Analysis of Maritime Costs, Delays, and Tariff Impacts: A Quantitative Analysis Using Bills of Lading Data

Blair Jones

Culminating Experience

Master of Applied Mathematics

University of Colorado, Boulder

December 9, 2025

Contents

1 Executive Summary

International trade powers the global economy and significantly impacts GDP growth and individual purchasing power. In the modern economy, the political and social implications of international trade are vast and frequently changing.

United States importers are challenged with predicting costs and delays of their goods, which affects inventories and trade cycles. In addition, importers face uncertainties as tariff announcements and impacts have shifted significantly in 2025. These costs materially impact procurement budgets and business decisions.

At the same time, US Customs Enforcement is tasked with identifying potential tariff evasion and maintaining a functioning trade system while handling millions of annual international shipments.

This project intends to show that machine learning and statistical analysis of international shipments could supplement vast domain knowledge in the field.

With over five years of available shipment data for footwear, modeling based on landed cost and delay days was performed using different machine learning techniques. In addition, statistical analyses based on historical and geographical trends inform the impacts of tariff announcements in 2025.

The primary findings are as follows:

- Costs (available for CIF shipments) are predictable with product type, supplier and volume information. An attention model neural network method produced a model with an R^2 of 87.5%, RMSE of $5.78, and average percentage error of only 5% on costs per kg.
- Shipment delay days are mildly predictable, with R^2 of 66% and RMSE of 3.3 days, using a random forest model and noting important features, such as carrier, route frequency and port.
- Statistical analyses of changes in shipment by departure country from pre and post tariff announcement periods were significant and indicate changes in business behavior post 2025 tariff announcements. Weight (kg) delivered from China from April through September 2025 decreased 31.8% compared with prior like-periods.

2 Technical Summary

Data was obtained from a paid service called ImportYeti.com (ImportYeti, 2025), which uses a Freedom of Information Act order to obtain sea shipment bill of lading information retained by US Customs Enforcement. Various search fields are available, like product description, supplier name, and carrier, to name a few. Data is obtained in tabular form (.csv format). This was then imported into Python for cleaning, exploration, analysis and modeling.

First, a regular expression extracted two words of context around footwear keywords in an attempt to clean the description field. For product categorization, the cleaned descriptions were embedded with `SentenceTransformer` (all-MiniLM-L6-v2 model) and an *OpenAI* API call to assist with embeddings was used to prepare a product type field, matching by category using cosine similarity. Plotting with *matplotlib* by type and time, along with analyzing large residuals from a simple linear regression found semantic mismatches (e.g., brake shoes rather than footwear), which were then removed from the analysis.

Once data was cleaned, Random Forest and XGBoost were used for prediction modeling, after which Pytorch neural net architectures (including MLP, ResNet, and Attention models) were built to compare. Neural networks used learned 8-dimensional embeddings for categorical variables (routes, companies, ports, suppliers), enabling the models to capture feature similarity that one-hot or label encoding cannot represent.

Finally, a tariff analysis on shipments was prepared using various statistical tests, including negative binomial counts (used over Poisson for over-dispersed data), the chi-squared test, Mann-Whitney U tests, and bootstrapped confidence intervals on mean difference.

3 Background

International imports of footwear into the United States represent a $28Bn industry as of 2024 (U.S. Census Bureau, 2024) .

This project intended to use the available international shipment information for select search terms to determine if various statistical and machine learning techniques can tell a story about the data.

A Bill of Lading is a contract of carriage and legal document which details exporter and importer information and responsibility. Fields from this document are parsed and stored in tabular form.

The following questions were of most interest throughout the project:

1. For shipments with CIF (Cost, Insurance, Freight) terms, can landed costs be predicted using a number of quantitative and qualitative features, including supplier name, shipment volume, product type etc.?
2. Can delays be predicted using temporal features and carrier information?
3. Due to recent tariff actions, is there a way to identify changes in shipment activity by weight / volume and departure location?

4. Are there certain techniques that model the data better than others, and which features of the data are the most important in those models?

Some important dates regarding tariff announcements in 2025 are as follows. Note that due to the high concentration of footwear imports from China, this specific timeline is significant (China Briefing, Arendse Huld, 2025).

