Yuzhi Liu
Jing Cheng
Guanyu Qu
Meixu Chen

Analyzing Different Factors Impacting the Unemployment Rate

INTRODUCTION

OVERVIEW OF THE DATA STRUCTURE

Dataset Source: U.S. Bureau of labor statistics

Dataset Composition

- Rows: Each row represents a year, starting from 1974.
- Columns: The first column lists the years, and the subsequent columns represent months from January to December, with data recorded for each month.

Relevance in Analysis

- Time Series Data: long-term view of economic levels, making it suitable for trend analysis over time
- Economic Indicators: unemployment level, unemployment rate, CPI are key indicators, useful for assessing economic conditions.

Application

Forecasting: Economists and analysts can use historical unemployment trends to forecast future labor market conditions.

Economic Research: Researchers can use this data to study the impact of economic policies, recessions, and expansions on employment levels.

Policy Making: Government agencies can analyze this data to inform policy decisions related to labor markets, unemployment benefits, and workforce development programs.

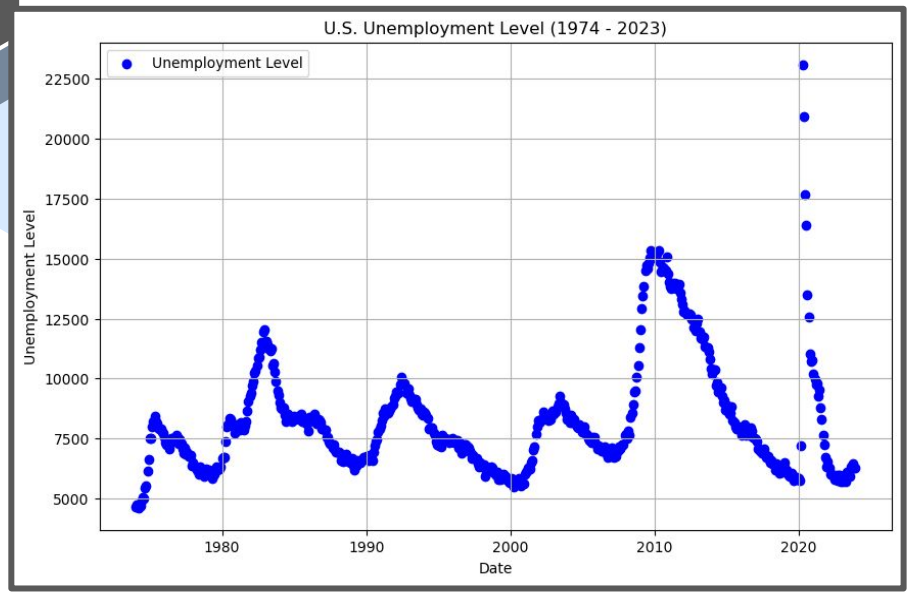
unemployment_rate

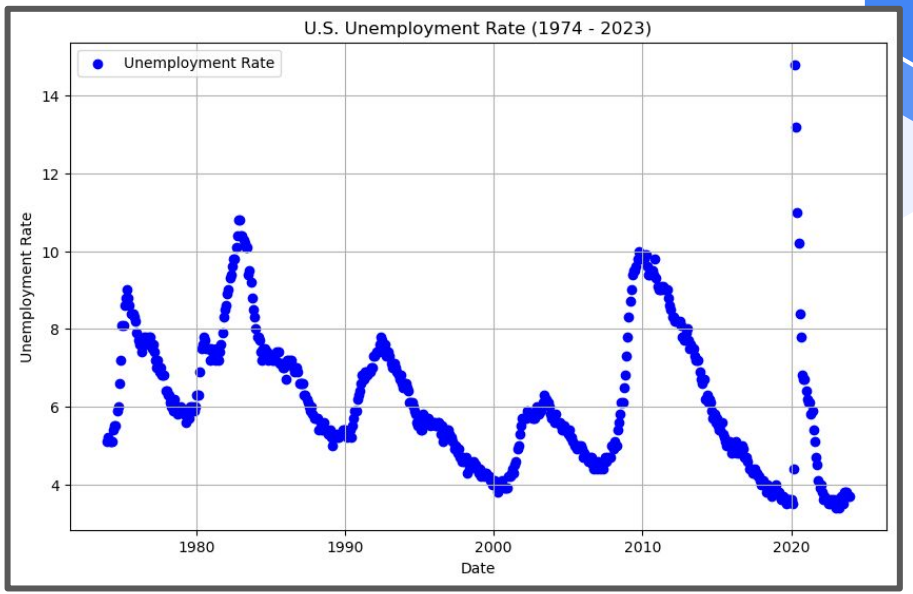
1	**Labor Force Statistics from the Current Population Survey**												
2	**Original Data Value**												
3													
4	**Series Id:**	LNS14000000											
5	**Seasonally Adjusted**												
6	**Series title:**	(Seas) Unemployment Rate											
7	**Labor force status:**	Unemployment rate											
8	**Type of data:**	Percent or rate											
9	**Age:**	16 years and over											
10	**Years:**	1973 to 2023											
11													
12	**Year**	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
13	**1973**	4.9	5.0	4.9	5.0	4.9	4.9	4.8	4.8	4.8	4.6	4.8	4.9
14	**1974**	5.1	5.2	5.1	5.1	5.1	5.4	5.5	5.5	5.9	6.0	6.6	7.2
15	**1975**	8.1	8.1	8.6	8.8	9.0	8.8	8.6	8.4	8.4	8.4	8.3	8.2
16	**1976**	7.9	7.7	7.6	7.7	7.4	7.6	7.8	7.8	7.6	7.7	7.8	7.8
17	**1977**	7.5	7.6	7.4	7.2	7.0	7.2	6.9	7.0	6.8	6.8	6.8	6.4
18	**1978**	6.4	6.3	6.3	6.1	6.0	5.9	6.2	5.9	6.0	5.8	5.9	6.0
19	**1979**	5.9	5.9	5.8	5.8	5.6	5.7	5.7	6.0	5.9	6.0	5.9	6.0
20	**1980**	6.3	6.3	6.3	6.9	7.5	7.6	7.8	7.7	7.5	7.5	7.5	7.2
21	**1981**	7.5	7.4	7.4	7.2	7.5	7.5	7.2	7.4	7.6	7.9	8.3	8.5
22	**1982**	8.6	8.9	9.0	9.3	9.4	9.6	9.8	9.8	10.1	10.4	10.8	10.8
23	**1983**	10.4	10.4	10.3	10.2	10.1	10.1	9.4	9.5	9.2	8.8	8.5	8.3
24	**1984**	8.0	7.8	7.8	7.7	7.4	7.2	7.5	7.5	7.3	7.4	7.2	7.3
25	**1985**	7.3	7.2	7.2	7.3	7.2	7.4	7.4	7.1	7.1	7.1	7.0	7.0
26	**1986**	6.7	7.2	7.2	7.1	7.2	7.2	7.0	6.9	7.0	7.0	6.9	6.6
27	**1987**	6.6	6.6	6.6	6.3	6.3	6.2	6.1	6.0	5.9	6.0	5.8	5.7

- Data loading
- Removing irrelevant rows
- Column selection
- Renaming Columns
- Reshaping data
- Data creation and filtering
- Sorting

