Rotman Datathon 2025: Analyzing the Impact of Rising Cost of Living Relative to Economic Development and Supply Chain Dynamics

Victor Badea: victorbadea22@gmail.com
Rotman School of Management, University of Toronto
BMO Financial Group Finance Research and Trading Lab
January 16th, 2025

Table of Contents

Model Specifications

Data Plan: Structuring of the Data

My data cleaning process began by bringing the data into Excel and calculating the percentage of missing data for each column. This helped me gauge which columns could be salvaged and which should be dropped due to a lack of available data. I set a threshold for >50% of data missing, anything else would be dropped from the table. One problem immediately apparent with this dataset is many of the columns have over 50% of data missing and the breadth is vast, but there is very little depth within the history of the data set. None of the countries give data over 13 years prior. That said, I knew I would have to focus on a small set of features to get the most out of this data. The features I chose to focus on would also make or break the results I got. We used the MICE algorithm to fill in the missing data points. MICE is a powerful algorithm that gives a 'robust, informative method of dealing with missing data in datasets. The procedure 'fills in' (imputes) missing data in a dataset through an iterative series of predictive models. In each iteration, each specified variable in the dataset is imputed using the other variables in the dataset. These iterations should be run until it appears that convergence has been met.' (Kunal, 2024). After filling in all the missing values, I felt comfortable in moving on to the analysis phase of the project.

Correlation Analysis and Predictive Model

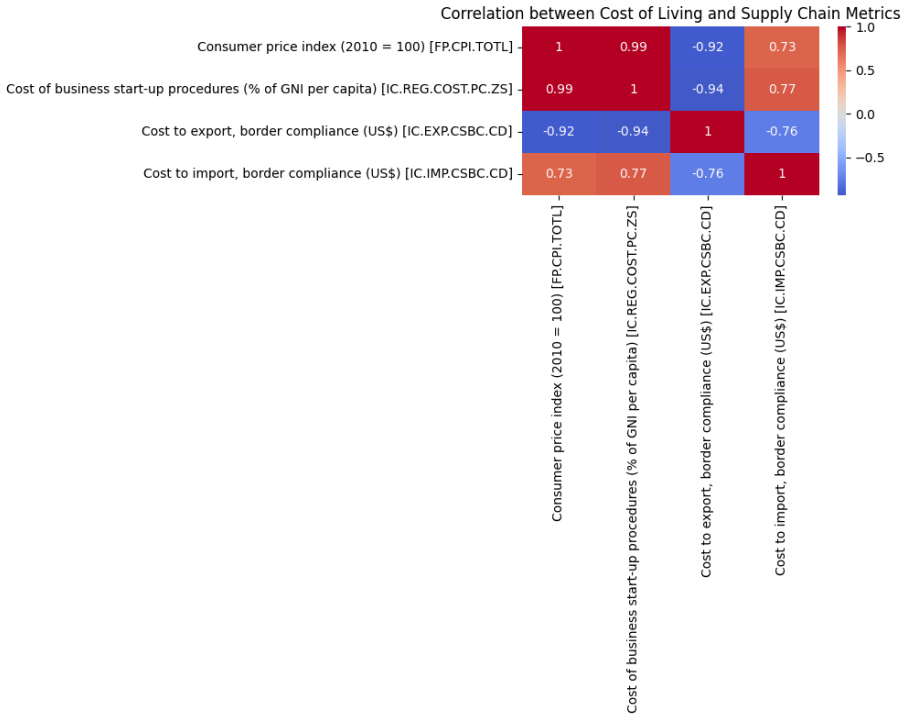
After cleaning the data, I began analyzing the data in Python to find any correlation between factors in economic data, supply chain data, and cost of living data. I used Google Collab along with libraries such as sklearn, pandas, numpy, matplotlib, and a few others to transform the data into actionable insights. First, I identified keywords in the data such as 'supply' and 'index' in the columns to help separate them into separate data frames under the categories of cost of living, economic, and supply chain. I then began exploring the correlation between cost of living and supply chain stability and costs. I used a correlation matrix to show the impact between different cost of living and supply chain metrics. Please find this matrix in the 'Cost of living impact on supply chain stability and costs' section. A similar approach was taken for answering the other two concerns of Economic and regional factors contributing to cost increases within supply chains and organizations adapting supply chain strategies. For the economic and regional factors, I also chose to use a bar chart to visualize separate features

importance of economic impact on export & import costs. These diagrams can also be found in the ‘Economic and regional factors contributing to cost increases within supply chains’ section. When it came to making recommendations, I used a cost sensitivity and cluster analysis plot to help find factors and suggest improvements. These diagrams can be found in the ‘How can organizations adapt supply chain strategies to mitigate the effects of rising costs on consumers?’ section.