U.S. Tariff Actions on Chinese Imports

Tariff Action	Announced Rate	Current Rate	Announced Date	Changed Date
Section 301 - List 4A	15%	7.5%	9/1/19	2/14/20
"Liberation Day"	34.0%	10.0%	4/2/25	5/14/25
"Fentanyl"	20.0%	10.0%	3/4/25	11/10/25

Section 301 tariffs existed prior to 2025 on a number of goods imported from China; footwear was on List 4A and is currently assessed a 7.5% rate (Trade Partnership Worldwide, 2023). Two updates to the *Liberation Day* tariffs occurred between April 8th and 9th, with reciprocal tariffs announced up to 125% before decreasing to the current rate of 10%, with several extensions of this rate continuing through today (China Briefing, Arendse Huld, 2025). Using April 2nd as the cut-off point for tariffs analysis is not entirely precise, given the additional *Fentanyl* tariffs of 20% were already in place on this day. However, this was the starting point for this analysis, with future enhancements based on other dates in consideration. It also should be noted that there are various lead times for both production and then shipment of goods from overseas vendors, so an overnight change is not realistic. However, general shifts would be likely given the US Administration's known intention to consider tariffs as part of its trade policy and the fact that importers have to use their domain knowledge to make decisions in real-time for potential future impacts.

In mid-2021, a supply chain crisis unfolded internationally. Closures in overseas factories, labor shortages at US ports and an increase in consumer demand for durable goods contributed to significant delays at ports. Delays in average days from overseas factory to US warehouse spiked to 76 in October 2021 from 48 in January 2020 (Siripurapu, 2021).

4 Data Description

As mentioned above, the raw data is obtained in .csv tabular form and requires cleaning due to semantic mismatches in search terms and messy product description fields. The raw data contains 56 fields, including many needed for specific search like BOL, container, and vessel numbers. Of interest to this project were fields like supplier (vendor), company (customer), and carrier names, departure and destination ports, weight (kg) and volume (in TEU – twenty-foot equivalent unit container size) and arrival date.

Sixty-nine months of data was available, January 2020 through September 2025.

Quantity / Units

There are fields for item quantity along with a quantity unit, such as CTN (carton), PCS (piece), and PKG (package). Due to inconsistencies with size of these units, weight and volume provided in standard units were preferred as features. One example of this inconsistency is the CTN unit which has a mean of 10kg but a max of 6,000+kg. One way to adapt for these inconsistencies is to derive features like quantity per TEU or weight per unit, normalizing for shipment size.

CIF

Cost, insurance and freight (CIF) is a standard incoterm where the seller agrees to pay for the freight and insurance of the goods while on the water. About 11-12% of the shipments contained these terms, meaning a value in the column Value in USD (CIF) that represented the cost of that shipment before import. For all others, "Different Incoterm Used" was noted. Some analyses used weight or shipment counts, so all cleaned data was available; others used CIF as the response variable, then with fewer observations available. Due to the low percentage of observations with CIF available, an attempt to fill in missing data was not feasible; rather, two different data sets were considered. Note that for items with CIF values, a cost per kg field was added to account for shipment size.

Feature Engineering

Simple additional features were added directly from the original data including extracting arrival month, year and day from the Arrival Date field as well as using that field to calculate delay days as Estimated Arrival Date was also available. To consider seasonality without the artificial jump from December (12) to January (1), cyclical encoding was applied to months:

$$\text{month\_sin} = \sin\left(\frac{2\pi \cdot \text{months}}{12}\right)$$
$$\text{month\_cos} = \cos\left(\frac{2\pi \cdot \text{months}}{12}\right)$$

Other fields were added using additional data containing port latitude and longitude, which then can be used to calculate shipping distance (National Geospatial-Intelligence Agency, 2024). A module called `searoutes` was imported and used to calculate true shipping distance, as a Haversine distance using latitude and longitude only captures "great circle" distance and is not realistic when considering shipments from Asia to the east coast of the United States.

As mentioned above, an *OpenAI* API call was used to attempt to classify certain items based on messy product descriptions. For the shoes data, these were descriptions like athletic shoes, dress shoes, boots, etc., to be used as an additional product type feature.

Supplier and company (customer) names were cleaned and combined using fuzzy matching and clustering via similarity graph and Top Customer and Top Supplier boolean fields were added based on the top 20. Similarly, route (departure to destination) and port frequency were included as additional shipment features.

For routes, in addition to looking at frequency, a K-Means clustering (K=15) was performed on latitude and longitude coordinates to consolidate geographically close ports, for example, the Ports of Los Angeles and Long Beach. Then, a `route_id` feature was created as a categorical variable.

