Oil Profitability Analysis

By
Ammanuel F. WoldeAregay

AGENDA

Introduction

Assess the production value data

Build a predictive model

Share insights

Final tips & takeaways

Problem Statement

The Primary Goal:

- Analyze oil and gas production across different countries other than U.S.
- Determine which areas are the most profitable for future investment.
- The analysis will consider like production trends in those areas.

Raw Data Sample

	country_name	type	product	flow	year	value
0	Australia	Balance	Crude oil	Industrial Production	2021	18029.678
1	Austria	Balance	Crude oil	Industrial Production	2021	561.852
2	Belgium	Balance	Crude oil	Industrial Production	2021	0.000
3	Canada	Balance	Crude oil	Industrial Production	2021	266630.180
4	Chile	Balance	Crude oil	Industrial Production	2021	340.997

2376 Rows
6 Columns

Data Wrangling

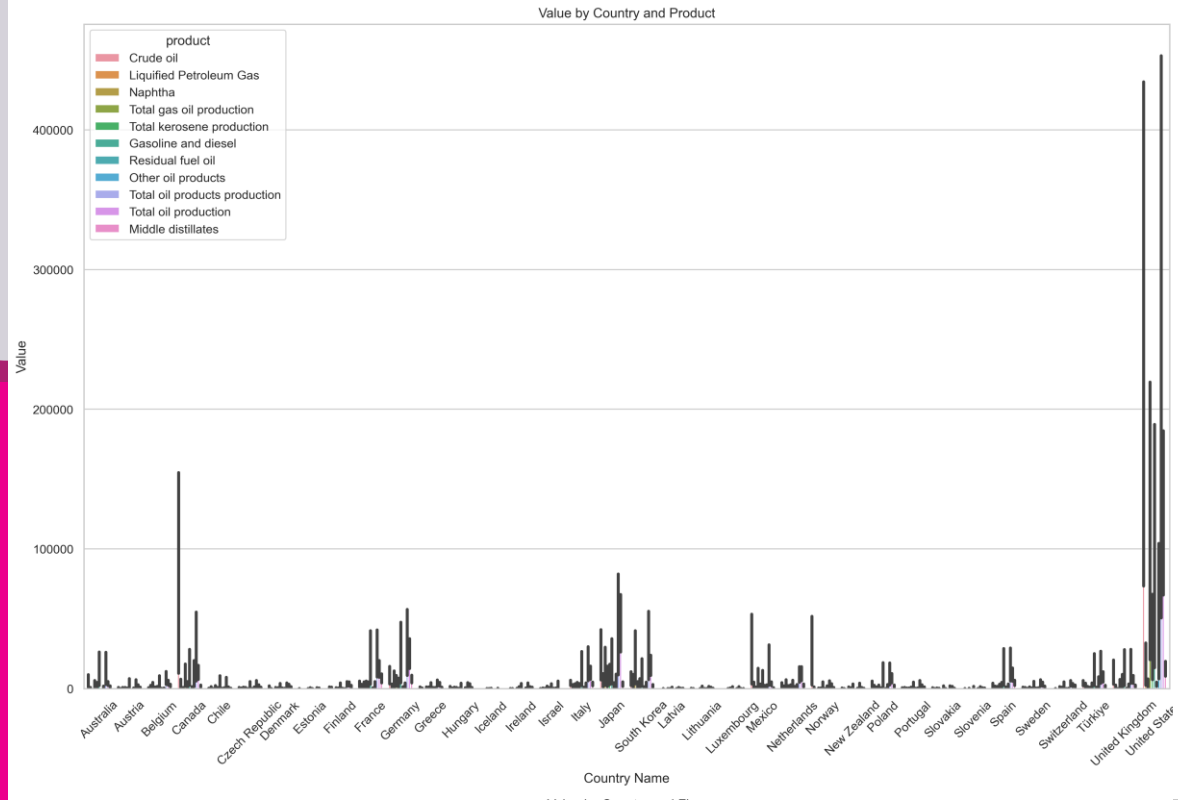
Original dataset had 2376 rows and Rows

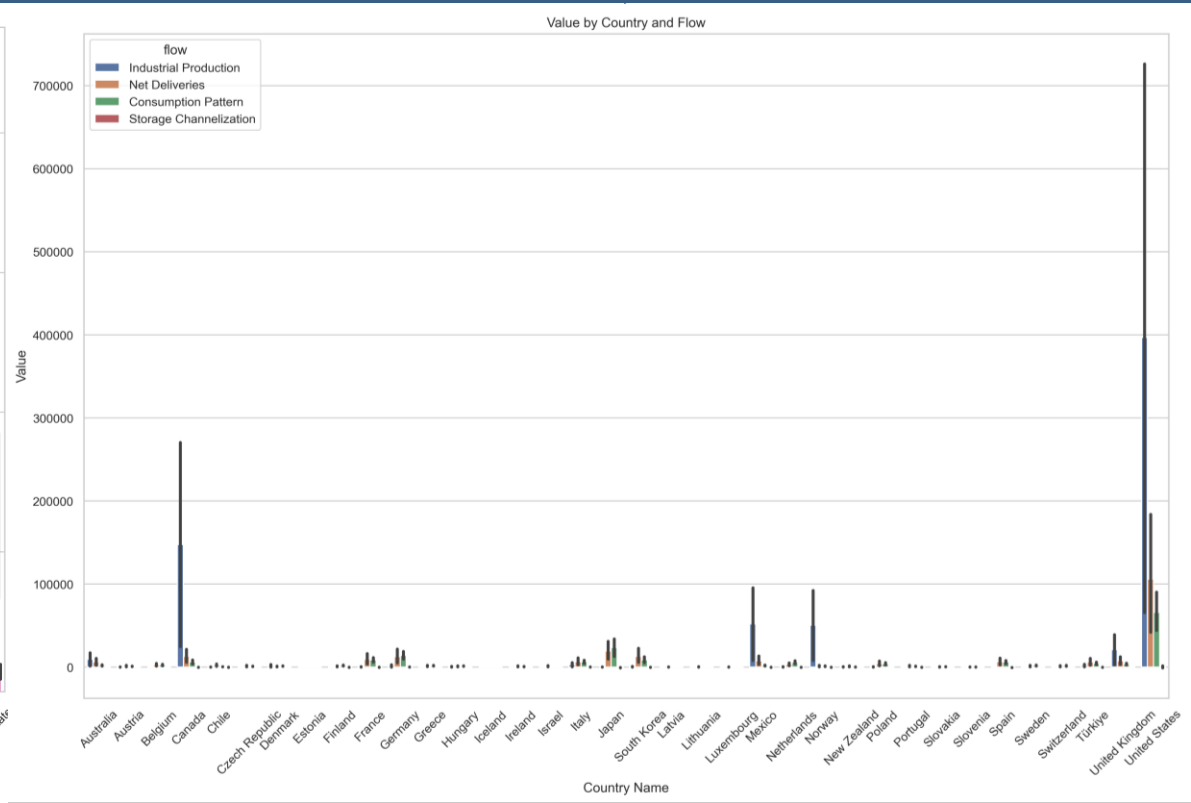
6 columns

- Drop unnecessary columns
- Assess the raw data
- Display data samples to understand its structure
- Data Cleaning: Identify and handle missing values
- Target Variable: Production value

