TASK-2

1. Learn how to read CSV files and manipulate

data frames using Pandas.

pip install pandas

import pandas as pd

# Reading a CSV file

df = pd.read\_csv('path\_to\_your\_file.csv')

# Display the first few rows of the dataframe

print(df.head())

# Display the dataframe info

print(df.info())

# Display summary statistics

print(df.describe())

**2. Performing Simple Data Cleaning Tasks**

**Handling Missing Values**

# Check for missing values

print(df.isnull().sum())

# Drop rows with any missing values

df\_dropped = df.dropna()

# Fill missing values with a specific value

df\_filled = df.fillna(0)

# Fill missing values with the mean of the column

df\_filled\_mean = df.fillna(df.mean())

# Check for duplicate rows

print(df.duplicated().sum())

# Drop duplicate rows

df\_no\_duplicates = df.drop\_duplicates()

### 3. Practicing Basic Data Manipulation Operations

#### Filtering Data

# Filter rows based on a condition

filtered\_df = df[df['column\_name'] > some\_value]

# Filter rows with multiple conditions

filtered\_df = df[(df['column\_name1'] > some\_value) & (df['column\_name2'] < another\_value)]

# Sort by a single column

sorted\_df = df.sort\_values(by='column\_name')

# Sort by multiple columns

sorted\_df = df.sort\_values(by=['column\_name1', 'column\_name2'], ascending=[True, False])

# Group by a single column and calculate the mean of each group

grouped\_df = df.groupby('column\_name').mean()

# Group by multiple columns and calculate the sum of each group

grouped\_df = df.groupby(['column\_name1', 'column\_name2']).sum()

import pandas as pd

# Reading the CSV file

df = pd.read\_csv('path\_to\_your\_file.csv')

# Viewing the first few rows and summary information

print(df.head())

print(df.info())

# Handling missing values

df = df.fillna(df.mean()) # Fill missing values with the mean

# Removing duplicates

df = df.drop\_duplicates()

# Filtering data

filtered\_df = df[df['some\_column'] > some\_value]

# Sorting data

sorted\_df = df.sort\_values(by='some\_column', ascending=False)

# Grouping data

grouped\_df = df.groupby('another\_column').sum()

# Display the results

print(filtered\_df.head())

print(sorted\_df.head())

print(grouped\_df.head())