Seasonality Using weights and looking at arrivals per week, a heatmap representing changes from overall average was created, to highlight weeks of the year that have higher or lower arrival weights. This is in line with expected slow-downs in China over the Chinese New Year holiday (January or February yearly), which show up in arrival data in the 12th-14th weeks of the year.

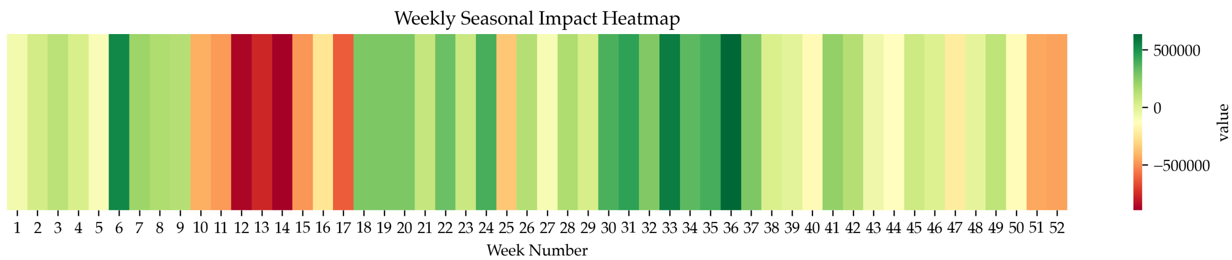

Figure 1: Weekly Shipments (kg) compared with Mean

Analysis by Country

For tariff-related analysis, shipment counts and volume by country (using the `departure_port` field) were summarized. China topped the counts list, representing 61% of all shipments for the full data period reviewed. Because of this, particular focus was given to Chinese-based suppliers when analyzing shipment location post tariff announcements.

Shipment Counts by Country

Country	**Count**	**Percent (%)**
China	98,403	61.04
Vietnam	28,194	17.49
Singapore	8,555	5.31
South Korea	3,712	2.30
Italy	2,309	1.43
Hong Kong	2,121	1.32
Malaysia	1,842	1.14
Taiwan	1,619	1.00
Germany	1,592	0.99
Belgium	1,477	0.92
Indonesia	1,268	0.79

5 Methods

5.1 Data Cleaning

Data cleaning is a process that is performed throughout the project. The original shoe data was over 192,000 shipments long. A cursory review of the data revealed some issues with product search, for example, “brake shoes” included, which then can be removed initially. However, due to the large data size, many of these mismatched product descriptions were not able to be identified until an initial regression was performed using *weight* as a response, in order to find potential issues with the data. After running a simple linear regression, top residuals were printed and matched to the original data fields where a number of other ambiguous search terms were included, such as shoe “shelving / furniture” and other industrial wording like “aluminum”, “iron” and “wheel.” The process of removing these unintended items needed to be completed a number of times before the data was ready for analysis. The final dataset, representing footwear shipments from January 2020 to September 2025, was 161,214 lines.

The data from the source does not contain significant missing values. As noted above, CIF (cost) is only available on certain shipments with those specific incoterms; therefore, the data was split into two differently sized data sets depending on the task. The cleaned CIF data was 14,250 lines.

5.2 Prediction Models for Costs

All footwear shipments were filtered by line items that contained a value in the CIF field and separated into a new dataset which was used for cost prediction analysis. After filtering for significant outliers (over 99.5% percentile on the higher-end), the dataset was 14,250 shipments long. Because all shipments are different sizes, a cost per kg field was created as the response variable, as weight in kg is a consistent field across all line items.

An initial Ordinary Least Squares model was run as a baseline to compare to other models. Features were split by numerical, boolean and categorical, with the categorical variables then being one-hot-encoded. Numerical features include volume, distance, and cyclical-month coding, among others. Boolean features include top company / supplier and additional product information, i.e. if the product description contained the word ”garment” or ”handbag”, indicating a mixed shipment. Categorical variables included product type (e.g. athletic shoes, dress shoes), and quantity unit of measure.

For Random Forest and XGBoost, the previous features were included along with some additional categorical columns, like company name, route ID (given based on specific departure to arrival port), port names and supplier country. These were added after first label encoding into integer values and then included as a numerical feature. Because of the nature of these models, categorical features encoded as numbers do not affect performance as the models do not look at magnitude necessarily, but rather use an order-based split to rank predictions. Doing this instead of one-hot encoding prevents an explosion of the feature space when the number of unique categories is significant (e.g. company names). Because of the additional features, the comparison from OLS to tree-based models is not exactly direct; however, the flexibility of the tree-based models is simply noted and expected to improve performance as training is built out for this data.