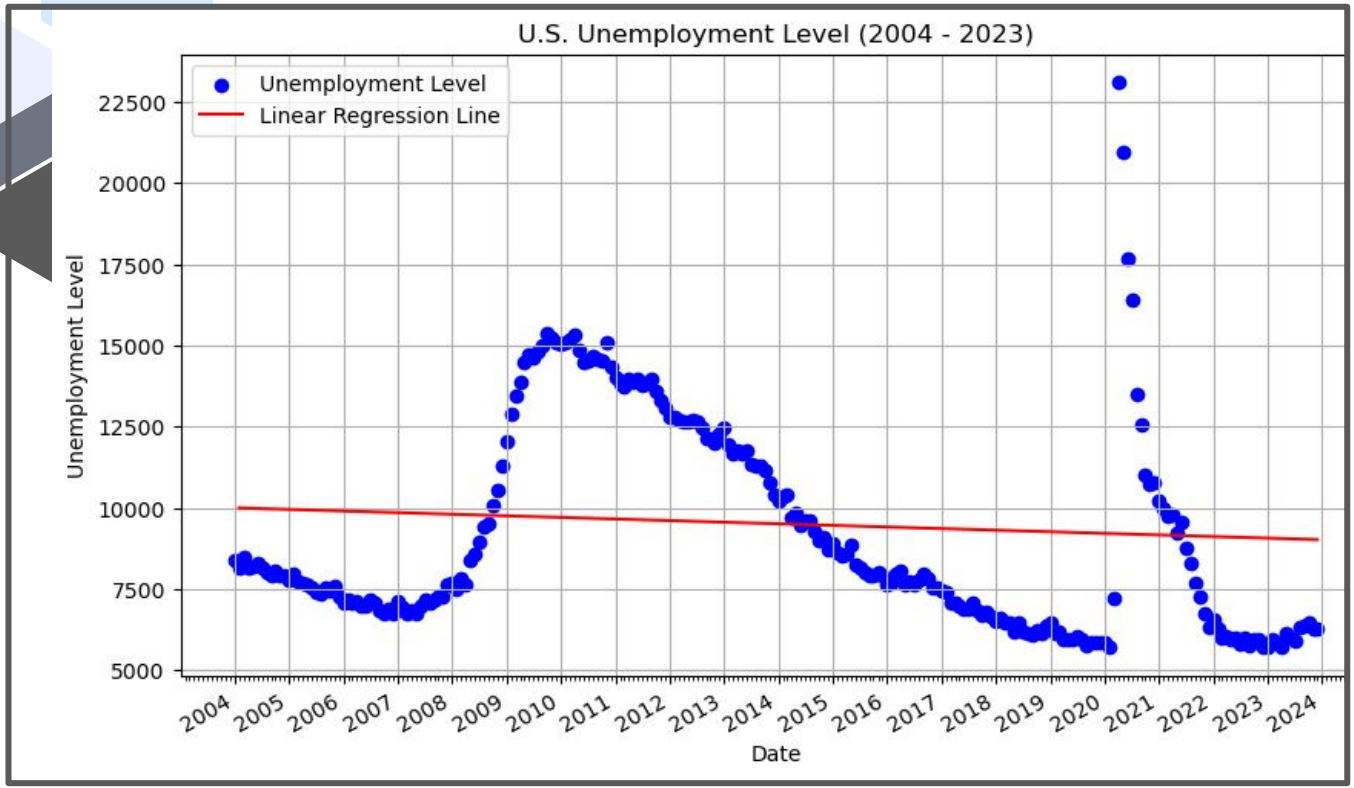
CLEAN DATA

10	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
11	1974	5.1	5.2	5.1	5.1	5.1	5.4	5.5	5.5	5.9	6.0	6.6	7.2
12	1975	8.1	8.1	8.6	8.8	9.0	8.8	8.6	8.4	8.4	8.4	8.3	8.2
13	1976	7.9	7.7	7.6	7.7	7.4	7.6	7.8	7.8	7.6	7.7	7.8	7.8
14	1977	7.5	7.6	7.4	7.2	7.0	7.2	6.9	7.0	6.8	6.8	6.8	6.4
15	1978	6.4	6.3	6.3	6.1	6.0	5.9	6.2	5.9	6.0	5.8	5.9	6.0
16	1979	5.9	5.9	5.8	5.8	5.6	5.7	5.7	6.0	5.9	6.0	5.9	6.0
17	1980	6.3	6.3	6.3	6.9	7.5	7.6	7.8	7.7	7.5	7.5	7.5	7.2
18	1981	7.5	7.4	7.4	7.2	7.5	7.5	7.2	7.4	7.6	7.9	8.3	8.5



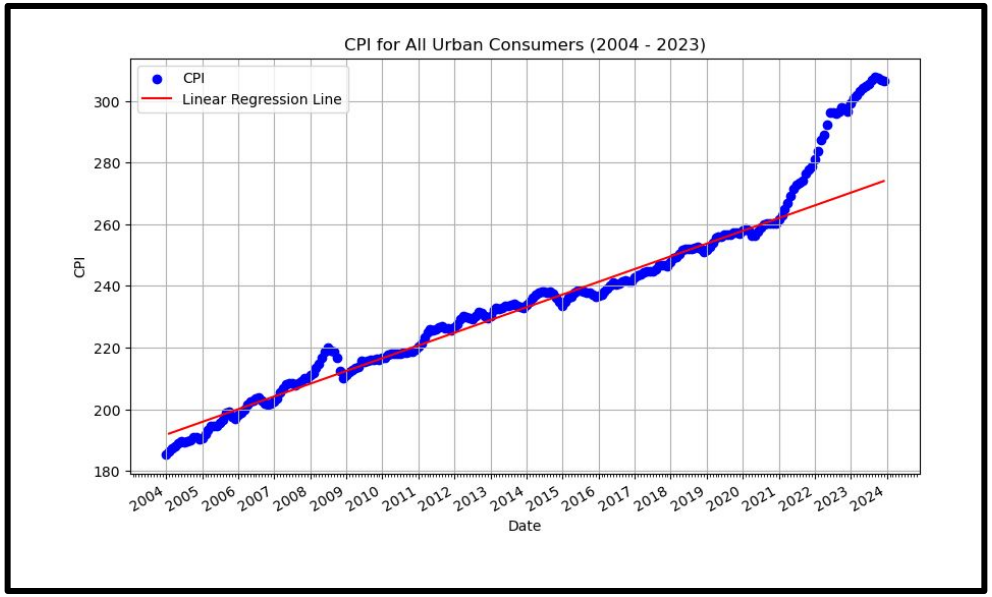


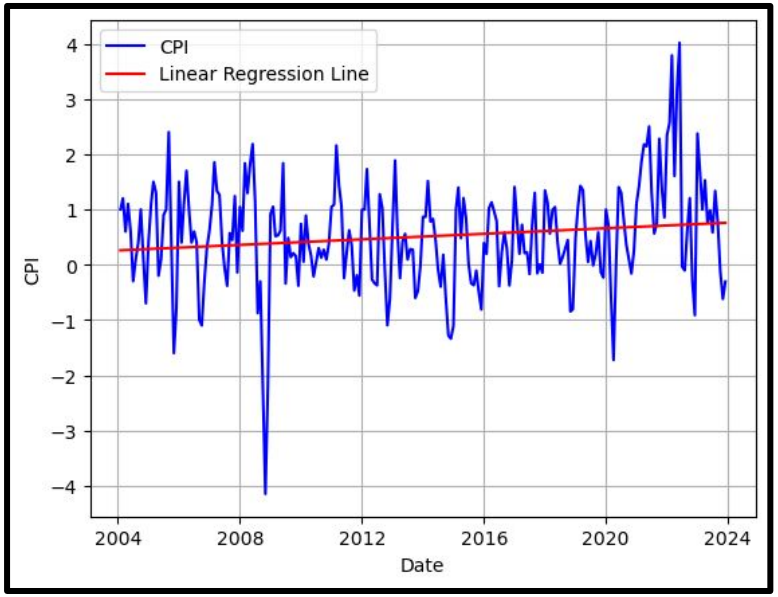
Nearly the SAME



Data:
unemployment level (2004 - 2023 Dec)
X Label:
Date
Y Label:
unemployment level
Train:
2004-2019
Linear regression

