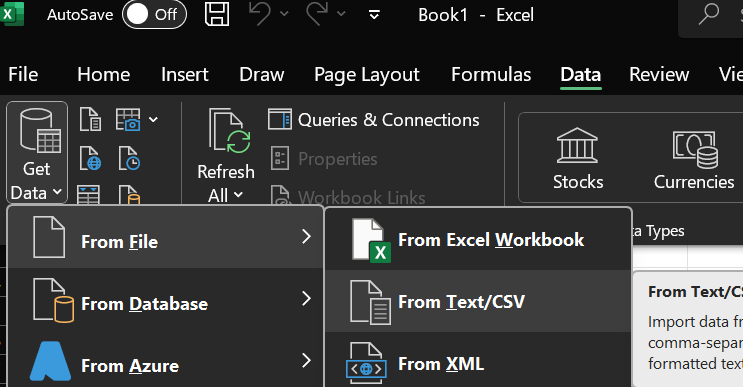
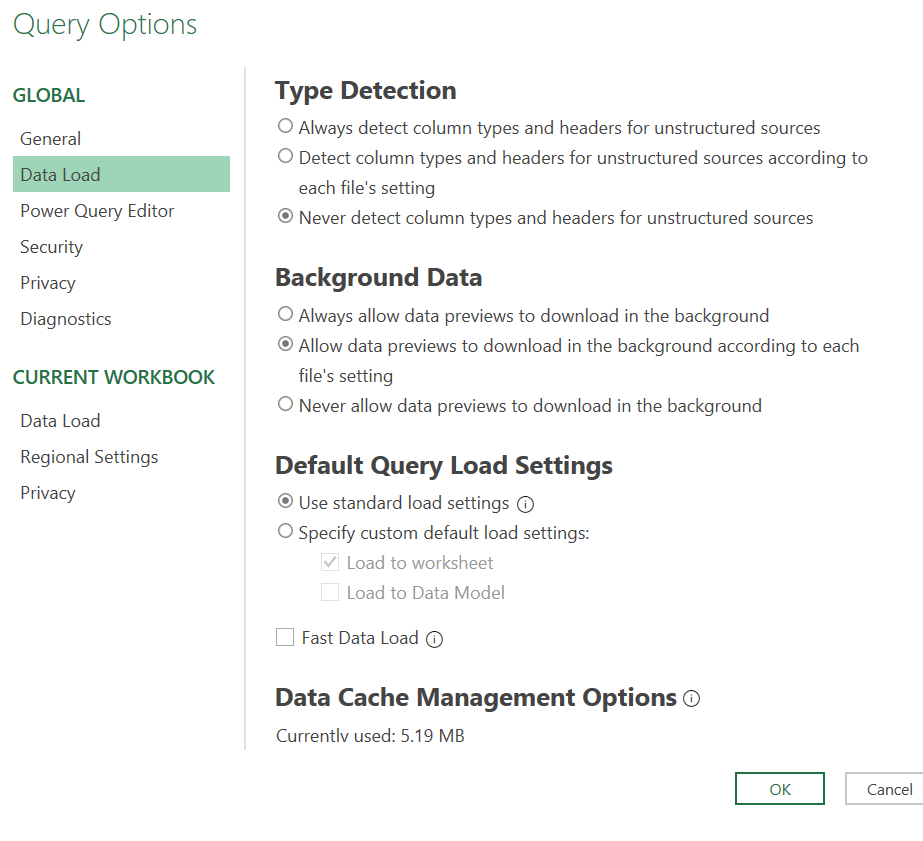
**GROUP 13- Rudrani Sakhare, Sophie Devekins, and Mohammad Al Najjar.**

**This is a Written Manual on Data Normalization in Power Query, converting CSV files to bis, and loading them on Microsoft Azure MySQL Server for Analysis through Powershell Or Command Prompt.**

DOWNLOAD DATA FILE IN CSV FORMAT.

