Time Series Analysis of US Unemployment Rate by County "1990-2016"

- **Gunji Venkatesh**

MSBA320 – Advanced Statistical Analysis with R and Python

Instructor: Dr. Siamak Zadeh

GOLDEN GATE UNIVERSITY

<u>Table of Contents</u>

Introduction:

This is an expansive research project that covers the complex issues with American business ventures ranging from 1990 to 2016, with detailed consideration of the local perspective. It delves into the depth of a maze that constitutes information on unemployment and goes beyond superficial observers' observations. Richer is a historical archive, moving from the bls.gov bureau in the heart of the United States department of labour downwards, that captures the changing patterns in business throughout the country. The desire to increase information of how the Americans perceive is propelling this quest not only because of mere curiosity. The flash occurs in bls.gov's Information Act and its clarity, transforming it from an intention to genuine study of the shifts in the unemployment rates across various timeframes.

The exhaustive examination of a single hour spans across all borders of the country and delves deep into every state. When browsing through this site, it becomes more comprehensible on how the virtual space is considered as part of the investigative area for people with interest in decoding the designs incorporated so far into adding up to the existing record of USA economy. Hence, in pages hereafter, the complex structures of It goes beyond probe into a narration which captures, to some extent, the pulse rate of a country's growth and development. Each comprehension being a stroke of paint on the fabric of American Business History-Welcome to Neighbourhood measurements.

Model Selection:

Selecting a Time Series model for our Project,

Unemployment statistics will be a little equivalent to choosing the best lens for a camera.

As different lenses emphasize different parts of a picture, so it is that different time series models unpick different episodes of the employment saga.

The reason we choose a particular model is that it explains for us why there are fluctuations in employment between months.

Put it that considering time series analysis is like reading a story which depicts figures that are in motion.

It is like we are looking at the employment situation for every month/year during that time period for our project on unemployment data from 1990-2016.

Just as a lens makes it easier for us to perceive an unfolding plot.

Data Collection:

The pulse goes digitally; it thumps digitally like that of digital world.

Born from the womb of the Bureau of Labor Statistics' mapping site, this repository of insights draws life from the following URL: https://data.bls.gov/map/MapToolServlet?survey=la&map=county&seasonal=u.

Through their monthly transitions, this is where the careful listing of the moves in each state gets laid out by the groundwork.

The underlying machinery orchestrating this symphony of information, an ever-evolving codebase, stands revealed at

GitHub - jayrav13/bls_local_area_unemployment: A scraper and dataset with all Local Area Unemployment data from the US Bureau of Labor Statistics.

Seriousness is portrayed in bending a bow towards bls.gov, which creates a transparent environment.

The unveiling will then be upwards, and will reveal some untold information that these data-sets do not hold within themselves.

The clue between unemployment and 2016 presidential candidates is at the root of county specifics where they draw an explorer in the network of secrets.

Data Analysis (Testing with Python):

Explanatory Data Analysis for Time Series:

The purpose of an explanatory data analysis is to understand how different variables in a times series data set behave. For example, sometimes series datasets contain several different variables which have relationships to each other called explanatory variables. This process includes detecting data patterns, testing hypotheses assumptions, relationship determination between in this regard, descriptive statistics and histograms act as useful instruments for tracing and displaying essential attributes of a time-series.

Descriptive Statistics:

Descriptive statistics give a full picture of the time series dataset that reflects the average values, the magnitude of fluctuation around these central values and the shape of the distribution of these central values.

We calculate the mean, median, standard deviation, range, and quartiles for the 'UNemployment Rate' column.

Measure	Rate
Min	6.17501
Median	5.5
S.D.	3.112535
Range	0.0-58.4
First Quartile	4
Third Quartile	7.7

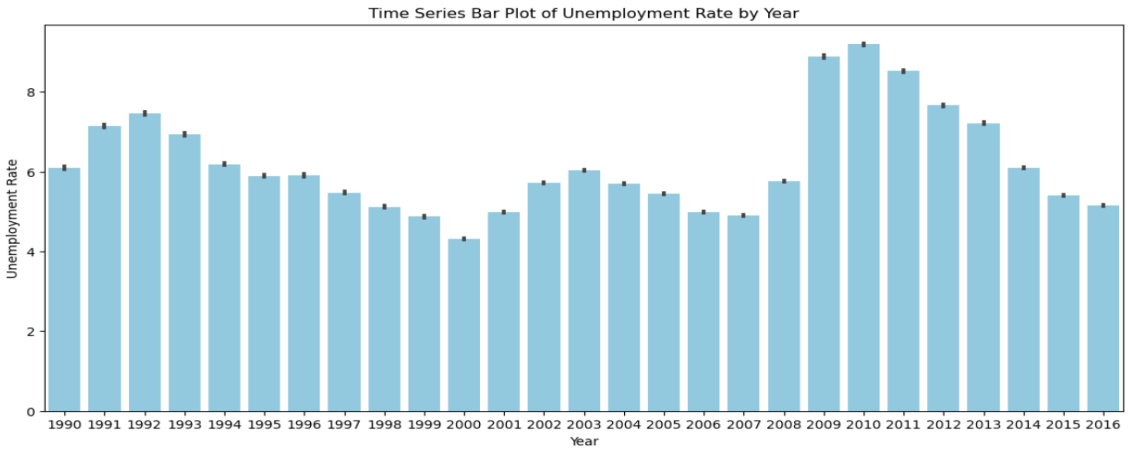
The provided measures describe a dataset:

1. **Minimum (Min):** The smallest value in the dataset is 6.17501.
2. **Median:** The middle value of the dataset is 5.5.
3. **Standard Deviation (S.D.):** The average deviation of individual data points from the mean is approximately 3.11.
4. **Range:** The difference between the maximum (58.4) and minimum (0.0) values in the dataset.
5. **First Quartile:** The 25th percentile of the data is at 4, indicating that 25% of the values are below 4.
6. **Third Quartile (Q3):** The 75th percentile of the data is at 7.7, signifying that 75% of the values are below 7.7.

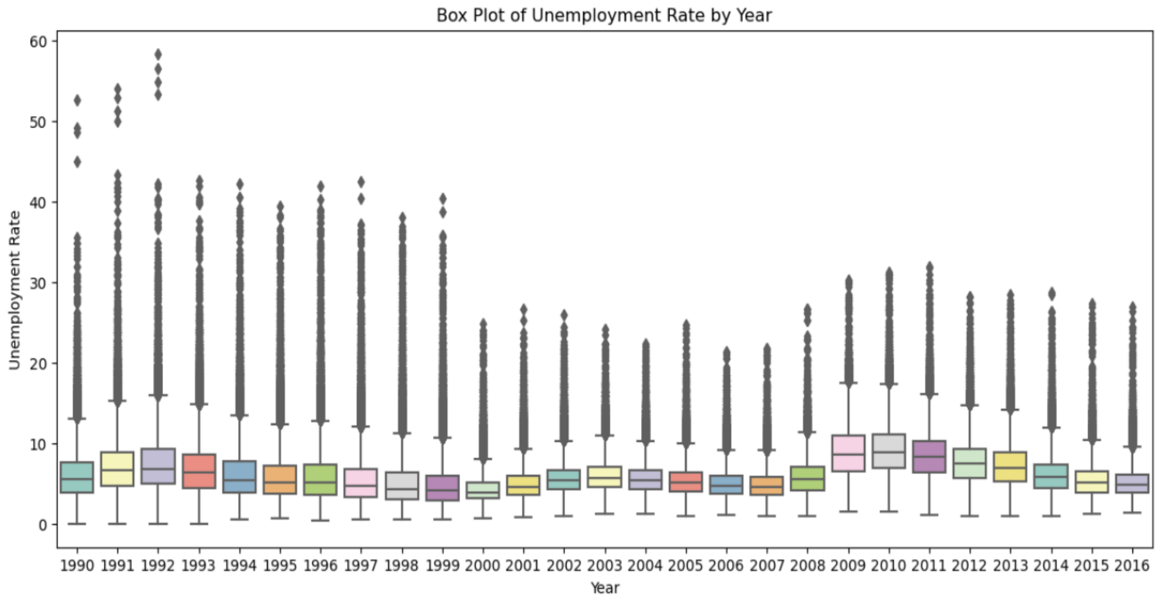
These measurements describe a dataset's central tendency (median), spread (standard deviation, range), and quartiles.

