**PRACTICAL NO.7**

**Write a Python program to read data from a CSV file, perform simple data analysis, and**

**generate basic insights. (Use Pandas is a Python library).**

**Code:**

import pandas as pd

file\_path = "sales\_data.csv"

data = pd.read\_csv(file\_path)

print("First 5 rows of the data:")

print(data.head())

# Summary statistics of numeric columns

print("\nSummary Statistics:")

print(data.describe())

# Check for missing values

print("\nMissing values in each column:")

print(data.isnull().sum())

# Data Types of columns

print("\nData Types of Columns:")

print(data.dtypes)

# Correlation between numeric columns

#print("\nCorrelation between numeric columns:")

#print(data.corr())

# Correlation between numeric columns

numeric\_data = data.select\_dtypes(include=['number']) # Select only numeric columns

print("\nCorrelation between numeric columns:")

print(numeric\_data.corr())

print("\n fill empty quantity")

x = data["Quantity"].mean()

#data["Quantity"].fillna(x, inplace = True)

data.fillna({"Quantity": x}, inplace=True)

print(data.head(15))

if 'Sales' in data.columns:

max\_sales = data['Sales'].max()

print(f"\nMaximum Sales value: {max\_sales}")

#the most frequent category in a categorical column (e.g., 'Product' column)

if 'Product' in data.columns:

most\_frequent\_product = data['Product'].mode()[0]

print(f"Most frequent product: {most\_frequent\_product}")

import matplotlib.pyplot as plt

import seaborn as sns

# Plotting a histogram of a numeric column like 'Sales'

if 'Sales' in data.columns:

plt.figure(figsize=(8, 6))

sns.histplot(data['Sales'], kde=True)

plt.title('Sales Distribution')

plt.xlabel('Sales')

plt.ylabel('Frequency')

plt.show()

**Output:**



