Contents

[Introduction 1](#_Toc158384946)

Strategic Overview of the Business Problem [1](#_Toc158384947)

Project Plan [2](#_Toc158384947)

Business Understanding [3](#_Toc158384947)

Data Understanding [3](#_Toc158384947)

Data Preparation [4](#_Toc158384947)

Findings & Recommendations [5](#_Toc158384947)

Visualizations [7](#_Toc158384947)

[Conclusion 11](#_Toc158384950)

References  [12](#_Toc158384947)

# Introduction

This report Is the second stage of our year-long capstone project, specifically addressing the Data Understanding and Data Preparation phases of the project as outlined in the CRISP-DM methodology. The initial goals at this stage are to understand the business problem we are trying to address, outline the project management methodology we will be using, and to begin understanding and preparing our chosen dataset.

For this project on Foreign Direct Investment (FDI) risk analysis, I will be Using historical GDP growth data from the World Bank (1960–2023) to explore trends, patterns, and anomalies across EU economies. This stage lays the groundwork for deriving key metrics such as economic volatility, and establishes an initial Stability Index, which will be used to classify countries into low, medium, and high-risk investment categories in subsequent phases of the project.

By documenting these phases in detail, the report provides transparency into the methods and tools used to clean, transform, and prepare the dataset for further analysis. This ensures that subsequent findings and recommendations are rooted in extensive data preparation, enhancing the credibility and accuracy of the overall project outcomes. The focus here is not only on identifying growth and stability trends, but also on establishing a repeatable process for economic analysis that aligns with the broader objectives of the capstone project.

# Strategic Overview of the Business Problem

Foreign Direct Investment (FDI) decisions are shaped by the need to balance growth potential with manageable risks. The European Union (EU), a region of economic diversity, exemplifies the challenges and opportunities for investors (World Bank, 2023). Stable economies, such as Germany and Belgium, are attractive for their predictable returns, while emerging markets like Latvia and Croatia offer higher growth potential but increased volatility.

GDP growth is a critical indicator of economic performance (OECD, 2023), but excessive volatility can signal instability, reducing the appeal of a market for FDI. For example, the 2008 financial crisis and the 2020 COVID-19 pandemic exposed vulnerabilities in high-risk economies like Greece, while more stable countries like Poland and Malta demonstrated resilience (IMF, 2020). Identifying such patterns is essential for effective resource allocation.

This report addresses the strategic need for FDI planning by analysing GDP growth trends and volatility across EU member states from 1960 to 2023. The Stability Index, which combines average GDP growth with volatility, provides a framework for categorizing countries into low, medium, and high-risk investment opportunities. This approach supports investors in navigating the complexities of the EU’s economic landscape by aligning economic performance metrics with strategic decision-making.

The analysis aims to deliver actionable insights, guiding long-term investments in stable markets and highlighting strategies for managing risk in more volatile economies.

# Project Plan

This report is a key component of the broader capstone project, which follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology (Chapman et al., 2000). CRISP-DM provides a structured framework for managing all stages of the project, from problem identification to deployment, ensuring a systematic approach to achieving the capstone's objectives.

**CRISP-DM in the Capstone Project**

*Business Understanding:*

* Define the overarching problem: Assess economic stability in EU member states to guide Foreign Direct Investment (FDI) decisions.
* Objectives: Classify countries into risk categories using data-driven insights.

*Data Understanding:*

* Collect and explore historical GDP data for trends, patterns, and anomalies. This report focuses on documenting this phase.

*Data Preparation:*

* Clean, transform, and engineer the dataset, including metrics like volatility and the Stability Index. These steps are detailed in this report.

*Evaluation (future phase):*

* Validate findings from this analysis and integrate them into broader capstone deliverables, such as strategic models and future reports.

*Deployment (future phase):*

* Present insights as actionable recommendations for stakeholders in the context of FDI strategies.

**Report’s Role**

This report addresses the Data Understanding and Data Preparation phases, providing a foundation for subsequent modelling and deployment efforts in the capstone project.

**Timeline and Execution**

The capstone project spans the academic year, with this report representing the second stage of the project. As I mentioned in my project proposal, the timeline for project completion is as follows:

1. Data collection and treating: **3-4 weeks**
2. Building the risk assessment model: **4-5 weeks**
3. Testing the model and examining results: **3-4 weeks**
4. Finalizing the project and displaying results: **4-5 weeks**

Following the CRISP-DM methodology, we are currently in the Data Preparation stage, or stage one of the initial timeline. This phase ensures the dataset is clean, structured, and ready for analysis, setting the foundation for model development and subsequent evaluation in the coming weeks.

# Business Understanding

Economic stability is a cornerstone of Foreign Direct Investment (FDI) decisions (OECD, 2023). Countries with consistent GDP growth and low volatility provide predictable environments for investors, minimizing risks and maximizing returns. Conversely, high volatility signals economic instability, susceptibility to external shocks, or structural weaknesses, making investments riskier and less predictable.

The European Union (EU) exemplifies these dynamics, offering a mix of highly developed nations alongside emerging markets. Stable economies often attract long-term investments due to their resilience, while volatile markets may present higher potential returns but also significant risks (Eurostat, 2021). The risk of possible future crises makes this a challenging issue for potential investors.

