****

#### ALY6030-Mod1: Assignment 1- Tech crunch

#### Integrated Data Warehousing and SQL

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

2nd June 2023

Submitted By: Srishti Singh

**Problem 1. Normalization**

**1) What is a good choice for a primary key here? In contrast, give an example of an attribute (or composite) that would *not* be a valid primary key.**

**Answer:** The "fund\_id" column is a good choice for a primary key in this dataset because it uniquely identifies each funding event and can serve as a unique identifier for each row. To confirm that it satisfies the definition of a primary key, you can sort the data by the "fund\_id" column in Excel and ensure that there are no duplicate values.

On the other hand, the "company" column would not be a valid primary key because it can have duplicate values. Since multiple funding events can be associated with the same company, it does not uniquely identify each row in the table.

**2) Does the table satisfy 1NF? Why or why not?**

Since there are no repeated groups or multiple values in any of the columns, the table meets the first normal form, or 1NF. One attribute is represented by each column, which satisfies the criteria of 1NF.

**3) Does the table satisfy 2NF? Why or why not?**

The dataset's "numEmps" column, which denotes the number of workers, has a primary key relationship with the "company" column rather than the "fund\_id" column. This suggests that the identification of the firm, rather than the identification of the financing event, determines the number of workers.In the context of normalization, it denotes a partial dependence when a nonkey column (in this example, "numEmps") solely depends on a portion of the main key (in this case, "fund\_id").The dataset should be broken up into different tables in order to eliminate partial dependencies in order to meet 2NF.As a result, it seems from the supplied dataset that the claim that the table does not meet 2NF because the number of workers depends on the business ID rather than the fund ID is false.

**4) Does the table satisfy 3NF? Why or why not?**

Due to the table's failure to meet 2NF and transitive dependencies between non-key attributes, the 3NF (Third Normal Form) constraint is not satisfied. The supplied table contains transitive dependencies, which means that every non-key attribute doesn't rely only on the primary key. For instance, columns like "number of employees" appear to depend more on the fund\_id than the firm.

5)

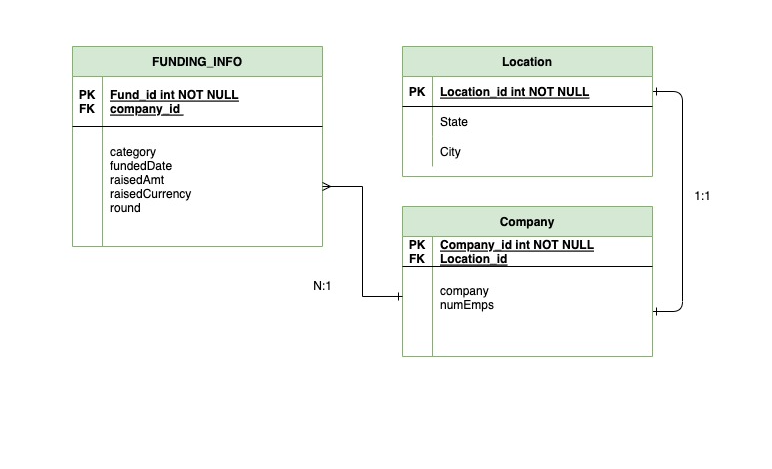
**To attain the 3NF in the provided Dataset, we can define the relationships between the tables as follows:**

**One-to-One Relationship:**

Location table and Company table: Each company is associated with one location, and each location is linked to one company. This is a one-to-one relationship because each company can have only one location, and each location can be associated with only one company.

