Assignment 1: Data Manipulation and Analysis using Pandas

Name: Ashutosh Shinde

Roll no: 282006

Batch: B1

Problem Statement

a) Read data from different formats (like CSV, Excel).

b) Find the shape of the data.

c) Find missing values.

d) Find the data type of each column.

e) Find zeros in the data.

f) Indexing and selecting data, sorting data.

g) Describe attributes of data, checking data types of each column.

h) Count unique values, format of each column, convert variable data types (e.g., from long to short, vice versa).

Objective

To introduce the Pandas library and its basic functions for data manipulation.

To familiarize with data cleaning and preprocessing techniques.

To enhance skills in handling data in various formats for analysis.

Methodology

1. Data Collection and Exploration

Dataset Used: Customers dataset (CSV format).

Tasks Performed:

Load the dataset into a Pandas DataFrame.

Explore the structure of the data (shape, columns, data types).

2. Data Preprocessing

Handle missing values.

Clean the data (remove duplicates, correct errors).

3. Data Manipulation

Sort data.

Select specific rows/columns.

Convert data types.

Conclusion

This assignment demonstrated the use of Pandas for data manipulation tasks such as reading data, handling missing values, sorting, and converting data types. These foundational skills are essential for efficient data analysis and preprocessing in real-world projects.