

**NAME: NEETI PANJWANI**

**NUID: 001382039**

**NORTHEASTERN UNIVERSITY**

**MPS ANALYTICS**

**SUBJECT: ALY 6030 Data Warehousing and SQL**

**CRN NO.: 20712**

**ASSIGNMENT 4**

**INSTRUCTOR NAME: PROF. Na Yu**

**Part 1**

1. Fact Table

Additive: Copay and insurancepaid (fact\_paymentByPatient)

Semi Additive: None

Non-Additive: None

1. Grain basically represents granular level of data. I have created only one fact table called PaymentByPatient which says which patient has paid for which drug. It has a field called copay and insurancepaid where in copay means payment made by the patient and insurance paid means payment made by insurance as per their cover plan

**Part 2**

1. Primary Key Designation and type

|  |  |  |
| --- | --- | --- |
| Table Name | Primary Key | Type of Primary Key |
| fact\_paymentByPatient | Id | Surrogate Key |
| dim\_drugBrand | drug\_brand\_generic\_code | Natural Key |
| dim\_drugDetails | drug\_ndc | Natural Key |
| dim\_drugType | drug\_form\_code | Natural Key |
| dim\_patientDetails | member\_id | Natural Key |

1. Foreign Key and its corresponding Primary Key

|  |  |  |  |
| --- | --- | --- | --- |
| Table Name | Foreign Key | Referenced To | PrimaryKey for the table |
| fact\_paymentByPatient | member\_id | dim\_patientDetails(member\_id) | Id (Surrogate Key) |
| fact\_paymentByPatient | drug\_ndc | dim\_drugDetails(drug\_ndc) | Id (Surrogate Key) |
| dim\_drugDetails | drug\_form\_code | dim\_drugType(drug\_form\_code) | drug\_ndc (Natural Key) |
| dim\_drugDetails | drug\_brand\_generic\_code | Dim\_drugBrand(drug\_brand\_generic\_code) | drug\_ndc (Natural Key) |

1. I have used Restrict when assigning my foreign key because my data is tightly dependent on all the table. One entry deletion from the master dimension table containing unique entry can create a lot of issue due to the structure I have defined for my dataset. The other reason for using restrict option is also that all the entry are related to healthcare industry, and according to government compliance, we have to retain all the old data in case of audits and other federal check. Hence, using restrict option is safe and recommended for the industry we are dealing with.

**PART 3**

A screenshot of a cell phone

Description automatically generated

**Part 4**

1. SQL Query:

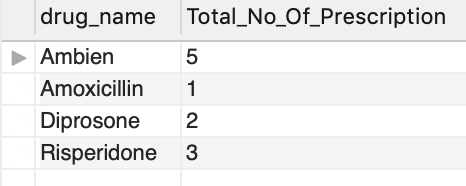
SELECT drug\_name, COUNT(\*) AS Total\_No\_Of\_Prescription FROM dim\_drugDetails

inner join fact\_paymentByPatient ON dim\_drugDetails.drug\_ndc=fact\_paymentByPatient.drug\_ndc

GROUP BY drug\_name

ORDER BY drug\_name asc;

Output:



5 prescription were filed for drug Ambien.

1. SQL Query:

SELECT COUNT(fact\_paymentByPatient.id) AS total\_prescription, COUNT(DISTINCT fact\_paymentByPatient.member\_id) AS member, SUM(fact\_paymentByPatient.copay) AS sum\_copay,SUM(fact\_paymentByPatient.insurancepaid) AS sum\_insurancepaid,

CASE WHEN dim\_patientDetails.age > 65 then "Older than 65"

WHEN dim\_patientDetails.age < 65 then "Younger than 65"

end as age\_filter

FROM fact\_paymentByPatient

INNER JOIN dim\_patientDetails ON fact\_paymentByPatient.member\_id = dim\_patientDetails.member\_id

GROUP BY age\_filter;

Output:

A screenshot of a cell phone

Description automatically generated

There is only 1 unique member older than 65 and they have filled 6 prescription.

1. SQL Query:

CREATE TABLE dim\_intermediatetable AS

SELECT fact\_paymentByPatient.member\_id,fact\_paymentByPatient.fill\_date,fact\_paymentByPatient.insurancepaid,dim\_patientDetails.first\_name as member\_first\_name,dim\_patientDetails.last\_name as member\_last\_name,dim\_drugDetails.drug\_name,row\_number() OVER (PARTITION BY fact\_paymentByPatient.member\_id ORDER BY fact\_paymentByPatient.fill\_date DESC) AS ROWID

FROM fact\_paymentByPatient

LEFT JOIN dim\_patientDetails ON fact\_paymentByPatient.member\_id=dim\_patientDetails.member\_id

LEFT JOIN dim\_drugDetails ON fact\_paymentByPatient.drug\_ndc=dim\_drugDetails.drug\_ndc;

SELECT member\_id, member\_first\_name, member\_last\_name, drug\_name, fill\_date,insurancepaid FROM dim\_intermediatetable WHERE ROWID=1;

Output:

A screenshot of a cell phone

Description automatically generated

Most listed drug for 10003 was Ambien and their insurance paid was 322.