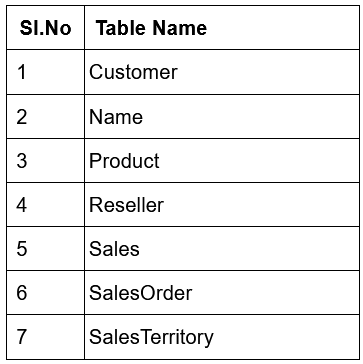
***ADVENTURE WORKS [Beginner] Problem Statement: -***

The AdventureWorks sales excel workbook captures all the details of a fictitious, multinational manufacturing company called Adventure Works Cycles, including the details of customers, products, sales, territory etc. Load, transform and build a visualization report from the following pointers.  
  


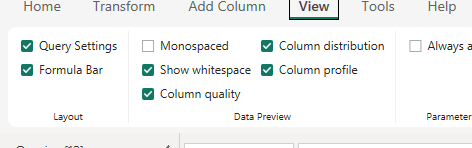
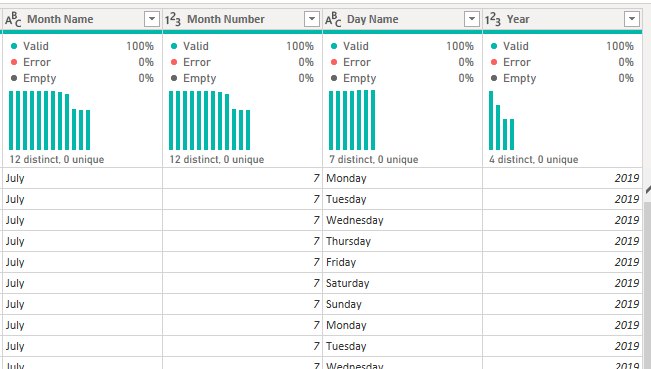
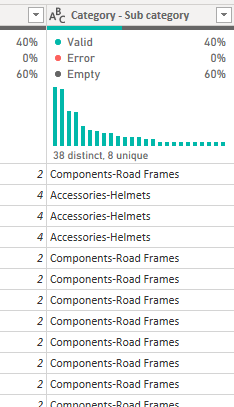
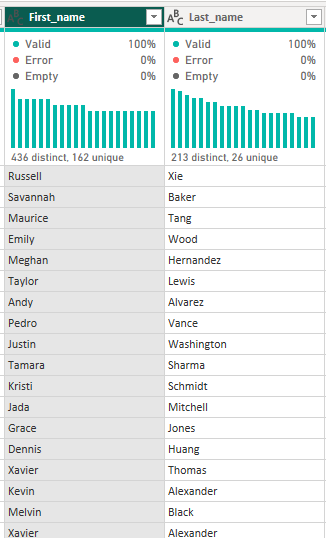