1. **Open Excel, Click the DATA Tab to** 
2. **Make Sure to change the Type Detection Settings on Power Query Editor:**

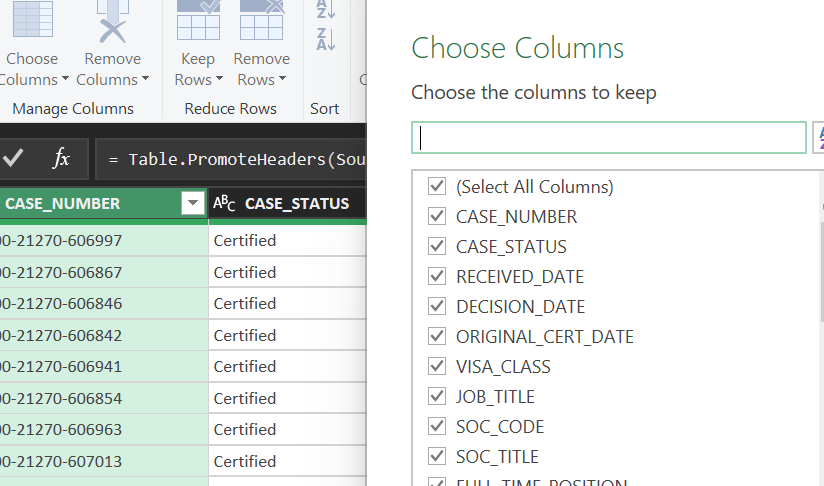
**Click on File > Options & Settings > Query Options**

1. **Make a Duplicate of the main file to work on.**

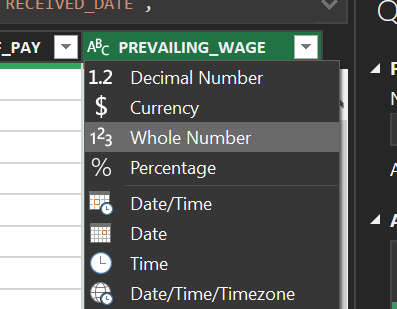
**Read and Understand the data available in the file all columns and their Descriptions.**

