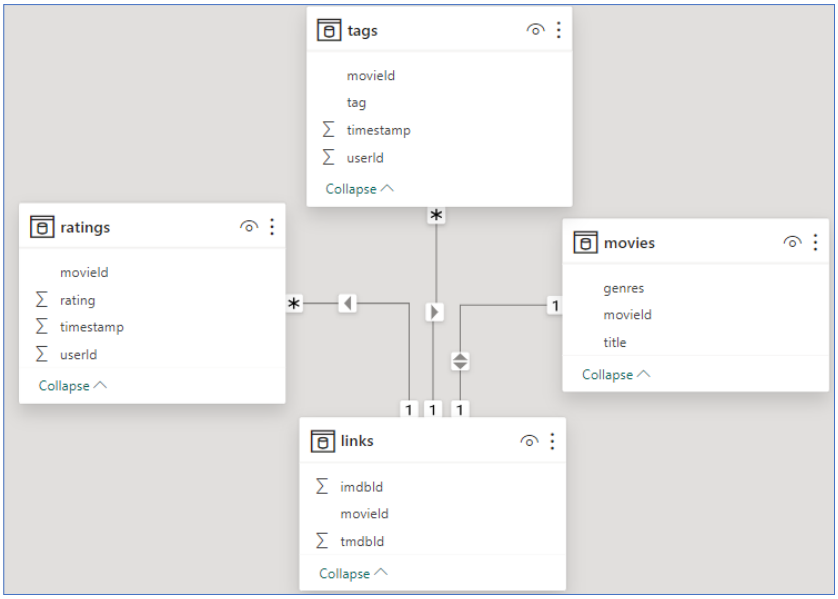
Dipika Jothinathan – Homework Day 3

2)

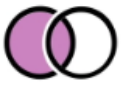
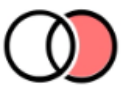


3) Explain why Data Modeling is key?

Data modeling makes it easier for developers, data architects, business analysts, and other stakeholders to view and understand relationships among the data in a database or data warehouse. In addition, it can reduce errors in software and database development.

4) Explain different types of data joins?

A join kind specifies how a merge operation will be performed. The following table describes the available join kinds in Power Query.

Join kind	Icon	Description
Left outer		All rows from the left table, matching rows from the right table
Right outer		All rows from the right table, matching rows from the left table
Full outer		All rows from both tables
Inner		Only matching rows from both tables
Left anti		Only rows from the left table
Right anti		Only rows from the right table

5) How to set up relationships among Tables / Entities?

If two or more tables are queried at the same time, when the data is loaded, Power BI Desktop attempts to find and create relationships *automatically*. Power BI Desktop looks at column names in the tables queried to determine if there are any potential relationships. If there are, those relationships are created automatically. However, the *Manage relationships* dialog box can be used to manually create or edit relationships.

6) Explain the difference between Regression vs Forecasting? When to use Forecasting and when to use Regression Techniques?

Linear regression is a method used to model the relationship between a dependent variable and one or more independent variables. It is used to quantify the relationship between variables to make predictions. Linear regression assumes the relationship between the variables is linear and is often used for predictive modeling and understanding the impact of one variable on another.

Time series forecasting is a technique used to make predictions based on historical time-ordered data points. Data is collected at regular time intervals and the goal is to predict future values based on patterns and trends observed. Time series forecasting considers the time component of the data and is used to make predictions about future values of a variable based on its past behavior.

In summary, while linear regression is used to model the relationship between variables and make predictions based on that relationship, time series forecasting is specifically focused on making predictions based on time-ordered data points.

7) Explain what is seasonal effect and how does it impact the forecasting?

Seasonality is a time series feature in which data experiences predictable variations that repeat every calendar year. Seasonal patterns are defined as predictable fluctuations that recur over a period of one year. In a business context, seasonality refers to predictable commerce changes that occur based on a particular season.

When correctly used, a seasonality forecast can decrease business costs and increase bottom-line profitability. Predicted seasonal variation can inform business decisions ranging from inventory levels to staff hiring initiatives. Seasonality forecasting methods can assist business professionals with stock and economic trend analysis. Seasonal forecasting provides businesses with critical data to assist with preparing for upcoming business operations.

8) Explain why Star model is useful in OLAP?

A star schema is a type of denormalized data model that is widely used in data warehousing and OLAP systems. It consists of a central fact table that contains the measures or metrics of interest, and multiple dimension tables that contain the attributes or dimensions of the data. The fact table has a foreign key for each dimension table, and each dimension table has a primary key that uniquely identifies each row. The star schema resembles a star, with the fact table in the center and the dimension tables as the points.

A star schema has several benefits for data analysis and query performance, such as simplifying the data structure and reducing the number of joins required to retrieve data from multiple tables. Additionally, it enables fast aggregation and filtering of data along different dimensions, which is useful for OLAP operations like slicing, dicing, drilling down, and rolling up. Moreover, it supports dimensional modeling, a logical approach to design data for business intelligence and analytics that focuses on the business questions and perspectives. Furthermore, it facilitates the use of pre-computed summary tables or indexes that can speed up query execution and analysis.

9) What is Time-Series Analysis?

Time series is a series of data points indexed at successive equally spaced points in time. Time series analysis comprises methods for analyzing time series data to extract meaningful statistics and other data characteristics. Time series data analysis can show how variables change over time. It provides an additional source of information and a set order of dependencies between the data.

Time series analysis typically requires many data points to ensure consistency and reliability. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting-predicting future data based on historical data.