

# **GDP Growth Prediction Analysis**

**Macro Economics for Data Science**



Jehan Nirmal

## Table of Contents

1. Introduction .....	2
1.1: Dataset .....	2
1.2: Data Exploration .....	3
1.3: Data Loading and Cleaning .....	3
2. Descriptive Statistics and Visualization.....	4
2.1: Descriptive Statistics .....	4
2.2: Visualization .....	5
2.2.2: Time series visualization of each variable .....	6
2.2.3: Visualization of correlation between each numerical variable .....	7
3. Analysis and Interpretation .....	8
3.1: OLS Regression Analysis (Relationship check) .....	9
3.3.1: OLS Regression Analysis on BI (\$Mn).....	9
3.3.2: OLS Regression Analysis on EI.....	11
3.2. Predicting upcoming next five-year GDP using regression .....	13
4. Conclusion and Communication .....	14

## Table of Figures

Figure 1 (Dataset Head) .....	3
Figure 2(Descriptive statistic on each numerical column) .....	4
Figure 3(Visualization of the distribution of each numerical variable) .....	5
Figure 4(Time series visualization of each variable) .....	6
Figure 5(Correlation Heat-map).....	7
Figure 6(OLS Regression Analysis on BI (\$Mn)) .....	9
Figure 7(Relationship of GDP with BI) .....	10
Figure 8(OLS Regression Analysis on EI) .....	11
Figure 9(Relationship between EI and GDP) .....	12
Figure 10(Prediction) .....	13
Figure 11(predicted values vs real values).....	13

# 1. Introduction

## 1.1: Dataset

The dataset offered includes data on several economic indices for a country. Subsidies granted per tax collected (ST), Exports to Imports ratio (EI), Per Capita income per worker hour (IH), Gross Domestic Product (GDP) in millions of dollars, and Business Investments (BI) in millions of dollars are among the metrics. These data points provide useful insights into the region's economic dynamics throughout this time period, focusing on factors such as government subsidies, trade balance, worker productivity, general economic output, and corporate investments. This information may be used to assist academics and policymakers in better understanding the economic patterns and possible variables impacting the region's economic health during this time period.

**ST (Subsidies per Taxes):** This reflects the ratio of government subsidies to taxes collected. It assesses how much the government subsidizes economic activity in relation to its tax income.

**EI (Exports to Imports):** EI is an abbreviation for the Exports to Imports ratio. It measures a country's balance of exports (goods and services supplied to other nations) and imports (goods and services purchased from other countries). A ratio larger than one implies that exports outnumber imports, whereas a ratio less than one indicates the reverse.

## 1.2: Data Exploration

Python was chosen as the programming language for the data exploration phase and all subsequent studies, and a Jupyter Notebook in Visual Studio Code (VS Code) served as a platform for processing the data, cleaning, and different analytical activities. For the above processes, I've imported the Python libraries numpy, pandas, matplotlib, seaborn, scipy, sklearn, statsmodels, itertools & and sklearn.

## 1.3: Data Loading and Cleaning

After loading the dataset I moved to the cleaning part for cleaning part first I checked for any duplicated values in the year column so that we could ensure that there were no repetitions in the year column and it came as zero. Then I checked for null values which also showed as zero, telling me that all the data points got a value. And with all that there's no need for any further data cleaning is required. Just to get a look below are the first five rows of the dataset.

	Year	ST	EI	IH	GDP (\$Mn)	BI (\$Mn)
0	1983	0.335	0.623	2.06	167.50	40.70
1	1984	0.330	0.616	2.10	298.40	78.65
2	1985	0.335	0.647	2.17	102.33	32.12
3	1986	0.330	0.652	2.21	124.24	38.14
4	1987	0.334	0.680	2.23	146.15	56.14

Figure 1 (Dataset Head)

## 2. Descriptive Statistics and Visualization

### 2.1: Descriptive Statistics

We use fundamental statistical metrics such as mean, median, and standard deviation to determine and define the main trends and variations in our data. We additionally use the graphical appeal of time series plots and histograms to create a clear view of trends, patterns, and distributions.

As mentioned above basic statistical measures for each numerical column is shown below.

