



Oil Profitability Analysis

By
Ammanuel F. WoldeAregay

A decorative graphic in the top-left corner of the slide. It consists of several overlapping geometric shapes: a light blue circle, a dark blue square with concentric circles, a purple triangle, a pink square with concentric circles, and a grey square with concentric circles. A small dark blue circle is positioned at the intersection of the purple triangle and the pink square.

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



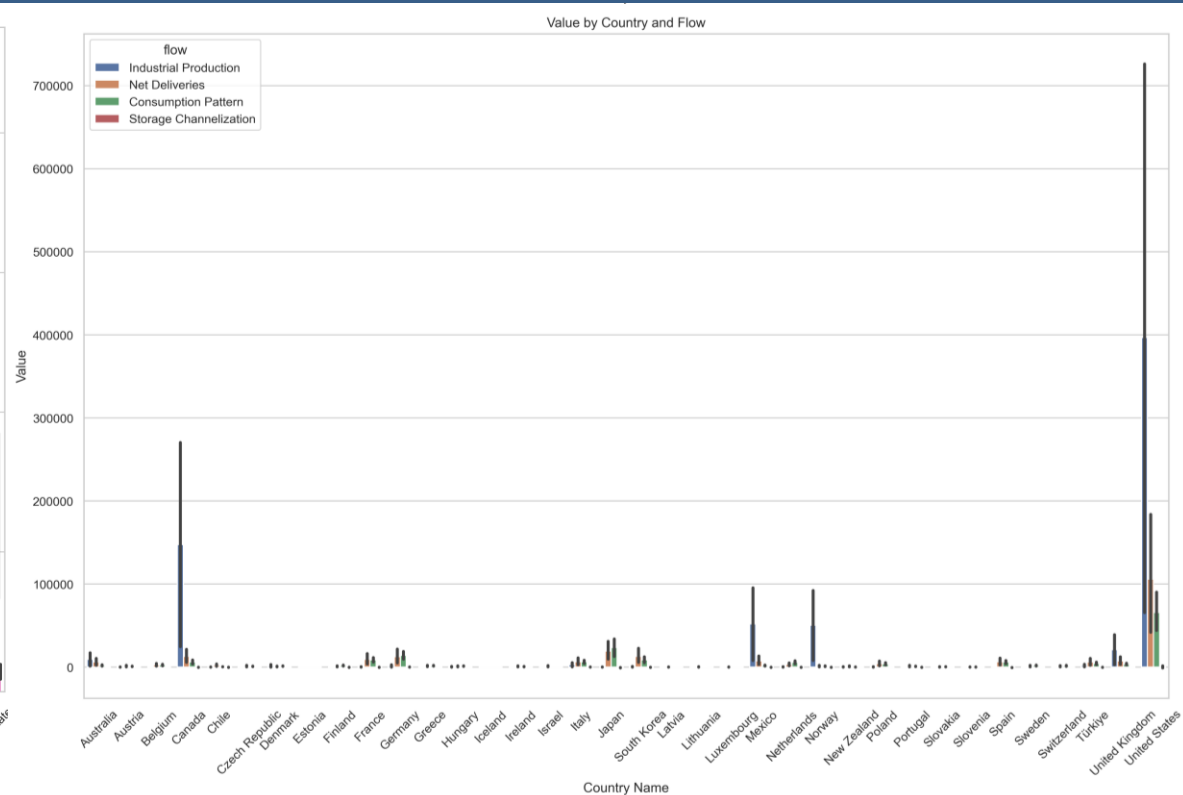
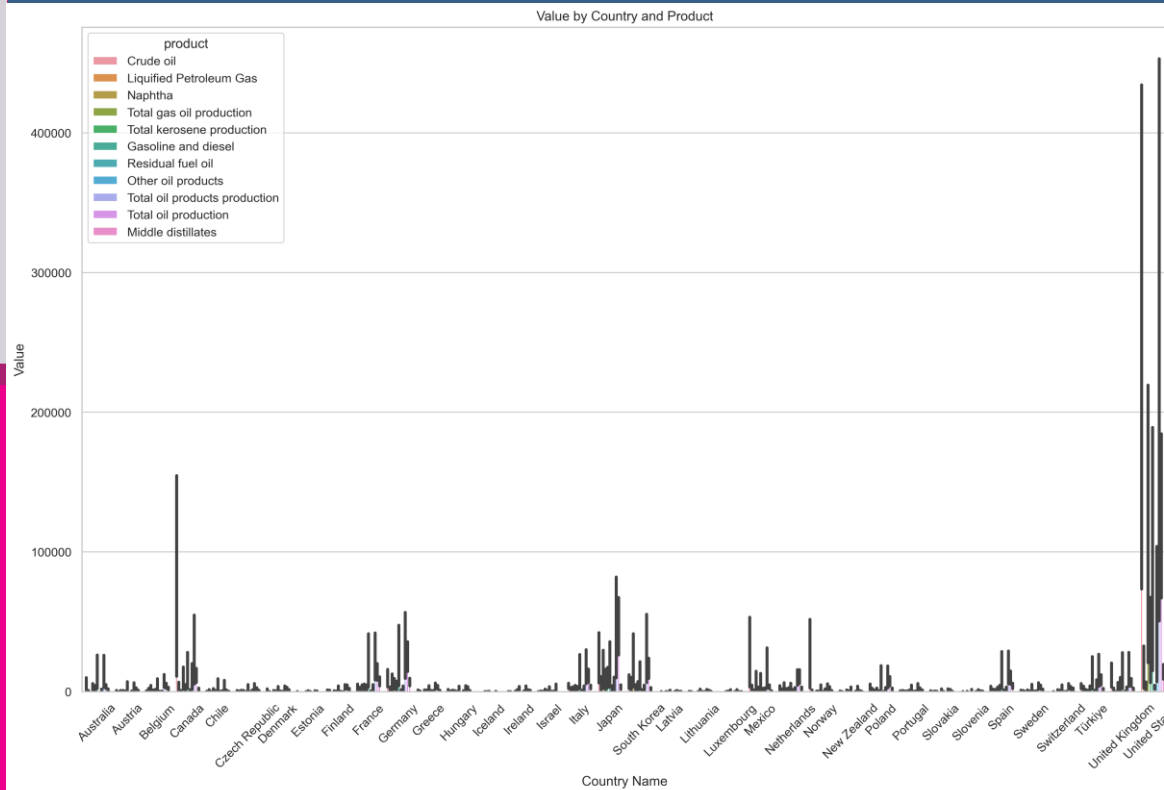
Data Wrangling

