The Impact of Economy, Geography, and Education on Unemployment Rate in Indonesia: A Deep Analysis and Approach

011DSC24

Introduction

The persistent decline in Indonesia's economy and high unemployment rate have quietly grown to be a serious, yet unaddressed issue. Indonesia has made some progress in its economy over the years, but not much. Regarding unemployment, it may have a negative impact on both the economy of the nation and the wellbeing of specific people. Additionally, one may surmise that a number of uncontrollably detrimental elements to an individual's well-being could influence the economy and the unemployment rate.

The economic progress of this nation throughout the years has led to an increasing Gross Domestic Product (GDP), placing it among the largest in Southeast Asia. Nevertheless, not everyone has benefited equally from the improving economic trend, especially in rural areas where unemployment rates are still disproportionately high. It is evident that Indonesia's economic growth has not sufficiently resulted in improved living standards or decreased unemployment rates, particularly among young people, when comparing GDP growth with GDP per capita. When contrasted to nations like Singapore, which have a substantially lower population but a far greater GDP per capita, this discrepancy is further accentuated.

These issues are critical because Indonesia wants to equip its youth, who make up a large part of the population, with the knowledge and opportunities necessary to prosper in a manufacturing world that is becoming more automated and driven by artificial intelligence. By empowering the next generation of workers, AI-driven solutions can significantly reduce unemployment and address economic issues.

This dashboard's data originates from a number of sources, including national databases like BPS (Badan Pusat Statistik), Wikipedia, and international institutions like the World Bank and International Labour Organization (ILO). Key economic variables for all ASEAN nations are included in the collection, including GDP per capita, unemployment rates, and public spending on education. The majority of the data is quantitative, providing a thorough analysis of trends in unemployment in both urban and rural areas as well as showing regional variations.

Numerous preprocessing procedures were carried out in order to guarantee the data's relevancy and accuracy. This required normalizing the datasets to enable precise cross-country comparisons, filtering for post-2012 data to guarantee a focus on current trends, and cleaning the data to remove any inconsistencies or missing numbers. Furthermore, the data was partitioned to illustrate unemployment rates in urban and rural areas, providing a more lucid picture of regional differences in Indonesia. This division is important because it highlights hidden variables like population density, the movement of people from rural to urban areas, and the effect of COVID-19 on employment losses, particularly in urban industries like retail and hospitality.

The dashboard's in-depth study attempts to clarify why Indonesia's unemployment rate, particularly among young people, remains a critical concern in spite of the country's robust GDP growth. Additionally, it will show how funding for digital infrastructure and education, especially in rural areas, can aid in closing the gap between job possibilities and economic growth. These data-driven insights are intended to educate stakeholders and policymakers so they can implement focused interventions that can lower unemployment and boost economic resilience in a global market that is becoming more automated and AI-driven.

Data Analysis and Results

1. **Average Unemployment Rate compared with GDP per Capita of ASEAN Countries**

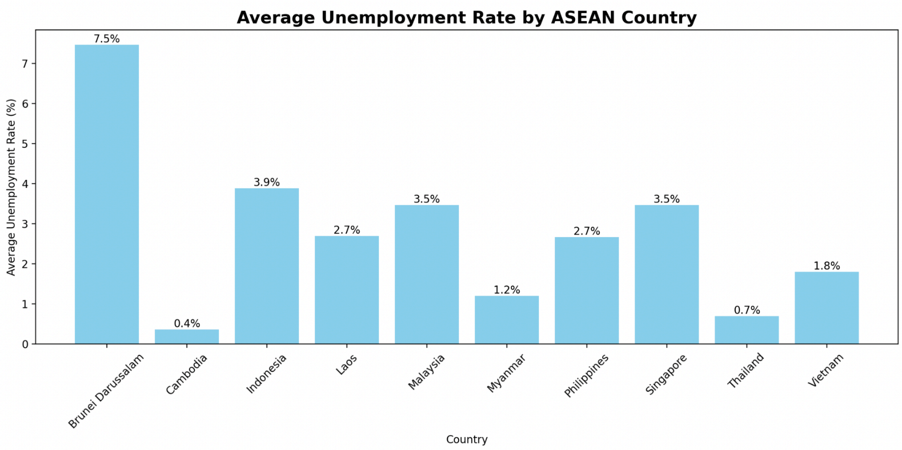


Fig 1.1

The graph shown is the unemployment data for ASEAN countries, filtered by year post-2012. The average unemployment rate was calculated for each country by aggregating unemployment values over the years from post-2012 and dividing the result by the total number of years provided. We use barcharts for a straightforward comparison. The result is that Indonesia's unemployment rate appears more concerning, with the fact that the country's population is the highest among the other ASEAN countries. Ideally, the higher one's population, the higher the working force, thus, the higher the GDP per Capita.

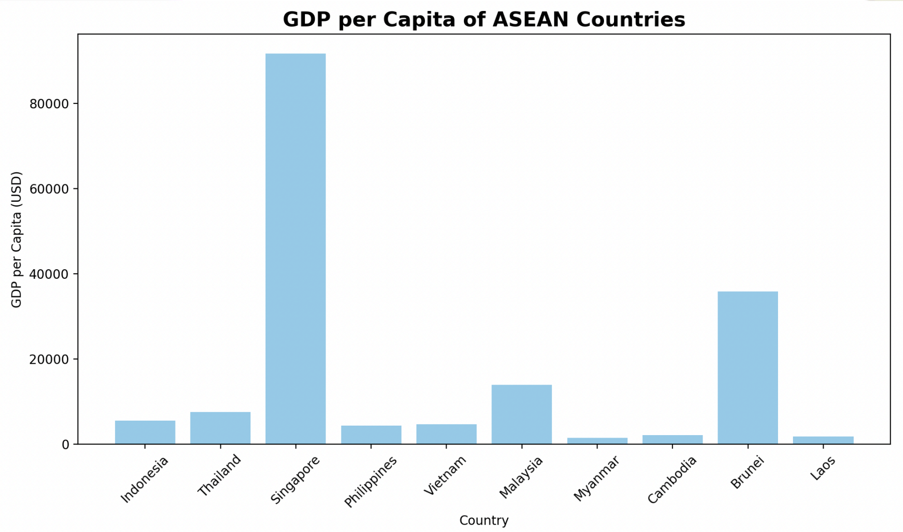


Fig 1.2

However, compared with Singapore, Indonesia's population is much more significant. Yet this population advantage has not translated into better employment outcomes or a proportionally higher GDP per capita. Singapore, despite having a much smaller population, has managed to maintain a significantly higher GDP per capita. This disparity between the two countries highlights the efficiency of economic management and labor market strategies in Singapore, where education, technology, and infrastructure have played key roles in driving both productivity and employment.

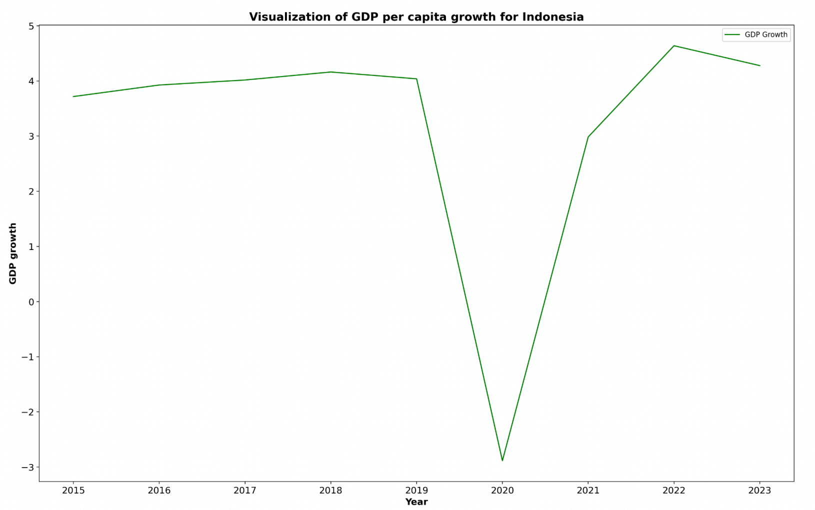


Fig 1.3

Looking more closely at Indonesia's GDP only, we can see there is a sharp downward trend between the years of 2019 and 2020. This sharp downward trend is most likely caused by the covid-19 pandemic. This shows a sharp decline of Indonesia's economy. After 2020, Indonesia's GDP is shown to be stabilizing with the recent job supply increase.

2. Urban vs Rural Unemployment Rate in Indonesia

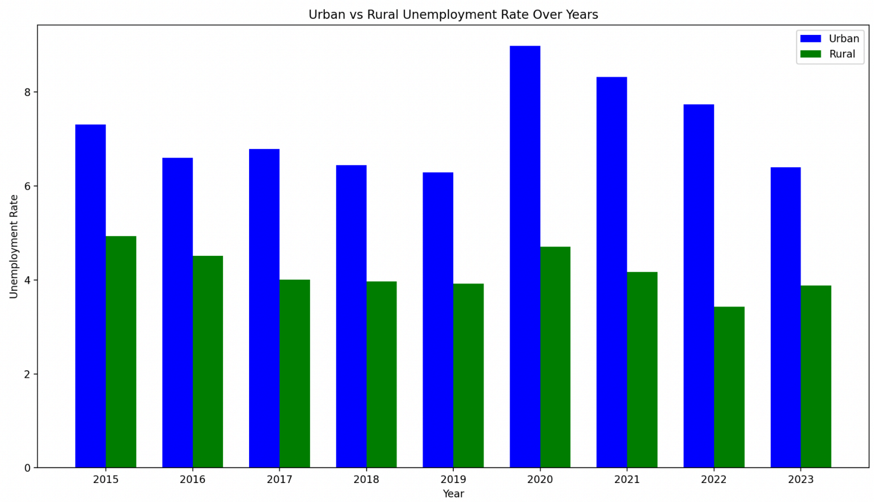


Fig 2.1

The graph above presents the unemployment rate data for Indonesia, categorized into urban and rural areas in post-2015. We use bar graphs for a straightforward visualization, and also helped identify periods with high disparities between urban and rural unemployment rates. This period has been crucial in understanding the differential impacts of economic policies and external factors on employment across different regions.

