**Introduction**

Energy consumption plays a crucial role in understanding a country's economic and environmental performance. This project focuses on analyzing Canada's energy consumption and comparing it with global trends to identify areas of strength, weakness, and opportunities for improvement. The analysis aims to provide actionable insights and recommendations to improve energy efficiency and sustainability, while using machine learning to forecast future energy demands.

The project is divided into three parts:

Detailed Analysis of Canada: This phase examines Canada's energy consumption at the national and provincial levels to uncover regional trends, disparities, and key drivers of energy usage.,

Global Energy Analysis: The second phase investigates energy consumption patterns across various countries worldwide to identify broader trends and benchmarks.

Comparative Analysis: In the final phase, Canada's energy performance will be compared to other key countries, highlighting where Canada excels and where improvements are needed.

To achieve these objectives, we are using three datasets:

1. World Energy Consumption Dataset (Source): Includes 131 variables and 22,000 rows with global metrics such as total energy consumption, GDP, and energy production by country.
2. Statistics Canada Energy Data (Source): Provides detailed monthly energy consumption data at both the national and provincial levels in Canada. Its comprised of about 1000 rows and 8 variables
3. IEA Monthly Electricity Statistics (Source): Contains 145,000 rows with energy production and balance metrics for various countries, categorized by key products like electricity, oil, and gas.

By combining these datasets, the project will analyze energy trends, identify critical insights, and utilize machine learning models to predict future energy consumption and gain further insight in the data. The findings will offer practical recommendations to improve Canada's energy efficiency and support sustainable energy strategies.

# Part 1 – Canada Analysis :

## Difficulties

Initially, analyzing three datasets with varying predictors, formats, and units was overwhelming. Handling missing values and merging the data required careful adjustments and the creation of aggregated datasets for organization. Identifying the best visualizations and crafting a clear, coherent story within a 10-page limit added complexity. Additionally, we faced challenges in avoiding overfitting and ensuring the selection of relevant features to produce meaningful and interpretable results.

## Data Analysis and Cleaning Process

We combined three datasets—World Energy Consumption (Kaggle), Government of Canada Statistics, and IEA Monthly Electricity Statistics—to analyze Canada's energy consumption. After merging and aggregating the data, we filtered for Canada, reducing it to 131 rows, and selected key variables like year, population, GDP, and energy metrics (fossil fuels, renewables, nuclear).

For cleaning, we removed rows with missing energy values, leaving 58 rows, and addressed gaps using forward fill and backward fill. We standardized column names, ensured chronological order, validated units (TWh), and performed multiple additional checks to prevent errors later on.

The final dataset, spanning 58 years (1965–2022), revealed trends like declining coal usage, hydroelectric dominance, and the gradual rise of solar and wind energy.

## Monthly Data:

We analyzed electricity consumption by month to identify seasonal patterns. The graphs revealed that energy usage peaks in January and February due to increased heating demands during Canada's cold winters. In contrast, consumption is significantly lower during summer months (June to August) when heating needs are minimal, while transitional months like April and October show moderate demand. These trends align with seasonal temperature variations, highlighting the importance of preparing for peak energy needs during winter.

### Energy Efficiency, System Losses, and Environmental Impact

To evaluate energy efficiency in Canada, we compared electricity production with final consumption to identify system losses. A large disparity between the two indicates significant inefficiencies in energy delivery.

We began by analyzing distribution losses over time, revealing a sharp decline between 2013 and 2014, followed by stagnation in recent years. Additionally, we assessed the net difference between production and consumption for key energy types, focusing on oil, gas, coal, and hydroelectricity.

Our findings highlighted that oil is the least efficient energy source, with increasing net losses since 2015, while other energy types exhibit smaller but still notable losses. This analysis underscores the need to transition toward cleaner and more efficient energy alternatives to minimize losses and reduce environmental impact.

## Gas Emissions:

After evaluating energy production and consumption efficiency, we turned our focus to the environmental impact by analyzing greenhouse gas emissions over time.

From 2000 onwards, emissions declined significantly from 130 $MtCO_2$ to a low of 75 $MtCO_2$ by 2022, with a slight uptick in the same year. While this progress is encouraging, understanding the key drivers of emissions is essential to ensure continued improvements.

