Assignment - 3

Divya, Nupur, Ranadheer, Varun

Data Description:

The dataset consists of structured data with 9 parameters (i.e., columns) and 708 entries (i.e. rows). Data talks about the patterns of recession in the US economy from 1959 to 2017. Various significant parameters are taken into consideration in this regard that are as follows:

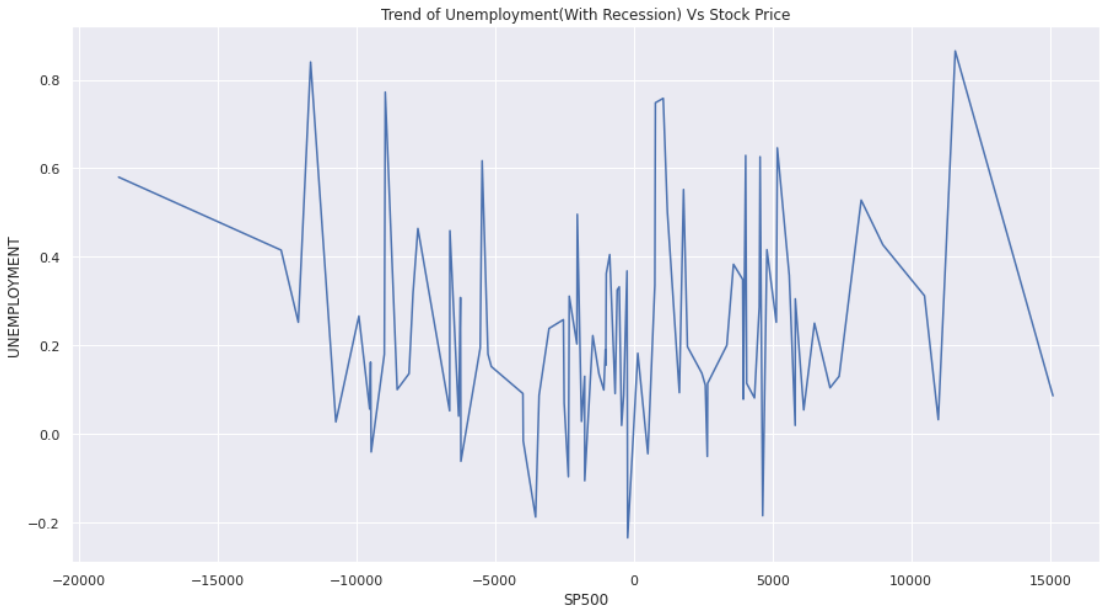
1. *Consumer Price Index (CPI)* – It is the mathematical measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
2. *Federal Funds Effective Rate (FEDFUNDS)* -Another mathematical volume-weighted median rate of overnight federal funds transactions reported in the FR 2420 Report of Selected Money Market Rates.
3. *Industrial Production: Total Index (INDPRO)* – Economic indicator that measures real output for all facilities located in the US manufacturing, mining, and electric, and gas utilities and other industrial sectors.
4. *Total Nonfarm Payroll (PAYEMS)*- It is mathematical measure of under non-farm activities which the sum of all payments and benefits paid or provided to or for the benefit of an employee.
5. *Yield curve (YC)* - Yield curve, in economics and finance, a curve that shows the interest rate associated with different contract lengths for a particular debt instrument (e.g., a treasury bill). It summarizes the relationship between the term (time to maturity) of the debt and the interest rate (yield) associated with that particular term.

Data Cleaning:

Year has been extracted from the date and placed into a column of its own in order to assess the changes in the pattern on a yearly basis. Column name ‘DATE’ has been capitalized for improved consistency across the data set. The position of the decimal point in the columns ‘CPI’, ‘SP500’, ‘FEDFUNDS’, ‘INDPRO’ and ‘YC’ has been shifted by a few places for a better ease of understanding.

