Project Report on

Pollution Data Analysis

Submitted by:

Imrana Khanam

Sourav

In partial fulfillment of completion of the course

Advanced Diploma in IT, Networking and Cloud Computing.

Under Guidance of:

Year 2022-2023

Abstract

Unemployment is a critical economic indicator that reflects the health of a nation's labor market. This study aims to conduct a thorough analysis of unemployment data, exploring trends, identifying causes, and examining the implications for individuals and the broader economy. The research utilizes a combination of quantitative methods, statistical tools, and qualitative insights to provide a comprehensive understanding of the multifaceted nature of unemployment.

Acknowledgement

At this juncture of our journey, we wish to express our heartfelt gratitude to all those who have contributed to the creation and success of **" Unemployment Analysis with Python ".** This project has been a labor of passion and dedication, and it would not have been possible without the unwavering support and guidance we have received.

First and foremost, we offer our thanks to the boundless creativity and inspiration that flows from the universe. We are grateful for the opportunity to embark on this venture.

We extend our sincerest appreciation to our mentors, **Mrs. Mala Mishra & Ms. Ankita Shukla**, whose wisdom and guidance have been instrumental in shaping the vision of **"Pollution Data Analysis".** Your support at every crucial turn has illuminated our path and fueled our determination to create a meaningful platform.

To our dedicated team of developers, designers, and content creators, we extend our deepest gratitude. Your tireless efforts, innovation, and creativity have breathed life into **" Unemployment Analysis with Python ".** It is your collective dedication that has made this project a reality.

Our appreciation also goes to our colleagues and friends who provided invaluable insights and feedback during the development process. Your input has been instrumental in refining our ideas and enhancing the user experience.

We acknowledge the contributions of the broader IT community, whose open-source ethos has been a wellspring of knowledge and inspiration. The collaborative spirit of this community has been a guiding light.

Last but not least, we owe a debt of gratitude to our families and friends who have stood by us throughout this journey. Your unwavering support, encouragement, and belief in our vision have been our constant motivation.

ADVANCE DIPLOMA IN IT NETWORKING & CLOUD COMPUTING

The Advanced Diploma in IT Networking and Cloud Computing program offered by NSTI (W) Noida in collaboration with Edunet Foundation is a comprehensive course designed to equip students with advanced skills in information technology and cloud computing. This program covers a wide range of topics, including Computer Networking, Database Management, Virtualization, Cloud Technologies, and Cybersecurity. Students will gain hands-on experience through practical labs, workshops, and real-world projects, enabling them to excel in the rapidly evolving IT industry. Upon completion of the program, Graduates will have a strong foundation in both IT Fundamentals and Cloud Computing, making them highly sought-after professionals in the field.

Project Requirements

Project Name	**Pollution Data Analysis**
Languages Used	**Python**
Editor	**Jupyter Notebook, Google Colab**
Python Library	**Numpy, Pandas, Matplotlib,pyplot.**