To achieve this, we implemented two machine learning models to identify the primary predictors of greenhouse gas emissions. The technical details and results are discussed in the following sections.

## Model Selection For Gas Emission:

### Model 1: Random Forests

To predict greenhouse gas emissions and identify key drivers, we employed the Random Forest Regressor for its ability to handle multiple predictors and reduce overfitting. Unlike single decision trees, Random Forest builds multiple trees on random subsets, improving accuracy and robustness. Hyperparameters were optimized using RandomizedSearchCV with 10-fold cross-validation to ensure generalizability.

The model achieved strong performance with minimal errors, validating its reliability. Feature importance analysis revealed the top contributors: wind_share_energy, carbon_intensity_elec, and solar_elec_per_capita, emphasizing the role of renewable energy and carbon intensity in influencing emissions. Economic factors like GDP and energy efficiency were also significant, showing that energy policies and economic growth both impact emissions.

To validate these findings, we applied OLS Regression while implementing techniques to mitigate overfitting and improve interpretability, further ensuring the robustness of our results.

### Model 2: Linear, Ridge, and Lasso Regression

To identify key predictors of greenhouse gas emissions, we applied Linear Regression, Ridge Regression (L2 regularization), and Lasso Regression (L1 regularization). Feature selection (e.g., backward stepwise) eliminated irrelevant predictors, and models were optimized using 10-fold cross-validation with hyperparameter tuning.

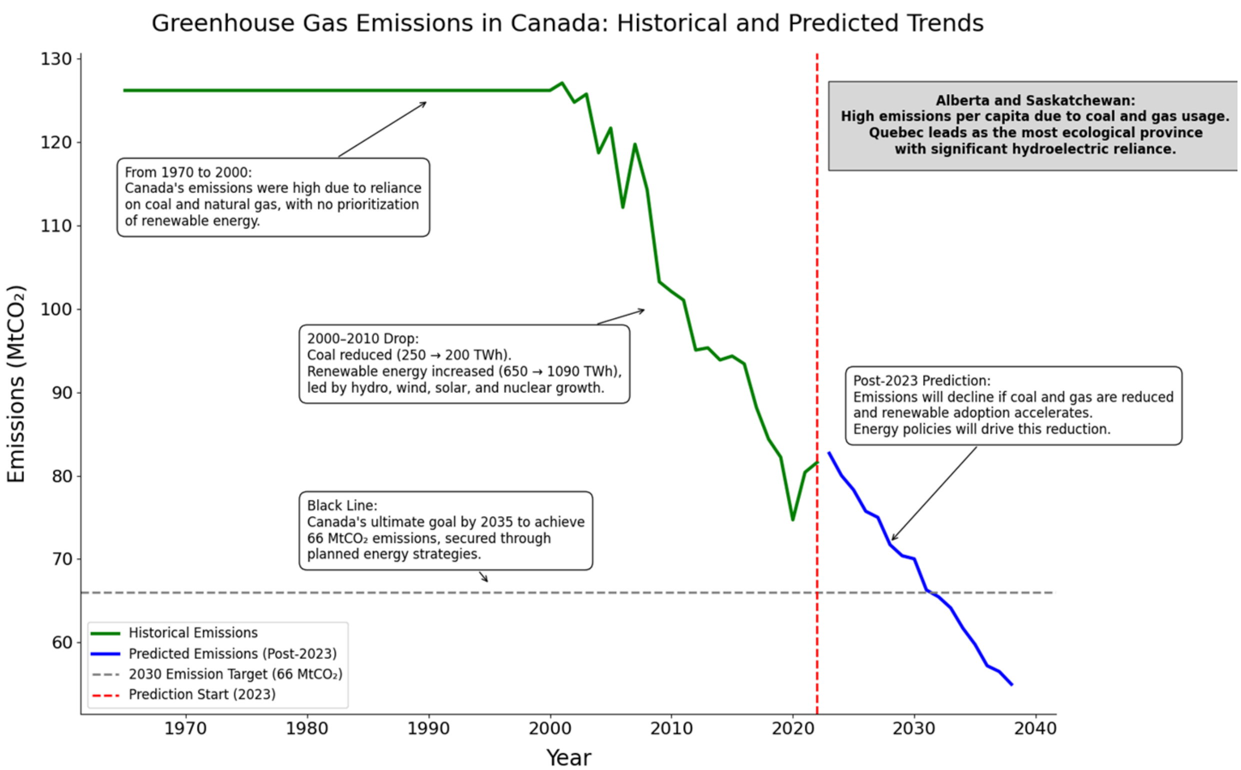
Lasso Regression achieved the best performance, highlighting critical predictors such as coal consumption per capita, fossil electricity per capita, and gas production per capita, emphasizing the role of fossil fuels in driving emissions.