Original dataset had
2376 rows and Rows
6 columns

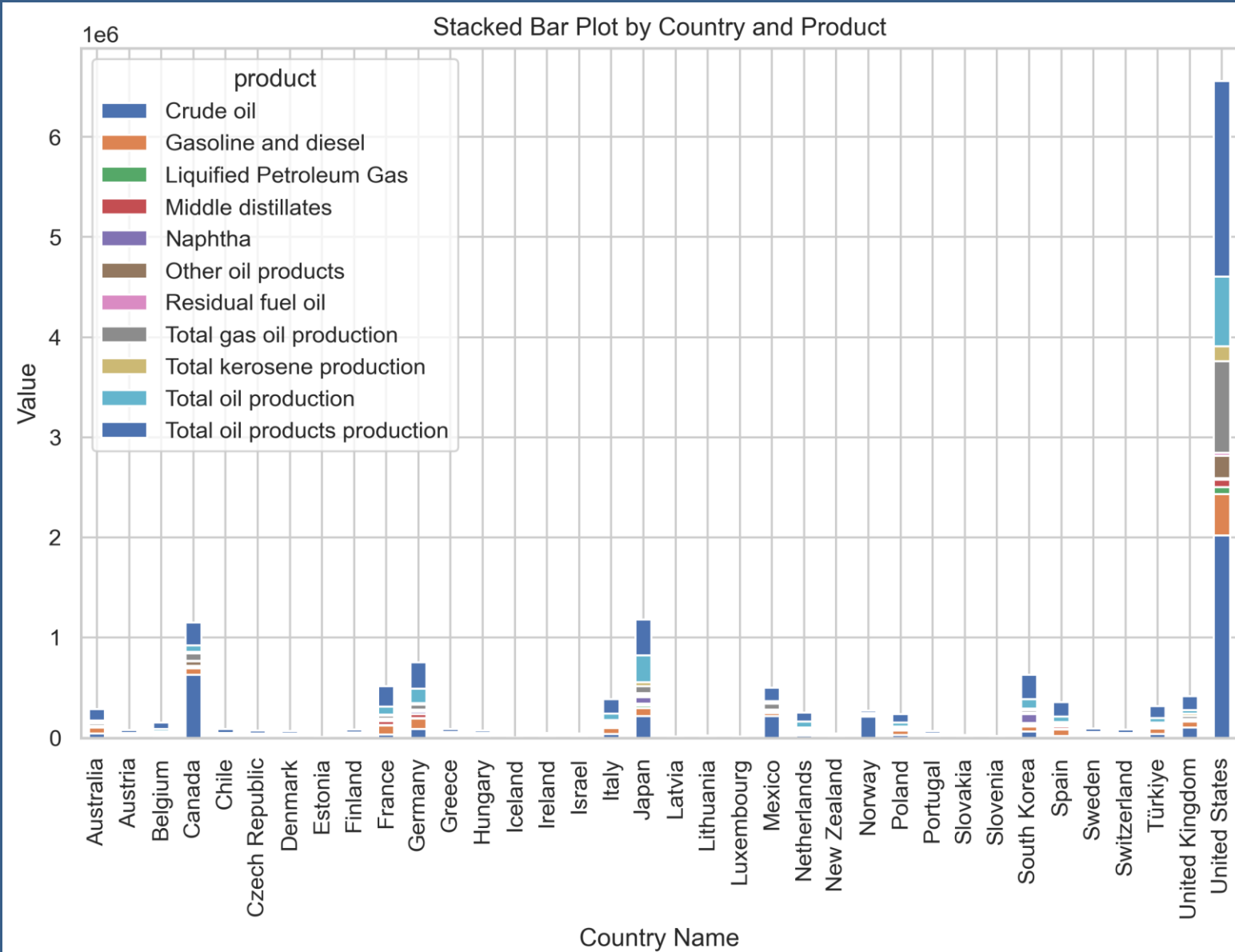
- ❖ Drop unnecessary columns
- ❖ Assess the raw data
- ❖ Display data samples to understand its structure
- ❖ Data Cleaning: Identify and handle missing values
- ❖ Target Variable: Production value

