



GROUP 8

SOLVING UNEMPLOYMENT

Low Hui Yi

22004790

Kong Yan Hao

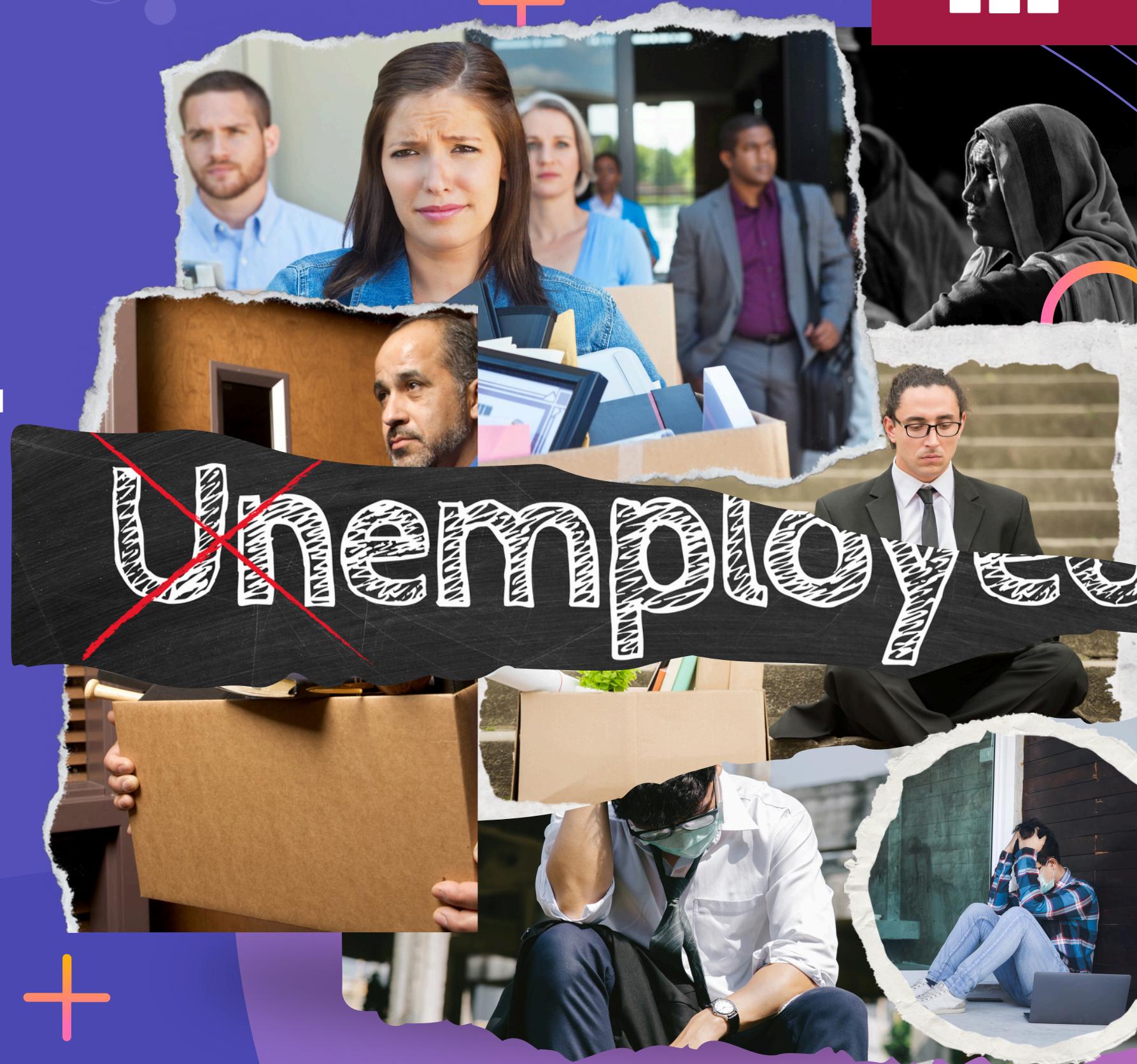
S2178056

Oon Yee Sem

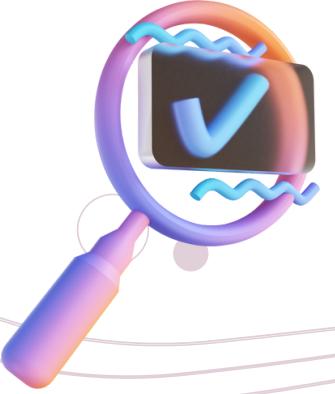
S2143263

Chai Jin Xin

22004735



PROBLEM STATEMENT



Unemployment is a persistent issue that affects economic stability and the well-being of individuals in Malaysia. This is often caused by economic fluctuations, sector-specific downturns, and a mismatch between labour market demands and workforce skills.

OBJECTIVES

- To identify the key economic indicators
- To forecast unemployment rate
- To classify the unemployment rate category

RESEARCH QUESTIONS

1. What are the key indicators of economic in Malaysia that affect decent work?
2. How do economic policies influence the quality of employment and working conditions in Malaysia?



ABOUT OUR DATASET

DATA SOURCE

- [Malaysia's National Statistics Organization OpenDOSM](#)
- Economy Data: Malaysian Economic Indicators, CPI by Division, PPI, IPI
- Decent Work Data: Monthly Principal Labour Force Statistics, Monthly Unemployment by Duration

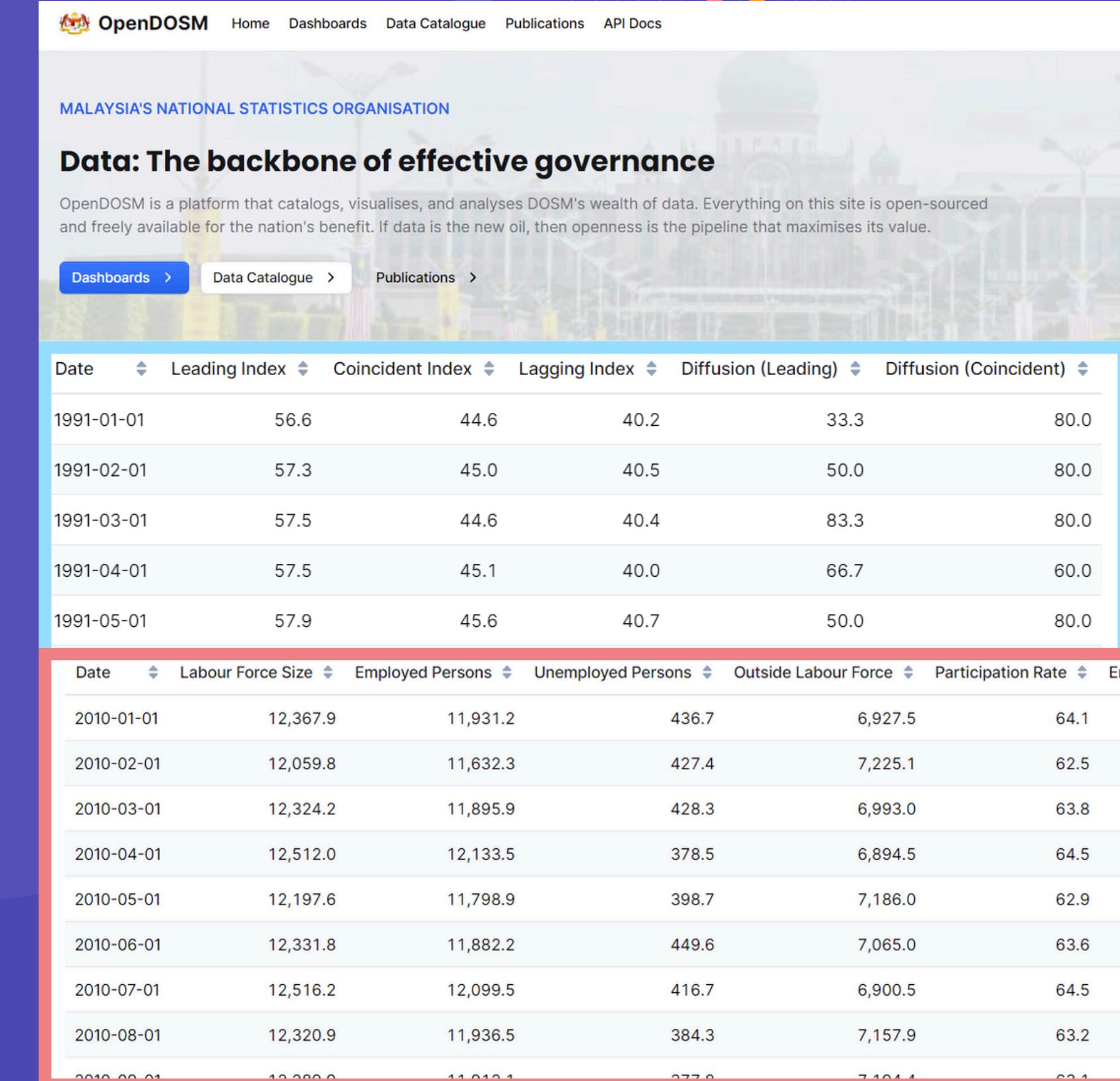
JUSTIFICATION

1. Comprehensive Data Coverage:

- A wide range of datasets on economic indicators, employment and more

2. Reliability:

- As an official government platform, the data is accurate, up-to-date, and trustworthy.