Ultimately, Linear Regression was selected for its simplicity, interpretability, and strong performance, aligning with domain expectations and effectively capturing the relationship between predictors and emissions.

### Scenario Analysis: Projecting Future Emissions

To simulate future emissions from 2023 to 2038, we used the Linear Regression, Ridge, and Lasso models. The Lasso model identified the two most significant predictors: coal consumption per capita and fossil electricity per capita. These predictors were projected forward using Linear Regression with a manual reduction rate of 1% per year, while all other features were held constant at their 2022 levels.

The predictions showed a clear downward trend in emissions over time. Lasso Regression produced the most stable results by focusing on critical predictors, while Linear and Ridge Regression demonstrated similar patterns with minor variations. A red dashed line representing 2022 emissions served as a baseline, emphasizing that reducing key predictors can drive significant and sustained reductions in greenhouse gas emissions, supporting long-term sustainability goals.



## Energy Consumption by Province

We analyzed total energy consumption across Canadian provinces to identify regional trends and inefficiencies. Ontario, Quebec, Alberta, and British Columbia emerged as the top energy consumers, with Ontario leading at over 500 million gigajoules and smaller provinces like New Brunswick consuming significantly less.

To further investigate, we plotted individual line graphs for the top four provinces:

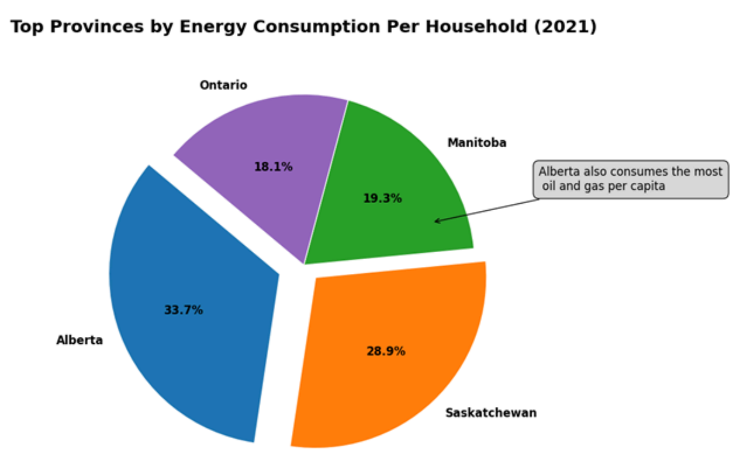
- Ontario showed a gradual increase until 2019, followed by a sharp decline in 2021.
- Quebec displayed a steady downward trend since 2013, suggesting improved efficiency.
- Alberta experienced fluctuations driven by industrial activity, peaking in 2019.
- British Columbia maintained stable consumption with minor variations.

Next, we evaluated energy efficiency by analyzing energy consumption per household to account for population differences. Data preparation included filtering relevant metrics, scaling features, and visualizing patterns for clarity.

The results highlighted key inefficiencies:

- Alberta, Saskatchewan, and Manitoba lead in per-household energy consumption, reflecting higher energy demands relative to population size.
- Natural Gas: Alberta and Saskatchewan rely heavily on natural gas, a major emissions contributor.
- Heating Oil: Prince Edward Island and Newfoundland show high heating oil consumption, signaling a need for cleaner alternatives.
- Quebec stands out for its reliance on hydroelectricity, positioning it as the most environmentally sustainable province.

In summary, we identified critical inefficiencies in Alberta, Saskatchewan, and Manitoba due to fossil fuel reliance and per-household consumption, while Quebec serves as a model for clean energy practices. This analysis underscores the need for targeted interventions to improve energy efficiency and transition to renewables in underperforming regions.