Exploratory Data Analysis

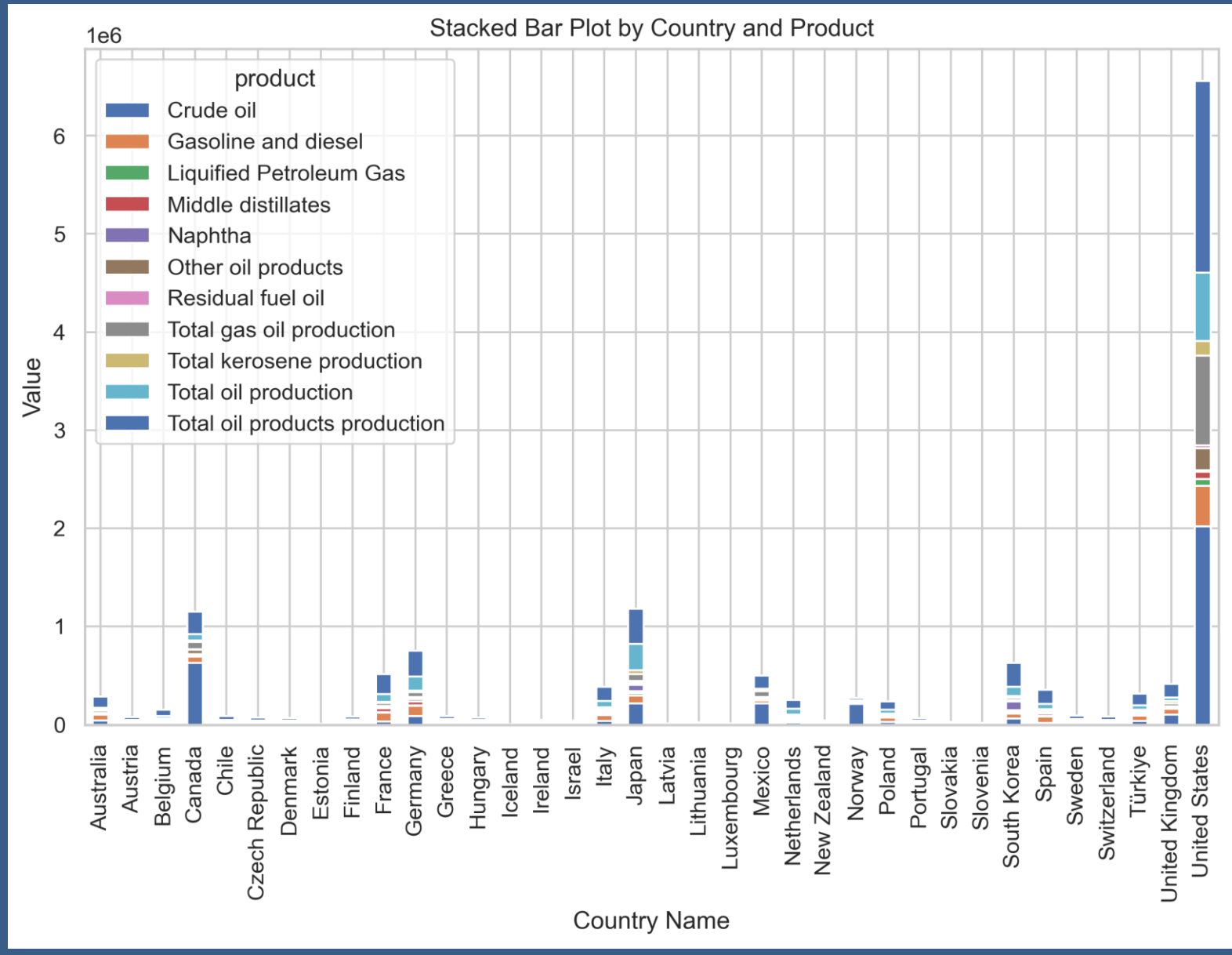
- Analyzing Feature (Independent Variables) Distributions





