**Name: Zeeshan Ahmad**

**Email:** [**Zeeshan.ahmad124586@gmail.com**](mailto:Zeeshan.ahmad124586@gmail.com)

**Contact: 03024920639**

**Submitted to: Digital Empowerment Pakistan**

**Task 1:**

**Analysis on Retail Store Dataset**

**Data Cleaning and Preprocessing:**

* Data Wrangling

1. Gathering Data
2. Assessing Data
3. Cleaning Data

* EDA
* Draw Conclusion

1. **Data Wrangling**
2. Gathering Data

Collect relevant data from various sources such as

* CSV files
* APIs
* Web Scrapping
* Databases

1. **Assessing Data**

Evaluating the quality, accuracy, and completeness of the data concisely and effectively.

* 1. Manually

Manually scroll through the dataset in Excel or a spreadsheet, observe the data, identify any issues, and note them.

* 1. Programmatically

Analyze by these Pandas methods

- head and tail

- sample

- info

- is null

- duplicated

- describe

**Note - Assessing Data is an Iterative Process**

**Write issues with your data.**

1. Then Clean and preprocess.

* Fill Missing Data (e.g. mean).
* Remove duplicates data.
* Correct incorrect datatypes.
* **Feature engineering:**
  1. Removing Outliers.(box plot)
  2. Merging Data frames.
  3. Adding new columns

**EDA:**

Finding Correlations.

Doing univariate and multivariate analysis.

Plotting Graphs.

Drawing conclusions.