Exploratory Data Analysis

❖ Analyzing Feature (Independent Variables) Distributions

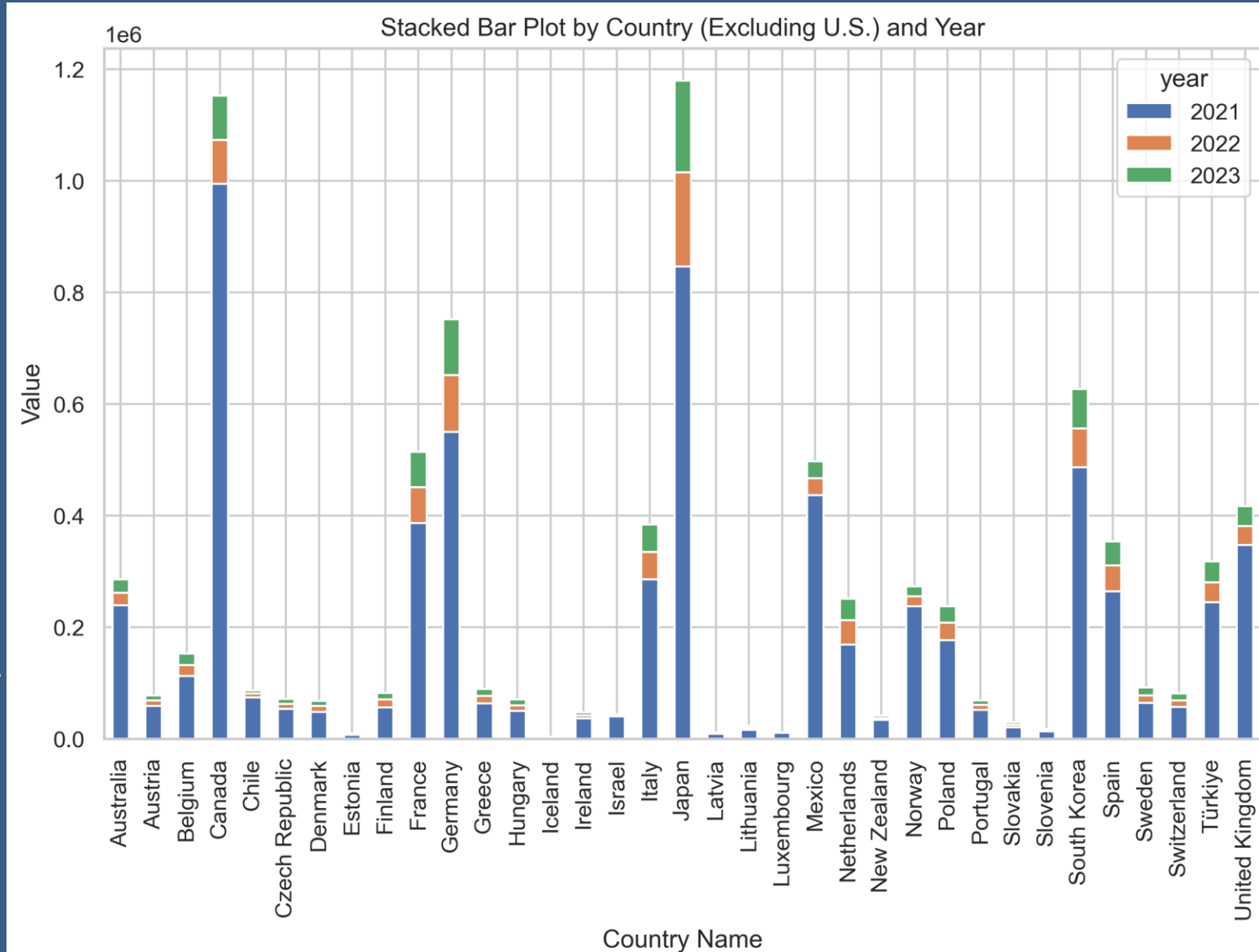


❖ Analyzing Feature (Independent Variables) Distributions



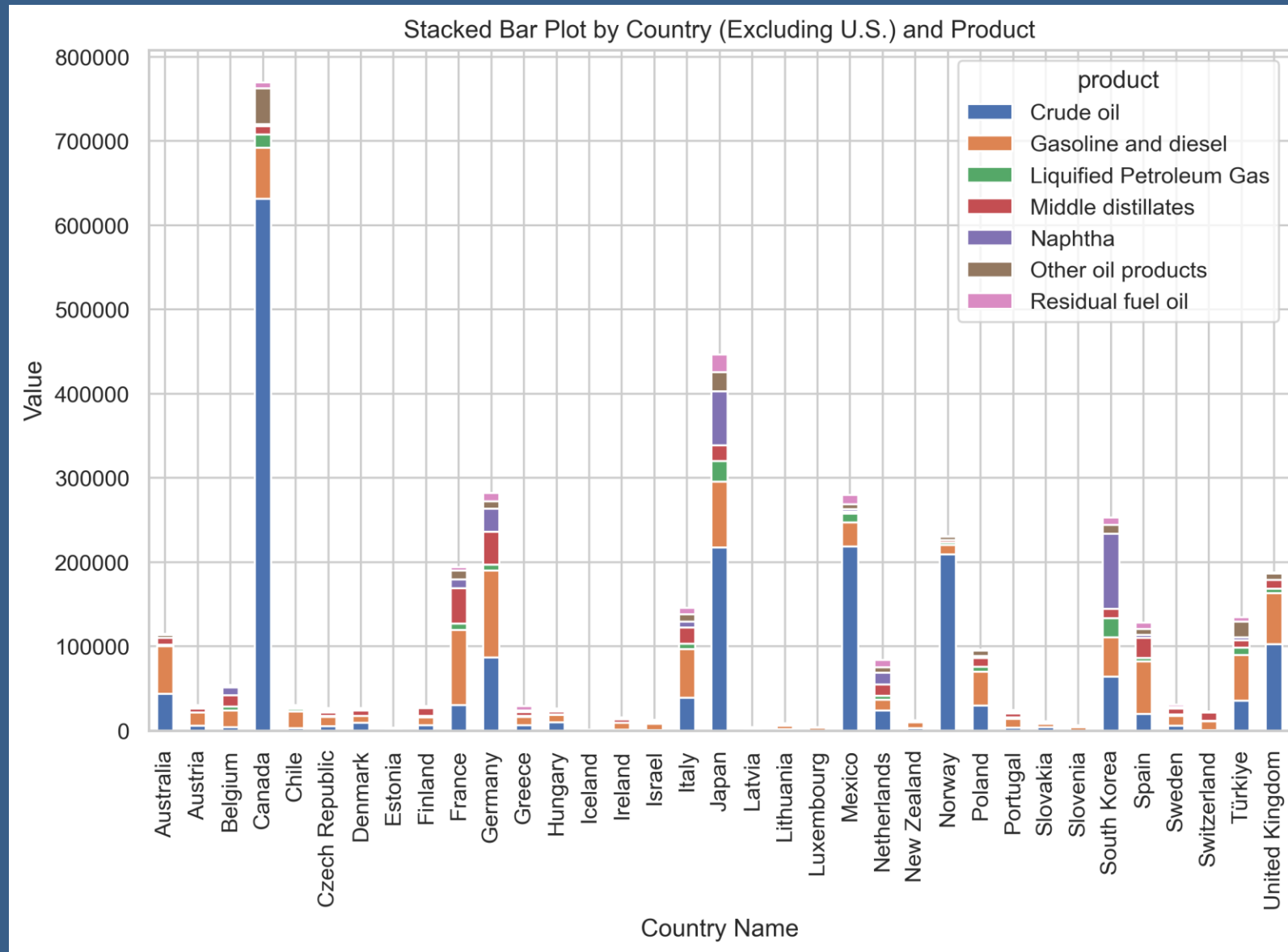