The screenshot shows the OpenDOSM website interface. At the top, there is a navigation bar with the logo of Malaysia's National Statistics Organisation, followed by links to Home, Dashboards, Data Catalogue, Publications, and API Docs. Below the navigation, a banner features the text "MALAYSIA'S NATIONAL STATISTICS ORGANISATION" and "Data: The backbone of effective governance". A sub-banner below it states: "OpenDOSM is a platform that catalogs, visualises, and analyses DOSM's wealth of data. Everything on this site is open-sourced and freely available for the nation's benefit. If data is the new oil, then openness is the pipeline that maximises its value." The main content area contains two data tables. The first table, titled "Leading Index", shows monthly data from 1991-01-01 to 1991-05-01. The second table, titled "Labour Force Size", shows monthly data from 2010-01-01 to 2010-09-01. Both tables include columns for Date, Leading Index, Coincident Index, Lagging Index, Diffusion (Leading), Diffusion (Coincident), Labour Force Size, Employed Persons, Unemployed Persons, Outside Labour Force, Participation Rate, and Employment Rate.

Date	Leading Index	Coincident Index	Lagging Index	Diffusion (Leading)	Diffusion (Coincident)
1991-01-01	56.6	44.6	40.2	33.3	80.0
1991-02-01	57.3	45.0	40.5	50.0	80.0
1991-03-01	57.5	44.6	40.4	83.3	80.0
1991-04-01	57.5	45.1	40.0	66.7	60.0
1991-05-01	57.9	45.6	40.7	50.0	80.0

Date	Labour Force Size	Employed Persons	Unemployed Persons	Outside Labour Force	Participation Rate	Employment Rate
2010-01-01	12,367.9	11,931.2	436.7	6,927.5	64.1	64.1
2010-02-01	12,059.8	11,632.3	427.4	7,225.1	62.5	62.5
2010-03-01	12,324.2	11,895.9	428.3	6,993.0	63.8	63.8
2010-04-01	12,512.0	12,133.5	378.5	6,894.5	64.5	64.5
2010-05-01	12,197.6	11,798.9	398.7	7,186.0	62.9	62.9
2010-06-01	12,331.8	11,882.2	449.6	7,065.0	63.6	63.6
2010-07-01	12,516.2	12,099.5	416.7	6,900.5	64.5	64.5
2010-08-01	12,320.9	11,936.5	384.3	7,157.9	63.2	63.2
2010-09-01	12,289.9	11,912.1	277.8	7,104.4	62.1	62.1

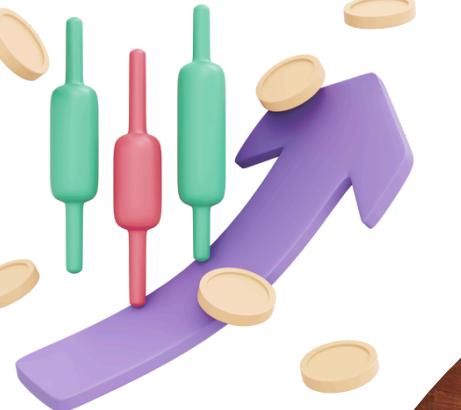
PROCESSES INVOLVED

DATA
PREPROCESSING

EXPLORATORY
DATA
ANALYSIS

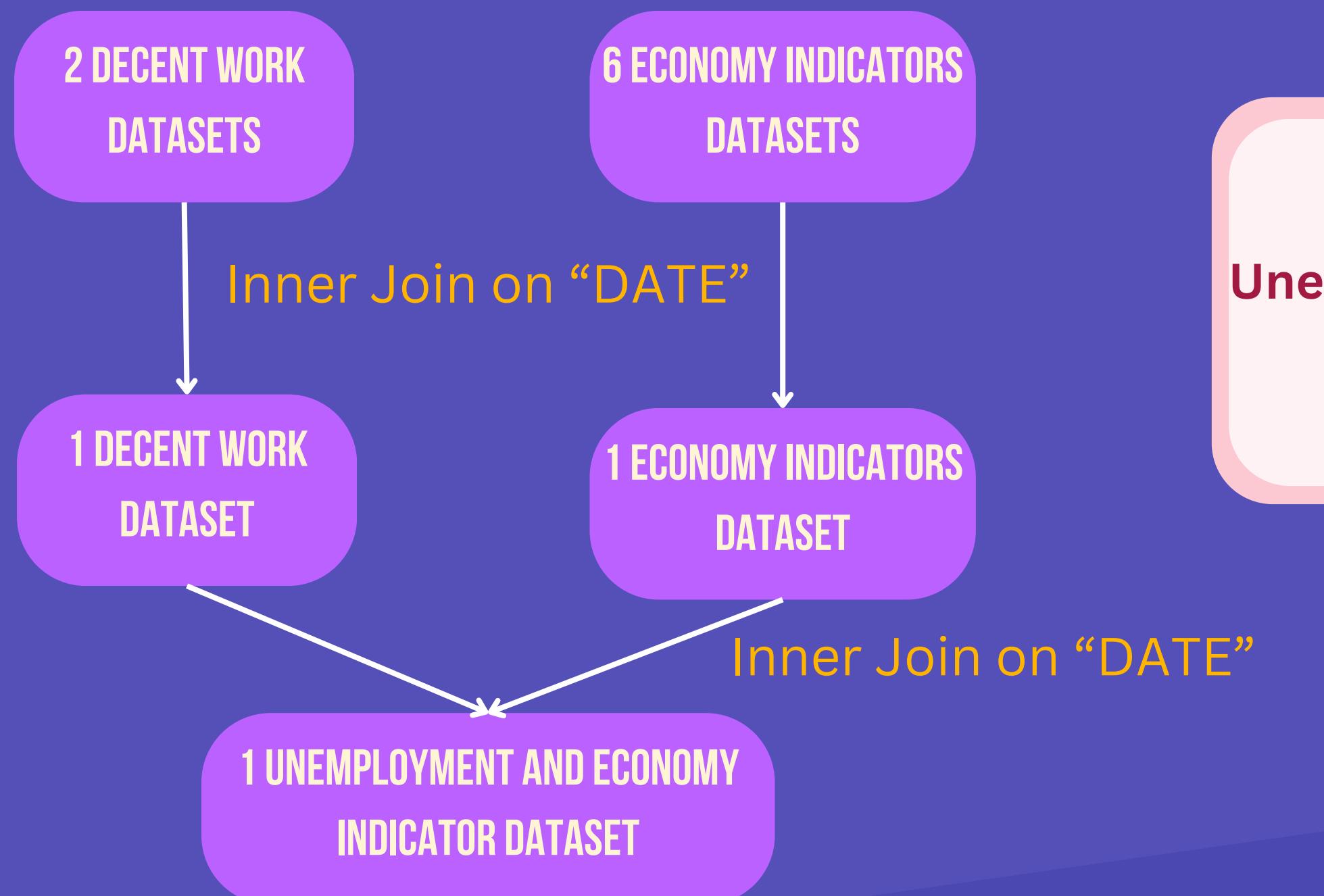
DATA
MODELLING

MODEL
EVALUATION



DATA PREPROCESSING

Merging dataset



Feature engineering

Unemployment rate

Low(<Q1)

Medium(>Q1 & <Q3)

