****

**Pharmacy Claims**

#### ALY6030-Mod6: Final Assignment

#### Integrated Data Warehousing and SQL

Instructor: [Kayal Chandrasekaran](https://northeastern.instructure.com/courses/140410/users/121784)

29th June 2023

**INTRODUCTION**

Sample pharmacy claims for fictitious insurance company members are handled by a Pharmacy Benefit Manager (PBM) in the "ALY 6030 Final Project Data Set.xlsx" file. When the whole claims data is ready, the goal of this project is to build up a test database and pre-program SQL queries to match reporting requirements. It is necessary to standardize the data and transform it into relational tables (3NF). The MySQL database will be given primary and foreign keys, and an entity relationship diagram (ERD) will be made to show the table structure. Additionally, example queries for analytics and reporting purposes will be created.

**QUESTIONS**

**PART 1:**

Three distinct files are made using the following dimensions and fact tables, along with the relevant column names, in order to transform the raw data into relational tables that adhere to 3NF standards:

1. Dimension Table: dim\_member

- member\_id (Primary Key)

- member\_first\_name

- member\_last\_name

- member\_birth\_date

- member\_age

- member\_gender

2. Dimension Table: dim\_drug

- drug\_ndc (Primary Key)

- drug\_name

- drug\_form\_code

- drug\_form\_desc

- drug\_brand\_generic\_code

- drug\_brand\_generic\_desc

3. Fact Table: fact\_prescription

- member\_id (Foreign Key referencing dim\_member.member\_id)

- drug\_ndc (Foreign Key referencing dim\_drug.drug\_ndc)

- fill\_date

- copay

- insurance\_paid

Three different CSV files with the names "dim\_member.csv," "dim\_drug.csv," and "fact\_prescription.csv" are used to save these tables. Copay and insurance\_paid are fact variables in the fact\_prescription table that may be combined across dimensions like member, drug, and fill date. They are additive variables.

**PART 2:**

The specified column names are expected to be adequate for uniquely identifying the rows with regard to the main keys and foreign keys. The following are the desired actions for deletion or update for the foreign keys in the fact\_prescription table:

1. For the foreign key member\_id referencing dim\_member:

- Deletion: CASCADE

- Update: CASCADE

2. For the foreign key drug\_ndc referencing dim\_drug:

- Deletion: CASCADE

- Update: CASCADE

It is ensured that when a referred entry in the parent table is removed or updated, the related rows in the child table (fact\_prescription) will likewise be deleted or updated by choosing CASCADE for both deletion and update operations. As a result, the consistency and referential integrity of the tables are preserved.

**PART 3:**

Please refer to the ERD diagram (Star Schema Representation) in Figure 1 for a visual representation of the tables and their relationships.

