Data Driven Decision Making

Immigration. There is much talk that immigration hurts the United States. Is this true? Come up with a model that predicts the economic outcome of immigration. Use a training set to show how well it works. What are the critical factors?

A -

Our approach is divided in 3 parts -

1. **Descriptive Questions:** We first understand the macro level effects that can influence the economic growth of any country. Next, we understand the metrics used to measure the economic growth of the country to help us understand the impact of immigration in similar metrics. Third, we gather data for the important metrics and refine it.
2. **Comparative Questions:** There are two kinds of immigration, legal and illegal, and what is their respective effect on the US economy.
3. **Relationship:** Do any of the metrics impact each other? Based on the relationships, we create a predictive model that will answer the question of whether immigration hurts the US economy or not. More specifically, we shall look at the Immigration numbers provided in the Kaggle database and correlate them with the important metrics to answer the question.

# Factors affecting Economic Growth of US:

The factors that influence economic growth of a country are complex and multifaceted. Based on general economic principles, some key factors that influence a country's economic growth include:

1. Human Capital: The skills, knowledge, and productivity of the workforce
2. Institutional Factors: Stable government, rule of law, and effective economic policies.
3. International Trade: Access to global markets and participation in the global economy.
4. Entrepreneurship and Innovation: The ability to create new businesses and introduce new ideas.
5. Demographic Factors: Population growth, age structure, and trifurcation of society.
6. Macroeconomic Stability: Inflation, GDP

While immigration can play a role in economic growth, particularly through its impact on human capital and labor force growth, it is just one of many factors that contribute to a country's overall economic performance.

# Economic indicators for tracking economic growth of US:

Economic indicators are crucial metrics used to assess and track the economic growth of a country. Following are the key economic indicators for tracking economic growth:

1. **Gross Domestic Product** (GDP): GDP is the primary indicator of economic growth, measuring the total value of goods and services produced within a country's borders in a given period.
2. Employment Figures: Employment-related indicators provide crucial insights into the labor market and economic health.
3. Unemployment rate: A low rate can indicate a strong economy, while a high rate may suggest economic weakness.
4. Consumer Spending: It accounts for approximately two-thirds of the U.S. GDP and is a vital indicator of economic growth. It reflects the health of consumer spending and can help businesses anticipate demand for their products or services.
5. Industrial Production**:** This indicator measures the output of manufacturing-based industries, including those producing goods for consumers and businesses. It provides insights into the health of the manufacturing sector and overall economic activity.
6. Inflation: Inflation, often measured by the Consumer Price Index (CPI), reflects the general price level rise of goods and services in an economy. Monitoring inflation is crucial for businesses to adjust pricing strategies and account for rising costs.
7. Trade Balance: The trade balance measures the difference between a country's exports and imports. A positive trade balance, where exports exceed imports, can contribute to economic growth and improve a nation's position in international trade relationships.
8. Investment Rates: Investment rates play a crucial role in sustaining economic growth. Higher investment rates drive productive economic activities and can lead to long-term economic expansion.
9. Government Policies and Regulations**:** While not a traditional economic indicator, changes in government policies and regulations can significantly impact businesses and overall economic growth. It's important for businesses to stay informed about potential policy changes that could affect their operations and the broader economy.

From these economic indicators, we choose GDP as the dependent variable of the Immigrants Obtaining Lawful Permanent Resident Status, Refugee Arrivals, Noncitizen Apprehensions, Noncitizen Removals, and Noncitizen Returns independent variables.

