**New Zealand's Unemployment Rate:** **A Macroeconomic Analysis and Predictive Study**

**DATA 422 Data Wrangling Group Project**



**Team Members:**

**Lin (Lin)Song -11199728**

**Huiyuan (Daisy) Xie -97451727**

**Peng Dong – 56950925**

**Song Gao -**

**Introduction:**

Labor market conditions have long captured the interest of policymakers, economist, and investors. Several factors can impact Unemployment Rate, including population structure, macroeconomics, religion, and politics (Laanani & Rey, 2015). Within the realm of macroeconomics, in theory, the relationship between macro-economic situation and unemployment demonstrates a negative correlation (Kalinová & Kroutlová, 2023). Therefore, our project aims to analyze the intricate connections between selected macroeconomic indicators in New Zealand. Simultaneously, we also observe the New Zealand Unemployment Rate among high-income countries within the East & Asia Pacific region (the same region where New Zealand is located), allowing us to gain a deeper understanding of which selected macroeconomic indicators exert a more powerful influence on the Unemployment Rate in New Zealand and whether the real-world trend in this relationship aligns with theoretical expectations. Finally, we model the data based on the indicators that have had a significant impact on past unemployment and use a time series model to forecast future unemployment.

**Data Sources and Data Selection Rationale:**

***Data Source:***

* **World Bank:** GDP, GDP Growth Rate, Foreign Investment, CPI and Unemployment Rate across worldwide were obtained from The World Bank via Julia and R. The World Bank is a well-respected international financial institution that provides financial and technical assistance to developing countries around the world. One of its valuable resources is the vast collection of economic and social data that it collects and maintains, which encompasses countries from around the world (The World Bank, 2018). Therefore, the primary factors driving our choice to source data from this institution are reliability, credibility, and comprehensiveness. It is worth note that there are packages was offered in both Julia ((WorldBankData)and R (WDI), which simplifies our data extraction process.
* **Interest.co.nz:**The New Zealand Currency Exchange Rate (USD/NZD) was collected from the Interest.co.nz. This New Zealand-based platform provides valuable insights into mortgages, interest rates, exchange rate, investments, and personal finance. It serves as a valuable resource for gaining a deeper understanding of the New Zealand financial market and economy.
* **Marcotrends:** We sourced our New Zealand Export data from Marcotrends. It is a reputable platform renowned for its comprehensive collection of economic and financial data. Also, it is a valuable resource frequently utilized by researchers and analysts worldwide due to its commitment to providing reliable and up-to-date data.
* **OECD:** This website provides New Zealand Business Confidence Index (BCI) and Consumer Confidence index (CCI). The Organization for Economic Co-operation and Development (OECD) is a prestigious international organization recognized for its comprehensive economic analysis and data collection. Comprising member countries from around the world, the OECD is dedicated to promoting economic growth, stability, and improved living standards (Kenton, 2020). It essential to note that disparities in the BCI index and the CCI index between OECD and other institutions (e.g., ANZ-Roy Morgan Consumer Confidence) can raise from serval factors such as methodological variances, survey timing, sample size, etc. (Piger, 2003). However, the overall trend appears to be similar after comparing the trend of OECD and other instantiations.

***Data (******Economic Indicator) Selection:***

In this project, we have chosen two primary datasets. The first dataset, the East Asia & Pacific dataset, primarily comprises the following variables: "Country," "Country Code," "Year," "GDP," "GDP Growth," "CPI," and "Unemployment Rate." The second dataset, specific to New Zealand, includes a broader range of variables, encompassing "Country," "Country Code," "Year," "GDP," "GDP Growth," "CPI," "Mean Exchange Rate" (yearly average), "Export," "Mean BCI" (yearly average), "Mean CCI" (yearly average), and "Unemployment Rate."