1. **Select the columns that you want to work on**

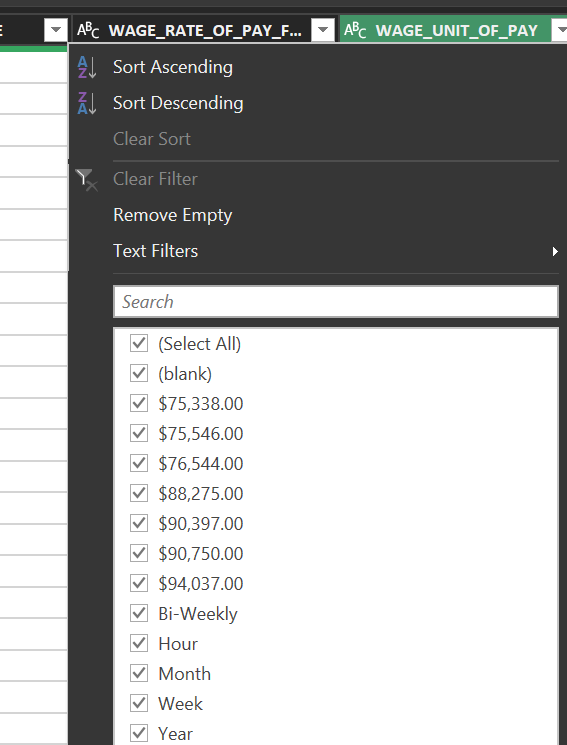
**Click Home > Choose Columns**



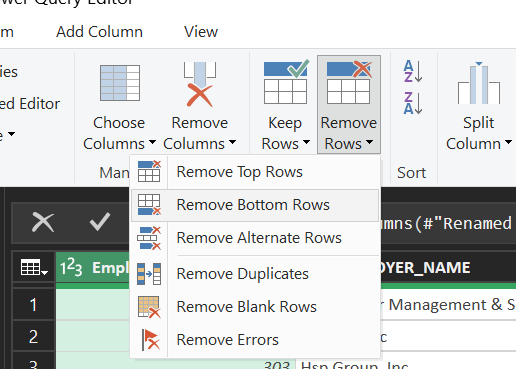
1. **Change Data Types where needed**

**Right-click on Column Name**

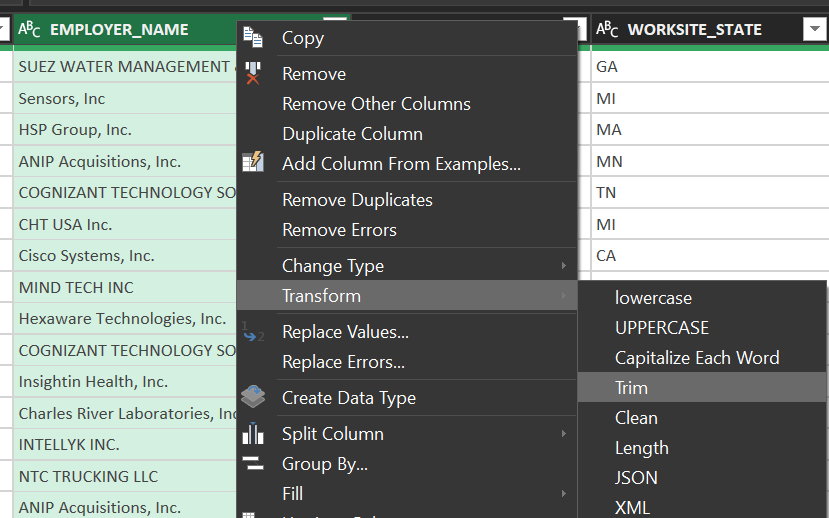
1. **Filter the columns with the Data relevant to your Analysis or Hypothesis.**

**Click the Arrow Next to the Column Name.**

1. **Or Use Remove rows from the Bottom or Top to Clean the Data**

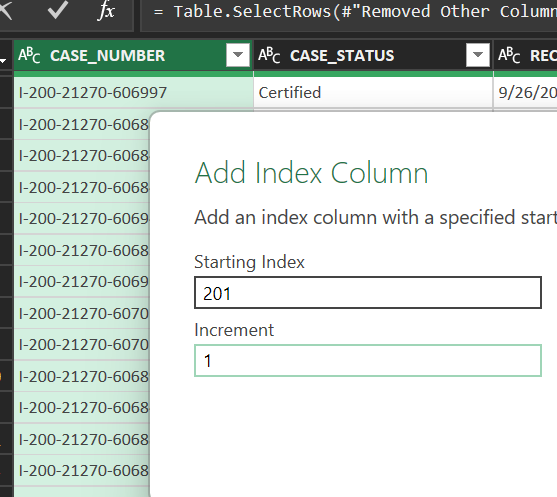
**Home > Remove Rows**

1. **Transform Columns For more Uniformity, This will make it easier when you try to remove duplicates or give Indexes.**