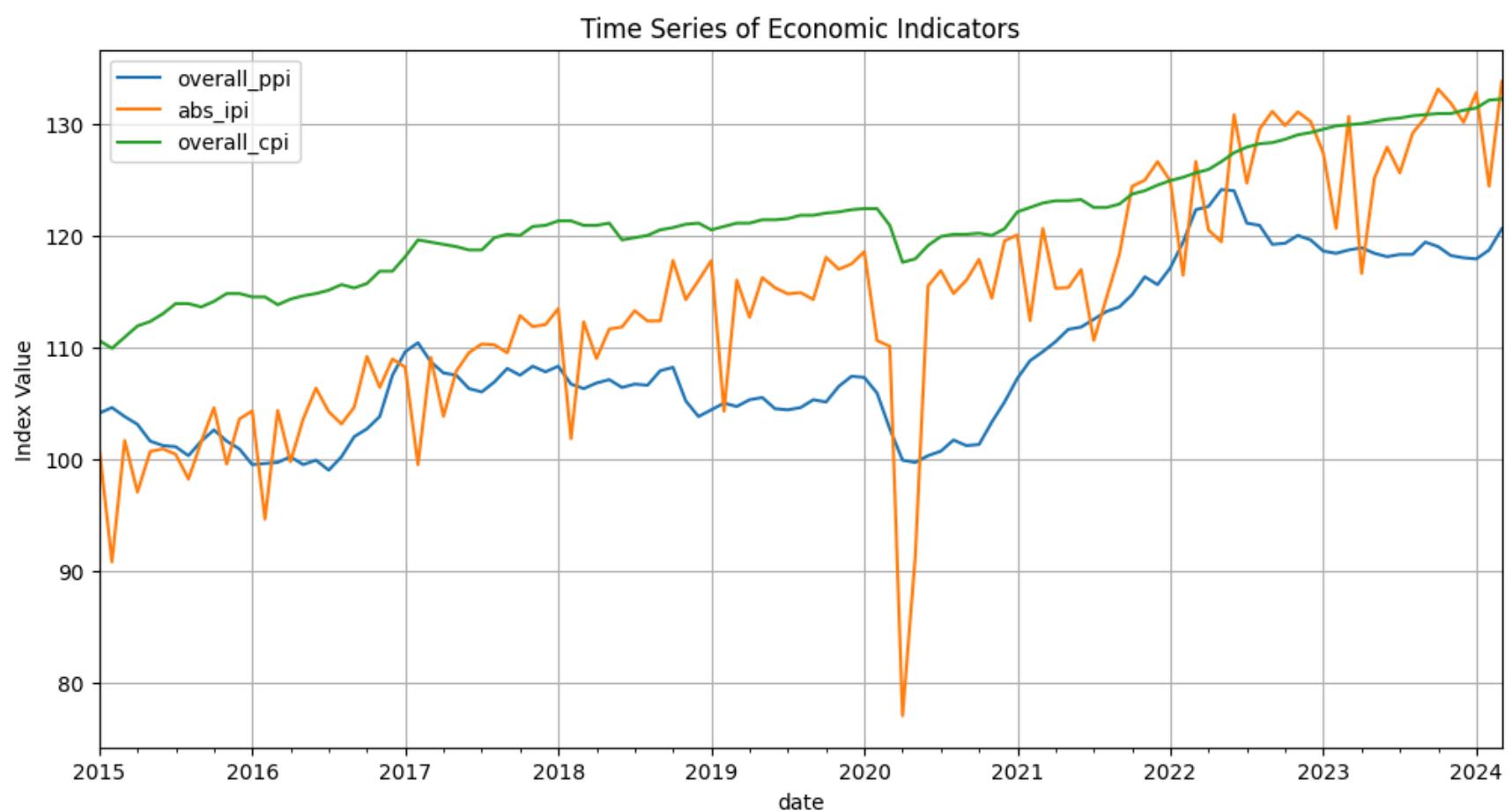
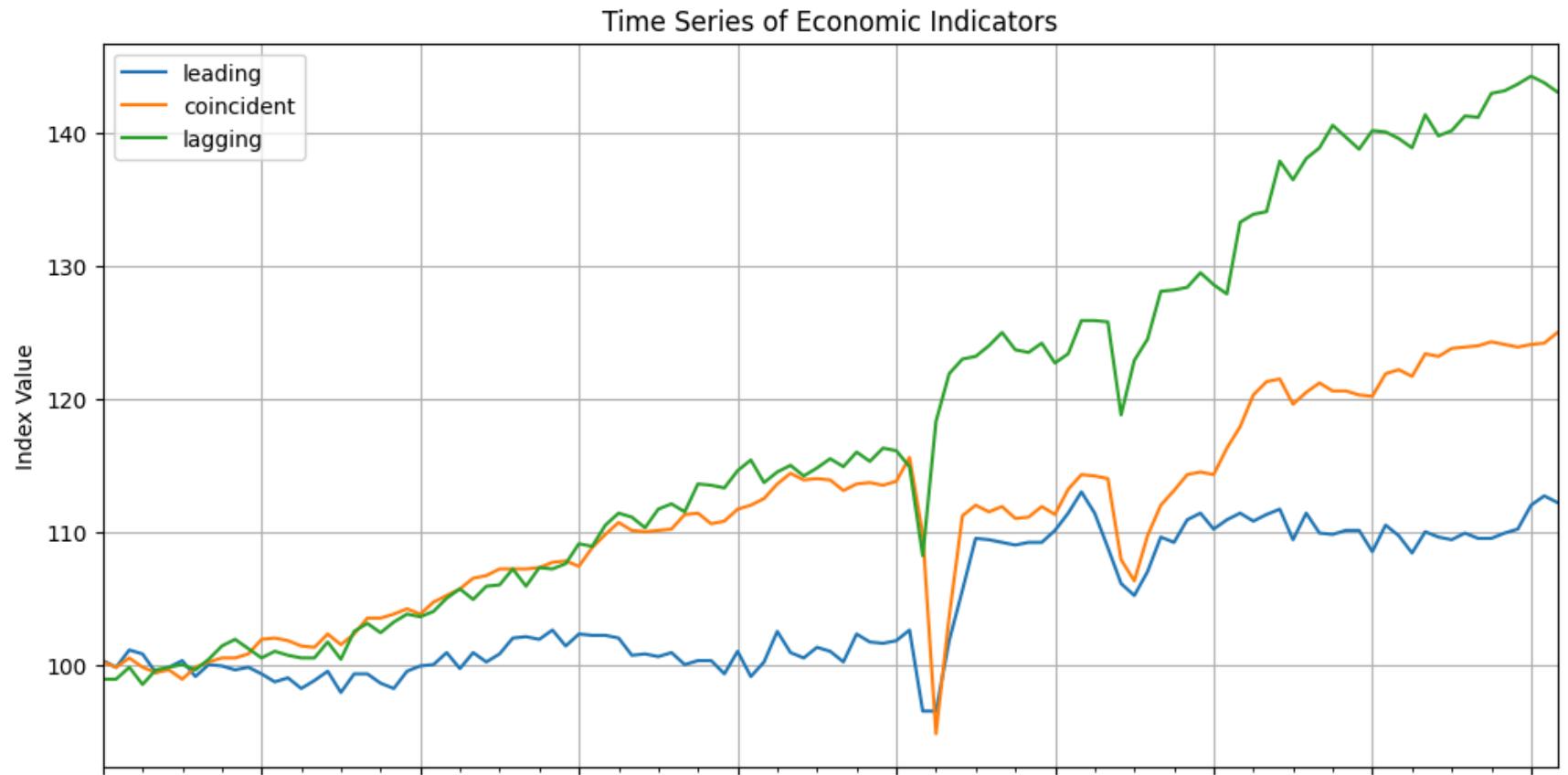
High(>Q3)