* **Gross Domestic Price (GDP):** GDP measures the monetary value of goods and services produced within a country's borders during a specific time period, such as the annual GDP we have chosen. It functions as a comprehensive report card for evaluating the overall health of a country's economy. In theory, a strong GDP often correlating with job creation and reduced unemployment. (Fernando, 2023).
* **Gross Domestic Price Growth (GDP Growth Rate):** An economic growth rate indicates the percentage shift in the total value of goods and services generated within a country over a given time frame (the annual growth rate we have chosen), in comparison to a previous period. A greater growth rate signifies an expansion of the economy, influencing employment opportunities (Chen, 2023).
* **Consumer Price Index (CPI):** The CPI measures the average change in prices paid by consumers for goods and services and is one of the most widely used metrics for tracking inflation and deflation. Central banks use CPI data to assess people's purchasing power and formulate economic policies. Typically, a healthy CPI falls within the range of 1% to 3%. If the CPI falls below 1%, there is a risk of deflation. Deflation occurs when consumer and asset prices decrease over time, leading to increased purchasing power. When consumers expect lower prices, they delay their purchases in the hope of securing better deals in the future. However, reduced consumer spending can result in lower income for producers, potentially leading to unemployment and slower economic growth. Conversely, there is a risk of inflation when the CPI exceeds 3%. Inflation represents a rise in prices, resulting in a decrease in purchasing power over time. This reduction in purchasing power affects the cost of living for the general public and can ultimately lead to a slowdown in economic growth, potentially resulting in unemployment and higher interest rates (Fernando, 2023).
* **Foreign investment:** Foreign investment refers to the act of a foreign investor investing in the assets and enterprises of a host country. Large multinational corporations actively pursue fresh avenues for economic expansion by establishing branches and broadening their investments abroad. Foreign investment is widely recognized as a catalyst for future economic growth and employment creation (Chen, 2023).
* **Currency Exchange Rate**: Export is a backbone in New Zealand and Exchange Rate affect a country's export and import competitiveness. A favorable exchange rate might boost exports, leading to economic growth and job creation.
* **Export:** High volume and value of a New Zealand’s exports drives economic growth, in other words, strong export performance might create more employment opportunities.
* **Business Confidence Index (BCI):** New Zealand is the heaven of small business and small business dominate some industries more than others. The BCI reflects how businesses feel about their firm's future and overall economic performance. Generally, high business confidence tends to lead to increased investments and expansion, which might generate more opportunities in the labor market.
* **Consumer Confidence Index (CCI):** CCI measures the optimism of consumers. When consumers are confident about the economy, they tend to spend more, leading to increased demand for goods and service, which provide significate insight into New Zealand economic conditions (Ganti, 2023).

Despite these indicators are not always the case in every economy, they collectively provide valuable insights into how economic conditions influence unemployment, offering a more comprehensive understanding of the complex relationship between economic factors and labor market conditions (Laanani & Rey, 2015).

**Target:**

The selected indicators encompass a broad spectrum of macroeconomic variables that could potentially influence the unemployment rate. Therefore, our goal is to leverage the power of data analysis tools (R and Julia) to clean and analyze these indicators, seeking to identify which of them exerts the most pronounced influence on unemployment. By employing these analytical tools, we aim to uncover valuable insights into the complex relationship between New Zealand's labor market conditions and a multitude of economic factors.

After conducting exploratory data analysis on the clean dataset, we have selected the economic indicators that significantly impact Unemployment Rates for our time series research. We will use these indicators to predict the employment rate for the next two years by fitting a time series model.

**Techniques:**

***Code Name and Description Searching:***

* Prior to data collection, thorough information searching is a crucial step in our project. During this phase, we utilized Julia and employed the "WorldBankData" package to retrieve descriptions and code names of our selected indicators, namely GDP, GDP growth, CPI, and Foreign Investment, from the World Bank database.

***Data collection and scraping:***

* We employed the "WDI" package in R to gather worldwide data from the World Bank, using the code names of the indicators. Additionally, we created a Julia function to scrape the New Zealand exchange rate (USD/NZD) from the website interest.co.nz, obtaining a CSV file. Further, we extracted New Zealand's Export data from macrotrends.net using R. We also downloaded the Business Confidence Index (BCI) and Consumer Confidence Index (CCI) from the OECD.