Relationship between COVID-19 and Unemployment Level

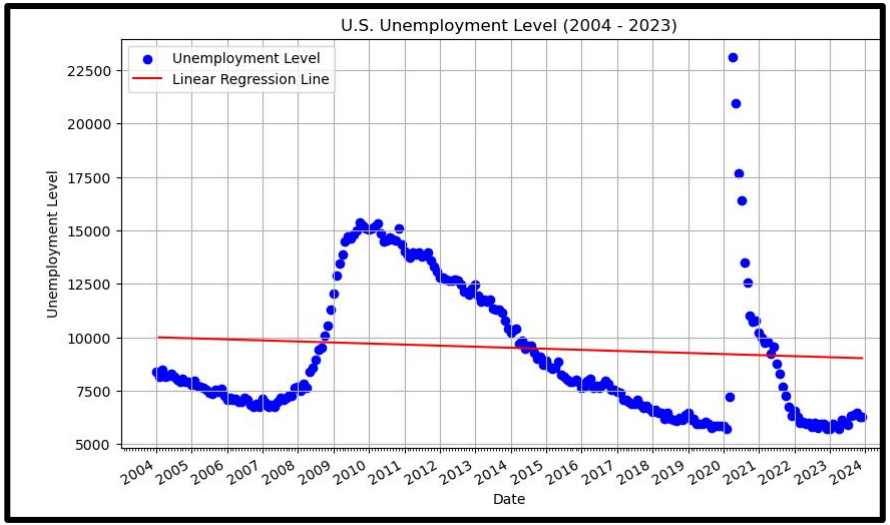


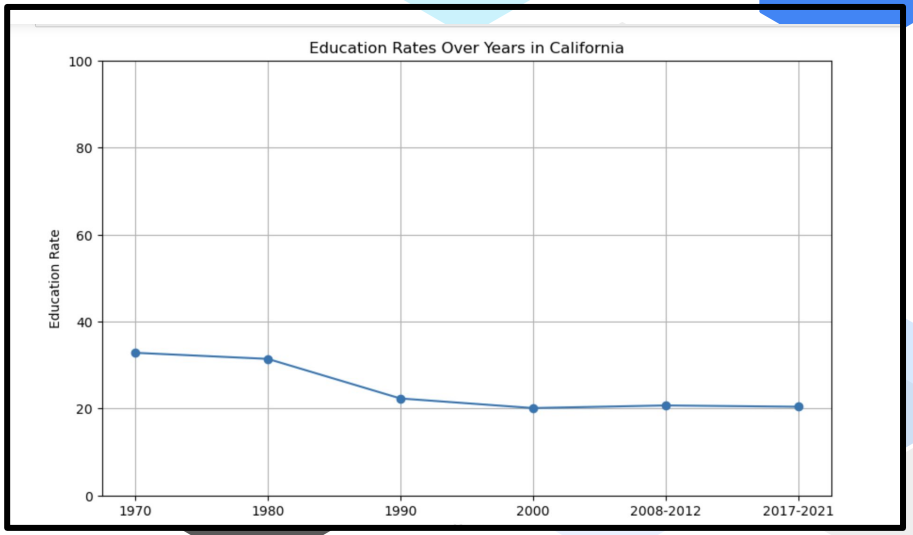


Relationship between COVID-19 and Consumer Price Index (CPI)

- Data: CPI for all consumers
- X Label: Date (2004-2023)
- Y Label: consumer price index
- Train: 2004-2019
- Linear Regression
- First-Order Difference

Unemployment Rate VS. Education Data





The regression line was found to be decreasing, but overall in a horizontal line.

SO DOES THE EDUCATION RATE!!

EducationReport_colldge

	Total					
Name	1970	1980	1990	2000	2008-2012	2017-2021
United States	10.7%	16.2%	20.3%	24.4%	28.5%	33.7%
Alabama	7.8%	12.2%	15.7%	19.0%	22.3%	26.7%
Alaska	14.1%	21.1%	23.0%	24.7%	27.5%	30.6%
Arizona	12.6%	17.4%	20.3%	23.5%	26.6%	31.2%
Arkansas	6.7%	10.8%	13.3%	16.7%	19.8%	24.3%
California	13.4%	19.6%	23.4%	26.6%	30.5%	35.3%
Colorado	14.9%	23.0%	27.0%	32.7%	36.7%	42.8%
Connecticut	13.7%	20.7%	27.2%	31.4%	36.2%	40.6%
Delaware	13.1%	17.5%	21.4%	25.0%	28.5%	33.6%
District of Columbia	17.8%	27.5%	33.3%	39.1%	51.2%	61.4%

Hypothesis:

The spread of tertiary education has an upward impact on employment rates

Data:

High school education rate (1970 - 2021)

College education rate (1970 - 2021)

Unemployment rate (1974 - 2023)

	Name	1970	1980	1990	\
	NaN	NaN	NaN	NaN	
	NaN	Total	NaN	NaN	
	Name	1970	1980	1990	
District of Columbia		26.2%	25.5%	21.2%	
California		32.8%	31.4%	22.3%	
Colorado		34.4%	34.6%	26.5%	
Washington		36.2%	37.4%	27.9%	
Oregon		35.0%	37.1%	28.9%	
Utah		35.9%	36.0%	27.2%	
Massachusetts		34.9%	36.4%	29.7%	
Arizona		31.6%	34.3%	26.1%	
Virginia		25.2%	28.4%	26.6%	
Minnesota		34.5%	38.6%	33.0%	
Maryland		28.5%	32.5%	28.1%	
Texas		25.1%	28.8%	25.6%	
North Carolina		21.7%	27.8%	29.0%	
New York		31.2%	34.1%	29.5%	
Illinois		31.9%	35.1%	30.0%	
Nebraska		37.2%	40.6%	34.7%	
Kansas		35.8%	39.2%	32.8%	
New Mexico		30.0%	34.1%	28.7%	
North Dakota		27.6%	31.3%	28.0%	
Connecticut		31.7%	34.4%	29.5%	
New Jersey		31.8%	35.9%	31.1%	
United States		31.1%	34.6%	30.0%	
Idaho		34.6%	36.5%	30.4%	
Hawaii		35.9%	35.1%	28.7%	
Georgia		22.4%	28.5%	29.6%	
New Hampshire		34.4%	37.2%	31.7%	
Nevada		38.9%	40.4%	31.5%	
Montana		34.0%	38.0%	33.5%	
Rhode Island		29.0%	32.8%	29.5%	

After Cleaning the DATA

Name	1970	1980	1990	\
Name	1970	1980	1990	
United States	10.7%	16.2%	20.3%	
Alabama	7.8%	12.2%	15.7%	
Alaska	14.1%	21.1%	23.0%	
Arizona	12.6%	17.4%	20.3%	
Arkansas	6.7%	10.8%	13.3%	
California	13.4%	19.6%	23.4%	
Colorado	14.9%	23.0%	27.0%	

Name	1970	1980	1990	\
NaN	NaN	NaN	NaN	
NaN	Total	NaN	NaN	
Name	1970	1980	1990	
District of Columbia	26.2%	25.5%	21.2%	
California	32.8%	31.4%	22.3%	
Colorado	34.4%	34.6%	26.5%	
Washington	36.2%	37.4%	27.9%	
Oregon	35.0%	37.1%	28.9%	
Utah	35.9%	36.0%	27.2%	
Massachusetts	34.9%	36.4%	29.7%	

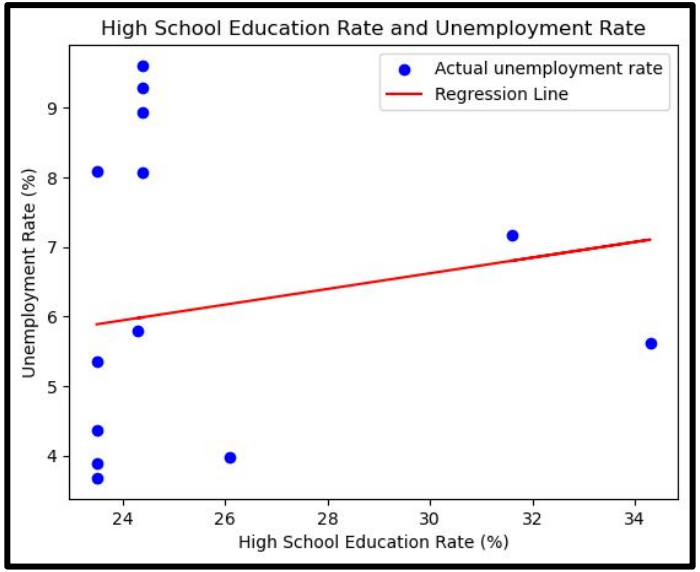
WE ARE MISSING U.S. DATA

The state that best represents the national high school education average is: California.

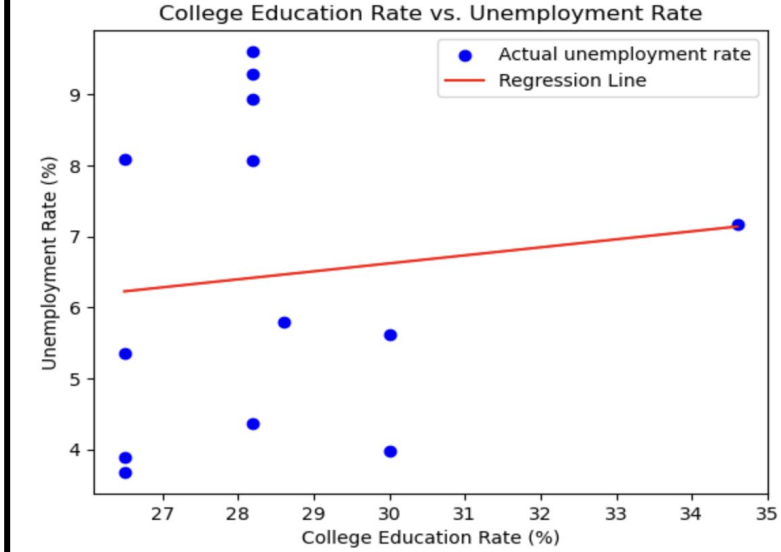
Code Methodology:

- Use pandas convert file into data frame / Clean the data
- Manhattan Distance
- Linear Regression

RESULT



X - Features:
High school Education rate
College Education Rate
Y - Label:
Unemployment rate by year



Conclusion:
They **DON'T** have correlation with each other.