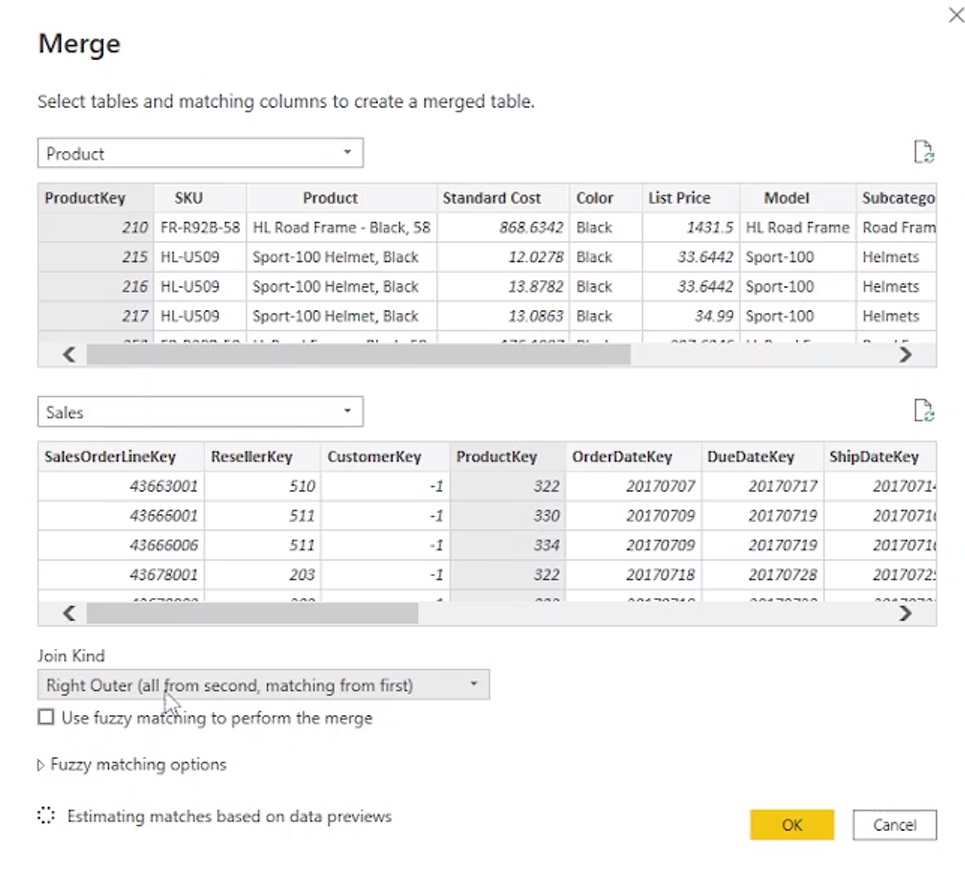
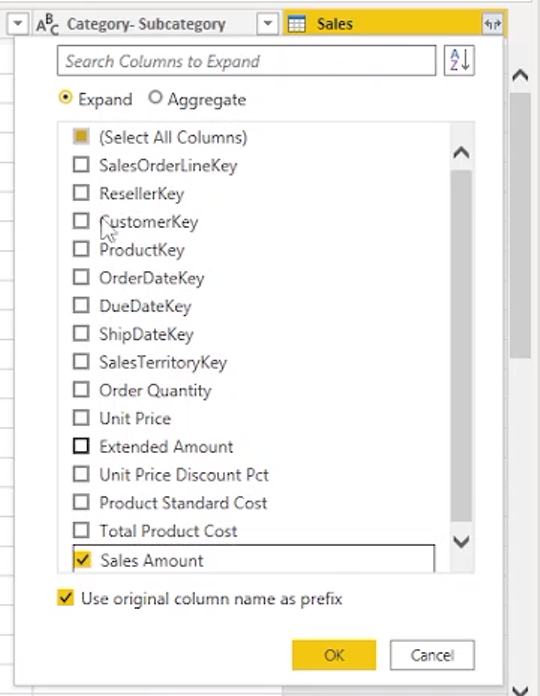
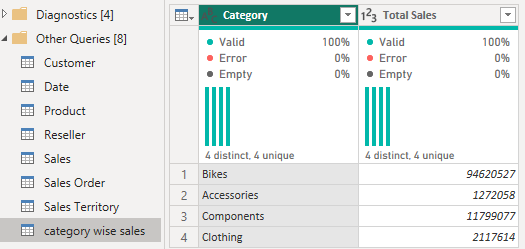
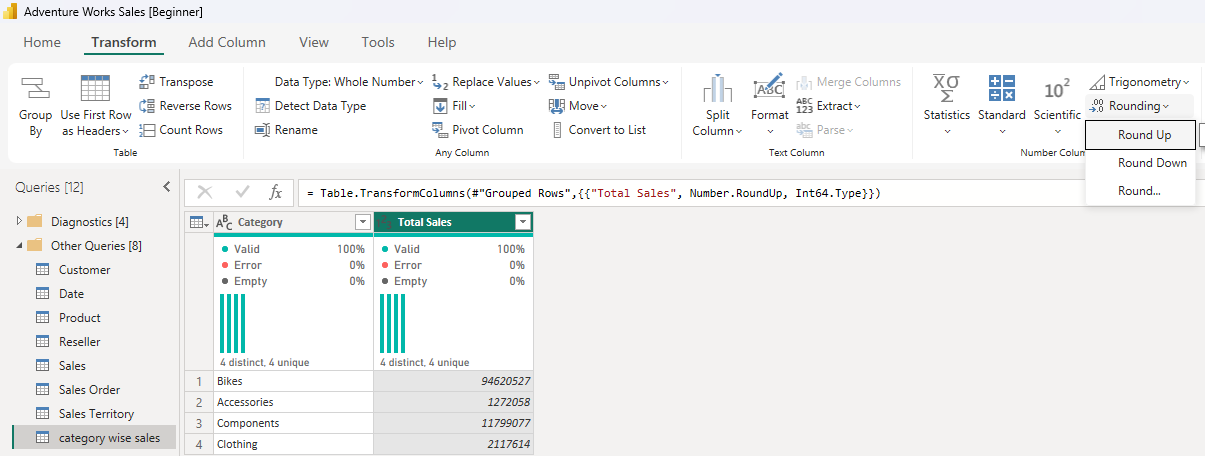
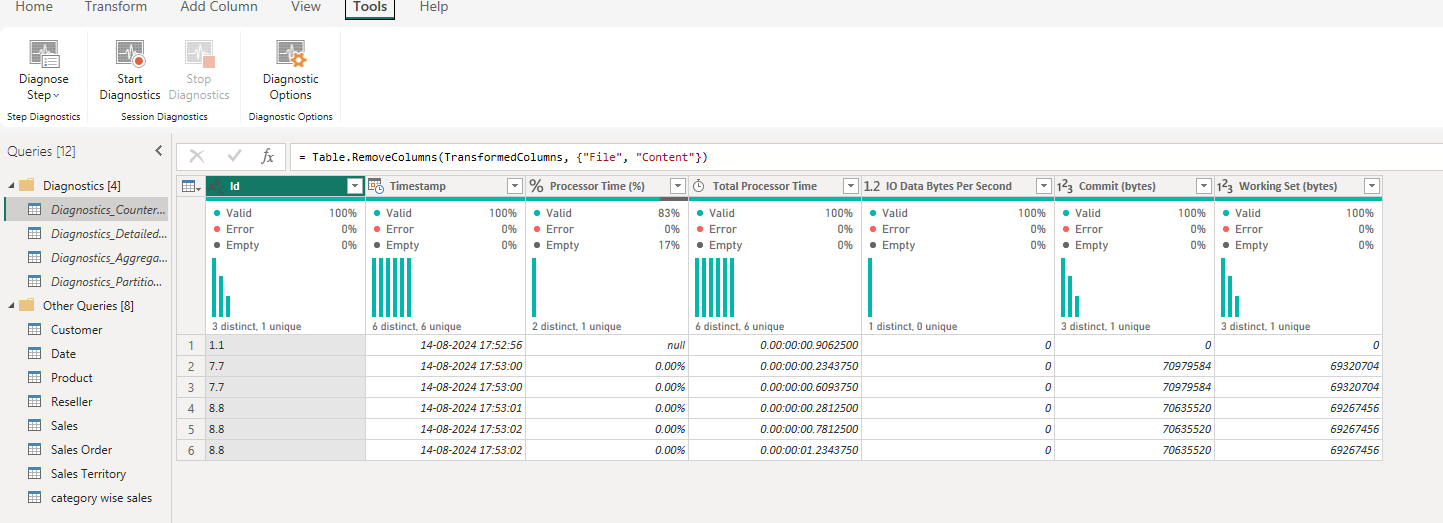