Note that once neural networks were trained, these label encoded categorical features were removed as simple numerical features, since there is no sense of actual numerical value. Rather, they were converted to embeddings, which would take either a previously sparse one-hot-encoded vector or non-meaningful label encoded vector and convert those features to dense representations which can relate within; meaning, features with similarities should end up close in the embedding space and provide meaningful relationships when learned through training.

Hyperparameters were tuned by random grid search on both Random Forest and XGBoost models. The hyperparameters considered were number of trees, maximum depth, minimum split and minimum leaf size for the random forest models, with additional tuning for regularization and learning rate for the XGBoost models.

Several *PyTorch* neural networks were built including a simple feed-forward neural network (MLP), a residual block network (ResNet) and an Attention model. The MLP model is a standard fully connected layer network with batch normalization, ReLU activation and dropout regularization. The ResNet model extends the MLP with skip connections that add each block's input to its output, allowing the network to learn from residuals. The attention model adds a feature attention mechanism that learns to weight features based on relevance. This allows the model to emphasize different features dynamically.

The loss function for all prediction models was root mean squared error:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The coefficient of determination R^2 was also used to compare model performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

5.3 Prediction Models for Delays

All shipment data regardless of the inclusion of the CIF field were considered for delay analysis. When reviewing delays over time, significant increases were found in mid-2021 through mid-2022. These were related to the Supply Chain crisis post Covid lockdowns, in which some Chinese factories faced closures, U.S. ports and trucking faced labor shortages and consumer demand was surging.

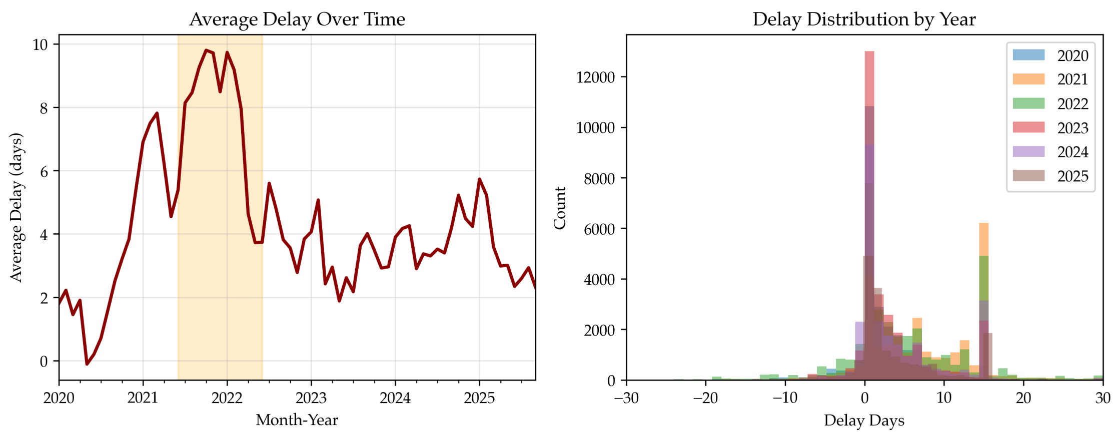

Figure 2: Delays Over Time

Because of this clear trend, delay prediction was performed post June-2022, to allow for the model to train on shipments that arrived in a more normalized time-frame.

A baseline random forest and tuned RF and XGBoost models were trained on the data. A random grid search for hyperparameter tuning, similar to cost predictions, was also performed.

Features were similar to costs for the tree-based models in that some categorical features were label encoded to avoid a significant increase to the column space. Because the delays analysis uses all shipment data, the number of potential unique values is well above the already large number seen in the CIF data set. Therefore, this analysis stuck with tree-based models and some label-encoded categorical variables.

5.4 Tariff Analysis

5.4.1 Data Preparation

A large part of the project was attempting to identify potential business changes after 2025 tariff announcements. As per the chart in Data Description, the majority of footwear imports arrive from China (61% of total shipments within the period reviewed, January 2020 through September 2025).

Although there have been a number of announcements and updates to various tariff rates and impacted countries, a significant announcement was made on April 2nd, 2025 (*Liberation Day*), including all trade partners; therefore, the attempt to quantify tariff impacts into a "pre" and "post" period used this day as delineation.

