**Title: Unemployment Analysis in India**

**Introduction**  
This project aims to analyze the unemployment data in India, providing insights into the estimated unemployment rates, employment distributions, and labor participation across different states and regions. The analysis utilizes Python libraries such as pandas, NumPy, matplotlib, seaborn, and Plotly for data manipulation and visualization. Additionally, a Streamlit app is developed to allow dynamic exploration of the data.

**Data Overview**  
The dataset used in this project, unemployment.csv, contains information about unemployment rates in various states of India, including the estimated number of employed individuals, labor participation rate, region, and geographic coordinates (longitude and latitude). Key columns include:

* **States**: Name of the state
* **Date**: Date of the record
* **Frequency**: Frequency of data collection
* **Estimated Unemployment Rate**: Unemployment rate in percentage
* **Estimated Employed**: Number of employed individuals
* **Estimated Labour Participation Rate**: Percentage of the labor force participation
* **Region**: Geographical region
* **Longitude** and **Latitude**: Coordinates for mapping

**Python Scripts**

**Script 1: Data Analysis and Visualization**

1. **Import Libraries**  
   The script starts by importing necessary libraries such as pandas, NumPy, matplotlib, seaborn, and Plotly.

python

Copy code

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

1. **Load and Prepare Data**  
   The dataset unemployment.csv is read into a pandas DataFrame. Columns are renamed for easier handling, and any missing data is checked.

python

Copy code

data = pd.read\_csv("unemployment.csv")

print(data.head())

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

data.columns= ["States", "Date", "Frequency", "Estimated Unemployment Rate", "Estimated Employed", "Estimated Labour Participation Rate", "Region", "longitude", "latitude"]

1. **Data Visualization**  
   Several visualizations are created using seaborn and matplotlib:
   * A correlation heatmap to understand relationships between numerical variables.
   * Histograms to show the distribution of employed individuals and unemployment rates by region.

python

Copy code

plt.style.use('seaborn-whitegrid')

plt.figure(figsize=(12, 10))

sns.heatmap(data.corr())

plt.show()

...

1. **Interactive Plot with Plotly**  
   A sunburst chart is generated to visualize the unemployment rate by region and state using Plotly Express.

python

Copy code

unemploment = data[["States", "Region", "Estimated Unemployment Rate"]]

figure = px.sunburst(unemploment, path=["Region", "States"], values="Estimated Unemployment Rate", width=700, height=700, color\_continuous\_scale="RdY1Gn", title="Unemployment Rate in India")

figure.show()

**Script 2: Streamlit Application for Interactive Analysis**

1. **Introduction to Streamlit**  
   Streamlit is a Python library used to create web apps for data science and machine learning projects. This script sets up a Streamlit app for interactive analysis of the unemployment data.

python

Copy code

import streamlit as st

import pandas as pd

import plotly.express as px

1. **Load and Preprocess Data**  
   Similar to the first script, the data is loaded and pre-processed to ensure it's ready for analysis.

python

Copy code

data\_path = 'unemployment.csv'

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

...

1. **Streamlit App Layout and Widgets**  
   The app’s title and markdown description are set, and sidebar filters are added to allow users to select states, regions, and a date range.

python

Copy code

st.title('Unemployment Analysis in India')

st.markdown("Explore the unemployment data dynamically by filtering by state, region, and time period.")

...

1. **Data Filtering and Visualization**  
   The filtered data is displayed as a table. Several interactive visualizations are created using Plotly to show:
   * Unemployment Rate over Time
   * Employment Distribution by Region
   * Unemployment Rate by Region and State

python

Copy code

st.subheader('Unemployment Rate Over Time')

fig = px.line(

filtered\_data,

x='Date',

y='Estimated Unemployment Rate',

color='States',

title='Estimated Unemployment Rate Over Time'

)

st.plotly\_chart(fig)

...

1. **Show Raw Data Toggle**  
   A checkbox allows users to toggle the display of raw data.

python

Copy code

if st.checkbox("Show Raw Data"):

st.write(filtered\_data)

**Conclusion**  
The analysis provides a comprehensive view of the unemployment landscape across different states and regions in India. Using the Streamlit app, users can dynamically explore the data, visualize trends over time, and better understand regional differences in unemployment rates.

**Future Work**  
Future work could include integrating more datasets, such as economic indicators or demographic data, to provide a more holistic view of factors influencing unemployment. Additional visualizations and predictive modeling could also be added to enhance the app’s capabilities.