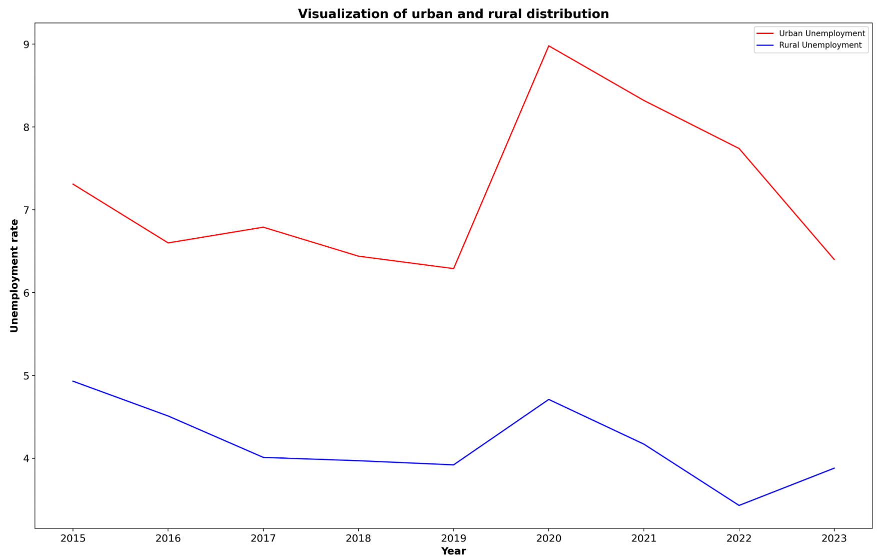


Fig 2.2

From graph fig 2.2, we can see the distribution of unemployment rate between rural and urban areas. From the urban trendline (red color), we can see post-2020, unemployment rate seems to be going down and the same goes for rural trendline (blue color). Numerous important variables are responsible for the continually higher unemployment rate in metropolitan regions. Density of population is one major factor. Due to the dense population in cities, there is fierce rivalry for the few available jobs. In addition, it is also uncommon for people to migrate from rural to metropolitan areas in pursuit of greater job possibilities. This flood of job searchers frequently exceeds the supply of open positions, making urban unemployment even worse.

Furthermore, we can see that during 2019-2020, we can see a spike in both rural and urban areas. Most likely due to the COVID-19 epidemic, which disproportionately affected urban areas, made matters worse. There were major disruptions in the retail, hotel, tourism, and other service-based businesses that were primarily focused in cities. An abrupt increase in urban unemployment was brought on by widespread layoffs and company closures as a result of quarantine regulations. The majority of the nearly 81 million jobs lost in Asia-Pacific during the pandemic, according to the International Labour Organization (ILO), were in metropolitan areas.

On the other hand, the direct effects of lockdowns had less of an impact on rural economies, which are mostly dependent on the informal sector and agriculture. Despite long-standing economic issues including lower salaries and less access to services, rural areas' reliance on agriculture served as a safety net against the employment losses that devastated metropolitan sectors. For example, throughout the pandemic, rural unemployment stayed largely steady because of the ongoing demand for agricultural products, although urban unemployment rates skyrocketed.

This discrepancy emphasizes the necessity of focused urban labor strategies that tackle the problems posed by dense populations as well as the financial susceptibility of urban areas to outside shocks like pandemics. Simultaneously, rural communities stand to gain from sustained investment in economic diversification, which would increase formal job opportunities and lessen the forces that encourage migration to urban areas.

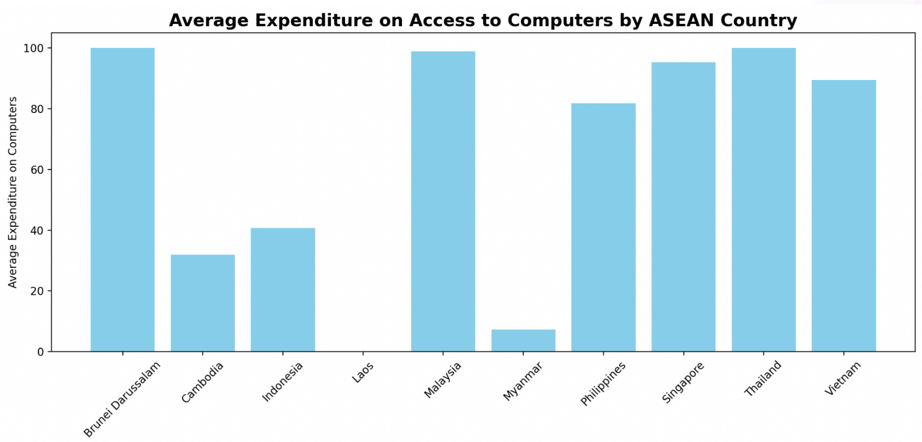


Fig 2.3

In this part, we will see the role of technology and its effects on the unemployment rate and GDP per capita of each ASEAN country. Furthermore, we will explore each country's average expenditure in technology with reference to fig 1.2 and fig 1.1.

Let's take a look at the correlation between high expenditure in technology and GDP growth. Countries like Brunei, Singapore and Thailand have the most expenditure in technology. These countries tend to have higher GDP per capita as evident in fig 1.2. Supposedly, investment in technology, for this instance, increases productivity rate and thus reduces unemployment rate as a whole. Increasing productivity rates are one of the drivers for increasing GDP growth.

Now let's take a look at the correlation between lower unemployment rates and high expenditure in technology. Countries, like Singapore, have lower unemployed individuals because investment in technology creates new job supplies for high-skill sectors, such as IT, finance, manufacturing, and computer science. Not only is it highly correlated with unemployment rates, but also in education as the more availability to technology is in one's country, the better the education and the upskilling opportunities.

What does this mean for low expenditure in technology? Countries like Myanmar, Cambodia and Indonesia, may experience slower GDP growth and higher unemployment rates. As some areas are not familiar with advanced technology, this significantly reduce productivity

3. Prediction for Further Analysis

In this part of the paper, we are going to look deeper into the visualization of what will happen in the next decade using machine learning approaches. Below are the results.

Fig 3.1

Fig 3.2

As shown in fig 3.1, we can see a downwards trend for the unemployment rate in Indonesia. By using the data from 2015 to 2021 from the unemployment rate dataset, we can predict that the coming years will lead to a decrease in the unemployment rate in Indonesia. This is confirmed when using the same dataset to confirm for the year 2023, which shows a percent of 3.27% as evident with the graph, Fig 3.2, and the same downward trend as shown in fig 3.1. This means that there may be an increase in the job supply and higher education rate post-2024 which may lead to the decrease in the unemployment rate.

Now let us take a look closer at Indonesia's geography factor and its correlation with unemployment rate and its predictions.

Fig 3.3

For the urban unemployment rate, we can see post-2024, we can expect a continuous downward trend, possibly lower than rural unemployment rate should we fast forward another 10 years. This could be likely due to more job supply with competent high-skill in demand. Furthermore, decreasing unemployment rate can mean boosting productivity rate. If we refer back to fig 1.2, we might catch up to Singapore if no third factors come into play, such as another pandemic.

For the rural unemployment rate, we can see post-2024, we can expect a slight increase in the unemployment rate up to 2026. However, going from 2026 forward, we can expect a plateau of unemployment rate value around the value of 4.3%. This could be likely due to the steady job supply with standard skills that only the rural areas are familiar with, such as farming, selling goods and merchants. This could be a problem if there is no increasing job supply, productivity rate in the rural areas may be steady. Another factor as mentioned before in the description of fig 2.2, there might be a possibility for urban migration, which could also explain the plateau unemployment rate post 2024 in rural areas and the high unemployment rate in urban areas.

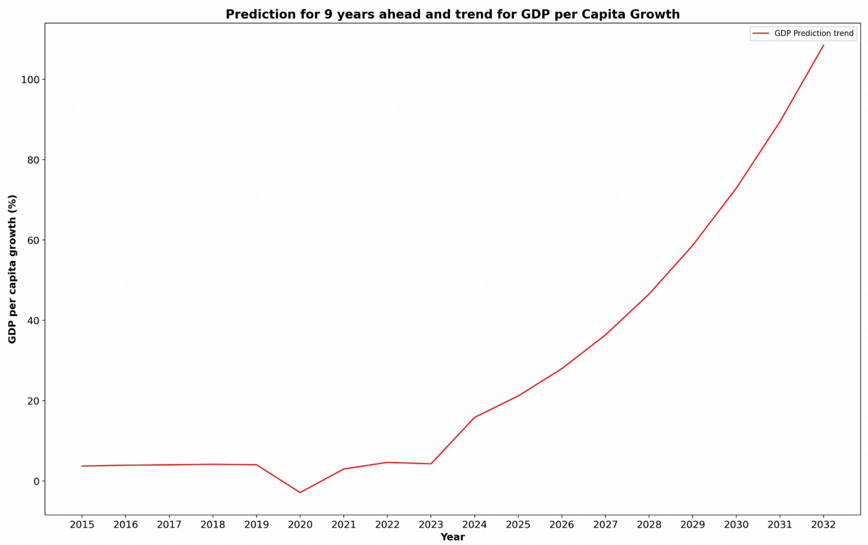


Fig 3.4

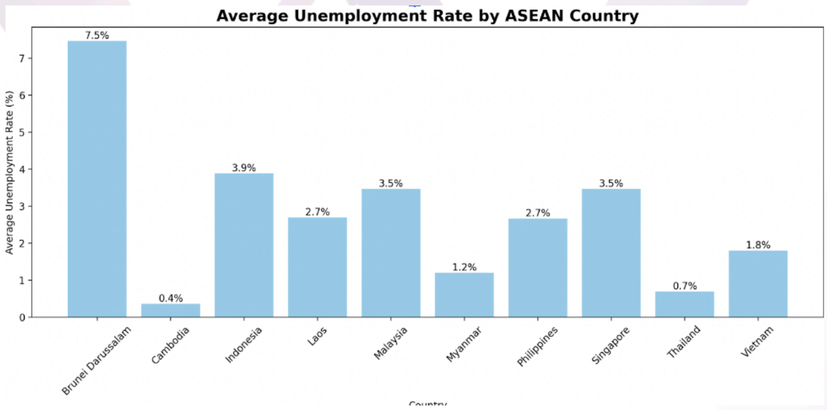
Decreasing unemployment rate can mean increasing productivity rate. Decreasing unemployment rate means that the job supply is increasing for both low-skill and high-skill individuals. As more people get into the workforce, the economy increases as the production of both goods and services increase. Increasing the economy means increasing GDP as a whole. Increased GDP means higher GDP per capita.