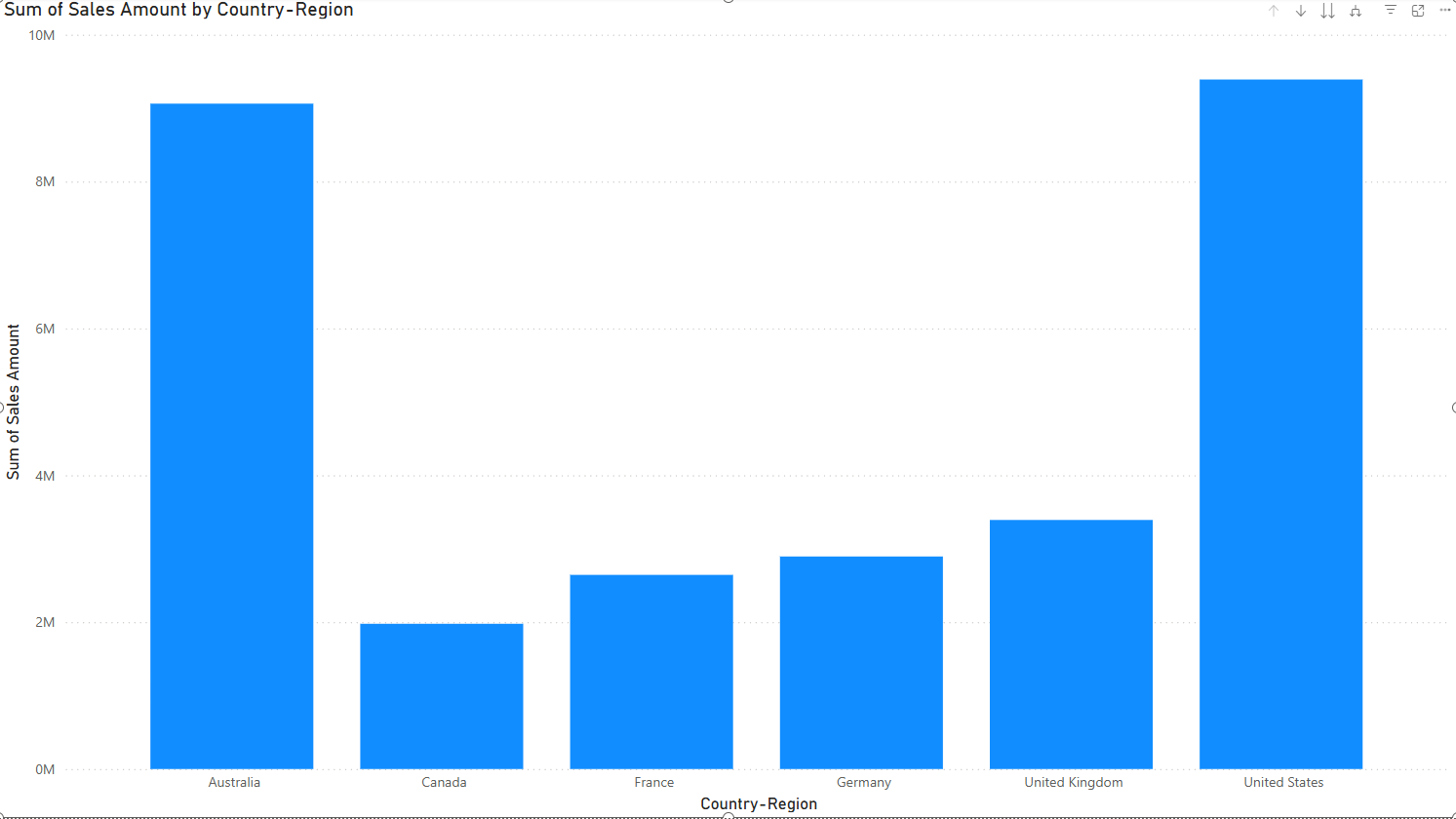
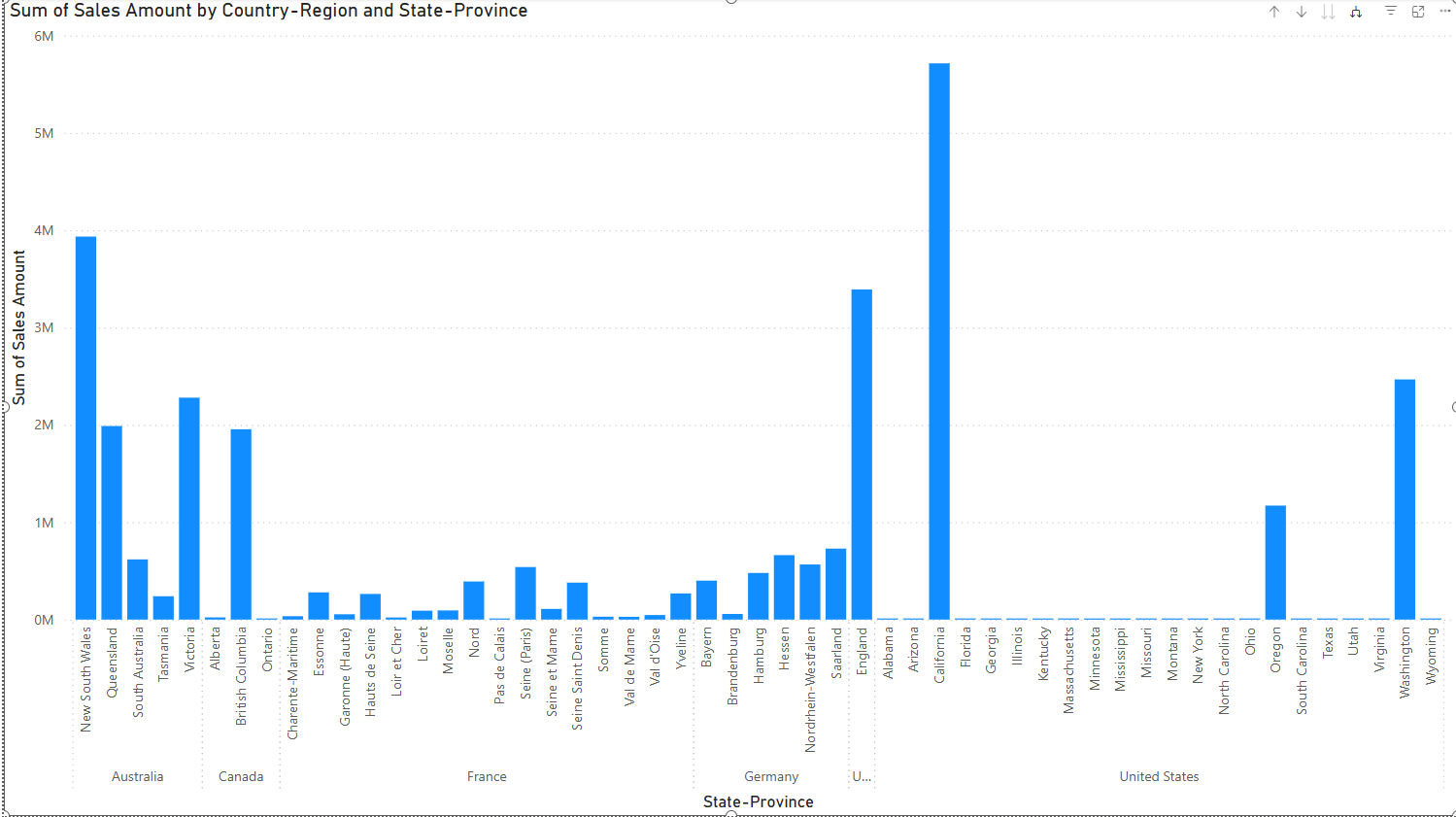
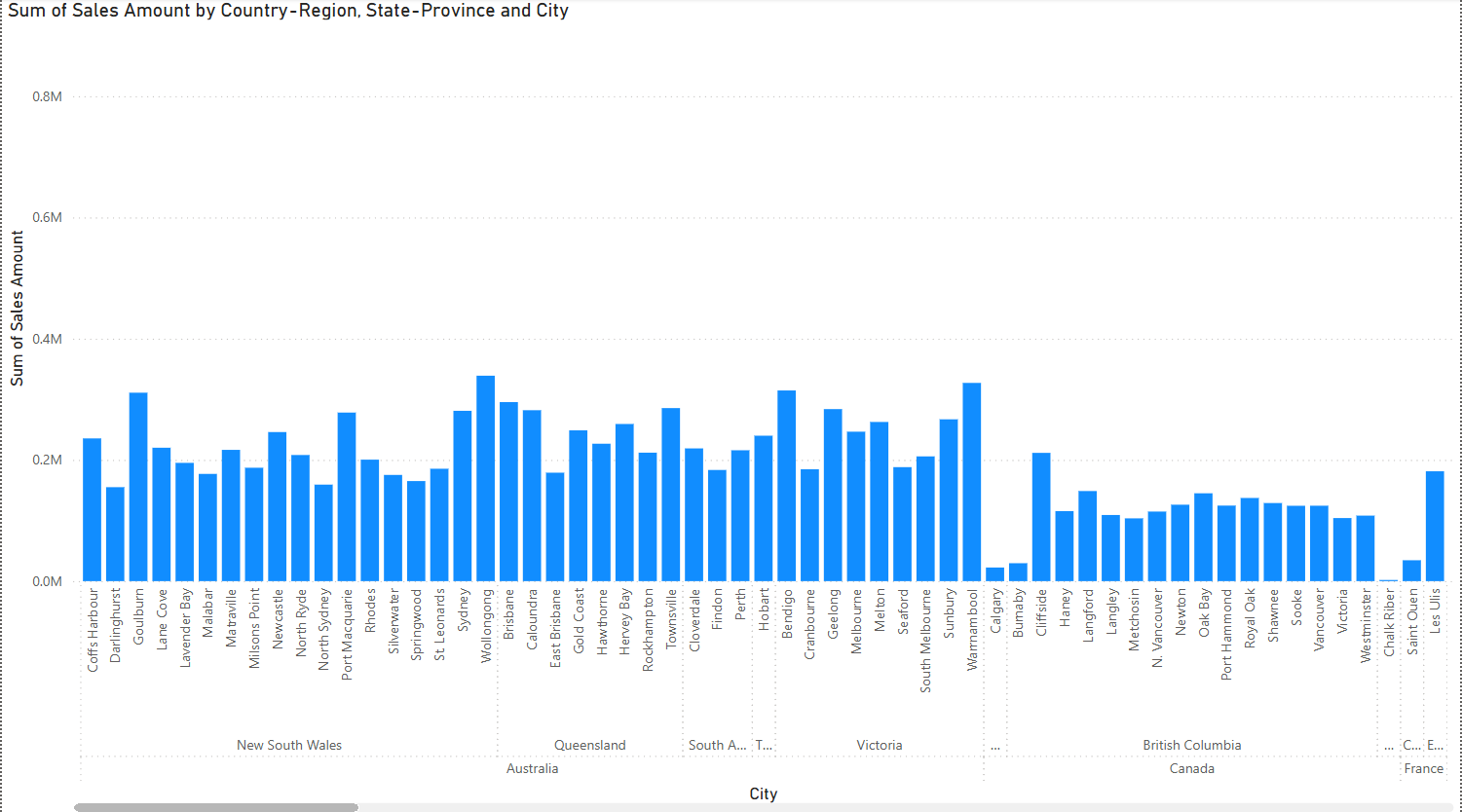
**Data transformation, using Power query editor:**