	ST	EI	GDP (\$Mn)	IH	BI (\$Mn)
count	41.000000	41.000000	41.000000	41.000000	41.000000
mean	0.341317	0.905317	206.748293	2.631463	54.330512
std	0.017647	0.201472	76.969493	0.329815	15.125019
min	0.312000	0.616000	102.330000	2.060000	32.120000
25%	0.330000	0.729000	148.910000	2.470000	40.890000
50%	0.337000	0.874000	192.340000	2.630000	56.261000
75%	0.354000	1.034000	232.000000	2.810000	64.075000
max	0.397000	1.296000	498.180000	3.330000	95.445000

Figure 2(Descriptive statistic on each numerical column)

## 2.2: Visualization

### 2.2.1: Visualization of distribution of each numerical variable

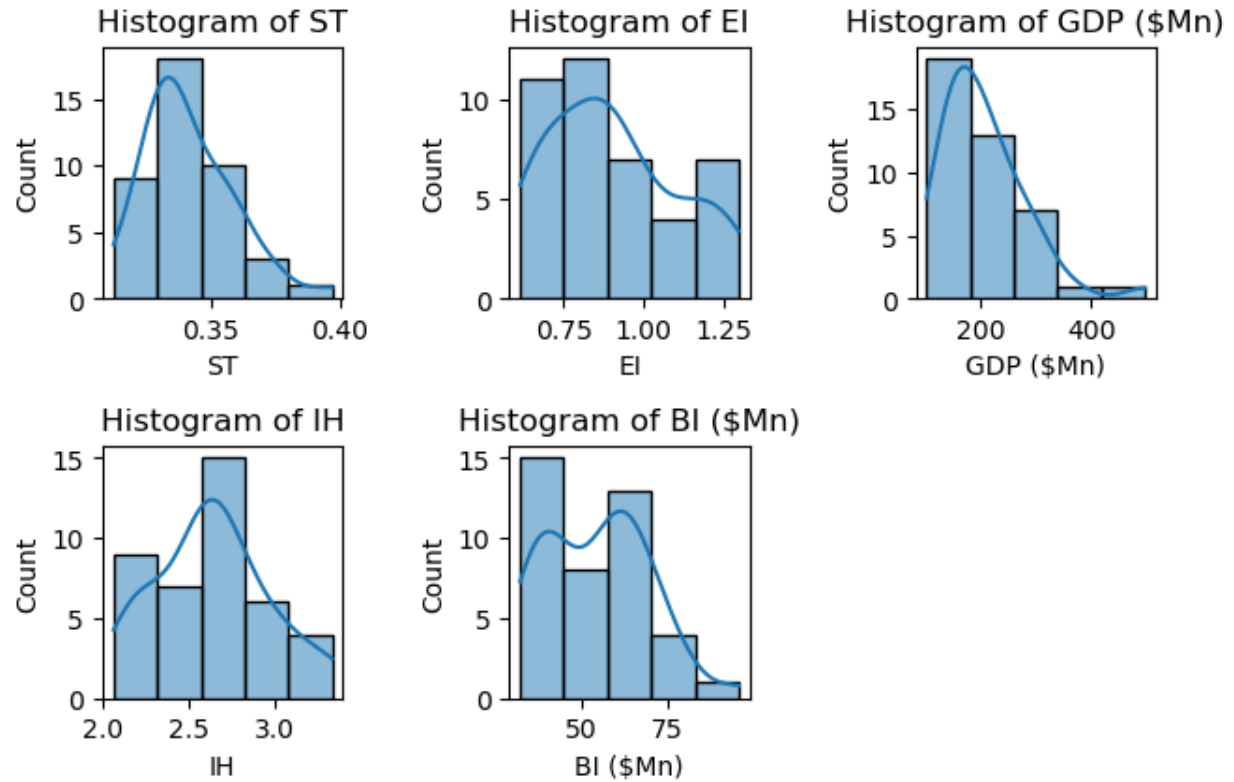


Figure 3(Visualization of the distribution of each numerical variable)

I've created the above set of histograms in order to visualize the numerical distribution of values of each numerical column so we can get an overall idea about the distribution,

## 2.2.2: Time series visualization of each variable

Since the given data is based on time yearly basis and the data set is about the macro economy of a country 'A' it's important to visualize the data according to the time. So, I've visualized the variation of each numerical variable with the year as a line plot.