Team Composition and Workload Division

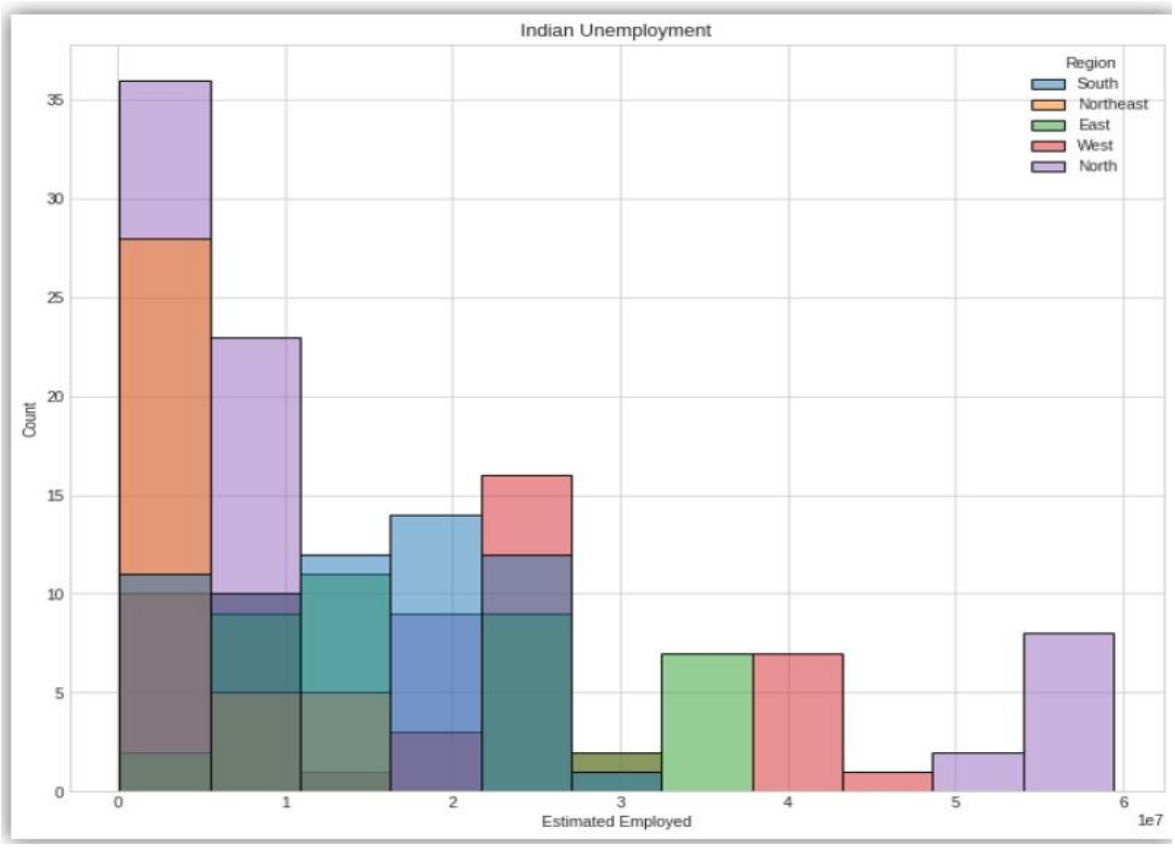
Imrana Khanam	Data Analysis, Synopsis
Sourav	Data Analysis, Synopsis

Tables of Content

1. Introduction to Problem

Unemployment is a complex economic phenomenon that requires a comprehensive analysis to understand its dynamics, identify influencing factors, and propose effective solutions. In this project, we aim to perform an in-depth analysis of unemployment data using Python programming language. The analysis will involve exploring