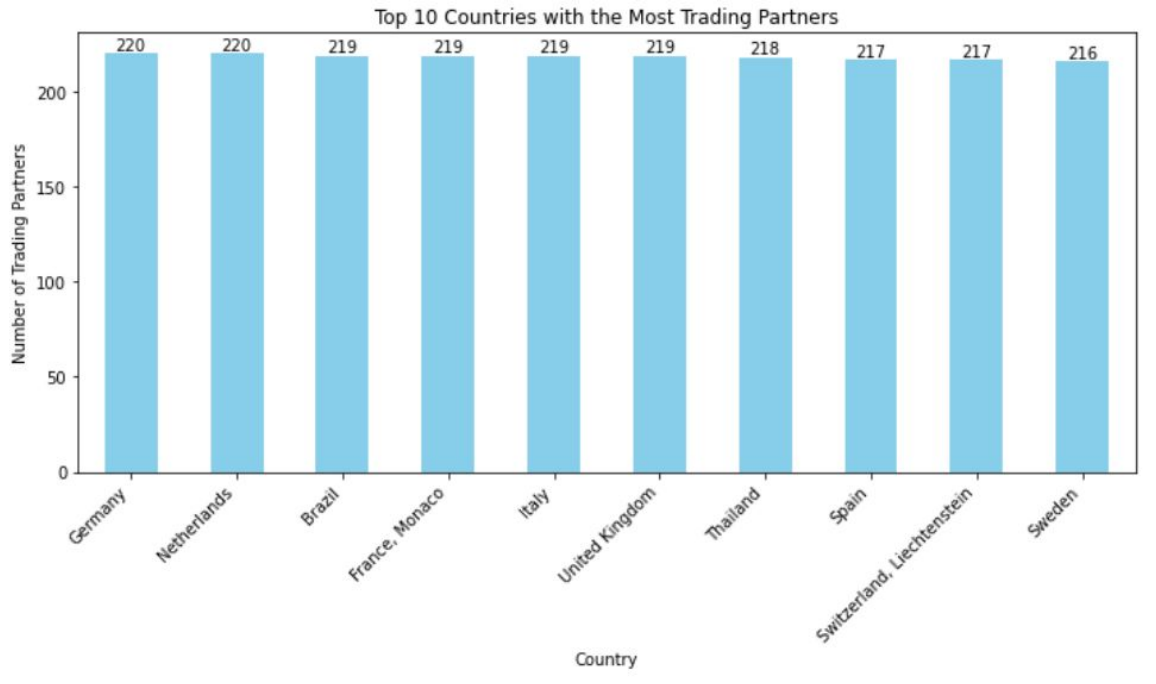
Mini Project II
Trade Flows between Countries

Descriptive Analysis 1 - Trading Partners



8 out of 10 are European countries:

- Geographical Proximity

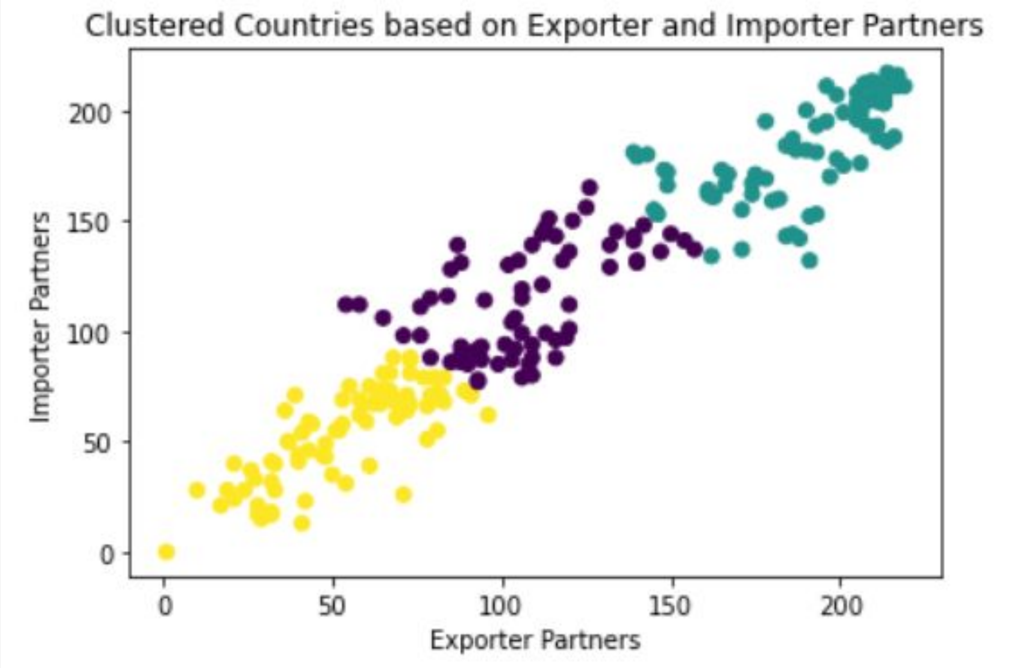
Brazil and Thailand are on the list:

- Regional Trading Hubs

China and USA are missing:

- Bilateral Trade Dominance

Trading Pattern 1 - Countries share similar Num of Trade Partners



Cluster 1

Exporter Partners=188.92

Importer Partners=182.68,

Countries with **high levels** of trade partners, with **extensive trade networks**

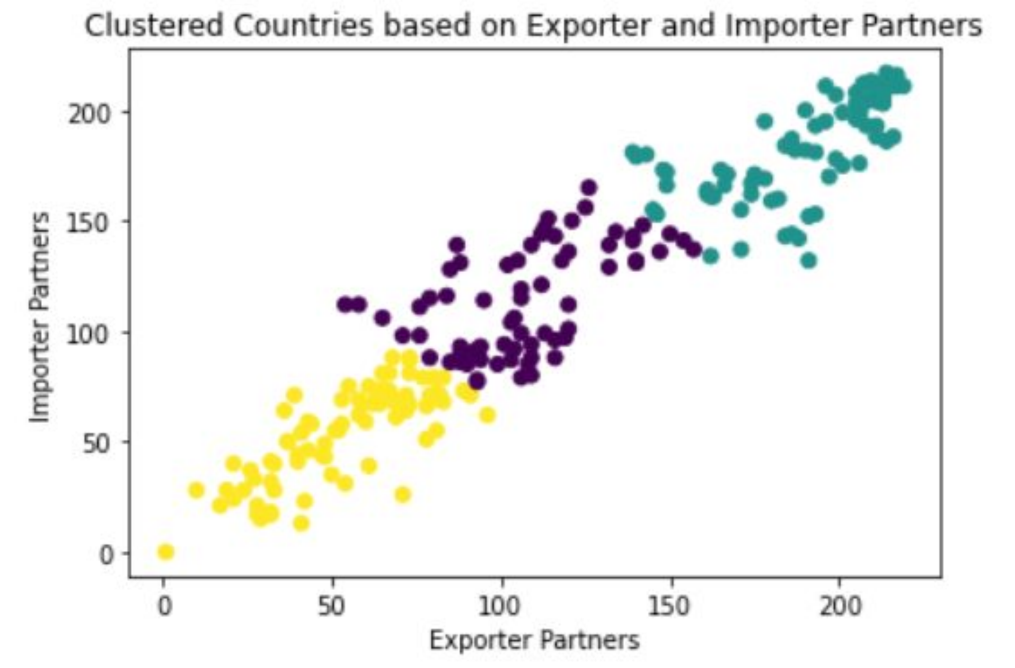
- **Moderate-large sized Developed country**:

Canada, Australia, USA

- **Large sized Developing country:**

India, China

Trading Pattern 1 - Countries share similar Num of Trade Partners



Cluster 2

Exporter Partners=107.03

Importer Partners=114.21

Countries with **moderate levels** of trade partners, with **balanced trade networks**

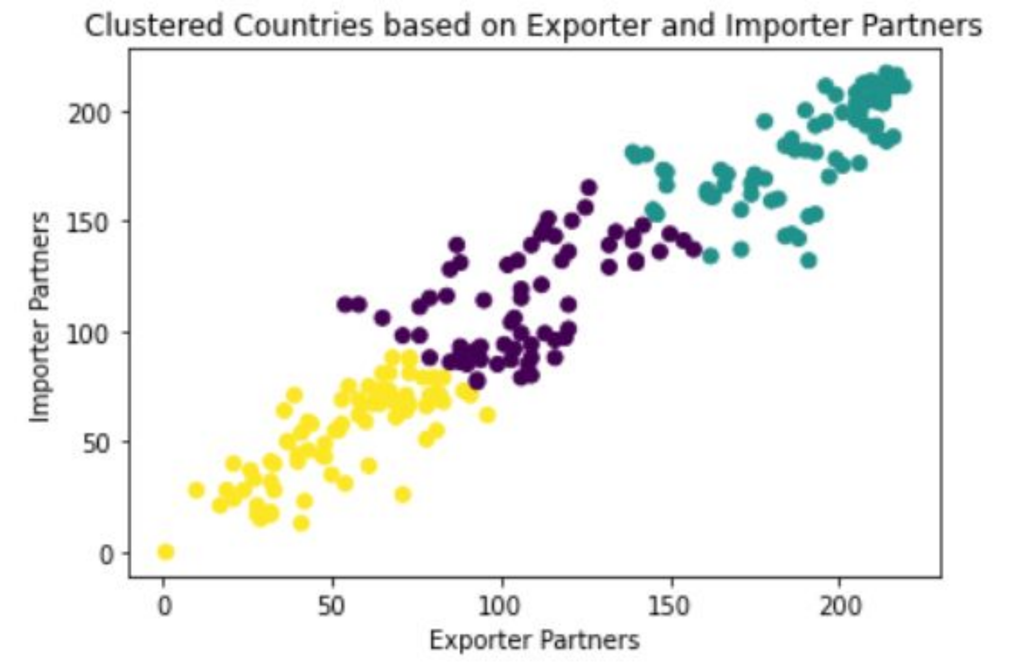
- **Moderate-sized Developing country**:

Cuba, Bangladesh, Vietnam

- **Small-sized Developing country**:

Iceland

Trading Pattern 1 - Countries share similar Num of Trade Partners



Cluster 3

Exporter Partners=53.69

Importer Partners=53.51

Countries with **low levels** of trade partners, with l**imited engagement** in global trade

- **Small-sized Developing country:**

African countries

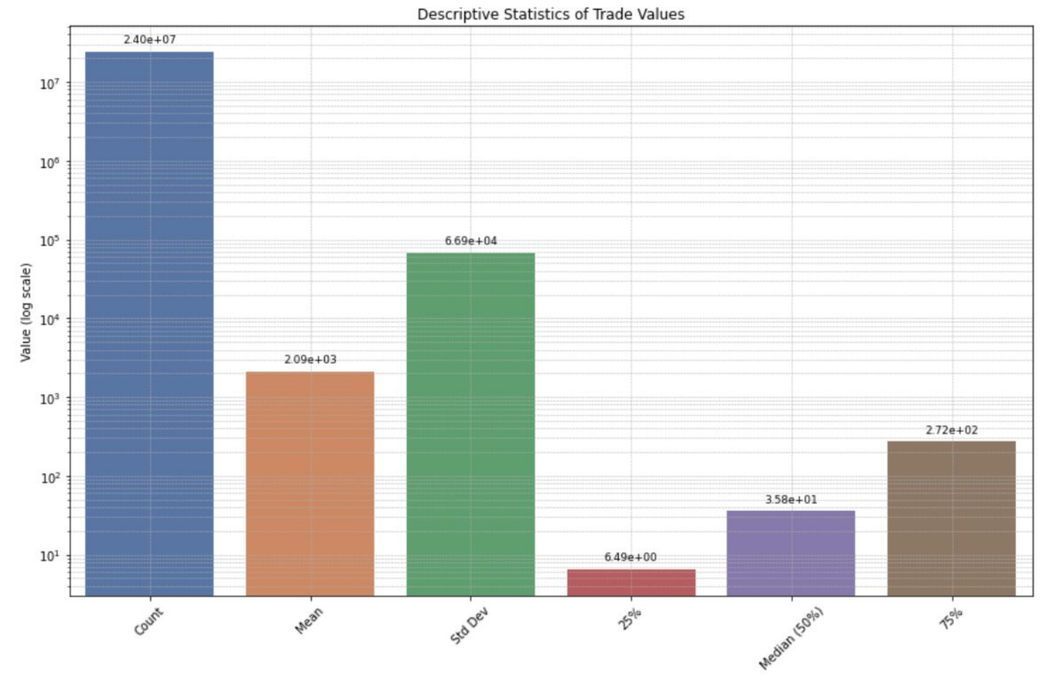
- **Regions do not count on trade:**

Macao Special Administrative Region

- Regions with **geopolitical concerns:**

North Korea

Descriptive Analysis 2 - Trading Volume(2016-2018)



Standard Deviation is very high:

Different countries have vastly different trade patterns and behaviors

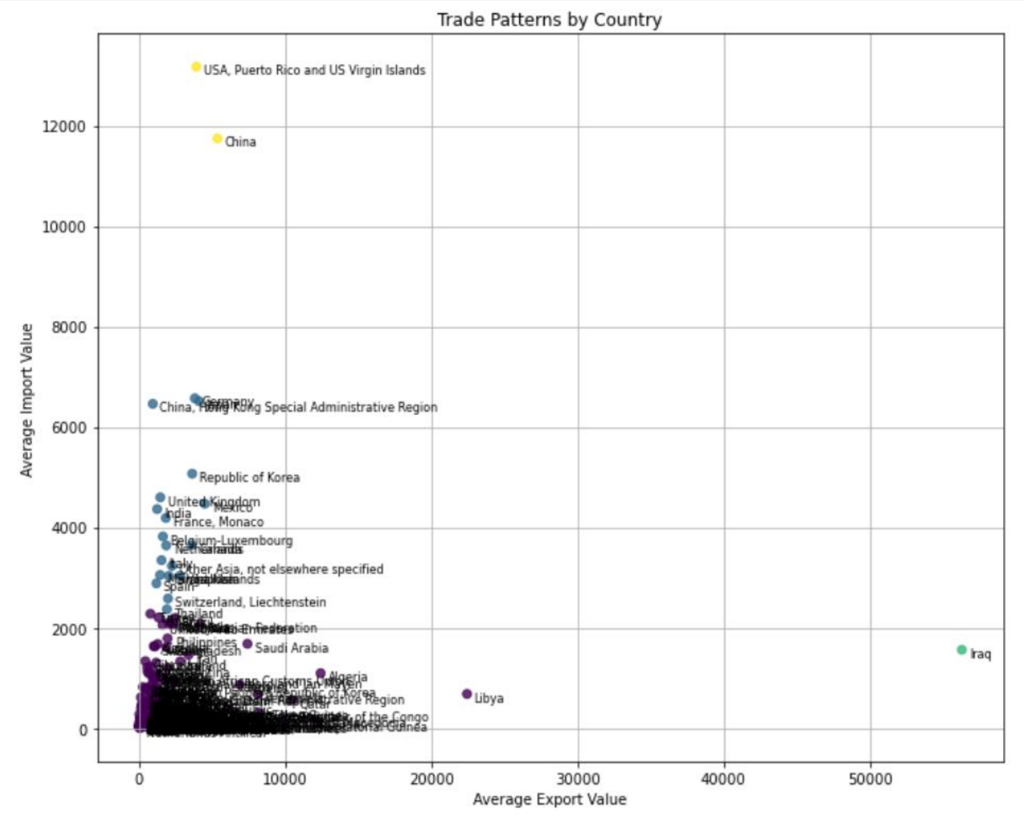
Median is much lower than the mean:

majority of trade values are not very large and the mean is influenced by some very large values

75th Percentile (272.413): 75% of the transactions have trade values below $272,413

>> further illustrates that most of the trade values are on the lower end, with only a few very large transactions