1. **Give Unique Keys or Indexes for Columns of Significant Data**

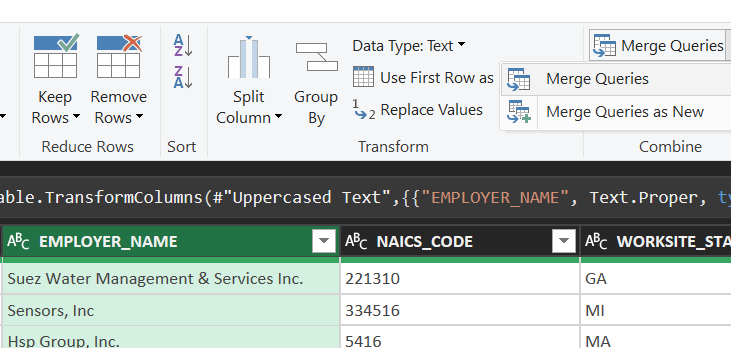
**NOTE: DO NOT FORGET TO REMOVE DUPLICATES BEFORE ASSIGNING KEYS.**

**Select Coloumn> Add Column Tab > Add Custom Index**

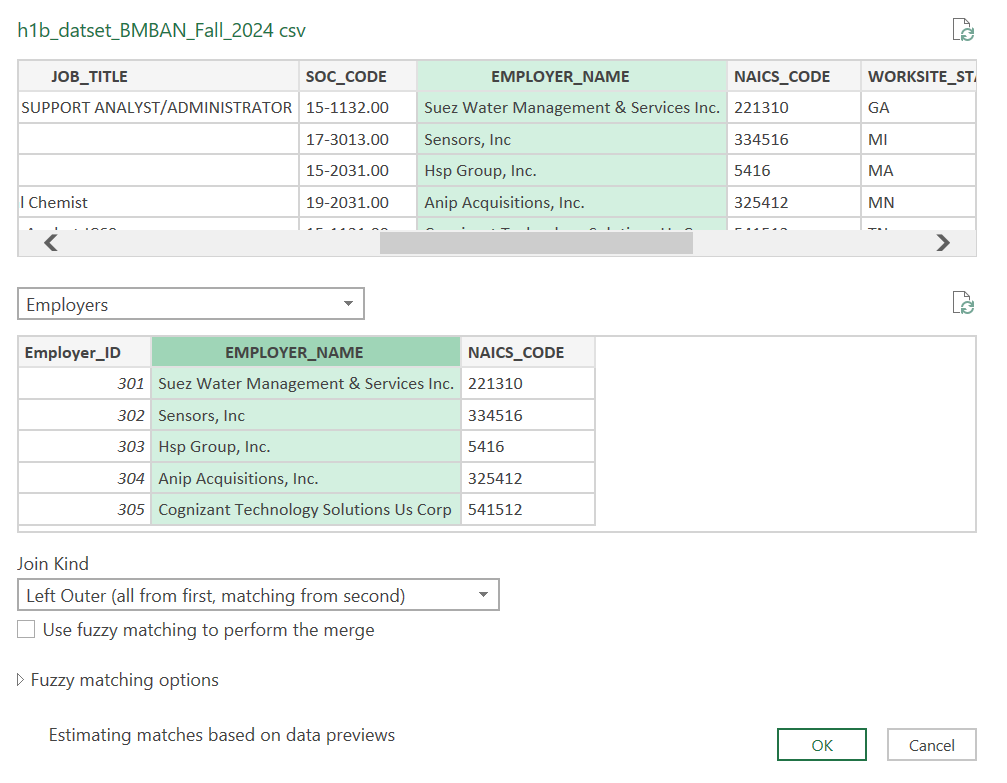
1. **Giving Unique Indexes or Keys Helps you Merge Queries From different Tables**

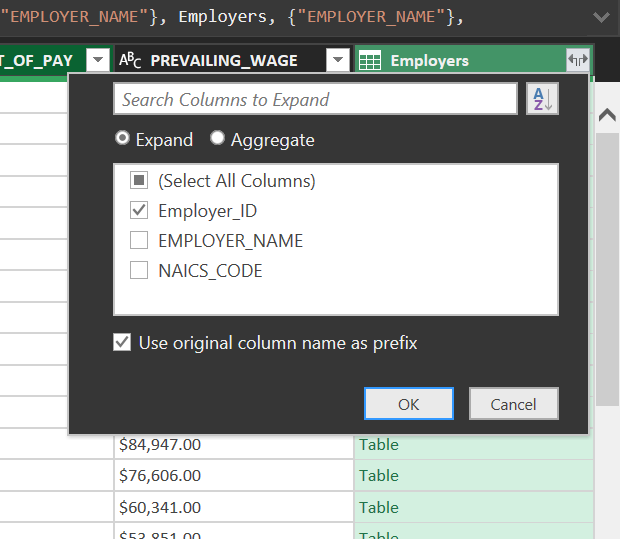
**Note: Make sure you normalize and transform the data that you will be joining the two**