1. Load AdventureWorks Sales data in Power BI Desktop.  Do not select the sheets (labelled \_data).
2. Check the data type of all columns and do a column profiling [Hint: Power Query editor ->View->Check Column profile, column quality, column distribution]
3. Extract Month Name, Month No, Day Name and Year from Date Column
4. Create a new column ‘Category - Subcategory’ by merging the respective columns in the Product table, using hyphen (-) as separator
5. Separate Customers first name and last name using customer column from Customer table
6. Create a new table “category wise sales’ with Product Category(Grouped) and sum of Total sales. Identify the two tables that are to be joined and the common column. Which Join kind is most suitable in this scenario?
7. Round the Sales Amount to appropriate decimals
8. Check the query performance using power query diagnostics and comment on it [Hint: Power Query Editor -> Tools -> Start Diagnostics-Stop Diagnostics].
9. Now, create a hierarchy in the dashboard by using the given dataset. (Country-region > State-Province > City)

[AdventureWorks Sales.xlsx](https://s3.amazonaws.com/grey_campus/production/system/OdinSchool/Data+Science+Backup/AdventureWorks+Sales.xlsx) (file Provided Alongwith this document)

Tags (Note: For internal purpose): PowerBI components - Desktop, Data Preparation and Transformation using Power Query(Extract, merge columns, split columns), Data modeling, Power BI Visuals (Basics)  
  
