Industry Investment Predictions by GDP & Income

Michael Speer, Elijah Taber, Caleb Trimble

May 25, 2025

# **Introduction**

Economic forecasting is essential for strategic investment, providing insights into market trends and industry shifts. By analyzing quarterly GDP proportions, this project aims to anticipate economic developments, highlight promising investment sectors, and assess market volatility and stability.

Predicting industry-level GDP trends allows investors to navigate downturns and seize emerging opportunities. Forecasting wages further enhances insight into labor market conditions and economic growth. By integrating these factors into five- and ten-year investment outlooks, this approach improves confidence in identifying industries poised for sustainable expansion. Through a data-driven methodology and risk assessment framework, this forecasting model refines decision-making and optimizes ROI, empowering investors to move beyond speculation toward a proactive strategy.

# **Business Problem/Hypothesis**

Investing without a clear understanding of future economic trends adds unnecessary risk. Traditional industry selection often relies on historical performance and subjective insights, failing to account for macroeconomic shifts and labor market fluctuations. Without a reliable forecasting model, investors risk allocating capital to declining industries, reducing returns and increasing portfolio volatility.

This project aims to predict industry-level GDP trends to guide investment decisions. By analyzing quarterly GDP proportions, forecasting growth, and incorporating wage trends, we provide a data-driven approach to identifying high-potential industries. A risk scoring system further refines uncertainty assessment, enabling smarter investment choices for 5- and 10-year forecasts.

# **Methods/Analysis**

The analysis leveraged quarterly U.S. GDP and wage data (2005–2024) from the BEA, focusing on 83 industries. Data preprocessing in Python (pandas, NumPy) standardized TimePeriod quartiles and IndustryDescription categories, with GDP proportions calculated to isolate sector-specific trends. Missing values were interpolated, and anomalies (e.g., 2020 downturns) were flagged to test model robustness. Wage data was aligned to GDP timelines to assess labor market linkages.

