The primary objectives of this study are to explore the relationship between an aging population, taxation, and unemployment, and to provide insights into the formulation of social security policies under these conditions.

Using the domain knowledge of the economics, I try to make an educated guess and generate hypotheses before constructing a machine learning model.

According to the basics of unemployment theory, unemployment rate is directedly related to Consumer Price Index, Deflator, GDP growth rate per captita, Labor Participation Rate etc.

Table

Description automatically generated

Ideally, feature selection can be done by univariate or multivariate analysis.

The problem is, the more variables are, the longer it takes to conduct analysis.

Here is a quick solution.

Consider using a different approach, as during my Univariate Analysis, I generated a series of inquiries that I attempted to verify or refute using basic visual aids.

Follow-up questions that need to be addressed

Q1: Does aging rate cause a rise in unemployment rate?

Q2: Does high CPI cause a rise in unemployment rate?

Q3: Does low Deflator lead to a less unemployment?

Q4: Do Real GDP per captita means contribute to unemployment rate?

Q5: Does a fast GDP growth rate per captita means a lower unemployment rate?

Q6: Is a high Labor Participation Rate causing an increase in unemployment rate?

Besides, as there are lagging or leading effects in terms of the features on unemployment, similar questions can be asked.

Q7: Does high CPI lagged term cause a rise in unemployment rate?

Q8: Does low Deflator lagged term lead to less unemployment?

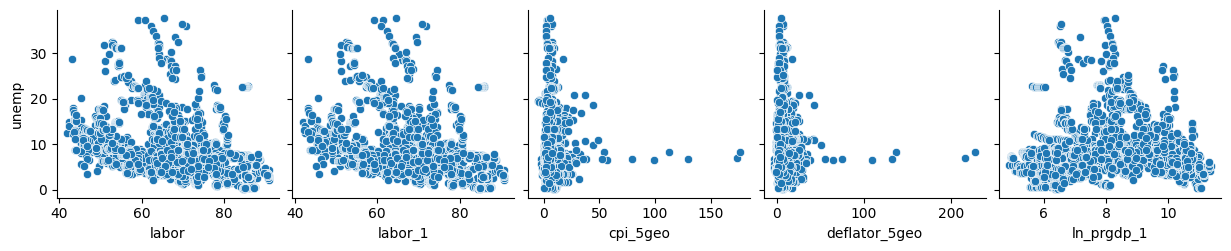
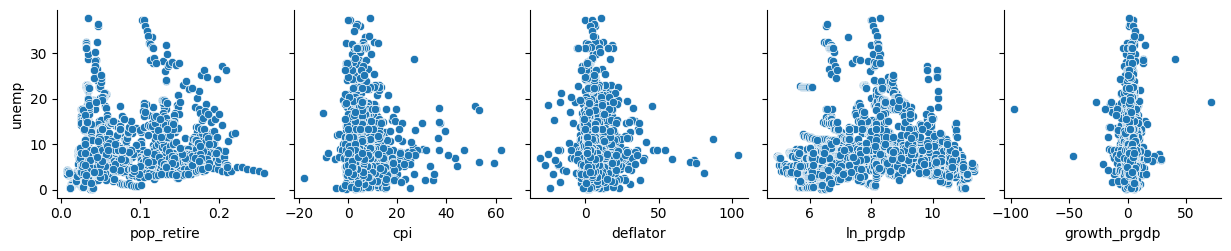
Q9: Does Real GDP per captita means Play a role in unemployment rate?

Q10: Does a lagged term of GDP growth rate per captita contribute to unemployment rate?

Q11: Is a high Labor Participation Rate lagged term causing an increase in unemployment rate?

Q12: Do unemployment lagged term contribute to unemployment rate?

Then, explore the relationship between the unemployment rate and each of the features mentioned above. This can be easily achieved by plotting a pair plot between target variable and calculating the corresponding correlation matrix.

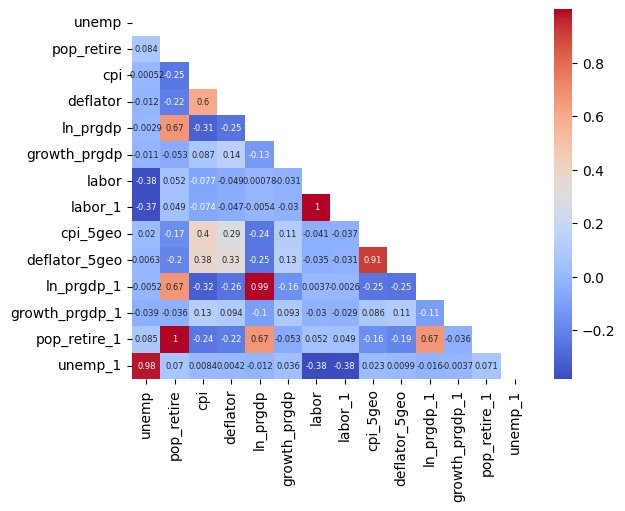


Chart, scatter chart

Description automatically generated

It can be seen that the answers to question 1, 4, 6, 9, 11 and 12 are Yes.

This can be confirmed by checking the correlation matrix of the features and the target variable.



Conclusion: 1, 4, 6, 9, 11 and 12

Q1: Does aging rate cause a rise in unemployment rate? Yes

Q2: Does high CPI cause a rise in unemployment rate? NO

Q3: Does low Deflator lead to a less unemployment? NO

Q4: Do Real GDP per captita contribute to unemployment rate? Yes

Q5: Does a fast GDP growth rate per captita means a lower unemployment rate? NO

Q6: Is a high Labor Participation Rate causing an increase in unemployment rate? Yes

Q7: Does high CPI lagged term cause a rise in unemployment rate? NO

Q8: Does low Deflator lagged term lead to less unemployment? NO

Q9: Does Real GDP per captita means Play a role in unemployment rate? Yes

Q10: Does a lagged term of GDP growth rate per captita contribute to unemployment rate? NO

Q11: Is a high Labor Participation Rate lagged term causing an increase in unemployment rate? Yes

Q12: Do unemployment lagged term contribute to unemployment rate? Yes