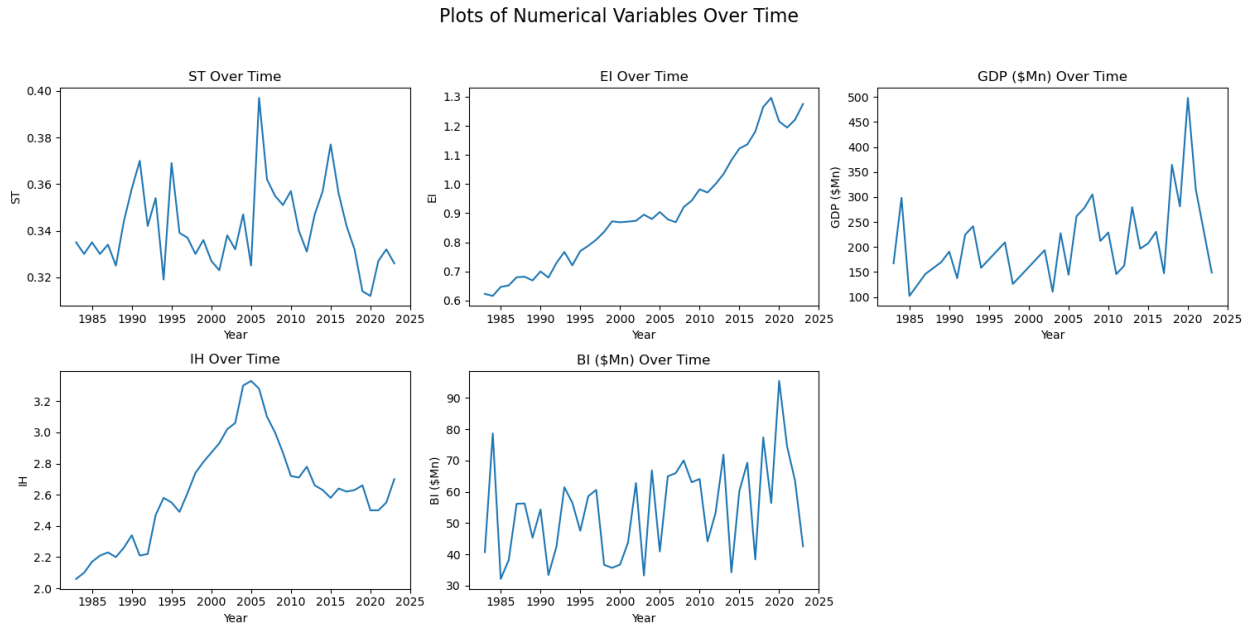


Figure 4(Time series visualization of each variable)

In the first chart, the variation of the ratio of Subsidies provided per tax collected is visualized over time and there is no specific trend while the second graph visualizes the ratio of Exports to Imports over time and it shows there's a huge increasing trend in the ratio over the time from 1985 to 2025. The third graph shows the variation in GDP value over time and the trend is not recognizable since there is no any flowing pattern. In the fourth graph, I've visualized the variation of Per Capita income per worker hour over time and it shows a recognizable peak in 2005 while in the fifth graph, I've visualized the variation of Business Investments over time within the given yearly range. *The main observation to be drawn from the above 5 graphs is that the variation of GDP is almost equal to the variation of BI over the same given time variation.*