In correlation with increasing productivity, individuals who have joined the workforce will generate income and this can contribute to two other effects, higher disposable income and increased tax revenues. Higher disposable incomes contribute to GDP as a whole as consumer spending increases. This means that supply-and-demand for goods and services is shifting fast, thus boosting economic growth. For the government, higher income means increased tax revenues. As more and more people enter the workforce, thus reducing the unemployment rate, governments can collect more taxes, which can fund the country's growing infrastructure as a whole. Furthermore, in correlation with the higher disposable income, consumer spending can also boost productivity in the industry sector.

The upward trend may mean that Indonesia is starting to invest in the technology sector. As analyzed before in fig 2.3, higher expenditure in technology means higher GDP growth per capita. This also means that higher job supply for high-skill in demand and better education. Numerically speaking, increasing GDP growth for Indonesia means that we are on progress to rival Thailand's, Malaysia's, Brunei's, and Singapore's GDP per capita as mentioned in fig 1.2, though it may take us a few decades to catch up should no other third factor come into play and disrupt the economy and unemployment rate.

User Manual

1. Average Unemployment Rate compared with GDP per Capita of ASEAN Countries



- Overview

  This chart displays the average unemployment rate across several ASEAN countries. The data has been aggregated from post-2012 values, and the unemployment rates are represented as percentages on the vertical axis, and the countries are shown on the horizontal axis. We preprocessed the data to only show ASEAN countries' unemployment rate from post-2012. The goal of this visualization is to give users a comparative understanding of unemployment rate among ASEAN countries, and also highlight Indonesia's position relative to other nations in the ASEAN regions.

This analysis allows for a better understanding of Indonesia's unemployment rate in the context of its economic standing. Indonesia, being the largest country in terms of population within ASEAN, presents a unique case where a higher population might ideally result in a higher workforce. However, the chart shows that despite this advantage, Indonesia still has a relatively high unemployment rate compared to some of its regional peers. This relatively high unemployment rate suggests that Indonesia faces challenges in fully utilizing its labor force to stimulate economic growth.

- Chart Description

The title of the chart is "Average Unemployment Rate by ASEAN Country", and we used a barchart to make a straightforward comparison. Each bar represents the average unemployment rate for each country, providing an easy comparison. The highest unemployment rate is Brunei Darussalam at 7.5%, and the lowest unemployment rate is Cambodia at 0.4%. Disclaimer, this rate also depends on each country's population, but from this graph, we see that Indonesia stands at 3.9%, which is relatively high compared to other large nations like Vietnam (1.8%).

- Observation

The chart shows that Brunei Darussalam has the highest unemployment rate at 7.5%, while Cambodia has the lowest rate at 0.4%. Countries like Indonesia (3.9%), Singapore (3.5%), and Malaysia (3.5%) are shown to have mid-range unemployment rates. Larger populations such as Indonesia are expected to have a larger working force, but still have relatively high unemployment rates compared to smaller nations like Singapore.

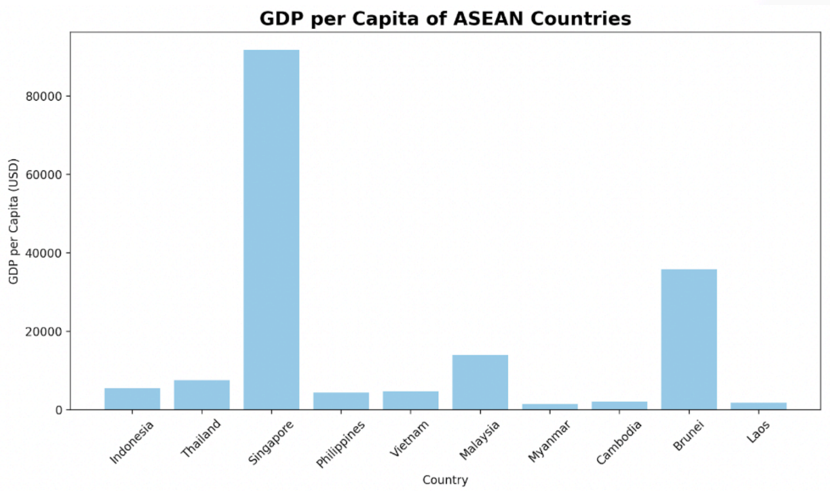
- How to Create

First, we filtered the data from the website, and took only the unemployment rates among ASEAN countries. We use Python for the preprocessed step. To recreate this chart, the following Python libraries are required: Matplotlib for generating the bar chart, and Pandas to manage the data. Below is a Python code example to generate the chart:

```python
class AseanUnemploymentRate:
    def __init__(self, file_path):
        self.file_path = file_path
        self.asean_data = []
    1 usage
    def read_and_filter_data(self):
        with open(self.file_path, 'r') as csvfile:
            reader = csvfile.readlines()
            for row in reader[1:]:
                row_split = row.split(';')
                country = row_split[1]
                year = row_split[2]
                series = row_split[3]
                value = row_split[4].strip()
                if country in asean_countries and year >= "2012" and "unemployment rate" in series.lower():
                    data = LabourForceData(country, year, series, value)
                    self.asean_data.append(data)
                    data.display()
    1 usage
    def calculate_average_unemployment_rate(self):
        country_data = {country: [] for country in asean_countries}

        for data in self.asean_data:
            country_data[data.country].append(data.value)

        average_unemployment = {country: (sum(values) / len(values) if values else 0) for country, values in country_data.items()}
        return average_unemployment

file_path = "(UN data) ASEAN unemployment rate.csv"
asean_labour_force = AseanUnemploymentRate(file_path)
asean_labour_force.read_and_filter_data()
average_unemployment = asean_labour_force.calculate_average_unemployment_rate()
```

```python
import matplotlib.pyplot as plt

2 usages
def visualize_average_unemployment(average_unemployment):
    countries = list(average_unemployment.keys())
    averages = list(average_unemployment.values())

    plt.figure(figsize=(12, 6))
    bars = plt.bar(countries, averages, color='skyblue')
    plt.xlabel('Country')
    plt.ylabel('Average Unemployment Rate (%)')
    plt.title( label: 'Average Unemployment Rate by ASEAN Country', fontsize=16, fontweight='bold')
    plt.xticks(rotation=45)

    for bar in bars:
        yval = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, yval,  s: f'{yval:.1f}%', va='bottom', ha='center')

    plt.tight_layout()
    plt.show()
```

2. GDP per Capita for ASEAN Countries



- Overview

This chart displays the GDP per capita of ASEAN countries. The main goal is to offer a comparative visualization of economic performance by showing how much income (per person) each country generates, with a particular focus on Indonesia's position relative to other nations. We use bar charts for straightforward comparison.

- Chart Description

Each bar represents the GDP per capita of a specific ASEAN country. The chart reveals significant economic disparities between the countries. From this chart, we know that the highest GDP per capita among ASEAN countries is Singapore ($85,900 USD) and the lowest is Myanmar, while Indonesia's GDP per capita is displayed at around $3,890 USD which is significantly lower than few ASEAN countries according to the chart.

- Observation

Although Indonesia has the largest population in the ASEAN region, its GDP per capita remains relatively low. This suggests that despite its large workforce potential, the economy may not be fully utilizing its population's capacity to produce wealth. This lower GDP per capita indicates structural economic challenges, such as unemployment, underemployment, or reliance on low-value sectors like agriculture.

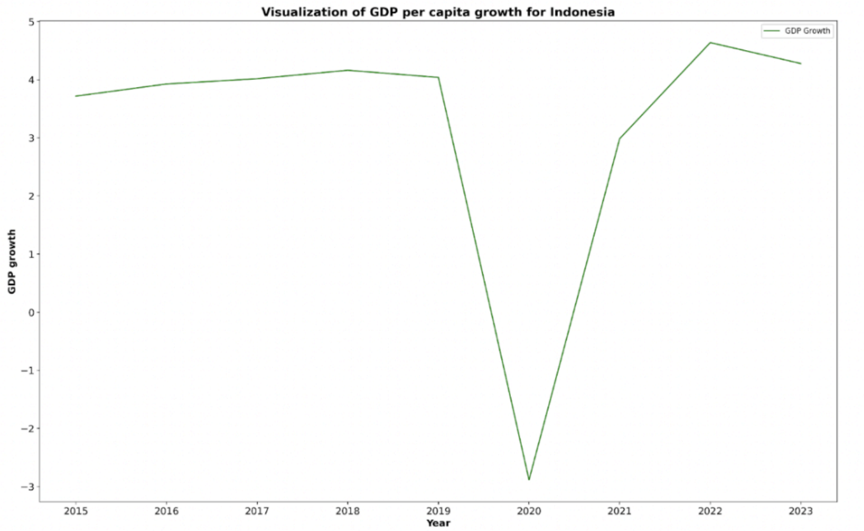
It also points to issues such as inequality in wealth distribution and the need for further development in industrial and technological sectors to boost productivity. In comparison to more developed countries like Singapore, where a smaller population produces much higher per capita income due to advanced industries like finance and technology, Indonesia's economy appears to be lagging behind. This reflects the need for economic reforms to improve workforce productivity and drive GDP growth.