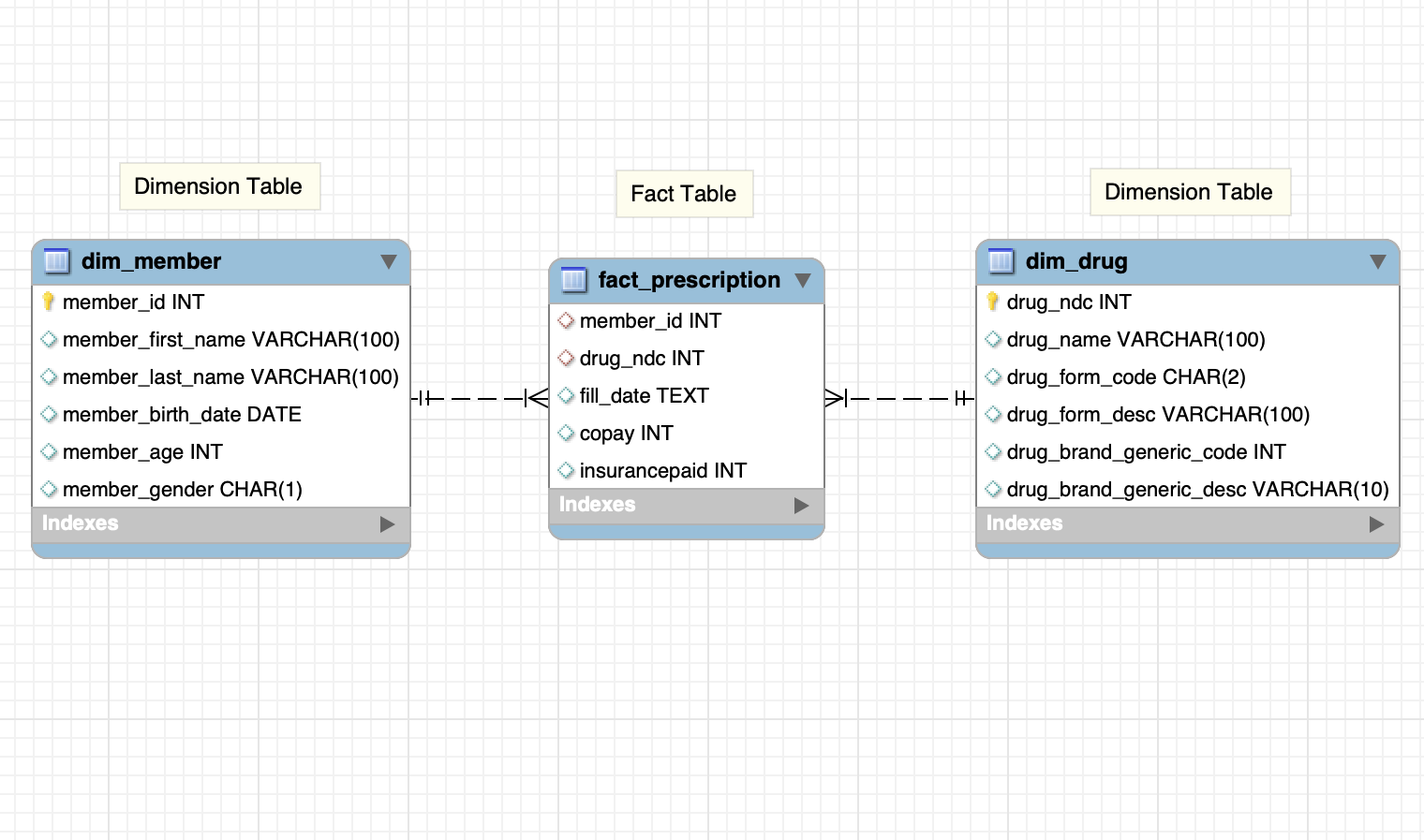


Fig1

**PART 4:**

**1:**

To determine the number of prescriptions by drug name, the following query can be used:

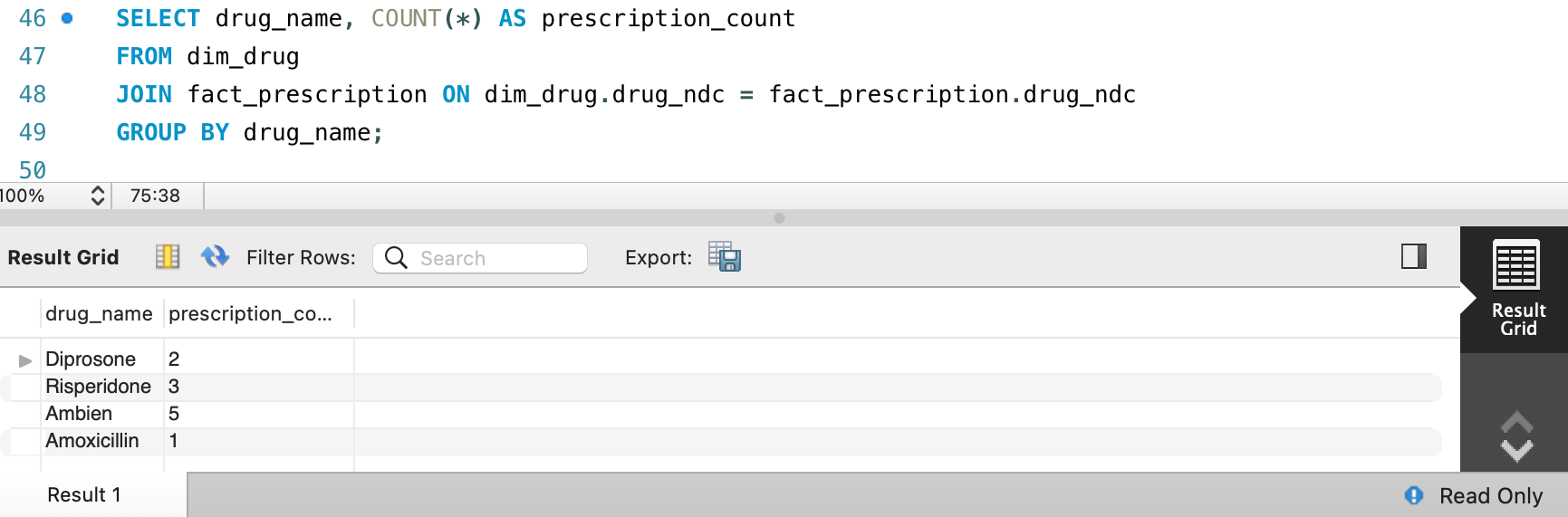


Figure 2 shows the result of the query, indicating that there are 5 prescriptions filled for the drug "Ambien."

**2:**

The following search may be used to get data on age, total prescriptions, unique members,total copays, and total insurance paid:

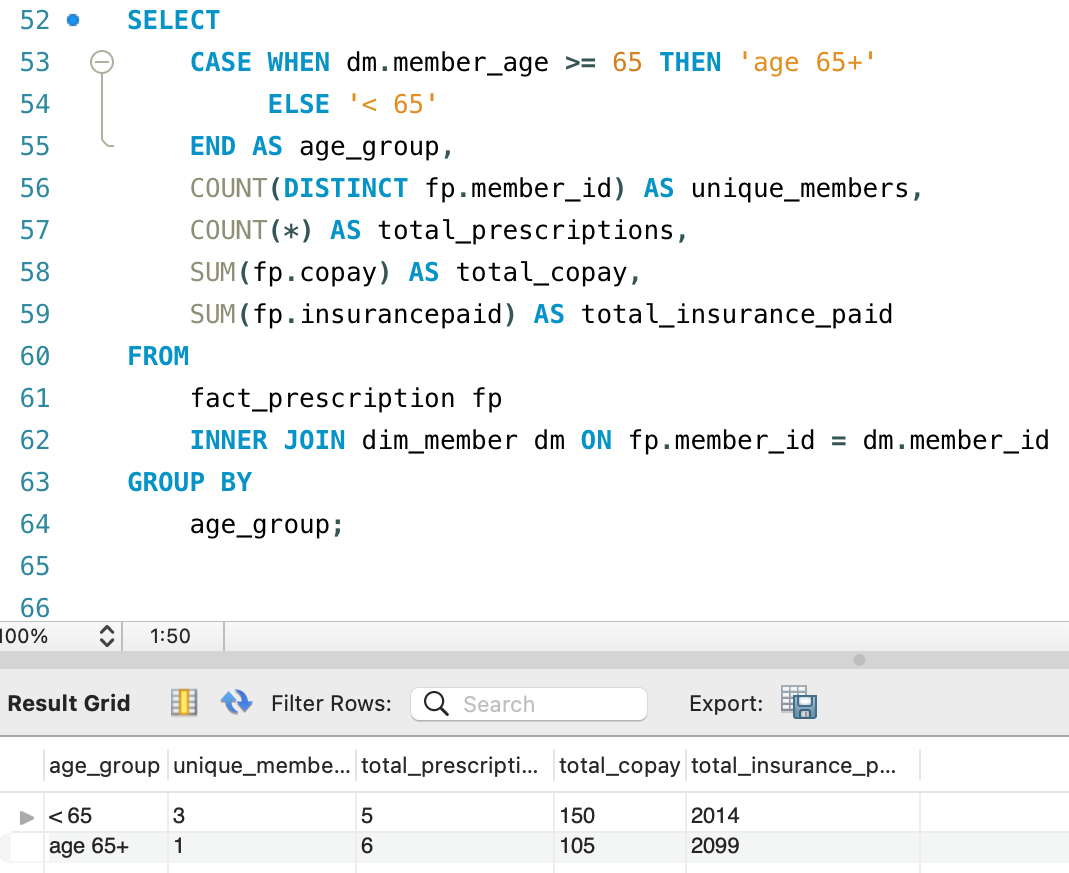
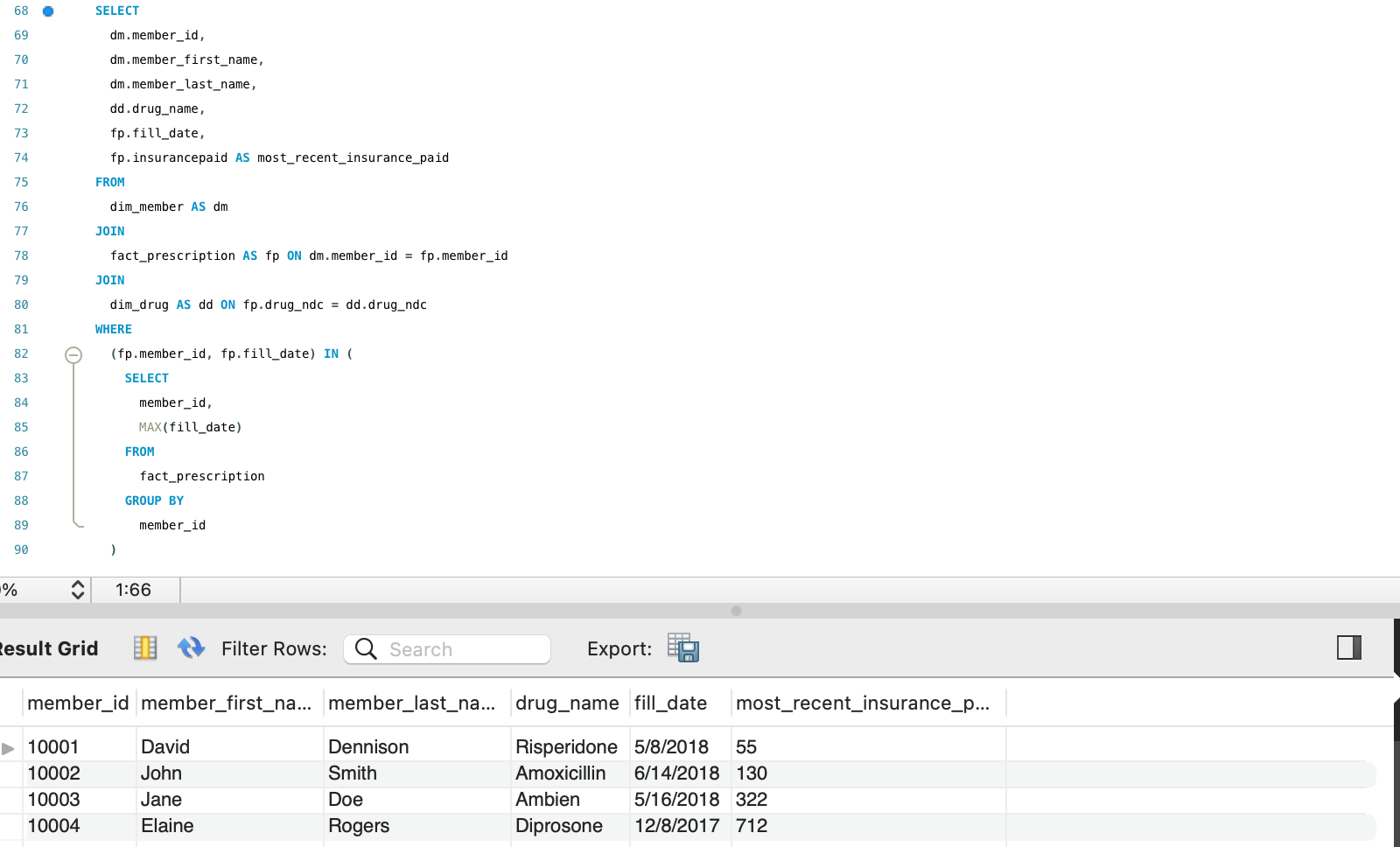


Figure 3 presents the result of the query, showing the age group over 65 has 1 unique member and a total of 6 prescriptions filled.

**3:**

The following query may be used to get the sum paid by insurance for the most recent prescription fill date for a given member:



**CONCLUSION**

In this project, we successfully converted the unprocessed pharmaceutical claims data into a 3NF-compliant normalized relational database. Two dimension tables (dim\_member and dim\_drug) and one fact table (fact\_prescription) were constructed, and they were all saved as distinct CSV files. In order to guarantee data integrity and referential integrity between the tables, MySQL assigned primary and foreign keys. To illustrate the database's star schema organization, an entity relationship diagram (ERD) was made. Additionally, we created model SQL queries to evaluate the data, revealing statistics for various age groups, the number of prescriptions written for specific drugs, and the amount of money paid by insurance for the most recent prescription fill date. Once the complete claims data is available, these queries will act as a starting point for further analysis and reporting.