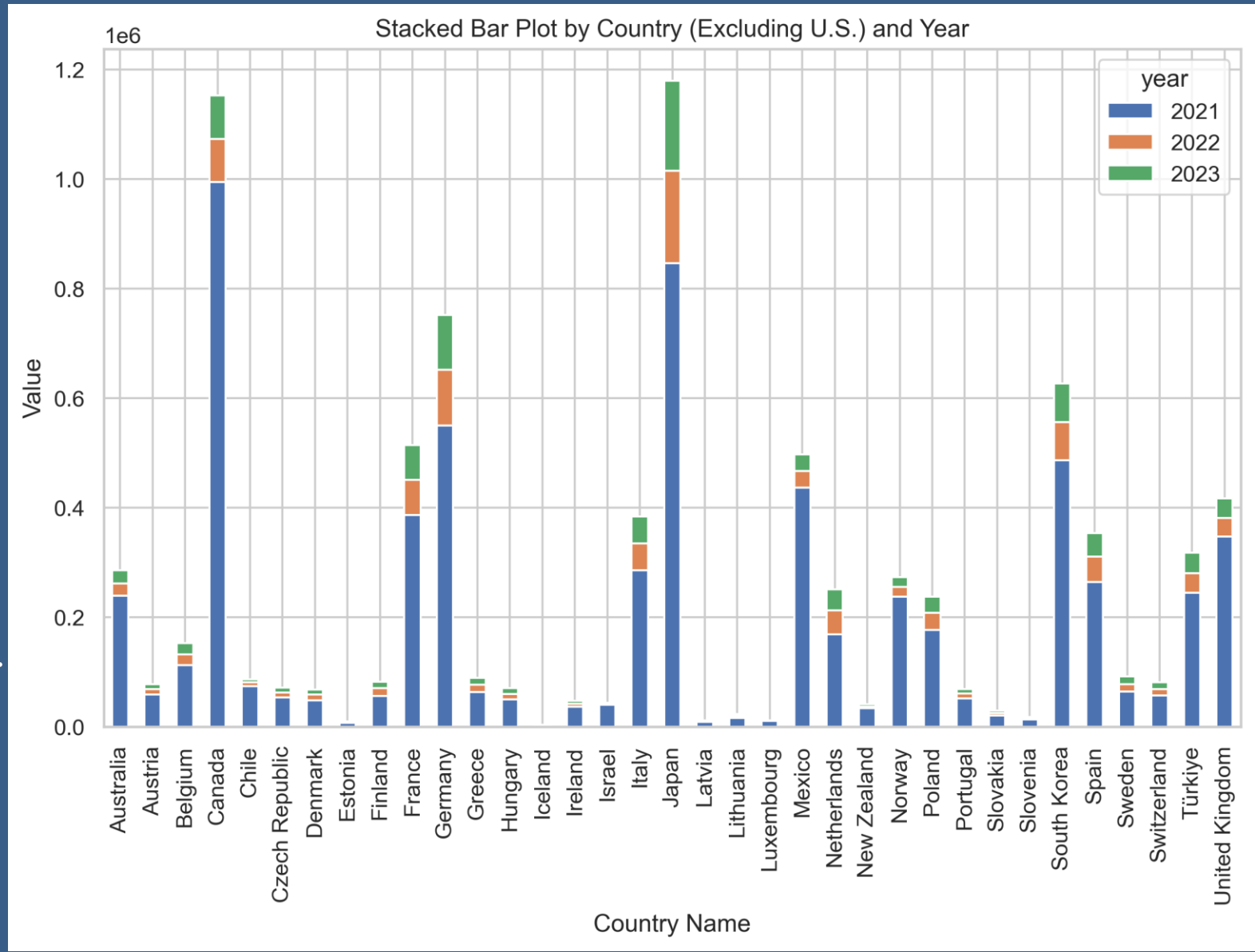
Exploratory Data Analysis

- Analyzing Feature (Independent Variables) Distributions



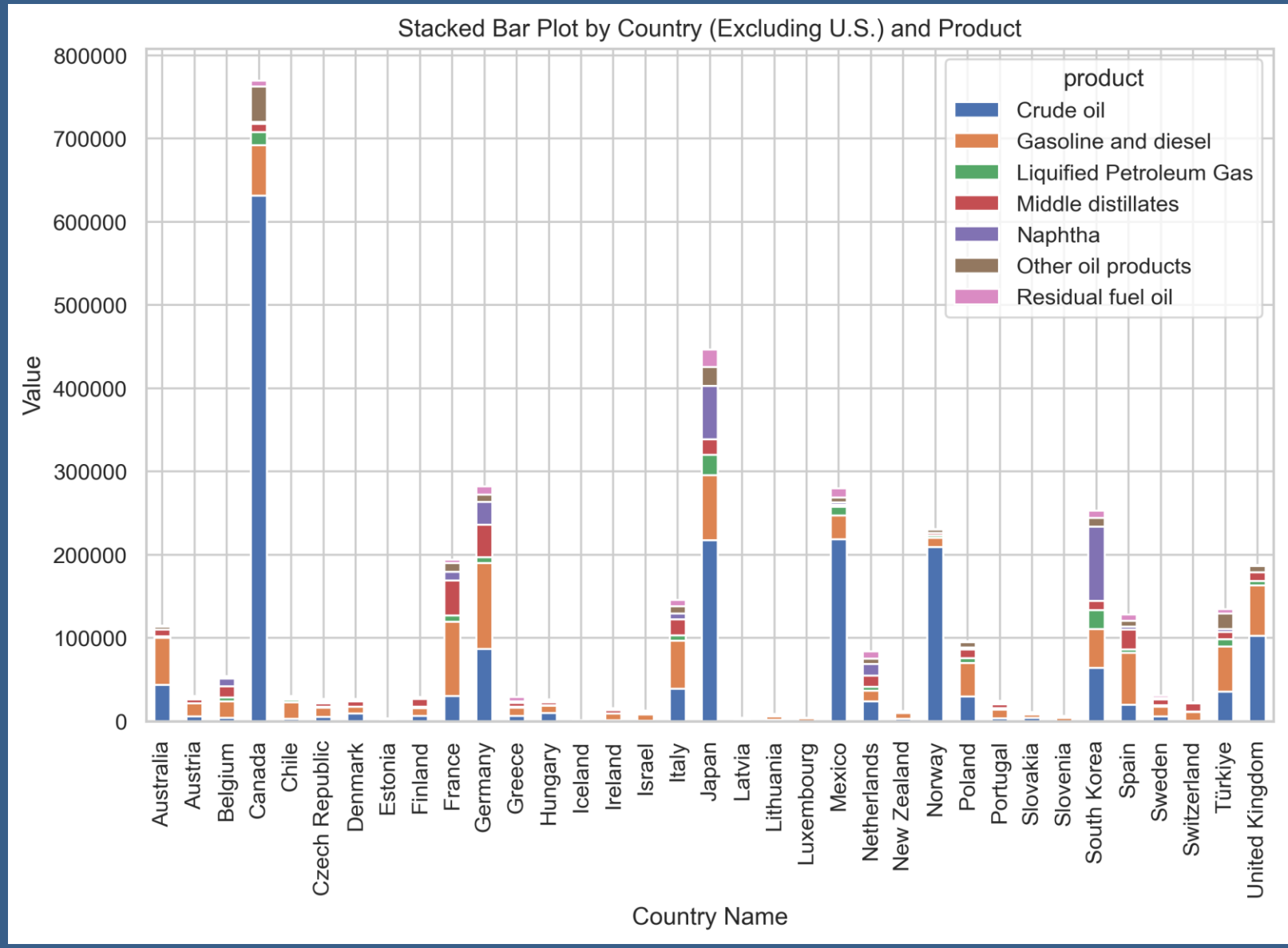
Exploratory Data Analysis

- Analyzing Feature (Independent Variables) Distributions outside of U.S.