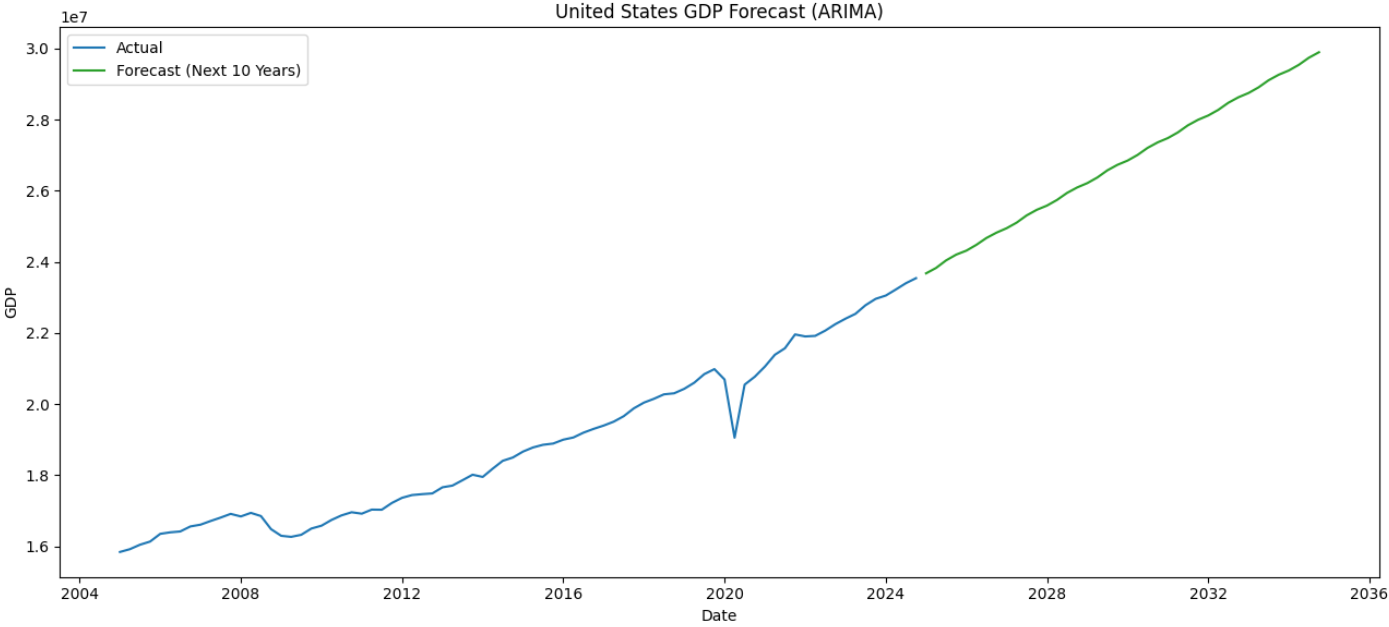
# 

**Figure 1: GDP growth trends by quarter showing consistent upward trajectory from 2005-2023, with quarterly distributions remaining stable across Q1-Q4 periods.**

Forecasts used ARIMA for long-term GDP trends and cyclical patterns, alongside Linear Regression for aggregate growth. Industry-level predictions employed a MultiOutput Regressor (scikit-learn) and individual ARIMA models, validated via walk-forward testing and MAPE. Risk scores integrated forecast variance, wage stability, and historical shocks to refine sector-specific uncertainty, resolving disparities (e.g., ARIMA’s 20.1% growth vs. MultiOutput’s contractions).

# **Results**

The GDP forecasting models provide insights into both overall U.S. economic growth and individual industry performance. The 10-year ARIMA forecast projects sustained expansion, offering a more nuanced perspective than linear regression, which struggles with anomalies like the 2020 downturn due to it high bias, low variance architecture.



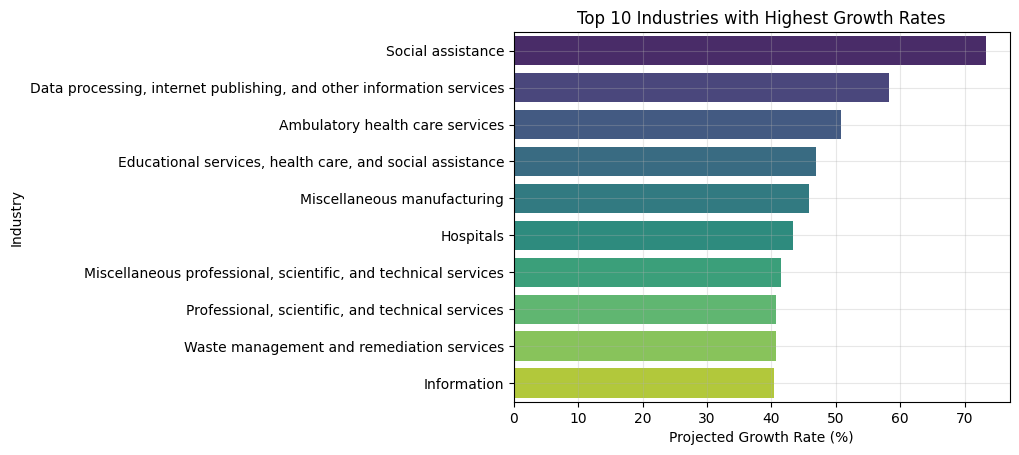
**Figure 2: Projected 10-Year U.S. GDP Growth (ARIMA Model). The figure shows the forecasted trajectory of the total U.S. over the next decade, based on ARIMA’s forecast.**

At the industry level, forecasts were generated using a MultiOutput Regressor and individual ARIMA models for 83 industries. The MultiOutput model suggests contractions in major sectors such as **Real Estate and Manufacturing**, with projected declines from $3.91T to $2.81T and $2.88T to $2.22T, respectively. In contrast, ARIMA forecasts a **more optimistic outlook**, with an average industry growth rate of **20.1%**, and **74 out of 83 industries** expected to expand.