This report focuses on analysing GDP growth trends and volatility across EU member states from 1960 to 2023. The objective is to classify countries into low, medium, and high-risk investment categories using a Stability Index, a metric combining average GDP growth and volatility. This analysis aims to answer the following questions:

1. Which EU countries provide the most stable environments for FDI?

2. How do global crises impact GDP growth across member states?

3. What strategies can investors adopt to navigate high-risk and medium-risk markets?

By addressing these questions, the report provides actionable insights to guide FDI strategies. It serves as a critical step in the broader capstone project, enabling data-driven decision-making and aligning economic analysis with investor priorities.

# Data Understanding

**Dataset Overview**

* Source: The dataset is sourced from the World Bank (World Bank, 2023) and includes historical GDP growth data for over 250 countries from 1960–2023.
* Attributes:
  + Country Name
  + Year: Annual data points from 1960 to 2023.
  + GDP Growth: Annual percentage change in GDP, representing economic performance.

For the purposes of this report, I filtered the dataset to only include EU member states. This filtered dataset provides a comprehensive historical view of economic trends, enabling analysis of both short-term and long-term stability across EU countries.

**Exploratory Data Analysis (EDA)**

* Trends:
  + Stable economies such as Germany and Belgium exhibit consistent GDP growth over decades.
  + Emerging markets like Latvia and Croatia show higher volatility, with growth rates fluctuating significantly over time.
* Impact of Crises:
  + The 2008 financial crisis caused GDP contractions across almost all EU member states, highlighting vulnerabilities in high-risk economies like Greece and Ireland.
  + The 2020 COVID-19 pandemic resulted in widespread GDP volatility, particularly in tourism-dependent countries such as Spain and Malta.
* Outliers:
  + Ireland's exceptional growth during its tech boom represents an outlier, with GDP growth exceeding average EU trends in certain periods.
  + Negative growth periods in economies like Greece emphasize the challenges of economic recovery following prolonged recessions.

**Key Insights from EDA**

1. Volatility varies significantly between countries, with some demonstrating resilience (e.g., Poland) and others exhibiting sensitivity to external shocks (e.g., Lithuania).
2. Stable economies generally recover faster from global crises, while high-risk markets often experience prolonged instability.
3. Trends over time reveal that periods of high volatility coincide with major global events, such as the oil crises of the 1970s, the 2008 crash, and COVID-19.

This phase of the analysis lays the groundwork for subsequent data preparation, enabling the derivation of metrics like volatility and a Stability Index. By understanding historical trends and patterns, the dataset provides the foundation for categorizing economies into risk tiers and informing FDI strategies.

# Data Preparation

The data preparation phase ensured that the dataset was clean, structured, and suitable for deriving actionable insights. Key metrics like volatility and the Stability Index were engineered to enable meaningful analysis and categorization of EU economies.

**Steps Taken**

1. *Data Cleaning:*

* Removed irrelevant columns such as country codes and metadata not necessary for analysis.
* Addressed missing values by excluding incomplete records for specific years or countries, ensuring consistent data quality.

1. *Data Transformation:*

* Reshaped the dataset from wide format (years as columns) to long format (one row per country-year observation). This structure facilitated time-series analysis and trend visualization.
* Normalized GDP growth values to identify patterns across countries with different economic scales.

1. *Feature Engineering:*

* Volatility: Calculated as the standard deviation of GDP growth for each country over the analysis period. This metric quantifies economic instability.
* Stability Index: Derived as the ratio of average GDP growth to volatility, providing a single measure of economic resilience.
* Risk Categories: Countries were classified into low, medium, and high-risk groups based on Stability Index thresholds, facilitating actionable insights for FDI strategy.

1. *Grouping by Decades:*

* Aggregated data by decades to identify long-term trends and periods of high or low volatility. This approach highlighted the impact of global economic cycles on EU economies.

**Tools Used**

• Pandas: Data cleaning, reshaping, and feature engineering.

• Matplotlib & Seaborn: Visualizations for identifying trends and outliers.

• NumPy: Statistical calculations for volatility and averages.

The prepared dataset served as the basis for visualizations and findings in subsequent phases. These transformations ensured that the analysis aligned with the objectives of classifying economic stability and informing FDI decisions.

# Findings & Recommendations

**Findings**

*Stability Index Results:*

1. Low-Risk Countries:

Countries such as Poland, Malta, and Belgium were identified as low-risk, with Stability Index values exceeding 1.0. These countries combine moderate GDP growth with low volatility, making them highly attractive for long-term investments.

1. Medium-Risk Countries:

Countries like Germany, Spain, and Finland fall into the medium-risk category (Stability Index between 0.6 and 1.0). These economies exhibit moderate stability with occasional periods of volatility, often linked to external factors such as global crises.

1. High-Risk Countries:

Finally, countries like Latvia, Lithuania, and Croatia were categorized as high-risk, with Stability Index values below 0.5. These countries experience frequent and significant GDP growth fluctuations, making investments less predictable.

**Volatility Analysis:**

* Smaller and emerging economies such as Latvia and Croatia are more volatile due to their sensitivity to external shocks and lack of economic diversification.
* Stable economies like Belgium and Germany have consistently low volatility, reinforcing their appeal for risk-averse investors.

**Impact of Economic Crises:**

* The 2008 financial crisis resulted in significant GDP contractions across the EU, with long-term instability in Greece and Ireland.
* The 2020 COVID-19 pandemic led to widespread economic disruption, particularly in countries dependent on tourism.
* Countries like Poland demonstrated resilience, with a faster recovery from crises compared to high-risk economies.