Exploratory Data Analysis

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



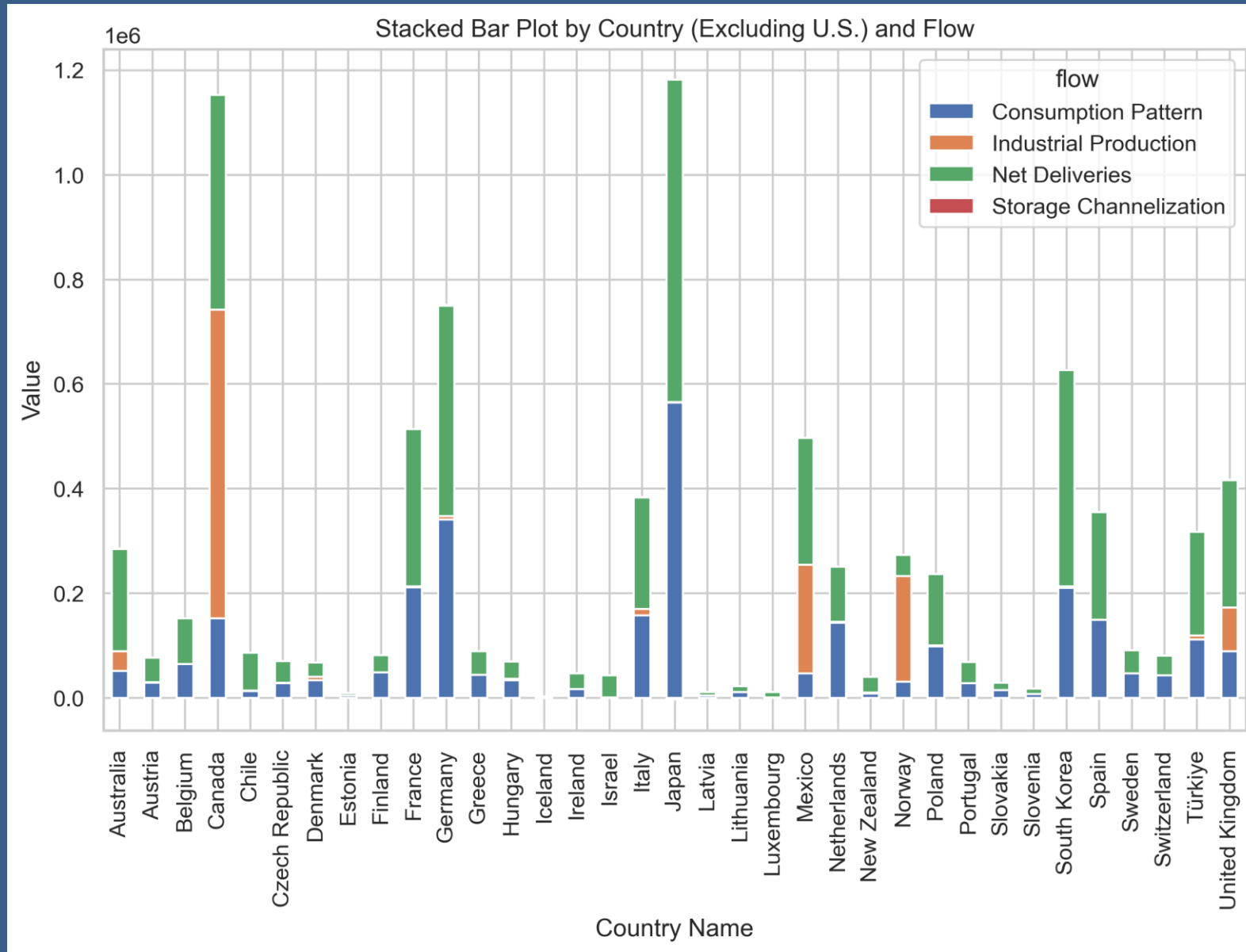
Exploratory Data Analysis

- ❖ Comparing values by product outside of U.S.