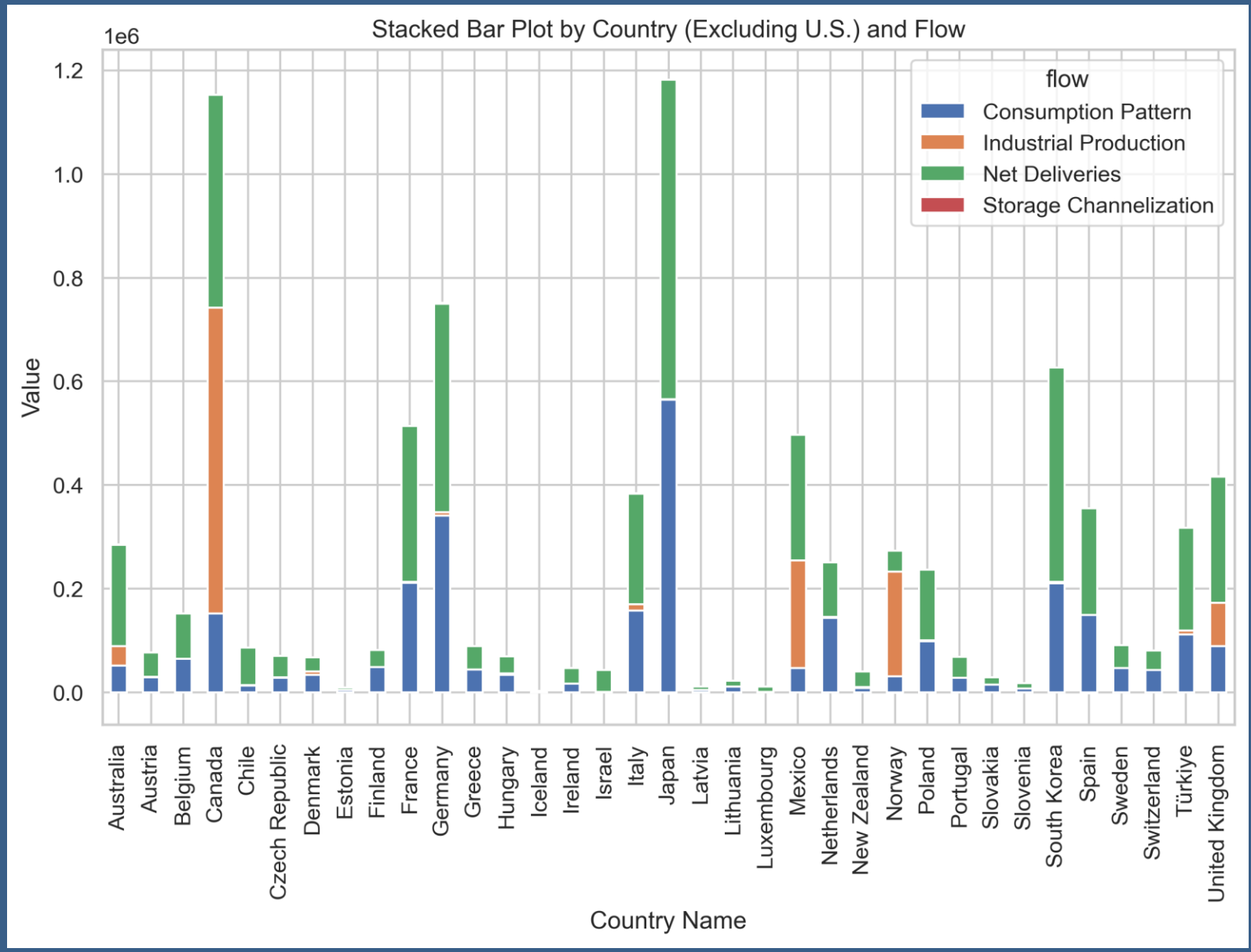
Exploratory Data Analysis

- Comparing values by product outside of U.S.



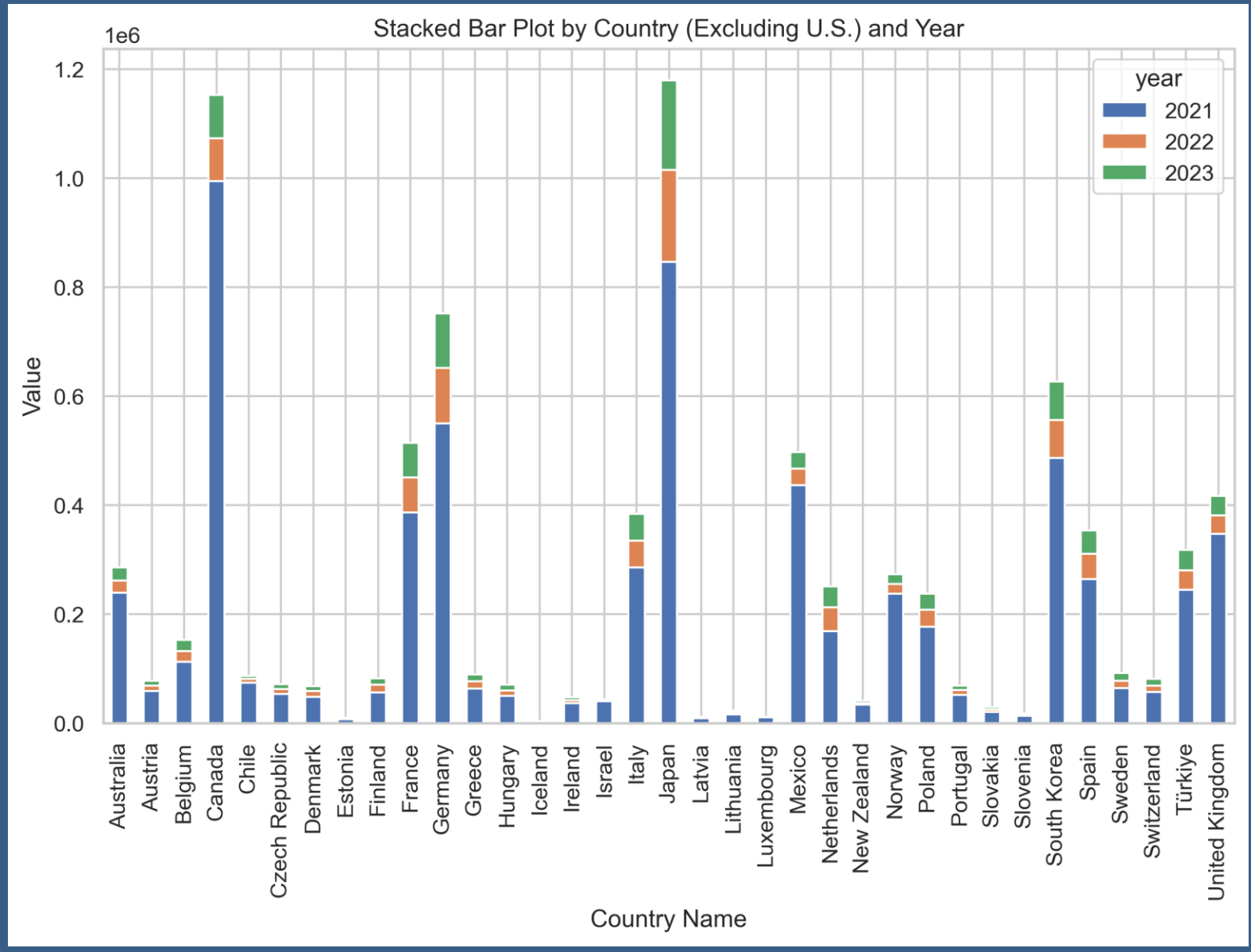
Exploratory Data Analysis

- Comparing values by Flow features outside of U.S.