**Decadal Trends:**

* Volatility was notably higher in emerging markets during the 1990s and 2000s, reflecting economic transitions in Central and Eastern Europe.
* The 2020s highlighted the fragility of high-risk economies during global shocks.

**Recommendations**

1. Low-Risk Countries:

* Strategy: Focus FDI on stable markets such as Poland, Belgium, and the Netherlands. These countries provide consistent returns and long-term growth potential.
* Key Sectors: Infrastructure, renewable energy, and financial services, which align with their stable economic environments.

1. Medium-Risk Countries:

* Strategy: Explore sector-specific opportunities in Germany, Spain, and Finland. These markets require strategic timing to mitigate risks and capitalize on growth periods.
* Risk Mitigation: Monitor external factors, including political developments and global economic conditions, for informed decision-making.

1. High-Risk Countries:

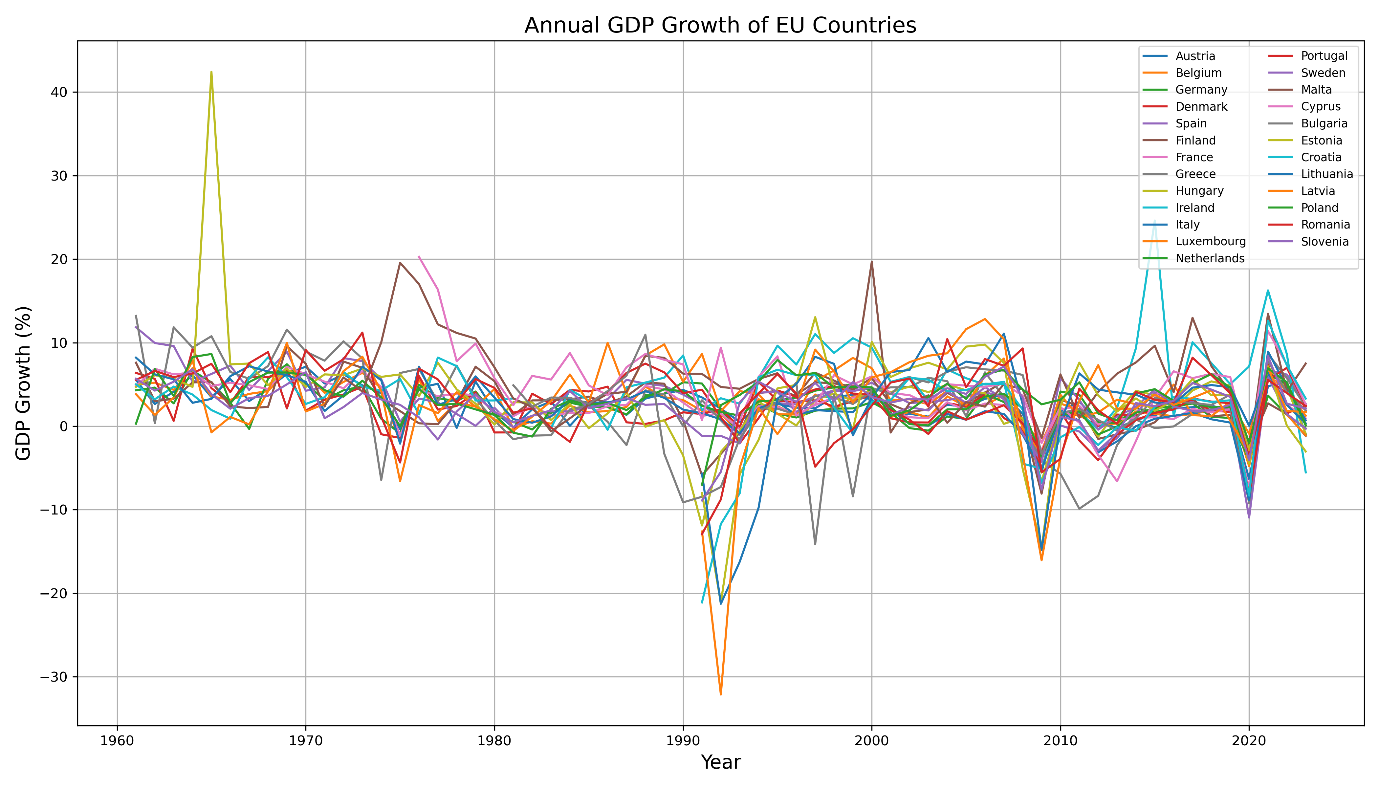
* Strategy: Approach markets like Latvia, Croatia, and Lithuania cautiously. Focus on high-reward sectors such as technology or tourism, emphasizing short-term investments.
* Diversification: Reduce exposure by spreading investments across multiple regions or sectors.