### 2.2.3: Visualization of correlation between each numerical variable

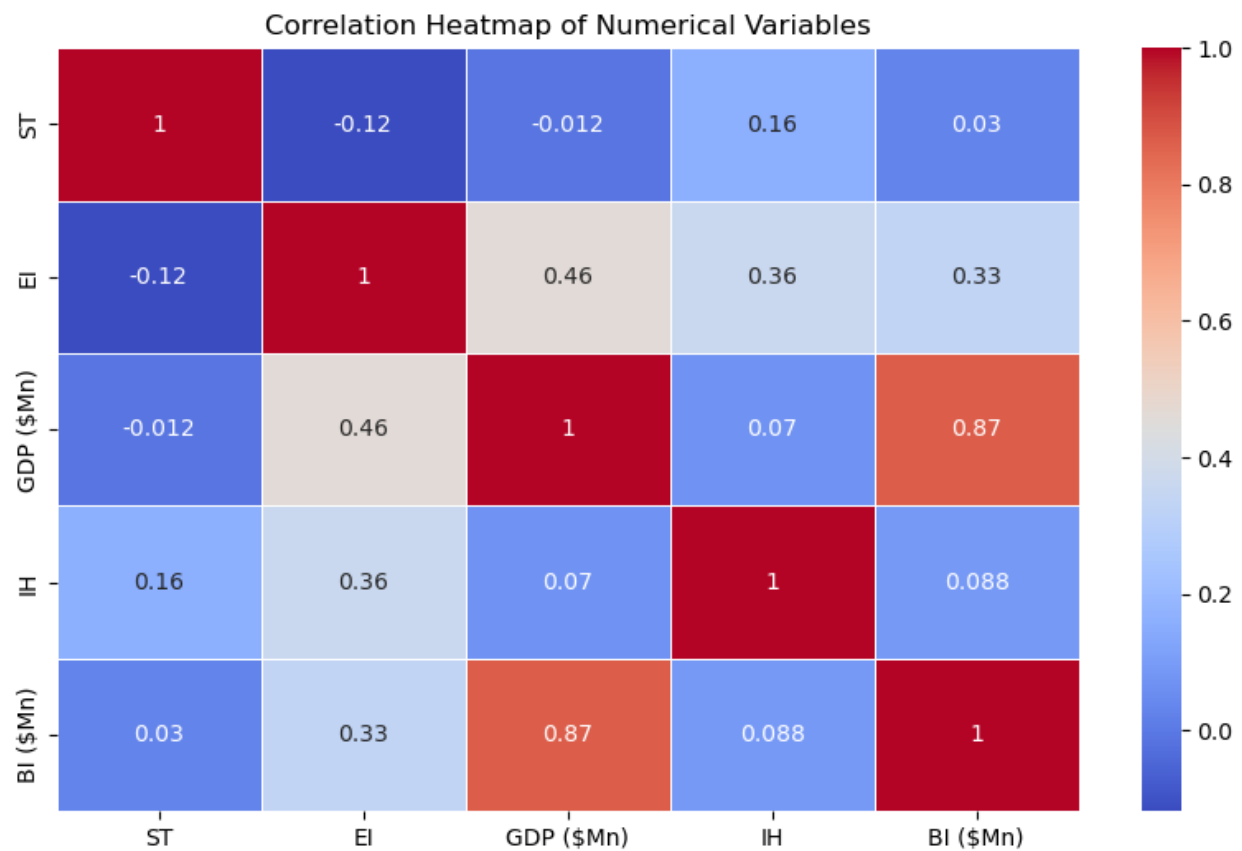


Figure 5(Correlation Heat-map)

Then as the next step in identifying the macroeconomic factors that affect the GDP, I've looked into the correlation of the GDP with all other numerical factors. Simply the correlation is the relationship between any two variables. So, in the above figure, each square represents a correlation value related to the variable from the X and Y axes.

So, let's start discussing the factors affecting the GDP from the highest correlation value of 0.87 which is with BI (\$Mn) where BI is the business investments. Where business investments have a greater impact on the GDP of a country. And the second highest GDP is with the Exports to Imports ratio (EI), which is 0.46. Then we have IH which is 0.07 and ST -0.012 which is a negative correlation with GDP.

When the correlation value is positive it means that when the value of one factor increases it affects the increase of the dependent factor while a negative correlation is a connection where when the value of one factor decreases it causes an increase in the value of the dependent variable.