historical trends, identifying key determinants of unemployment, and developing visualizations to communicate findings effectively.

2. E-R Model



3. Requirements

3.1 Technology Stack

Python: High-level programming language used for server-side scripting.

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text, providing an interactive and collaborative environment for data science and analysis.

3.2 Hardware

Laptop/ Computer

3.3 Software

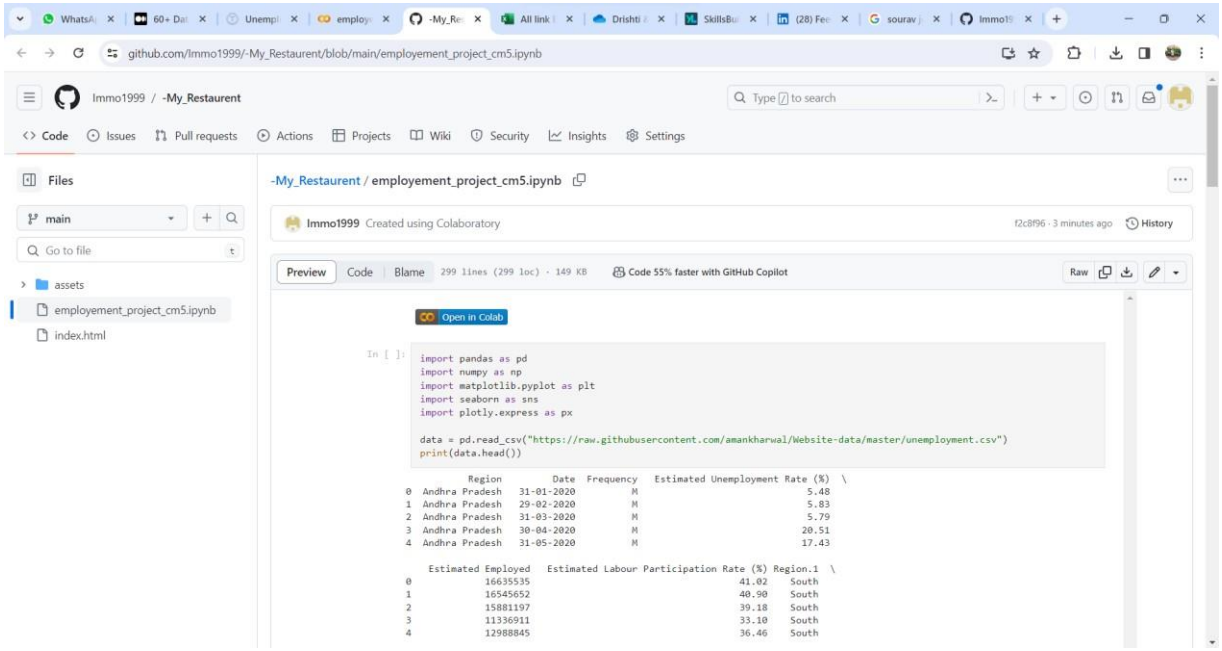
Operating System (OS)