Is there a cycle in unemployment rate, civilian labor force, and employed civilians?

Hypothesis: There is an annual cycle in unemployment rate, civilian labor force, and the number of employed civilians.

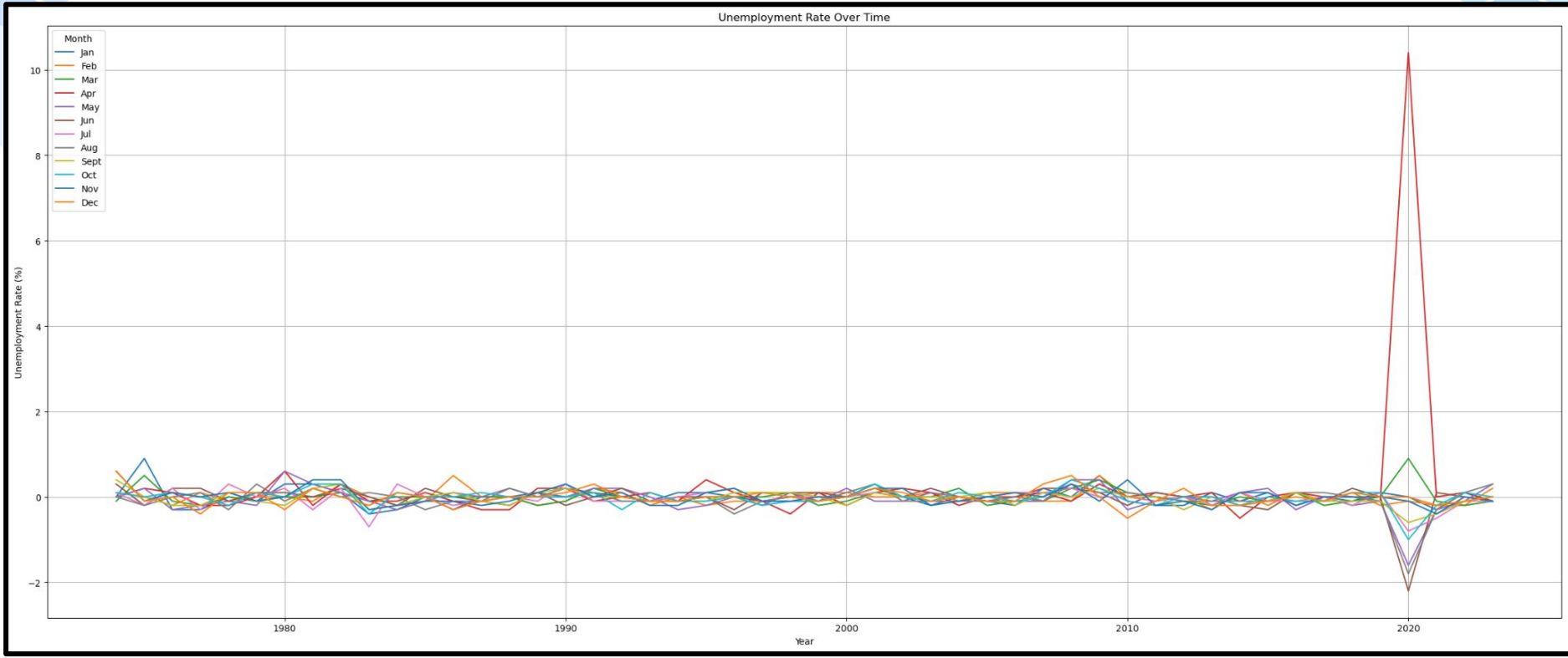
Methodology:

- Take the first order difference of each data
- Plot by month
- Plot autocorrelation and lag

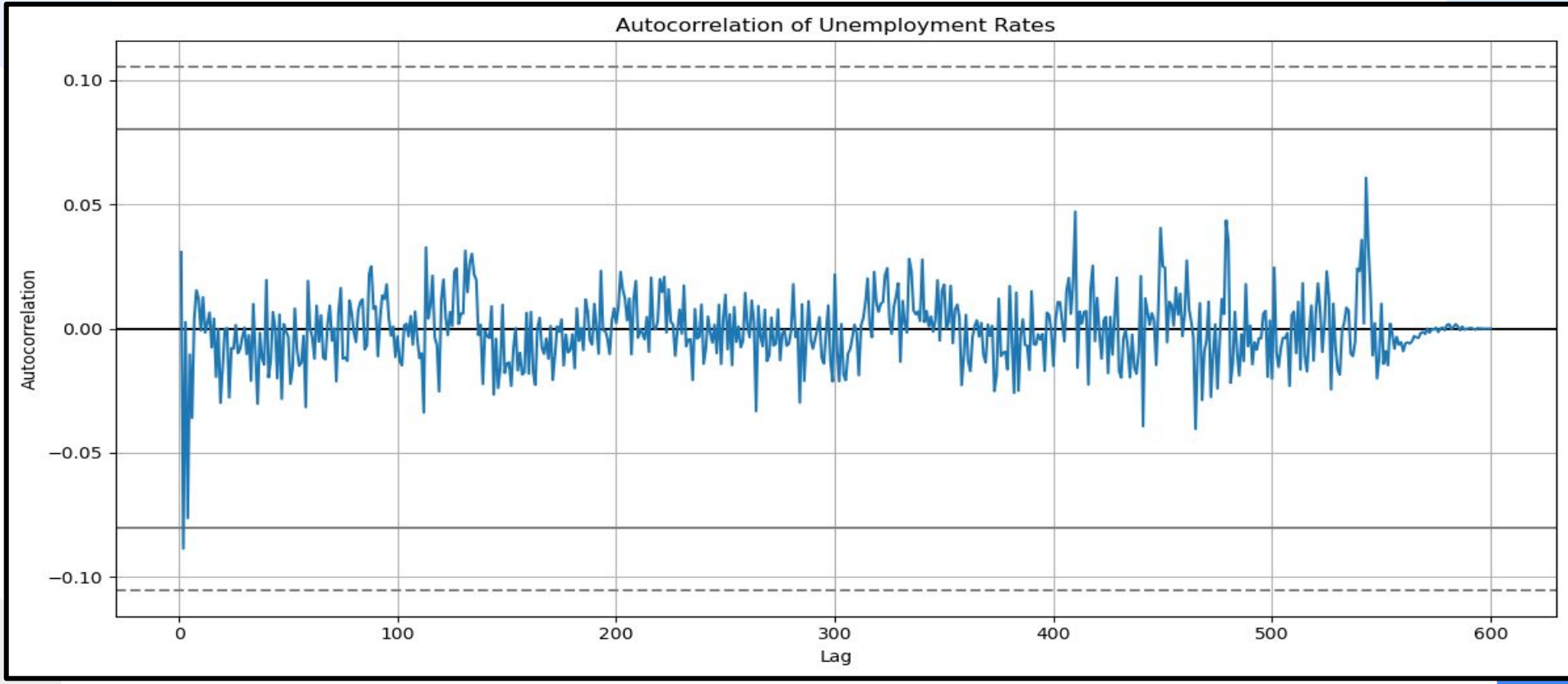
Data:

- Unemployment Rate (1974 - 2023)
- Civilian Labor Force Level (1974 - 2023)
- Employed Civilian Level (1974 - 2023)