### 3. Analysis and Interpretation

To answer the issue of whether there is a meaningful association between changes in investments and GDP growth, I choose to use Ordinary Least Squares (OLS) regression analysis as opposed to conventional hypothesis testing. OLS regression gives a more thorough and precise explanation of the connection than hypothesis testing, which can only provide a binary response to the study issue. OLS regression provides a wealth of detail that enables a deeper examination and interpretation of the data, ultimately leading to more robust and nuanced insights. It not only evaluates the relationship's significance but also quantifies its strength and direction.

Below are the main key points we look into when we get an OLS Regression Results sheet.

- Dependent Variable: *The dependent variable is the one we are attempting to explain or forecast.*
- R-squared: *R-squared measures how well the regression model fits the data.*
- F-statistic: *The F-statistic is a test statistic that aids in determining the regression model's overall significance. The model may be statistically significant if it has a higher F-statistic and a lower p-value.*
- Log-Likelihood: *Lower numbers indicate a better fit for the model, according to the log-likelihood metric, which measures the model's goodness of fit.*

By looking at above data from the results sheet we can get a much clear idea about the relationship between the data.

### 3.1: OLS Regression Analysis (Relationship check)

After selecting the main two variables ( *BI(\$Mn)* & *EI* ) that affect the GDP I've implemented an OLS regression in both of the variables and below I'm going to discuss each of them and how would they affect the GDP change.

#### 3.3.1: OLS Regression Analysis on BI (\$Mn)

OLS Regression Results						
=====						
Dep. Variable:	GDP (\$Mn)	R-squared:	0.753			
Model:	OLS	Adj. R-squared:	0.747			
Method:	Least Squares	F-statistic:	118.9			
Date:	Fri, 08 Sep 2023	Prob (F-statistic):	2.08e-13			
Time:	10:49:10	Log-Likelihood:	-207.08			
No. Observations:	41	AIC:	418.2			
Df Residuals:	39	BIC:	421.6			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	-33.1740	22.819	-1.454	0.154	-79.329	12.981
BI (\$Mn)	4.4160	0.405	10.905	0.000	3.597	5.235
-----						
Omnibus:	5.442	Durbin-Watson:	1.296			
Prob(Omnibus):	0.066	Jarque-Bera (JB):	4.148			
Skew:	0.720	Prob(JB):	0.126			
Kurtosis:	3.597	Cond. No.	213.			
=====						

Figure 6(OLS Regression Analysis on BI (\$Mn))

The association between business investments (\$Mn) and the gross domestic product (\$Mn) in millions of dollars is examined by the above-mentioned OLS regression analysis. With an R-squared value of 0.753, the model shows an excellent fit, with changes in business investments causing around 75.3% of the variance in GDP. The model's overall statistical significance is highlighted by the high F-statistic of 118.9 and the exceptionally low p-value, which confirms that business investments have a large influence on GDP. Additionally, the model's strong fit to the data is highlighted by the log-likelihood value, which is noticeably low at -207.08.

Now let's discuss how BI affects the GDP of a country,

Business investments (BI) often refer to the money spent by companies on infrastructure, machinery, and other productive assets. These investments are essential to the development and growth of a nation's economy. Changes in BI can have the following effects on Country A's GDP:

- **Economic Expansion:** *Increasing BI frequently triggers economic growth. Businesses can boost economic activity by making investments in new machinery, growing their activities, or constructing new facilities. As an example, employment is created through*

construction projects, and industry production is increased by investing in new equipment. As a result, Country A's overall GDP tends to increase.

- **Job Creation:** *BI may result in the creation of jobs. Businesses frequently employ additional personnel when they grow and invest in new initiatives. As a result, unemployment rates drop and consumer spending rises, which, in turn, drives economic growth.*
- **Multiplier Effect:** *The multiplier effect describes how BI's impacts may spread across the economy. When companies invest, they frequently buy supplies, services, and inputs from other industries. The multiplier effect, which boosts expenditure and economic growth, results in higher profits for suppliers and employees.*
- **Long-Term Growth:** *Increases in BI that are steady and long-lasting can result in long-term economic expansion. Businesses build a foundation for future economic growth and raise their GDP over time when they consistently invest in infrastructure, R&D, and innovation.*

Below is a scatter chart with trend highlights which I've created in order to visualize the relationship between BI(\$Mn) and GDP