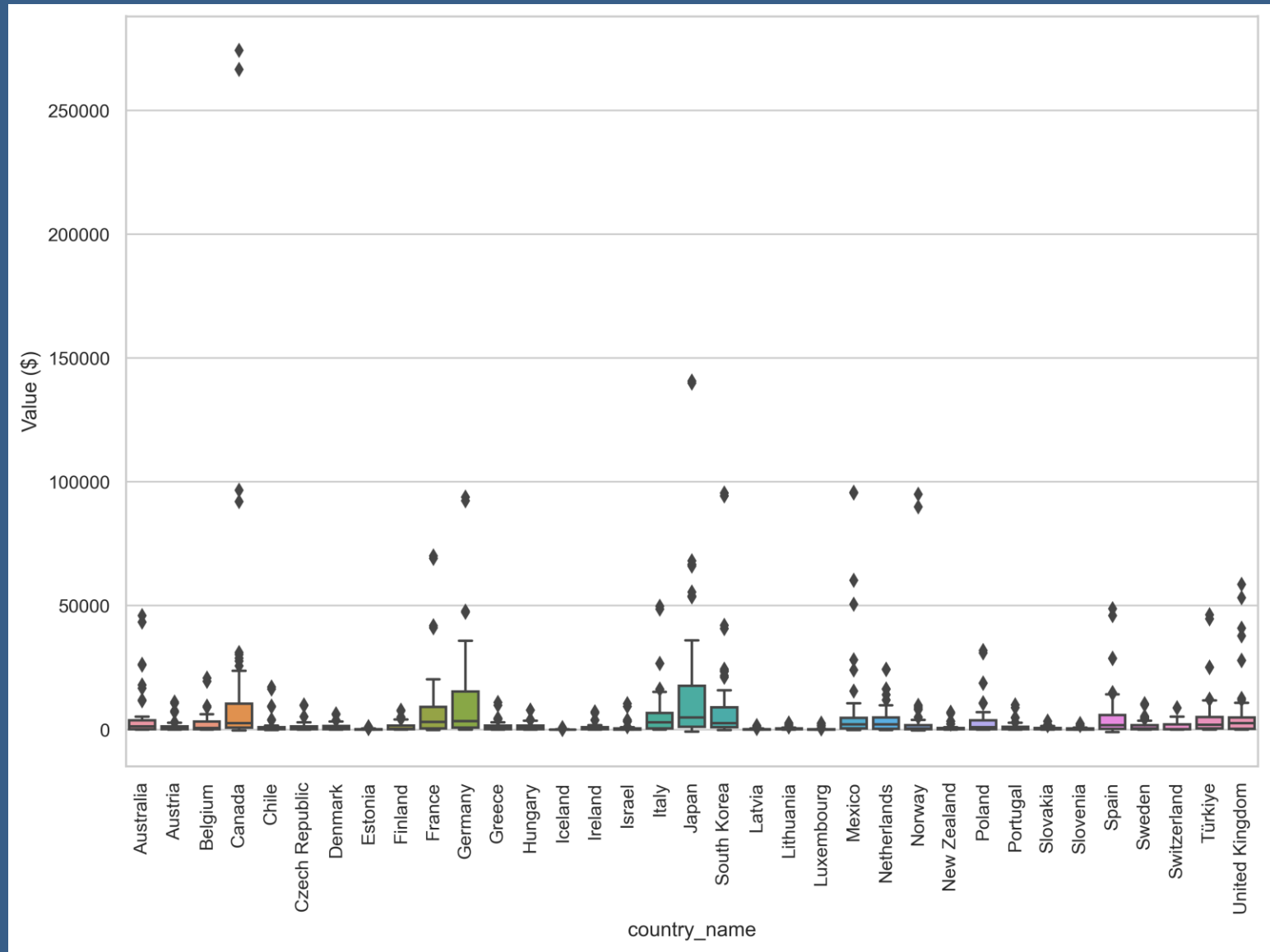
Exploratory Data Analysis

- Yearly distributions of production values outside of U.S.



Exploratory Data Analysis

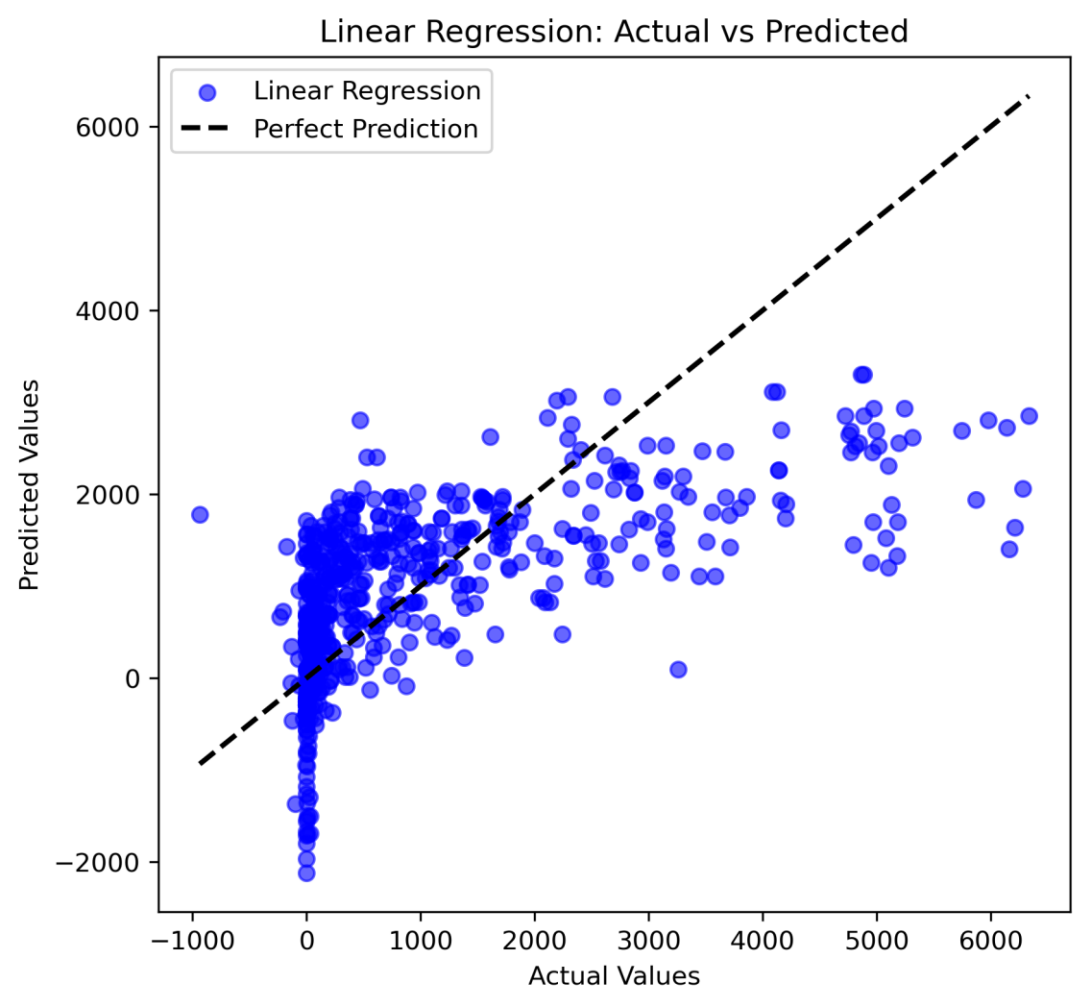
- Analyzing features for outliers
  - Canada has exaggerated outlier compared to other countries