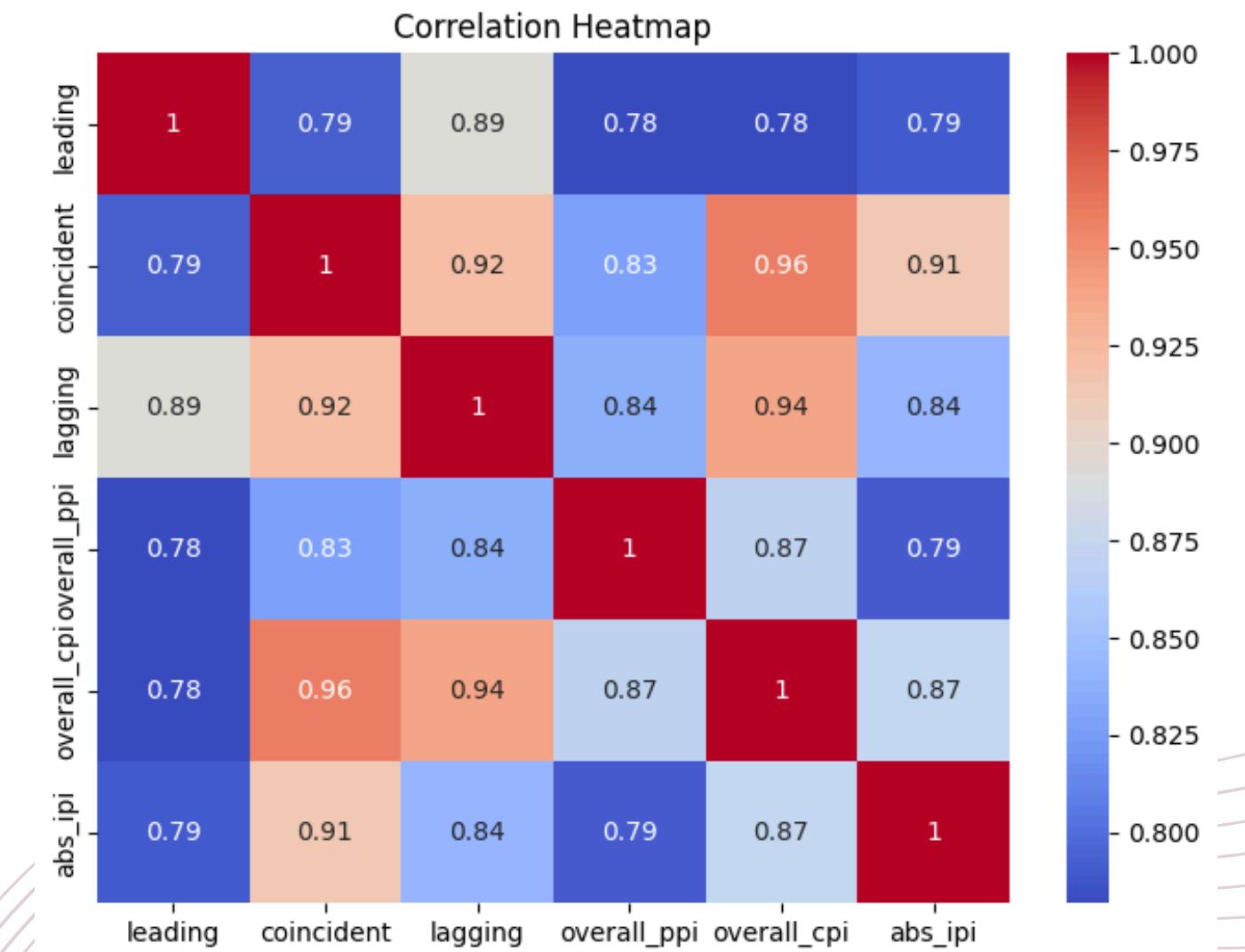
NEED
A
JOB

Unemployment

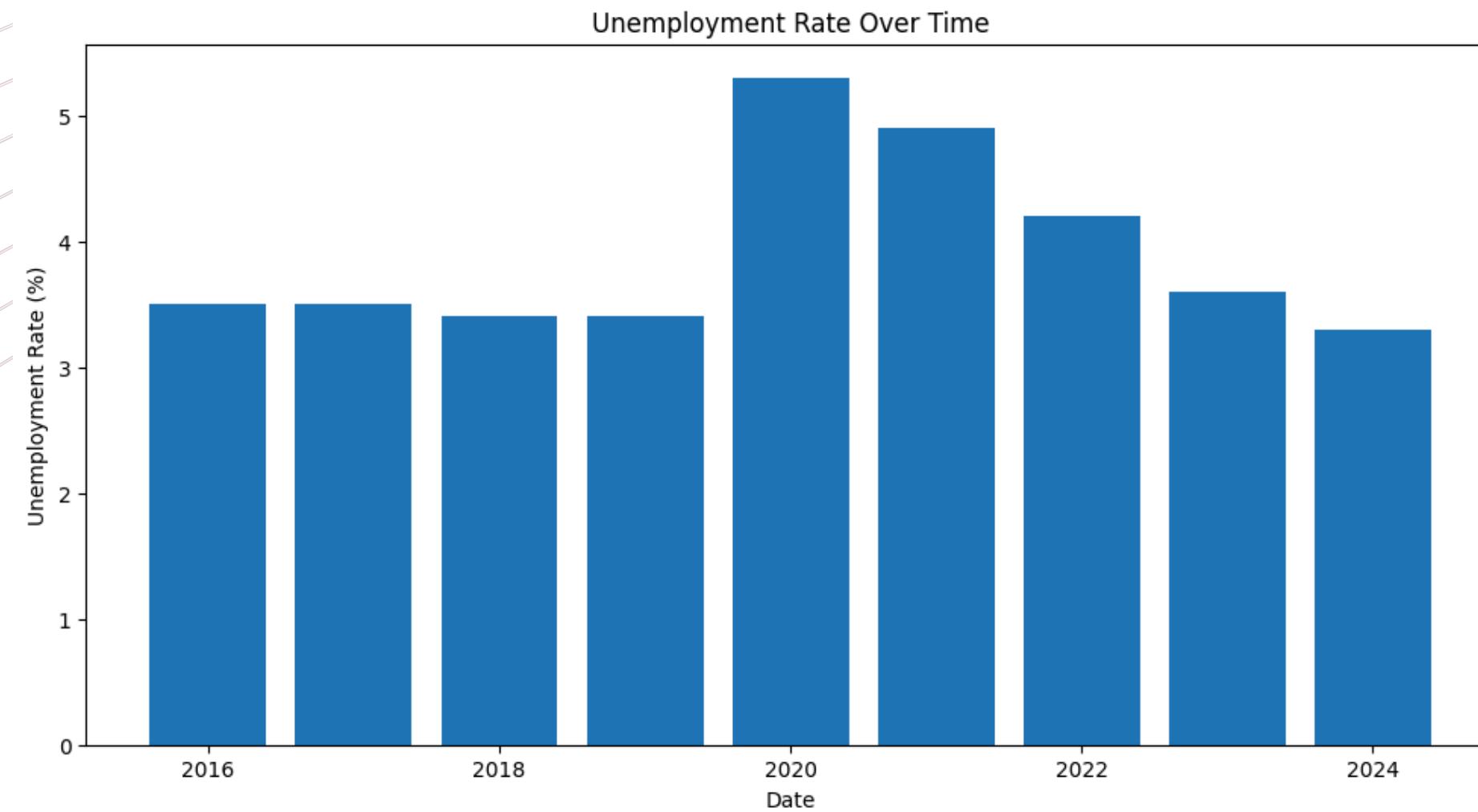
EXPLORATORY DATA ANALYSIS



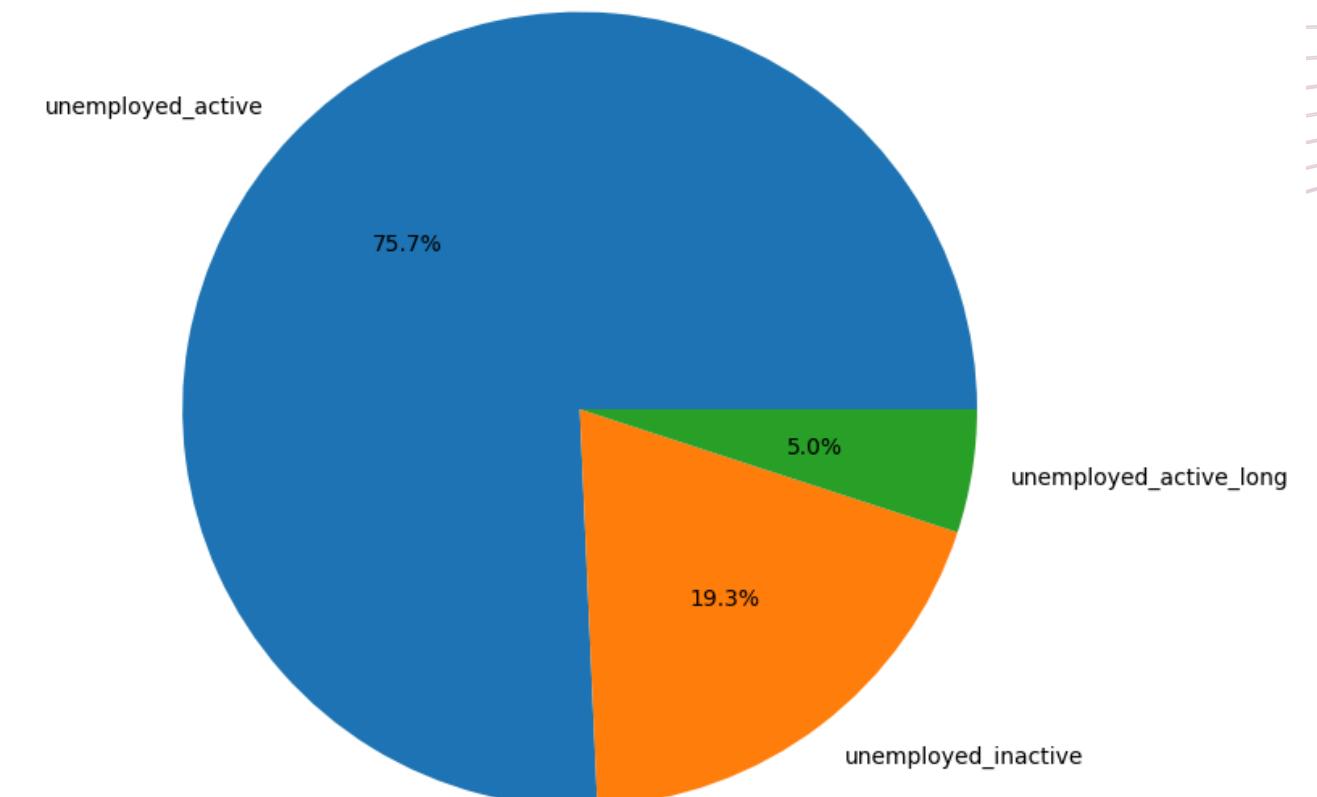
- Leading indicators, which predict future economic activity, show a steady increase with a notable dip in 2020, followed by a recovery.
- Economic indicators experienced significant volatility around 2020 due to the pandemic, with signs of recovery in subsequent years
- Steady increase in CPI suggests rising consumer prices (inflation)



EXPLORATORY DATA ANALYSIS

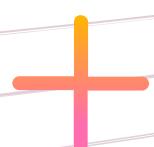


Breakdown of Unemployment Categories



- After 2020, the unemployment rate shows a decreasing trend, suggesting economic recovery and improvements in the job market
- Post-2020, recovery efforts and reopening of businesses reduces unemployment
- Reskilling and upskilling initiatives succeeded to make workforce more adaptable to changing job market demands