- How we create

To recreate this chart, the following Python libraries are required: Matplotlib for generating the bar chart, and Pandas to manage the data. Below is a Python code example to generate the chart:

```python
import pandas as pd
import matplotlib.pyplot as plt

file_path = '(WIKIPEDIA) list gdp of asean.csv'
df = pd.read_csv(file_path)
df_countries = df[df['Country'] != 'ASEAN']

plt.figure(figsize=(10, 6))
plt.bar(df_countries['Country'], df_countries['GDP per capita (USD)'], color='skyblue')
plt.title('GDP per Capita of ASEAN Countries', fontsize=16, fontweight='bold')
plt.xlabel('Country')
plt.ylabel('GDP per Capita (USD)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

3. GDP per Capita in Indonesia



- Overview

This chart displays the GDP per capita growth for Indonesia over a period from 2015 to 2023. The objective of this visualization is to provide users with an understanding of how Indonesia's economy, specifically in terms of GDP per capita, has grown or contracted in recent years, with a special focus on the economic impact of the COVID-19 pandemic.

- Chart Description

The X-axis represents the years, ranging from 2015 to 2023, while the Y-axis displays the GDP per capita growth rate (percentage) where positive values indicate growth and negative values represent economic contraction. The line graph shows the fluctuation in GDP per capita growth for Indonesia over time, but a sharp decline is observed between 2019 and 2020, highlighting the negative impact of the Covid-19 pandemic on Indonesia's economy, but the post-pandemic recovery is represented by a steep upward trend from 2020 to

2021, showing how Indonesia's economy bounced back. In post-2021, the growth levels stabilize, but remain below pre-pandemic levels.

- Observation

In 2015 to 2019, the chart shows a steady, relatively modest GDP per capita growth, indicating that Indonesia's economy was growing at a stable pace before the pandemic. From 2019 to 2020, the most prominent feature is the sharp drop in 2020, where the GDP per capita growth plunges into negative territory (around -3%). This corresponds to the economic downturn caused by the COVID-19 pandemic, when lockdowns, restrictions, and global trade disruptions affected economic productivity.

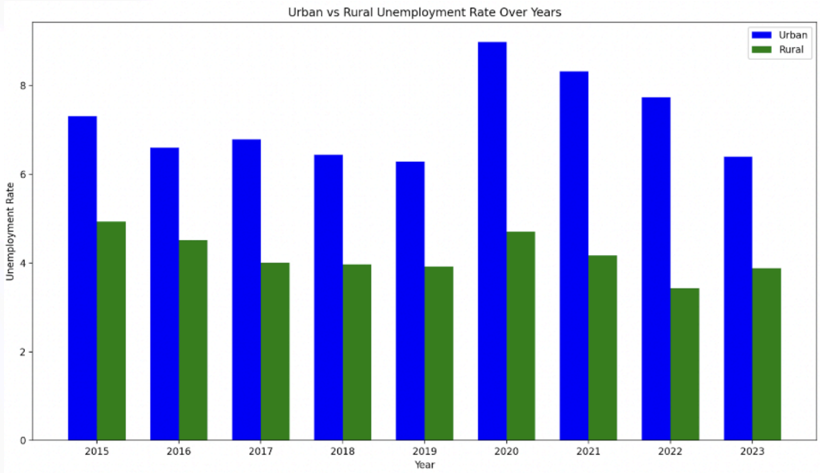
However, there is a strong recovery in GDP per capita growth post-2020, with the graph showing a sharp rise as economic activity resumes and growth bounces back. And from 2022 to 2023, the growth appears to stabilize, but still hasn't reached the pre-pandemic levels of 2019.

- How We Create

To recreate this chart, the following Python libraries are required: Matplotlib for generating the bar chart, and Pandas to manage the data. Below is a Python code example to generate the chart:

```
import pandas as pd
import matplotlib.pyplot as plt

year_arr = [str(year) for year in range(2015, 2024)]
GDP_growth_world_df = pd.read_csv( filepath_or_buffer: "GDP GROWTH INDONESIA - WORLD BANK.csv", delimiter=",")
GDP_growth_indonesia = GDP_growth_world_df[(GDP_growth_world_df["Country Name"] == "Indonesia")].drop(columns=['Unnamed: 68'])
GDP_growth_indonesia_per_year = GDP_growth_indonesia.loc[:, [year for year in GDP_growth_indonesia.columns if year.isdigit() and int(year) >= 2015]
GDP_growth_indonesia_per_year = GDP_growth_indonesia_per_year.transpose()
GDP_growth_arr = GDP_growth_indonesia_per_year.values.flatten().tolist()

plt.plot( *args: year_arr, GDP_growth_arr, color="green", label="GDP Growth")
plt.title( label: "Visualization of GDP per capita growth for Indonesia", fontweight="bold", fontsize=16)
plt.xlabel( xlabel: "Year", fontsize=13, fontweight='bold')
plt.ylabel( ylabel: "GDP growth", fontweight="bold", fontsize=13)
plt.legend()
plt.show()
```

4. Urban vs Rural Unemployment Rate in Indonesia



- Overview
  This chart provides a comparative analysis of the unemployment rates in urban and rural areas over a span of several years (2015–2023). The aim is to highlight the differences between urban and rural unemployment trends and how these have evolved over time.

- Chart Description
  The X-axis represents the years from 2015 to 2023, while the Y-axis represents the unemployment rate as a percentage. The blue bars in the chart represent the urban unemployment rate in Indonesia for each year, and the green bars represent the rural unemployment rate.

  Key insights from this chart include the noticeable gap between urban and rural unemployment. Urban areas consistently have a higher unemployment rate compared to rural areas, suggesting that Indonesia's urban job markets face more challenges in absorbing labor than rural areas, where the informal sector, especially agriculture, often provides employment.

  A significant event shown in the chart is the impact of the COVID-19 pandemic in 2020, during which both urban and rural unemployment rates spiked. However, the effect was particularly severe in urban areas, reflecting the heightened vulnerability of cities to economic shocks. Post-pandemic data (2021–2023) indicates a decline in both urban and rural unemployment rates, signifying the start of Indonesia's economic recovery.

- Observation

  This chart highlights challenges in the urban job market, with urban areas showing consistently higher unemployment rates. This could be due to migration from rural to urban areas in search of better opportunities, or structural issues within urban industries that struggle to create sufficient employment. Rural unemployment rates, although lower, reflect the stability of agriculture and informal sectors, but also point to potential limitations in rural development and infrastructure. The impact of the pandemic, reflected in the spike in 2020, serves as a reminder of the vulnerability of urban economies to global crises.

For policymakers, this data can inform strategies to address regional employment disparities in Indonesia. Urban areas may require job creation programs, particularly in new or underdeveloped industries, to accommodate their growing populations, while rural areas may benefit from modernization initiatives in agriculture and infrastructure to further reduce unemployment. Additionally, the chart provides a clear view of Indonesia's post-COVID recovery, showing how different regions are bouncing back at varying rates.

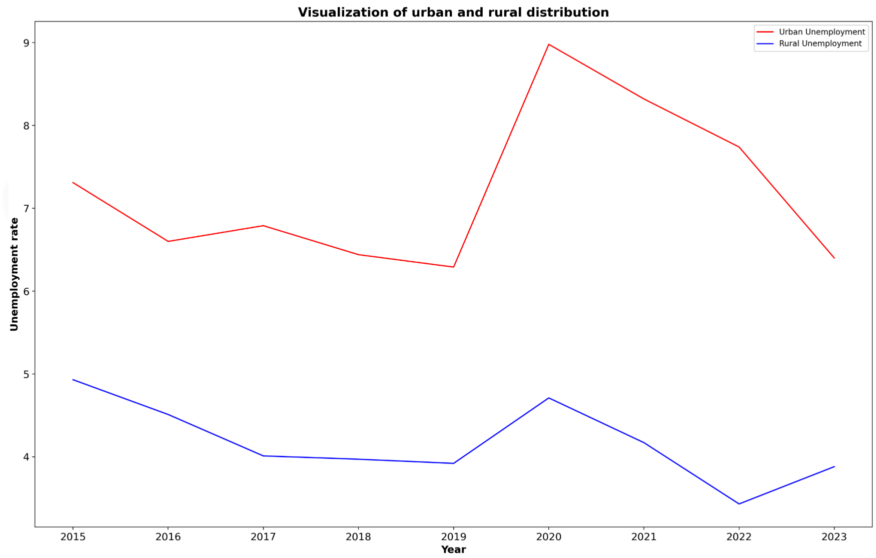
- How We Create

  To recreate this chart, the following Python libraries are required: Matplotlib for generating the bar chart, and Pandas to manage the data. Below is a Python code example to generate the chart:

```python
def read_unemployment_file(file_path):
    unemployment_data_list = []

    with open(file_path, 'r', encoding='utf-8-sig') as file:
        reader = csv.reader(file, delimiter=';')
        next(reader)
        next(reader)
        header = next(reader)[1:]

        for row in reader:
            if row[0]:
                area = row[0]
                years_data = {year: row[i+1].replace( __old: ",", __new: ".") for i, year in enumerate(header)}
                unemployment_data = UnemploymentData(area, years_data)
                unemployment_data_list.append(unemployment_data)

    return unemployment_data_list

file_path = "Unemployment Rate by Urban-Rural Classification.csv"
unemployment_data_list = read_unemployment_file(file_path)
for unemployment_data in unemployment_data_list:
    unemployment_data.display_data()
visualize_unemployment_bar_all_years(unemployment_data_list)
```