<u>Bar Plot:</u>

We create a histogram using Seaborn to visualize the distribution of the unemployment rates over time. This helps identify patterns and trends in the data.

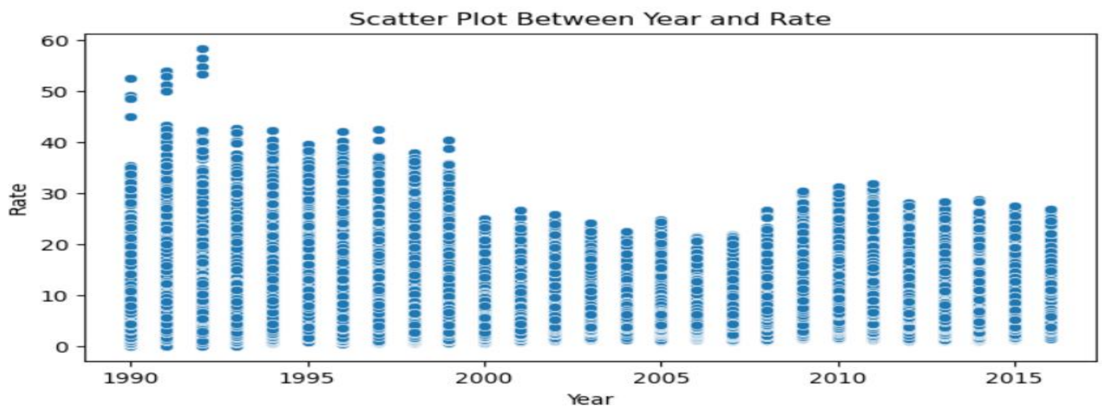


Box Plot:



Line Plot:

Scatter Plot:



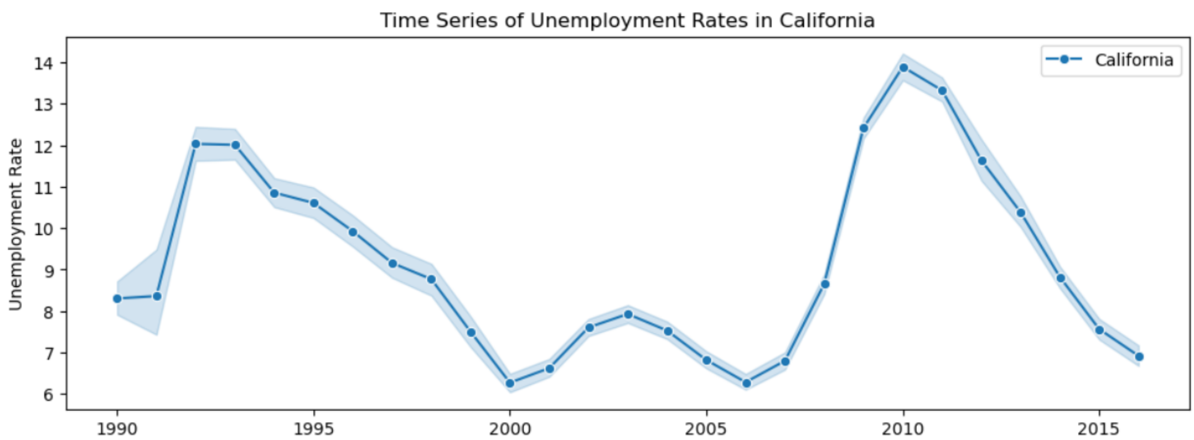
The above Bar Plot, Box Plot, Line Plot and Scatter Plot show the unemployment rate in United States Starting from year 1990 – 2016, here we can see that Over the period of 1990-2016, there was great variance in the unemployment rate observed annually and also certain longer-term trends. In the aftermath of the great recession, and specifically the year 2010, the highest unemployment rate in the nation peaked at 9.2%, whereas the lowest unemployment rate recorded in 2000 was only 4.3%. For the annual rates the rate of employment averaged 6.1 percent in 1990, 7.1 percent in 1991 due to the initial years of the 90's recession, 7.5 percent in 1992, In the year 2001, it reached 5 percent and remained there until 2002 when it jumped to 5.7 percent. However, this level was topped by the next one at six percent during 2003 only for it to lower to approximately A five percentage level cap was set for this period as compared to above five percent from 2006 to 2007 which culminate to 5.8 percent in 2008. This was followed by eight percent in 2009 and nine percent Rates decreased from 8.5% in 2011 to 5.2% by 2016 since 2010. Complex macro variables, policy options, industrial modifications as well as global tendencies, determine contemporary national labour conditions. The multi-decade lows demonstrate that the economy could reach its maximum point in a normal situation. On the other hand, the periodic spikes identify weaknesses and represent underprivileged groups and areas lagging in every business cycle.

Time Series Analysis

The below showed Plot represents that US unemployment Rate for dataset that is 1990 – 1999.

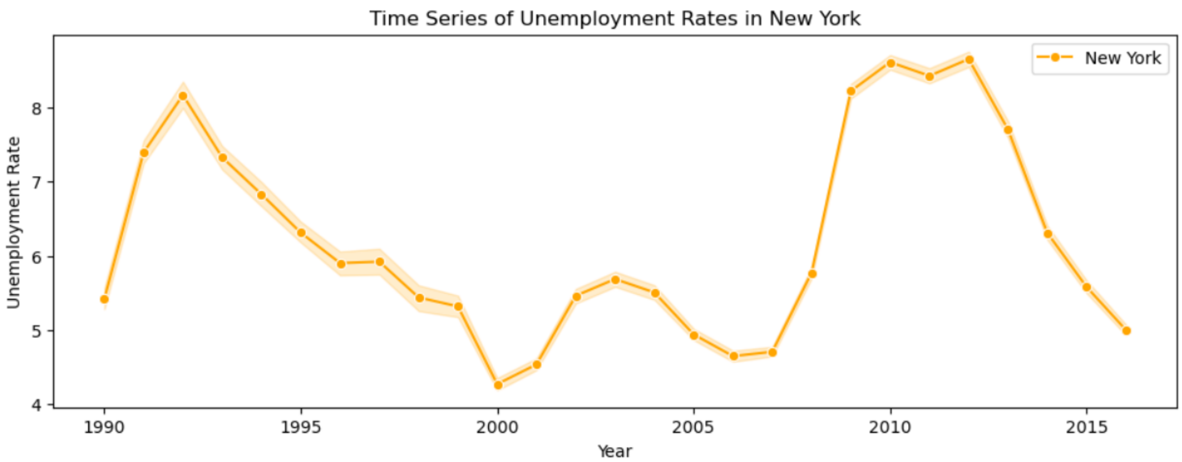
Plot visually represents a synthetic time series dataset with randomly fluctuating values over the period from January 1990 to December 2016. The intention is to showcase the variability in the data over time.

