

Assignment: Data Analytics 2452

Objective:

To analyze a dataset using Excel and create visual insights through charts, presented in a formal report.

Instructions:

1. Download the Dataset:

- Visit Kaggle
[Find Open Datasets and Machine Learning Projects | Kaggle](#)
Datasets and choose a dataset relevant to your analysis. Ensure the dataset is comprehensive enough to generate meaningful insights.

2. Data Preparation:

- Import the Dataset into Excel:
 - Download the dataset from Kaggle.
 - Open Excel and go to File > Open to import the downloaded dataset (usually in CSV format).
- Clean the Data:
 - Handling Missing Values: Identify any missing values in the dataset. Use the Filter option to quickly find empty cells. Decide whether to fill in missing values with a placeholder (e.g., 0 or "Unknown") or remove rows/columns with missing data.
 - Removing Duplicates: Go to the Data tab and use the Remove Duplicates feature to eliminate any duplicate rows.
 - Ensuring Data Consistency: Check for any inconsistencies in data formats (e.g., date formats, numerical values) and standardize them. Use the Text to Columns feature if needed to split data into separate columns.

3. Data Analysis:

Calculate Basic Statistics:

AVERAGE: This function calculates the mean of a range of values.

Example: =AVERAGE(B2:B100) calculates the average of values in column B from row 2 to 100.

Steps:

Click on an empty cell where you want the result to appear.

Type =AVERAGE(.

Select the range of cells you want to average (e.g., B2 to B100).

Close the parenthesis) and press Enter.

MEDIAN: This function finds the middle value in a range of numbers.

Example: =MEDIAN(B2:B100) calculates the median of values in column B from row 2 to 100.

Steps:

Click on an empty cell.

Type =MEDIAN(.

Select the range of cells (e.g., B2 to B100).

Close the parenthesis) and press Enter.

MODE: This function returns the most frequently occurring value in a range.

Example: =MODE(B2:B100) finds the mode of values in column B from row 2 to 100.

Steps:

Click on an empty cell.

Type =MODE(.

Select the range of cells (e.g., B2 to B100).

Close the parenthesis) and press Enter.

STDEV: This function calculates the standard deviation, which measures the amount of variation or dispersion in a set of values.

Example: =STDEV(B2:B100) calculates the standard deviation of values in column B from row 2 to 100.

Steps:

Click on an empty cell.

Type =STDEV(.

Select the range of cells (e.g., B2 to B100).

Close the parenthesis) and press Enter.

Identify Key Variables and Relationships

Using PivotTables:

PivotTables are powerful tools for summarizing, analyzing, and exploring data.

Creating a PivotTable:

Steps:

Select the range of data you want to analyze.

Go to the Insert tab and click on PivotTable.

In the Create PivotTable dialog box, ensure the correct data range is selected.

Choose where you want the PivotTable to be placed (e.g., a new worksheet or an existing one).

Click OK.

Configuring the PivotTable:

The PivotTable Field List will appear on the right side of the Excel window.

Drag and drop fields into the following areas:

Rows: Place the field you want to categorize your data by (e.g., Region).

Columns: Place the field you want to create columns for (e.g., Product).

Values: Place the field you want to summarize (e.g., Sales). By default, Excel will sum the values, but you can change this by clicking on the field in the Values area and selecting Value Field Settings (e.g., to calculate the average, count, etc.).

Filters: Place any fields you want to use to filter the data (e.g., Year).

Example:

Suppose you have a dataset with columns Region, Product, and Sales.

You want to analyze the total sales by region and product.

Drag Region to the Rows area, Product to the Columns area, and Sales to the Values area.

- The PivotTable will display a summary of sales for each product in each region.

4. Visualization:

- Create Charts:
 - Bar Chart: Select the data you want to visualize, go to Insert > Bar Chart, and choose the desired bar chart type.
 - Line Chart: Select the data, go to Insert > Line Chart, and choose the desired line chart type.
 - Pie Chart: Select the data, go to Insert > Pie Chart, and choose the desired pie chart type.
- Customize Charts:
 - Use the Chart Tools to add titles, labels, and legends. Right-click on chart elements to format them (e.g., change colors, fonts).
 - Example: To add a title, click on the chart, go to Chart Tools > Layout > Chart Title, and enter your title.

5. Formal Report:

- Write a formal report that includes the following sections:
 - Title Page: Title of the report, your name, date, and course details.
 - Table of Contents: List of sections and page numbers.
 - Introduction: Brief overview of the dataset and the objectives of the analysis.
 - Methodology: Explanation of the data preparation and analysis process.
 - Results: Presentation of the charts and a detailed explanation of each insight.
 - Discussion: Interpretation of the insights and what understanding can be obtained from them.
 - Conclusion: Summary of findings and any recommendations based on the analysis.
 - References: List of any sources or references used.

Submission:

- Include the formal report as a separate document.

Evaluation Criteria:

- Accuracy and completeness of data cleaning and preparation.
- Appropriateness and clarity of the charts.
- Depth of insights and quality of the written report.

- Adherence to the formal report format.
- Overall presentation and organization of the assignment.