**tables on**

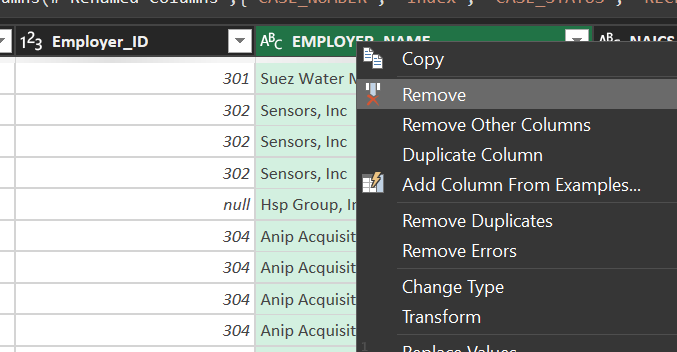
**Click Home > Merge Queries**

1. **Select The Columns you want to join the two tables on.**

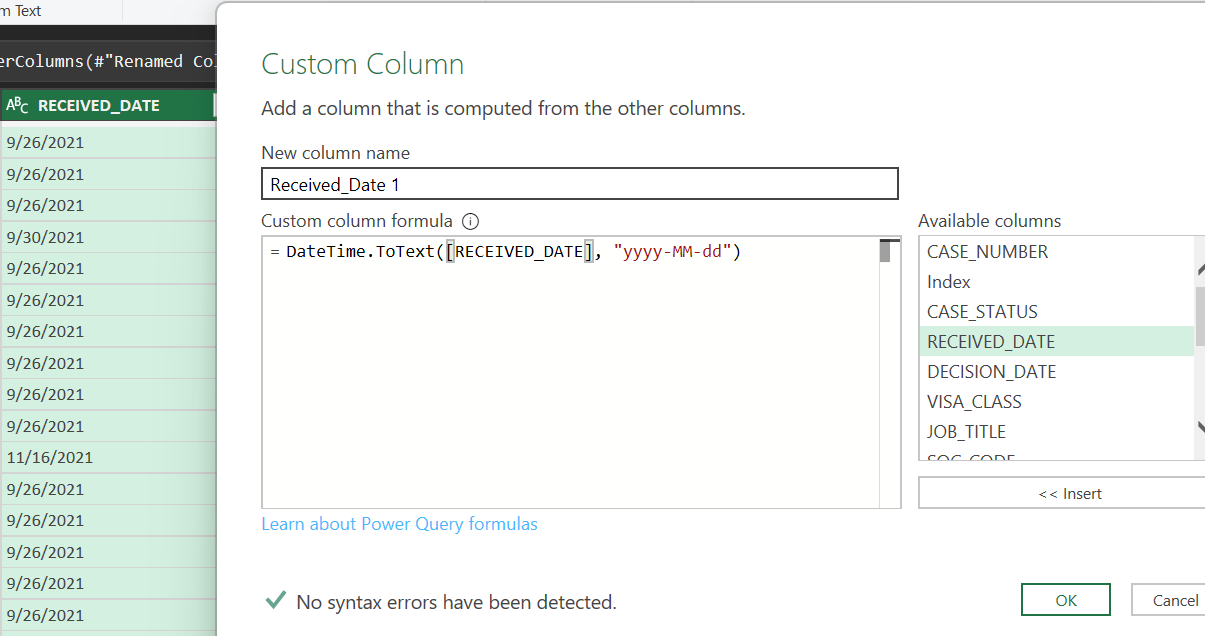
**Click on the Column from the first Table > Select another table > Select the column from the other table** 

1. **Select the column you want from the other table**
2. **Remove the columns that are Unnecessary after merging.**

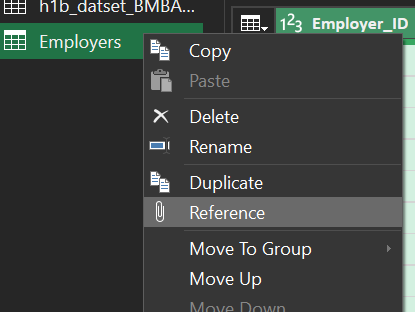
**Right-click on the Column Name > Remove**



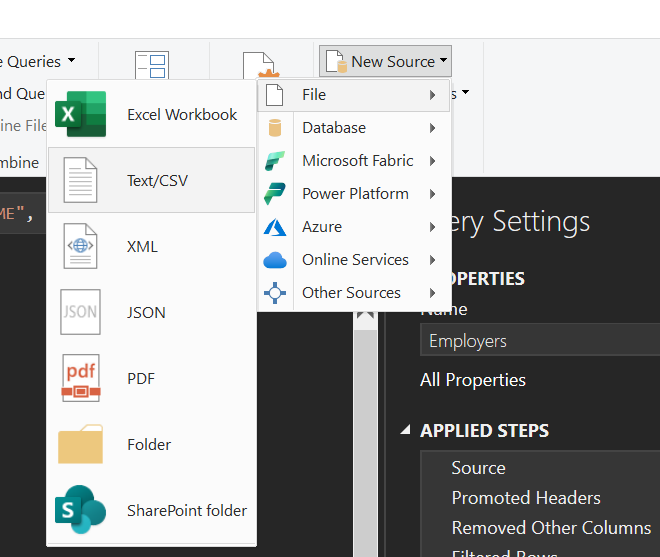
1. **Transform Dates Through the Custom column and Use the Formula to change the Date Format to yyyy-MM-dd. This can be used when the data is large otherwise Column By**

**Examples can be used.**

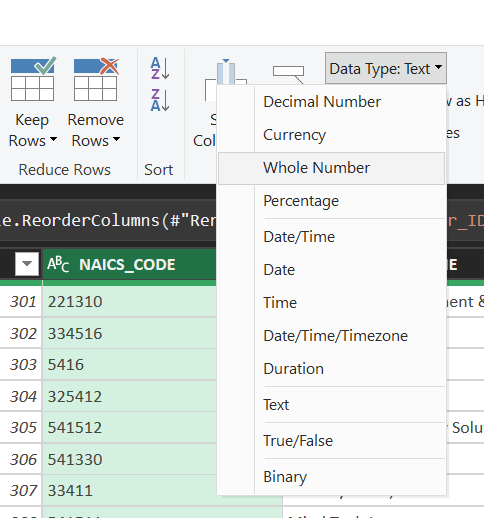
1. **Make a point to Use Referencing while making new Tables for Data Reusability, Performance Optimization, and Modular Workflow.**