# Part 2 – World Analysis

### Data Overview and Insights

After cleaning, the dataset spans 133 countries over 58 years (1965–2022) with key statistics: the mean population is 207.8 million (max 4.72 billion), and the mean GDP is $838 billion USD (max $26.97 trillion). Energy trends show coal consumption averaging 1270 TWh (max 36,847 TWh), reflecting industrial reliance, while oil consumption averages 1793 TWh, highlighting its dominance. Gas consumption averages 1128 TWh, driven by its growing role, and renewables average 416 TWh, with hydro leading at 313 TWh. Solar (42.5 TWh) and wind (80.6 TWh) are recent additions. Nuclear energy averages 212 TWh, peaking at 6627 TWh during its peak years.

### Key Observations

Key observations include significant global population growth, peaking at 4.72 billion across 133 countries in 2022. Fossil fuels dominate energy consumption, with oil averaging 1793 TWh globally, while renewables like solar and wind are gaining traction but remain far below fossil fuel levels. Hydro leads renewables with an average of 313 TWh. Nuclear energy, with peaks reaching 6627 TWh, highlights its role as a clean energy alternative.

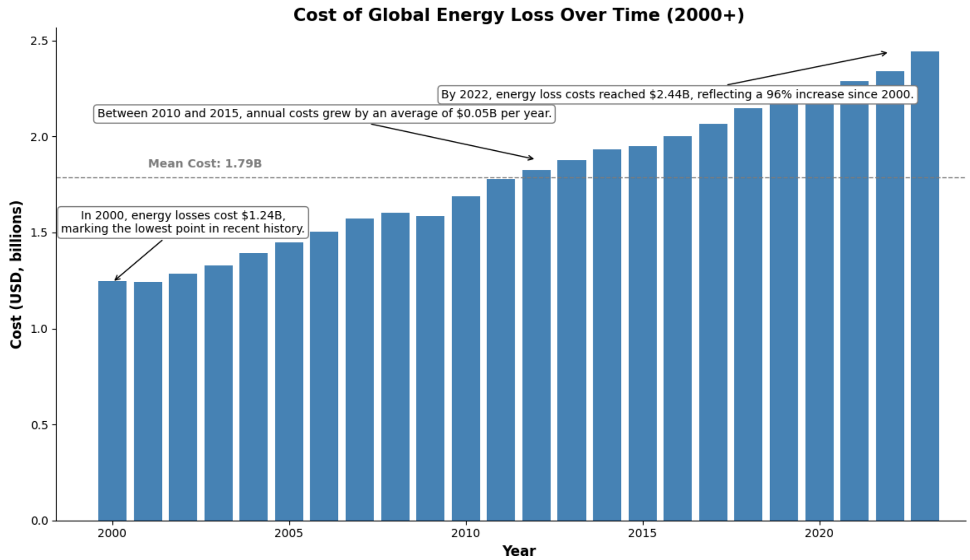
## Part 2 – World Analysis

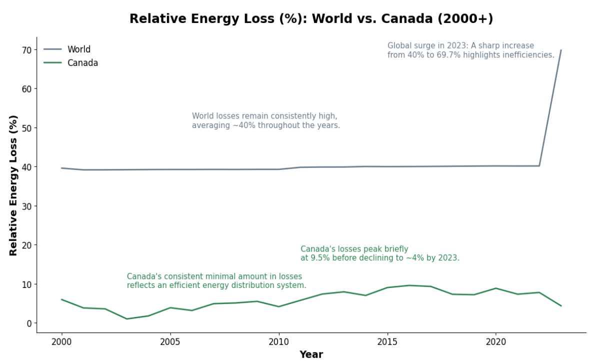
The dataset spans 133 countries over 58 years (1965–2022), highlighting global energy trends. Key statistics include an average population of 207.8M (peaking at 4.72B in 2022) and GDP averaging $838B (max $26.97T). Energy consumption remains dominated by fossil fuels, with coal averaging 1270 TWh (max 36,847 TWh), oil at 1793 TWh, and gas at 1128 TWh, reflecting industrial reliance. Renewables average 416 TWh, primarily driven by hydroelectricity (313 TWh), while nuclear energy averages 212 TWh, peaking at 6627 TWh during its prime years.