***Data Cleaning and Wrangling:***

In the data cleaning and wrangling phase, we carried out the following tasks:

* Filtered the datasets for New Zealand, as well as the East Asia & Pacific and high-income regions.
* Renamed variables with unclear names to ensure clarity and conciseness.
* Conducted data reshaping and transformation to convert the data from a long format to a wide format, enabling the observation of economic indicators and their year-by-year details.
* Conducted checks for missing values pattern and identified the first year of unemployment record.
* Addressed missing values by removal or imputation based on specific circumstances.
* Merged the BCI, CCI, Export and Currency Exchange Rate datasets to create a comprehensive New Zealand dataset, which is prepared for subsequent visualization, analysis and modelling.

Undoubtedly, we encountered certain obstacles in the process of adapting the data source to our target data model. The World Bank data spans from 1960 to 2022, but after the data reshaping, a significant number of missing values emerged in the unemployment rate column (see Picture 1).

A graph showing a graph of a number of data

Description automatically generated with medium confidence

*Picture 1: The missing value patten in the East Asia & Pacific and high-income regions dataset.*

Consequently, we identified that the Unemployment Rate records were available from 1991 onwards by using a function (see Picture 2).

A screenshot of a computer code

Description automatically generated

*Picture 2: Define a function to find the first column of Unemployment Rate with non-missing values.*

As a result, the analysis period for the East Asia & Pacific dataset and the New Zealand dataset was limited to the years 1991 to 2022 and 1991 and 2021 respectively. This reduction in data volume of New Zealand dataset might potentially generate limitations to our analysis. To address this, we included additional variables (Currency Exchange Rate, Export, BCI, CCI) that might influence the Unemployment Rate during modeling stage in New Zealand dataset. Moreover, during our investigation, we observed a time series trend in Foreign Investment from 2000 to 2022 (see Picture 3).

A graph showing the time and year

Description automatically generated with medium confidence

*Picture 3: Foreign Investment Over Time*

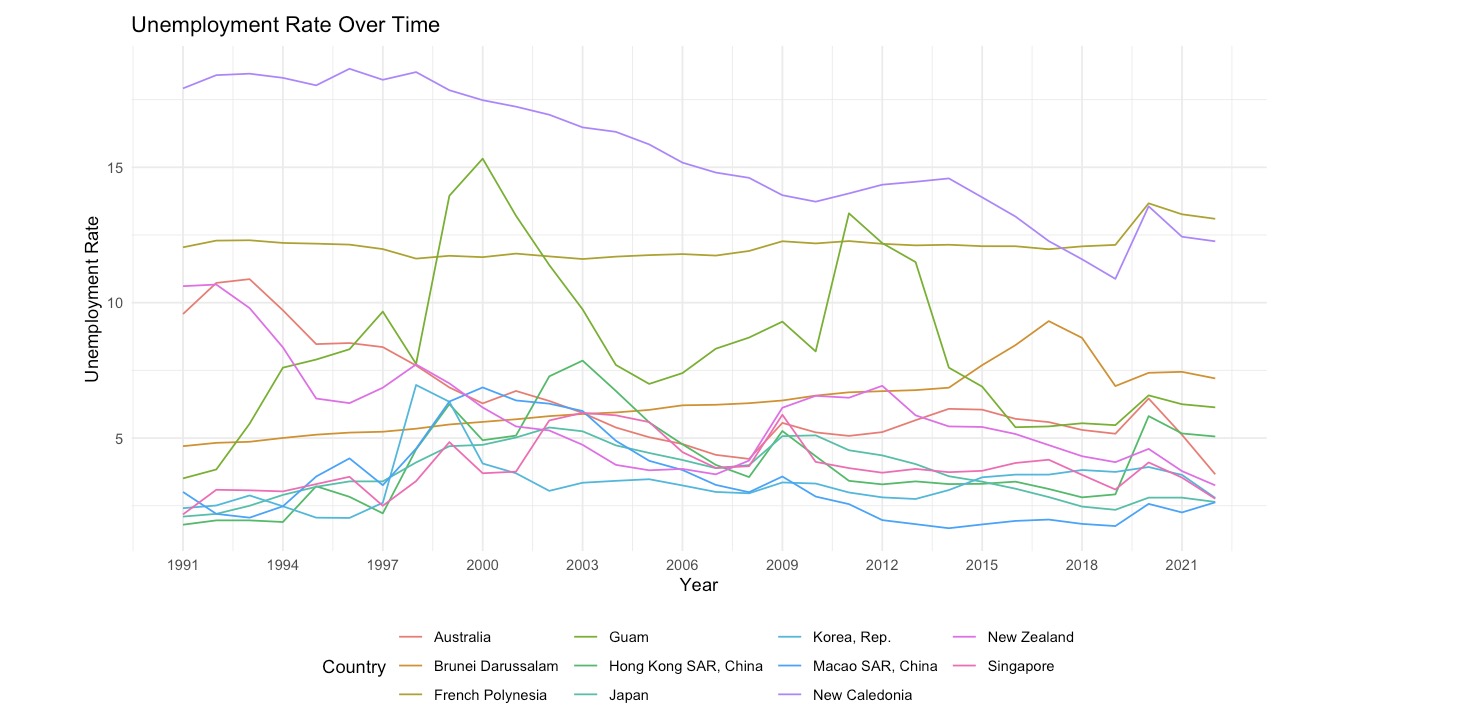
At regular intervals, there is a modest upturn in Foreign Investments, succeeded by a more substantial surge. Subsequently, there is a sharp and significant decline. This sequence repeats itself three times between 2000 and 2021. Consequently, we can employ this recurring pattern to fill in the missing values predating the year 2000. To impute the missing values in the time series variable "Foreign Investment" from 1990 to 2000, we employed the "na.kalman" function from the "ImputeTS" package. The Kalman filter imputation method is statistically efficient, it considers not just the visible data but also the inherent structure and patterns within the time series, enabling it to accurate estimates for missing value (Boiko et al., 2019). This approach significantly enhances the reliability of our dataset by mitigating data gaps. From the data frame obtained through the previous steps, all variables have the potential to be utilized for model fitting and data prediction. However, in practice, not all may be necessary; some could prove redundant. To streamline computation and avoid overfitting, it's essential to carefully select the variables to be included in the model.