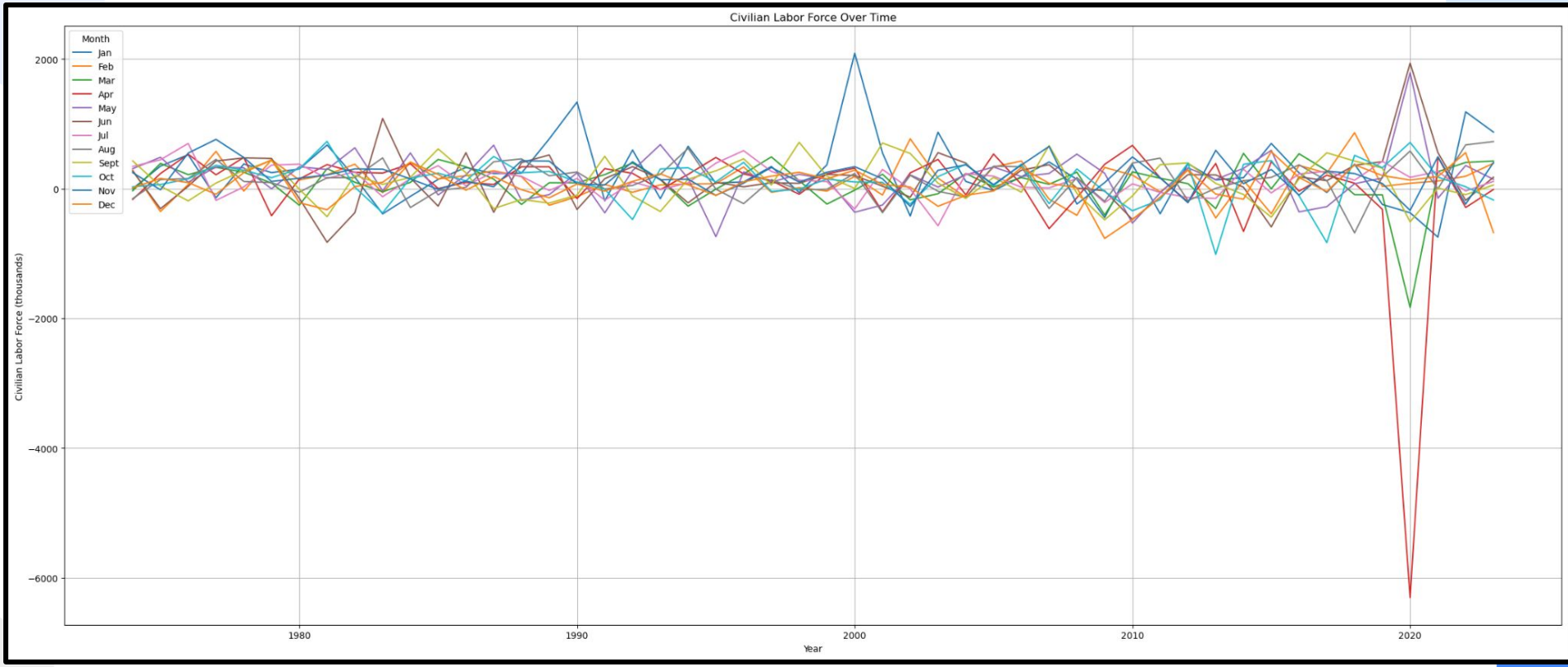
Change in Unemployment Rate



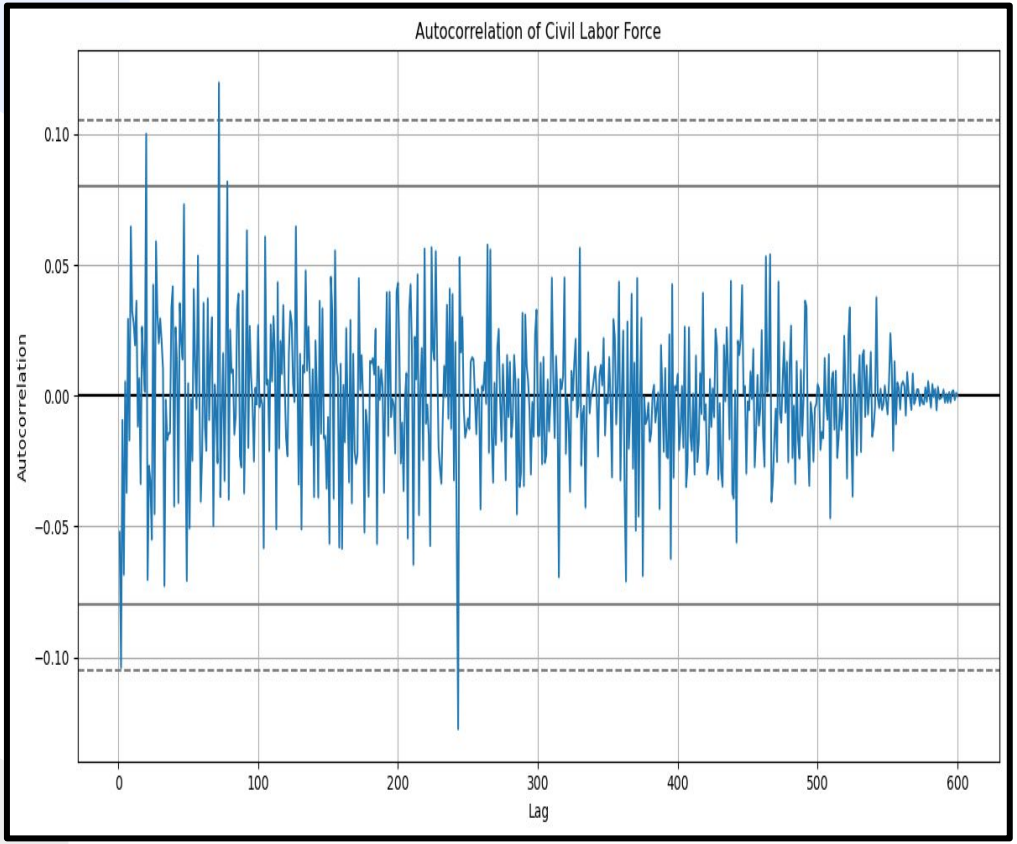
Based on the autocorrelation plot,
there is no correlation between unemployment rate and itself.

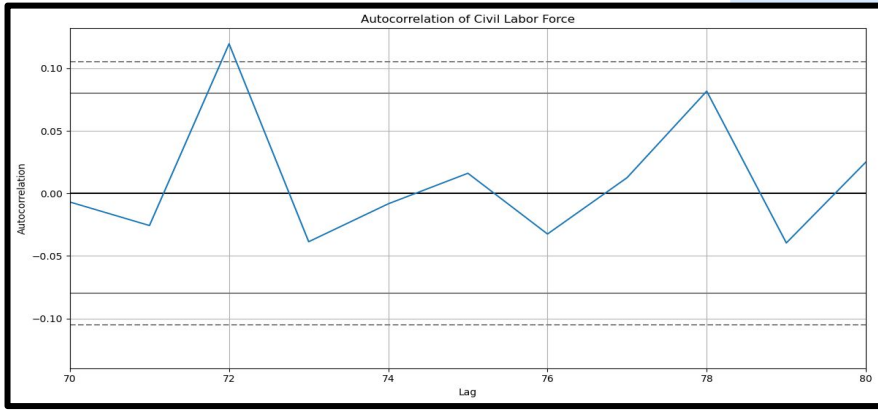


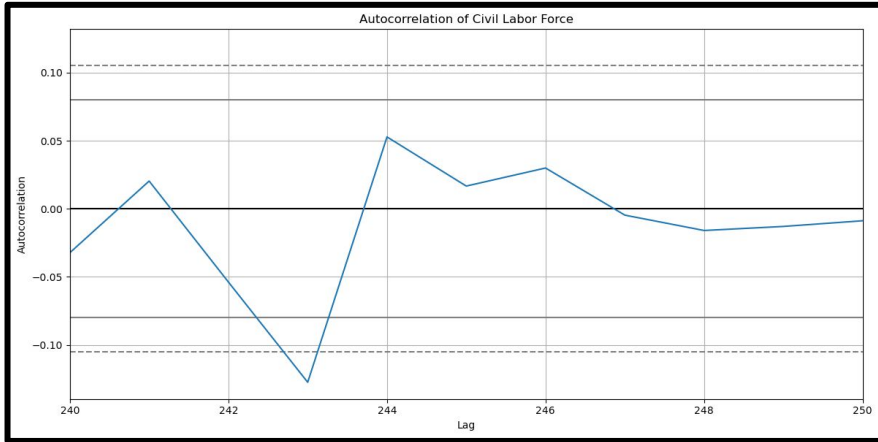
Change in Amount of Civilian Labor Force Participants