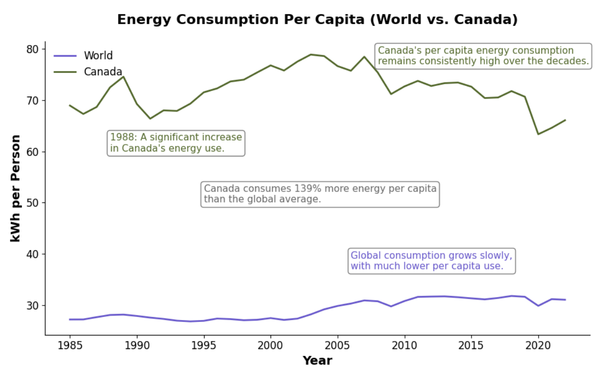
Fossil fuels remain the dominant global energy source, with oil leading at 314,791 TWh in 2022, followed by coal (269,890 TWh) and gas (238,567 TWh). Renewables, though growing, remain far behind at 144,000 TWh, led by hydroelectricity (46%), wind (24%), solar (15%), and biofuel (5%). Despite growth, solar and wind contribute a modest share compared to fossil fuels. Meanwhile, nuclear energy declined to 41,733 TWh by 2022, signaling reduced reliance globally.

### Methodology

To analyze energy trends, global energy data was aggregated while isolating Canada-specific trends. Energy loss costs were calculated as:
Cost of Loss (USD) = (Net Loss (TWh) × Cost per GWh) / 10^9
with Cost per GWh = $40,000. Per capita energy consumption was derived as:
Per Capita (kWh/person) = (Fossil Fuel Consumption (TWh) / Population) × 10^6.
Relative energy loss percentages were compared globally and for Canada to evaluate efficiency. Anomalies, such as the 2023 spike, were identified and excluded to ensure data accuracy.




## Key Findings

Global energy loss costs increased by 96%, rising from $1.24B in 2000 to $2.44B in 2022. Canada maintained minimal loss costs (< $0.01B annually), demonstrating superior efficiency. Per capita energy consumption reveals that Canada averages 71.96 kWh/person, 139% higher than the global average of 30.09 kWh/person. This difference may be attributed to Canada's harsh climate, industrial reliance, and abundant hydroelectric/renewable energy infrastructure.

Relative energy loss further highlights global inefficiencies, with the world averaging ~40% and peaking at 69.7% in 2022. In contrast, Canada consistently reduced energy losses to ~4% by 2022, showcasing the effectiveness of its infrastructure.

## Analysis and Skills Applied

Our approach involved detailed data processing to compute per capita energy use, relative losses, and energy loss costs. Visualizations included annotated bar and line graphs to emphasize trends and anomalies. Critical reasoning was applied to identify irregularities such as Canada's 1988 consumption spike and the 2020 drop possibly linked to the global pandemic.

## Recommendations

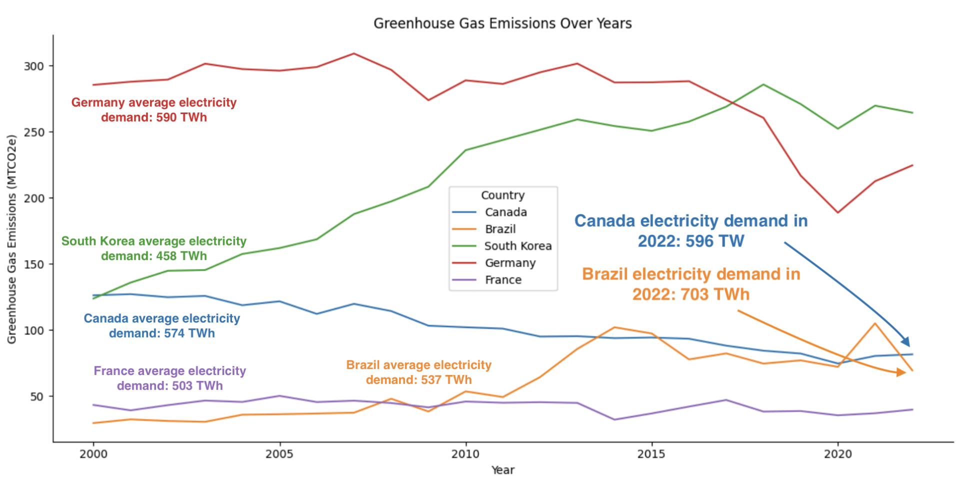
To address global inefficiencies, investments in energy system improvements are essential to reduce the persistent near 40% energy loss. Best practices, such as Canada's success in leveraging efficient infrastructure and hydroelectric resources, should be studied and adopted. Future analyses could compare energy consumption and efficiency with Nordic countries to identify scalable solutions. Additionally, further investigation into anomalies like the 1988 spike and 2020 drop is recommended to uncover root causes and potential mitigation strategies.