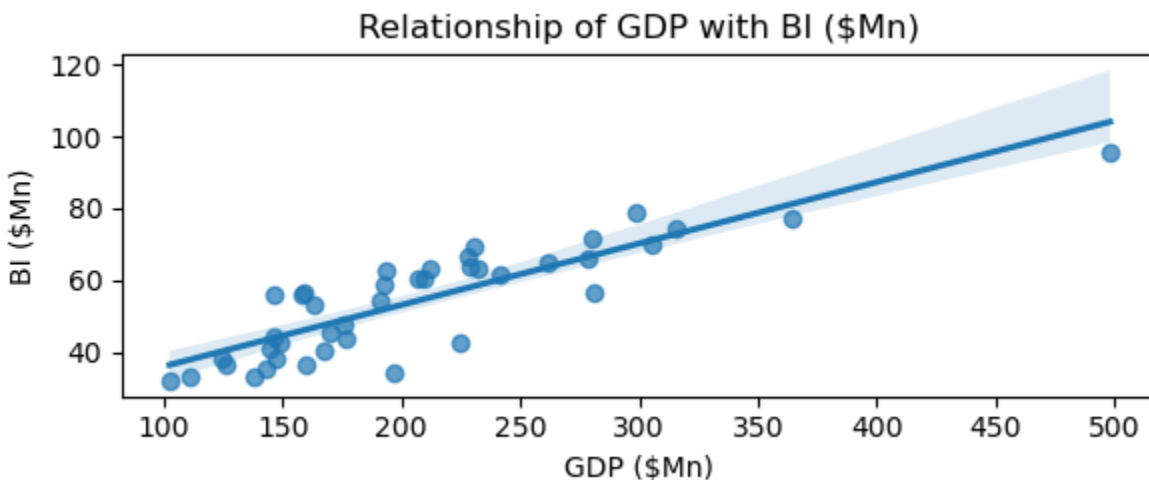


Figure 7(Relationship of GDP with BI)

The link between BI and GDP is not linear, and other elements, like governmental regulations, consumer spending, and general economic conditions, can also have an impact on GDP. Furthermore, BI's effectiveness and quality, as well as the business climate in Country A, are key factors in deciding how much of an influence it will have on GDP growth. Therefore, in order to maximize the beneficial influence on a country's GDP, it is crucial to promote a climate that is conducive to BI and ensure that investments are routed properly.

### 3.3.2: OLS Regression Analysis on EI

OLS Regression Results						
=====						
Dep. Variable:	GDP (\$Mn)	R-squared:	0.211			
Model:	OLS	Adj. R-squared:	0.190			
Method:	Least Squares	F-statistic:	10.40			
Date:	Fri, 08 Sep 2023	Prob (F-statistic):	0.00255			
Time:	10:49:56	Log-Likelihood:	-230.90			
No. Observations:	41	AIC:	465.8			
Df Residuals:	39	BIC:	469.2			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	48.0452	50.382	0.954	0.346	-53.863	149.953
EI	175.3011	54.354	3.225	0.003	65.360	285.243
=====						
Omnibus:	13.411	Durbin-Watson:	1.730			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	15.506			
Skew:	1.068	Prob(JB):	0.000430			
Kurtosis:	5.124	Cond. No.	9.23			
=====						

Figure 8(OLS Regression Analysis on EI)

The GDP (\$Mn) is the dependent variable in the results above, and the important independent variable is the exports-to-imports ratio (EI). According to the R-squared value of 0.211, variations in EI may account for around 21.1% of the variation in GDP. It implies a diffident correlation between EI and GDP, although not very strong. The model is statistically significant at a 0.25% significance level, as shown by the F-statistic of 10.40 and low p-value (0.00255), which together show that changes in EI have a large overall influence on GDP. In spite of being negative, the log-likelihood value shows that the model fits the data rather well. So, according to these results, variations in the Exports to Imports ratio (EI) are statistically significant in explaining variations in GDP.