# Trend analysis:

LPR’S:

The number of immigrants obtaining Lawful Permanent Resident (LPR) status in the United States has shown a general upward trend from 1980 to 2021, marked by significant fluctuations. In the 1980s, LPR numbers averaged around 600,000 annually, but the early 1990s saw a dramatic spike, peaking at 1,826,595 in 1991. After stabilizing above 800,000 per year in the late 1990s and early 2000s, another surge occurred in the mid-2000s, with over 1 million LPRs annually from 2005 to 2007. Throughout the 2010s, LPR numbers consistently hovered around 1 million per year, demonstrating a sustained high level of legal immigration. However, the COVID-19 pandemic significantly impacted these trends, causing a notable decrease to 707,362 in 2020, followed by a slight recovery to 740,002 in 2021. This overall pattern reflects the complex interplay of U.S. immigration policies, global events, and economic factors influencing migration over the past four decades.

Refugee arrivals:

Refugee arrivals in the United States have exhibited significant fluctuations over the past four decades. The early 1980s saw a peak of 207,116 arrivals in 1980, followed by a sharp decline. Numbers rebounded in the late 1980s and early 1990s, consistently exceeding 100,000 annually from 1989 to 1995. The late 1990s and early 2000s witnessed a general decline, with arrivals dropping below 70,000 by 2000. From the mid-2000s to mid-2010s, refugee arrivals stabilized, ranging between 50,000 to 85,000 annually. However, recent years have seen a dramatic decrease, particularly in 2020 and 2021, with only 11,840 and 11,454 arrivals respectively, likely due to the impact of the COVID-19 pandemic and changes in immigration policies.

Non citizens apprehensions:

Noncitizen apprehensions in the United States have shown dramatic fluctuations over the past four decades, reflecting changing immigration patterns and enforcement policies. The mid-1980s saw a significant surge, peaking at 1,767,400 apprehensions in 1986. This was followed by a period of decline and stabilization in the late 1980s to early 1990s, with annual numbers ranging from 1 to 1.3 million. Another surge occurred in the late 1990s to early 2000s, reaching a high of 1,814,729 in 2000. The 2000s witnessed a general downward trend, with apprehensions dropping below 1 million by 2007. The 2010s were characterized by fluctuating numbers, ranging from a low of 596,560 in 2015 to a high of 1,175,841 in 2019. Most recently, apprehensions saw a sharp drop to 609,265 in 2020, likely due to the COVID-19 pandemic, followed by a dramatic increase to 1,865,379 in 2021, the highest number recorded in the dataset.

Non citizen removals:

The trend analysis of non-citizen removals from 1980 to 2021 reveals significant fluctuations and notable patterns over the decades. In the early 1980s, removals were relatively low, starting at 18,013 in 1980 and gradually increasing to 34,427 by 1989. The 1990s saw a dramatic rise, with removals escalating from 30,039 in 1990 to 183,114 by 1999, reflecting heightened enforcement policies during this period. The trend continued upward into the 2000s, where removals reached a peak of 379,739 in 2009. The early 2010s marked the highest levels recorded, peaking at 432,201 removals in 2013, indicating a period of intensified immigration enforcement. However, starting in 2017, there was a noticeable decline in removals, dropping to 284,298 that year and further decreasing to just 89,191 by 2021. This recent decline can be attributed to various factors, including changes in immigration policy and the impact of the COVID-19 pandemic on enforcement operations. Overall, the data illustrates a complex interplay of immigration enforcement trends that have evolved significantly over the past four decades.

Non Citizen returns:

The trend analysis of non-citizen returns from 1980 to 2021 reveals significant fluctuations over the decades. In the early 1980s, returns started at 719,211 in 1980 and steadily increased, reaching over 1 million by 1985. The late 1980s saw a slight decline, but numbers surged again in the 1990s, peaking at 1,675,876 in 2000. This period likely reflected intensified border enforcement policies. The 2000s initially maintained high levels of returns, but a downward trend began mid-decade. This decline accelerated sharply after 2007, dropping from 891,390 to 471,798 by 2010. The trend continued downward throughout the 2010s, reaching a low of 100,454 in 2017. Recent years show a slight uptick, with returns increasing to 178,227 by 2021, though still far below historical peaks. This overall declining trend in non-citizen returns over the past two decades may reflect changes in immigration policies, enforcement priorities, or broader migration patterns.