```python
import matplotlib.pyplot as plt
import numpy as np

2 usages
def visualize_unemployment_bar_all_years(unemployment_data_list):
    years = list(unemployment_data_list[0].years_data.keys())
    x = np.arange(len(years))
    width = 0.35

    urban_totals = []
    rural_totals = []

    for year in years:
        urban_total = sum(float(data.years_data[year]) for data in unemployment_data_list if "Urban" in data.area)
        rural_total = sum(float(data.years_data[year]) for data in unemployment_data_list if "Rural" in data.area)
        urban_totals.append(urban_total)
        rural_totals.append(rural_total)

    plt.figure(figsize=(12, 7))
    plt.bar(x - width/2, urban_totals, width, label='Urban', color='blue')
    plt.bar(x + width/2, rural_totals, width, label='Rural', color='green')
    plt.xlabel('Year')
    plt.ylabel('Unemployment Rate')
    plt.title('Urban vs Rural Unemployment Rate Over Years')
    plt.xticks(x, years)
    plt.legend()
    plt.tight_layout()
    plt.show()
```

5. Urban and rural distribution trendline graph



- Overview
  This chart provides a trendline to the unemployment rate in urban areas and rural areas. The data provides insight to analyze the correlation between urban and rural areas in terms of unemployment rate. Through this data, we can analyze what may cause the high urban unemployment rate with the low unemployment rate in rural areas.

- Chart description
  The X-axis represents the countries within the year from 2015-2023, while the Y-axis shows the unemployment rate, either in percentage terms or as a normalized value for easier comparison. Each line in this chart depicts the urban unemployment rate (in red) and rural unemployment rate (in blue).

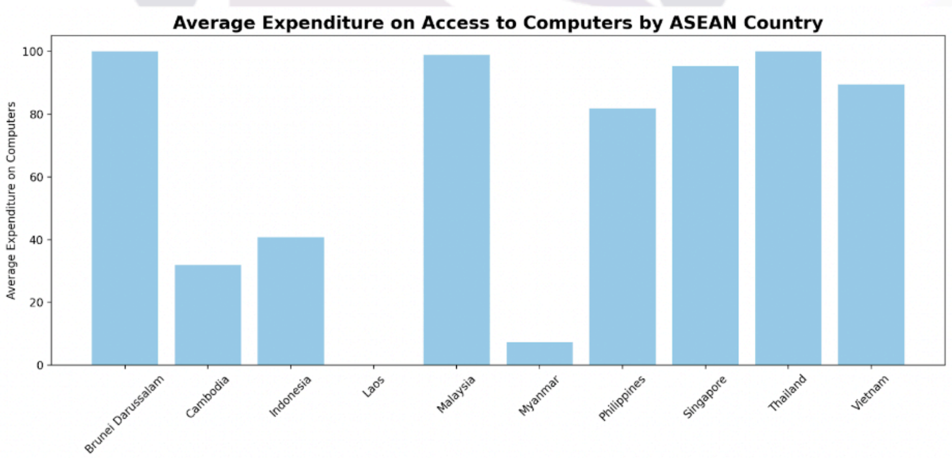
  From the chart, we may gain insight that urban unemployment rate is higher than rural unemployment rate. This means that there must be factors as to why this is happening

- Observation
  From the graph above, we can see that the urban trendline (red color) is slowly decreasing over the years. However, we can also see that the rural trendline (blue color) may increase after the year of 2023. This may be due to several factors as analyzed before, such as available job supplies and urban migration.

- How we create
  To recreate this chart, the following Python libraries are required: Matplotlib for generating the bar chart, and Pandas to manage the data. Below is a Python code example to generate the chart:

```
main_df = matrix_collation()
previous_urban_arr = main_df['Urban Unemployment'].values.flatten().tolist()
previous_rural_arr = main_df['Rural Unemployment'].values.flatten().tolist()
year_arr = main_df['Year'].values.flatten().tolist()
plt.plot( *args: year_arr, previous_urban_arr, color="red", label="Urban Unemployment")
plt.plot( *args: year_arr, previous_rural_arr, color="blue", label='Rural Unemployment')
plt.title( label: f"Visualization of urban and rural distribution", fontweight='bold', fontsize=16)
plt.xlabel( xlabel: f"Year", fontweight='bold', fontsize=13)
plt.ylabel( ylabel: f"Unemployment Rate", fontweight='bold', fontsize=13)
plt.xticks(fontsize=13)
plt.yticks(fontsize=13)
plt.legend()
plt.show()
```

6. Average Expenditure on Access to Computers by ASEAN Country



- Overview
  This chart illustrates the average expenditure on access to computers across various ASEAN countries. The data provides insights into the disparity in spending on technology access within the region, offering a comparative view of how different countries prioritize or afford computer access.

- Chart Description

  The X-axis represents the countries within the ASEAN region, while the Y-axis shows the average expenditure on access to computers, either in percentage terms or as a normalized value for easier comparison. Each bar in the chart corresponds to one ASEAN country, illustrating its relative expenditure on computer access.

  From the chart, we know that Brunei Darussalam and Malaysia exhibit the highest average expenditures on computer access, while Indonesia display relatively lower expenditure, suggesting either limited access to technology or fewer resources allocated to such purchases

- Observation
  Indonesia remains in the mid-to-low range in terms of expenditure on computer access among ASEAN countries, suggesting potential economic constraints or differing national priorities. Despite being one of the largest

economies in the region, Indonesia's investment in technological access appears modest compared to its peers. This may be attributed to several factors, including uneven infrastructure development across its vast archipelago, where rural areas often lack adequate digital connectivity.

Additionally, the government may prioritize other pressing socio economic issues, such as poverty reduction, education, and healthcare, over digital inclusion. While Indonesia is gradually investing in technology, its relatively lower expenditure on computer access indicates that more effort is needed to bridge the digital divide, particularly as the global economy becomes increasingly reliant on digital skills and infrastructure.

- How We Create

  This code leverages the Matplotlib library for creating bar charts and Numpy for handling numerical data efficiently.

```python
def read_expenditure_csv(file_path):
    expenditure_data_list = []

    with open(file_path) as csvfile:
        reader = csvfile.readlines()

        for row in reader:
            data_split = row.split(";")
            country = data_split[1]
            year = data_split[2]
            series = data_split[3]
            value = data_split[4].strip()
            if year >= "2012" and "Basic access to computers by level of education" in series:
                expenditure = EducationExpenditure(country, year, series, value)
                expenditure_data_list.append(expenditure)

    return expenditure_data_list

1 usage
def calculate_average_expenditure(expenditure_data_list):
    asean_countries = ['Brunei Darussalam', 'Cambodia', 'Indonesia', 'Laos',
                       'Malaysia', 'Myanmar', 'Philippines', 'Singapore',
                       'Thailand', 'Vietnam']

    country_expenditure = {country: [] for country in asean_countries}

    for expenditure in expenditure_data_list:
        if expenditure.country in asean_countries:
            country_expenditure[expenditure.country].append(expenditure.value)

    average_expenditure = {country: (sum(values) / len(values) if values else 0) for country, values in
                           country_expenditure.items()}
    return average_expenditure
```

```python
def visualize_average_expenditure(average_expenditure):
    countries = list(average_expenditure.keys())
    averages = list(average_expenditure.values())

    plt.figure(figsize=(12, 6))
    plt.bar(countries, averages, color='skyblue')
    plt.xlabel('Country')
    plt.ylabel('Average Expenditure on Computers')
    plt.title( label: 'Average Expenditure on Access to Computers by ASEAN Country', fontsize=16, fontweight='bold')
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()

expenditure_data_list = read_expenditure_csv("Public expenditure.csv")
average_expenditure = calculate_average_expenditure(expenditure_data_list)
visualize_average_expenditure(average_expenditure)
```

Before going any further for the user manual on how to generate the prediction graph using machine learning, this paper would like to give a brief insight of the machine learning used to predict the next data values. This paper will be mainly using linear regression, Auto Regressive Integrated Moving Average (ARIMA), and multiple regression to predict the coming decade.

First machine learning approach is linear regression. Linear regression, as suggested by the name, is to draw a linear relationship between an independent feature and a dependent target. The aim of the linear regression is to form a formula.

$$y = a + bx$$

If we were to plot these features and targets into a graph, the features would be the x-axis and the target would be the y-axis. The value of a will be the intercept and the value of b will be the slope. Through the formula, we will be able to predict the value of the target from the value of the feature that we want (Schneider et al., 2010, 777). This approach is implemented in fig 3.2.

Second machine learning approach is multiple regression or as known as multivariable linear regression. Different from linear regression, this type of machine learning aims to form a formula.

$$y = a + \sum_{i=1}^{n} b_i x_i$$

This type of linear regression uses n independent features and is added with the y-intercept a. The value of b represents the slope for each features and the value of y is the dependent target (Schneider et al., 2010, 779). Through this formula, we may be able to account for various features and adjust the target accordingly. This approach is implemented in fig 3.1.

Third machine learning approach is ARIMA. This model is different from linear regression as it depends solely on the past value and the current value of the target (dependent variable) to create a prediction based on timestamps accurately. This approach is most suited for observations based on time series that are correlated to the target statistically speaking (Permata & Habibi, 2023, 33). This approach is implemented in fig 3.3 and fig 3.4.

Before moving on to the graph implementation, there are some preprocessing steps that we need to collate all the information needed into one data table. The steps are as shown below.

First step is to read the filename “Unemployment Rate by Urban-Rural Classification” and “GDP GROWTH INDONESIA - WORLD BANK” using the pandas library. For the GDP per capita growth csv, we filter the country name to Indonesia and remove any unnecessary column from the dataframe. After that, we will filter all columns to only show years 2015 and above. Below is the code.

