PROBLEM STATEMENT:

Title: PulseCheck Pharmaceuticals: A Prescription for Profit

Context: You’re a junior analyst at PulseCheck Pharmaceuticals, a mid-sized drug manufacturing firm based in Hyderabad. The firm has launched a new Ayurvedic syrup called KofSoothe, claiming faster relief in dry coughs. It has been released across 6 major Indian cities. The leadership wants to assess the **product’s monthly sales performance**, **doctor endorsement trends**, and **returns due to expiry over the last 12 months**.

Your Task: You are to help them build a performance dashboard using descriptive analytics. You are provided or will generate a dataset with the following columns:

• City – Name of the city (e.g., Mumbai, Chennai, Kolkata, Delhi, Bangalore, Ahmedabad)

• Month – January to December

• Units\_Sold – Monthly bottles sold

• MRP\_per\_Unit – Fixed price per bottle (e.g., ₹120)

• Units\_Expired – Bottles returned after expiry

• Doctor\_Endorsements – Count of doctors in each city who recommended the syrup that month

Metrics to Compute:

1. Total Revenue per city/month Formula: Total Revenue = Units\_Sold × MRP\_per\_Unit
2. Expiry Rate (%): Formula: Expiry Rate = (Units\_Expired / Units\_Sold) × 100
3. Mean, Median, Mode, and Std Dev of Units\_Sold and Doctor\_Endorsements
4. Top 3 months with highest revenue
5. Visualize monthly trend in sales using a line plot (optional)
6. City with highest average doctor endorsements

Product Planning according to Output:

1. Kolkata has highest revenue while, Mumbai has lowest total revenue. So Mumbai needs more sales and better sales strategy
2. Kolkata has the highest Average expiry rate % while Bangalore has the least. Kolkata needs better inventory management to avoid cost and surplus.
3. Unit sold has positive skewness meaning few months have unusually high sales with pulls up the average and Doctor endorsements has negative skewness which means there were few low endorsement months. Also, standard deviation is high for both Units sold and Doctor endorsements which means high variability. Some month had very high sales and some very low. Similarly, for doctor endorsements it was very high for some cities and very low for some.
4. March, September and May have the highest revenues.
5. Sales dropped very low for the month of April.
6. This means that, across all 12 months, Mumbai consistently had more doctors recommending the syrup than any other city.

CODE INPUT:

# Create Synthetic Dataset

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# Set seed for reproducibility

np.random.seed(42)

cities = ['Mumbai', 'Chennai', 'Kolkata', 'Delhi', 'Bangalore']

months = ['January', 'February', 'March', 'April', 'May', 'June',

          'July', 'August', 'September', 'October', 'November', 'December']

# Generate synthetic data

data = []

for city in cities:

    for month in months:

        units\_sold = np.random.randint(800, 1500)

        units\_expired = np.random.randint(0, 100)

        endorsements = np.random.randint(10, 50)

        data.append([city, month, units\_sold, 120, units\_expired, endorsements])

df = pd.DataFrame(data, columns=[

    'City', 'Month', 'Units\_Sold', 'MRP\_per\_Unit', 'Units\_Expired', 'Doctor\_Endorsements'

])

print(df.head(10).to\_string(index=False))

df.to\_excel("Sales\_Data\_Final.xlsx", index = False)

# Total Revenue per City/Month

df['Total\_Revenue'] = df['Units\_Sold'] \* df['MRP\_per\_Unit']

print(df[['City', 'Month', 'Units\_Sold', 'Total\_Revenue']].head())

grouped\_1= df.groupby('City')['Total\_Revenue'].sum().sort\_values(ascending=False).reset\_index()

print('\nTotal Revenue by City:\n', grouped\_1)

# Expiry Rate (%)

df['Expiry\_Rate (%)'] = (df['Units\_Expired'] / df['Units\_Sold']) \* 100

grouped\_2 = round(df.groupby('City')['Expiry\_Rate (%)'].mean().sort\_values(ascending=False).reset\_index(), 2)

print(df[['City', 'Month', 'Units\_Sold', 'Units\_Expired', 'Expiry\_Rate (%)']].head())

print('\nAverage Expiry Rate by City(%):\n', grouped\_2)

# Descriptive Statistics

stats = round(df[['Units\_Sold', 'Doctor\_Endorsements']].agg(['mean', 'median', 'std']),2)

mode\_units = df['Units\_Sold'].mode()[0]

mode\_endorsements = df['Doctor\_Endorsements'].mode()[0]

print("\nDescriptive Statistics:\n", stats)

print(f"Mode of Units Sold: {mode\_units}")

print(f"Mode of Doctor Endorsements: {mode\_endorsements}")

# Correlation between Units Sold and Doctor Endorsements

correlation = df['Units\_Sold'].corr(df['Doctor\_Endorsements'])

print(f"\nCorrelation between Units Sold and Doctor Endorsements: {round(correlation,2)}")

# Top 3 Months with Highest Revenue

monthly\_revenue = df.groupby('Month')['Total\_Revenue'].sum()

top\_months = monthly\_revenue.sort\_values(ascending=False).head(3)

print('\nTop 3 months with Highest Revenue:\n', top\_months)

# City with Highest Average Doctor Endorsements

city\_avg\_endorsements = round(df.groupby('City')['Doctor\_Endorsements'].mean().sort\_values(ascending=False), 2)

print('\nCity with Highest Average Doctor Endorsements:\n', city\_avg\_endorsements.head(1))