# ARGUMENT:

Based on the trend analysis and economic indicators for tracking economic growth of the US, we plan to correlate the immigrants’ data attributes with GDP since the GDP is the best economic indicator of a country.

# DATA ANALYSIS:

Data Understanding:

The given dataset provides a comprehensive analysis of immigration status in the United States. The dataset has five attributes namely Immigrants obtaining LPR, Refugee arrivals, Noncitizen apprehensionship, Noncitizen removals, Noncitizens returns.

A new attribute has been added, namely GDP(Gross domestic product), in order to correlate the economic growth and how other independent variables are acting upon dependent variables(GDP).

Data preparation:

* The model is being split into testing (2011-2021) and training data( data from 1980 - 2010).
* The independent variables are LPR, Refugee arrivals, Noncitizen apprehensionship, Noncitizen removals, Noncitizens returns
* The dependent variable is GDP Data Modeling:
* CORRELATION ANALYSIS:

The correlation analysis is implemented using =CORREL() function in order to understand how each of the variables is being dependent on the independent variable.

|  |  |
| --- | --- |
| Correlation analysis |  |
| GDP VD LPR | 0.37 |
| GDP VD REFUGEE ARRIVALS | -0.64 |
| GDP VS NON CITIZEN APPREHENSIONS | -0.36 |
| GDP VS NON CITIZENS REMOVAL | 0.81 |
| GDP VS NON CITIZEN RETURNS | -0.70 |

GDP VD LPR

The weak positive correlation between GDP and Lawful Permanent Residents (LPRs) suggests that as the number of LPRs increases, there's a slight tendency for GDP to increase as well. This relationship, while not strong, can be explained by several factors like LPRs often join the workforce and contribute to economic productivity and output, they become consumers, potentially increasing overall consumer spending in the economy, and they bring diverse skills and expertise, potentially filling labor market gaps and contributing to innovation. Some LPRs may start businesses, creating jobs and contributing to economic growth. However, the weak correlation (0.37) indicates that this relationship is not strong or consistent, suggesting that other factors likely have more significant impacts on GDP growth.

GDP vs Refugee Arrivals

The moderate negative correlation between GDP and refugee arrivals is due to refugee arrivals requiring immediate government support for housing, healthcare, and education, potentially straining public resources in the short term. Refugees might face difficulties in quickly integrating into the workforce due to language barriers or skill mismatches, temporarily limiting their economic contributions. The positive economic contributions of refugees often materialize over time as they integrate, which may not be captured in short-term GDP figures.

GDP vs Non-Citizen Apprehensions

The weak negative correlation between GDP and non-citizen apprehensions might be due to increased apprehensions that may reflect higher spending on border control and immigration enforcement, potentially diverting resources from other productive economic activities. Apprehensions might reduce the available workforce in certain sectors, particularly those relying on undocumented labor, potentially impacting economic output. Higher apprehension rates could create uncertainty in industries dependent on immigrant labor, potentially affecting investment and growth.

GDP vs Non-Citizens Removal

The strong positive correlation between GDP and non-citizen removals is explained by Removals might coincide with periods of economic growth when there's less reliance on immigrant labor, rather than directly causing GDP increase. Increased removals may reflect broader policy changes that coincide with economic growth periods, rather than being a direct cause of GDP increase. Removal of undocumented workers might lead to job opportunities for documented workers, potentially resulting in higher reported economic activity.

GDP vs Non-Citizen Returns

The strong negative correlation between GDP and non-citizen returns might be due to Returns representing a reduction in the available workforce, potentially impacting industries reliant on immigrant labor. Departing non-citizens may lead to decreased local consumer spending, affecting economic activity. Economic downturns might coincide with increased returns as job opportunities decrease, rather than returns directly causing GDP decline.

Note: It's crucial to note that these correlations do not imply causation. The relationships between immigration factors and GDP are complex and influenced by numerous economic, social, and policy variables not captured in simple correlations.