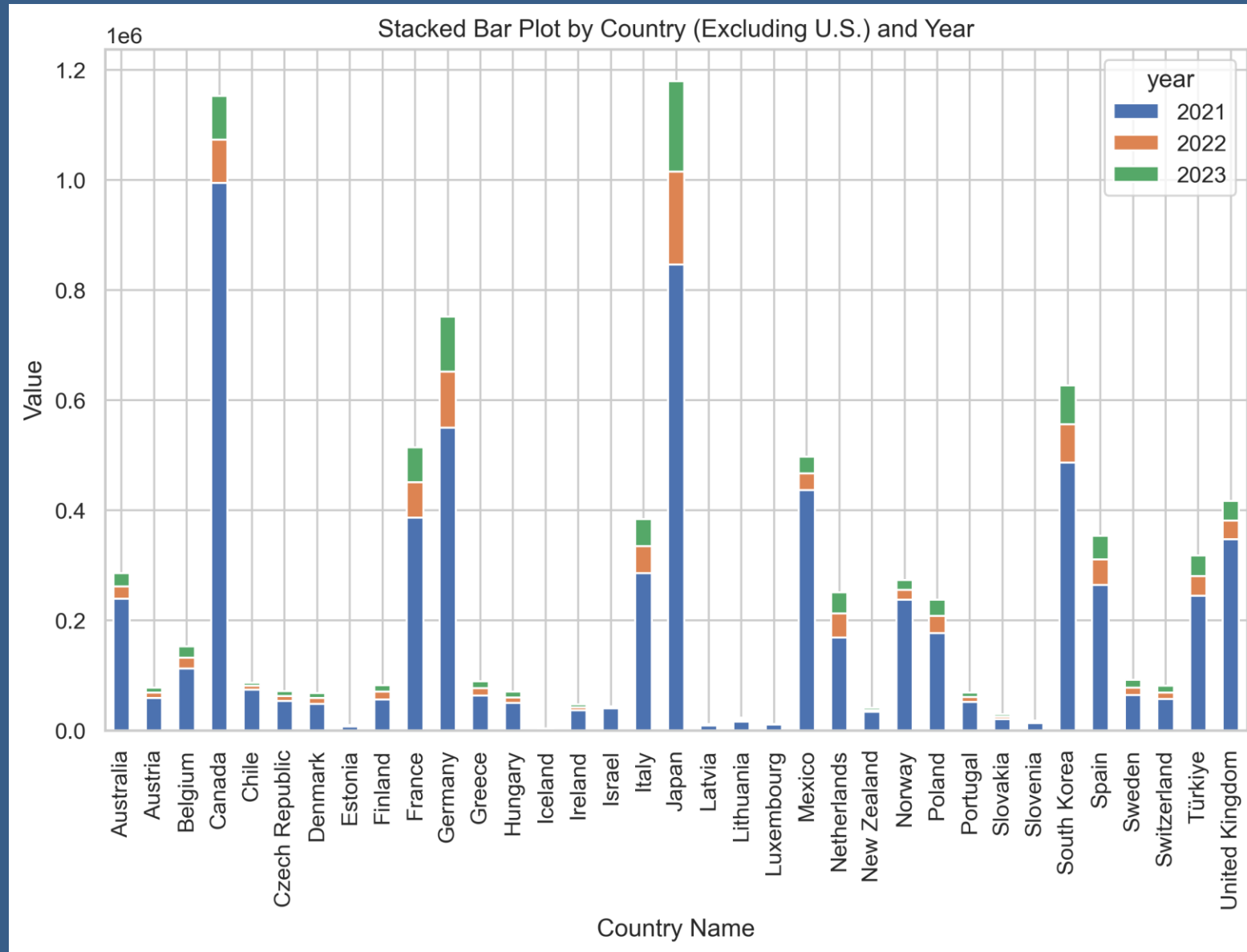
Exploratory Data Analysis

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



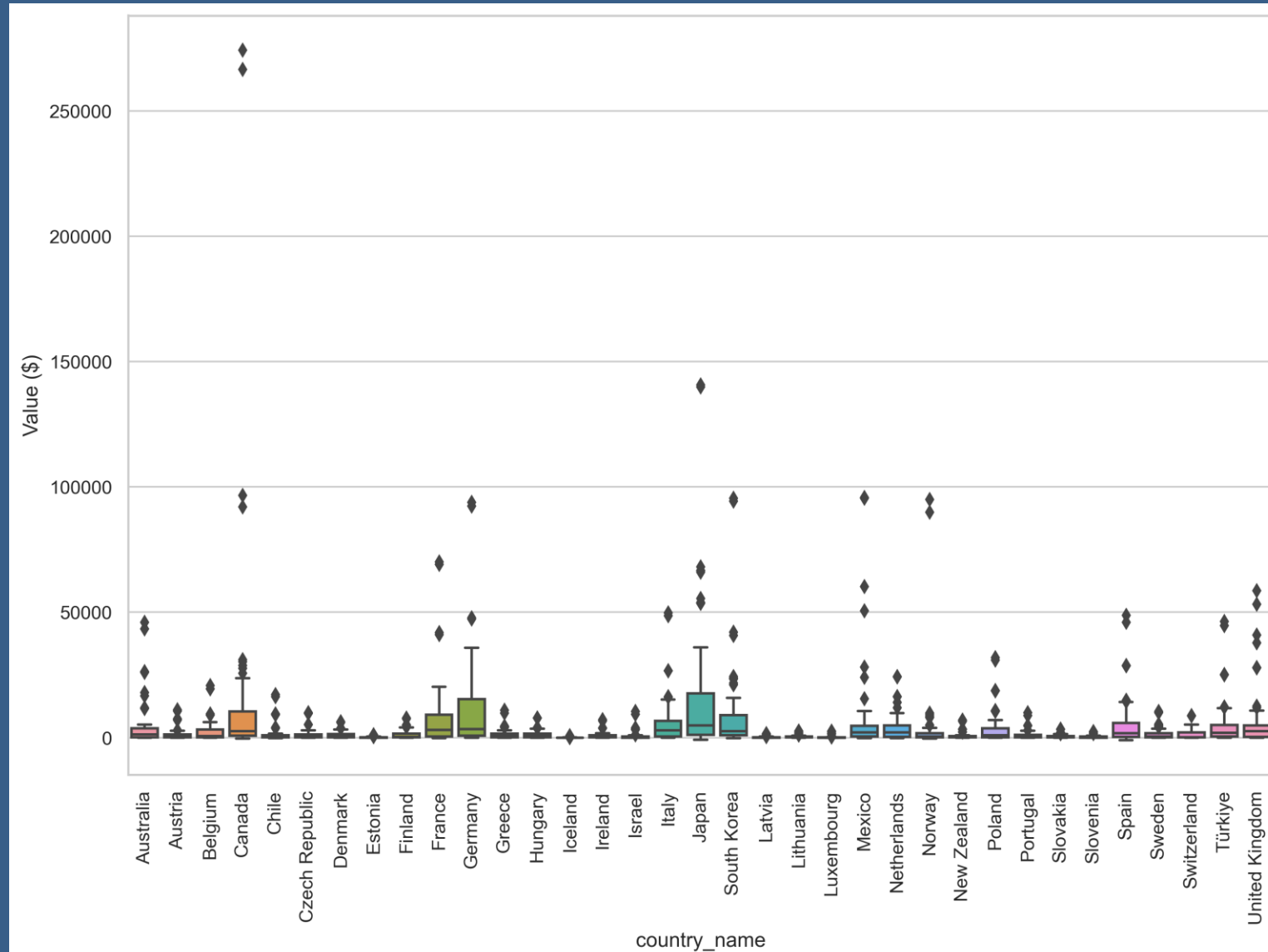
Exploratory Data Analysis

- ❖ 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