***Data Visualization:***

* **East Asia & Pacific high-income countries dataset:**

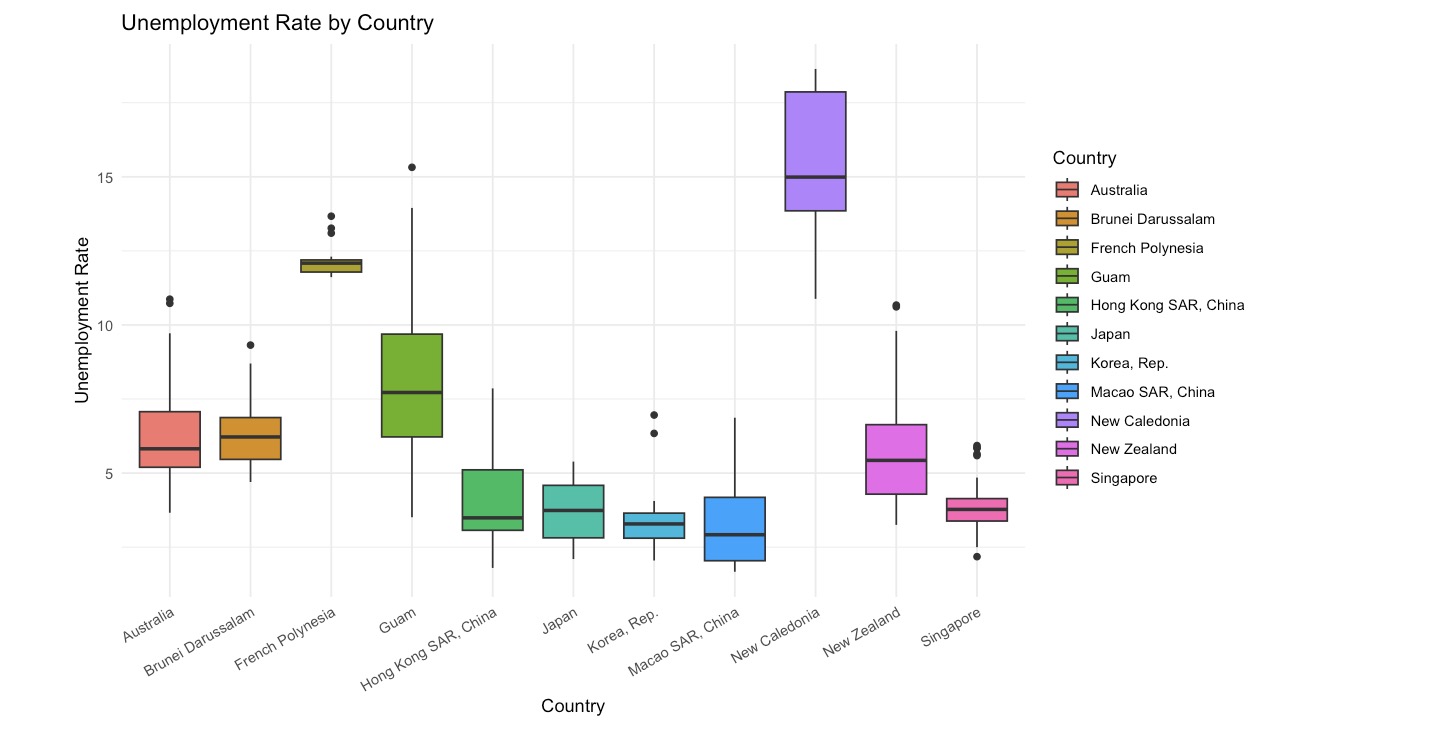
In our analysis of the East Asia & Pacific high-income countries dataset, we harnessed data visualization tools such as 'ggplot2' to conduct a comprehensive examination of the data structure, with a particular focus on analyzing the Unemployment Rate in New Zealand. To gain a deeper understanding of the patterns related to missing values, we utilized a missing value plot (as depicted in Picture 1).