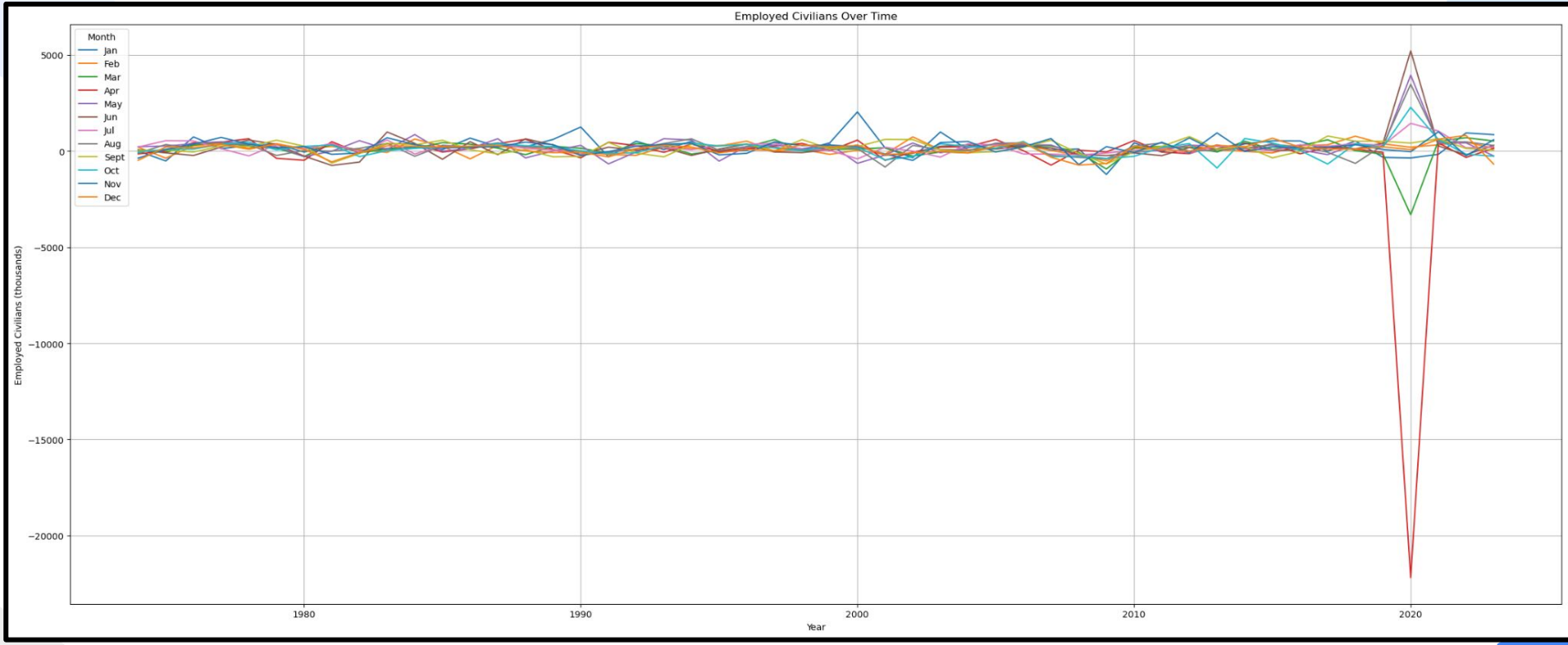
When we zoom in on the autocorrelation plot, we can see that there is a positive correlation every 72 months and a negative correlation between the change in amount of civilian labor force participants and itself every 243 months.