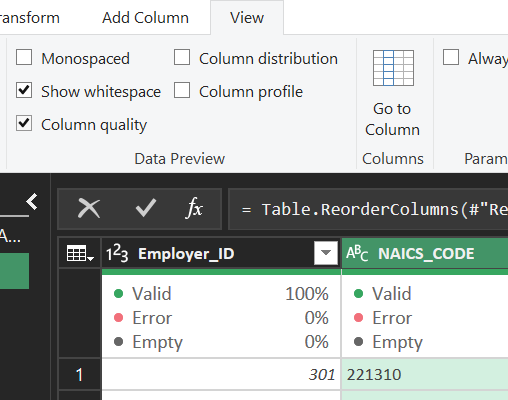


1. **To add New Files, Workbooks, and Folders to the existing Power Query**

**Click Home > New Source.** 

1. **Another Way to Change the Data Types**

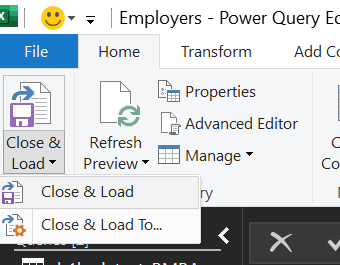
**Home > Data Type:’ ‘.**

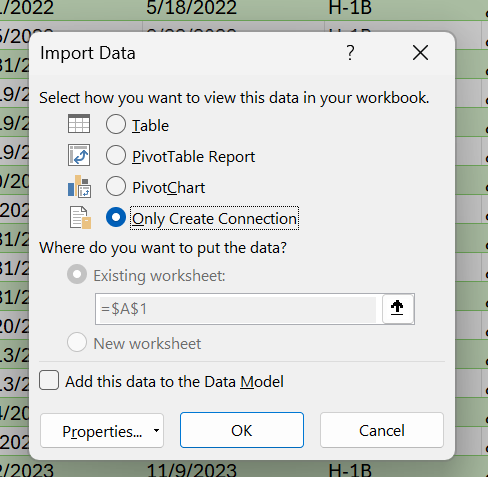
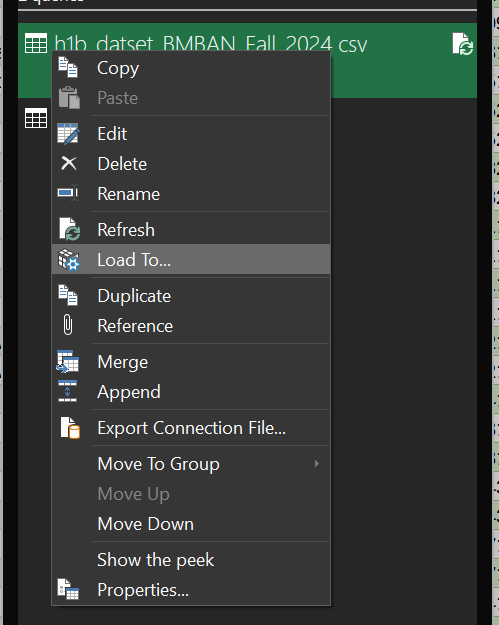
1. **Another practice that can be followed to Check the Data Quality and Integrity is by using** 

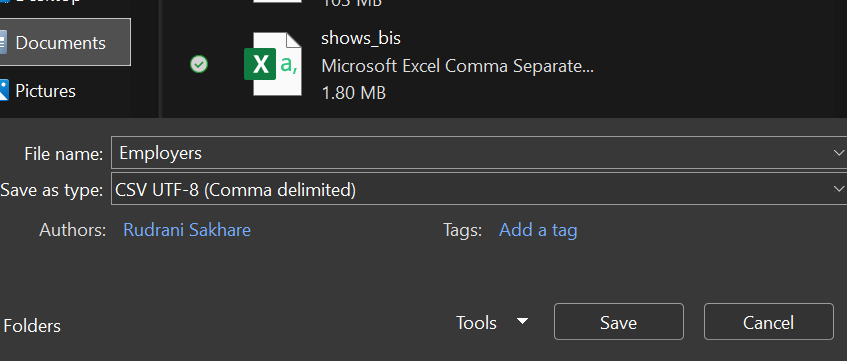
**Column Quality**

**View > Check Column Quality.**

1. **Check all Columns and the Data to make sure it is as you want it to be to run the analysis.**