Understanding the Visualizations:

1. Line Plot (Chart)-

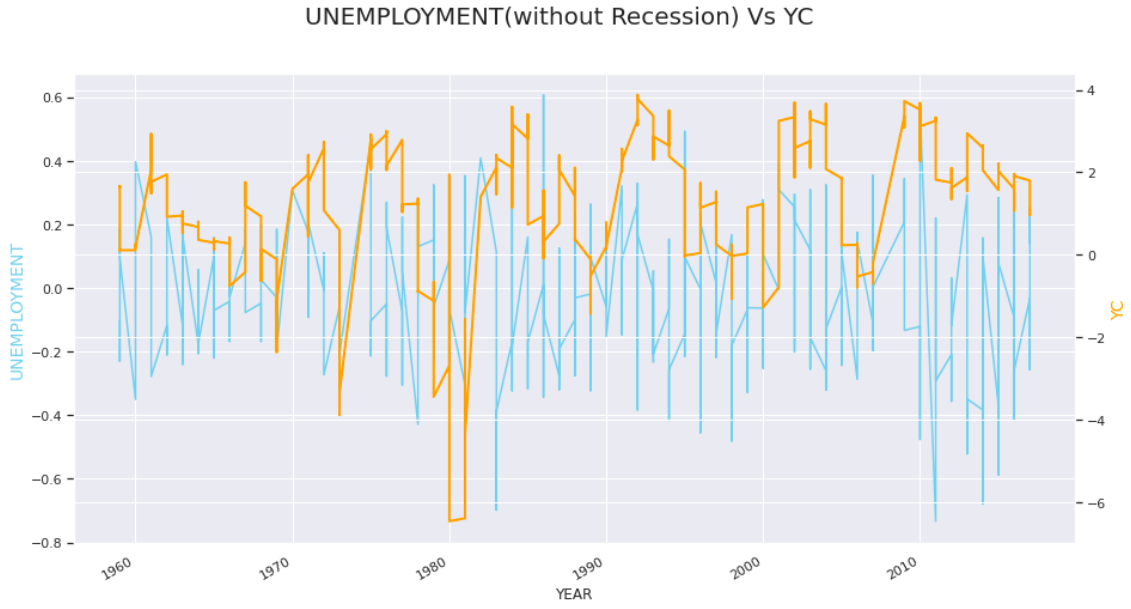


The above-mentioned two-Line plots are bifurcated on the basis of recession, in which the trend of Unemployment rate over the Stock Price Index in past 5 decades with recession and without recession is visualized. Former shows random spikes with regular fluctuations while the later is more densely fluctuated with irregularities.

2. Area Chart-

The above-mentioned shaded Line Plot are evident enough to understand the change in unemployment rate from 1960 to 2017 with and without recession in USA. While the former shows stability and a natural trend of increase and decrease in unemployment rate, the later is differing at regular intervals.