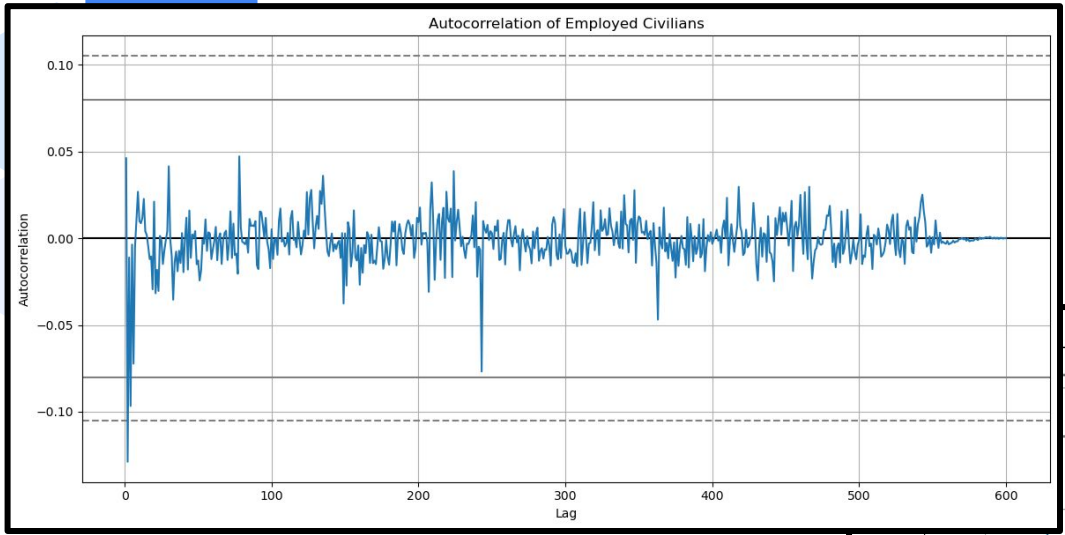


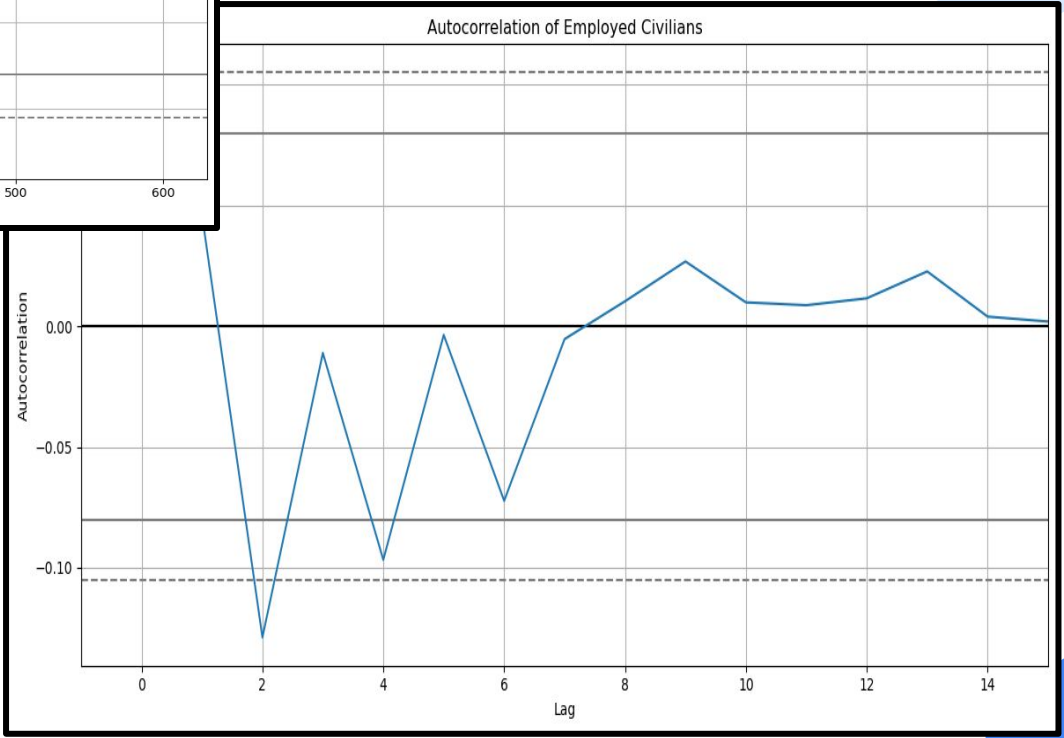




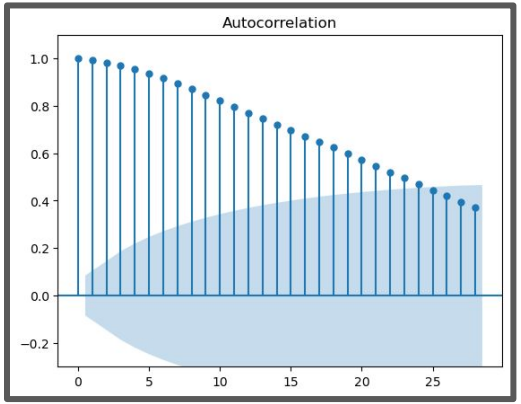
Change in Amount of Employed Civilians

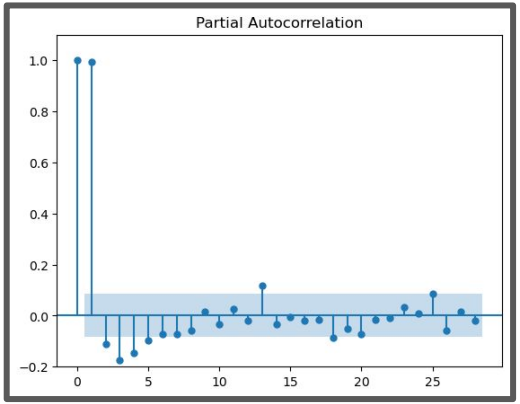




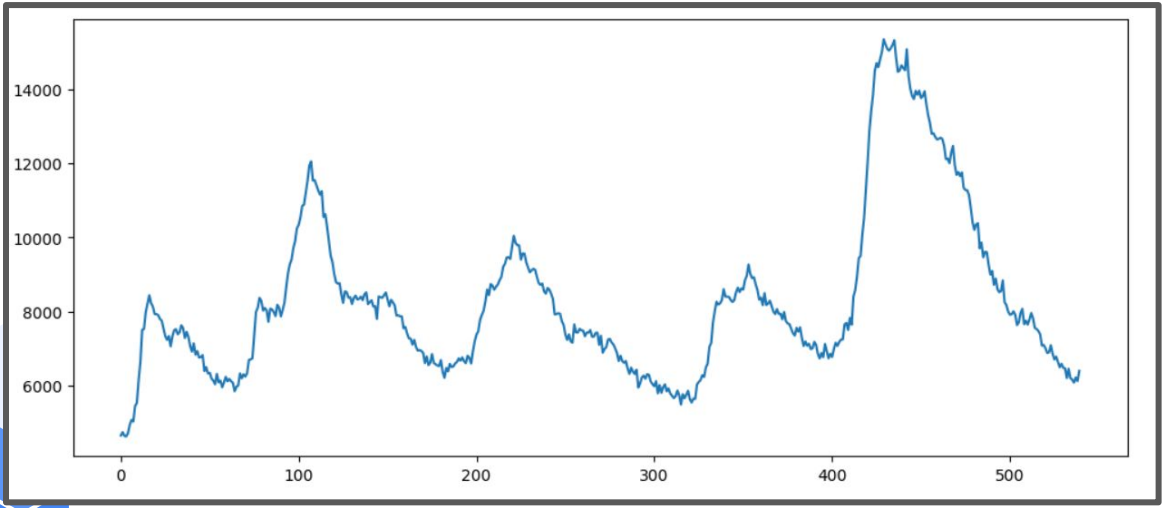


When we zoom in on the autocorrelation plot, we can see that there is a negative correlation between the change in amount of employed civilians and itself every 2 months



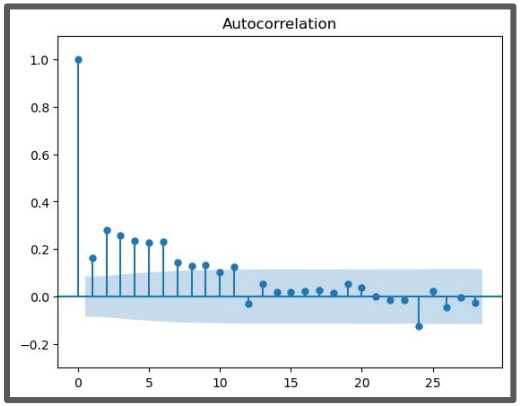


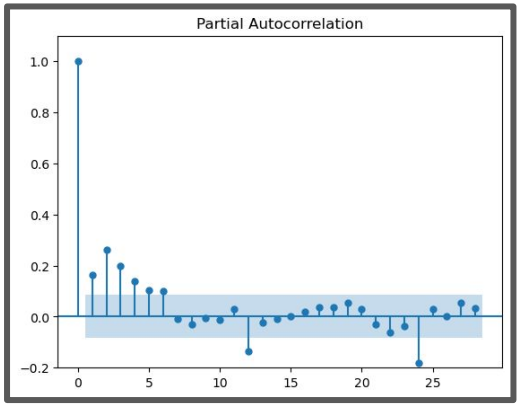

```
ADF test result:
Test Statistic                  -3.255830
p-value                          0.016972
Number of Lags Used              6.000000
Number of Observations Used    533.000000
Critical Value (1%)             -3.442678
Critical Value (5%)             -2.866978
Critical Value (10%)            -2.569666
dtype: float64
(-0.2, 1.1)
```



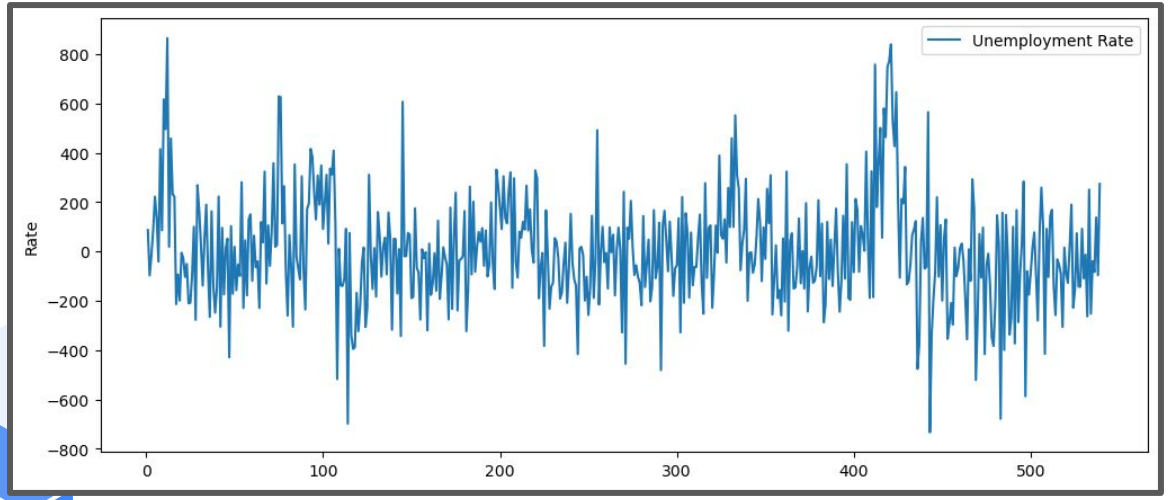
- The original data failed to pass the ADF test.
- Hence, we had to perform First- Order differencing.

Analyzing trends for Original data






```
ADF test result:
Test Statistic                  -5.168018
p-value                          0.000010
Number of Lags Used              5.000000
Number of Observations Used    533.000000
Critical Value (1%)             -3.442678
Critical Value (5%)             -2.866978
Critical Value (10%)            -2.569666
dtype: float64
(-0.2, 1.1)
```



- The differentiated data passed the ADF test. With p-value = 0.00001
- Therefore, data is now good for ARIMA training

Analyzing trends for Differentiate data

Model Fitting and ARIMA Parameter Selection

We divided the data into two segments:
90% is allocated for training, and the remaining
10% is designated for testing

P = 0
D = 1
Q = 2

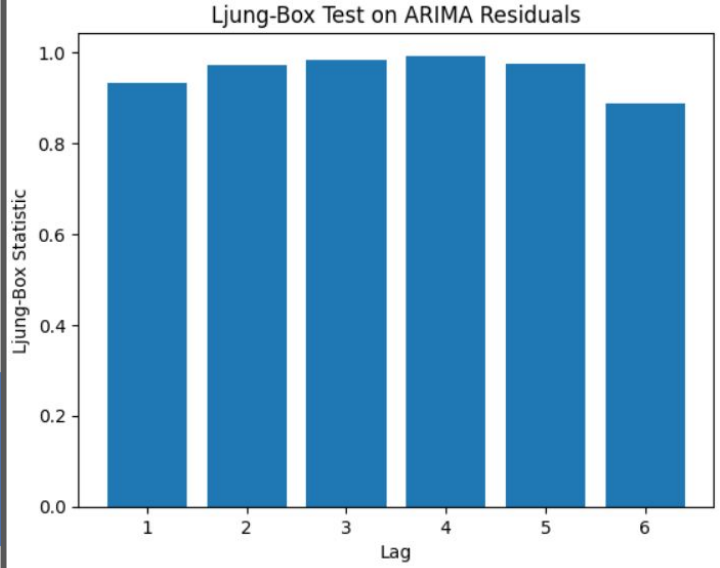
	0	1	2	3	4
0	7634.084431	7304.423878	7300.462765	[illegible]	7312.920515
1	7429.188471	7300.524471	7306.690920	7313.012658	7319.242155
2	7362.688254	7306.663328	7312.977827	7319.294861	7301.061814
3	7337.743871	7312.950544	7319.254451	7309.507196	7305.838512
4	7327.195955	7309.216846	7307.417197	7314.980747	7326.924467
5	7320.682815	7325.714899	7313.900106	7322.480149	7307.317635
6	7326.203848	7310.404968	7315.921427	7325.248899	7314.532136