The line chart we utilized (refer to Picture 4) illustrates the trend in Unemployment Rates among East Asia & Pacific high-income countries from 1991 to 2021. New Zealand displays a decreasing unemployment rate with some fluctuations compared to its regional counterparts.



*Picture 4:* East Asia & Pacific high-income countries Unemployment Rate from 1991 to 2022 line chart

A boxplot (see Picture 5) was also generated to provide a clear and concise visualization of the data distribution. New Zealand's Unemployment Rate remains stable at approximately 5%, positioning it at a moderate level within the East Asia and Pacific region. It surpasses the rates observed in Hong Kong, Macau, Japan, Singapore, and South Korea but falls below the rates in numerous other countries in the Pacific region.



*Picture 5:* East Asia & Pacific high-income countries Unemployment Rate from 1991 to 2022 boxplot chart

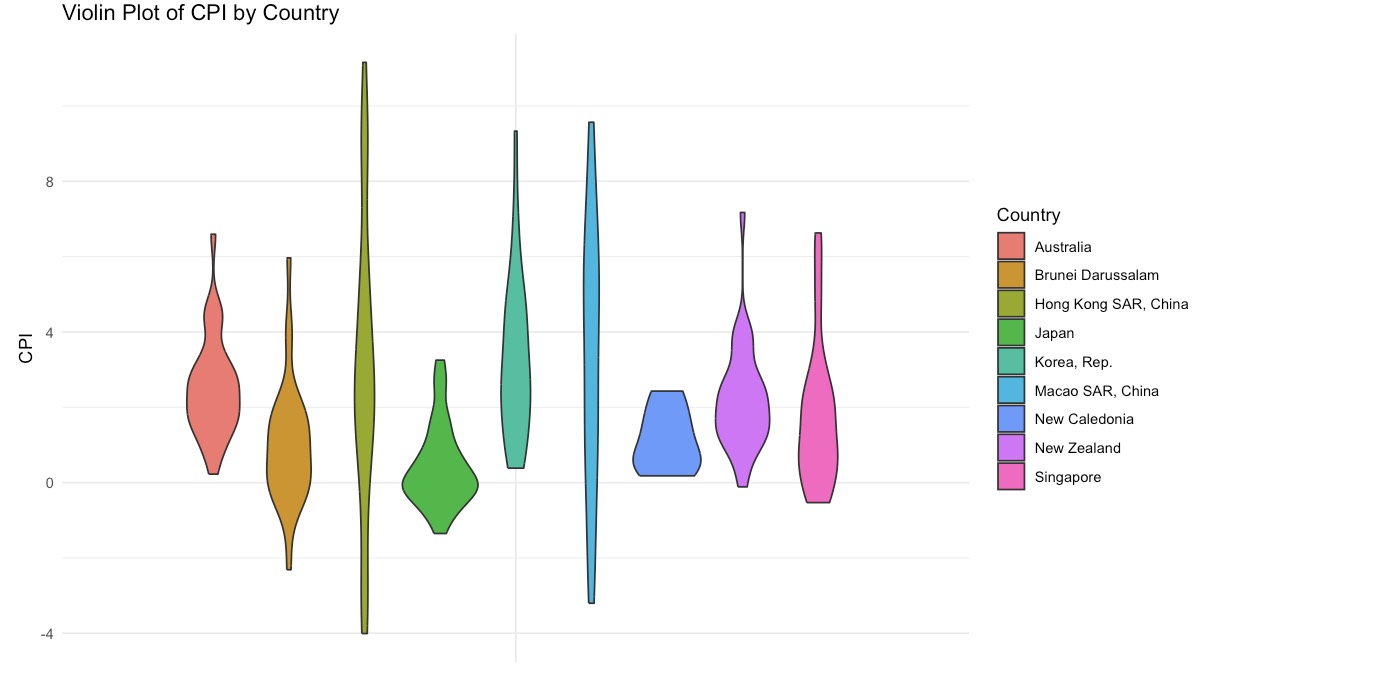