Because of changes in seasonality, the *pre* periods include April through September only, so as to not confuse differences in volumes to ongoing temporal factors. Therefore, the data in this section is restricted to April 2022 forward.

5.4.2 Shipment Counts (Negative Binomial)

It was noted that in terms of shipment counts, the data is overdispersed, meaning the variance exceeds the mean. Because of this, a traditional Poisson test for count data was not appropriate. The negative binomial distribution, $\mathrm{NB}(r, p)$, models count data with overdispersion because an additional dispersion parameter r accounts for excess variance.

$$P(Y = k) = \binom{k + r - 1}{k} p^r (1 - p)^k$$

with mean $\mu = r(1 - p)/p$ and variance $\sigma^2 = \mu + \mu^2/r$

By factoring in larger variances, the changes over time periods should be less likely to be overstated in significance as there is an expectation of changes anyway (high variance).

5.4.3 Weight Analysis (Mann-Whitney U Test)

Changes in weight per shipment were tested using the Mann-Whitney U test, which is a non-parametric statistical test that asseses whether groups originate from the same distribution. Due to the skewed nature of the data, this was selected over other standard tests.

The U statistic is computed as:

$$U = R - \frac{n(n + 1)}{2}$$

Where R is the sum of ranks for one group.

The rank-biserial correlation (r) was computed as an effect size measure, where $|r| > 0.1$ indicates a small effect, $|r| > 0.3$ a medium effect, and $|r| > 0.5$ a large effect.

5.4.4 Route Shifting (Chi-Squared)

The original data contains both a departure country (port) field and supplier country field. These were compared and highlighted if a non-match. A contingency table was prepared and a Chi-squared test was performed. Note that suppliers can be multinational with factories in many countries and the available supplier address may be located in a place for specific contact or payment reasons.

5.4.5 Chinese Supplier Analysis (Mann-Whitney / Bootstrap C.I.)

Due to the concentration of Chinese imports, a more specific Chinese supplier analysis was performed, which looked at imports specifically from suppliers which have a Chinese country code. Two groups were formed; those shipments which were shipped from a Chinese port, and those that were shipped from other countries' ports.

Daily shipment counts and weights for each routing category were compared between periods using two methods:

1. **Mann-Whitney U test:** Non-parametric comparison of daily rate distributions
2. **Bootstrap confidence interval:** 1,000 bootstrap samples were drawn to form 95% confidence intervals for the difference in mean daily rates; significance was concluded when the interval excluded zero

6 Results

6.1 Prediction Models for Costs

Model	**RMSE ($)**	**R^2**
OLS	11.7511	48.14%
Random Forest	6.3425	84.89%
XGBoost	5.9535	86.69%
MLP	5.9194	86.84 %
ResNet	6.1717	85.70%
Attention	**5.7805**	**87.45%**
Mean	24.27	–
Std Dev	16.47	–

As expected, the tree-based models performed well on the data, which include quantitative, categorical and boolean features. The flexibility of these models, along with randomness of feature selection and regularization, allow for a large variety of features to be included without overfitting and poor generalization.

Initially, the tree-based models outperformed the neural networks. Then, some categorical variables, like company name, route ID, supplier country, and port names were converted to embeddings which are learned throughout training. Rather than encode as boolean, sparse vectors, the dense embeddings are able to find similarities within the feature space and provide meaningful information to the model. This practice allows the various categorical features, that are natural to this type of data, to be adapted fully for machine learning modeling.

6.2 Delay Analysis and Predictions

6.2.1 Analysis

As discussed above, arrival delays were reviewed with considerable differences in timeframe.