```python
import pandas as pd

year_arr = [str(year) for year in range(2015, 2024)]

def matrix_collation():
    # Read the matrix of unemployment rate by urban-rural classification
    unemployment_urban_rural_df = pd.read_csv("Unemployment Rate by Urban-Rural Classification.csv", delimiter=";")
    unemployment_urban_rural_df = unemployment_urban_rural_df.transpose()
    urban_arr = unemployment_urban_rural_df["Urban"].tolist()
    rural_arr = unemployment_urban_rural_df["Rural"].tolist()

    dataframe_data = {
        'Year': year_arr,
        "Urban Unemployment": urban_arr,
        "Rural Unemployment": rural_arr
    }
    unemployment_urban_rural_df = pd.DataFrame(dataframe_data)

    # Read the matrix of GDP Growth INDONESIA
    GDP_growth_world_df = pd.read_csv("GDP GROWTH INDONESIA - WORLD BANK.csv", delimiter=",")
    GDP_growth_indonesia = GDP_growth_world_df[(GDP_growth_world_df["Country Name"] == "Indonesia")].drop(columns=['Unnamed: 68'])
    GDP_growth_indonesia_per_year = GDP_growth_indonesia.loc[:, [year for year in GDP_growth_indonesia.columns if year.isdigit() and int(year) >= 2015]]
    GDP_growth_indonesia_per_year = GDP_growth_indonesia_per_year.transpose()
    GDP_growth_arr = GDP_growth_indonesia_per_year.values.flatten().tolist()
    dataframe_data = {
        "Year": year_arr,
        "GDP Per Capita Growth": GDP_growth_arr
    }
    GDP_growth_indonesia_df = pd.DataFrame(dataframe_data)

```

Second step is to define a function for the different levels of education, primary (SD), junior high school (SMP) and senior high school (SMA) and set it to return a dataframe containing the year and the level. After that, we will read three csv files based on those different levels and then merge them into one dataframe.

The final step is to collate all dataframe, urban-rural and education, to one main dataframe. For the second step and the final step, the code is as shown below.

```python
    # Read education df
    # education_df = pd.read_csv("BPS STATISTIK TINGKAT PENDIDIKAN.csv", delimiter=";")
    # education_df.to_excel("BPS STATISTIK EDUCATION.xlsx")
    def read_education_df_per_category(filename, type):
        file_df = pd.read_csv(filename, delimiter=";")
        average_arr_per_year = []
        for year in year_arr:
            average_value = file_df[year].mean()
            average_arr_per_year.append(average_value)

        column_2 = f"Average Education Completion {type}"

        dataframe_data_dict = {
            'Year': year_arr,
            column_2: average_arr_per_year
        }

        return pd.DataFrame(dataframe_data_dict)


    SD_education_completion_dataframe = read_education_df_per_category("BPS STATISTIK EDUCATION - SD.csv", "SD")
    SMP_education_completion_dataframe = read_education_df_per_category("BPS STATISTIK EDUCATION - SMP.csv", "SMP")
    SMA_education_completion_dataframe = read_education_df_per_category("BPS STATISTIK EDUCATION - SMA.csv", "SMA")
    SD_to_SMP = pd.merge(SD_education_completion_dataframe, SMP_education_completion_dataframe, on=['Year'])
    SD_SMP_SMA_df = pd.merge(SD_to_SMP, SMA_education_completion_dataframe, on=['Year'])

    # Collate all dataframe into 1 dataframe
    urban_rural_to_GDP = pd.merge(unemployment_urban_rural_df, GDP_growth_indonesia_df, on=['Year'])
    main_df = pd.merge(urban_rural_to_GDP, SD_SMP_SMA_df, on=['Year'])

    return main_df

```

Not only that, but we need to prepare the data frame for the unemployment dataframe. The first step to prepare it is to read the “unemployment analysis” csv file and then filter the dataframe to only show the country name, Indonesia. After that, we filter the year column to only show the years from 2012 and above. After that, we read the “ASEAN unemployment rate” csv file and remove any unnecessary columns. Before collating both data frames to complete the data frame, we need to preprocess the ASEAN unemployment rate to only show the unemployment rate total per year and then collate it with the main dataframe. The steps are as shown below.

```python
def find_indonesia_unemployment_rate():
    main_dataframe = pd.read_csv("unemployment analysis.csv", delimiter=",")

    indonesia_dataframe = main_dataframe[main_dataframe["Country Name"] == "Indonesia"]

    indonesia_main_df = indonesia_dataframe.loc[:, [year for year in indonesia_dataframe.columns if year.isdigit() and int(year) >= 2012]]

    with open("ASEAN unemployment rate.csv", "r+") as unemployment_file:
        # Skip the first two lines
        next(unemployment_file)
        next(unemployment_file)

        # Initialize the main dataframe
        main_dataframe = pd.read_csv(unemployment_file, delimiter=";")
        main_dataframe = main_dataframe.drop(columns=['Unnamed: 5'])

    indonesia_dataframe = main_dataframe[(main_dataframe['Region'] == "Indonesia") & (main_dataframe['Year'] == 2023)]
    indonesia_unemployment_df = indonesia_dataframe[indonesia_dataframe["Series"].str.contains('Unemployment rate')]

    average_per_year = indonesia_unemployment_df.groupby('Year')['Value'].mean().astype("str")

    indonesia_main_df[2023] = round(float(average_per_year[2023]),2)

    year_arr = list(indonesia_main_df.keys().astype("str"))

    unemployment_values_arr = indonesia_main_df.values.flatten().tolist()

    df = pd.DataFrame(unemployment_values_arr, index=year_arr)

    df = df.reset_index()

    df.columns = ['Year', 'Unemployment Rate']

    return df
```

After preparing the function, we are ready to predict!

7. Prediction on rural and urban unemployment rate: Linear regression

- Overview
  This chart illustrates two insights. One insight is the actual trendline for the current unemployment rate and the other is the predicted trendline for the coming years. This chart can provide insight for the predicted trend and the current trend to compare both values and validate each other.

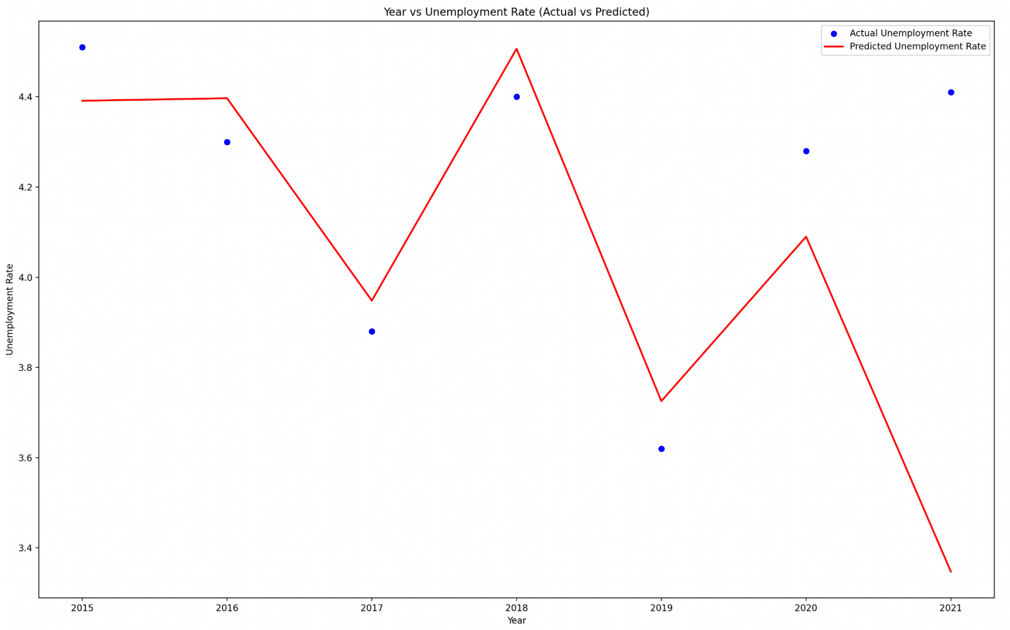
- Chart Description
  The x-axis represents the year, while the y-axis represents the unemployment rate. The blue line depicts the actual unemployment rate value trend and the orange line is depicted with the linear regression line/ the predicted trendline.

- Observation
  From this chart, we can see the actual unemployment trendline that is going up and down, but eventually decreasing in the long run. This is further supported by using the linear regression line, which shows the predicted trend and possible values for the coming years. This means that increased job supply can be one of the factors for the decreasing unemployment rate in the long run. Consequently, this means higher GDP per capita for Indonesia.

- How we create
  The graph is made possible using the linear regression machine learning approach. The steps are shown below.

```python
import numpy as np

from ML_PREDICTION_2 import *
import matplotlib.pyplot as plt
import sklearn.model_selection as ms
import sklearn.linear_model as lm

year_arr, unemployment_arr, main_df = first_analysis()

X_idx = np.array(main_df.index).reshape(-1, 1)
unemployment_df = pd.DataFrame(unemployment_arr)

X_train, X_test, Y_train, Y_test = ms.train_test_split( *arrays: X_idx, unemployment_df, test_size=0.5, random_state=0)

# Find the value of c and m
model = lm.LinearRegression()
model.fit(X_train, Y_train)

m = model.coef_.flatten().astype("float")
m = float(m[0])
c = model.intercept_.flatten().astype("float")
c = float(c[0])

def visualize(x_arr = year_arr, y_arr = unemployment_arr, coef = m, intercept = c):
    plt.title("Unemployment Rate and Prediction Trend in Indonesia")
    plt.plot( *args: x_arr, y_arr, marker='x')
    plt.xlabel("Year")
    plt.ylabel("Unemployment rate")
    x1 = np.linspace( start: 0,  stop: 10)
    y1 = intercept + coef * x1
    plt.plot( *args: x1, y1)
    plt.show()

visualize()
```

8. Prediction on rural and urban unemployment rate: Multiple regression



- Overview

  This trendline illustrates the actual unemployment rate of Indonesia as a whole and its predicted trendline. This data provides insight to the possibility of Indonesia's unemployment rate and the actual values of the unemployment rate.

- Chart Description
  The x-axis represents the year and the y-axis represents the unemployment rate as a percentage value. The scatter line in the graph represents the actual unemployment rate and the trendline with the red color is the multivariable regression line.

- Observation
  From the graph, we can see the various scores and the improved trendline by using education completion per level as features to predict the trendline. From there, we can see a correlation between education completion and unemployment rate as a whole in Indonesia without accounting for geography. This may mean that the decrease in unemployment rate may be caused by more individuals completing their studies at the bare minimum.

- How we create
  The graph is made possible using the multiple regression machine learning approach. The steps are shown below.