Trading Pattern 2 - Countries share similar Trade Volume



Cluster 1

Number of countries: 2

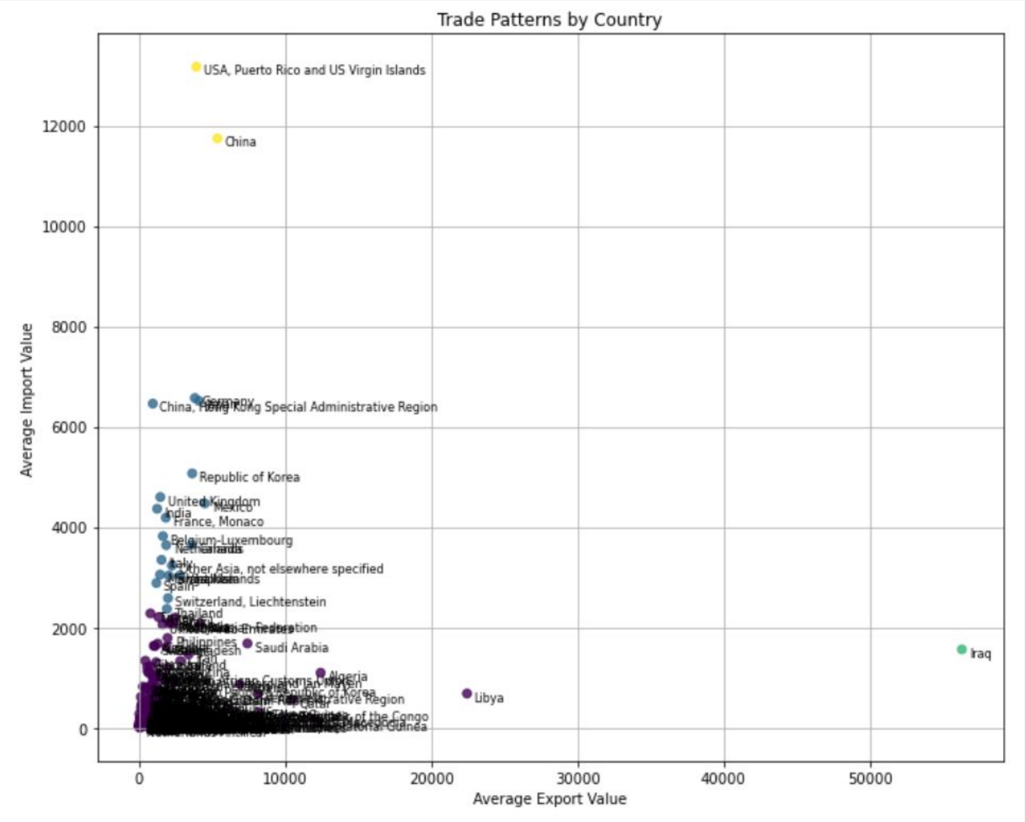
Avg Export value: 4664.229702

Avg Import value: 12459.145678

Feature: **High import and export value**

- **highly developed country:** USA
- **major global economy:** China

Trading Pattern 2 - Countries share similar Trade Volume



Cluster 2

Number of countries: 1

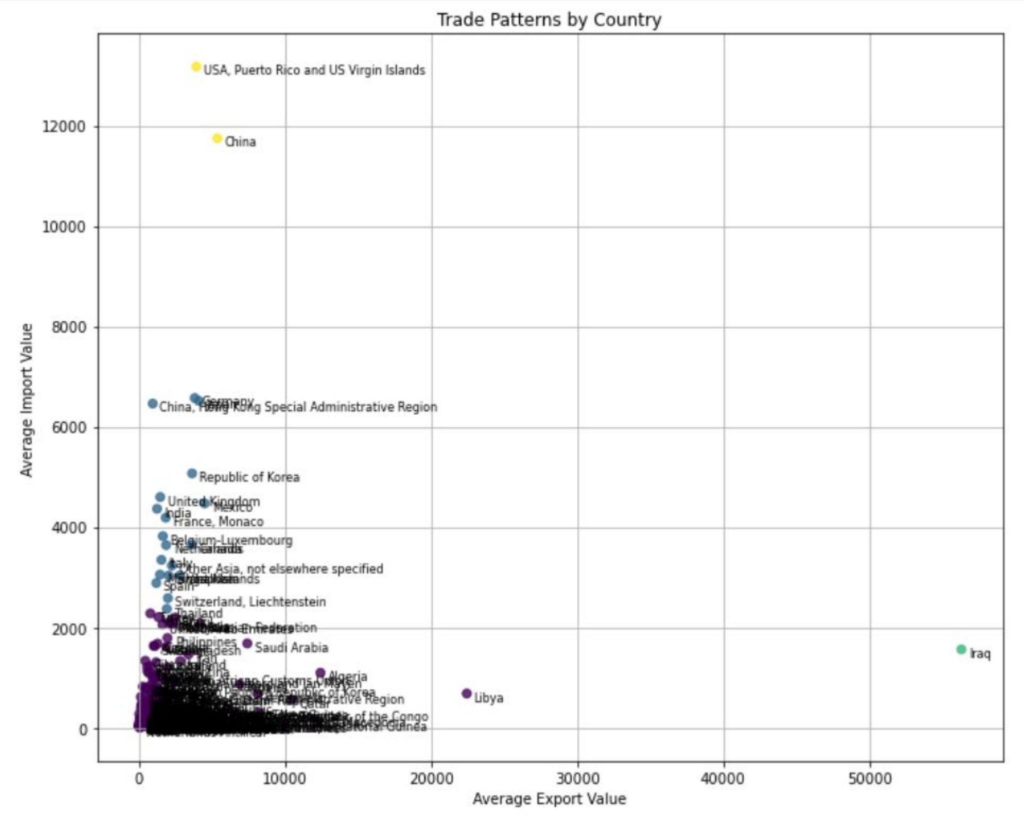
Avg Export value: 56294.536931

Avg Import value: 1563.499801

Feature: **Extremely high export**

- **highly developed country:** USA
- **major global economy:** China

Trading Pattern 2 - Countries share similar Trade Volume



Cluster 3

Number of countries: 19

Avg Export value: 56294.536931

Avg Import value: 1563.499801

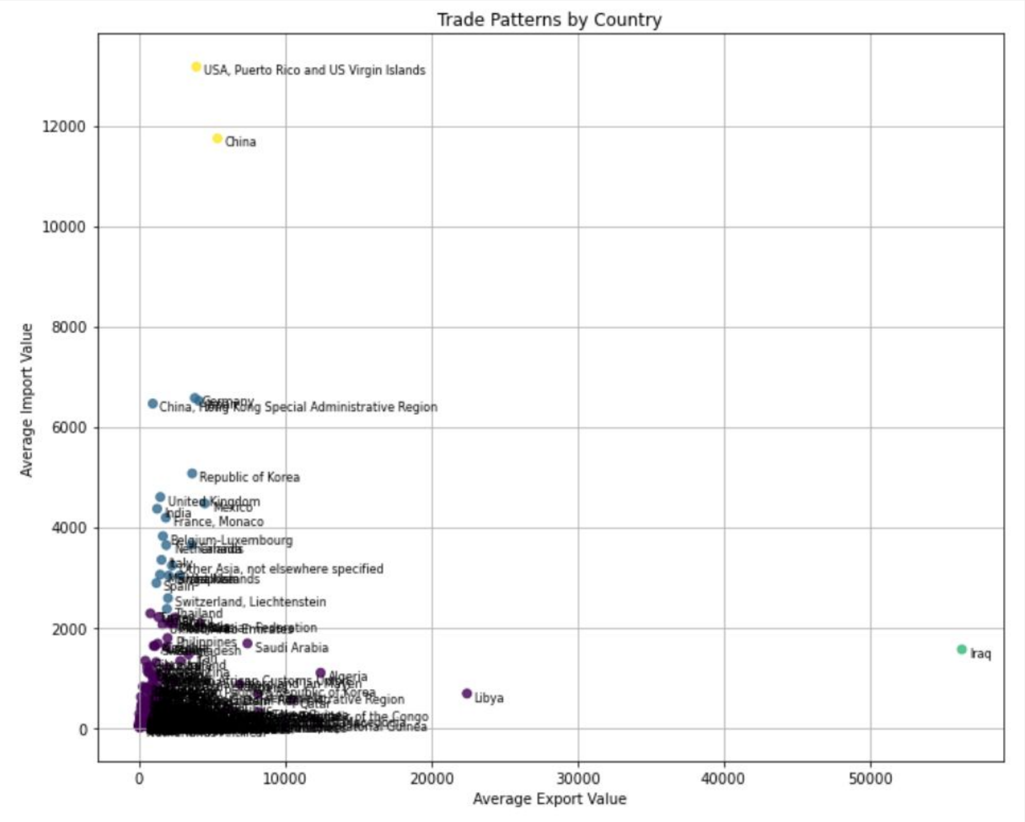
Feature: **moderate trade value**

- Emerging or mid-sized economies: India, Mexico

with Import larger than export

- Rely on import to meet domestic demands, which are not produced: Japan, South Korea
- Wealthy countries with high consumer demand: France, UK

Trading Pattern 2 - Countries share similar Trade Volume



Cluster 4

Number of countries: 200

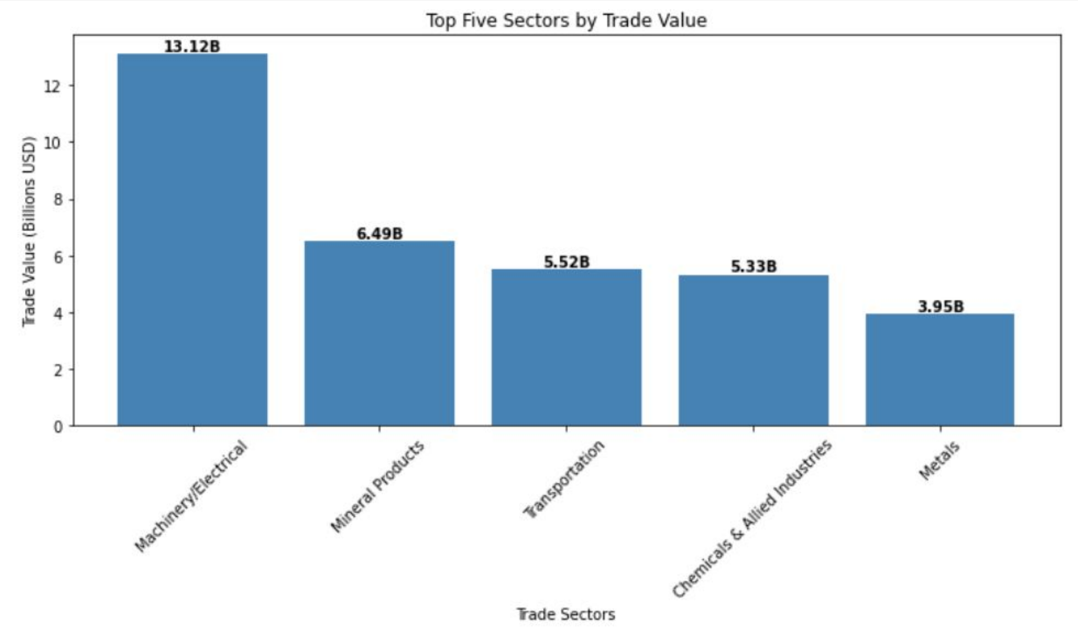
Avg Export value: 1548.9607

Avg Import value: 479.741741

Feature: **low trade value**

- **Smaller economies with less global trade activities**

Descriptive Analysis 3 - Trading Product



Based on The Harmonized System, classifying trade product into 14 sectors(simplified), based on the first two codes