  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Solutions:-***

1. There are multiple ways to load the data into the table but what I find best suitable one is to just click on excel workbook under the home ribbon and a pop-up window appears where we will be able to navigate through our computer and load the respective dataset, select the .xlsx file and load all the tables.
2. To perform check for columns of each table, we can turn on column quality, column distribution under View ribbon in Power Query, and column profile. Follow the given hints - Power Query editor -> View -> Check Column profile, column quality, column distribution.  
     
   
3. Go to the Date table and click on Date column, select it then click on Date under add columns option at the top, a dropdown menu will appear, select the corresponding options and make the necessary renaming changes.  
     
   
4. Select category then sub-category, under product and go to add columns and then merge them [know that hitting merge under transformation tab will merge the existing columns together without making sure to keep the originals together – we want it in a new column without doing any transformation on the existing columns], then select a delimiter, on this case, “-” (custom) and press enter.  
     
   
5. Click on the customer name column in customer table, then under transform, select split column, and it is by default by delimiter, which is space. Then rename the new columns as first and last name as per requirement.  
     
   
6. We need to merge the two respected columns together via Product key to get the respected result. As the total sales amount is in the sales table and the category is in the Product table. So what we do is join these two tables together, so in order to do so, Select the product table then under home tab, on the extreme right hand side, select merge query (select merge as new query) [also if you want to add columns, we go for merge and if add rows, we go for append], now a dialogue box appears and we need to select the matching column, over here, as mentioned above, Product key was the matching one, so select it on both the tables in the dialogue box “matching column”.   
     
   Left outer join (by default – all from first and then matching from second) had few of the rows missing, so it was wise to use Right Outer Join (all the values from the sales table and some matching from the product table – so all the rows would be matching).  
     
     
     
   Now, rename the newly made table as per requirement and then on the extreme right of this table there is sales from where we can select which rows to show, so show the amount column from the selection now the question asks us to keep only the columns of the sales amount and product category together, so select those two and right click them and select to remove all other columns.  
     
     
     
   Now, use Group by function after selecting category and select new column name as Total Sales, with Operation as Sum and Column would be Sales.Sales Amount hit okay and we’re done. Now, the category has been grouped with the Total Sale.  
     
   
7. Now select the ‘Total Sale’ column in the previous Table and go to Transform tab and select Rounding option, underneath it, there is a Round Up Function. Go ahead and select it to round it up.  
     
   
8. To do diagnostic, just select Tools and hit ‘Start Diagnostic’, to see how much time it took to perform each step. Click on Applied steps and just go to different tables etc. It basically records all of the time it took to perform those steps. After doing it all, hit Stop Diagnostic to get a Grouped Table of results on how the Power BI performed etc.  
   
9. Now close and apply from the power query {few errors might be there but it’s a data fault skip the error} and come back to report view. Now, go to customer table, click on country column and right click on it and select create hierarchy, and state and city must be added into the same hierarchy [Add to hierarchy >> click on the hierarchy initially created]. Now take a clustered column chart and add Sales Amount (y-axis) by Country-Region (x-axis), we can remove ‘non applicable’ for sales in the filters section, to remove incorrect data and then we can drill down to find more states and cities in the same visualisation via the help of hierarchy, which we created.   
     
    (Country - Region)  
    (State - Province)  
     
    (City)  
     
     
     
     
   Project By – Bidarth Kr. Singh

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*