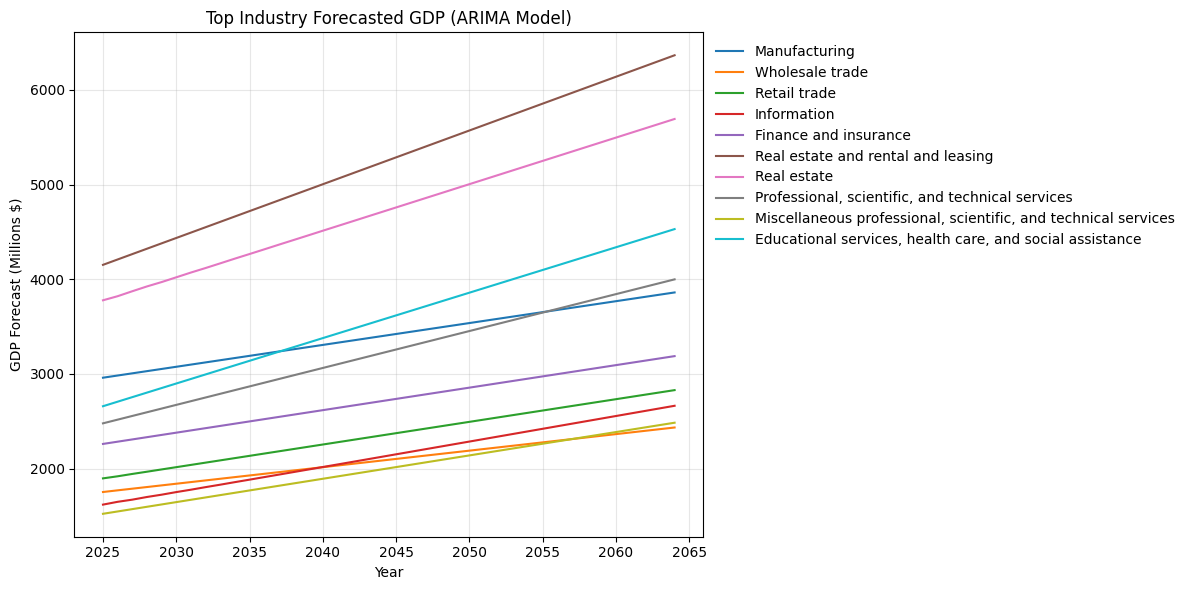
Notable growth projections include **Real Estate (34.5%)** and **Educational Services & Healthcare (46.9%)**, along with high-growth industries like **Social Assistance (73.3%)** and **Data Processing (58.2%)**. The contrast between models highlights uncertainties in industry trajectories, reinforcing the need for diversified investment strategies.

# **Recommendations**

Industry-level forecasts highlight sectors with strong growth potential, particularly those projected by ARIMA models. High-growth industries such as **Social Assistance (73.3%)**, **Data Processing & Internet Services (58.2%)**, and **Ambulatory Health Care (50.8%)** present substantial investment opportunities, driven by demographic shifts, technological advancements, and evolving societal needs.

**Figure 3: Top 10 Industries with Highest Growth Rates. This figure shows which industries have the highest projected growth by percentage.**

Beyond percentage growth, industries with large projected GDP values offer stability and strong economic contributions. **Real Estate and Rental Leasing ($5.26T), Real Estate ($4.73T),** and **Educational Services, Health Care & Social Assistance ($3.59T)** remain foundational pillars of the U.S. economy. Effective investment strategies should balance high-growth emerging opportunities with established sectors for long-term stability.

**Figure 4: Top Industry Forecasted GDP (ARIMA Model). This figure shows the forecasted increase in GDP for the top industries for the next 40 years.**

These recommendations are based on model-driven forecasts, acknowledging that real-world economic conditions and unforeseen events may impact actual outcomes.