For a more concise overview of unemployment rates across various nations, we created a bar chart (see Picture 6). This chart highlights the range of unemployment rates, which span from near 0% to approximately 15%. New Zealand falls within the 5-10% range, indicating a scenario of moderate unemployment.

A graph of green squares

Description automatically generated

*Picture 6:* East Asia & Pacific high-income countries Average Unemployment Rate from 1991 to 2022 bar chart

Additionally, we used a violin plot (see Picture 7) to effectively display the distribution and density of the data. The violin plot showcases the CPI distribution in East Asia & Pacific high-income countries. New Zealand (depicted in light blue) exhibits a wider range of CPI values, suggesting higher variability in this economic indicator.



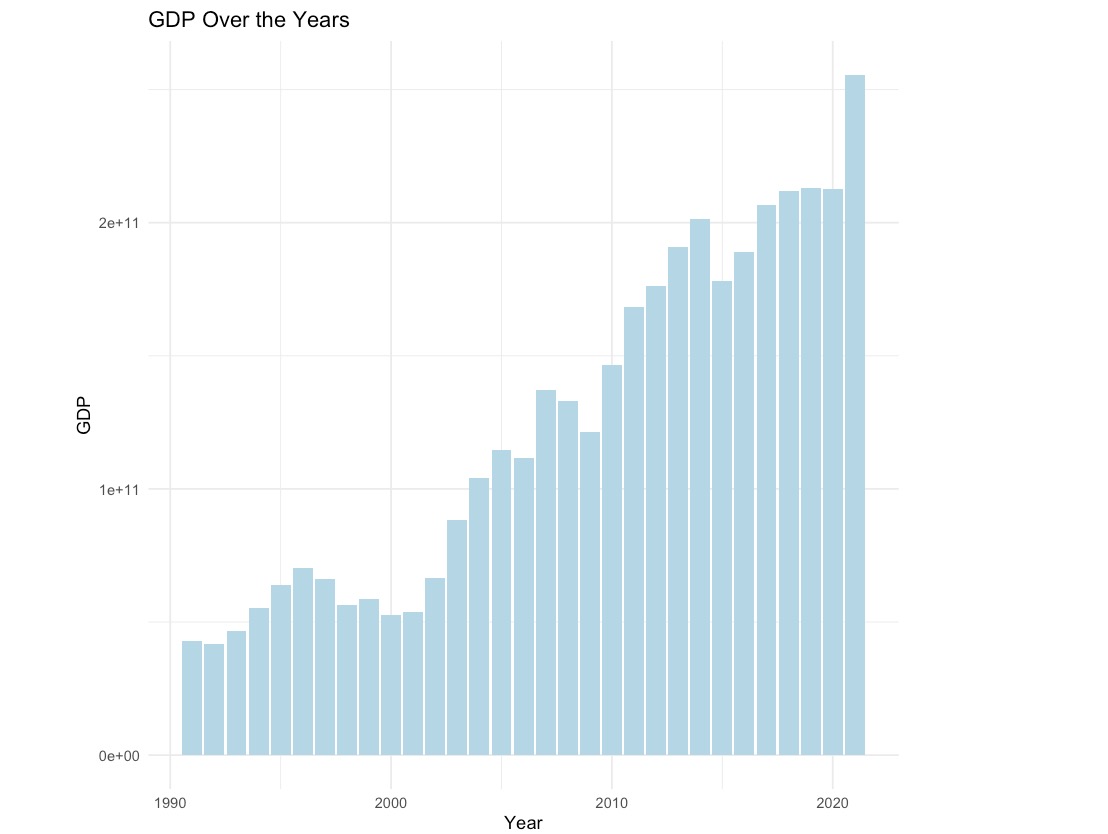
*Picture 7: East Asia & Pacific high-income countries CPI from 1991 to 2022 violin chart*