## Conclusion

Canada demonstrates exceptional energy efficiency despite higher per capita consumption driven by unique climatic and industrial needs. In contrast, persistent global inefficiencies highlight the need for targeted interventions and accelerated renewable energy adoption. This analysis underscores the importance of robust data handling, visualization, and critical evaluation in addressing global energy challenges.

# Part 3 - Comparative Analysis

To enhance our comparative analysis, we chose to evaluate Canada alongside several distinct groups of countries sharing similar characteristics. These groups include nations with comparable GDP, population size, winter climate duration, energy demands, membership in the G7, and those with the lowest greenhouse gas emissions. After our analysis we found a couple of interesting stories:



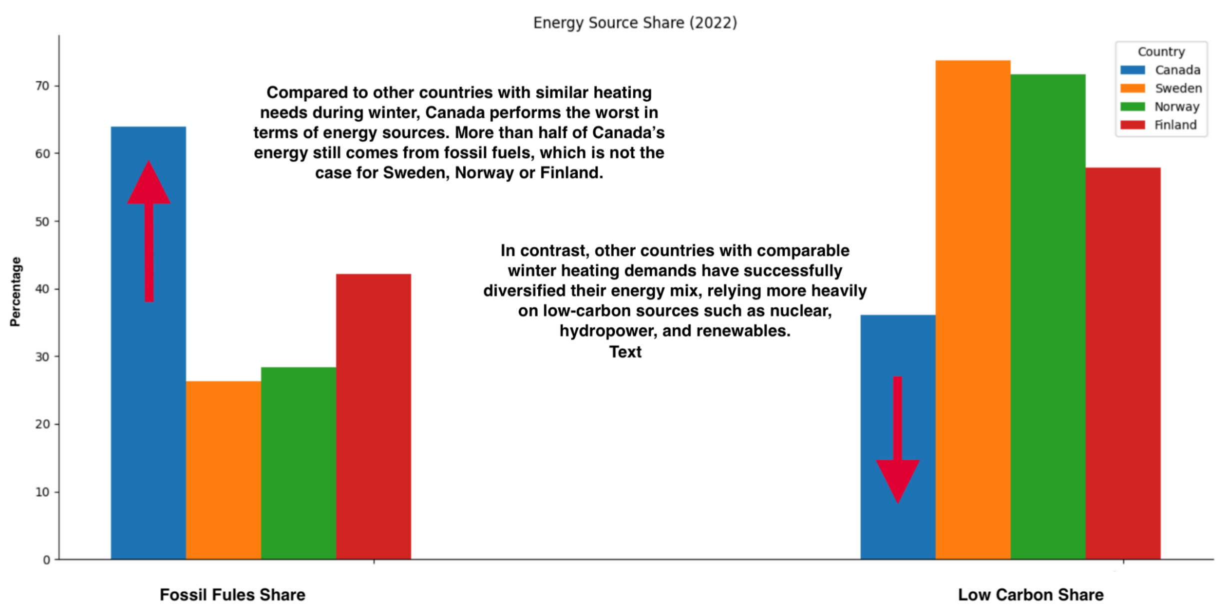

This graph compares Canada to five countries with similar electricity demands, illustrating their greenhouse gas emissions over the past 20 years. Overall, Canada performs well in managing its electricity needs. Despite having an average electricity demand higher than many countries, Canada has successfully kept its greenhouse gas emissions relatively low compared to most of the countries in this group. However, there are two notable exceptions: Brazil and France.

France has lower electricity needs compared to Canada. Brazil, on the other hand, had a much higher electricity demand than Canada in 2022 but still managed to achieve lower greenhouse gas emissions. Why is this?