Animal & Animal Products # 01-05

Vegetable Products # 06-15

Foodstuffs # 16-24

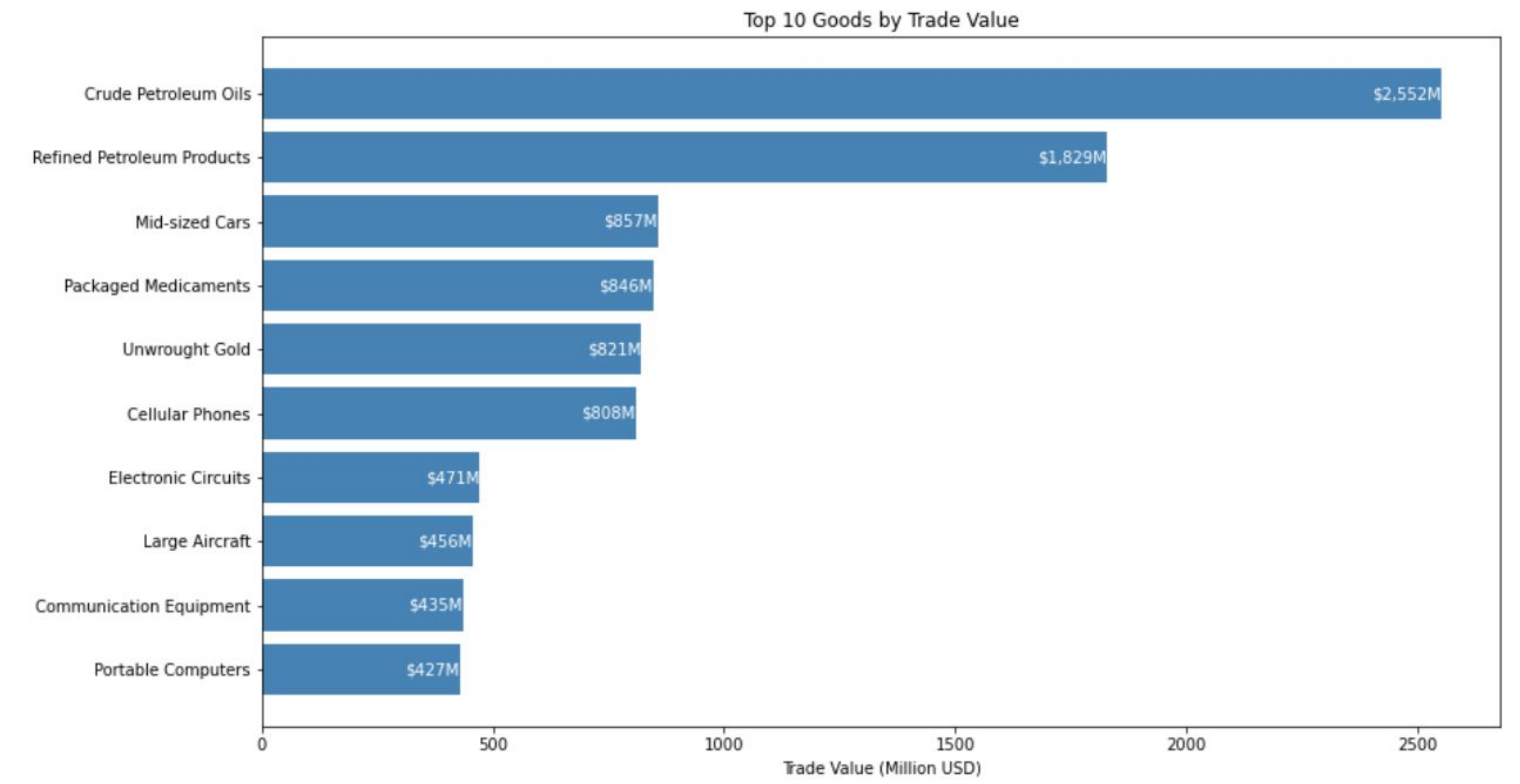
Mineral Products # 25-27

Chemicals & Allied Industries # 28-38

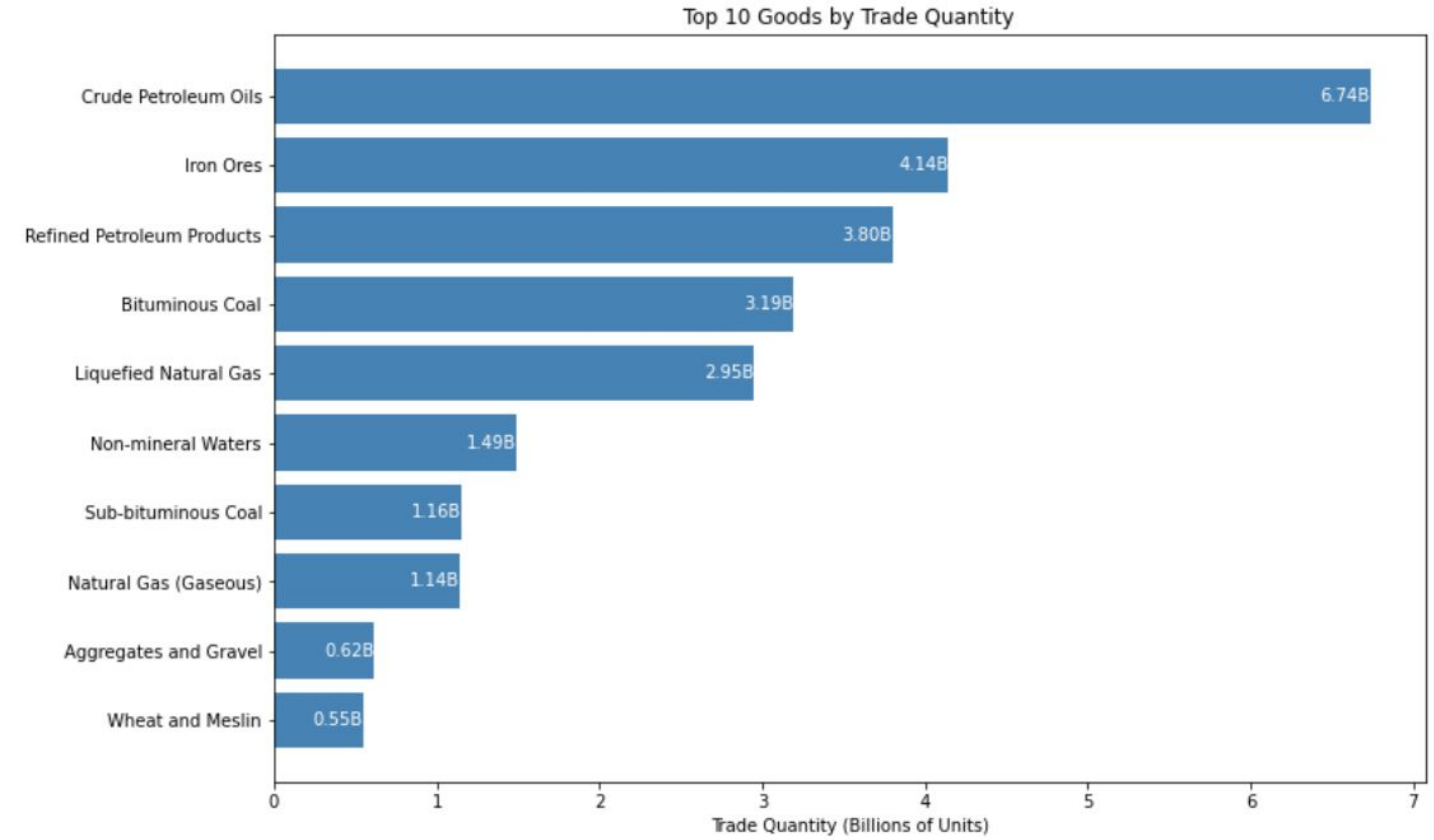
Plastics / Rubbers # 39-40

… …