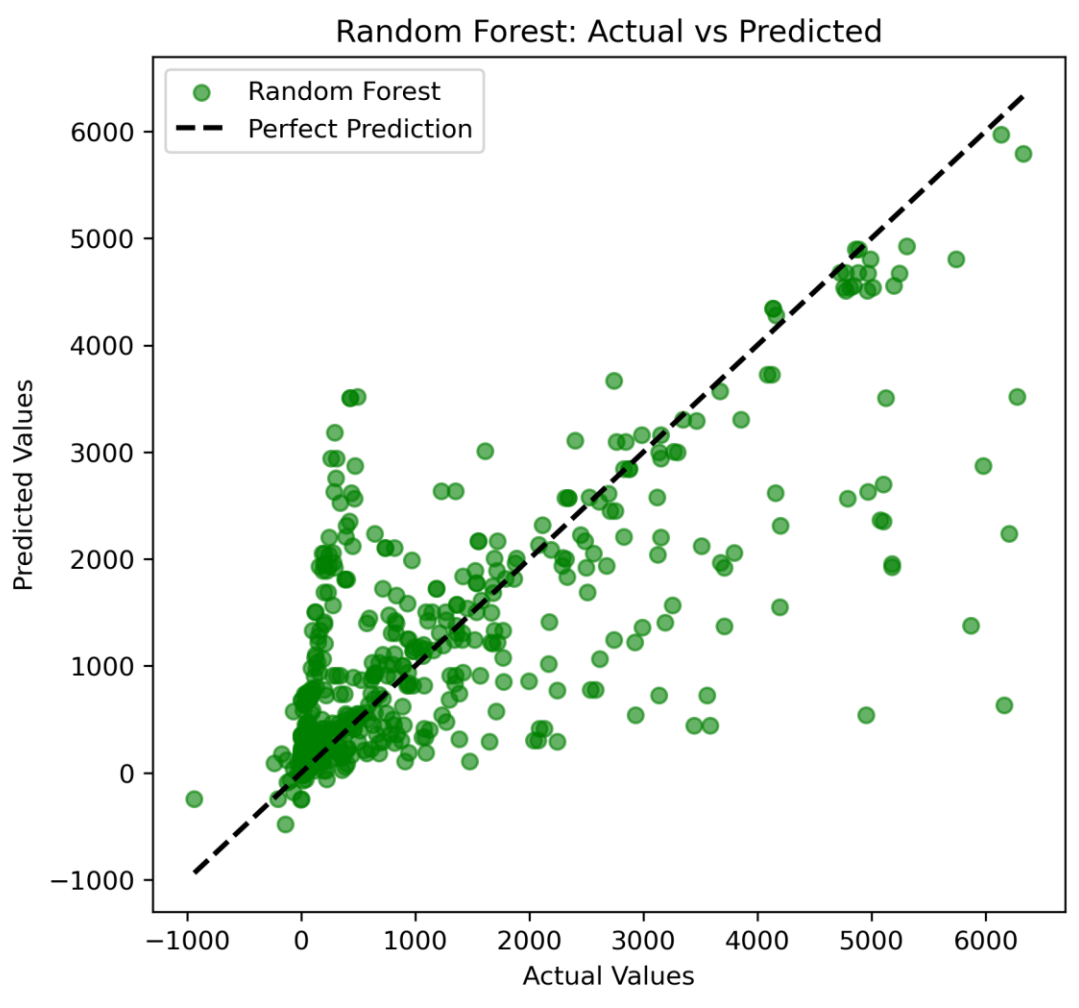


Modeling Evaluation

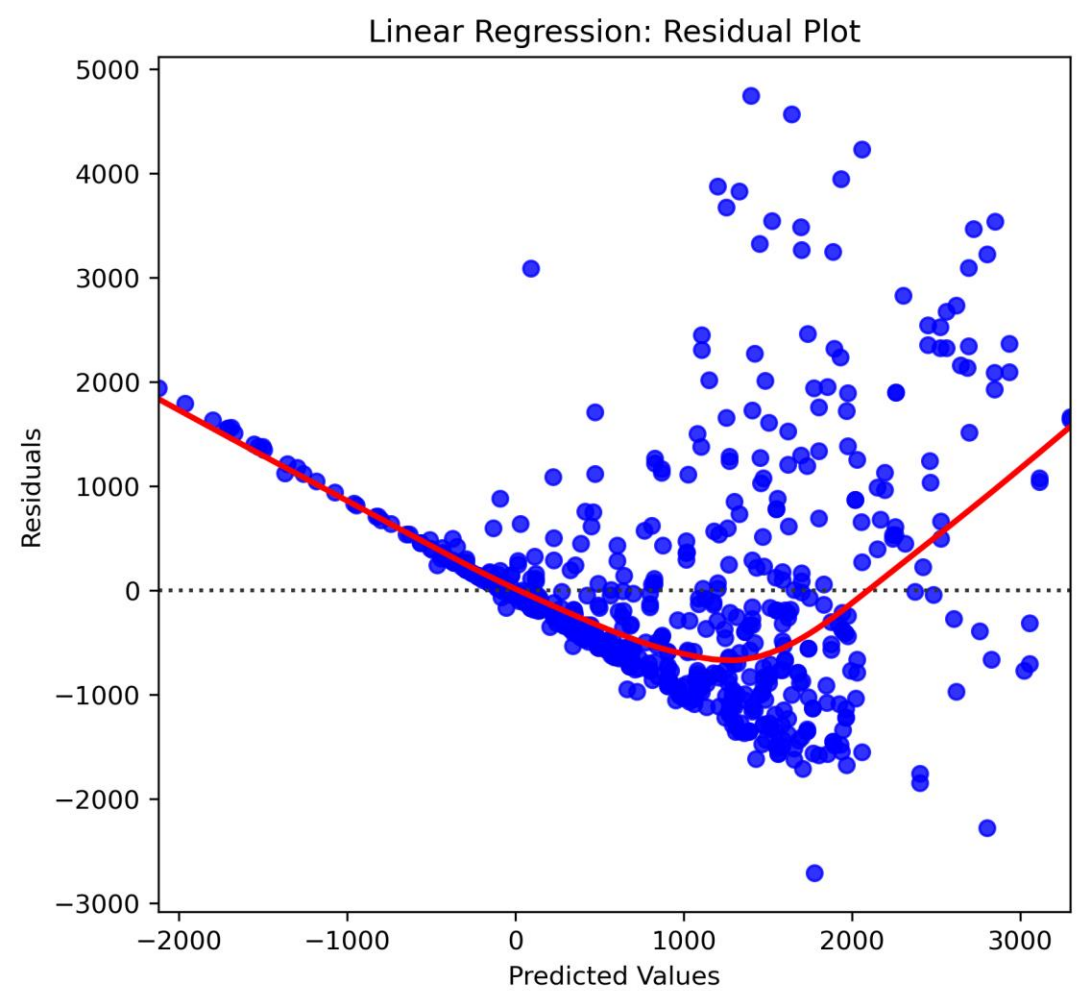
- **Compared models:**
- **Linear Regression**
  - **Evaluation:**
    - Mean Absolute Error: 803.25
    - Mean Squared Error: 1213187.81