Thus, these plots offer enhanced clarity when describing the New Zealand Unemployment Rate in comparison to other high-income countries in the East Asia & Pacific region.

* **New Zealand dataset:**

After cleaning data and merging data, we used comprehensive New Zealand dataset to observe and analyze the relationship between our selected macroeconomic indicators.

The bar chart of New Zealand GDP we created (see Picture 8) shows overall upward from 1991 to 2021.



*Picture 8: New Zealand GDP over 1991 to 2021 bar chart*

In order to observe the trend of Unemployment Rate

***Data Modelling and Prediction:***

**Achievement and Failure:**

**References:**

Boiko, Y., Lin. C., Kiringa, I &Yeap, T. (2019) "Navigational data imputation with GPS pinning

in compositional Kalman filter for IoT systems*," 2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE)* pp. 1-7, doi: 10.1109/ROSE.2019.8790427.

Chen, J. (2023). Economic Growth Rate: Definition, Formula, and Example. *Invetopedia.*

https://www.investopedia.com/terms/e/economicgrowthrate.asp

Chen, J. (2023). Foreign Investment: Definition, How It Works, and Types. *Invetopedia.*

*https://www.investopedia.com/terms/f/foreign-investment.asp#:~:text=Key%20Takeaways,their%20investments%20in%20other%20countries.*

Fernando, J. (2023). Gross Domestic Product (GDP): Formula and How to Use It. *Invetopedia.*

https://www.investopedia.com/terms/g/gdp.asp

Fernando, J. (2023). Consumer Price Index (CPI) Explained: What It Is and How It's Used.

*Invetopedia.* https://www.investopedia.com/terms/c/consumerpriceindex.asp

Ganti, A. (2023). Consumer Confidence Index (CCI): Definition and What It Indicates.

*Invetopedia*. https://www.investopedia.com/terms/c/cci.asp

Kalinová, E., & Kroutlová, K. (2023). The relationship between unemployment and

macroeconomic indicators. *SHS Web of Conferences.* https://doi.org/10.1051/shsconf/202316001008

Kenton, W. (2020). Organisation for Economic Co-Operation and Development (OECD).

*Invetopedia.* https://www.investopedia.com/terms/o/oecd.asp

Laanani, M., & Rey, G. (2015). Impact of unemployment rate and the economic crisis on

suicide mortality in western european countries (2000-2010). *European Journal of Public Health, 25(suppl\_3)* https://doi.org/10.1093/eurpub/ckv168.049

Piger, M, J. (2003). Consumer Confidence Surveys: Do They Boost Forecasters' Confidence?

*Federal Reserve Bank of St.Louis.* https://www.stlouisfed.org/publications/regional-economist/april-2003/consumer-confidence-surveys-do-they-boost-forecasters-confidence

The World Bank. (2018). Introducing the online guide to the World Development Indicators:

A new way to discover data on development. *The World Bank.* https://blogs.worldbank.org/opendata/introducing-online-guide-world-development-indicators-new-way-discover-data-development