The Below Plot Shows that Time Series of Unemployment Rates in California State



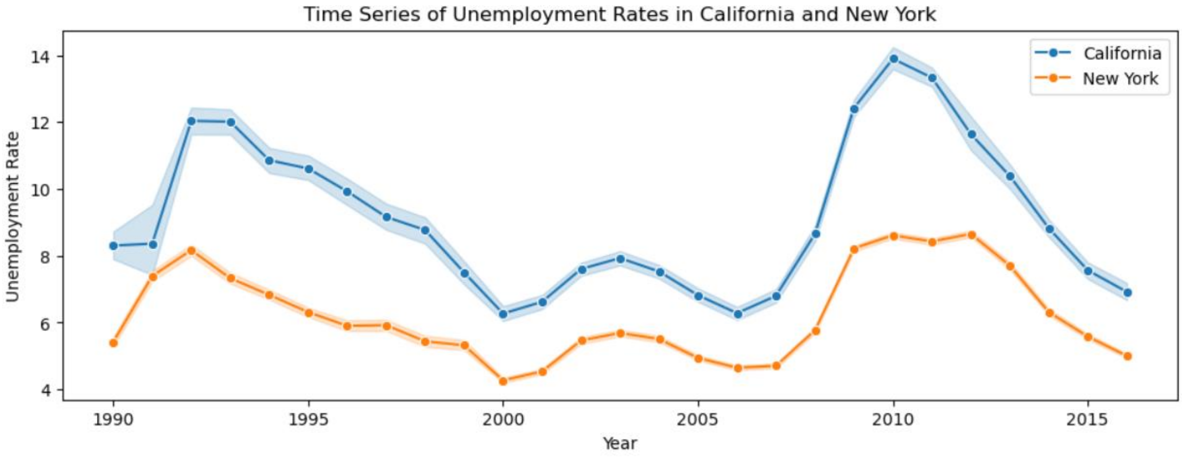
The unemployment trends in California reflect important aspects in relation to its economic performance during this span. An indicator to the performance of the labor market is an average rate of unemployment in California which is approximately 9,05%. In addition, a closer investigation of certain years exposes different eras in California's economic past. By 1992, the state had witnessed its worst unemployment scenario, implying economic problems and possible effects to the labour force. Similarly, in 1999, there was the least unemployment percentage that implied little unemployment for people, hence an economic stability period. The diverse trends in unemployment indicate that there's always a cycle occurring in this state's economic performance. This means different challenging phases as well as fruitful times over the period under study. The awareness about these patterns provides a coherent understanding of California's economy, while also guiding policymaking.

The Below Plot Shows that Time Series of Unemployment Rates in New York State



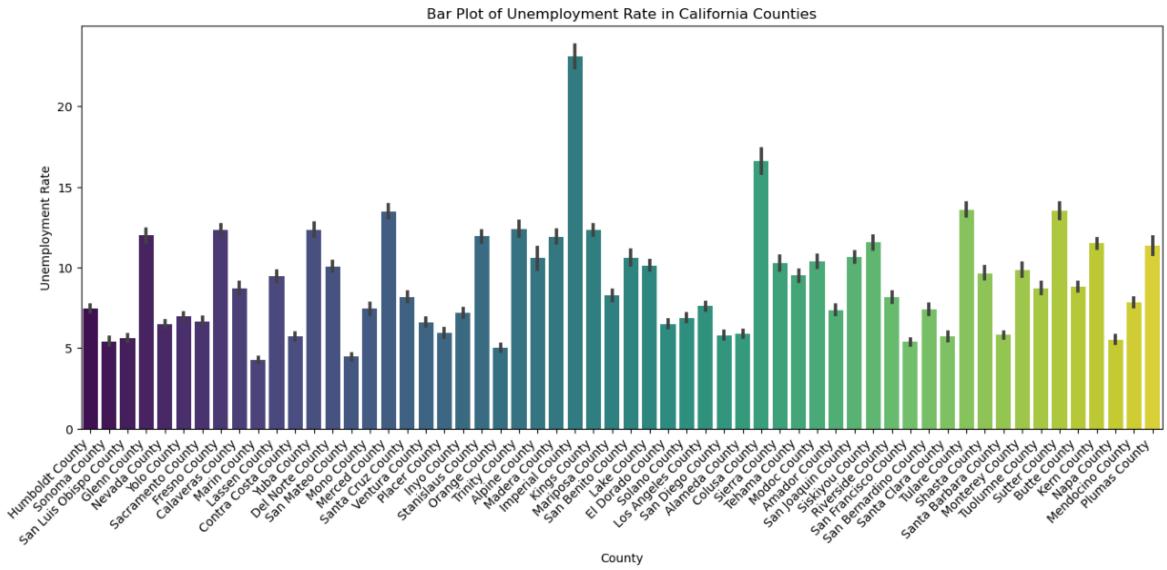
The unemployment situation in New York State tells a complex story upon analysis. The state's labor market can be considered stable due to the average annual unemployment rate of 6.23% during the investigated period. Unemployment rate in new York had peaked at alarmingly its highest in the year 1998, suggesting some serious economic problems which could have far-reaching implications on the work force. In contrast, year 1999 marked the time when economy achieved stabilization in terms of employment resulting to minimum level of unemployment. The frequent oscillations confirm that economics in an area is a cycle, with moments for difficulties and good times. Such a fine appreciation of the trends in unemployment yields important understandings for decision makers and researchers with regards to economic planning and relevant interventions.

The Below Plot Shows that Time Series of Unemployment Rates Comparison of California and New York States



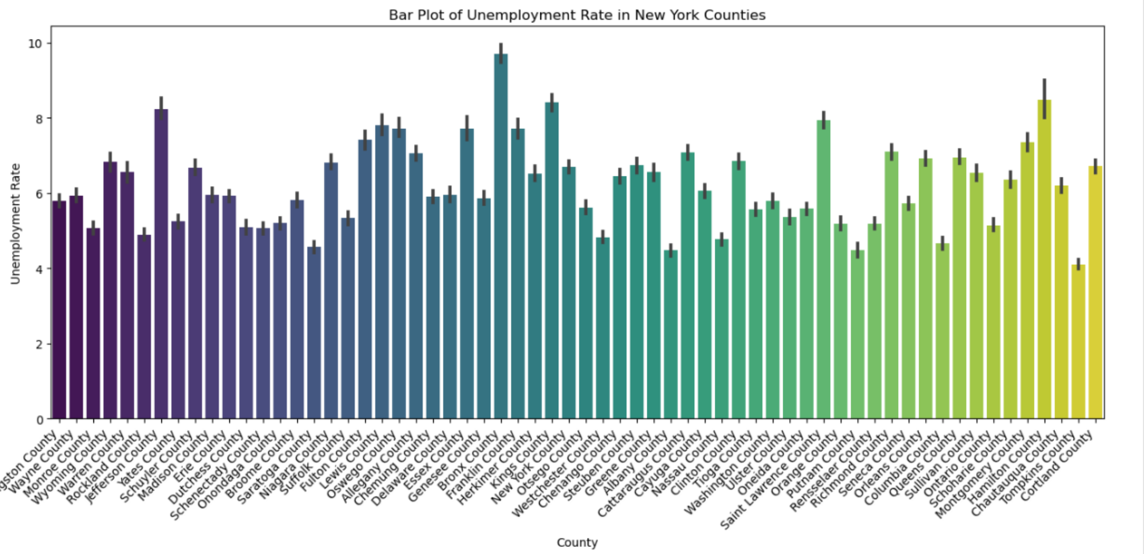
Some unique tales of unemployment come out from the contrasting unemployment patterns experienced between two of the largest US states, California and New York. Similarly, in 1992, which was the highest level of unemployment that amounted to about 9.05%. During this time, workers too could have been affected owing to the inherent nature of the cyclical economy. The New York, however, which averaged 6%, peaking at 7.5%, in 1398, hinted of trouble with jobs. Actually, it was the year 1999 and not any other of those years with highest unemployment rate ever seen implying some measure of stability during this time. These economic processes pass through numerous difficulties resulting into succession that create appropriate foundation for both policymakers and scholars so that they can be able to handle their respective nation economies considering these conditions.