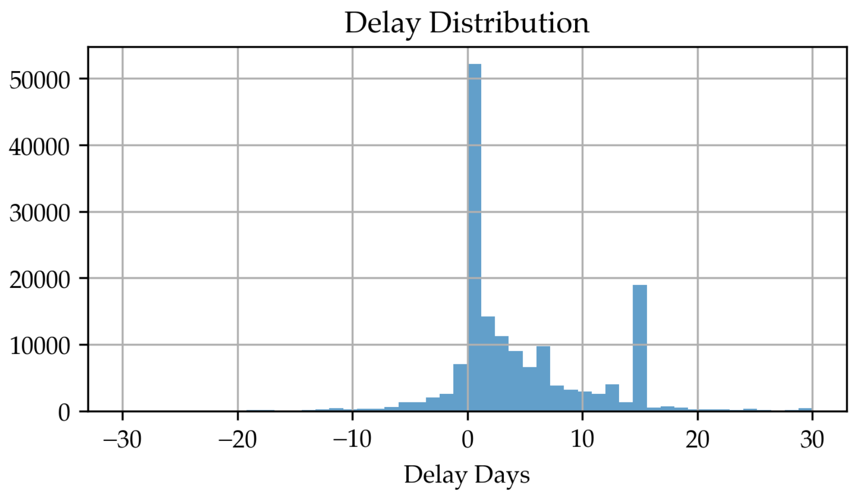

Figure 3: Delay Distribution

Delay Distribution

Category	**Count (% of Total)**
Early (< -1 days)	17,102 (10.7%)
On Time (-1 to 4 days)	86,578 (54.1%)
Slight Delay (4–9 days)	23,200 (14.5%)
Major Delay (> 9 days)	33,243 (20.8%)

Yearly Summary

Year	**Shipments**	**Avg Delay**	**Median**	**Major Delays**	**Major %**
2020	24,270	2.10	1.00	1,180	4.9%
2021	27,055	7.71	7.00	11,408	42.2%
2022	34,701	5.07	3.00	9,571	27.6%
2023	30,156	3.05	1.00	3,579	11.9%
2024	26,257	3.94	2.00	4,730	18.0%
2025	17,684	3.54	1.00	2,775	15.7%
TOTAL	160,123	4.33	2.00	33,243	20.8%

Statistical Test: 2021 vs Other Years. The average delay in 2021 was 7.71 days, compared to 3.64 days in other years. A t-test comparing 2021 to the other years yielded a t-statistic of 99.778 with a p-value ≈ 0, indicating that the difference is highly significant.

6.2.2 Prediction Models

Tree-based prediction models were built on data post June-2022.

Model Performance on Delay Prediction

Model	**RMSE (days)**	$\mathbf{R^2}$
Baseline RF	3.303	0.663
Tuned RF	3.406	0.642
Tuned XGB	3.712	0.575

Notes: Post June 2022. Delay statistics: mean = 3.60 days, std = 5.69 days.

The random forest model showed some predictive power as RMSE in days was 3.3 days, compared with 5.7 days standard deviation on the data.

Feature Importances for Delay Prediction

Feature	**Importance**
carrier	0.2285
day_of_year	0.1616
months_since_start	0.0770
arr_port_volume	0.0513
company_freq	0.0577
quantity_per_teu	0.0459

Important features include carrier, time-based features, port information and shipment volume.

6.3 Tariff Impacts

Note that specific countries listed in this section are based on the departure port field. Tariffs are to be assessed based on *Country of Origin*, which is not specifically listed on the bill of lading. Because of this, tariff evasion cannot be confirmed. Rather, this section looks at movements in business over the different periods.

Note that the periods reviewed in this section are April through September only, with the pre period representing years 2022-2024, and the post period representing 2025.

6.3.1 Shipment Counts

Because the data was overdispersed, in which individual country variances exceeded means by several multiples, negative binomial tests were used to assess significance of count changes as this distribution contains another parameter to adjust for overdispersion. The rates listed are shipments per day, to account for the different time periods pre and post.

Seasonally-Adjusted Negative Binomial Tests

Country	Pre→Post Rate	Change	Expected	Actual	p-value
Sri Lanka	0.29→1.50	+418.9%	52.4	272	0.0000***
South Korea	1.80→3.62	+100.7%	326.4	655	0.0000***
Belgium	0.75→0.05	-93.4%	136.5	9	0.0000***
Malaysia	1.15→1.93	+68.8%	207.4	350	0.0007***
Italy	1.15→0.40	-65.6%	209.0	72	0.0000***
Hong Kong	0.68→0.32	-53.2%	124.0	58	0.0001***
China	62.26→29.64	-52.4%	11269.5	5364	0.0000***
Spain	0.62→0.36	-42.4%	112.8	65	0.0006***
Indonesia	0.80→1.08	+35.7%	144.4	196	0.1564
Netherlands	0.31→0.20	-34.0%	56.0	37	0.7299

Sorted by absolute percentage change, the top eight countries noted significant p-values with respect to changes in shipment rates.

6.3.2 Weight Analysis

After shipment counts were reviewed, shipment weight (kg) analysis was added in case shipment weights had changed over the periods.