```
import pandas as pd

from MATRIX_MANIPULATION import matrix_collation, find_indonesia_unemployment_rate
import matplotlib.pyplot as plt
import sklearn.model_selection as ms
import sklearn.linear_model as lm
from sklearn.metrics import mean_squared_error

main_df = matrix_collation()
main_df = main_df.loc[:6]

unemployment_df = find_indonesia_unemployment_rate()
unemployment_df = unemployment_df.loc[3:9]

collation_matrix = pd.merge(main_df, unemployment_df, on=['Year'])

feature_X = collation_matrix[['Average Education Completion SD', 'Average Education Completion SMP', 'Average Education Completion SMA']]
target = collation_matrix['Unemployment Rate']

X_train, X_test, Y_train, Y_test = ms.train_test_split( *arrays: feature_X, target, test_size=0.2, random_state=10)

model = lm.LinearRegression()
model.fit(X_train, Y_train)

m = model.coef_.flatten().astype("float")
c = model.intercept_.flatten().astype("float")

Y_pred = model.predict(X_test)
# print(X_test)
# print(model_unemployment_pred_test)

error = mean_squared_error(Y_test, Y_pred)

score = model.score(X_test, Y_test)

# Step 1: Scatter plot of Year vs Actual Unemployment Rate
plt.figure(figsize=(8,6))
plt.scatter(collation_matrix['Year'].values.flatten().tolist(), target, label='Actual Unemployment Rate', color='blue')

# Step 2: Line plot showing the predicted values (if you have predictions)
y_pred = model.predict(feature_X)  # Assuming you've already trained the model and have predictions
plt.plot( *args: collation_matrix['Year'].values.flatten().tolist(), y_pred, color='red', lw=2, label='Predicted Unemployment Rate')

# Step 3: Customize plot
plt.xlabel('Year')
plt.ylabel('Unemployment Rate')
plt.title('Year vs Unemployment Rate (Actual vs Predicted)')
plt.legend()  # Adding a legend for clarity
plt.show()
```

9. Prediction on rural and urban unemployment rate: ARIMA

- Overview
  This trendline illustrates the unemployment rate in terms of the geographical location, urban and rural. The data provides insight as to what to expect for the coming decade for the urban areas and the rural areas.

- Chart Description
  The x-axis represents the year, while the y-axis shows the unemployment rate as a percentage value. Each line in the graph represents the unemployment rate trend. The urban unemployment rate is shown as the red line and the rural unemployment rate is shown as the blue line.

- Observation
  From fig 3.3, we can see both rural and urban unemployment rate values. Referring to fig 2.2, the predicted downward trend for urban unemployment rate can potentially be lower than rural unemployment rate probably in the next decade or more. Furthermore, the rural unemployment rate in the rural area may be of concern. This is because after the year 2026, the unemployment rate for rural areas has reached a plateau value. This may be caused by several factors as mentioned before in fig 2.2, this may be caused due to urban migration.

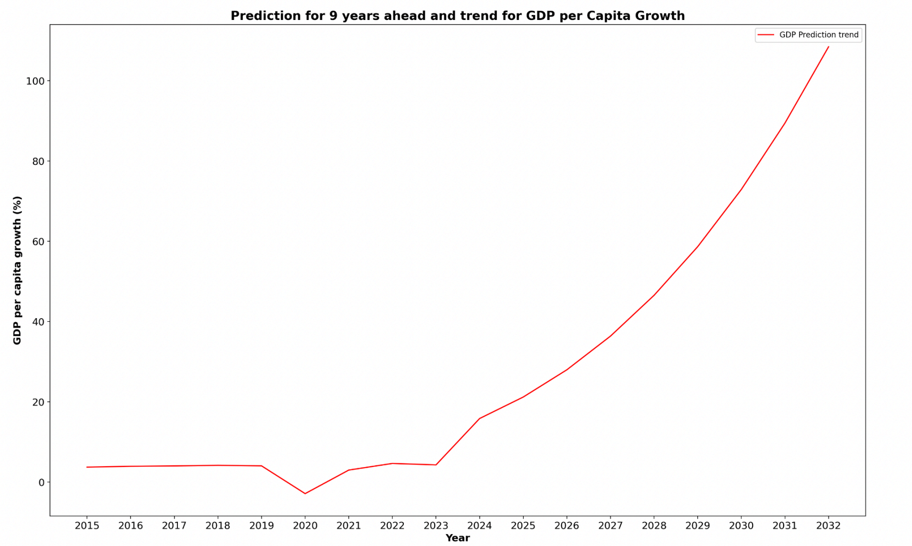
- How we create
  The graph is made possible using the ARIMA machine learning approach. The steps are shown below.

```python
import pandas as pd

from MATRIX_MANIPULATION import matrix_collation
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error

main_df = matrix_collation()

new_year_arr = [str(num) for num in range(2024, 2033)]

def find_value_of_d_model(key: str):
    column_of_interest = main_df[key]
    times_diff = 0

    while True:
        result = adfuller(column_of_interest)
        p_value = result[1]  # Get the p-value from the test result

        if p_value < 0.05:  # Check if the series is stationary
            # print(f"The series is stationary after {times_diff} differencing. {p_value}")
            break
        else:
            # Apply first difference and drop NaN values
            column_of_interest = column_of_interest.diff().dropna()
            times_diff += 1  # Increment the differencing count
            # print(f"Current p-value: {p_value}. Series is not stationary, applying differencing {times_diff}.")
    return times_diff

def visualize_for_pacf_and_acf_values(column_of_interest):
    fig, axs = plt.subplots(nrows: 1, ncols: 2, figsize=(12, 6))
    plot_acf(column_of_interest, ax=axs[0])

    plot_pacf(column_of_interest, ax=axs[1])

    plt.tight_layout()
    plt.show()

def model_score(column_of_interest, forecast):
    mse = mean_squared_error(column_of_interest, forecast)
    print(f"MSE: {mse}")


d_urban_value = find_value_of_d_model('Urban Unemployment')
urban_df = main_df['Urban Unemployment']

```

```python
model = ARIMA(urban_df, order=(1, 2, 1))
model_fit = model.fit()
forecase = model_fit.forecast(steps=9)

# For Urban unemployment, the p, d, q is 1, 2, 1

urban_forecast_data = {
    'Year': new_year_arr,
    'Urban Unemployment Prediction': forecase.values.flatten().tolist()
}
urban_prediction = pd.DataFrame(urban_forecast_data)

d_rural_value = find_value_of_d_model('Rural Unemployment')
rural_df = main_df['Rural Unemployment']
visualize_for_pacf_and_acf_values(rural_df)
# p = 1
# q = 1
model_2 = ARIMA(rural_df, order=(1, 0, 1))
model_2_fit = model_2.fit()
forecast = model_2_fit.forecast(steps=9)

# For rural unemployment, the p, d, q is 1, 0, 1

rural_forecast_data = {
    'Year': new_year_arr,
    'Rural Unemployment Prediction': forecast.values.flatten().tolist()
}
rural_prediction = pd.DataFrame(rural_forecast_data)

urban_rural_prediction = pd.merge(urban_prediction, rural_prediction, on=['Year'])

urban_arr = urban_rural_prediction["Urban Unemployment Prediction"].values.flatten().tolist()
rural_arr = urban_rural_prediction['Rural Unemployment Prediction'].values.flatten().tolist()

previous_urban_arr = main_df['Urban Unemployment'].values.flatten().tolist()
previous_rural_arr = main_df['Rural Unemployment'].values.flatten().tolist()
year_arr = main_df['Year'].values.flatten().tolist()

for values in urban_arr:
    previous_urban_arr.append(values)

for values in rural_arr:
    previous_rural_arr.append(values)

for values in new_year_arr:
    year_arr.append(values)

```

```python
main_dataframe = {
    "Year": year_arr,
    'Urban Unemployment': previous_urban_arr,
    'Rural Unemployment': previous_rural_arr
}

main_df = pd.DataFrame(main_dataframe)

def visualize_graph():
    plt.plot( *args: year_arr, previous_urban_arr, color="red", label="Urban trend")
    plt.plot( *args: year_arr, previous_rural_arr, color="blue", label="Rural trend")
    plt.title( label: f"Prediction for 9 years ahead and trend for Unemployment Rate", fontweight='bold', fontsize=16)
    plt.xlabel( xlabel: f"Year", fontweight='bold', fontsize=13)
    plt.ylabel( ylabel: f"Unemployment Rate", fontweight='bold', fontsize=13)
    plt.xticks(fontsize=13)
    plt.yticks(fontsize=13)
    plt.legend()
    plt.show()

visualize_graph()
```

10. Prediction on rural and urban unemployment rate: ARIMA



- Overview
  This chart illustrates the prediction for GDP per capita growth from the past 2023 and the predicted value for 2024-2032. The data provides insight as to what to expect for the coming decade and what may cause this uptrend.

- Chart Description
  The x-axis represents the year, while the y-axis shows the GDP per capita growth in percentage value. The red line represented in the graph illustrates the past and current value from 2015-2023 and the predicted trend in 2024 until 2032.

  From the chart, we can expect a positive uptrend for the coming decade though numerically speaking, we may not be even close to Singapore's GDP per capita.

- Observation
  From fig 3.4, we can see that Indonesia's potential for GDP per capita growth is exponentially high post-2024. Referring to fig 3.1, the predicted downward trend may be the main contribution to the increasing GDP per capita projected upward trend.

- How we create
  The graph is made possible using the ARIMA machine learning approach. The steps are shown below.