Investment Strategies

Cost of living impact on supply chain stability and costs

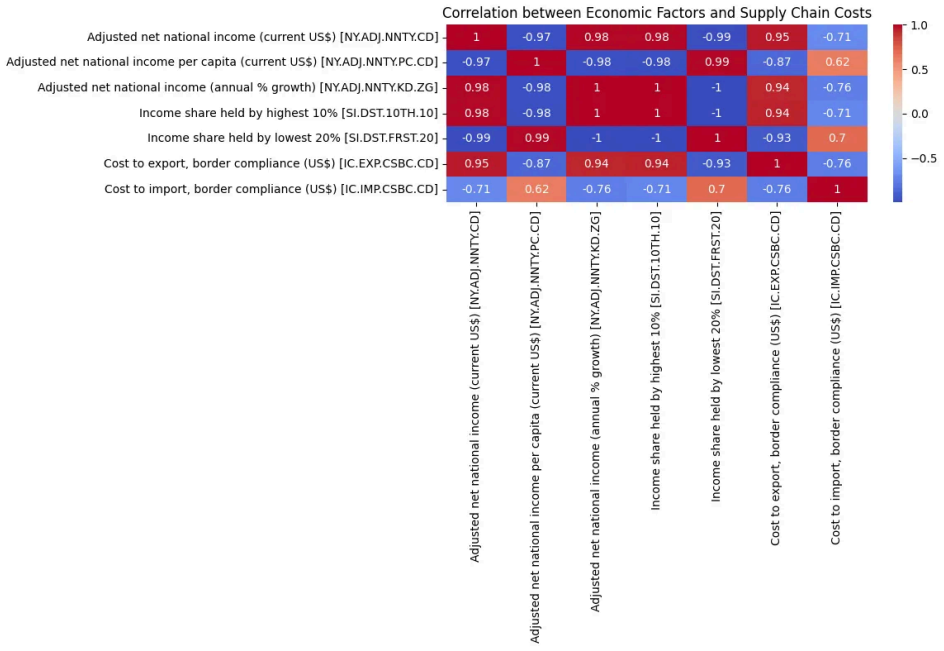
Below we see the correlation matrix between cost of living and supply chain metrics. I chose to analyse a few key features that I found to play the largest role. For Cost of Living, I chose CPI and cost of start-up procedures. For Supply chain metrics I chose cost to import and cost to export. I chose the cost of business start-up procedures because for many, this factors into the equality of a country. The higher the cost of business start-up procedures, the further lower and upper classes are. The cost of start-up procedures also provides a solid link to supply chain metrics because import and export prices can typically ‘make or break’ the feasibility of a business.