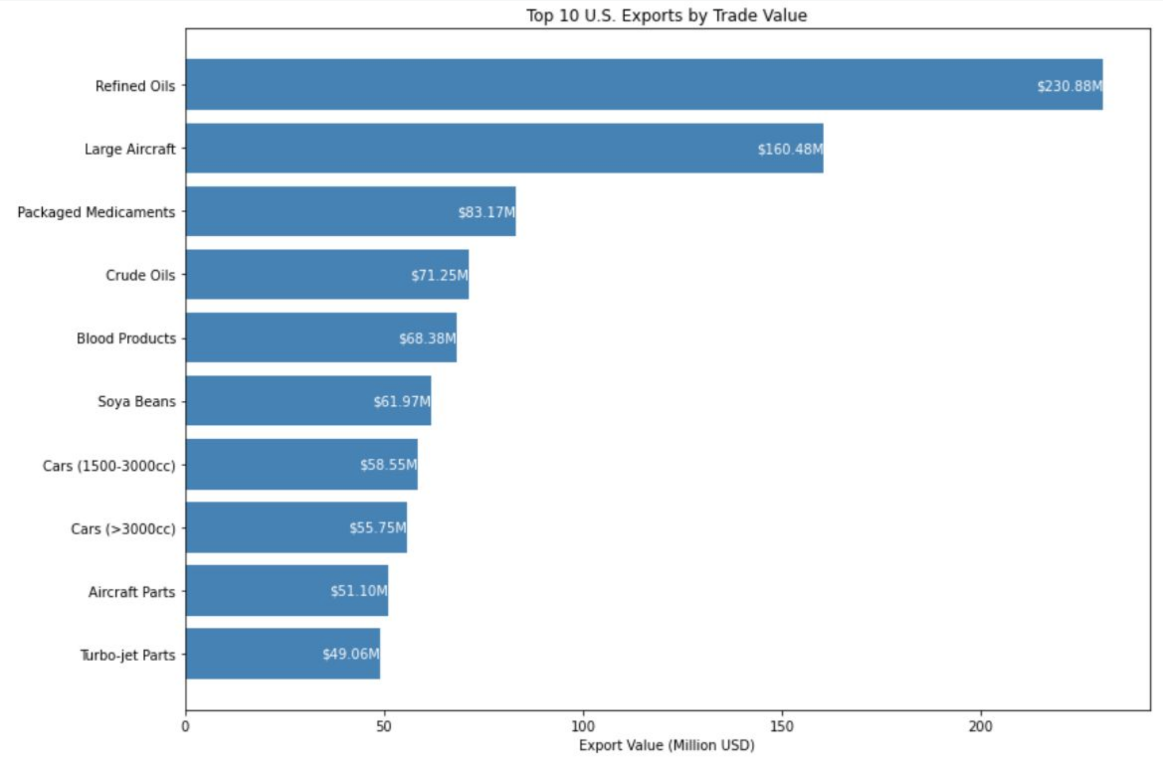
Descriptive Analysis 3 - Trading Product



Descriptive Analysis 3 - Trading Product



Descriptive Analysis 3 - Trading Product



Diversified Export Portfolio

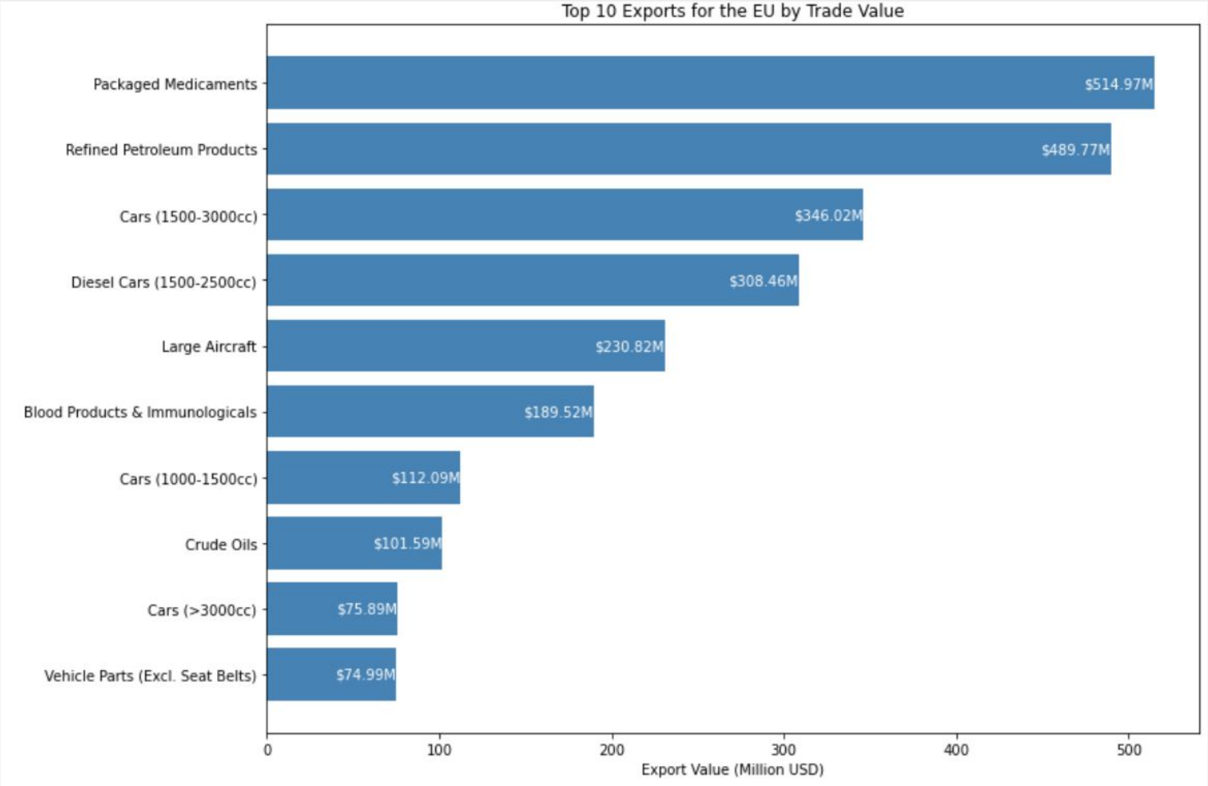
- Industry product
- Medicaments
- Agriculture product