3. Dual Y-axis Chart-

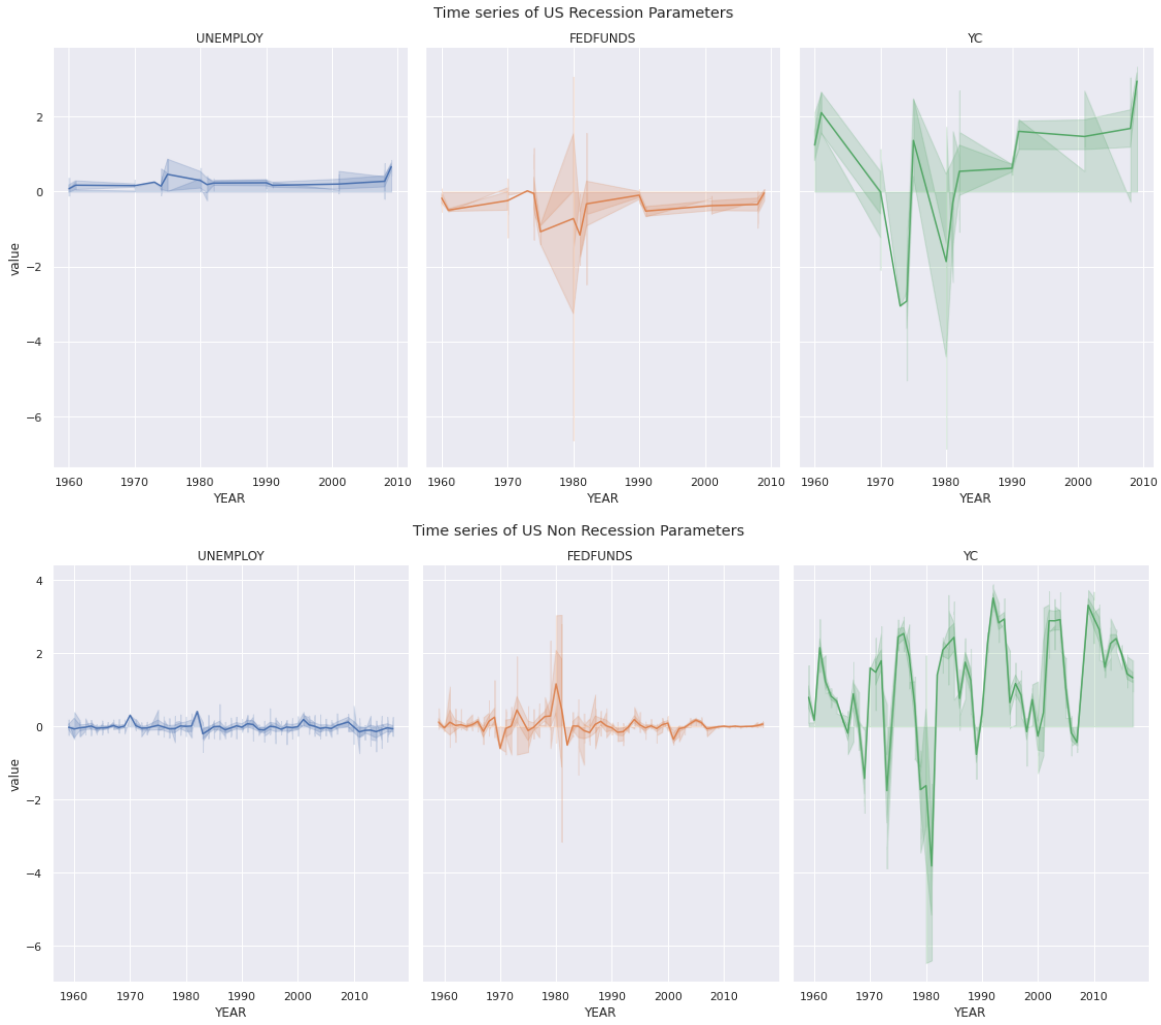


The above-mentioned charts are specifically Dual Y-axis plot having Unemployment rate on first y-axis while Yield Curve on second y-axis iterating over years on common x-axis with and without recession in USA. Taking insights on with recession plot there is a clear margin between YC and rate while without recession it was found that the data was dipping into each at several points.

4. Stacked Area Chart-

This Stack area charts are displaying the evolution of the numerical data that is Unemployment Rate with Yield Curve for US Recession as well as US Non-Recession Dataset over the years. The Yield Curve is stacked over the Unemployment rate, and is dominantly shaded over the axis x-axis while the rate has acquired some notable patches within the chart under YC.