**Then Close and Load the Data on Excel.**

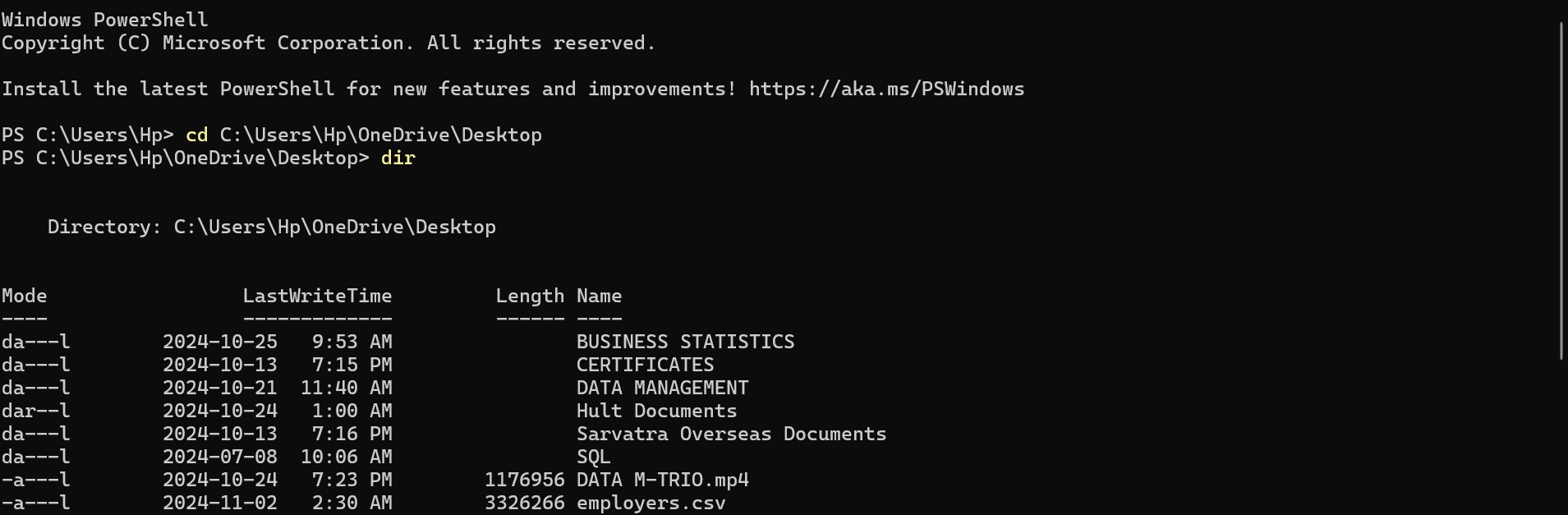
1. **To avoid errors and the Main Dataset make sure to Load it as a connection. This Helps us keep the data without deleting it.**
2. **Download the Excel Sheets in CSV- UTF 8 Format.**



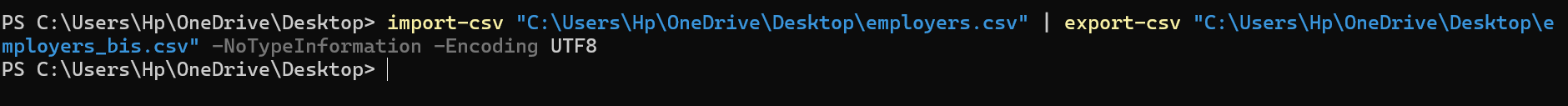
**Converting Files From CSV UTF-8 to Bis Files.**

1. **Open Powershell and Run the Command ‘cd’ with the path for the folder of your CSV file**