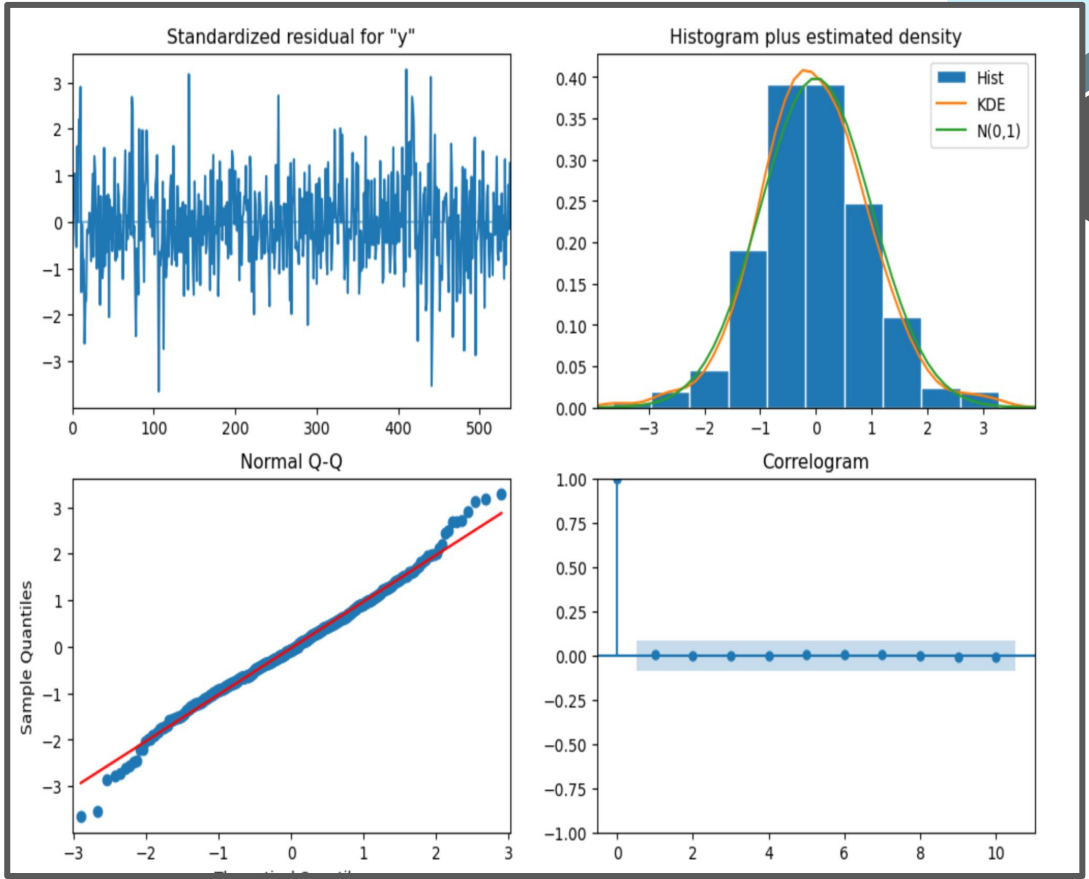
	5	6
0	7318.918172	7323.059521
1	7318.968940	7309.634910
2	7323.032606	7328.918837
3	7311.626945	7331.371962
4	7315.955210	7323.750536
5	7340.278885	7330.219276
6	7332.374020	7341.992170

Ljung-Box

P-value for Ljung-Box Test:

1	0.946438
2	0.990409
3	0.999291
4	0.999885
5	0.996276
6	0.954236

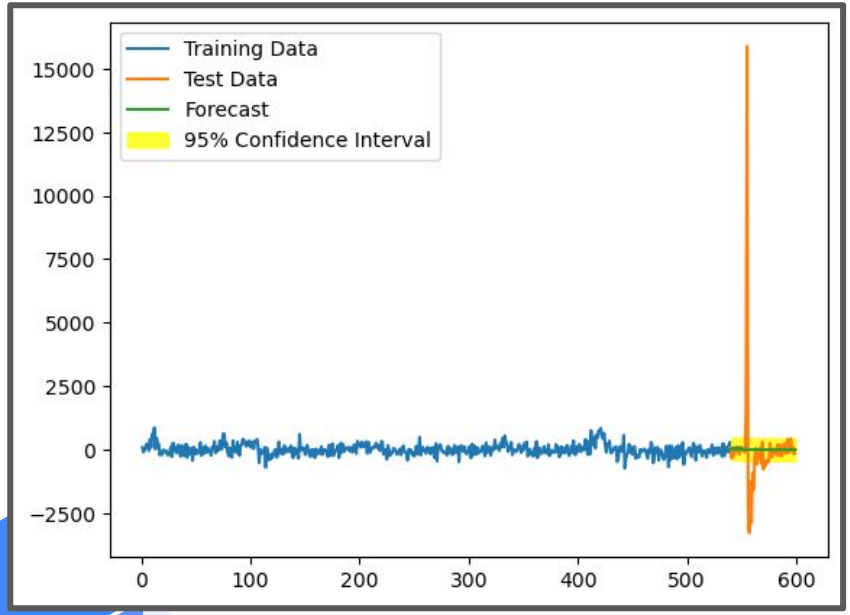


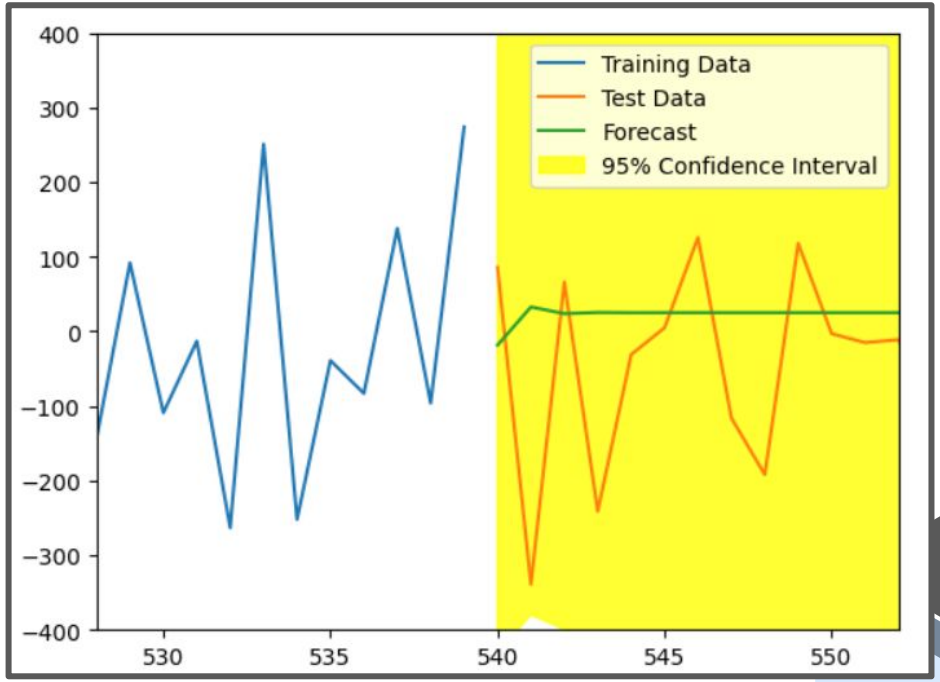


Check Residuals

Result of the Trends

```
Root Mean Squared Error (RMSE): 2187.591101690848
Mean Absolute Error (MAE): 683.7228488532103
10th Percentile AFE: -821.9133937193822
25th Percentile AFE: -339.11339371938215
50th Percentile AFE: -134.61339371938215
75th Percentile AFE: 30.63660628061784
90th Percentile AFE: 212.98660628061785
```





THANKS!