* MULTIPLE LINEAR REGRESSION:

Model Equation:

After running the regression, we would get an equation like:

GDP = β0 + (β1 \* LPRs) + (β2 \* Refugee Arrivals) + (β3 \* Noncitizen Apprehensions) + (β4 \* Noncitizen Removals) + (β5 \* Noncitizen Returns)

Where β0 is the intercept, and β1 to β5 are the coefficients for each variable. Model Evaluation:

The R-squared value to determine how well the model fits the training data. And the values for the testing and training data are:

R SQUARE: 0.977041882 (Training data) R SQUARE: 0.966844660 (Testing data)

Since the R Square values for test and train dataset is 97.7% and 96.6% irrespectively, it indicates that the model generalizes well from the training data to the testing data, suggesting it's not overfitting.

Hence, the relationship between immigration factors and GDP is consistently strong across both datasets.

Testing the Model:

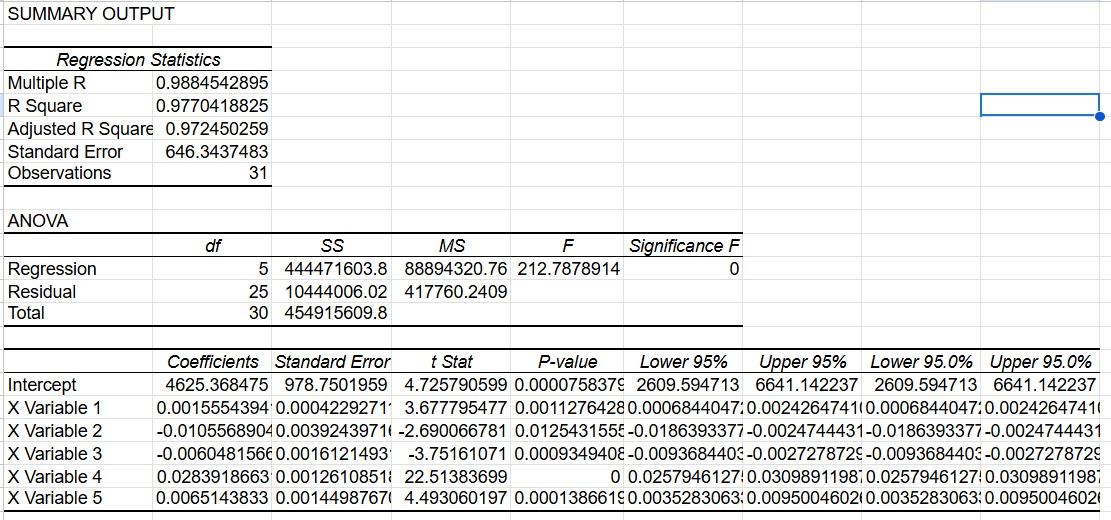
Apply the model to the training set(1980-2010) and testing set (2011-2021) using the Linear regression model.

Formula:

GDP = β0 + β1(LPRs) + β2(Refugee Arrivals) + β3(Noncitizen Apprehensions) + β4(Noncitizen Removals) + β5(Noncitizen Returns)