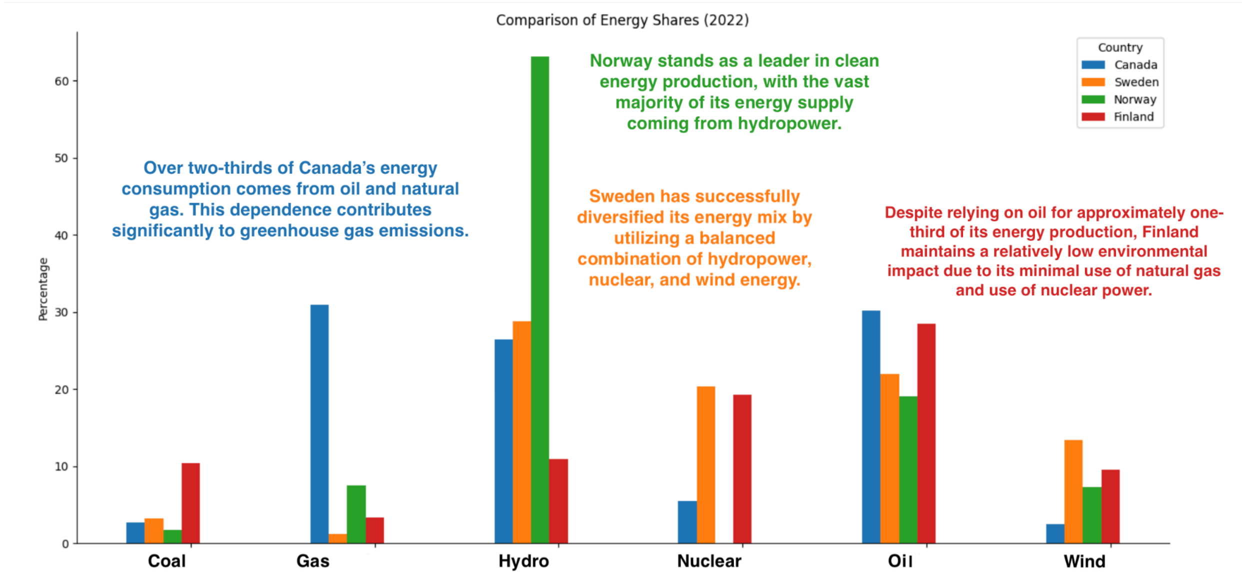
Brazil has a highly diversified electricity mix, predominantly relying on low-carbon energy sources such as hydropower and renewables. Similarly, Canada generates most of its electricity from low-carbon emission sources, with only about a 10% difference compared to Brazil, which isn't enough to explain our higher greenhouse emissions.

Canada relies on low-carbon sources, primarily hydropower, to produce electricity. However, its overall energy production heavily depends on fossil fuels. This is largely due to the unique situation where 90% of Canadians require heating during the winter, a highly energy-intensive process that is predominantly powered by fossil fuels.

We compared Canada to other countries with similar winter climates to assess how they manage their energy needs.




Canada's heavy dependence on fossil fuels for energy production contributes to higher greenhouse gas emissions, particularly during the cold winter months when energy demand for heating peaks. This highlights a critical area for improvement, as other nations with similar climatic challenges demonstrate that transitioning to cleaner, low-carbon energy sources is achievable while still meeting heating needs.




Scandinavian countries heat their homes using highly efficient, sustainable methods, with a focus on district heating powered by renewable energy like biomass, waste heat, and geothermal systems. Heat pumps are widely adopted, while Norway relies heavily on electric heating from abundant hydropower, and Iceland uniquely uses nearly 100% geothermal energy. These systems are integrated with strict building codes, smart energy management, and government incentives to prioritize sustainability. Compared to Canada, which primarily uses natural gas and electricity, Scandinavia relies more on centralized and renewable heating systems, whereas Canada's heating is often decentralized and still transitioning toward greener technologies.

## Conclusion:

In conclusion, our comparative analysis highlights Canada's strengths and challenges in managing energy demands and greenhouse gas emissions. While Canada performs well in maintaining low emissions relative to its high electricity demand, its reliance on fossil fuels for winter heating remains a significant contributor to overall emissions. To further reduce emissions, Canada could adopt centralized, renewable-based heating systems, improve building efficiency, and incentivize sustainable energy practices, aligning its energy strategies with those of global leaders in sustainability.