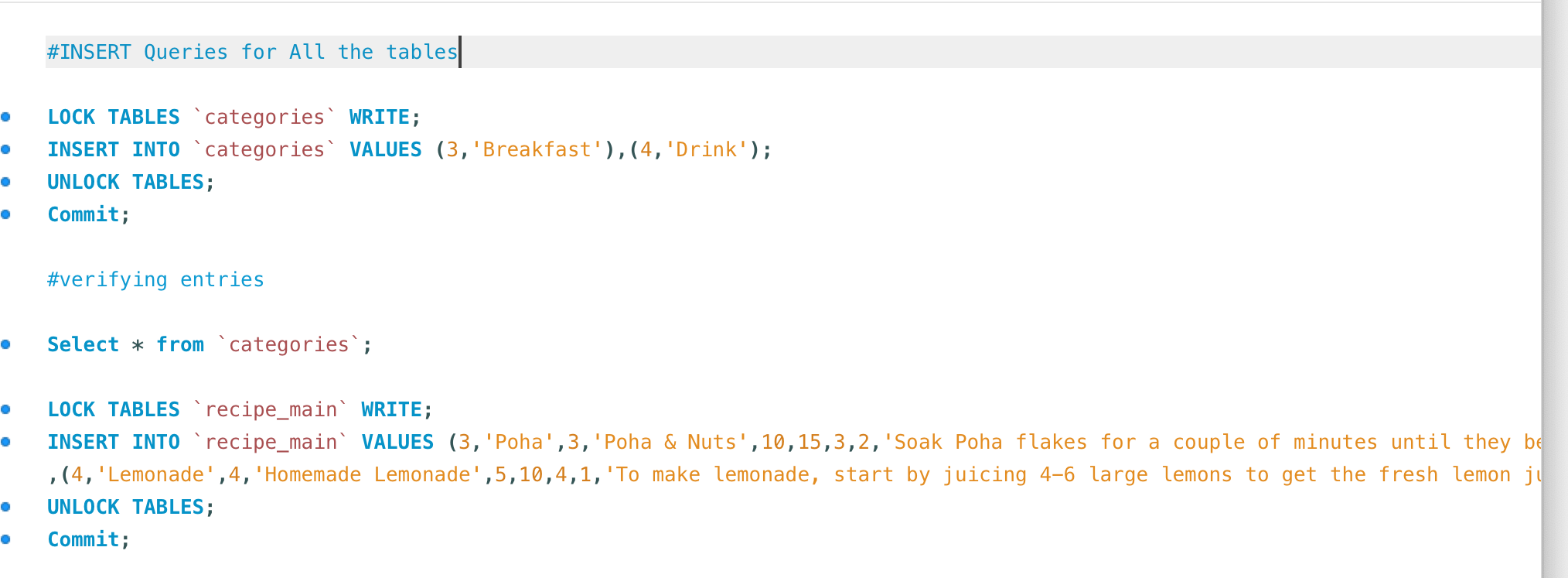
**One-to-Many Relationship:**

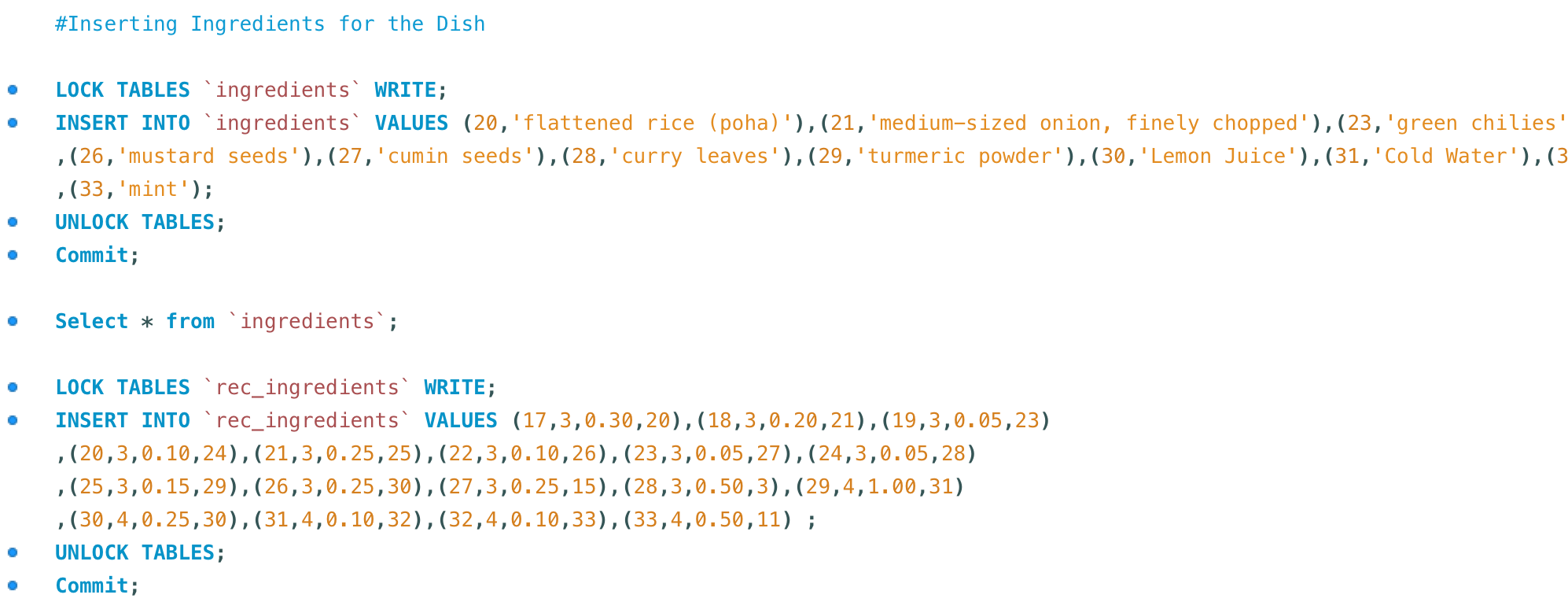
A firm may have several funds, but only one fund is linked to each company in the company table and the fund table. One firm may own numerous funds, but each fund may only be tied to one company, making this a one-to-many connection.