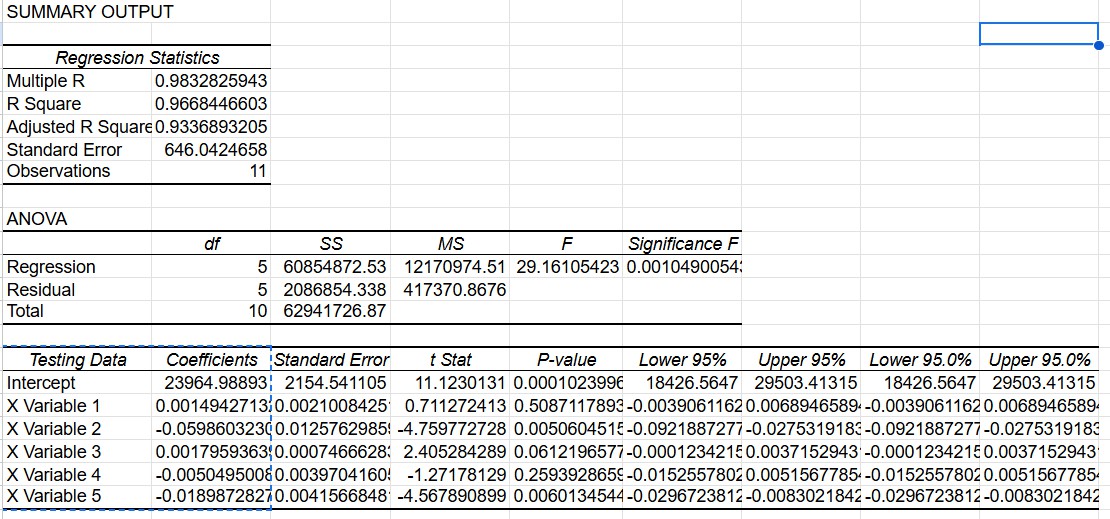
**For the Regression training dataset:**

|  |  |
| --- | --- |
| *Training Dataset* | *Coefficients* |
| Intercept | 4625.368475 |
| X Variable 1 | 0.001555439415 |
| X Variable 2 | -0.01055689045 |
| X Variable 3 | -0.006048156641 |
| X Variable 4 | 0.02839186631 |
| X Variable 5 | 0.0065143833 |

****

**For the Regression testing dataset:**

|  |  |
| --- | --- |
| *Testing Dataset* | *Coefficients* |
| Intercept | 23964.98893 |
| X Variable 1 | 0.001494271327 |
| X Variable 2 | -0.05986032303 |
| X Variable 3 | 0.001795936391 |
| X Variable 4 | -0.005049500854 |
| X Variable 5 | -0.01898728272 |

****

**Conclusion**:

We have developed a Correlation analysis model and Multi-linear regression analysis model to analyze the immigration dataset and compare it with the GDP to determine the economic outcome of the immigration.

As we can see from the Correlation analysis, the LPRs have a weak positive correlation with GDP suggesting that other factors likely have more significant impacts on GDP growth. The negative correlation of Refugee arrivals, non citizen apprehensions, and non citizen returns makes a case for the GDP supporting them and losing out on key talent during their return.

In the multi-regression analysis model, we trained it using the 2011-2021 dataset and applied the same on the remainder of the dataset. We found almost similar results in the R-squared values of the training and testing models.

R SQUARE: 0.977041882 (Training data) R SQUARE: 0.966844660 (Testing data)

Thus, we conclude that immigration does not hurt the US GDP in any major sense.

References:

1. Data to decode: The economics of being a business owner,

[https://www.jpmorgan.com/insights/economy/inflation/10-economic-indicators-ev](https://www.jpmorgan.com/insights/economy/inflation/10-economic-indicators-every-business-owner-should-know) [ery-business-owner-should-know](https://www.jpmorgan.com/insights/economy/inflation/10-economic-indicators-every-business-owner-should-know)

1. Economic Growth Indicators,

<https://hmarkets.com/learn-to-trade/macroeconomics/economic-growth/>

1. Top 10 U.S. Economic Indicators

[https://www.investopedia.com/articles/personal-finance/020215/top-ten-us-econo](https://www.investopedia.com/articles/personal-finance/020215/top-ten-us-economic-indicators.asp) [mic-indicators.asp](https://www.investopedia.com/articles/personal-finance/020215/top-ten-us-economic-indicators.asp)

1. US Leading Indicators

<https://www.conference-board.org/topics/us-leading-indicators>

1. Gross Domestic Product data

<https://www.bea.gov/data/gdp/gross-domestic-product>