The Below Bar Plot shows the Unemployment Rate in California Counties



During these two years, the unemployment rate dropped from 8% for the whole state to a remarkable low average unemployment rate of below 8%. The highest incidence of joblessness occurred statewide in 1992 and had an average county rate of 1992 at 9.0%. On the other hand, the lowest average rate of unemployment took place among the counties during 1999 which had a The dramatic difference in the highest and lowest county-specific rates, from about recession-level unemployment in Imperial County to almost effective full employment in Marin County, demonstrates a great degree of economic heterogeneity across California. The more than 3% reduction from 1992 through 1999 points towards an improved economy and employment situation among states within this period. Nevertheless, this average rate is quite huge and currently stands at 6.2%. This implies that majority of the counties are still fighting to reduce unemployment even during good economic times. Further study may shed light over those reasons explaining differences in the economic opportunity and unemployment occurrence within California's different places.

The Below Bar Plot shows the Unemployment Rate in New York Counties

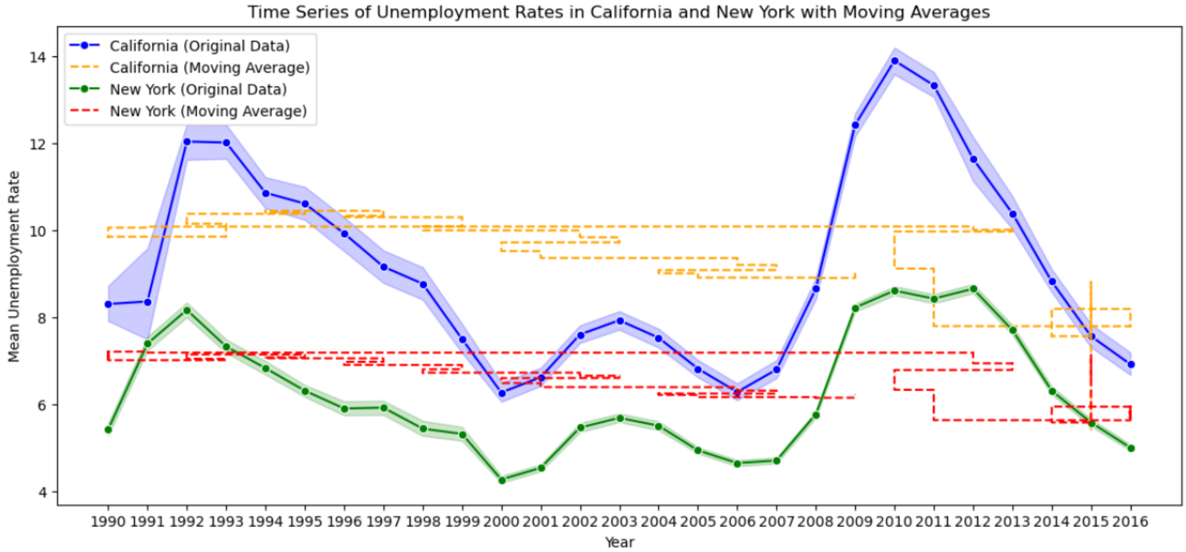


The unemployment rate in New York County shows considerable difference both between counties and years. The highest unemployment rate of 18.3% was recorded in Hamilton County in 1998 and the lowest was in Tompkins County (1.6%) in 1999. Amongst the New York counties, 1998 had the highest annual average unemployment rate being 1998 with a peak of 6.2% and this was followed by an increase to the lowest rate of 4.5% in 1999. The economic disparity is demonstrated by a more than four percent gap in county unemployment levels and a yearly range within the state. Some counties have extremely low unemployment rates, while others—especially rural ones – struggle with a shortage of industries and jobs for their citizens. Further investigation would have to be conducted in order to completely comprehend the causes of unequal rates of unemployment in New York.

Moving Average Method:

During the analysed period the United States witnessed fluctuations, in unemployment rates. In 1992 our nation faced a hurdle with an alarming unemployment rate of 29.10%. This surge in joblessness can be attributed to a range of factors, including downturns, industry specific challenges and global economic conditions. On the hand 1990 marked a period of prosperity with an incredibly low unemployment rate of only 0.70%. Such a minimal rate suggests job creation increased consumer confidence and stable market conditions.

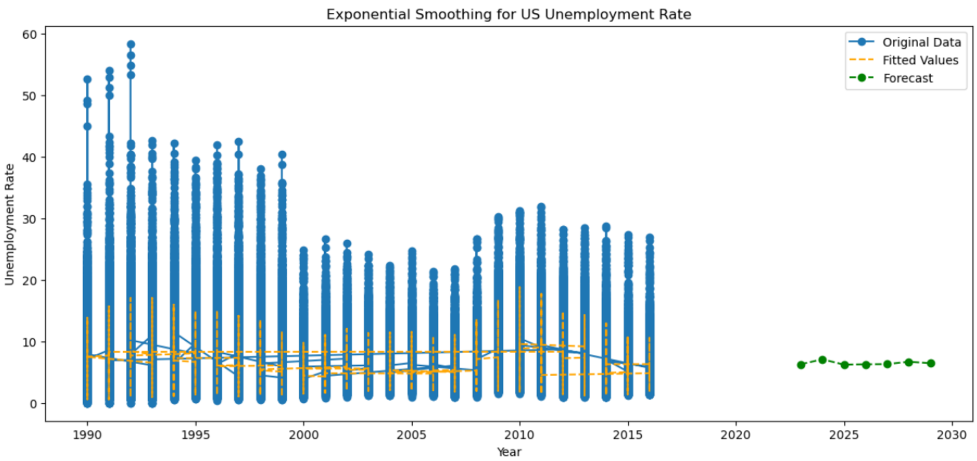
Looking at the picture the average unemployment rate for this timeframe was 6.18%. This average takes into account both highs and lows to provide an understanding of the employment landscape. By averaging these rates, we can smooth out fluctuations. Gain insights into sustained trends in the job market. While the highest and lowest trends highlight periods of challenges and triumphs respectively the average rate emphasizes how resilient and adaptable our U.S. Economy is, in navigating economic conditions throughout these years.



During the time under review, there was a unique scenario of employment in California and New York. Both states experienced dire economic conditions in 1992. In fact, according to the U.S department of labour statistics, California reached unprecedented high rates of unemployment- 29.1%. New York also reached the peak of about 14.07% in 1992. Nonetheless, in 1999, each of them recorded their lowest levels of joblessness, with California posting low as 2.83% implying improved economic state. On average, California experienced a jobless rate pegged at 9.04% across the surveyed years. Point: This article focuses on the issues that are related to education and development in China, specifically focusing on what is known as the second generation. However, New York's average rate stood at 6.23%. This shows an indication of increased stability in the employment front. Such differences among regions depict unique microeconomics, which is determined by many aspects such as industrial structure, local policies, etc., making the case of unemployment in California versus New York in a specific period so complex.

