|  |  |  |  |
| --- | --- | --- | --- |
| A picture containing drawing, stop, room  Description automatically generated | Business Intelligence  Practical #2 | | |
|  |  |  |  |
| **Name** | Farman Farooq Kazi | **Roll Number** | 21302B0049 |
| **Class** | TY BSc. IT | **Division** | A |
| **Subject/Course:** | Business intelligence | | |
| **Topic** | Perform Data Wrangling (ETL) | | |
|  |  |  |  |
| **What is ETL?** | | | |
| * ETL stands for **“Extract, Transform and Load”**. * ETL is also known as **Data Wrangling**. * Its’s a process that combines data from multiple sources into a centralized repository called a **Data Warehouse**. * ETL uses business rules to **clean** and **organized** raw data and prepare it for **storage**, **data analytics** and **Machine Learning (ML)**. * It's becoming increasingly necessary due to the rapid expansion of **data** and **data sources**. | | | |
|  | | | |
| **What is Power Query?** | | | |
|  | | | |
| * Power Query is data transformation and data preparation Engine. * Power Query comes with a **Graphical Interface** for getting data from source and a Power Query Editor for applying transformations. | | | |
|  | | | |
| **What are the components of Power Query Editor?** | | | |
|  | | | |
| 1. **Ribbon**: the ribbon navigation experience, which provides multiple tabs to add transforms, select options for your query, and access different ribbon buttons to complete various tasks. 2. **Queries pane**: a view of all your available queries. 3. **Current view**: your main working view, that by default, displays a preview of the data for your query. You can also enable the diagram view along with the data preview view. You can also switch between the schema view and the data preview view while maintaining the diagram view. 4. **Query settings**: a view of the currently selected query with relevant information, such as query name, query steps, and various indicators. 5. **Status bar**: a bar displaying relevant important information about your query, such as execution time, total columns and rows, and processing status. This bar also contains buttons to change your current view. | | | |
| **Write the steps to perform ETL in Power BI?** | | | |
| 1. Select the arrow next to **Get data** in the Power BI Desktop ribbon's **Home** tab, and then select **Excel workbook** from the **Common data sources** menu.      1. In the **Open** dialog box, navigate to and select the **Products.xlsx** file, and then select **Open**. In the **Navigator**, select the **Products** table and then select **Transform Data**.     *A table preview opens in the Power Query Editor, where you can apply ETL.* | | | |
| 1. In Power Query Editor, select **New Source** and then, from the **Common data sources** menu, select **OData feed**.      1. In the **OData feed** dialog box, paste the Northwind OData feed URL and Select **OK**.      1. In **Navigator**, check the **Orders** table, and then click **Transform the data** to load the data into Power Query Editor.        1. Scroll to the right in the **Orders** table until you see the **Order\_Details** column. Click on tow arrows **(****)** which indicates that it contains references to another table and not data.      1. Select “**Select All Columns**” to clear all columns and select **ProductID**, **UnitPrice**, and **Quantity**, and then select **OK**.     *After you expand the* ***Order\_Details*** *table, three new nested table columns replace the* ***Order\_Details*** *column. There are new rows in the table for each order's added data.*    *Creating a custom calculated column*   * + In the Power Query Editor's **Add Column** ribbon tab, select **Custom Column**.      * + In the **Custom Column** dialog box, type **LineTotal** in the **New column name** field. In the **Custom column formula** field after the **=**, enter **[Order\_Details.UnitPrice]** \* **[Order\_Details.Quantity]**. You can also select the field names from the **Available columns** scroll box and select **<< Insert**, instead of typing them and select **OK.** (*The new LineTotal field appears as the last column in the Orders table.*)      * + New **LineTotal** column has an **Any** data type. To assign a data type, right-click the **LineTotal** column header, select **Change Type** from the dropdown menu, and then select **Fixed decimal number**.    *Clean up the orders columns*  * Select the non-listed columns and use **Remove Other Columns**, right-click on one of them, and select **Remove Columns**.      * Double-click or tap and hold each column header, or right-click the column header, and select **Rename** from the dropdown menu. Delete the **Order\_Details.** prefix from each name.      * To make the **LineTotal** column easier to access, drag and drop it to the right, just to the left of the **Quantity** column.    *Review the query steps*   Ready to import it into Power BI Desktop **Report** view, select **Close & Apply** > **Close & Apply** in the **Home** ribbon tab's **Close** group. | | | |