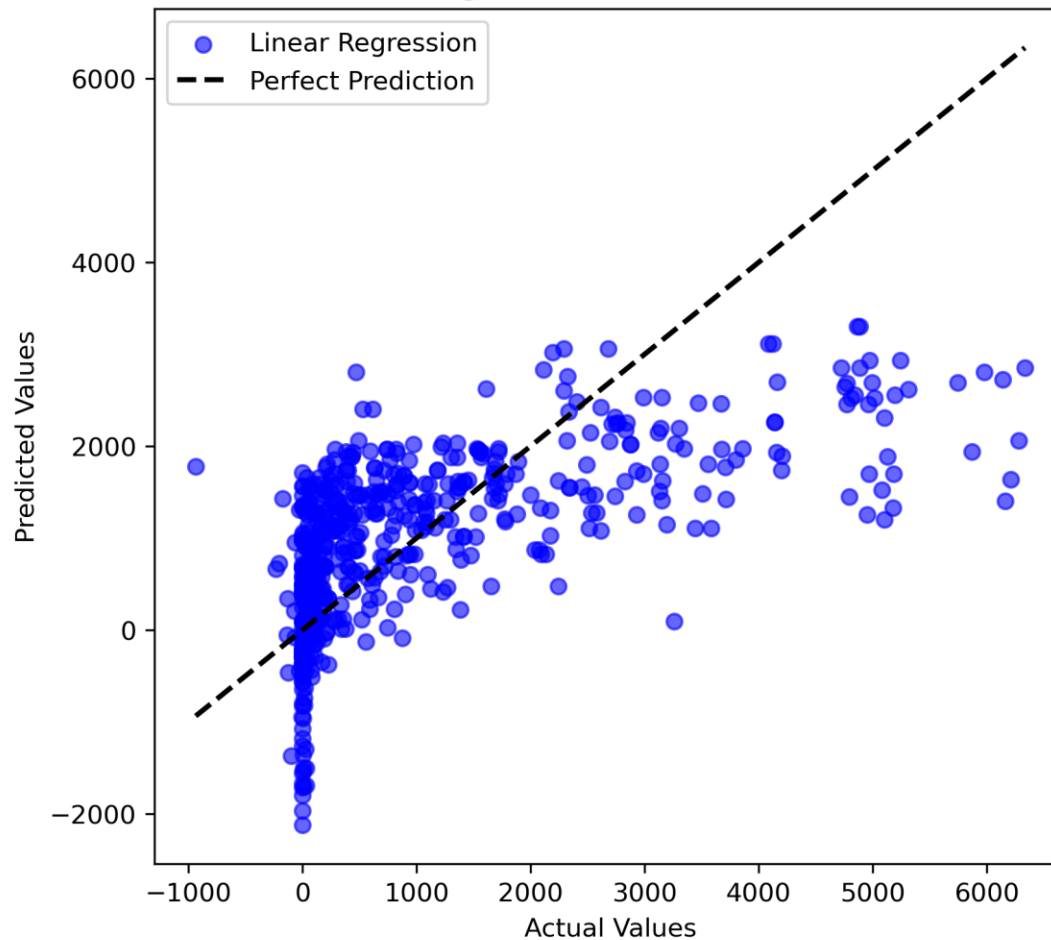
❖ Random Forest Regressor

❖ Evaluation:

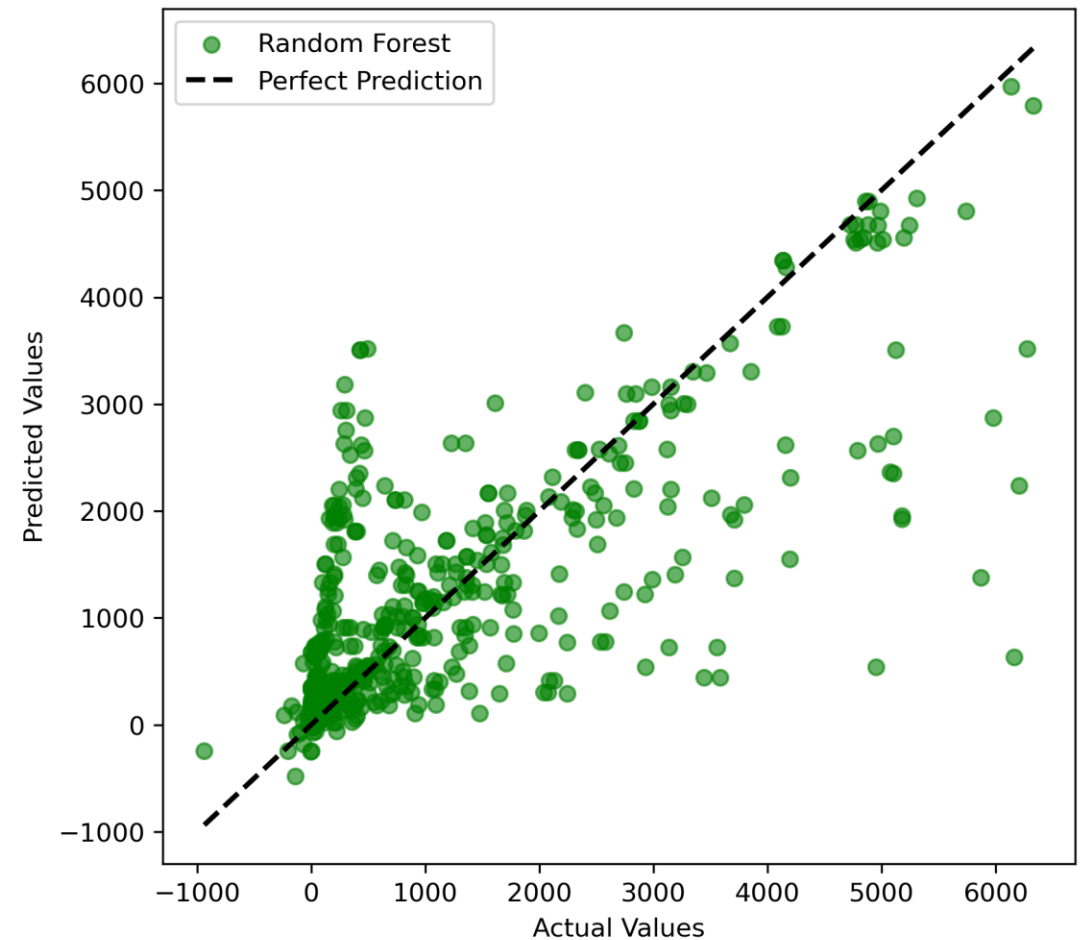
- Mean Absolute Error: 515.6194
- Mean Squared Error: 828804.4951
- R^2 Score: 0.59