Now let's discuss how EI affects the GDP of a country,

Exports to Imports Ratio: *The exports to imports ratio (EI) is an indicator of a nation's overall balance with international commerce. And also Exports to Imports Ratio is also called **Terms of Trade (TOT)**.*

$$\text{Terms of Trade (TOT)} = \frac{\text{Index of average export prices}}{\text{Index of average import prices}} \times 100$$

- Economic stability: *When exports overtake imports, there is a trade surplus, which is indicated by a positive EI. This extra revenue may help to stabilize the economy of Country A. It indicates that the nation is making more money from its foreign trade projects, which may have a favorable impact on its GDP.*

- *Demand Growing: An increase in the EI usually indicates an increase in consumer appetite for Country A's products and services on global marketplaces. In order to meet this demand, output may be expanded, which would raise GDP. Export-focused industries may grow, generating work and promoting economic expansion.*
- *Exchange Rates: A positive EI may also have an impact on exchange rates. Exports may become more expensive and imports more affordable when Country A sells more than it consumes. This may improve the international trade balance even further and promote economic expansion.*
- *Investment and Innovation: To remain competitive in international markets, domestic enterprises must invest in innovation and expansion as well as attract foreign investment. By raising output and productivity, this investment can promote economic growth.*
- *Diversity: A healthy EI promotes the global export marketplace's diversity. The nation's economy may become stronger due to changes in the world's economy, thereby lowering risk, by minimizing dependence on just one market.*

Below is a scatter chart with trend highlights which I've created in order to visualize the relationship between EI and GDP

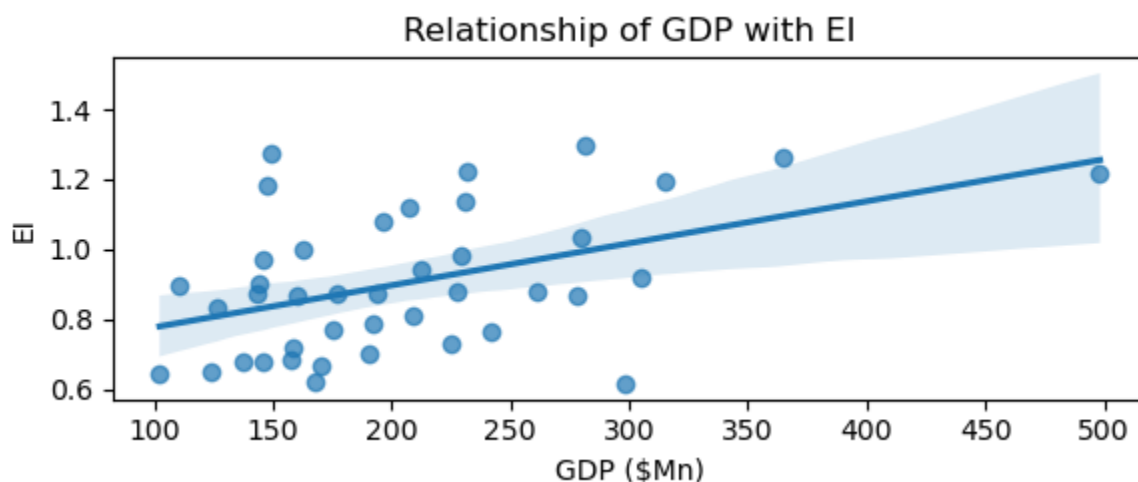


Figure 9(Relationship between EI and GDP)

It's necessary to remember that a favorable EI might have good benefits, GDP growth isn't just driven by it. The economic success of a nation is also significantly influenced by other variables, including governmental policies, consumption within the country, and international economic conditions. Additionally, to preserve the long-term advantages of a favorable EI for Country A's GDP and economic stability, maintaining a balanced and sustainable trade policy is essential.

### 3.2. Predicting upcoming next five-year GDP using regression

I did write a code in Python and through that I've created a regression model to forecast the increase of the Gross Domestic Product (GDP), a significant macroeconomic quantity, based on other pertinent economic variables. In the discipline of economic forecasting and analysis, where the goal is to understand and forecast the behavior of a complex economic system, this is a widely used technique. A linear association between GDP growth and variables like business investment (BI \$Mn) and external investment (EI) is established via the code's use of a linear regression model, a statistical approach. The dataset is divided into two subsets in order to guarantee the model's accuracy: a training set for teaching the model and a testing set for assessing its performance on hypothetical data. The ultimate objective is to accurately forecast future GDP growth.