# Monthly Sales Trend (Line Plot)

monthly\_sales = df.groupby('Month')['Units\_Sold'].sum().reindex(months)

plt.figure(figsize=(10, 6))

monthly\_sales.plot(kind='line', marker='o', color='blue', linewidth = 2)

plt.title('Monthly Sales Trend (Units Sold)', fontsize=14)

plt.xlabel('Month', fontsize=10)

plt.ylabel('Total Units Sold', fontsize=10)

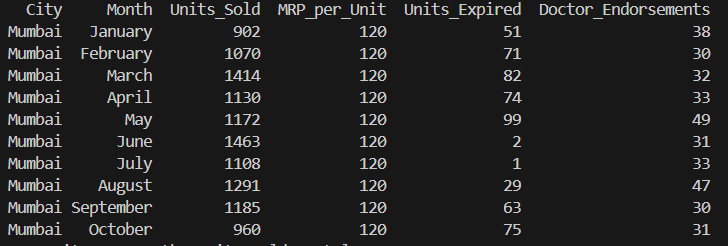
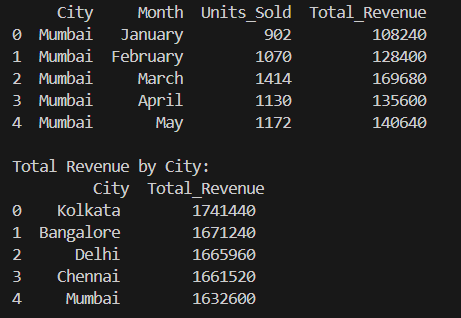
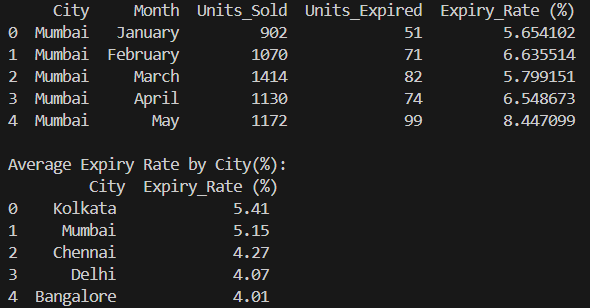
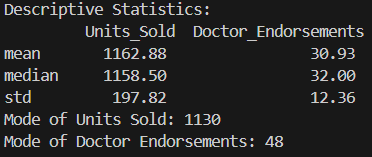
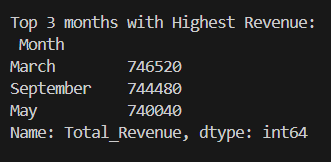
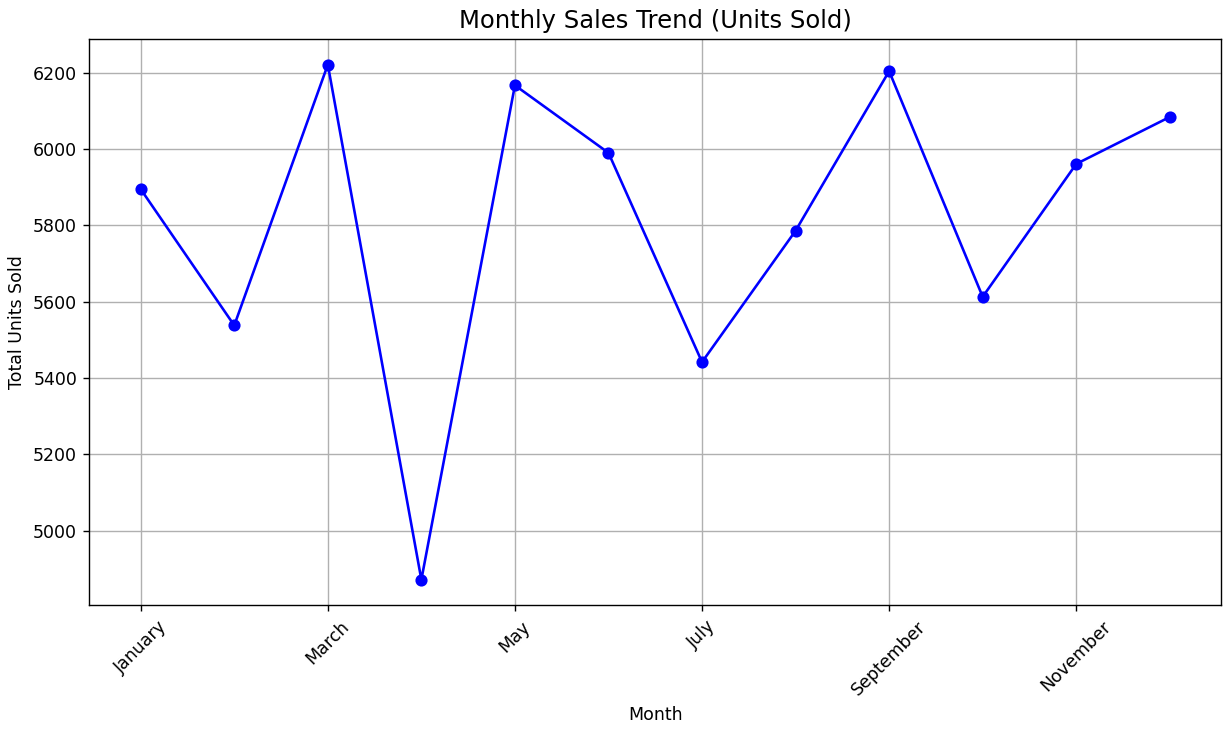
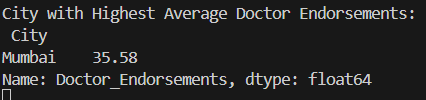
plt.grid(True)

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

CODE OUTPUT :

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 