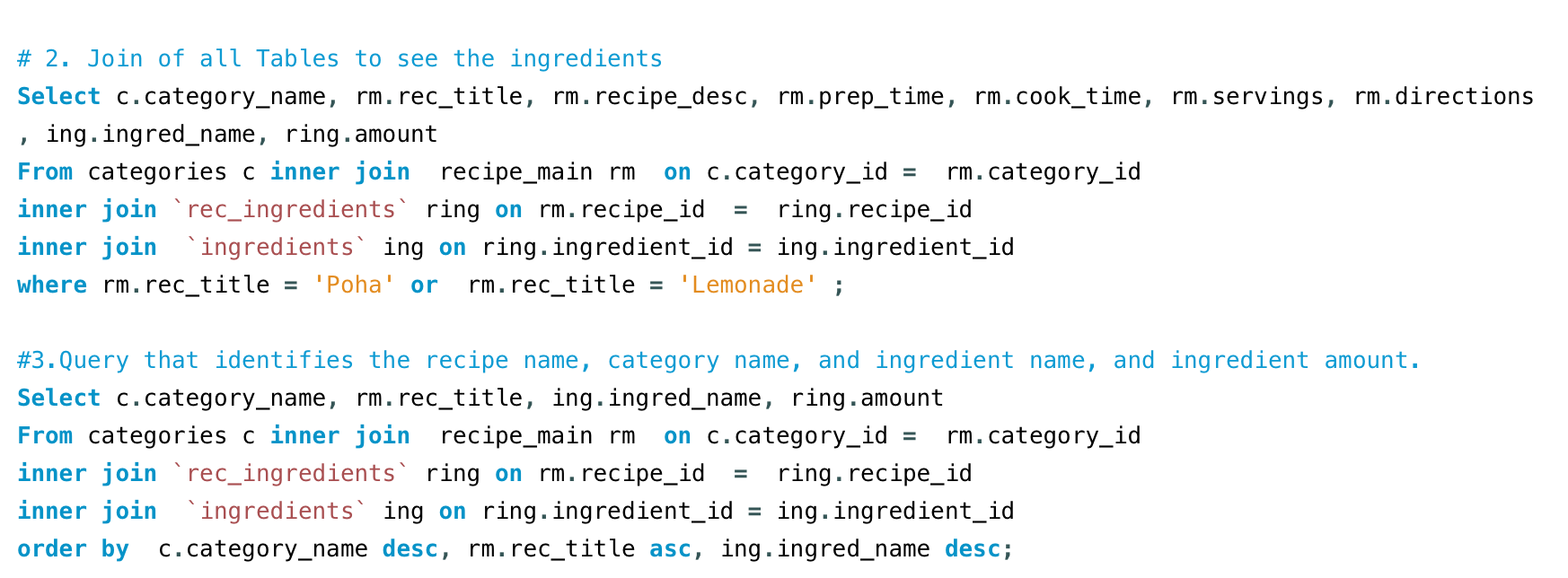
**Problem 2. Case study (adapted from Comeau, Chapter 9)**