5. Multi-Plot Grid (FacetGrid)-

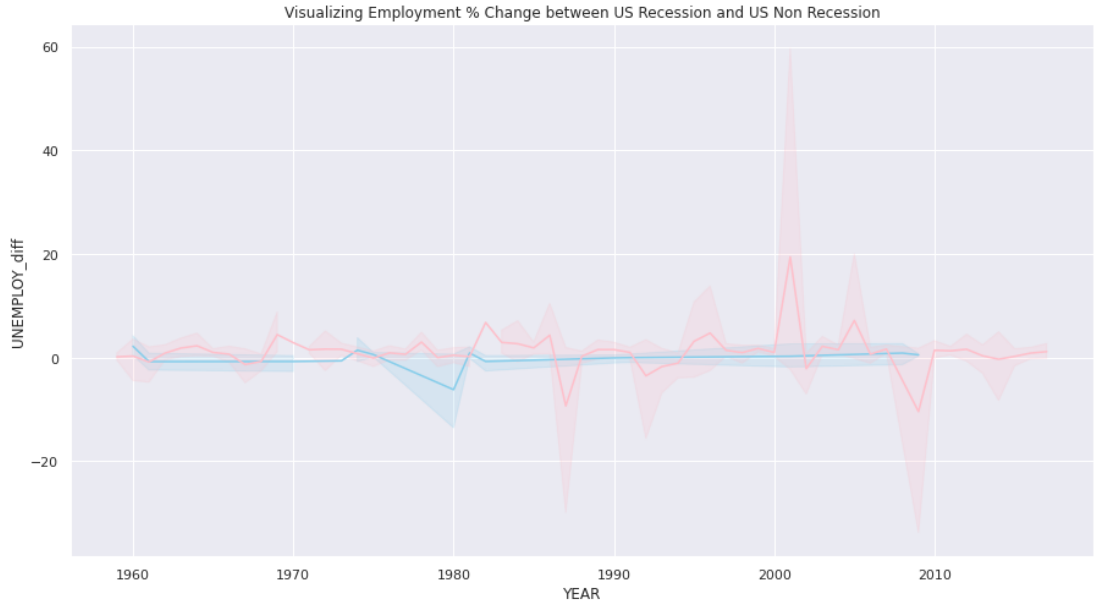


The above mentioned are FacetGrid Plot specifically from the Multi-Grid Plot under the time series in which we can simultaneously visualize multiple Parameters. Like-wise here we have visualized the unemployment rate, federal fund rate and yield curve yearly over the constraints of being recessive and non-recessive data in USA. The patterns are dense in non-recession while sparse in the recessive dataset.

6. Rolling Average with Line Chart-

The above-mentioned Line chart has three lines with initial line of color ‘skyblue’ being given Consumer Price Index, while the second line of color ‘red’ is the rolled average Consumer Price Index at a window of numerical 10, while the last line of color ‘green’ is again rolled average Consumer Price Index at a window of numerical 20. Plotting them altogether simultaneously shows the variation against the mean with specified window from original pattern over the past few years.

7. Normalizing with periodical change in Line Plot-



So, applying normalizing on the ground of percentage change on employment rate taking both recessive with skyblue color and non-recessive with pink color lines are visualized. While the rate was stable in recession, non-recession has some significant fluctuations.

So, applying normalizing on the ground of shift change on industrial production taking both recessive with skyblue color and non-recessive with pink color lines are visualized. While the production had few spikes in recession period, non-recession period had regular undulation in industry specific productions.

8. Decomposing into components-

Decomposing the plots into individual components under time series is a very vital visualization. As we can understand the unemployment rate under the grounds of level,

trend, seasonal and raised functionality in one go. It shows multi-dimensional view of the unemployment rate for getting profound knowledge of its trends.

9. **Other Visualizations-**

Other Visualizations include Scatter Plot of Unemployment rate with and without recession over the past few years, also the Scatter plot of Unemployment rate against Stock Price. Yearly trend of Stock Price with and without recession on Line-Plot, Shaded Line Plot of Yearly Federal Fund rate with and without recession, First Clubbed Box-plot for Unemployment rate, Total Nonfarm Payroll and Stock Price Index, Second Clubbed Box-plot for Unemployment rate, Federal Fund rate and Industrial Production on US Recession Dataset, Lineplot (lmplot) that is two-dimensional plot on unemployment rate against Total Non-farm payroll and Bar plot for year-wise un employment rate.

Conclusion-

US Recession dataset was the comprehensive application of the Time-Series. Based on the Recession many visualizations were created like Line Plot, Dual Y-axis, Multi-Plot etc. for understanding the variation in data with and without recession, few plots significantly compare the trend while a few interprets the insights from both recession and non-recession data collectively. Overall the terminologies and concepts of time series are perfectly blended in this Dataset giving profound knowledge and application of plots as well.