Version Control System

Text Editors and Integrated Development Environments (IDEs)

3.4 Deployment Environment

Github



4. Overview

This project involves a comprehensive analysis of unemployment data using Python. The objectives include acquiring and preprocessing relevant data, conducting exploratory data analysis, and exploring temporal and demographic trends. Correlation analysis with other economic indicators and the implementation of machine learning models for predictive analysis are key components. Sentiment analysis of social media or news data will complement the understanding of public perception. The project aims to provide actionable insights and policy recommendations for addressing unemployment challenges. Deliverables include a Python script or Jupyter Notebook, visualizations, and a detailed report summarizing key findings and recommendations. Proficiency in Python programming, data manipulation, and visualization tools is required, along with a basic understanding of machine learning concepts and sentiment analysis using NLP libraries. The ultimate goal is to contribute to informed decisionmaking and effective policy formulation in tackling unemployment issues.

5. Project Module

1. Import the required libraries.
2. Load/ Read the Dataset
3. Prepare EDA
4. Do Visualizations
5. Unemployment Analysis with Python
6. Prepare Heatmap/ Confusion Matrix
7. Prepare Profile Report

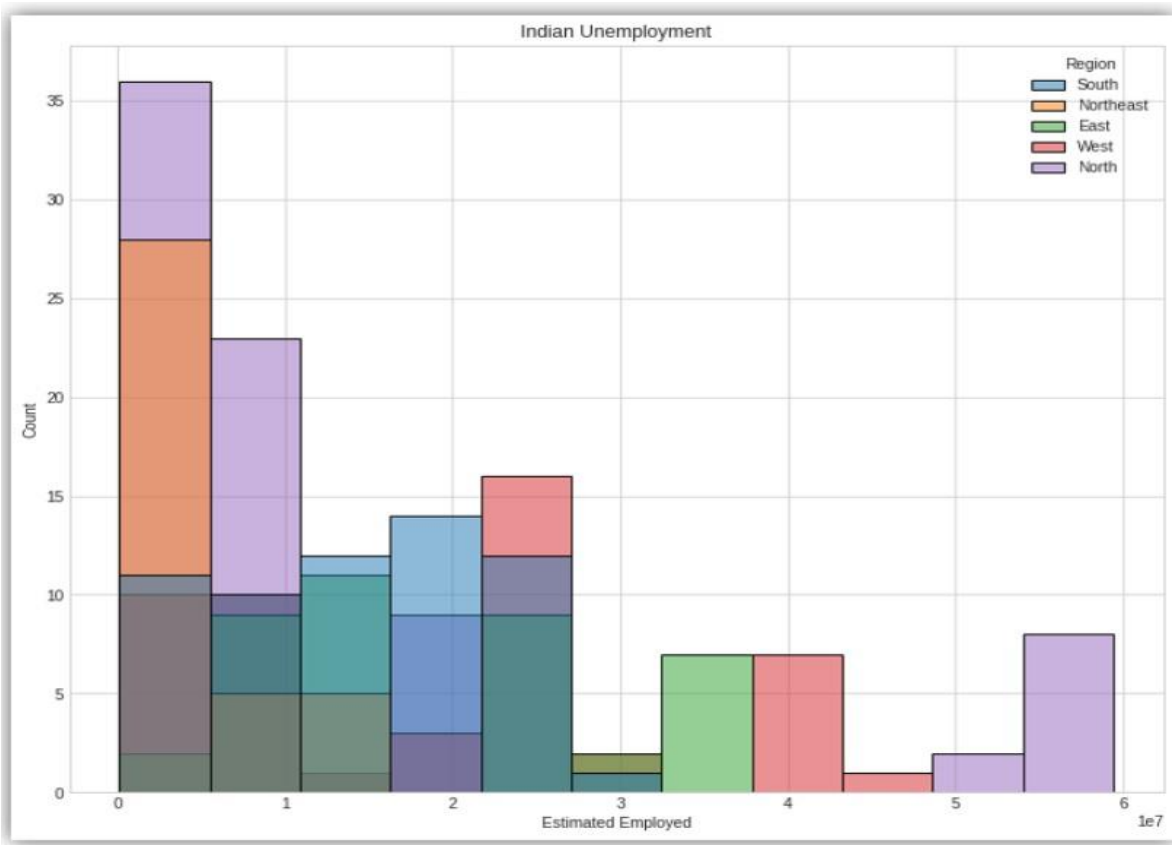
6 Sample Screenshots

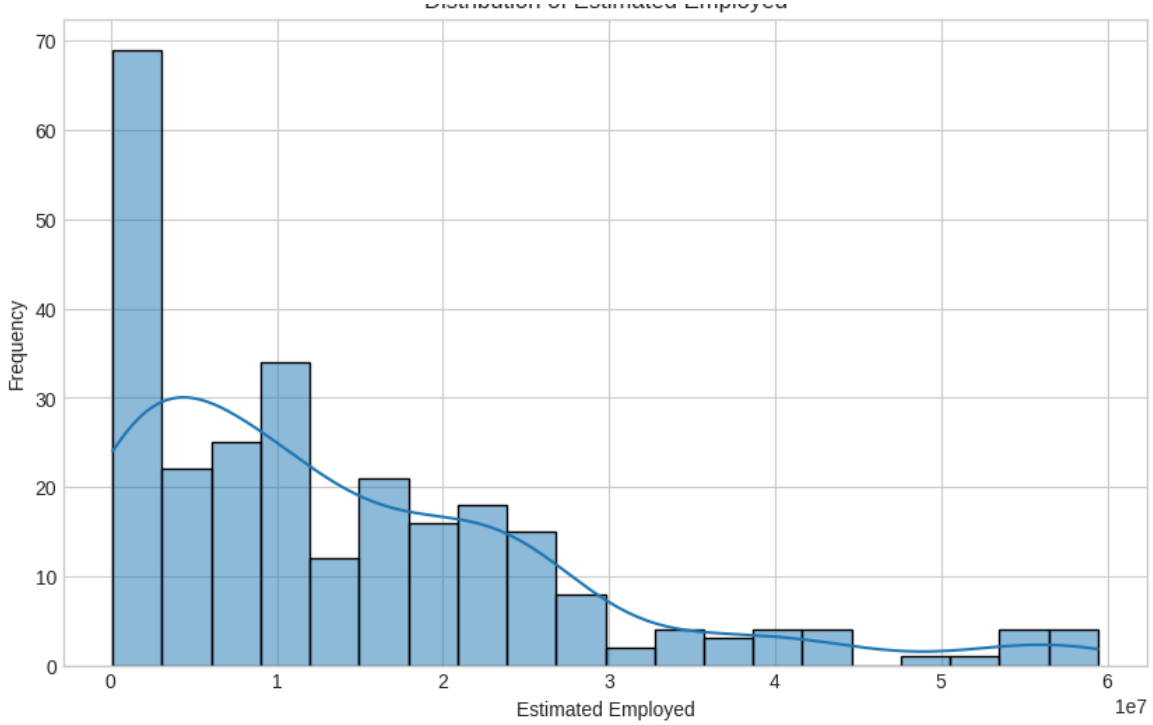
+ Code + Text

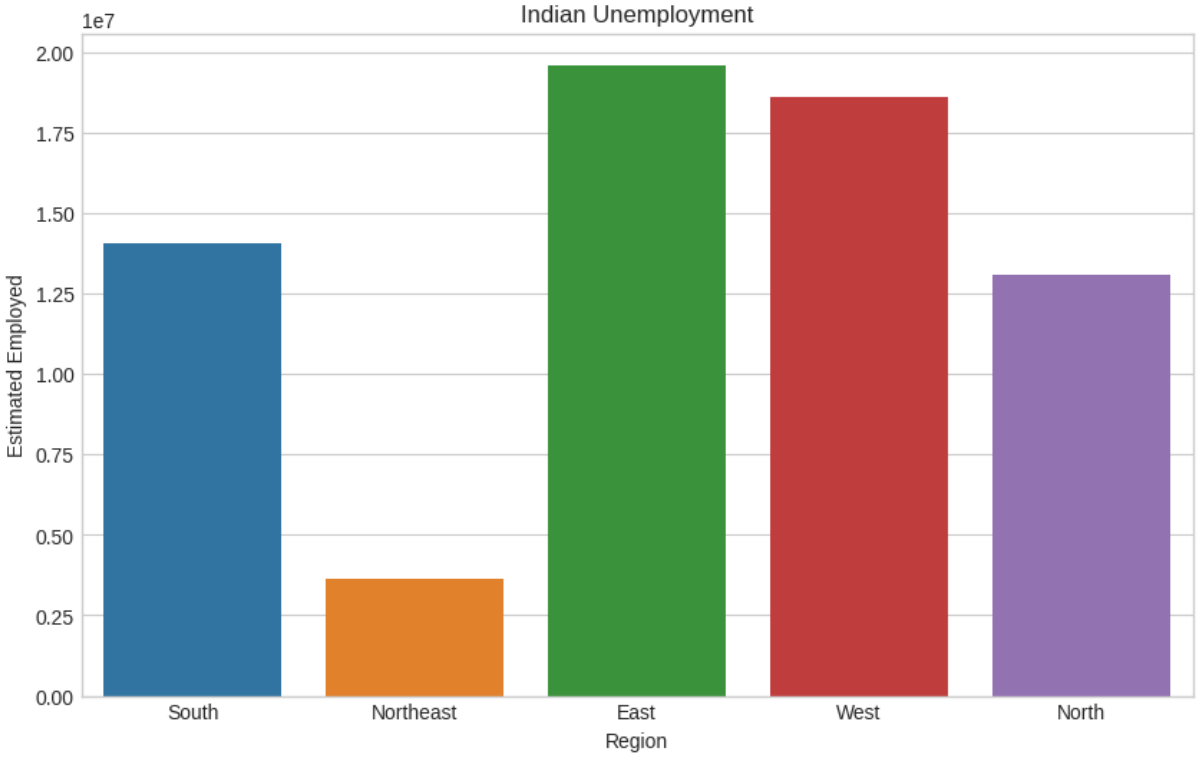
Unemployment Analysis with Python

Double-click (or enter) to edit

```
[ ] import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as sns
    import plotly.express as px

    data = pd.read_csv("https://raw.githubusercontent.com/amankharwal/Website-data/master/unemployment.csv")
    print(data.head())
```