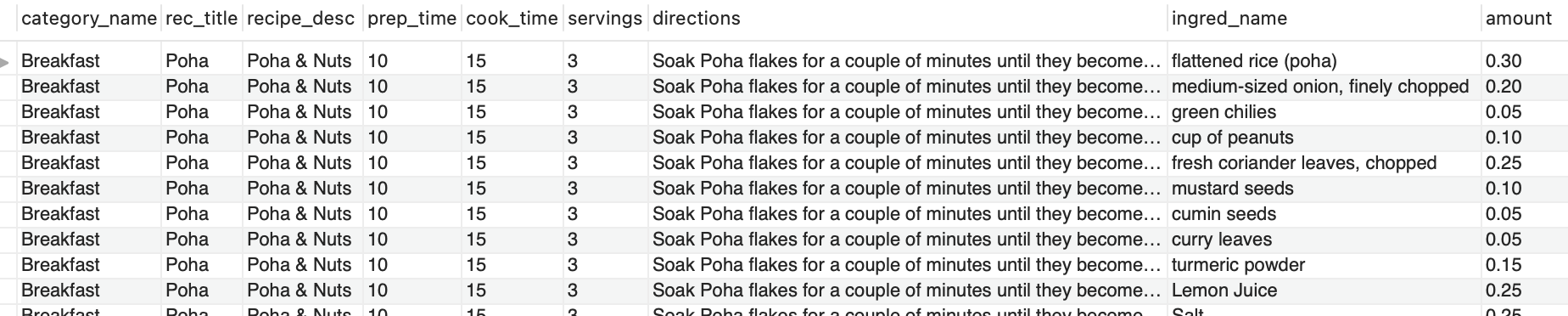
1. **Inserting entries into table for dishes and drinks i.e., Poha and lemonade**



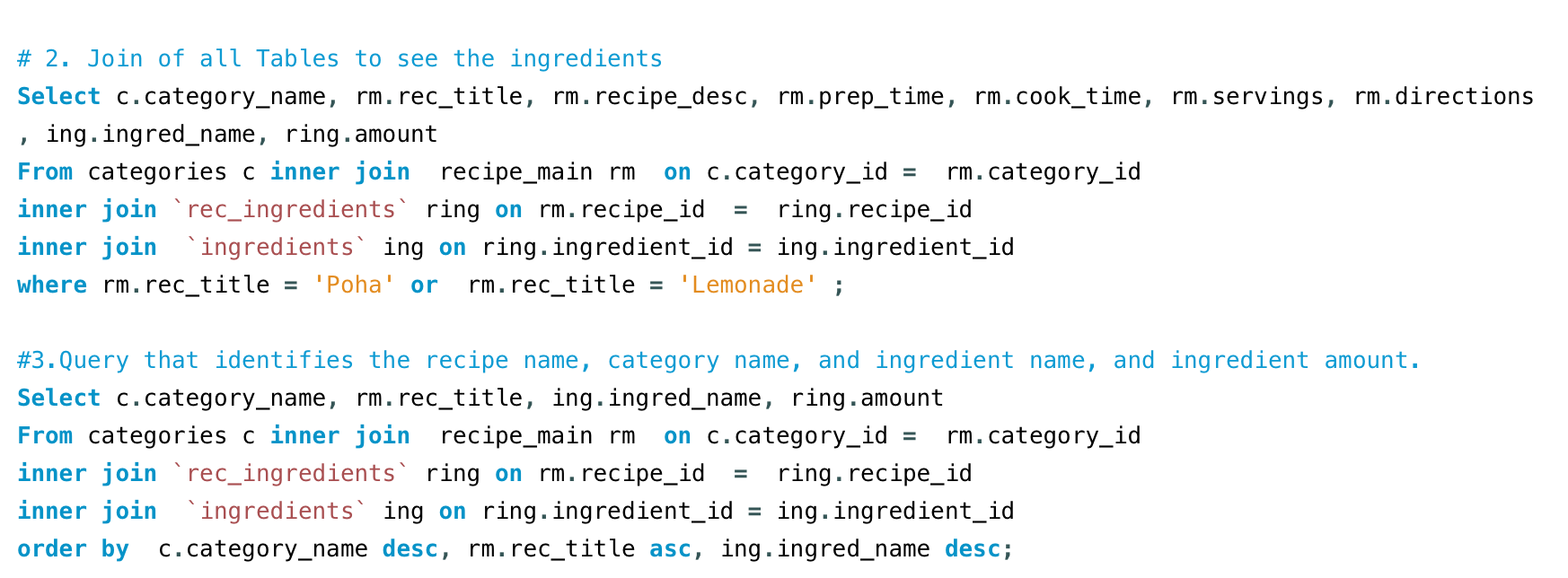


1. Query showing all relevant information from all four of the tables





1. **Final Query Result**

****

****