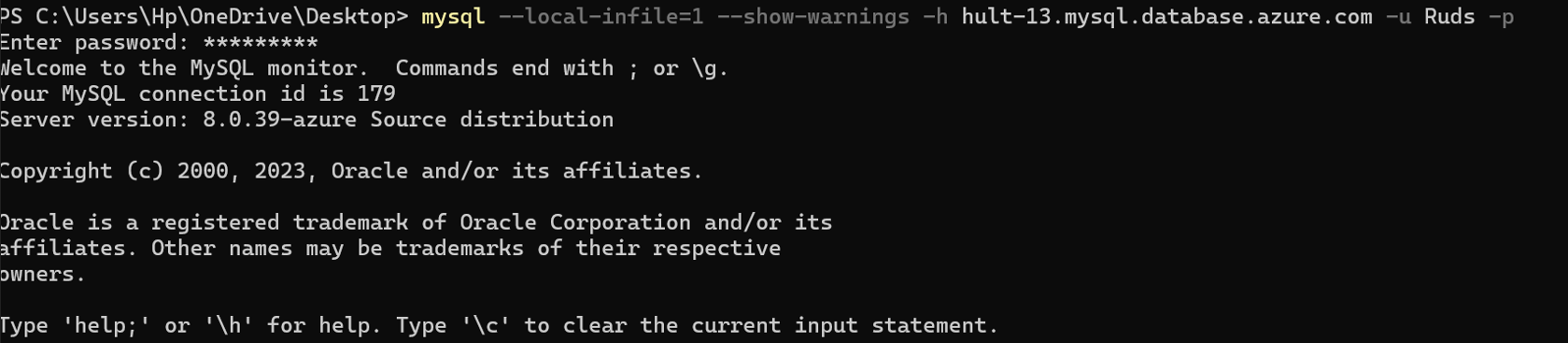
**Eg. cd C:\Users\Hp\OneDrive\Desktop**

**dir C:\Users\Hp\OneDrive\Desktop ( to view the files)**

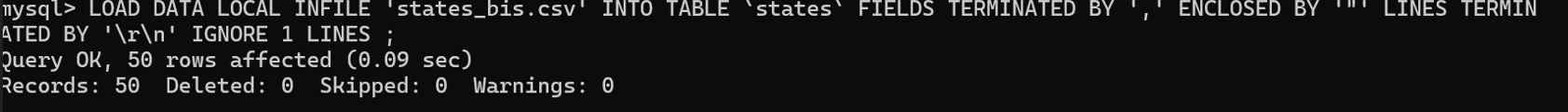
1. **Converting CSV UTF8 to Bis CSV.**



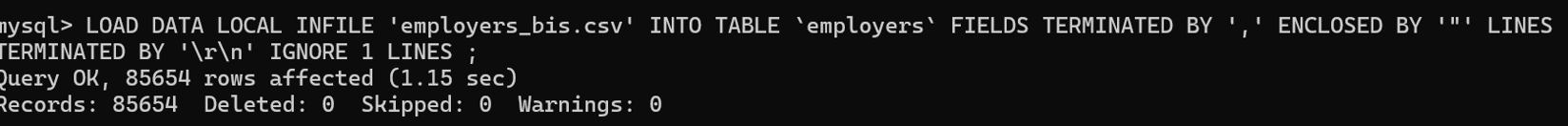
1. **Login to the MySql Server that you created in Microsoft Azure through Powershell.**

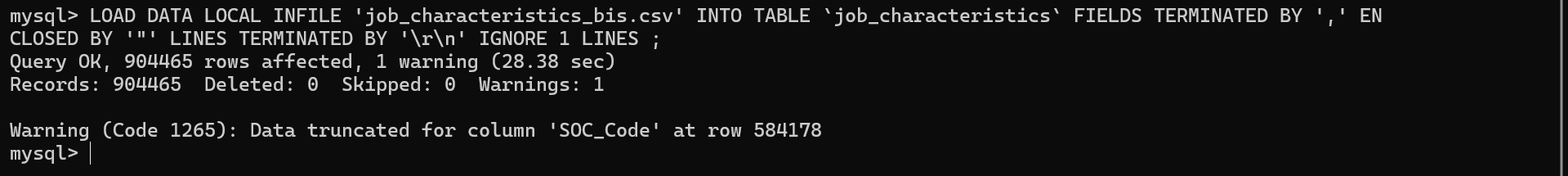
**NOTE: Add the correct Hostname, Assigned User Name, and Password.** 

1. **Use SQL Syntax to Choose the Database you are Working on.**
2. **Load all the Bis files one at a time in the sequence of Grandparents>Parents>Child table**
   1. **Loading of states table through PowerShell**



* 1. **Loading of employers tables through PowerShell.**



* 1. **Loading of job\_characteristics table through PowerShell.**

**Your MySQL Database is ready to do analysis.**

**Access to the Server:**

**User: prof\_luis**

**Password: 12345**

**Hostname: hult-13.mysql.database.azure.com**