Rate-Adjusted Weight Statistics (Per Day): Pre- vs Post-Tariff

Metric	Pre-Tariff	Post-Tariff	Change
Shipments/day	92.34	62.00	−32.9%
Total weight/day (kg)	571,530	501,318	−12.3%

Shipments per day were down 33% overall. Weights per day were also down; however, a smaller amount at only 12%.

Weight by Origin Country (Apr–Sep Comparison, Sorted by p-value)

Country	Ship/day	Change	Mean kg/ship	Change	Total kg/day	Change	p-value	r
Sri Lanka	0.3 → 1.5	+419.8%	14782 → 6090	-58.8%	4273 → 9152	+114.2%	0.0000***	+0.24
China	62.2 → 29.6	-52.3%	5906 → 8448	+43.0%	367264 → 250369	-31.8%	0.0000***	-0.29
Indonesia	0.8 → 1.1	+36.0%	4781 → 11510	+140.8%	3807 → 12464	+227.4%	0.0000***	+0.15
Italy	1.2 → 0.4	-65.5%	2362 → 2334	-1.2%	2727 → 928	-66.0%	0.0000***	-0.18
South Korea	1.8 → 3.6	+100.6%	7031 → 7898	+12.3%	12682 → 28580	+125.4%	0.0002***	+0.18
Vietnam	14.0 → 14.9	+6.5%	6670 → 8177	+22.6%	93602 → 122210	+30.6%	0.0003***	+0.18
Spain	0.6 → 0.4	-42.2%	5951 → 6269	+5.3%	3700 → 2251	-39.2%	0.0006***	-0.14
Hong Kong	0.7 → 0.3	-53.1%	5361 → 4913	-8.4%	3665 → 1574	-57.0%	0.0008***	-0.13
Netherlands	0.3 → 0.2	-33.9%	4836 → 3916	-19.0%	1495 → 800	-46.4%	0.1399	-0.04
Germany	0.8 → 0.8	-0.6%	9867 → 13159	+33.4%	7624 → 10106	+32.5%	0.1652	-0.06

The amounts were broken out by country. The highlighted rows were countries that, when considering weight in addition to shipment count, now have significant changes in volume in the post period. Both Indonesia and Vietnam had increased shipment counts that were further enhanced by larger shipment weights.

Most of the r values showed a light effect, with China roughly a medium negative effect.

6.3.3 Route Shifting

In the period reviewed (April through September from 2022-2025), the proportion of shipments in which the supplier country does not equal the country of the departure port represented 20.4%.

Route Shifting Rate Pre- vs Post-Tariff

Period	Route Shift	No Shift	Total	Shift Rate (%)
Pre-tariff	17,246	73,643	90,889	18.97
Post-tariff	3,550	7,672	11,222	31.63

Change: +12.66 percentage points
Chi-square test: $\chi^2 = 986.26$, $p = 0.0000$ ***

Given the significant p-value from the Chi-Square test, there has been an increase in route shifting in the post period.

6.3.4 Chinese Supplier Analysis

Country export analysis was completed on Chinese suppliers only.

Origin Distribution (Same-Month Apr–Sep Comparison)

Origin Country	**Pre/day**	**Post/day**	**Change**	**Change %**
China	57.45	24.64	−32.82	−57.1%
South Korea	0.74	1.67	+0.94	+126.9%
Germany	0.00	0.37	+0.37	NEW
Vietnam	0.00	0.28	+0.27	+7482.9%
Canada	0.00	0.09	+0.09	NEW
Malaysia	0.02	0.08	+0.05	+226.6%
Spain	0.03	0.07	+0.04	+146.4%
Singapore	0.02	0.06	+0.04	+175.7%
Taiwan	0.07	0.06	−0.02	−26.0%
Sri Lanka	0.00	0.04	+0.04	NEW

The table above shows the rate of shipments per day of Chinese suppliers in April through September, with the pre period reflecting 2022-2024 and the post period for 2025.

Non-China Origin Shifts of Chinese Suppliers

Category	**Pre (per day)**	**Post (per day)**	**Daily Change (%)**
Non-China origins	1.10/day	2.93/day	+1.82 (+165.3%)
China origin	57.46/day	24.64/day	−32.82 (−57.1%)

Chinese suppliers exporting goods from non-Chinese ports increased while Chinese ports decreased. Mann-Whitney U tests and bootstrapped 95% confidence intervals of group means were completed with significant results in all metrics.