Answers for the predicted next five years are as below:

```
Year 2024: Predicted GDP Growth = 255.10
Year 2025: Predicted GDP Growth = 217.74
Year 2026: Predicted GDP Growth = 105.90
Year 2027: Predicted GDP Growth = 276.18
Year 2028: Predicted GDP Growth = 198.12
```

Figure 10(Prediction)

I've created a chart in order to visualize the predicted values with the real values in the test set. So that relevant chart is shown below.

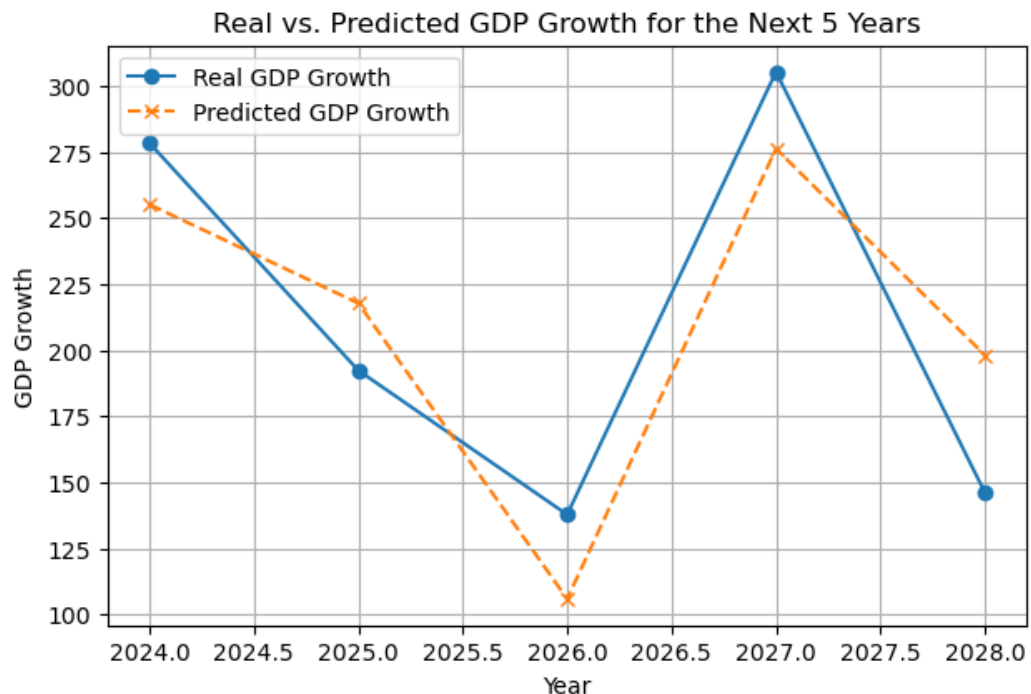


Figure 11(predicted values vs real values)

## 4. Conclusion and Communication

We looked at important economic variables to identify their effects on Gross Domestic Product (GDP) in our analysis of economic data for Country A. The collection includes statistics on things like business investments (BI), exports to imports ratio (EI), per capita income per worker hour (IH), and subsidies per tax collected (ST). We discovered a substantial positive association between business investments (BI) and GDP, demonstrating that higher investments frequently promote economic expansion. Higher BI is correlated with elements like employment creation, the multiplier impact, and long-term economic potential.

Even though the exports to imports ratio (EI) and GDP showed a positive association, the BI-GDP connection was stronger. However, a positive EI can promote investment, diversity, investment growth, improved exchange rates, and economic stability. Additionally, we developed a prediction model to project GDP growth over the following five years, which offers important insights into prospective economic trends.

The importance of business investments in promoting economic growth and the necessity of balanced trade policies to generate a positive EI must be emphasized while presenting these findings. These understandings can help stakeholders, entrepreneurs, and officials in Country A make wise decisions that will promote stability and economic growth.

\*\*\*\*