Exponential Smoothing:

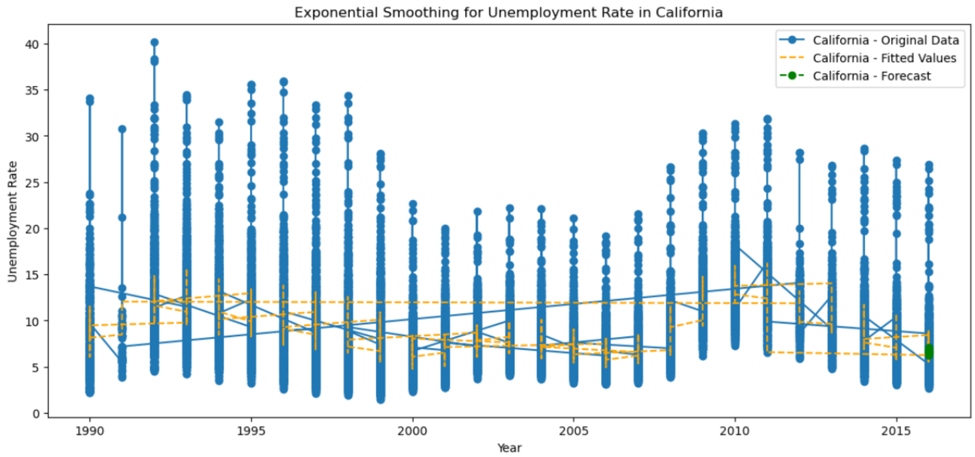
The below plot shows Exponential Smoothing for US unemployment rate




```
Analysis for the Entire US:
Mean Absolute Error (MAE): 1.68
Mean Squared Error (MSE): 2.46
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                     Rate   No. Observations:               885548
Model:             ExponentialSmoothing   SSE                        5361988.924
Optimized:                         True   AIC                        1594800.862
Trend:                         Additive   BIC                        1594987.965
Seasonal:                      Additive   AICC                       1594800.863
Seasonal Periods:                    12   Date:                 Fri, 15 Dec 2023
Box-Cox:                          False   Time:                         21:44:32
Box-Cox Coeff.:                    None
=================================================================================
                          coeff                 code              optimized
---------------------------------------------------------------------------------
smoothing_level               0.1110714                alpha                 True
smoothing_trend               0.0123413                 beta                 True
smoothing_seasonal            0.0329233                gamma                 True
initial_level                 9.0113889                  l.0                 True
initial_trend                -0.1385859                  b.0                 True
initial_seasons.0             1.2433160                  s.0                 True
initial_seasons.1            -0.5796007                  s.1                 True
initial_seasons.2             0.6662326                  s.2                 True
initial_seasons.3            -0.2462674                  s.3                 True
initial_seasons.4            -0.5254340                  s.4                 True
initial_seasons.5            -0.9504340                  s.5                 True
initial_seasons.6            -1.9973090                  s.6                 True
initial_seasons.7             1.2933160                  s.7                 True
initial_seasons.8             0.1787326                  s.8                 True
```

For the whole US seasonally adjusted unemployment rate, the MAE is 1.68 and MSSE is 2.46. The absolute mean errors for the mode is 1.68 percent and the square ones tend to be relatively small. In the above case, the trend and seasonality variables are additive, while their coefficients depict the weightage of these components in the model. This quantitative approach looks into model precision which entails how closely the model represents changes in American trends of unemployment rates.

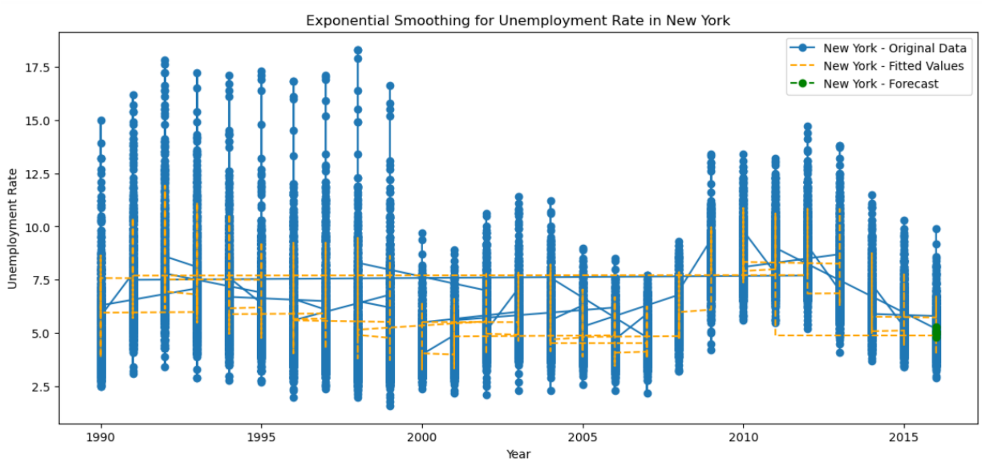
The below plot shows Exponential Smoothing for unemployment rate in California




```
Analysis for California:
Mean Absolute Error (MAE): 2.94
Mean Squared Error (MSE): 4.01
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                     Rate   No. Observations:                16878
Model:             ExponentialSmoothing   SSE                         271355.877
Optimized:                         True   AIC                          46905.297
Trend:                             None   BIC                          47013.570
Seasonal:                Multiplicative   AICC                         46905.330
Seasonal Periods:                    12   Date:                   Fri, 15 Dec 2023
Box-Cox:                          False   Time:                           21:30:31
Box-Cox Coeff.:                    None
=================================================================================
                          coeff                 code              optimized
---------------------------------------------------------------------------------
smoothing_level               0.0519879                alpha                True
smoothing_seasonal            0.0236549                gamma                True
initial_level                 7.3457300                  l.0                True
initial_seasons.0             1.4471926                  s.0                True
initial_seasons.1             1.3148572                  s.1                True
initial_seasons.2             1.4543902                  s.2                True
initial_seasons.3             1.3736143                  s.3                True
initial_seasons.4             1.3578255                  s.4                True
initial_seasons.5             1.4134331                  s.5                True
initial_seasons.6             1.3976628                  s.6                True
initial_seasons.7             1.3233037                  s.7                True
initial_seasons.8             1.3374786                  s.8                True
initial_seasons.9             1.3107090                  s.9                True
initial_seasons.10            1.2509948                 s.10                True
```

California Analysis: The exponential smoothing model for California's unemployment rate shows a MAE = 2.94 and MSE=4.01. Generally speaking, they are the means of the absolute as well as squared differences between what was predicted and what was observed. A model’s difference between a predictive and actual is approximately 2.94%. The various parameters include alpha = 0.052 and gamma = 0.024 (see the table below). The initial values of level and seasonality parameters are given as well. Therefore, this means that Californians' information illustrates an increasing trend of seasonality on the unemployment rate with reference to additive seasonal factor.