Technological-intensive

- Engineering
- Healthcare

Heavy Industry

Descriptive Analysis 3 - Trading Product

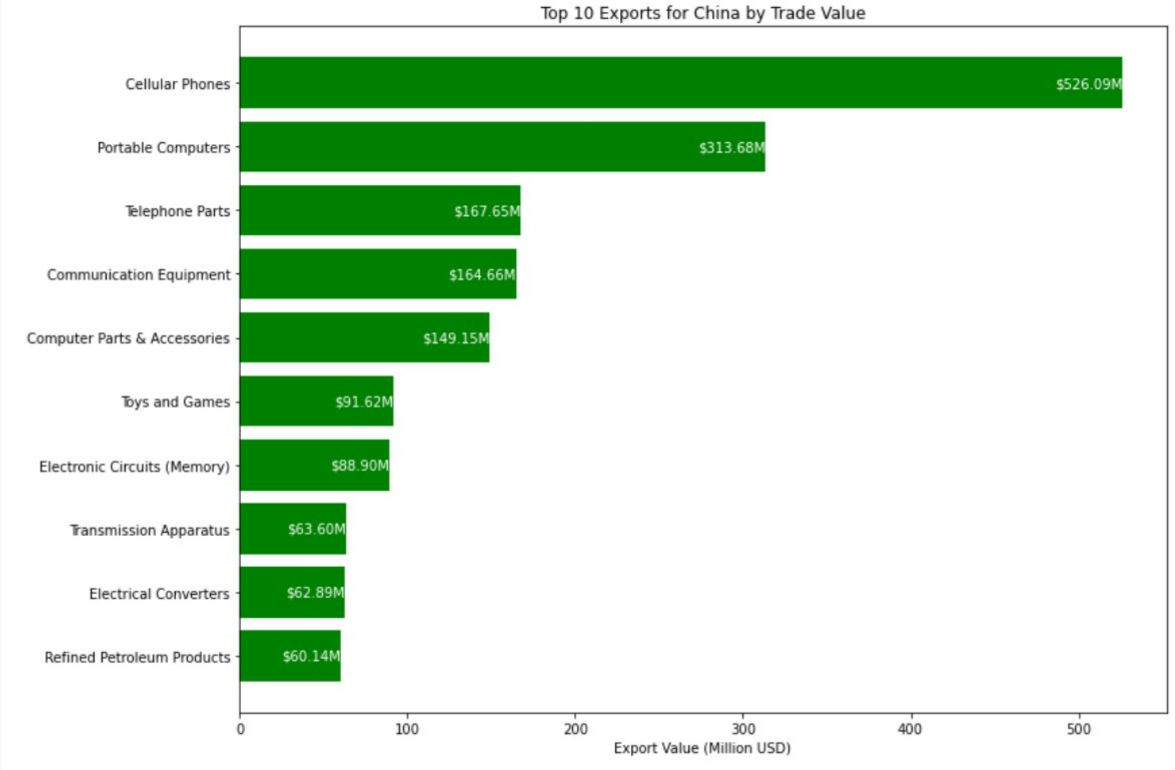


Specialization

Pharmaceutical Dominance: contribute to global healthcare

Robust automotive industry: known for manufacturing high-quality vehicles

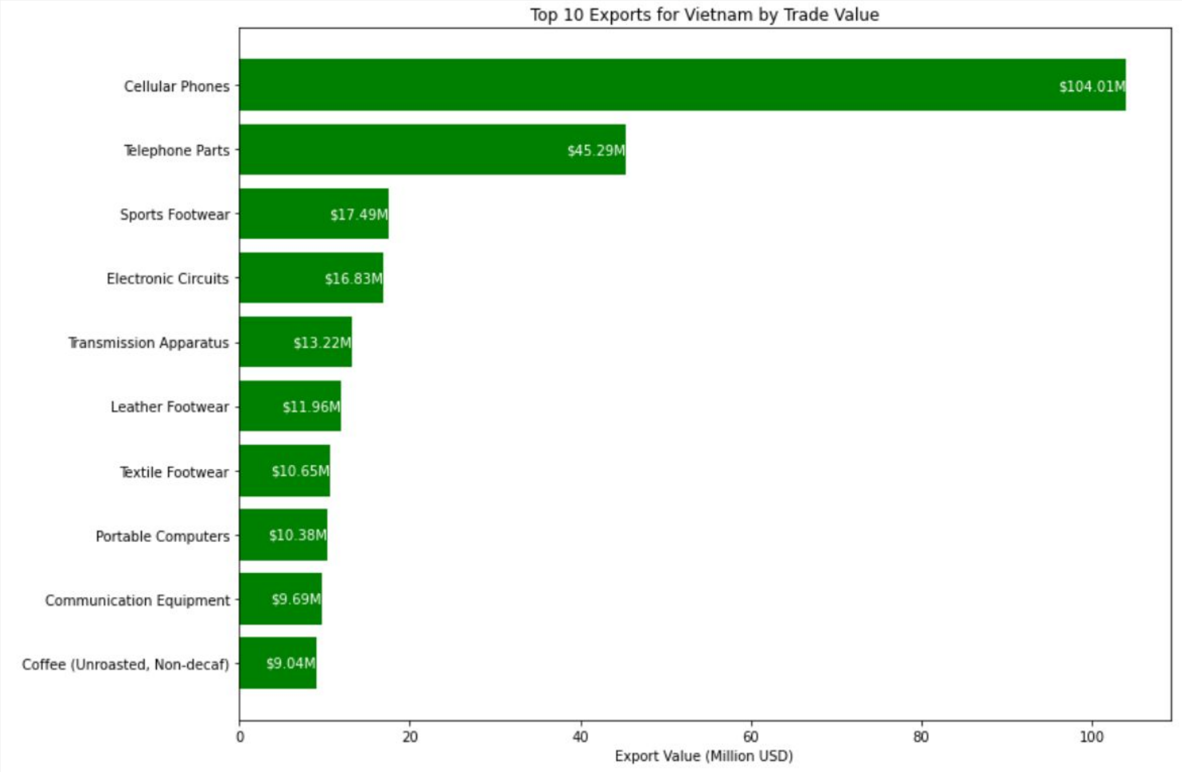
Descriptive Analysis 3 - Trading Product