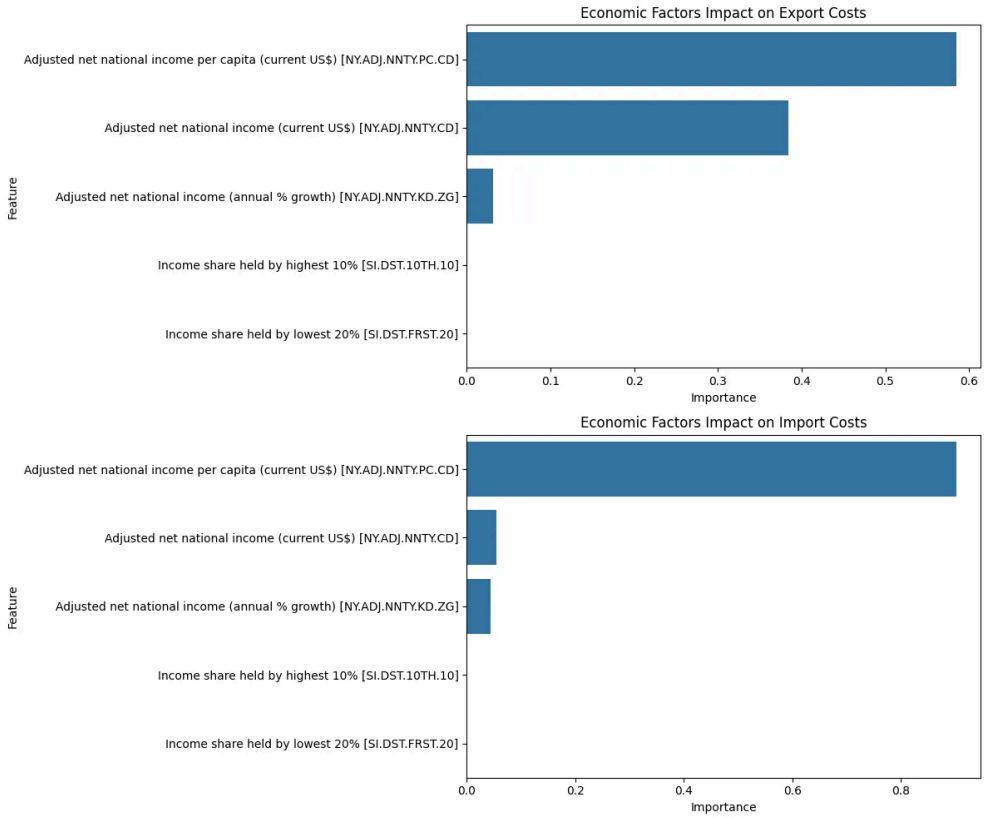


Here, we see a few key indicators of how the cost of living plays a role in supply chain costs. First, we find that there is a strong negative correlation with export costs. Business start-up costs have an even stronger negative correlation (-0.94) with export compliance costs. This suggests that as the general cost of living increases, border compliance costs for exports tend to decrease, which is an interesting and somewhat counterintuitive finding. Second, we see a moderately positive correlation with import costs. This indicates that as the cost of living rises, import costs tend to rise as well, though not as strongly as the export relationship. The negative correlation with export costs might suggest some regulatory or policy interventions in high-cost-of-living areas that make exports more efficient. Supply chain strategies might also need to be different for imports versus exports, as they respond differently to cost of living changes. This comes from the strong negative correlation (-0.76) between import and export compliance costs. As a side note, we are also able to see an extremely high positive correlation between CPI and start-up costs, proving that it does indeed act as a good measure for the cost of living.

Economic and regional factors contributing to cost increases within supply chains

Below we find the correlation matrix between economic and regional factors to cost increases in supply chains. We also find a bar chart showing which economic factors play the largest role in import and export costs.