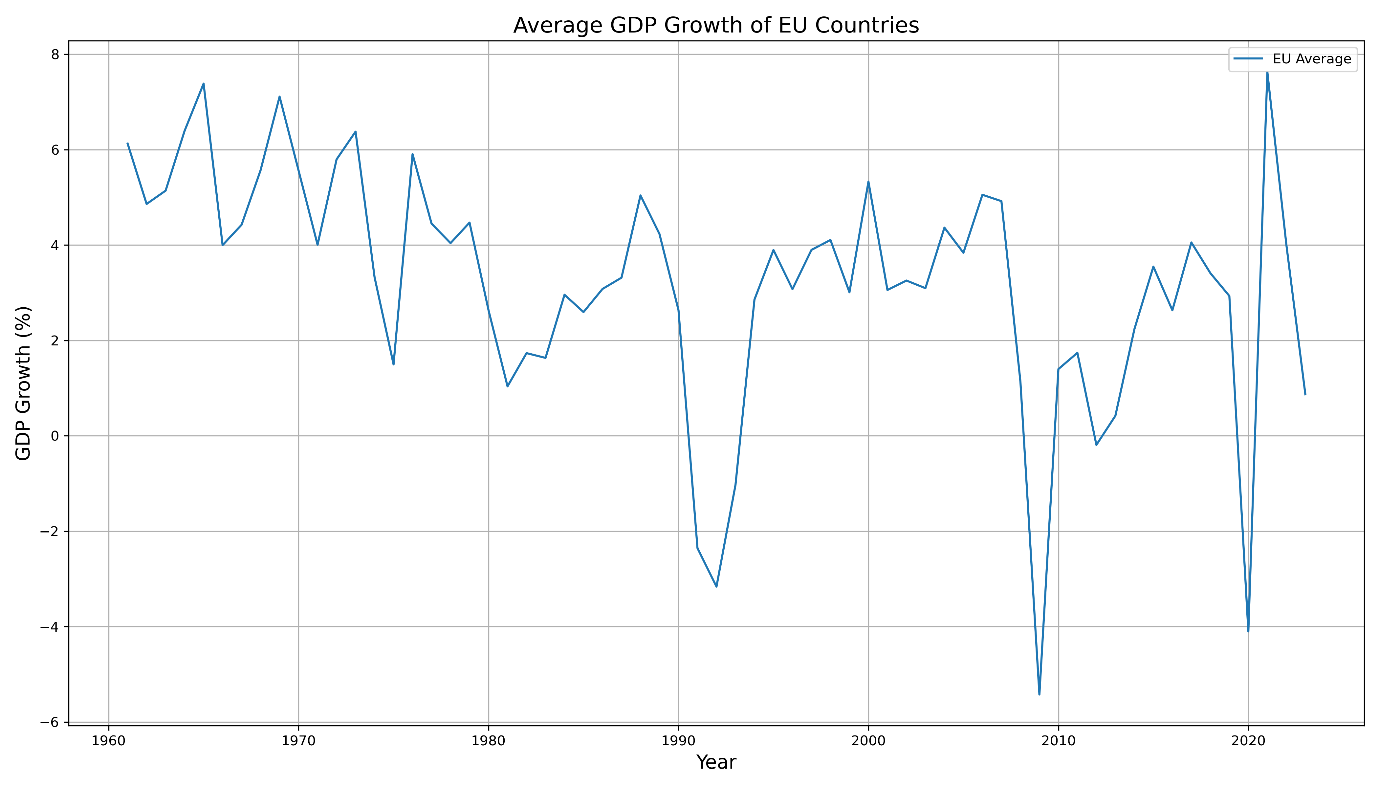
1. Future Analysis:

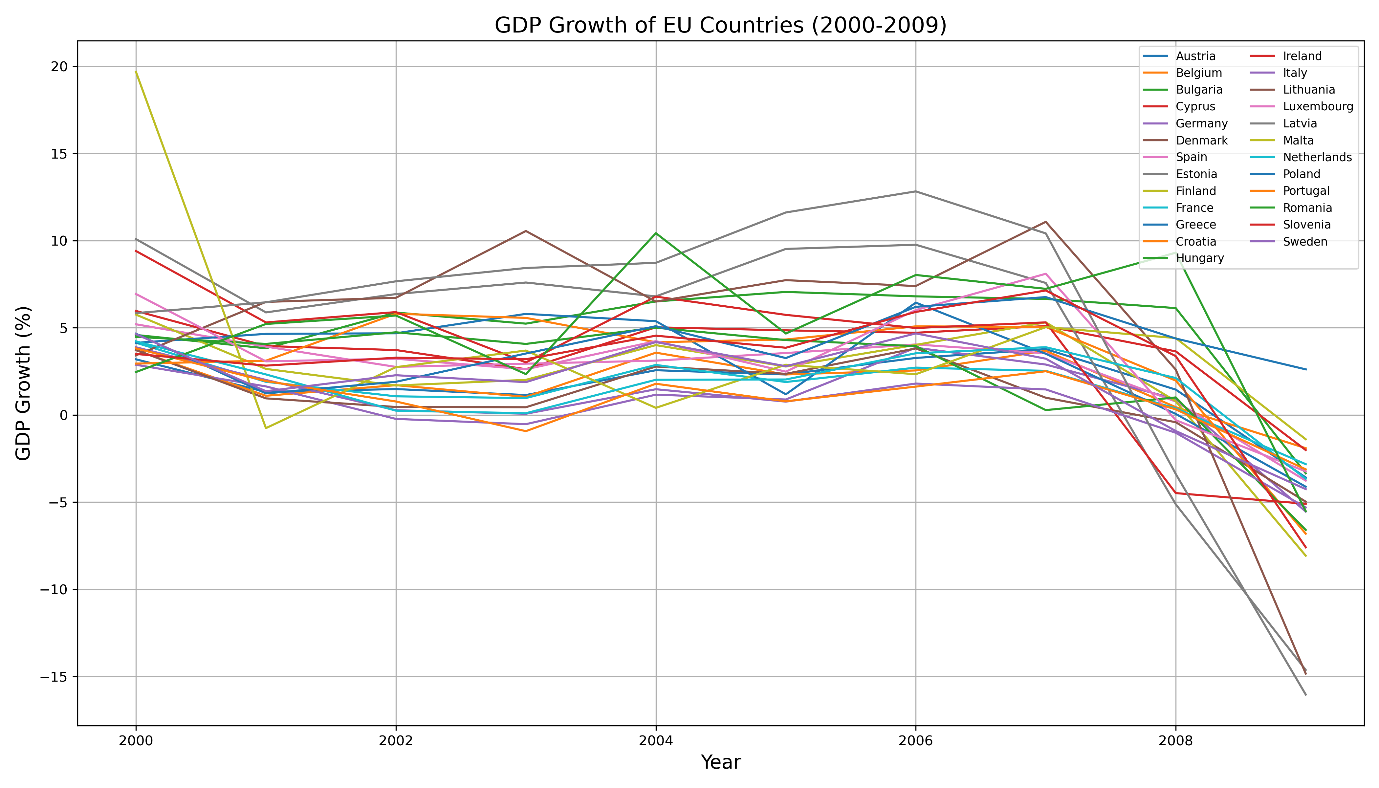
* Incorporate additional macroeconomic indicators such as inflation and unemployment to refine the Stability Index further.
* Leverage machine learning for clustering countries based on broader economic patterns.