Specialization in light industry

- Electronics Dominance
- Toys and Recreational Products

Descriptive Analysis 3 - Trading Product



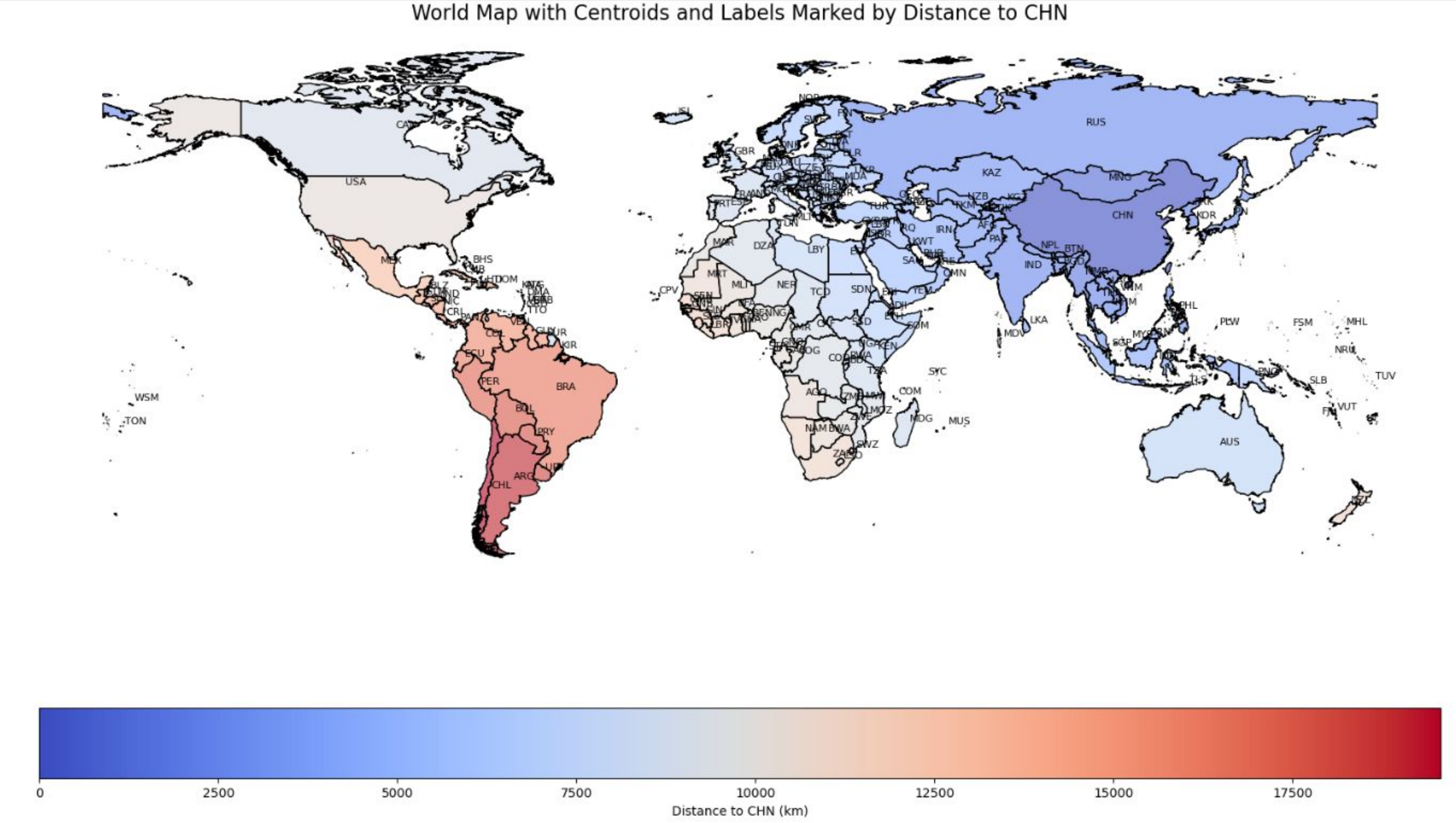
Specialization in light industry

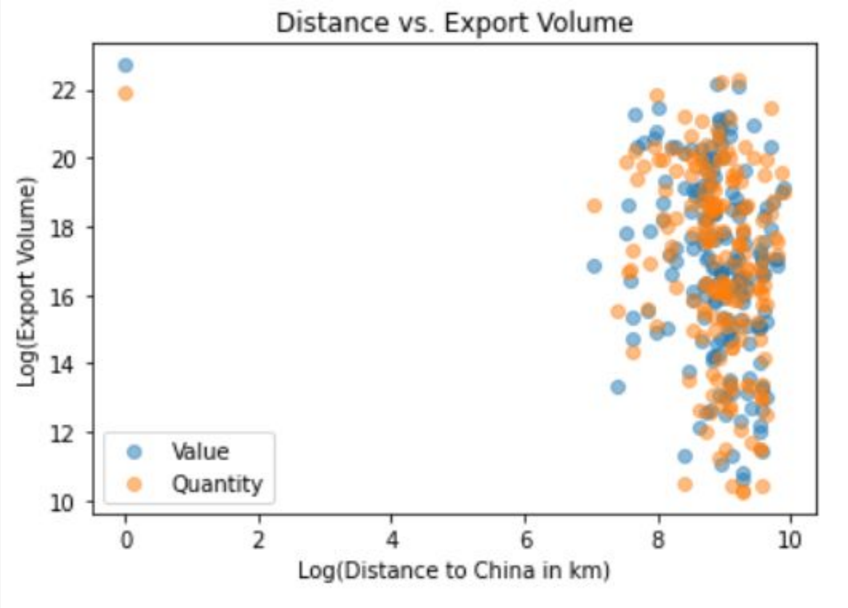
- Electronics Dominance
- Footwear

Major agricultural activity:

Coffee production

Descriptive Analysis 4 - Trading Distance





No negative correlation between distance and trade volume

>> distance might not be a significant factor for exporting goods to China

What makes distance less of a barrier:

- modern logistics
- trade agreements
- China's role as a major global trading partner
- China’s large market

Machine Learning - Predict Trading quantity based on GDP and Population

- Merge data: **GDP** and **Population** for exporter and importer country
  - EconMap is the database developed by the CEPII in 2010 to picture the world economy in the long term. It provides GDP at constant or current prices, as well as production factors and technical progress from 1980 to 2100 for 170 countries.
- Targeted variable: **Quantity**

Dask DataFrame Structure:

	t	i	j	k	q	gdp_exporter	population_exporter	gdp_importer	population_importer
npartitions=3									
	int32	int32	int32	int32	float32	float32	float32	float32	float32
	...	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...

- **Split**: 80% for training; 20% for testing
- **Scaled**: use the mean and standard deviation computed from the training data to standardize the testing data to improve convergence

Machine Learning - Predict Trading quantity based on GDP and Population

```
# Predictions
y_pred_rf = grid_rf.predict(X_test_scaled.compute())
y_pred_gb = grid_gb.predict(X_test_scaled.compute())
y_pred_nn = nn.predict(X_test_scaled.compute())

# Evaluation
mse_rf = mean_squared_error(y_test.compute(), y_pred_rf)
mae_rf = mean_absolute_error(y_test.compute(), y_pred_rf)
r2_rf = r2_score(y_test.compute(), y_pred_rf)

mse_gb = mean_squared_error(y_test.compute(), y_pred_gb)
mae_gb = mean_absolute_error(y_test.compute(), y_pred_gb)
r2_gb = r2_score(y_test.compute(), y_pred_gb)

mse_nn = mean_squared_error(y_test.compute(), y_pred_nn)
mae_nn = mean_absolute_error(y_test.compute(), y_pred_nn)
r2_nn = r2_score(y_test.compute(), y_pred_nn)

print("Random Forest - MSE:", mse_rf, "MAE:", mae_rf, "R^2:", r2_rf)
print("Gradient Boosting - MSE:", mse_gb, "MAE:", mae_gb, "R^2:", r2_gb)
print("Neural Network - MSE:", mse_nn, "MAE:", mae_nn, "R^2:", r2_nn)
```

```
Random Forest - MSE: 248434160117.4171 MAE: 3726.5339390363015 R^2: -0.7616149848191338
Gradient Boosting - MSE: 154185221037.16415 MAE: 6320.245772361284 R^2: -0.09330776286298836
Neural Network - MSE: 140923606551.18076 MAE: 5813.59412791831 R^2: 0.000728656244437853
```

Machine Learning - Predict Trading quantity based on GDP and Population

Performance metrics for models are not ideal

	MSE	MAE	R-squared
Random Forest	248434160117.41	3726.53	-0.76
Gradient Boosting:	154185221037.16	6320.24	-0.0936
Neural Network	140923606551.18	5813.59	0.0007

1. **Mean Squared Error (MSE):** measures the average squared difference between the predicted values and the actual values; Higher values indicate higher prediction errors
2. **Mean Absolute Error(MAE):** measures the average absolute difference between the predicted values and the actual values
3. **R-squared:** measures the proportion of the variance in the dependent variable predictable from the independent variables; Negative values indicate that the model fits the data worse than a horizontal line, which suggests that the model is not performing well.

“Thanks for Listening”