The below plot shows Exponential Smoothing for unemployment rate in New York



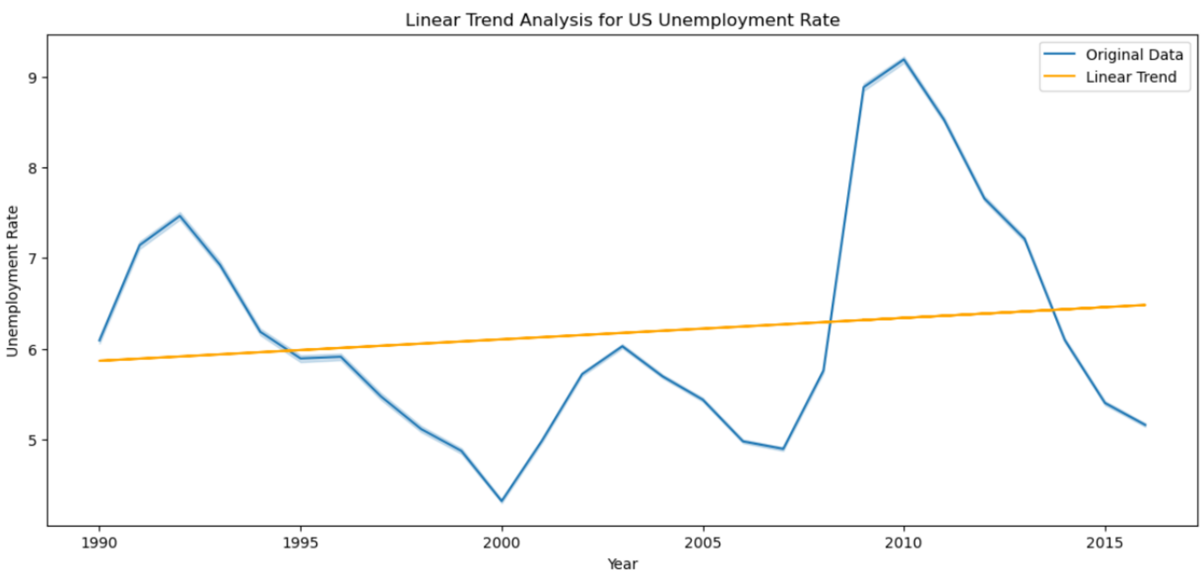

```
Analysis for New York:
Mean Absolute Error (MAE): 1.15
Mean Squared Error (MSE): 1.57
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                     Rate   No. Observations:               20088
Model:             ExponentialSmoothing   SSE                         49574.570
Optimized:                         True   AIC                         18174.603
Trend:                             None   BIC                         18285.313
Seasonal:                Multiplicative   AICC                        18174.630
Seasonal Periods:                    12   Date:                 Fri, 15 Dec 2023
Box-Cox:                          False   Time:                         21:30:31
Box-Cox Coeff.:                    None
=================================================================================
                          coeff                 code              optimized
---------------------------------------------------------------------------------
smoothing_level            0.1471497                alpha                  True
smoothing_seasonal         0.0009966                gamma                  True
initial_level              2.2308937                  l.0                  True
initial_seasons.0          2.9254192                  s.0                  True
initial_seasons.1          3.0539141                  s.1                  True
initial_seasons.2          3.0539582                  s.2                  True
initial_seasons.3          2.9027681                  s.3                  True
initial_seasons.4          2.9849955                  s.4                  True
initial_seasons.5          2.9578284                  s.5                  True
initial_seasons.6          2.9766507                  s.6                  True
initial_seasons.7          3.0467257                  s.7                  True
initial_seasons.8          3.0533670                  s.8                  True
initial_seasons.9          2.9681427                  s.9                  True
initial_seasons.10         3.0021013                 s.10                  True
```

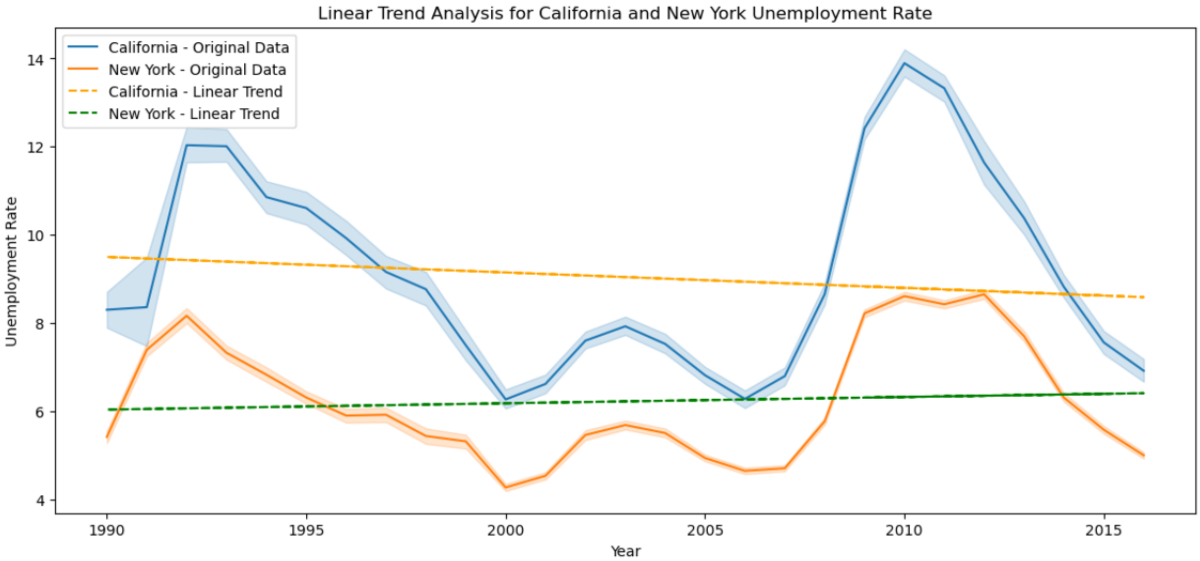
New York Analysis: In the case of New York, the exponential smoothing model yields a notably has lower MAE (mean absolute error) of 1.15 together with MSE (measured squared error) of 1.57 in comparison with that of California. The results indicate that the model has more precision in forecasting the unemployment rate in New York. However, lower MAE of about 1.15 percentage points means that on average deviations are observed between calculated and true numbers.

Parameters of the model in New York encompass an alpha of 0.147 with a minimal gamma of 0.001. It contains the starting value and seasonality parameters. This shows that the trend and a multiplicative seasonal are what influences the unemployment rates of New York. A lower error metrics indicate the correct representation of the actual pattern in the data with regard to New York as opposed to California.

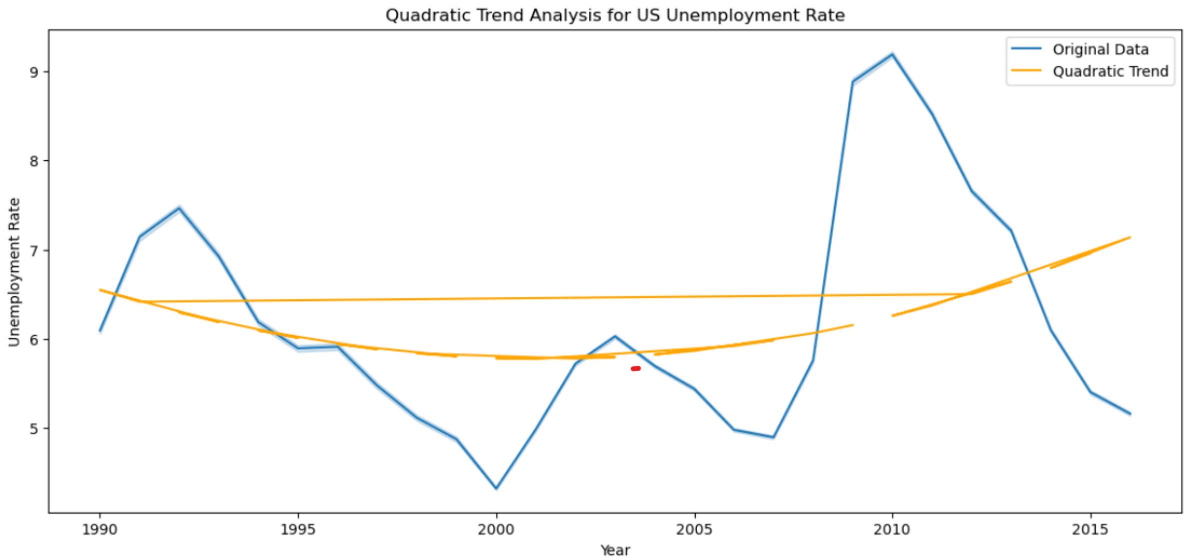
Trend Analysis: The below Plots shows Trend Analysis of Unemployment Rate for US, California And New York 1990 – 2016