These findings and recommendations offer a clear framework for balancing risk and reward in FDI strategies across the EU, empowering investors with actionable, data-driven insights.

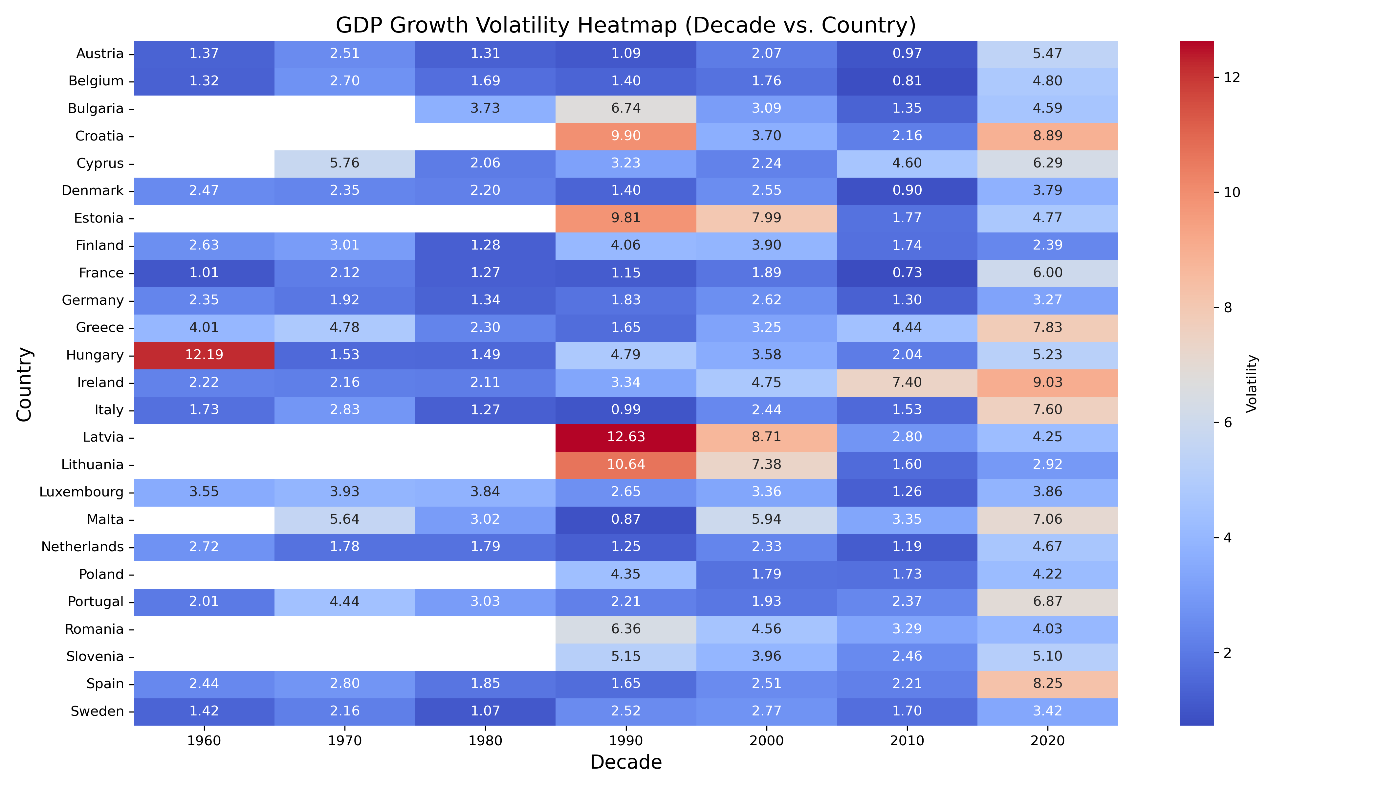
# Visualizations

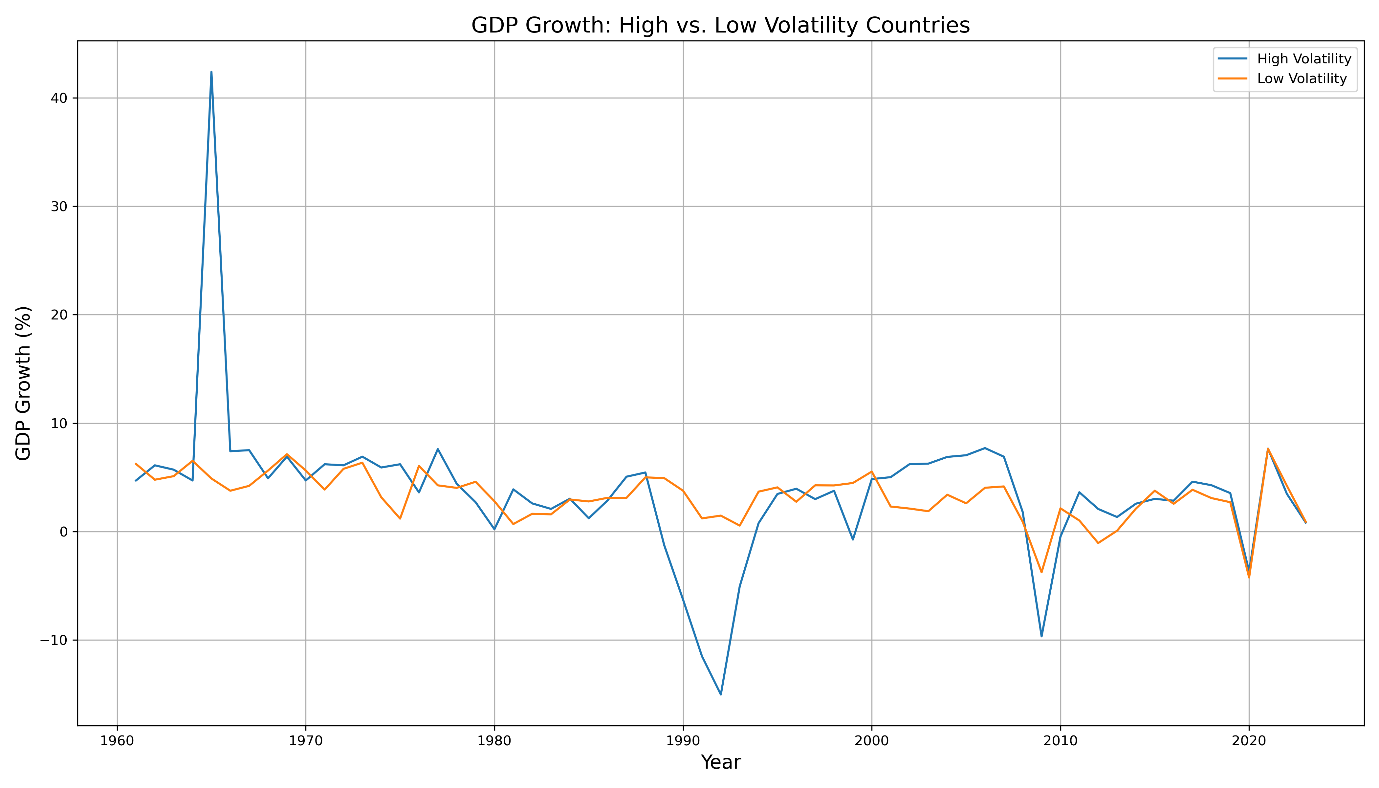


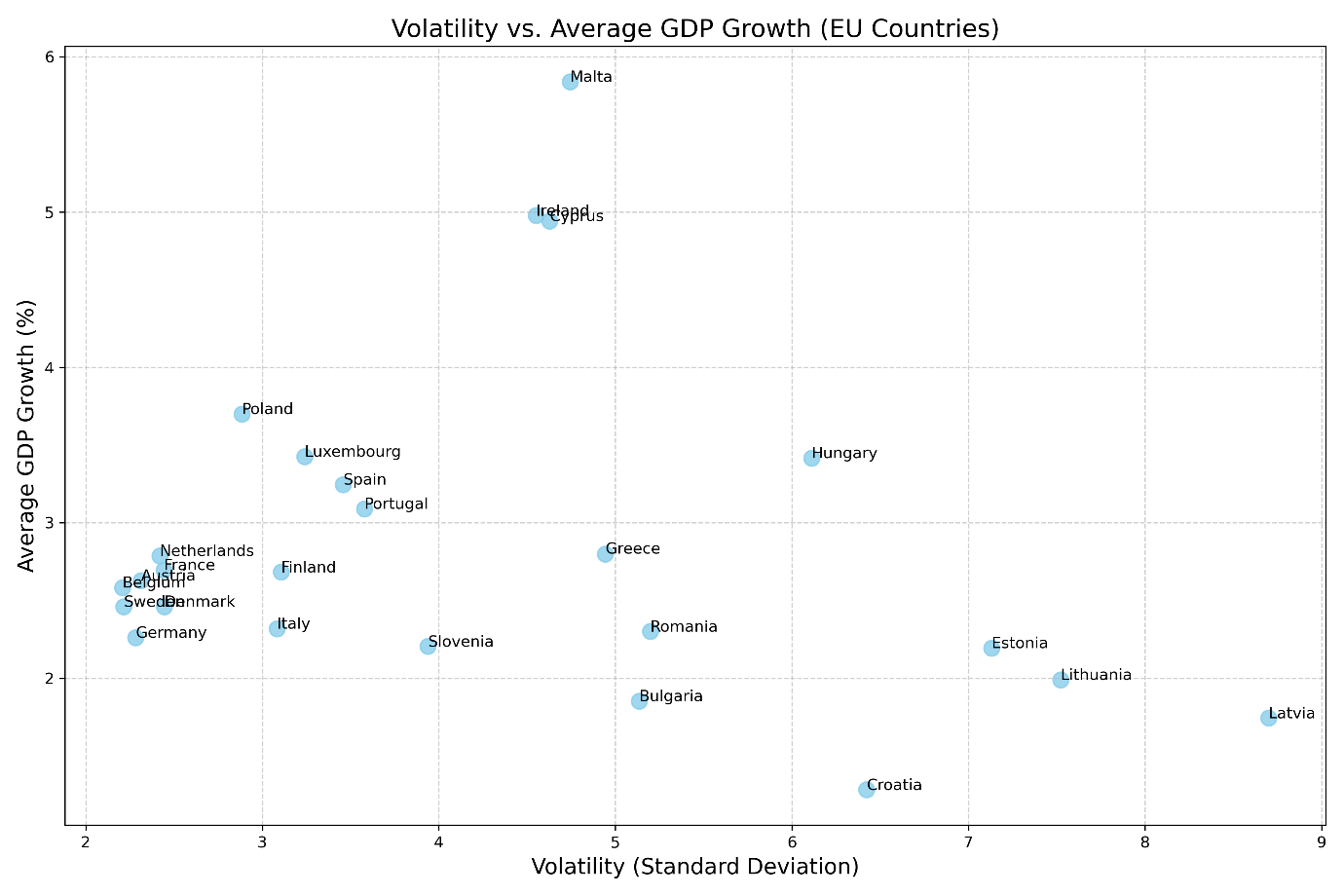


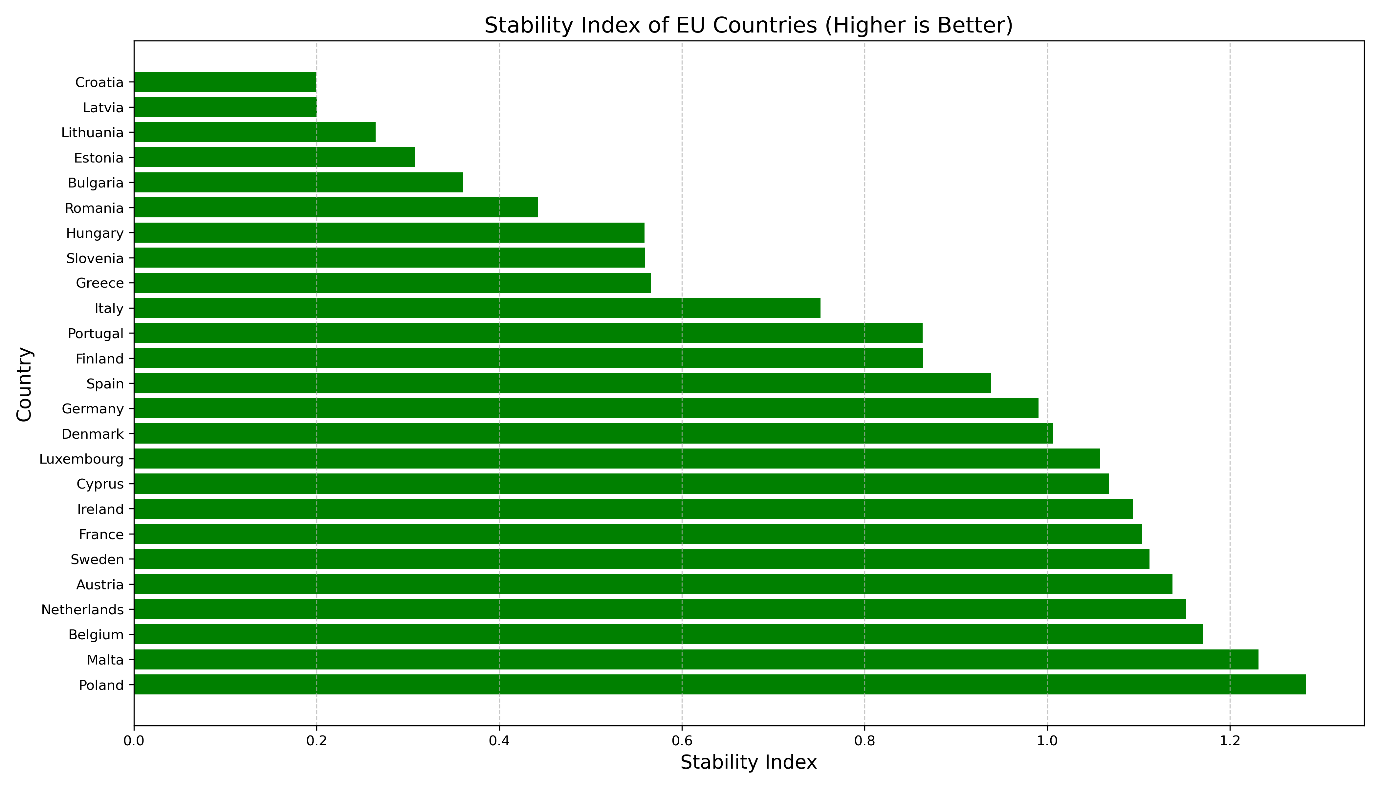


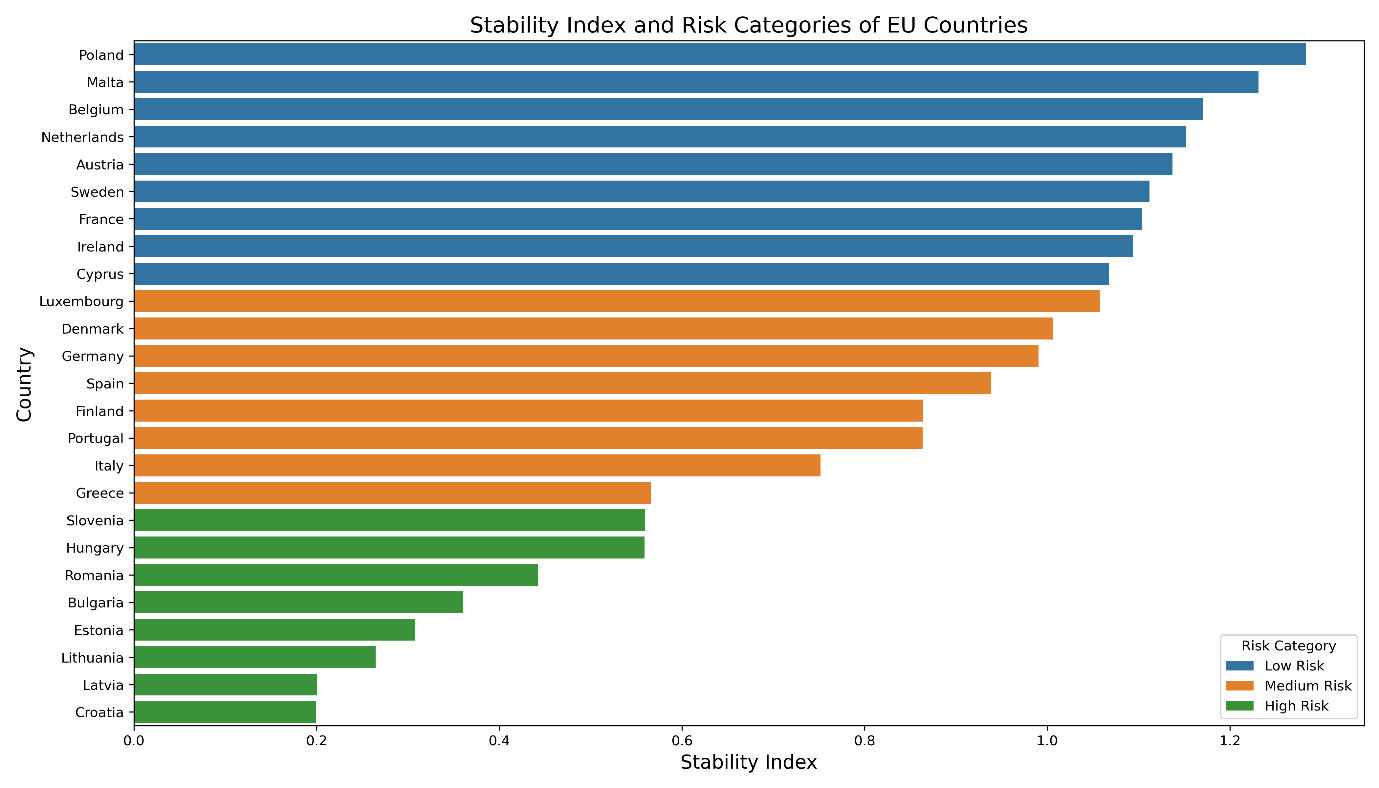












# Conclusion

This report analysed GDP growth trends and volatility across EU member states, categorizing countries into low, medium, and high-risk