**To:** Analytics

**From:** Kevin Regan

**Subject:** TEU Growth Rates

**Date:** 5/12/2025

**Abstract:** 90% of the world’s goods are transported on shipping containers which are incredibly heavy, weighing upwards of 4k to 9k pounds. As a result, repositioning, repairing, and storing a shipping container can be costly, if at the end of their contract or useful life are not redelivered to (i) high resale locations and (ii) low repair cost locations. This study seeks to determine whether the volume of Annual TEU (20ft Equivalent Shipping Containers) volume can be predicted accurately and to discover which ports offer the greatest growth potential.

**Motivation: I** am currently employed as a data scientist at Triton International, I thought it would be practical to research shipping port activities as part of graduate program.

**Research Questions:**

* Which ports and countries are growing the fastest in terms of annual TEU
* