AS InspectionRanked
WHERE
    SECOND_MR_INSPECTION_COST IS NOT NULL
    AND INSPECTION_COST > SECOND_MR_INSPECTION_COST
ORDER BY PHA_NAME;
```

	PHA_NAME	MR_INSPECTION_DATE	MR_INSPECTION_COST	SECOND_MR_INSPECTION_DATE	SECOND_MR_INSPECTION_COST	CHANGE
1	ADAMS METROPOLITAN HOUSING AUTHO	1/27/2014	32874	1/28/2014	24047	
2	Akron Metropolitan Housing Autho	9/24/2014	35447	9/25/2014	13871	
3	Akron Metropolitan Housing Autho	9/23/2014	36279	9/24/2014	35447	
4	Akron Metropolitan Housing Autho	8/6/2014	23194	8/8/2014	11472	
5	Akron Metropolitan Housing Autho	8/6/2014	26417	8/6/2014	23194	
6	Akron Metropolitan Housing Autho	8/4/2014	25943	8/5/2014	22940	
7	Akron Metropolitan Housing Autho	8/27/2014	22289	8/29/2014	11882	
8	Akron Metropolitan Housing Autho	8/25/2014	22730	8/26/2014	17208	
9	Akron Metropolitan Housing Autho	8/19/2013	38119	8/25/2014	22730	
10	Akron Metropolitan Housing Autho	8/12/2013	38177	8/19/2013	38119	
11	Akron Metropolitan Housing Autho	7/28/2014	17218	7/29/2014	11493	
12	Akron Metropolitan Housing Autho	7/18/2013	30030	7/28/2014	17218	
13	Akron Metropolitan Housing Autho	6/21/2013	23866	7/17/2013	22802	

	R_INSPECTION_DATE	MR_INSPECTION_COST	SECOND_MR_INSPECTION_DATE	SECOND_MR_INSPECTION_COST	CHANGE_IN_COST	PERCENT_CHANGE_IN_COST
1	27/2014	32874	1/28/2014	24047	8827	36.71
2	24/2014	35447	9/25/2014	13871	21576	155.55
3	23/2014	36279	9/24/2014	35447	832	2.35
4	6/2014	23194	8/8/2014	11472	11722	102.18
5	6/2014	26417	8/6/2014	23194	3223	13.90
6	4/2014	25943	8/5/2014	22940	3003	13.09
7	27/2014	22289	8/29/2014	11882	10407	87.59
8	25/2014	22730	8/26/2014	17208	5522	32.09
9	19/2013	38119	8/25/2014	22730	15389	67.70
10	12/2013	38177	8/19/2013	38119	58	0.15
11	28/2014	17218	7/29/2014	11493	5725	49.81
12	18/2013	30030	7/28/2014	17218	12812	74.41
13	21/2013	23866	7/17/2013	22802	1064	4.67