Unemployment Analysis Using Python

Submitted by: Omeerpal singh

Table of Contents

1. Project Description

This project analyzes unemployment data on a macro level, identifying patterns, seasonality, and trends using data science and machine learning methods. The goal is to provide deep insights into unemployment dynamics while emphasizing regional disparities and COVID- 19's impact on unemployment.

2. Technologies Used

Python 3.x, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn, XGBoost, Statsmodels

3. Dataset Overview

The dataset contains global unemployment statistics. Key columns include:
- date: Date of observation
- country: Country name
- unemployment_rate: Unemployment rate (%)
- gdp_per_capita: GDP per capita
- inflation_rate: Annual inflation (%)

- population: Total population
- labor_force_participation: Labor force %

4.Source Code

1.Data Loading and Preprocessing

```
import pandas as pd import numpy as np
import matplotlib.pyplot as plt import
seaborn as sns import plotly.express as px
import warnings
warnings.filterwarnings('ignore')

# Load data df = pd.read_csv("global_unemployment.csv")

# Standardize column names df.columns = df.columns.str.strip().str.lower().str.replace("
", "_")

# Convert date df['date'] = pd.to_datetime(df['date'])

# Handle missing values df = df.dropna(subset=['unemployment_rate'])

# Feature extraction df['year'] = df['date'].dt.year df['month']
= df['date'].dt.month

# Show first few rows  df.head()
```

2.Advanced EDA (Country-wise and Global Trends)

1. Unemployment Rate by Country Over Time plt.figure(figsize=(15,8)) countries

= df['country'].value_counts().nlargest(5).index

for country in countries:
 temp_df = df[df['country'] == country] plt.plot(temp_df['date'],

temp_df['unemployment_rate'], label=country)

plt.title("Unemployment Rate Over Time (Top 5 Countries)") plt.xlabel("Year")

plt.ylabel("Unemployment Rate (%)") plt.legend() plt.grid(True)

plt.tight_layout() plt.show()

2.Heatmap of Unemployment by Month and Year

pivot_table = df.pivot_table(values='unemployment_rate', index='month', columns='year', aggfunc='mean')

plt.figure(figsize=(12, 6)) sns.heatmap(pivot_table, cmap='YlOrRd', annot=True,

fmt=".1f") plt.title("Heatmap of Unemployment Rate (Monthly Average Across

Years)") plt.xlabel("Year") plt.ylabel("Month") plt.tight_layout() plt.show()

3.Time Series Decomposition

import statsmodels.api as sm

Aggregating worldwide unemployment
global_monthly =

```
df.groupby('date')['unemployment_rate'].mean().resample('M').mean().interpolate()


# Decomposition decomp = sm.tsa.seasonal_decompose(global_monthly,
model='additive') fig = decomp.plot() fig.set_size_inches(14, 8) plt.suptitle("Time Series
Decomposition of Global Unemployment Rate", fontsize=16) plt.tight_layout()
plt.show()
```

4.Feature Engineering and Outlier Detection

```
# Feature engineering df['unemployment_change'] =
df.groupby('country')['unemployment_rate'].diff() df['gdp_change'] =
df.groupby('country')['gdp_per_capita'].pct_change()


# Outlier detection using IQR
Q1 = df['unemployment_rate'].quantile(0.25)
Q3 = df['unemployment_rate'].quantile(0.75) IQR =
Q3 - Q1


outliers = df[(df['unemployment_rate'] < (Q1 - 1.5 * IQR)) | (df['unemployment_rate'] > (Q3 +
1.5 * IQR))]


# Visualize plt.figure(figsize=(10,6)) sns.boxplot(x='country', y='unemployment_rate',
data=df[df['country'].isin(countries)]) plt.title("Unemployment Rate Outliers by Country")
plt.xticks(rotation=45) plt.tight_layout() plt.show()
```

5.Feature Importance

```
importances = pd.Series(model.feature_importances_,
index=X.columns).sort_values(ascending=True)
```

```
plt.figure(figsize=(8, 5)) importances.plot(kind='barh',
color='steelblue') plt.title("Feature Importance in Unemployment
Prediction") plt.xlabel("Importance Score") plt.tight_layout() plt.show()
```

5. Data Cleaning and Preprocessing

Data is loaded, cleaned, and transformed before analysis, including handling missing values and extracting yearly and monthly features.

6. Advanced Exploratory Data Analysis

Unemployment Rate by Country Over Time

- Trends for top-performing and most-affected countries are analyzed using time-series visualizations**.**

Regional Disparities

- Comparative unemployment trends between continents (North America, Asia, Europe, Africa, etc.).
- Analysis of economic interventions affecting unemployment levels in different regions.

7. COVID-Specific Unemployment Analysis

Pre-COVID, Peak-COVID, and Post-COVID Trends:

Unemployment rate shifts before, during, and after the pandemic.
The Impact of lockdowns and stimulus packages on employment recovery across different economies.

- **Industry-Specific Analysis**:
- Sectors most affected (hospitality, aviation, retail, etc.).
- The role of government aid in different regions for employment resurgence.

8. Time-Series Decomposition

Using statistical decomposition to separate unemployment trends into seasonality, cyclic patterns, and residual noise.

9. Feature Engineering and Outlier Detection

Feature Enhancements

- Unemployment Change: Rate difference across months.
- GDP Change: Percentage shift in GDP affecting unemployment patterns.

Outlier Detection

Detecting unusual spikes in unemployment using Interquartile Range (IQR) and visualizing outliers at the regional level.

10. Machine Learning Modeling & Performance Evaluation

- Various models (Random Forest, XGBoost) used for unemployment prediction.
- Performance comparison through accuracy metrics (RMSE, R^2) with detailed justifications for model choices.

11. Feature Interpretation

- Importance ranking of GDP, inflation, and labor force participation in unemployment prediction.
- Impact of pandemic response policies on employment levels.

12. Final Insights

- **High inflation and low labor participation correlate with higher unemployment**, especially in economically weaker regions.
- **Global unemployment exhibits seasonal and cyclical behavior**, with COVID significantly disrupting expected trends.
- **Predictive models effectively forecast unemployment rates** using macroeconomic indicators.
- **Regional disparities highlight differences in policy effectiveness**, with developed nations recovering faster post-pandemic.

THANKYOU

-