Shipment Counts (Shipments/day)

Category	**U**	**p-value**	**95% CI**
Non-China routing	67,740	$5.25e^{-15}$	[1.25, 2.45]
China direct	26,386	$2.78e^{-21}$	[−39.28, −26.99]

Weight (KG/day)

Category	**U**	**p-value**	**95% CI (KG)**
Non-China routing	68,917	$1.21e^{-16}$	[14,487, 30,641]
China direct	31,715	$2.80e^{-13}$	[−156,149, −99,195]

Both shipment count and weight tell a similar story of changes in the pre and post periods for the two origin groups.

7 Conclusions

This project studied over one hundred thousand footwear shipments over the past five years to answer the following:

Can landed costs be predicted?
Yes, for the available shipments with CIF terms, both tree-based and neural network regression models were able to predict costs with high R^2 values (over 85%) and predictive power over the standard deviation ($5-$6 compared with $16 std. dev.). Both these model types allow for flexibility in numerous categorical data, which is important for this data set. Interpretability is higher for the tree-based models over the neural networks; but overall performance for costs was highest for the attention model.

Can delays be predicted?
Partially. Tree-based models were used given the high number of categorical features as well as larger data set (160K+ lines, compared with costs only using the CIF data at 14K+). The random forest model predicted delays within 3.3 days; compared with an overall standard deviation of 5.7 days.

Are there impacts to recent 2025 tariff announcements and implementations?
Yes, based on the negative binomial (counts) and Mann-Whitney U tests (weight), a number of countries displayed significant changes in shipment rates pre and post tariff announcements (both increases noted in some southeast Asian countries like Indonesia and Vietnam along with the decrease from China). For Chinese suppliers, the rate and volume in which shipments originating from China vs. Non-Chinese countries changed significantly in the post period.

Which models / techniques perform best on this data?
As expected, tree-based models performed well given the flexibility and the high number of categorical features. In addition, the random selection of features and regularization allow for these models to adjust to potential correlation between features. Initially, these outperformed neural network models until new embedded features (which are randomly initialized then learned during training) like company name and supplier country were added. It is thought that using one-hot-encoding for the large number of unique categorical features did not help these models learn the data, and only when the dense embeddings were formed did the neural networks perform above standard machine learning techniques.

7.1 Considerations

As mentioned above, *Country of Origin* is the determination of tariff rates and not *departure port*. Lacking the ability to obtain a commercial invoice which should state the proper origin, this analysis had to work with the available data which only includes the port in which the goods last left before arriving in the United States. It cannot be said with certainty that changes in departure ports post-2025 tariffs announcements are suppliers actively avoiding tariffs. This analysis was prepared to identify if there were any changes in business behavior that could be statistically tested.

Also, the post tariff period is only six months, so the observed patterns may be temporary.

7.2 Future Analysis

Additional product searches for other goods would be a natural continuation of this research. In addition, a more in-depth analysis of vendor and customer relationships and concentrations could deepen the analysis and help explain larger business trends. Given the success of some categorical feature embeddings within the neural networks, additional feature engineering should be considered. Also, continuing to track movement from exporting countries would provide a longer-term picture of tariff impacts once more time has passed, as well as looking at other dates as cut-off points given the ever-changing announcements at the beginning and through 2025.

References

China Briefing, Arendse Huld (2025). China briefing. https://www.china-briefing.com/news/us-china-tariff-rates-2025/. Accessed: 2025-11-15.

ImportYeti (2025). Importyeti. https://www.importyeti.com/. Accessed: 2025-09-01.

National Geospatial-Intelligence Agency (2024). World port index (pub 150). Maritime Safety Information, https://wwf-sight-maps.org/arcgis/rest/services/Global/Ports/MapServer/0/query. Accessed via WWF-SIGHT: https://wwf-sight-maps.org/arcgis/rest/services/Global/Ports/MapServer.

Siripurapu, A. (2021). What happened to supply chains in 2021? Council on Foreign Relations, https://www.cfr.org/article/what-happened-supply-chains-2021.

Trade Partnership Worldwide (2023). Impacts of section 301 tariffs on imports from china: Case studies of apparel, footwear, travel goods and furniture. Technical report, Trade Partnership Worldwide LLC.

U.S. Census Bureau (2024). Usa trade online. https://usatrade.census.gov. Accessed: 2025-11-15.