Step 1: Load & Clean the Data

import pandas as pd

# Load the dataset

df = pd.read\_csv('/mnt/data/Unemployment in India (1).csv')

# Display basic info

print(df.info())

print(df.head())

Step 2: Rename Columns (if needed)

df.columns = [col.strip().replace(" ", "\_").lower() for col in df.columns]

print(df.columns)

Step 3: Check for Missing Values and Data Types

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

print(df.dtypes)

Step 4: Convert Date Column to Datetime Format

df['date'] = pd.to\_datetime(df['date'])

df = df.sort\_values('date')

Step 5: Unemployment Rate Trends Over Time

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(12, 6))

sns.lineplot(data=df, x='date', y='estimated\_unemployment\_rate\_(%)', hue='region')

plt.title('Unemployment Rate Trends in Indian States (2016-2020)')

plt.xlabel('Date')

plt.ylabel('Unemployment Rate (%)')

plt.legend(bbox\_to\_anchor=(1.05, 1), loc='upper left')

plt.tight\_layout()

plt.show()

Step 6: Covid-19 Impact Analysis (2020 Focus)

covid\_df = df[df['date'].dt.year == 2020]

plt.figure(figsize=(12, 6))

sns.lineplot(data=covid\_df, x='date', y='estimated\_unemployment\_rate\_(%)', hue='region')

plt.title('Unemployment Rates During COVID-19 Pandemic (2020)')

plt.xlabel('Date')

plt.ylabel('Unemployment Rate (%)')

plt.legend(bbox\_to\_anchor=(1.05, 1), loc='upper left')

plt.tight\_layout()

plt.show()

Step 7: Monthly Seasonal Patterns

df['month'] = df['date'].dt.month

monthly\_avg = df.groupby('month')['estimated\_unemployment\_rate\_(%)'].mean()

plt.figure(figsize=(10, 5))

monthly\_avg.plot(kind='bar', color='orange')

plt.title('Average Monthly Unemployment Rate in India')

plt.xlabel('Month')

plt.ylabel('Unemployment Rate (%)')

plt.xticks(ticks=range(0, 12), labels=['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',

'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'], rotation=45)

plt.grid(axis='y')

plt.tight\_layout()

plt.show()

Step 8: Heatmap by State and Month

state\_month = df.copy()

state\_month['month'] = state\_month['date'].dt.strftime('%b')

pivot\_data = state\_month.pivot\_table(index='region', columns='month',

values='estimated\_unemployment\_rate\_(%)', aggfunc='mean')

plt.figure(figsize=(14, 10))

sns.heatmap(pivot\_data, cmap='coolwarm', annot=True, fmt=".1f")

plt.title('Monthly Average Unemployment Rate by Region')

plt.show()

Step 9: Key Insights & Suggestions

- \*COVID-19 Impact:\* Unemployment rate peaked around April–June 2020, indicating job loss due to lockdowns.

- \*State-Wise Variation:\* States like Tripura and Haryana had significantly higher unemployment rates.

- \*Seasonal Trends:\* Unemployment is slightly higher mid-year, possibly due to monsoon and post-academic transitions.

- \*Policy Recommendation:\*

- Initiatives for employment in regions with persistent high rates.

- Seasonal employment schemes for periods of expected rise.

- Monitor pandemic recovery efforts in employment sectors.