    - R^2 Score: 0.39
- **Random Forest Regressor**
  - **Evaluation:**
    - Mean Absolute Error: 515.6194
    - Mean Squared Error: 828804.4951
    - R^2 Score: 0.59

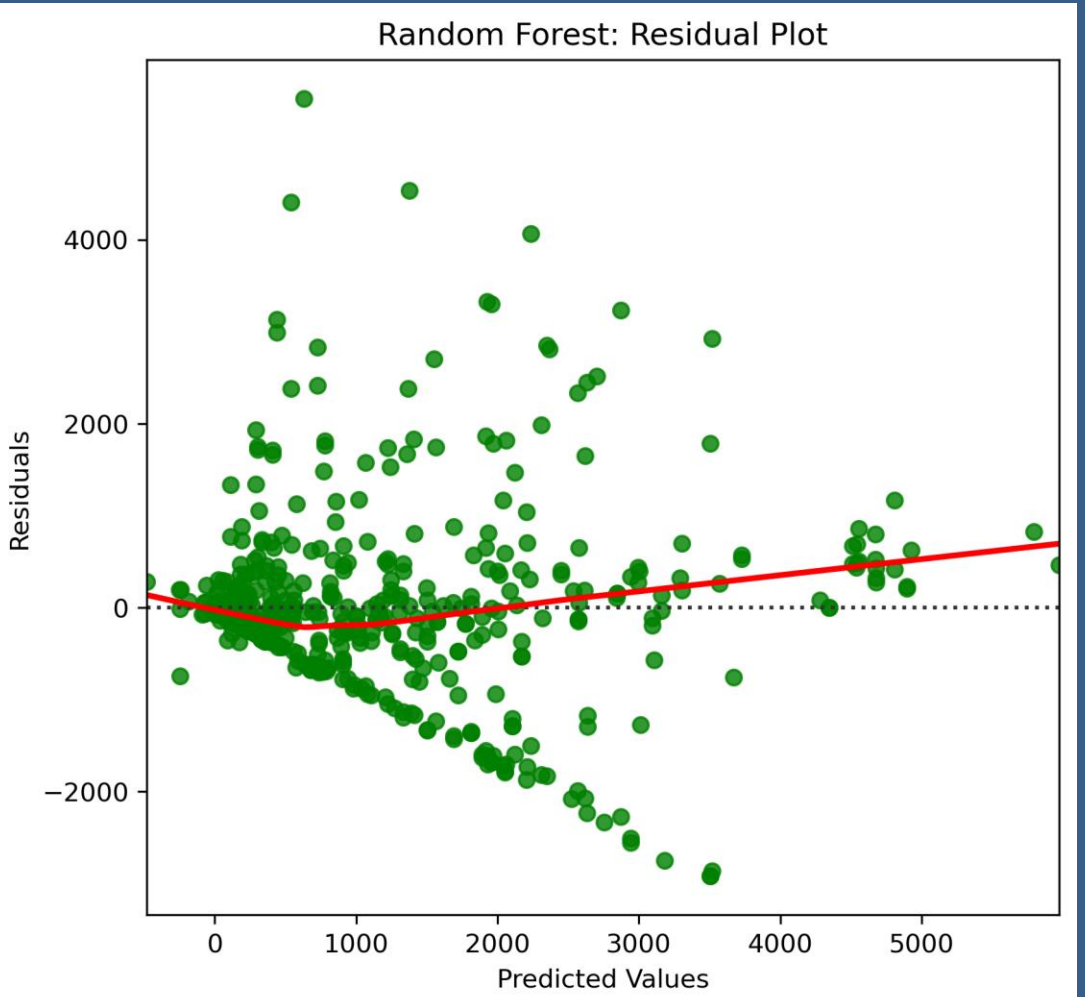
Predictions vs. Actual Values Evaluation