```python
import pandas as pd

from MATRIX_MANIPULATION import matrix_collation
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error

main_df = matrix_collation()

new_year_arr = [str(num) for num in range(2024, 2033)]

def find_value_of_d_model(key: str):
    column_of_interest = main_df[key]
    times_diff = 0

    while True:
        result = adfuller(column_of_interest)
        p_value = result[1]  # Get the p-value from the test result

        if p_value < 0.05:  # Check if the series is stationary
            print(f"The series is stationary after {times_diff} differencing. {p_value}")
            break
        else:
            # Apply first difference and drop NaN values
            column_of_interest = column_of_interest.diff().dropna()
            times_diff += 1  # Increment the differencing count
            print(f"Current p-value: {p_value}. Series is not stationary, applying differencing {times_diff}.")
    return times_diff

def visualize_for_pacf_and_acf_values(column_of_interest):
    fig, axs = plt.subplots( nrows: 1,  ncols: 2, figsize=(12, 6))
    plot_acf(column_of_interest, ax=axs[0])

    plot_pacf(column_of_interest, ax=axs[1])

    plt.tight_layout()
    plt.show()

def model_score(column_of_interest, forecast):
    mse = mean_squared_error(column_of_interest, forecast)
    print(f"MSE: {mse}")

```

```python
pred_model = ARIMA(column_of_interest, order=(1, 4, 1))
pred_model_fit = pred_model.fit()

forecast = pred_model_fit.forecast(steps=9)
new_GDP_arr = forecast.values.flatten().tolist()
previous_GDP_arr = column_of_interest.values.flatten().tolist()
previous_GDP_arr.extend(new_GDP_arr)

previous_year_arr = main_df['Year'].values.flatten().tolist()
previous_year_arr.extend(new_year_arr)

new_df = {
    "Year": previous_year_arr,
    'GDP Per Capita Growth': previous_GDP_arr
}
predicted_df = pd.DataFrame(new_df)

def visualize_graph():
    plt.plot( *args: previous_year_arr, previous_GDP_arr, color="red", label="GDP Prediction trend")
    plt.title( label: f"Prediction for 9 years ahead and trend for GDP per Capita Growth", fontweight='bold', fontsize=16)
    plt.xlabel( xlabel: f"Year", fontweight='bold', fontsize=13)
    plt.ylabel( ylabel: f"GDP per capita growth (%)", fontweight='bold', fontsize=13)
    plt.xticks(fontsize=13)
    plt.yticks(fontsize=13)
    plt.legend()
    plt.show()

visualize_graph()
```

Conclusion and Recommendations

- Conclusion

  In conclusion, there is a real relationship between Indonesia's unemployment rates, topographical features, financial circumstances, and educational attainment. These elements are intricately linked to one another, and they all have different effects on one another.

  Initially, unemployment rates are significantly influenced by geographical factors. Due to the rapid migration of people from rural to urban areas, unemployment rates are often greater, leading to an imbalance in job prospects. Economic inequalities are the primary cause of this rural-urban movement, with metropolitan areas providing more diverse industries and employment opportunities. However, this migration may result in the oversaturation of the urban labour market, driving up unemployment rates in metropolitan areas.

  Another important component that is closely linked to unemployment and economic growth is education. Higher educated communities typically have lower unemployment rates since their residents are more skilled and have the talents that modern industries need. On the other hand, because there are fewer qualified people available, rural areas with poorer access to high-quality education sometimes have greater unemployment rates. Higher levels of education typically translate into reduced unemployment in an inverse relationship between education and unemployment.

  Economic growth is also closely tied to both geographical factors and education. Regions with stronger economic performance, typically urban centers, tend to have more resources for education and job creation. This, in turn, fosters a more skilled workforce and a more stable job market. On the other hand, rural areas with weaker economies face challenges in improving education and creating sufficient job opportunities, leading to higher unemployment.

  In conclusion, the relationships between unemployment, geography, the economy, and education are distinct and complex. Long-term changes in employment and economic conditions require a comprehensive approach, as each component both influences and is influenced by the others, resulting in a complex web of interdependencies.

- Recommendations

  By referring to fig 3.3 and fig 2.2, the results show a downward trend prediction for people living in the urban areas. Through this data, it seems clear that the urban areas are performing better than rural areas. Possibly due to the large job space available in the urban areas and the steady job space available in the rural areas. One of the solutions to further decrease the unemployment rate in both rural and urban areas is to increase both job spaces to achieve results. Another factor to look at is urban migration. When people from rural areas move to urban cities, it may explain the plateau for rural unemployment, as explained in the analysis of fig 3.3

This is where AI comes into play. One of the proposed actionable solutions is to implement the idea of an AI-driven education and upskilling. This solution targets mainly rural youths, but can also be accessible to urban youths. We can implement this approach to e-learning platforms that provide customized educational programs. This approach can equip individuals with skills in demand that modern industries require. Not only does this aim to reduce rural unemployment rate, but also reduce urban migration that can potentially slow down the progress of reducing unemployment rate in the urban areas.

The limitations of implementing this approach may cause a technological dependence and rural areas unfamiliar with technology. As people get used to technology, especially in the rural areas, it may cause a risk of technological dependance where educational values may be looked down on. Furthermore, the cost of building an AI driven education platform is heavily costly. Moreover, rural areas that are used to traditional ways of living may be skeptical of this approach, thus may be causing resistance.

Further work for this approach is to use this learning platform in practice and introduce this approach physically to each rural area. At first, we may introduce this platform online, however, some rural areas have no access to the internet or may need convincing to use the platform. To tackle this problem, we can approach each rural area for those having technical difficulties and need convincing to introduce them to the AI driven learning platform to tackle the plateau of rural unemployment rate.

Referring to fig 1.3 and fig 3.4, this huge upward trend may not be equal to Singapore's GDP per capita growth. Although Indonesia may have the largest population, ideally this means greater GDP per capita growth for Indonesia. In fact, Singapore, with almost 100 times smaller the population than Indonesia, is able to generate such high GDP per capita, almost 100 times higher than Indonesia. This implies that the productivity rate in both countries are significantly different.

There may also be other factors intervening in this GDP growth, such as poor economic decisions, corruption (as suggested by the corruption possible index given to Indonesia), rural and urban unemployment (as suggested before) and education completion for primary, junior school, and high school. To tackle unemployment and education completion, it has been explained before to implement an AI driven learning platform. However, to tackle poor economic decisions and corruption, there are 2 solutions to this.

One solution is to implement AI in public policy and economic forecasting. This solution aims to introduce and/ or enhance decision-making for Indonesia's economy, especially in the economic planning and public policy, by predicting what will happen should the government implement policy changes on unemployment rate and GDP per capita growth. With this implementation, the government can make more effective job creation strategies to tackle the unemployment rate and make more effective economic decisions. This way, the prediction trend for the GDP growth for Indonesia can increase and therefore boost productivity and economy rate.

However, the downside of implementing this approach is that the government may be too reliant on AI predictions. Without the role of human consideration, choices made by AI may not achieve the results that we target. This may be caused by misinterpretations of AI-driven data or choices made by AI. Furthermore, not only do we not achieve the results that we target, but we may cause social problems.

Another solution is to implement AI in fraud detection systems in economic development funds. The software can adapt to new types of fraud by training the model with new datasets, although it could be slower to detect frauds in real-time. However, when implementing this approach, corruptors may find a loophole by minimizing transaction amounts that do not overlap the software's transaction limit, a group of people working together to "legally" extract money, thus bypassing the software's condition. Money-wise, this fraud detection software is costly and will be needing huge amounts of data to optimize the model.

Further work on this system is that we can also implement an AI system that can monitor the allocation and the utilization of government funds. This system targets to ensure that funds are properly allocated and utilized effectively for any government activities. In this case, for any and all youth employment and development programs to boost Indonesia's education and ideally, reduce the unemployment rate.

References

United Nations. (n.d.). GDP and GDP per capita. UNdata. Retrieved September 19, 2024, from https://data.un.org/

United Nations. (n.d.). Public expenditure on education. UNdata. Retrieved September 19, 2024, from https://data.un.org/

United Nations. (n.d.). Labour force and unemployment. UNdata. Retrieved September 20, 2024, from https://data.un.org/

World Bank. (n.d.). GDP per capita (current US$). World Bank Open Data. Retrieved September 20, 2024, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Pantanjali. (2022). Unemployment analysis [Data set]. Kaggle. Retrieved September 22, 2024, from https://www.kaggle.com/datasets/pantanjali/unemployment-dataset

Badan Pusat Statistik. (n.d.). Tingkat penyelesaian pendidikan menurut jenjang pendidikan dan provinsi. BPS - Statistics Indonesia. Retrieved September 23, 2024, from https://www.bps.go.id/id/statistics-table/2/MTk4MCMy/tingkat-penyelesaian-pendidikan-menurut-jenjang-pendidikan-dan-provinsi.html

Wikipedia contributors. (n.d.). List of ASEAN countries by GDP. *Wikipedia*. Retrieved September 24, 2024, from https://en.wikipedia.org/wiki/List_of_ASEAN_countries_by_GDP

Badan Pusat Statistik. (n.d.). Unemployment rate by urban-rural classification. *BPS - Statistics Indonesia*. Retrieved September 20, 2024, from https://www.bps.go.id/en/statistics-table/2/MTE3OCMy/unemployment-rate-by-urban-rural-classification.html

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, *107*(44), 776-782. 10.3238/arztebl.2010.0776

Permata, J. M. A. C., & Habibi, M. (2023). Autoregressive Integrated Moving Average (ARIMA) Models For Forecasting Sales Of Jeans Products. *Telematika: Jurnal Informatika dan Teknologi Informasi*, *20*(1), 31-40. 10.31515/telematika.v20i1.7868