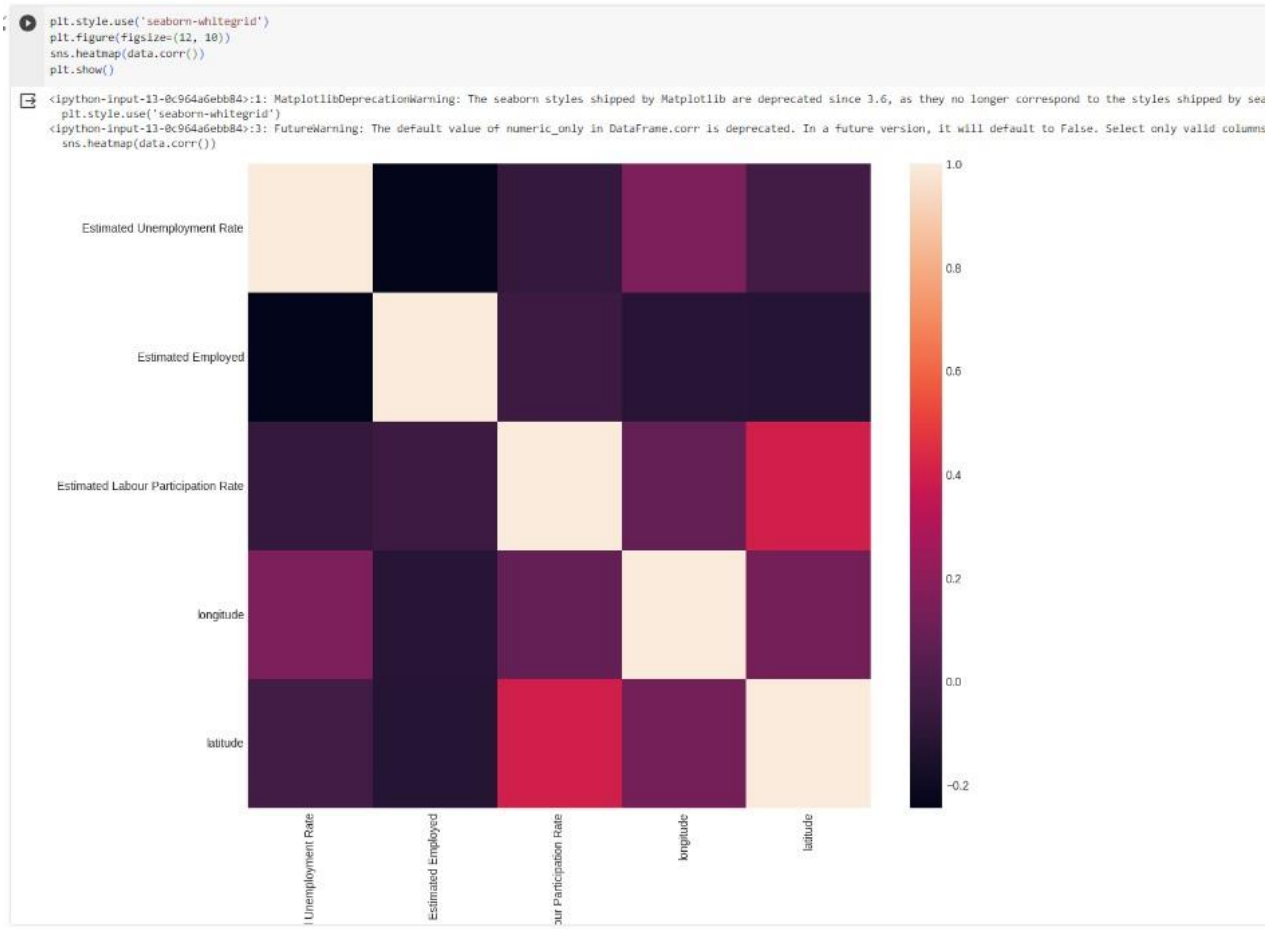



70
60
50
40
30
20
10
0
Frequency
0
1
2
3
4
5
6
1e7
Estimated Employed


Indian Unemployment
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Estimated Employed
South
Northeast
East
West
North
Region

7 Source Code

```
4    15.9129     79.74

[ ] print(data.isnull().sum())

Region                                0
Date                                  0
 Frequency                            0
 Estimated Unemployment Rate (%)      0
 Estimated Employed                   0
 Estimated Labour Participation Rate (%)  0
Region.1                              0
longitude                             0
latitude                              0
dtype: int64

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

[ ] plt.style.use('seaborn-whitegrid')
    plt.figure(figsize=(12, 10))
    sns.heatmap(data.corr())
    plt.show()

ipython-input-4-0c964a6ebb84>:1: MatplotlibDeprecationWarning: The seaborn styles shipped by Matplotlib are deprecated since 3.6, as they no longer correspond to the styles shipped by
  plt.style.use('seaborn-whitegrid')
ipython-input-4-0c964a6ebb84>:3: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid col
  sns.heatmap(data.corr())
```



8 Future Scope

The future scope of unemployment analysis with Python lies in leveraging advanced machine learning models, real-time data streams, and sentiment analysis for dynamic insights. Integration with emerging technologies like blockchain ensures data security, while collaboration with diverse data sources and interdisciplinary research enriches analysis perspectives. Cloudbased solutions offer scalability, and ethical considerations, along with fairness-aware machine learning, ensure responsible and equitable unemployment research. Augmented reality interfaces further enhance user engagement in exploring analysis outcomes.

9 Conclusion

This project aims to leverage Python's capabilities for data analysis to provide insights into unemployment patterns, correlations, and potential interventions, contributing to informed decision-making and policy development. This project aims to leverage Python as a powerful tool for unemployment analysis, providing stakeholders with actionable insights to address the challenges posed by unemployment and contribute to the formulation of effective policies.

10 References

https://www.kaggle.com/datasets

Thank you