Lab Work No. 6: Kaggle Dataset Exploration, Data Cleaning, and Pivot Table Creation

Objectives:

1. Search and download a dataset from Kaggle.
2. Import and clean the dataset in Excel (remove duplicates, handle missing values, etc.).
3. Create **at least 5 pivot tables** to analyze the data.

Part 1: Searching and Downloading a Dataset from Kaggle

1. Open your browser and go to https://www.kaggle.com.
2. Log in to your Kaggle account (create one if you don't have it).
3. In the **search bar**, type a keyword of interest (examples: "sales dataset", "movies dataset", "student performance dataset", "COVID-19 dataset").
4. From the search results, click on a dataset with:
    - **At least 500 rows** (so pivot tables are meaningful).
    - Data in **CSV format** (most convenient for Excel).
5. Click **Download** and save the dataset to your computer.

Part 2: Importing the Dataset into Excel

1. Open **Microsoft Excel**.
2. Go to **File > Open > Browse** and locate your downloaded CSV file.
3. Excel will automatically format the dataset into columns.
4. Save the file as an **Excel Workbook (.xlsx)** for further work.

Part 3: Data Cleaning

Before creating pivot tables, the dataset must be clean. Follow these steps:

1. **Check column headers:**
    - Ensure column names are clear (rename if necessary).

- Example: Change Stud_ID → Student ID.

2. **Remove duplicates:**
    - Select the dataset → Go to **Data > Remove Duplicates**.
    - Confirm that no repeated rows remain.
3. **Handle missing values:**
    - Use **Filter** (Data → Filter) to check if any blank cells exist.
    - Fill them with reasonable values (e.g., average score for missing grades) or replace with N/A.
4. **Check data types:**
    - Ensure numbers are formatted as **Number** or **Currency**.
    - Ensure dates are formatted as **Date**.
5. **Save the cleaned dataset.**

Part 4: Creating Pivot Tables

Now you'll create at least **5 pivot tables**. Choose analyses that fit your dataset. Here are examples:

Pivot Table 1: Frequency Count of Categories

- Insert → PivotTable.
- Rows: Category (e.g., Movie Genre, Product Type, Student Gender).
- Values: Category → Count.
- This shows how many entries exist for each category.

Pivot Table 2: Sum of Numerical Values per Category

- Rows: Category (e.g., Product, Subject).
- Values: Sales, Scores, or Ratings → Sum.
- Shows total values per category.

Pivot Table 3: Average per Group

- Rows: Region / Class / Country.

- Values: Sales, Scores, or Ratings → Average.
- Shows performance per region or class.

Pivot Table 4: Cross Analysis with Two Dimensions

- Rows: Category (e.g., Product, Movie Genre).
- Columns: Year / Month.
- Values: Sales, Revenue, or Scores → Sum or Count.
- This shows trends over time.

Pivot Table 5: Top 10 Analysis

- Rows: Customer, Product, or Student.
- Values: Total Sales, Total Score, etc. → Sum.
- Sort in descending order and filter to **Top 10**.

Part 5: Formatting and Interpretation

1. Apply **Pivot Table Design** (Report Layout → Show in Tabular Form).
2. Add **Number Formatting** (currency, percentage, etc.).
3. Insert **Charts** (Column, Bar, or Line, etc.) for better visualization.
4. Write a **1–2 sentence interpretation** under each pivot table. Example:
    - "The highest sales came from Product A with $50,000 revenue."
    - "Female students scored higher on average in Science than male students."

Deliverables (to be submitted):

1. Cleaned dataset (Excel file) : **LW6_LastName.xlsx.**
2. At least **5 pivot tables** with corresponding charts.
3. Short interpretations of each pivot table result.