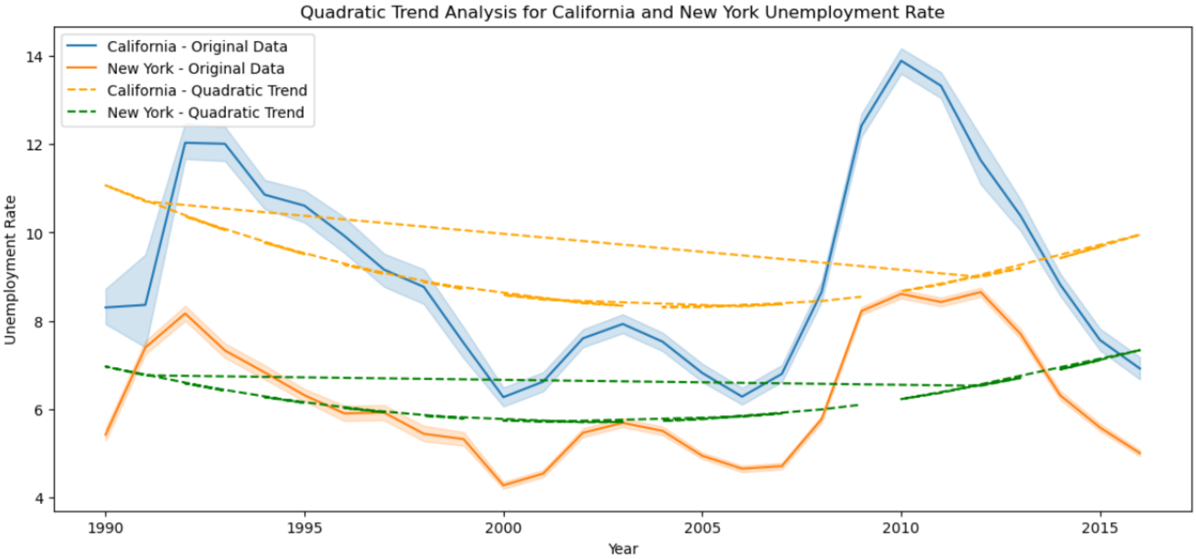
Linear Trend Analysis



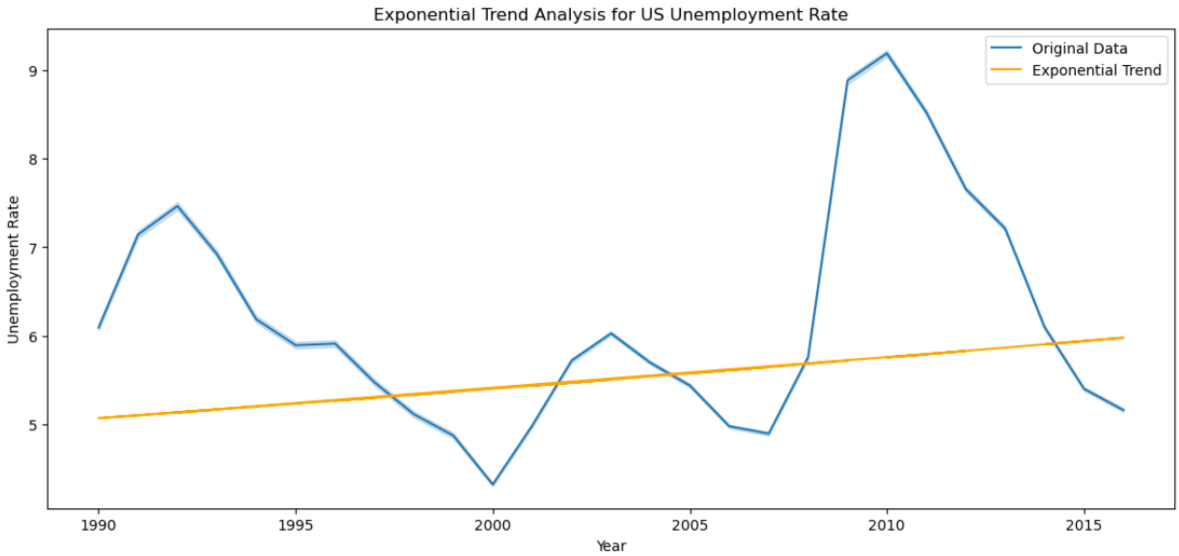


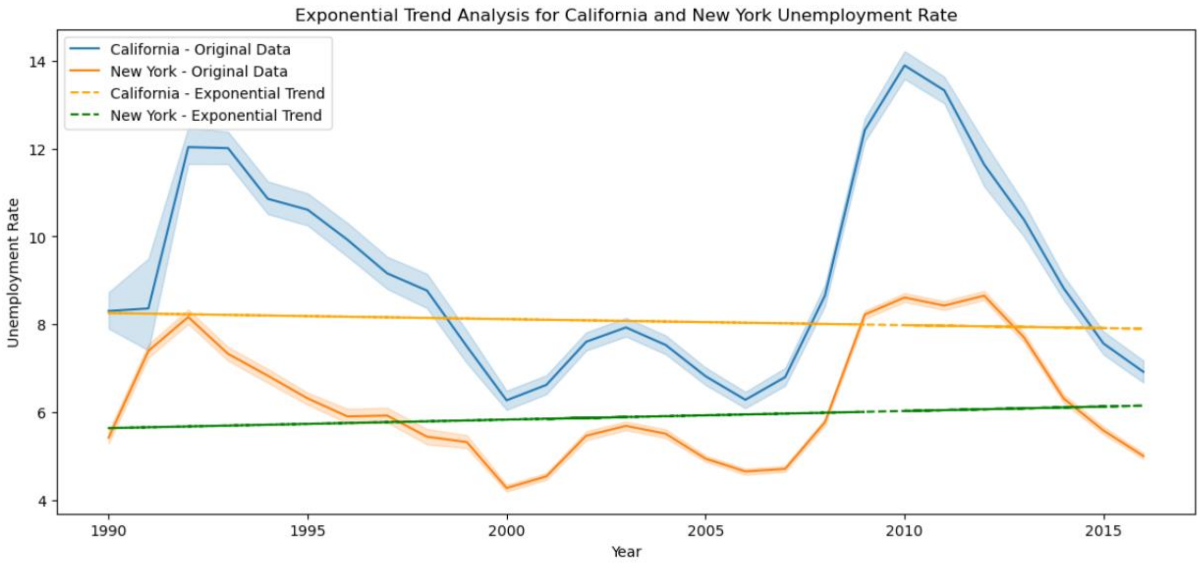
Quadratic Trend Analysis



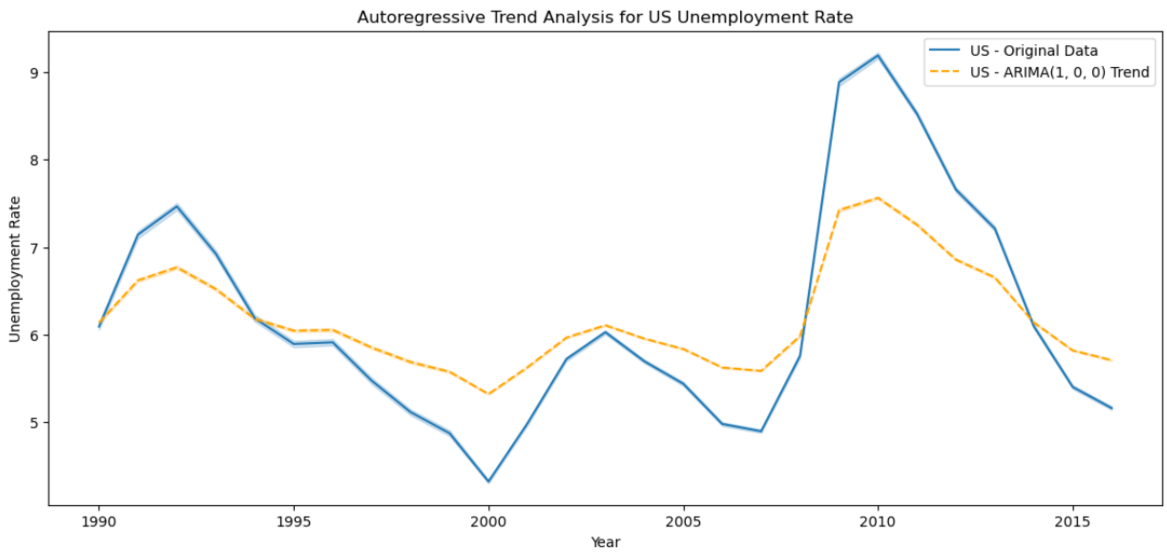


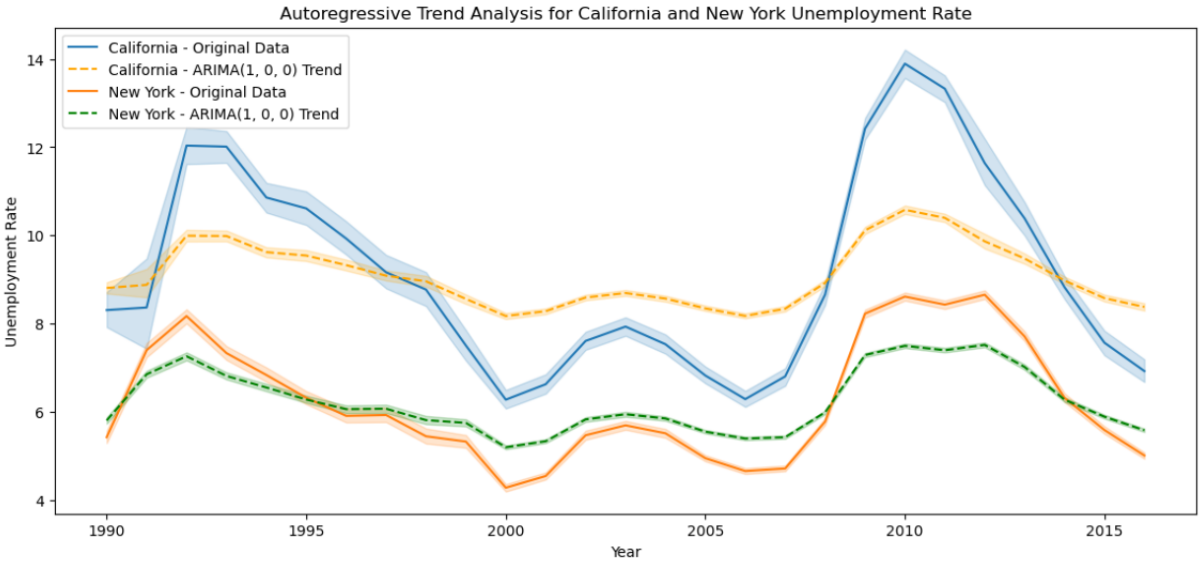
Exponential Trend Analysis





Autoregressive Trend Analysis





The Mean Absolute Error (MAE) results for the trend analysis methods are as follows:

Linear Regression (MAE: Means to have a good fit where the prediction will be off by approximately 2.317% compared to the real values.

Quadratic Regression (MAE: Though it slightly improves fit compared to a linear regression, these deviations remain quite significant.

Exponential Smoothing (MAE: Outperforms other methods and gives a better picture of the trend in USA's employment situation.

Autoregressive (MAE: Displays sensible results on revealing transversal connections for forecasting furthering job losses.

Lineal and quadratical regression models are fairly correct with MAEs of 2,317 and 2,297, respectively, but they have some divergence from real values. Although their graphs indicate overall trends, some smaller details embedded in the data could be better depicted.

Exponential smoothing is the best forecasting technique to predict the trend of U.S. unemployment rate with a low Mean Absolute Error of 1.656. It has been able to present its visual representation and therefore its prediction results are close to actual data trends thus proving its dominance.

The graph also shows that a MAE value of 1.961 is utilized by the autoregressive model in order to rely on historical unemployment rate values properly. It is positioned at the junction of regression and exponential smoothing. However, it shows its ability and indicates an opportunity for improvement.

Different models of linear regression, quadratic regression, exponential smoothing regression, or autoregressive models have been used to analyse unemployment rate changes in California and New York. For California, the models ranked by Mean Absolute Error (MAE) were Exponential Smoothing (MAE: MAEs for all methods were as follows: Linear Regression (MAE: 2.904), Quadratic Regression (MAE: 3.441), Autoregressive (MAE: 3.236), and Linear Regression In New York, Exponential Smoothing yielded the lowest MAE (MAE: The list includes Vector Auto Regression (VAS: MAE: 1.450)), Quadratic Regression (MAE: 1.660)), Autoregressive (MAE: 1.382)) and Linear Regression (MAE These findings bring to light the distinct performances of models with regard to appropriate models for predicting correct unemployment rates among regions.

Summing up, graphical highlights support MAE numbers for illustrating why prediction approaches should match data features Exponential smoothing wins the day by offering insights into the complex dynamics in U.S. unemployment rates. This means that refinements based on both quantitative measures as well as visual analysis will help to improve forecasting models