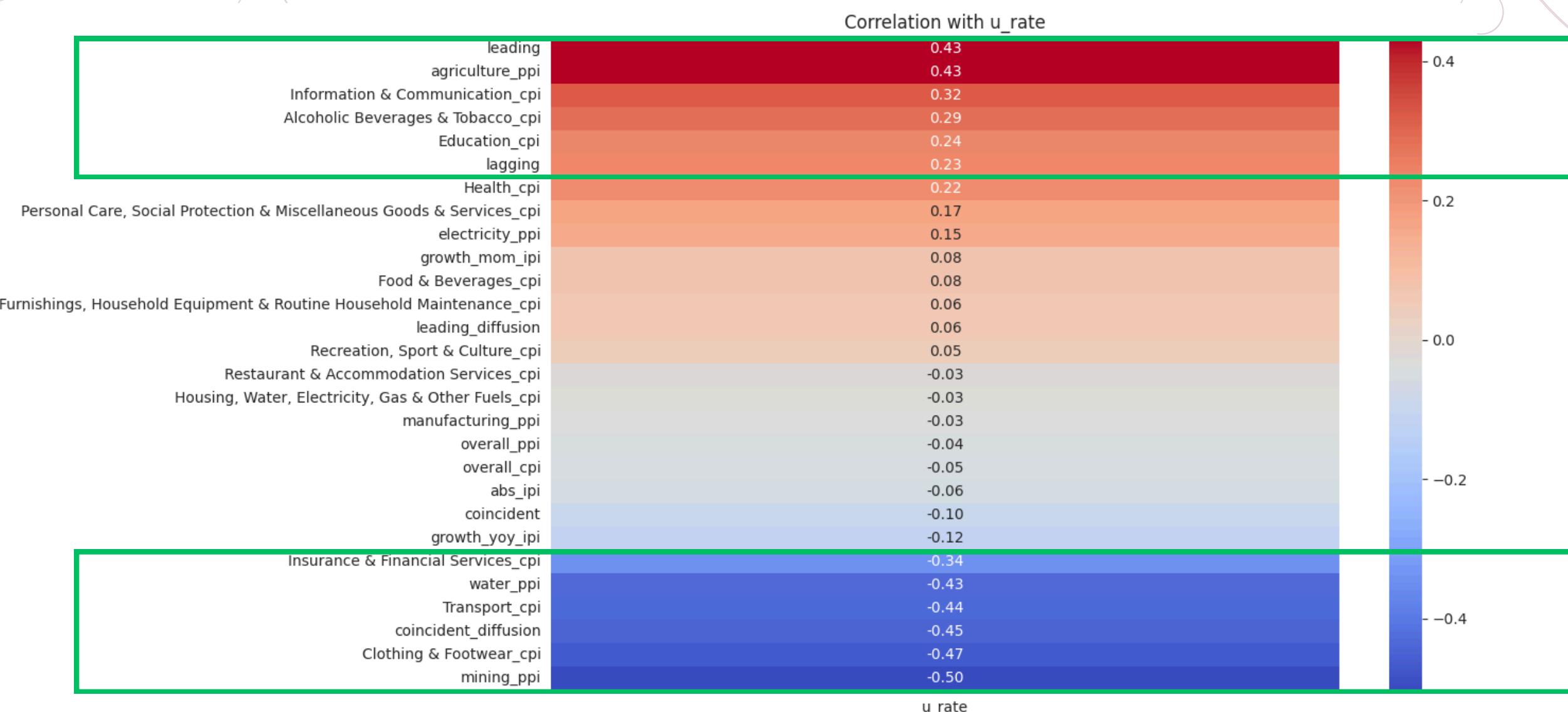
- The majority of the unemployed population falls into the "unemployed_active" category, indicating that most unemployed individuals are actively seeking work



EXPLORATORY DATA ANALYSIS



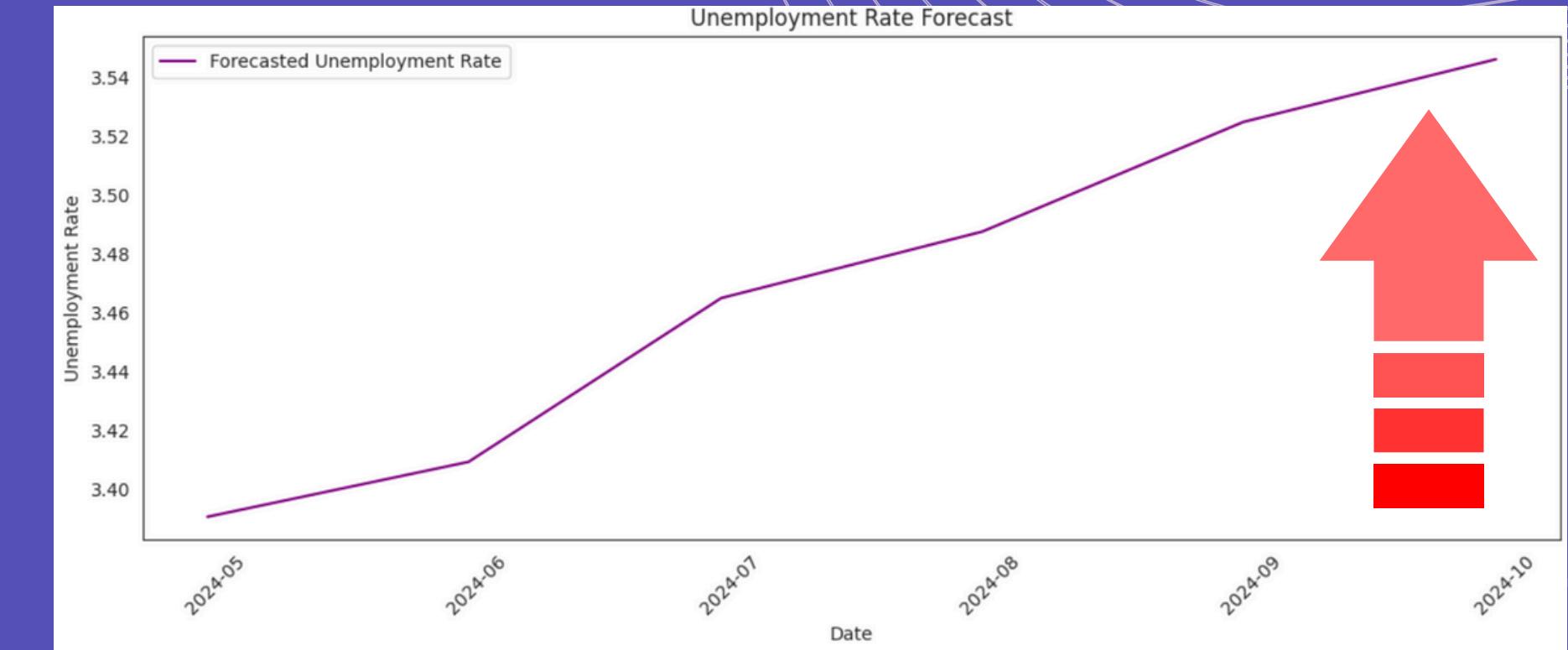
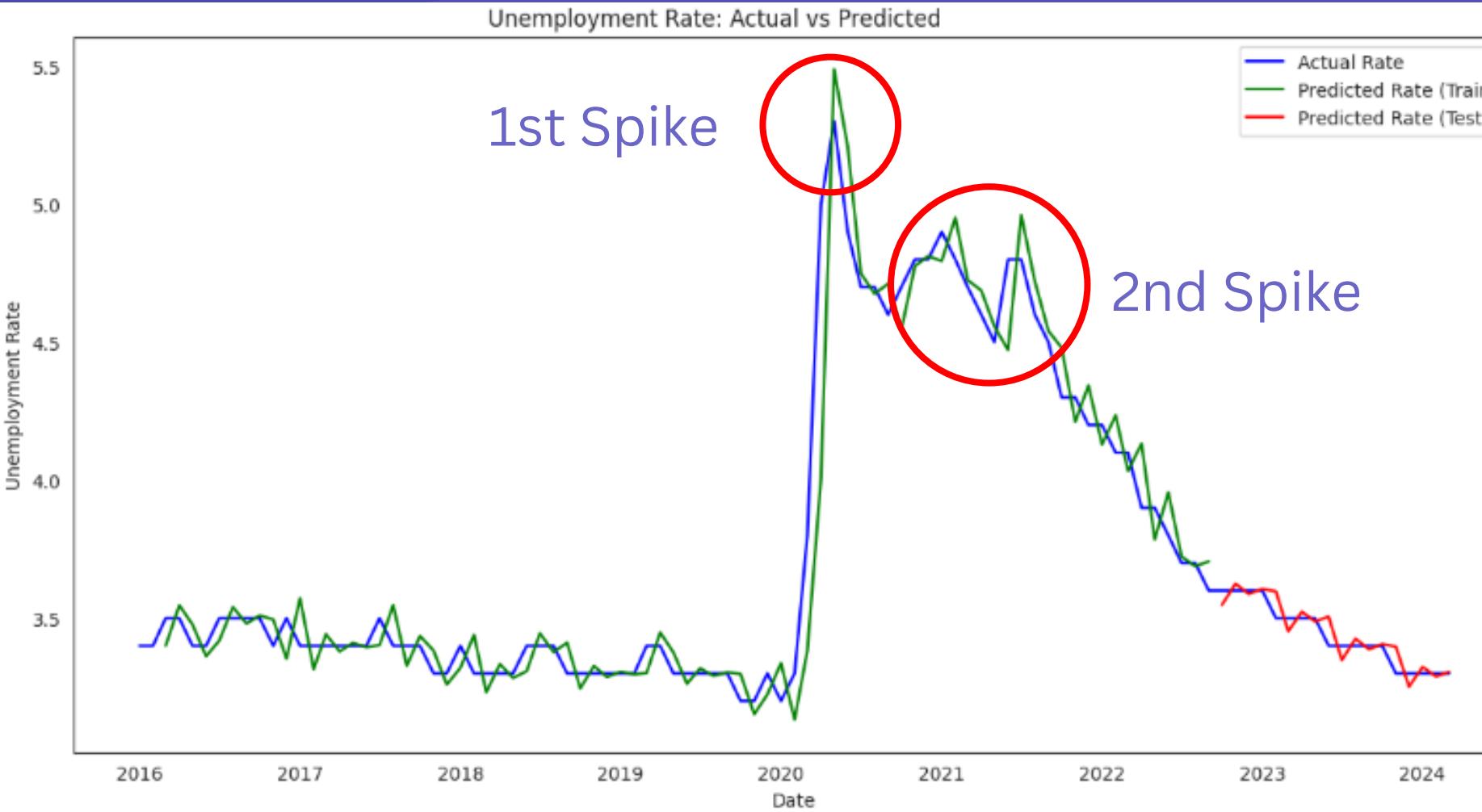
- Agriculture: Economic distress in rural areas.
- Information & Communication: Barrier to job search and increase business operational costs.
- Education: Higher cost leads to decreased educational attainment and skill acquisition.