# **Conclusion**

This project explored multiple time series forecasting models to predict U.S. GDP trends and industry performance over the next decade. Linear Regression and ARIMA models provided insights into aggregate GDP, while MultiOutput Regression and individual ARIMA models assessed 83 industries. The findings indicate **continued GDP growth**, with ARIMA models generally projecting **stronger sector-specific expansion** compared to the MultiOutput Regressor.

Key industries expected to thrive include **Social Assistance, Data Processing, and Ambulatory Healthcare**, highlighting promising areas for investment and policy considerations. While forecasting carries inherent uncertainties, the use of multiple methodologies provides a **data-driven framework** for evaluating economic trends. The divergence between models underscores the need for **multi-perspective analysis** when predicting complex economic systems.

# **References**

*GDP (current US$) - United States*. World Bank Open Data. (n.d.). <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US>

U.S. Bureau of Economic Analysis, "[Value Added by Industry](https://apps.bea.gov/iTable/?reqid=150&step=2&isuri=1&categories=gdpxind#eyJhcHBpZCI6MTUwLCJzdGVwcyI6WzEsMiwzXSwiZGF0YSI6W1siY2F0ZWdvcmllcyIsIkdkcHhJbmQiXSxbIlRhYmxlX0xpc3QiLCIxIl1dfQ==)" (accessed Sunday, May 18, 2025).

U.S. Bureau of Economic Analysis, "[Table 6.3A. Wages and Salaries by Industry](https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey&_gl=1*14pf2ql*_ga*ODU1MjA0OTQ4LjE3NDcxNTM3MDA.*_ga_J4698JNNFT*czE3NDc2MTAwMDMkbzMkZzEkdDE3NDc2MTAzMzMkajYwJGwwJGgwJGR0WVZtNDREVWttcEVVRl9GdGlQMk1ta1JWQmItVFY0UUZB#eyJhcHBpZCI6MTksInN0ZXBzIjpbMSwyLDNdLCJkYXRhIjpbWyJjYXRlZ29yaWVzIiwiU3VydmV5Il0sWyJOSVBBX1RhYmxlX0xpc3QiLCIxODYiXV19)" (accessed Sunday, May 18, 2025).

# **Appendix**

**Data Source 1**

* **Dataset**: GDPbyIndustry
* **Access Date**: May 2025
* **URL**: https://apps.bea.gov/iTable/?reqid=150&step=2&isuri=1&categories=gdpxind#eyJhcHBpZCI6MTUwLCJzdGVwcyI6WzEsMiwzXSwiZGF0YSI6W1siY2F0ZWdvcmllcyIsIkdkcHhJbmQiXSxbIlRhYmxlX0xpc3QiLCIxIl1dfQ==

**Data Source 2**

* **Dataset:** Wages and Salaries by Industry
* **Access Date:** May 2025
* **URL:** https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey&\_gl=1\*14pf2ql\*\_ga\*ODU1MjA0OTQ4LjE3NDcxNTM3MDA.\*\_ga\_J4698JNNFT\*czE3NDc2MTAwMDMkbzMkZzEkdDE3NDc2MTAzMzMkajYwJGwwJGgwJGR0WVZtNDREVWttcEVVRl9GdGlQMk1ta1JWQmItVFY0UUZB#eyJhcHBpZCI6MTksInN0ZXBzIjpbMSwyLDNdLCJkYXRhIjpbWyJjYXRlZ29yaWVzIiwiU3VydmV5Il0sWyJOSVBBX1RhYmxlX0xpc3QiLCIxODYiXV19

**Key Variables Used in Analysis**

* **GeoName**: State
* **TimePeriod**: Annual GDP broken down into Quartiles (2005-2024)
* **IndustrYDescription**: Industries broken down into 99 different categories
* **Frequency**: Quarterly results for GDP by industry
* **LineDescription**: Industries broken down into types of industries (Distributive, Goods-producing, etc.)
* **TimePeriod:** Quarterly income (1947-2000) by industry type.

**Tools & Libraries**

* Python: pandas, matplotlib, seaborn, scikit-learn, numpy, urllib, plotly
* Jupyter Notebook environment