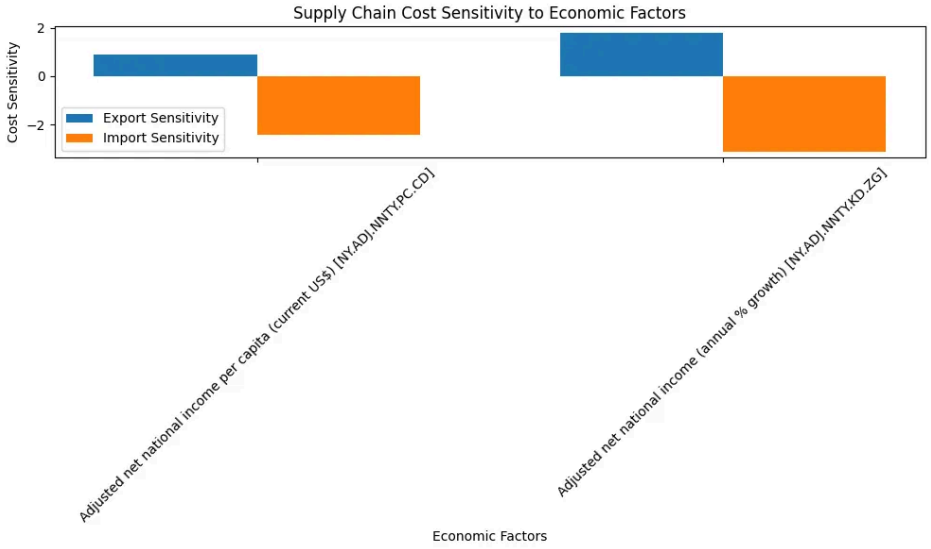


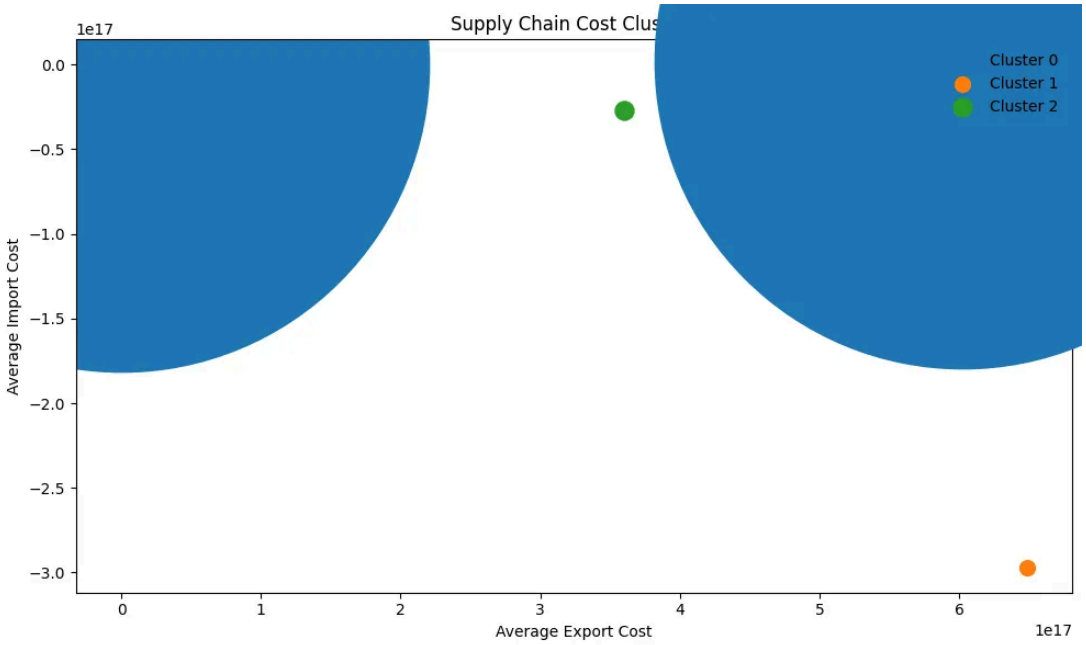


The five economic factors I chose to explore are adjusted net income per capita, adjusted net national income, adjusted net national income, income share held by the highest 10% and income share held by the lowest 10%. We see the strongest driver of export costs is the adjusted net national income per capita (0.585 impact score). This is followed by total adjusted net national income at a 0.384 impact score. The importance is calculated using XGBoost's feature importance metric. It calculates the feature importance by 'the amount that each attribute's split point improves the performance measure, weighted by the number of observations the node is responsible for. In other words, you are simply summing up how much splitting on each feature allows you to reduce the impurity across all the splits in the tree.' (Marsh 2023). As we can see from our charts, the absolute level of wealth (especially per capita) is more important for predicting export costs than economic growth rates. Companies should expect higher supply chain costs when operating in high-income markets.

How can organizations adapt supply chain strategies to mitigate the effects of rising costs on consumers?

Finally, we want to understand what to make of this analysis and how we can move forward to advise organizations on how to adapt supply chain strategies to mitigate the effects of rising costs on consumers. Below we see a cost sensitivity analysis and a cost clustering analysis.





Based on the analysis of economic factors' influence on supply chain costs, organizations must implement strategic adaptations to mitigate the impact of rising costs on consumers. The analysis identifies adjusted net national income per capita as the primary factor, with an importance score

of 0.58 in predicting export costs, followed by total adjusted national income at 0.38. This shows that organizations should develop market-specific strategies, with particular importance placed on high-income regions where supply chain costs show greater sensitivity to economic highs and lows. The research reveals significant negative correlations (-0.94) between business start-up costs and export compliance costs, suggesting that markets with elevated entry barriers frequently maintain more efficient export processes. Organizations can utilize these insights through three primary strategies: implementing import cost optimization in high-income markets, where the impact score reaches 2.414; establishing strategic regional distribution hubs to effectively manage costs across diverse economic environments; and deploying market-specific pricing mechanisms that reflect local economic conditions. The cluster analysis reinforces this approach, revealing distinct cost patterns that enable organizations to calibrate their strategies according to specific market characteristics, thereby maintaining competitive consumer pricing while ensuring supply chain sustainability.

References

Kunal. (2024, January 16). Multivariate imputation by chained-equations (mice). Medium. https://medium.com/@kunalshrm175/multivariate-imputation-by-chained-equations-mice-2d3efb063434

Marsh, E. K. (2023, February 28). XGBoost feature importance. Medium. https://medium.com/@emilykmarsh/xgboost-feature-importance-233ee27c33a4