- Mining: Improved trade balances, leading to stronger currency valuation and more capital inflows which can be reinvested in other sectors.
- Transport: Allow cheaper commute, more access to job opportunities.
- Insurance & Financial Service: Stable economic, encourage investment and job creation



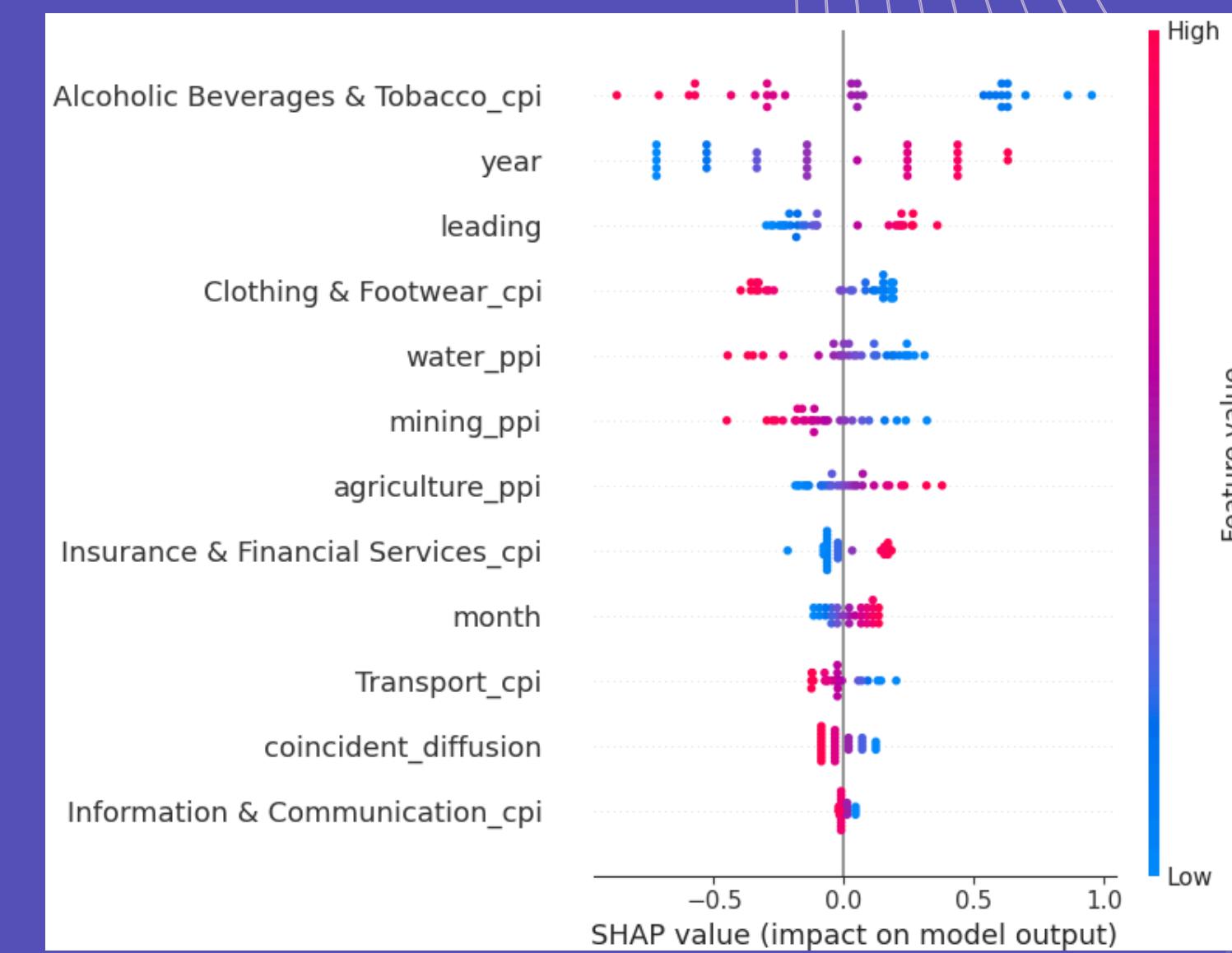
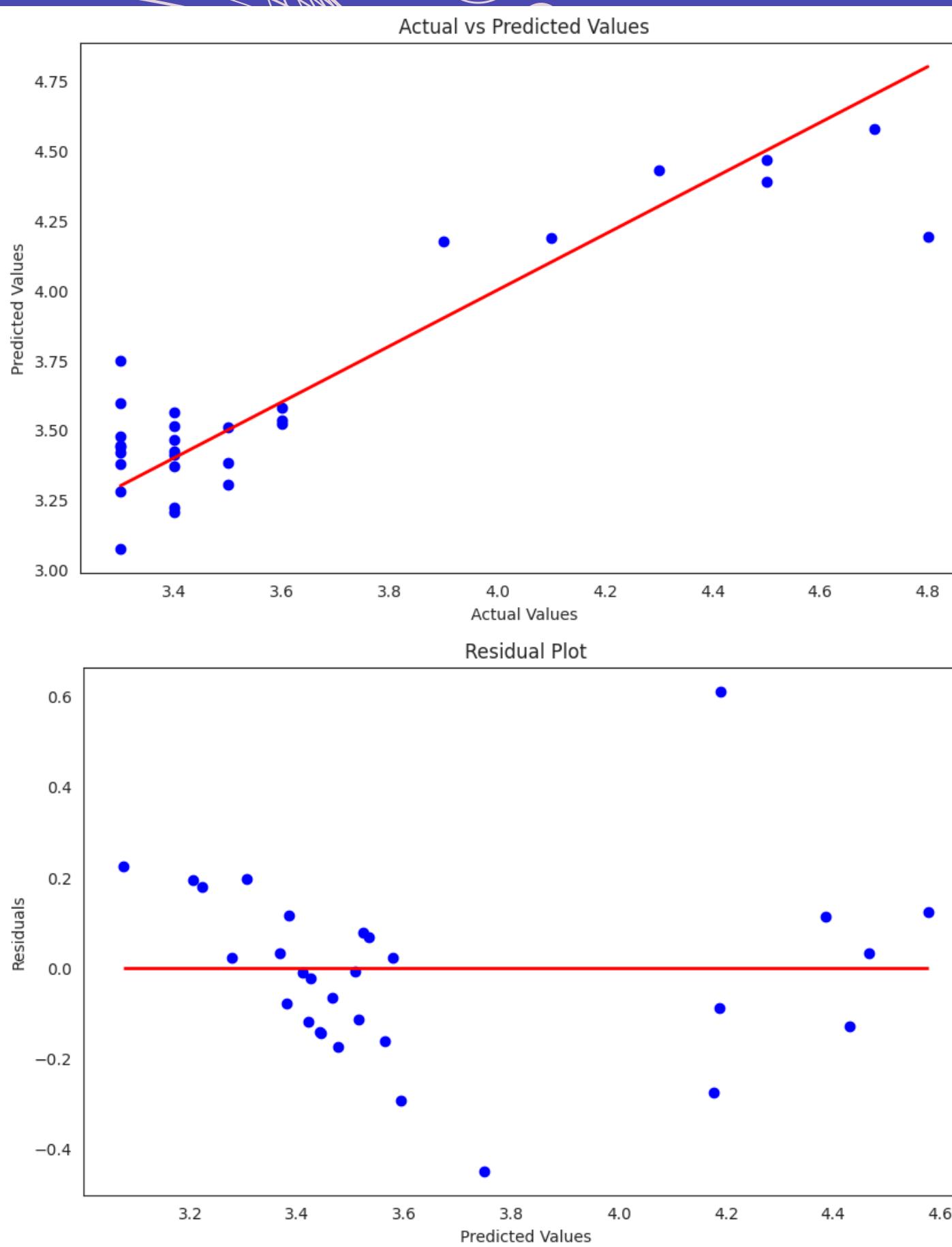
SEASONAL ARIMA MODEL