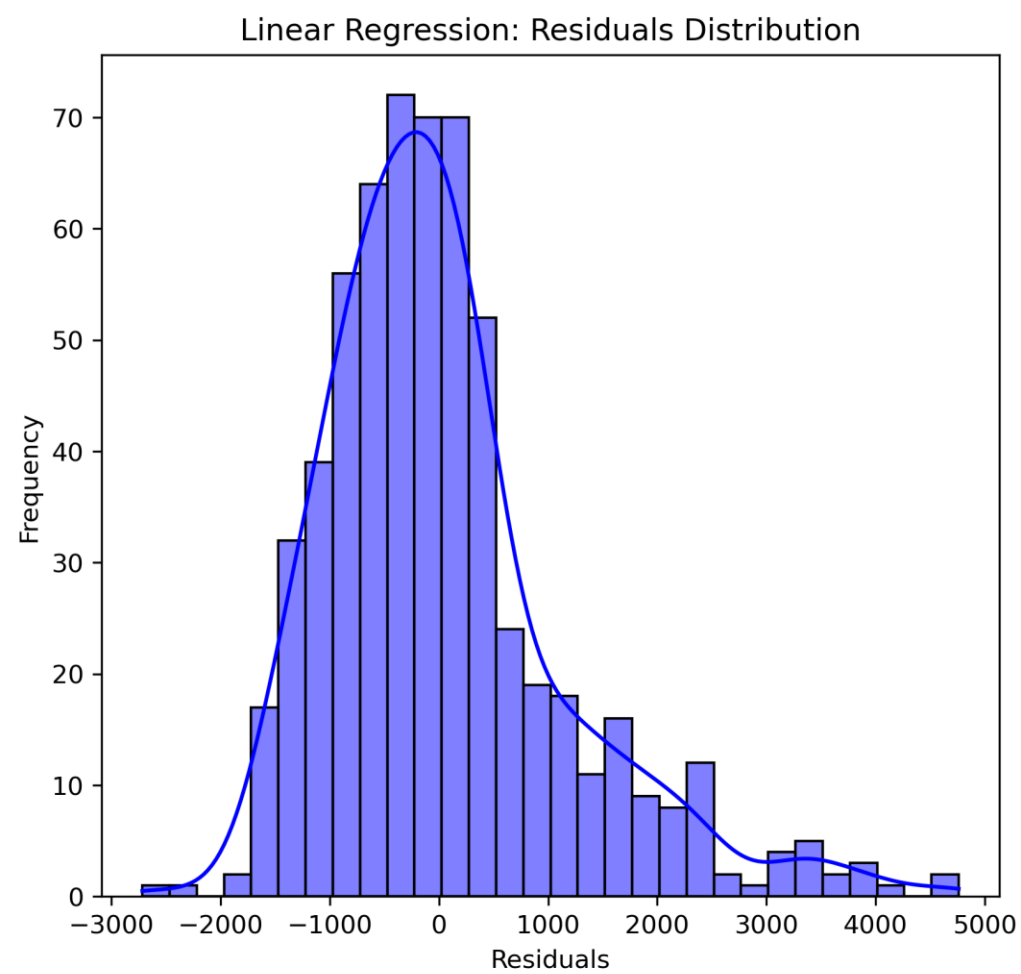


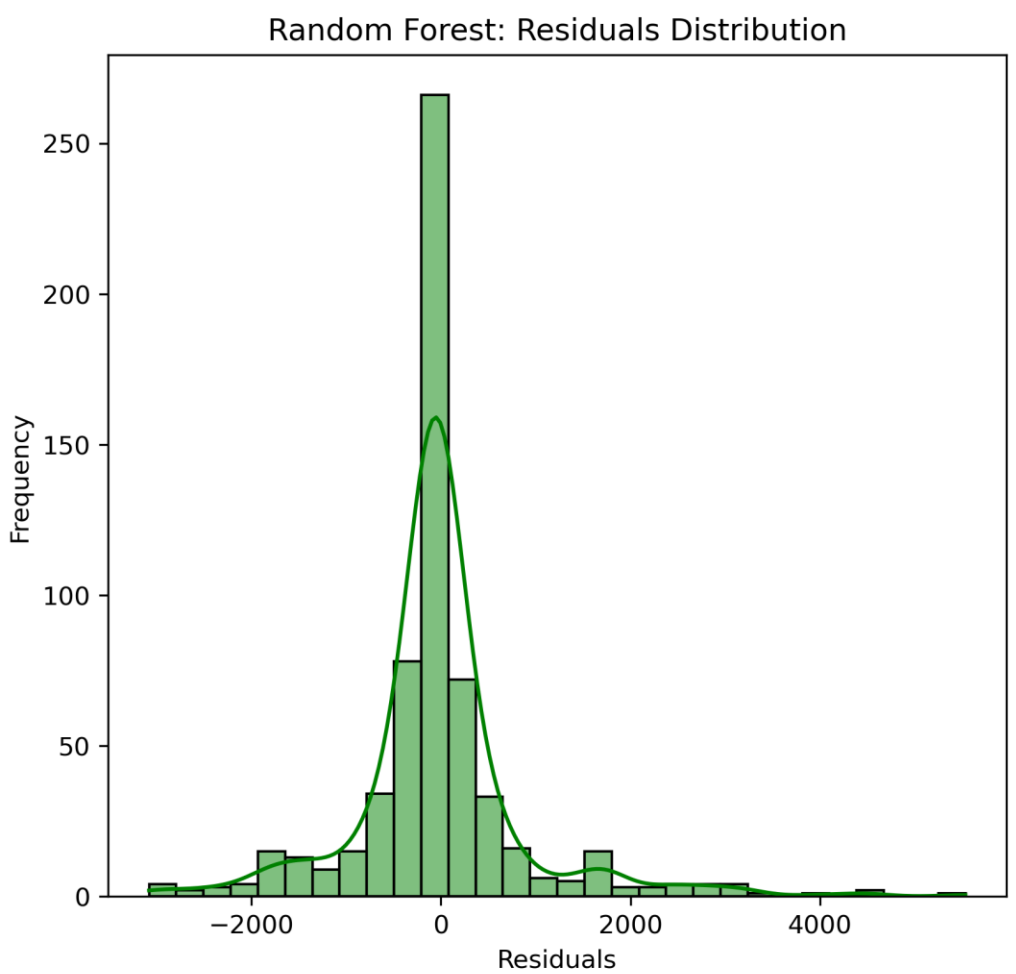
Model Residuals Plot Evaluation





Model Residuals Distribution Evaluation





Takeaways & Final Tips:
• Model Performance

- The Random Forest Regressor outperformed the Linear Regression model in terms of both R^2 score (0.59 vs. 0.39) and error metrics. The Random Forest model captured non-linear relationships and interactions between features better than the simpler linear model, making it more suitable for this type of analysis.

- The Linear Regression model provided a baseline but had a relatively low R^2 score, suggesting that linear assumptions might not be sufficient to model the complexities of the oil production data.

Takeaways & Final Tips:
• Investment Insights

- Countries with stable or growing production, such as Canada, Saudi Arabia, and Russia, are likely to be attractive investment targets.
- However, production declines in other countries could point to regions where there is potential for future growth if investment is directed toward improving infrastructure and technological innovation.

THANK YOU

Ammanuel F. WoldeAregay