Predictions vs. Actual Values Evaluation

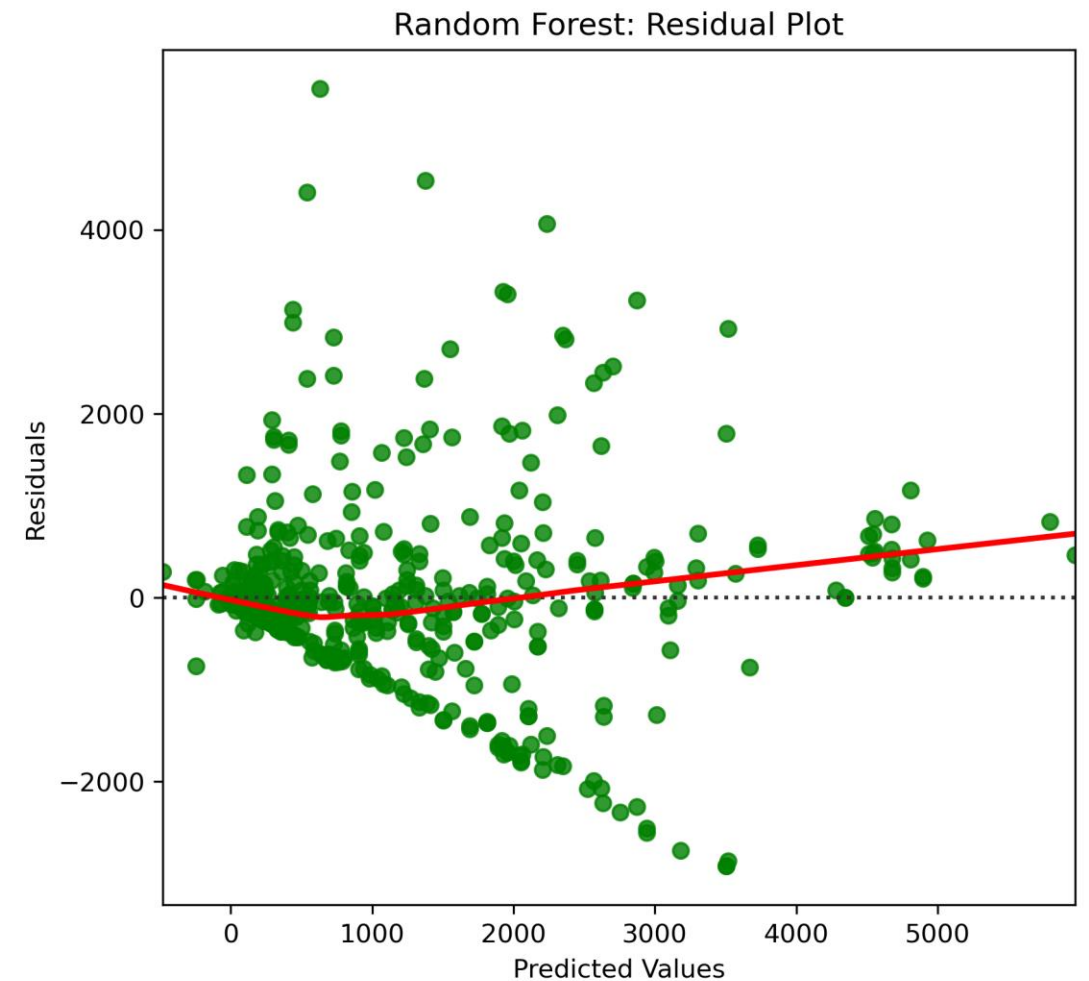
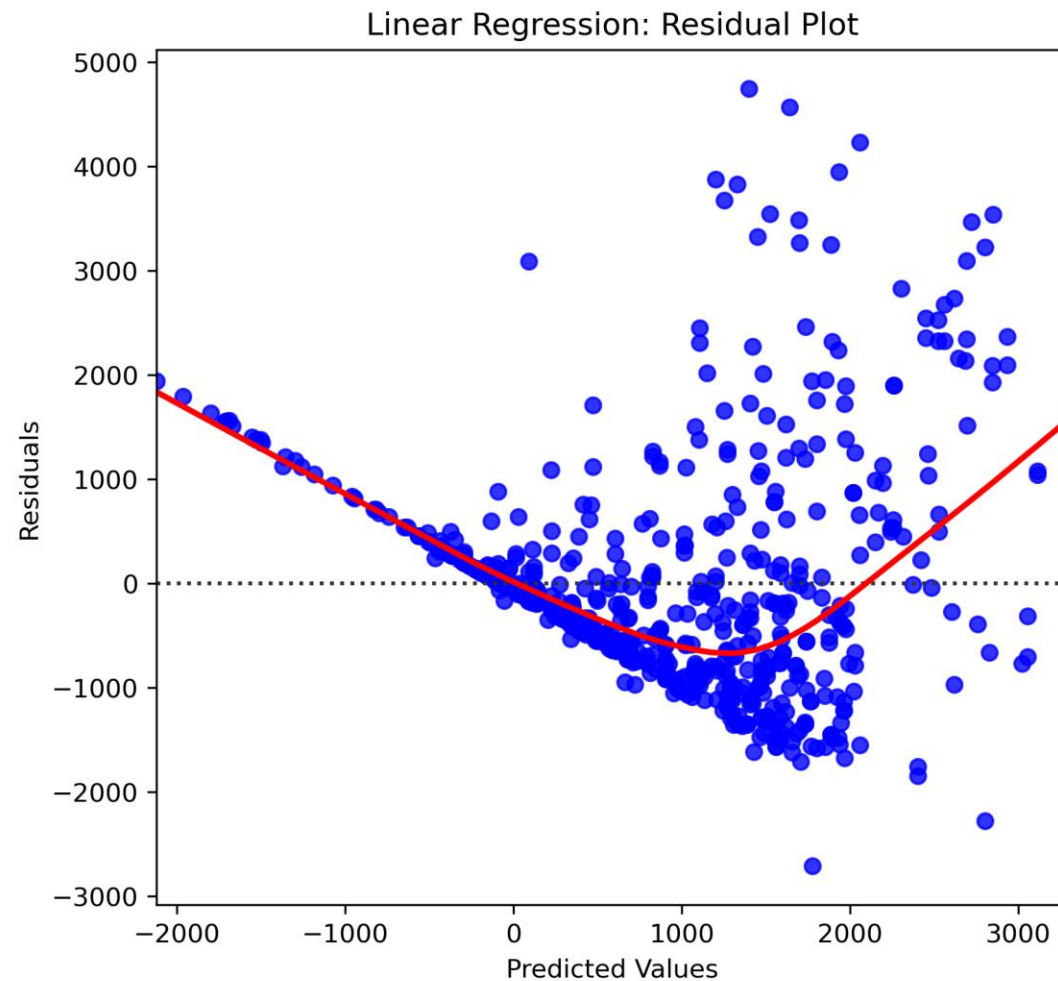
Linear Regression: Actual vs Predicted