- Significant spike in year 2020: COVID-19 pandemic which affect the job market
- 1st spike: Model does not anticipate crisis event, quite large deviation from actual spike
- 2nd spike (post-spike): Model recalibrated and adapted to new post-crisis conditions, align more closely with the actual rate

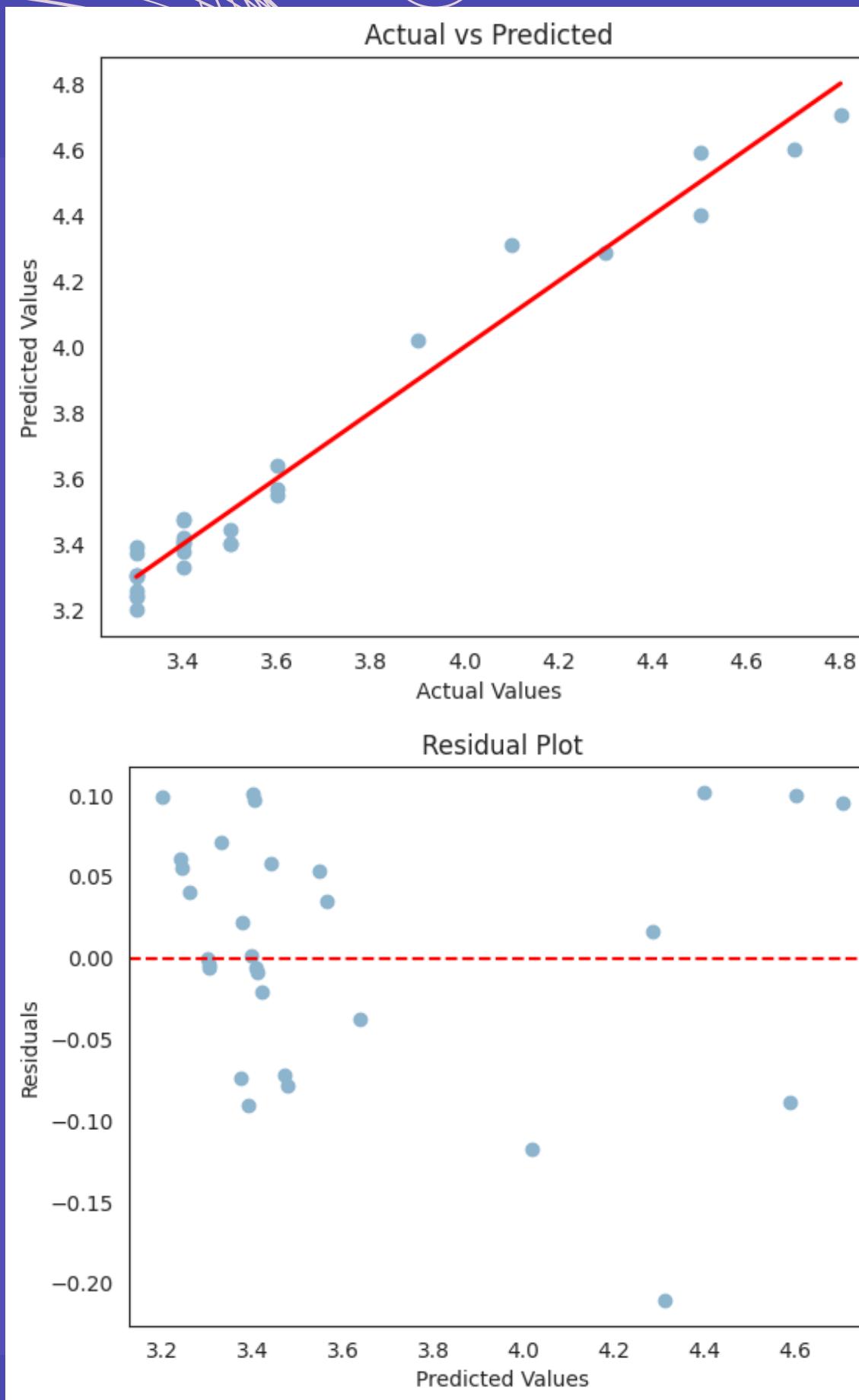
- Upward trend in unemployment rate for the next 6 months
- Structural labour market issues:
 1. Skill Mismatch
 2. Youth Unemployment: Difficulties in entering the workforce
- Gig economy provides flexible work but may not always offer stable employment

LINEAR REGRESSION MODEL

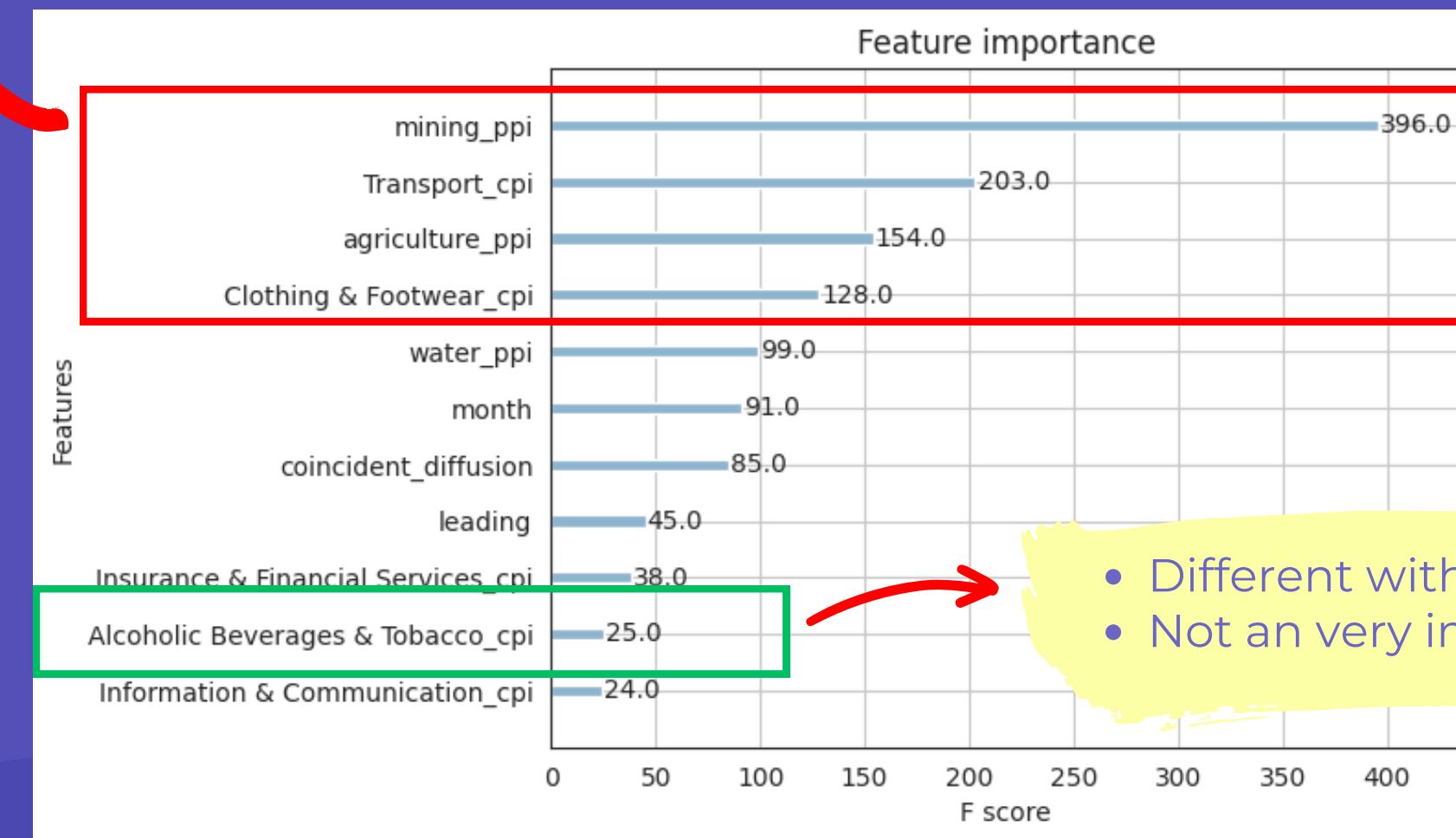


- The year impact the unemployment rate because the timeline span from before COVID, during COVID and after COVID
- Alcoholic Beverage & Tobacco: Health issues, leading to absenteeism and decreased work performance
- Clothing & Footware: Changing of customer preferences and financial status, and seasonal demand impact the whole manufacturing and supply chain