investment opportunities using the Stability Index. By applying the CRISP-DM methodology, the report systematically evaluated economic resilience and growth potential, providing actionable insights for Foreign Direct Investment (FDI) strategies, and laid the groundwork for future stages of the capstone project.

Key findings highlighted Poland, Malta, and Belgium as low-risk countries, offering stable environments for long-term investments. Medium-risk countries, such as Germany and Spain, presented moderate opportunities, while high-risk markets, including Latvia and Croatia, required cautious approaches. The impact of economic crises on GDP growth in the EU underscored the importance of resilience in driving investment decisions.

The recommendations emphasized focusing on stable markets for consistent returns, exploring sector-specific opportunities in medium-risk economies, and approaching high-risk markets with diversification strategies. Future analyses could enhance this framework by incorporating additional macroeconomic indicators or leveraging advanced machine learning techniques.

# References

* Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide. Retrieved from https://www.crisp-dm.org.
* Eurostat. (2021). GDP and Beyond: Measuring Economic Resilience. Retrieved from https://ec.europa.eu/eurostat.
* International Monetary Fund. (2020). World Economic Outlook: The Great Lockdown. Retrieved from https://www.imf.org/en/Publications/WEO.
* OECD. (2023). Economic Outlook 2023. OECD Publishing. Retrieved from
* https://www.oecd.org/economic-outlook.
* The World Bank. (2023). *GDP Growth (Annual %)* [Data set]. Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?most\_recent\_year\_desc=true.