Random Forest: Actual vs Predicted

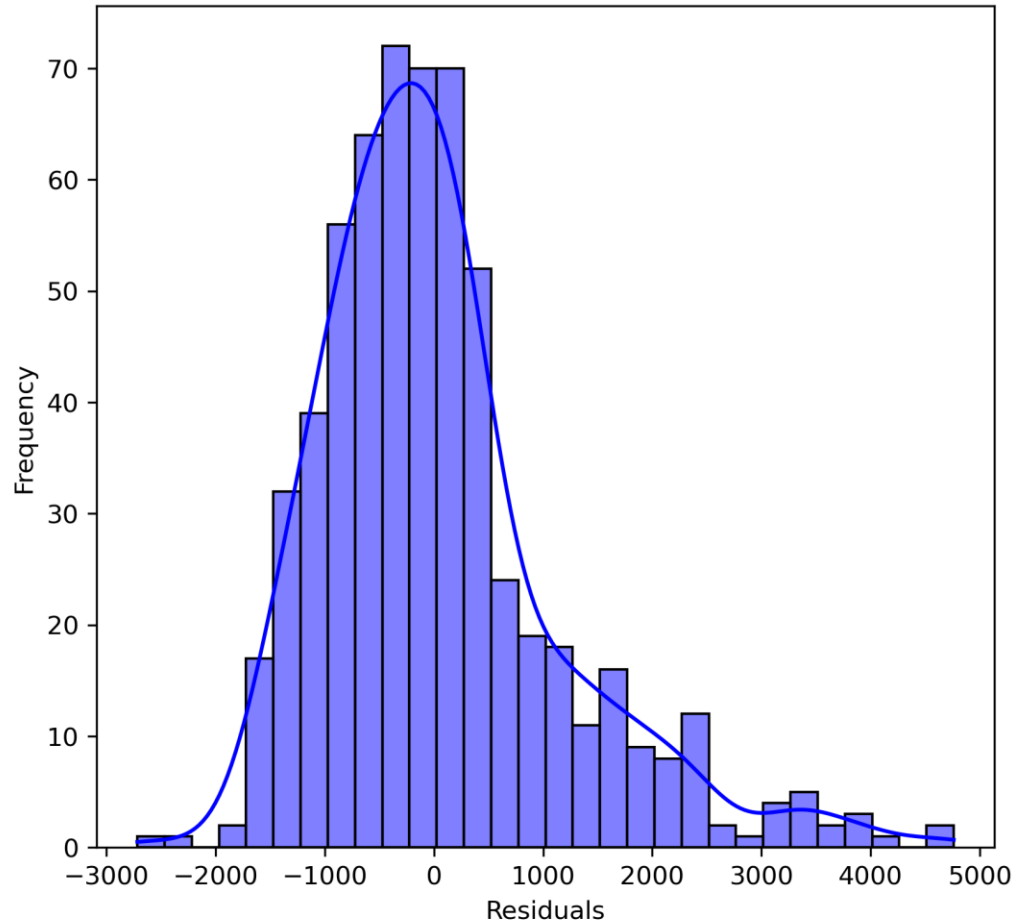


Model Residuals Plot Evaluation

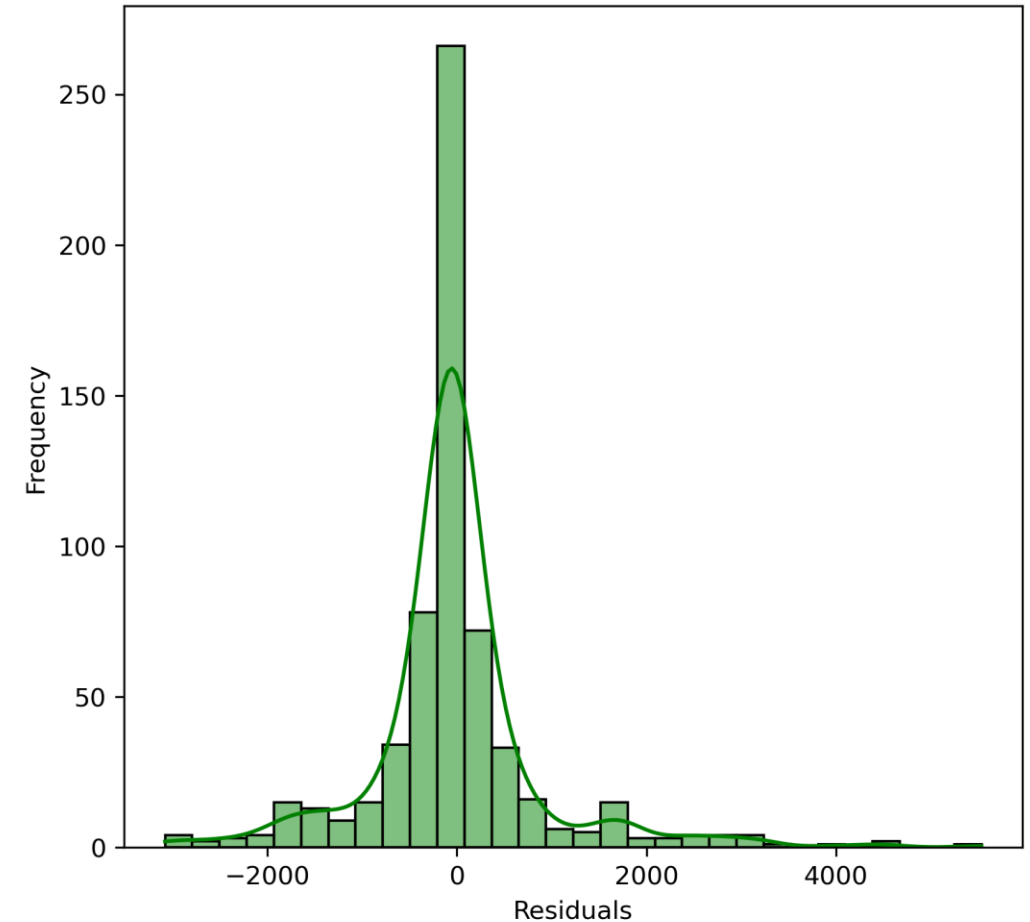


Model Residuals Distribution Evaluation

Linear Regression: Residuals Distribution



Random Forest: Residuals Distribution





Takeaways & Final Tips

- ❖ Random Forest Regressor is the preferred model with R^2 of 0.59.
- ❖ An R^2 score of 0.59 indicates that your model explains 59% of the variance in the target variable, which means the model is performing moderately well but leaves a significant amount of unexplained variance (41%). This can be acceptable in certain contexts, especially for complex real-world problems and when we have limited data, as in our case.
- ❖ Acquire more data for better model performance
- ❖ Test more parameters

An abstract geometric design on the left side of the slide. It features a dark blue background with various geometric shapes and patterns. A white circle is positioned near the top left. Below it, a light blue semi-circle is visible. To the right of the semi-circle, there is a pink triangle with diagonal lines. Further down, there is a pink square with a pattern of concentric lines. At the bottom, there is a pink triangle with a pattern of concentric lines. The overall design is modern and minimalist.

THANK YOU

Ammanuel F. WoldeAregay