XGBOOST REGRESSION MODEL



- More align with the correlation analysis in EDA
- Top 4 sectors: mining, transport, agriculture, clothing and footware
- Stimulate other sectors and create more job opportunities



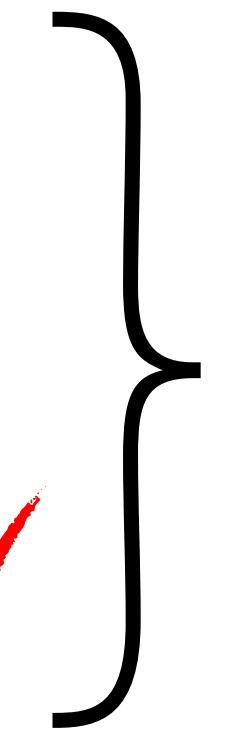
MODEL EVALUATION



MODELS	RMSE VALUE	R-SQUARED VALUE
SARIMA MODEL	0.1514	0.9312
LINEAR REGRESSION MODEL	0.0371	0.8214
XGBOOST MODEL	0.0760	0.9722



- High accuracy (low RMSE value)
- Captures the underlying structure and dynamics of the time series data effectively
- Reliable forecasting

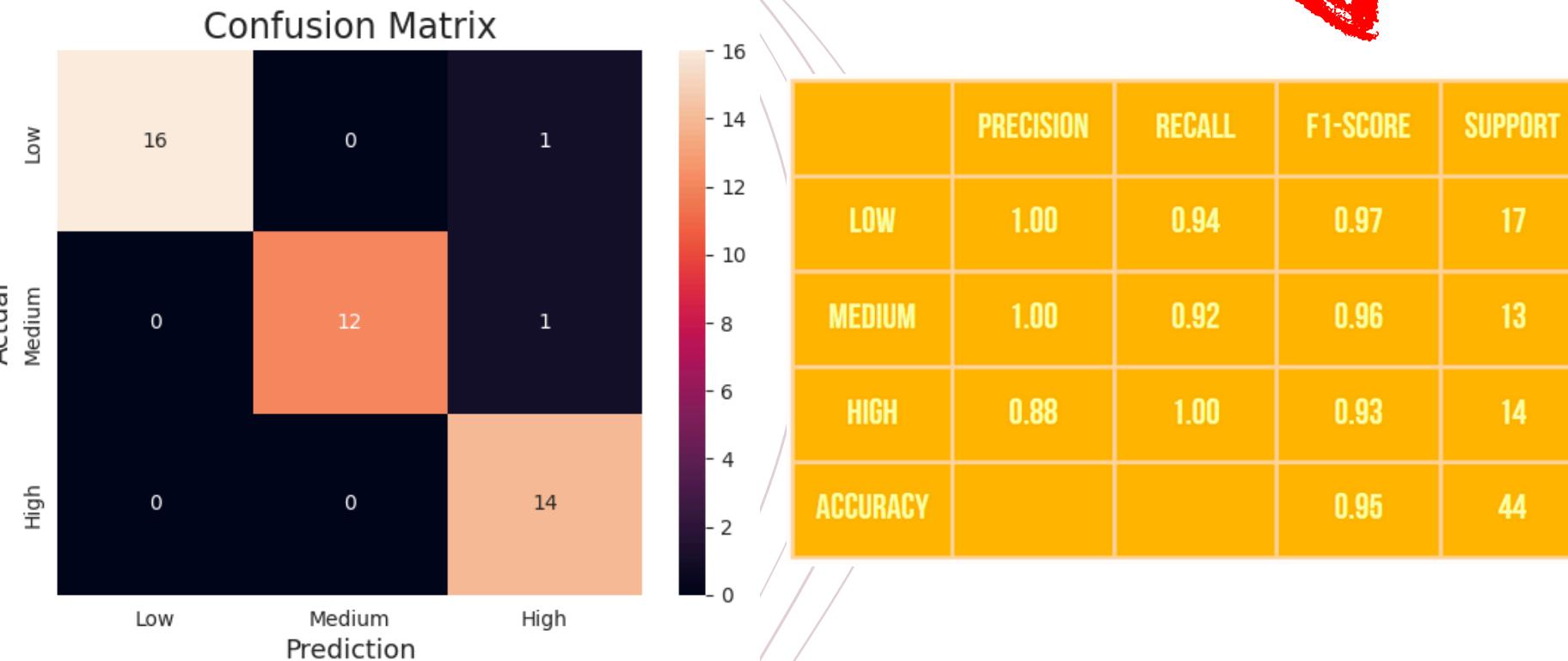


- Linear regression performs better in terms of minimizing prediction errors than XGBoost
- XGBoost captures larger significant portion of the underlying patterns in data
- Adapts better to various data distributions
- Ensemble approach gives more robust and stable results, reducing the risk of relying on a single model's biases

CLASSIFICATION MODELS



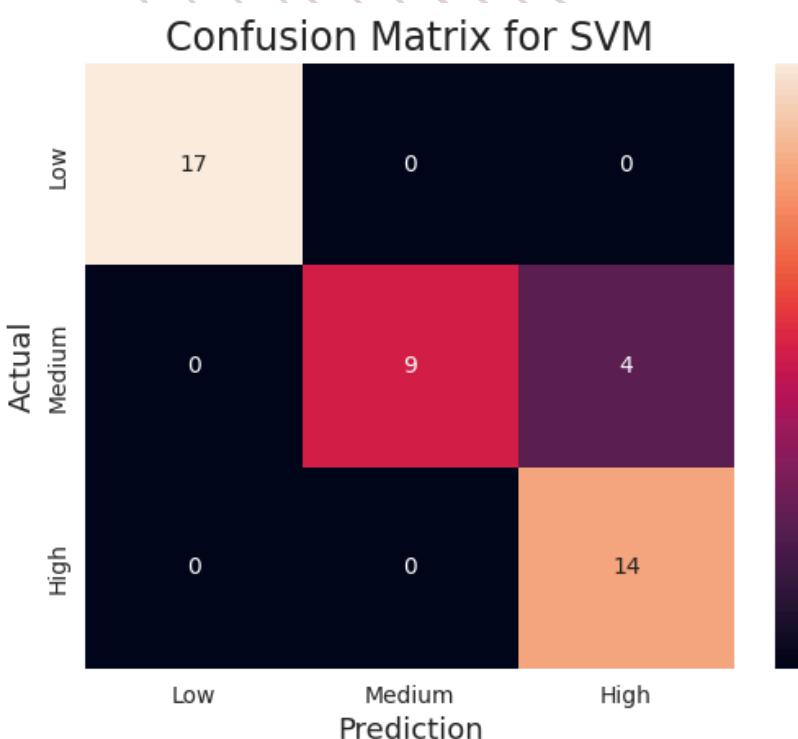
RANDOM FOREST CLASSIFIER



SVM CLASSIFIER

Confusion Matrix for SVM

		Low	Medium	High
Actual	Low	17	0	0
	Medium	0	9	4
High	0	0	14	



- High precision and recall across the categories: Robust and reliable model
- High F1-score: Balanced precision and recall, minimising false negatives and false positives
- High precision: Ensures efficient resource allocation
- High recall: Timely interventions and accurate decision making

RECOMMENDATIONS

Invest in Education and Skills Development

- Collaborate with private sector to provide more training and upskilling programs
- Create more programs similar to AI for citizens (AI untuk Rakyat)
- Enhance citizens digital literacy and understanding of AI
- Bridge the technology and skill gap between market demands and youth's skills

Focus on Key Industry Sector

- Invest in modern mining technologies and infrastructure
- Provide tax breaks, grants, and subsidies to businesses in the clothing and footwear industry
- Invest in public transport infrastructure to improve accessibility and reduce transportation costs



Enhance Employment Insurance Scheme (EIS)

- Continuously evaluate the effectiveness of EIS policies and adjust based on real-time data
- Involve labour unions in shaping EIS policies and advocate for workers' rights
- Strengthen the job placement services to help retrenched workers find new jobs quickly



CONCLUSION



Model	Suitability	Limitation
SARIMA	Incorporating seasonality, suitable for time-series data	Assumes that the time series is stationary
Linear regression	Provides interpretable coefficients	Assume linear relationship, not suitable for complex economic systems
XGBoost	Ensemble learning	May overfit without proper regularization
Random Forest Classifier	Ensemble learning	Less interpretable than linear model
SVM classifier	Can handle non-linear relationship	Understanding of kernel function to make selection

- Key economic indicators like CPI and PPI have significant impact on the unemployment rate.
- Combining policies, data-driven insights and collaboration yields the best result in combating unemployment.
- Continuously monitor and evaluate policies to adapt to changing economic condition.