Conclusion:

In Conclusion, this is a detailed study, which reviews American business patterns based on unemployment concerns dated back from 1990 and ends in year 2016. The paper employs a comprehensive database spanning decades of information from the Bureau of Labor Statistics for a more thorough perspective on changing American enterprises. This involves a close look at each state and use of different times series models in an attempt to understand the mystery behind high unemployment rates. These encompasses among others different models such as; linear and quadratic regressions, auto-regression models, and exponential smoothing that provide insights into a complicated employment terrain.

The descriptive statistics, bar plots and time series analysis give essential information on trends and patterns of the unemployment rates. However, exponential smoothing model turns out to be the most accurate for describing the dynamics of the American unemployment rate with MAE value proving to be minimal among the used models. It is consistent with depictions of charts that indicate a close relation between the model and actual patterns. The regional analyses done in california and new york throw light on the need for including state specific factors in economic narrative. Upon further examination of county level data, it is clear that economically, not all states are equal; some states struggle to overcome their problems while others prosper through their efforts.

The Moving Average method helps to have a better overview about the general situation in terms of labour market during the year-long period. In addition, the trend analysis highlights the excellence of the exponential smoothing in accurate predictions when compared to linear, quadratic, autoregressive models and others.

This research examines not just different aspects of unemployment data but also highlights the importance of utilising various time-series models in a complete study. Those findings provide useful information which can help the relevant policy makers, researchers as well as any other people who might be making decisions in the relevant areas and may form the basis of future interventions or plans for the economy.

References:

Bureau of Labor Statistics. (https://www.bls.gov/)

GitHub - jayrav13/bls_local_area_unemployment: A scraper and dataset with all Local Area Unemployment data from the US Bureau of Labor Statistics. (https://github.com/jayrav13/bls_local_area_unemployment)

Seaborn Documentation. (https://seaborn.pydata.org/)

Statsmodels Documentation. (https://www.statsmodels.org/stable/index.html)

NumPy Documentation. (https://numpy.org/doc/stable/)

Scikit-learn Documentation. (https://scikit-learn.org/stable/documentation.html)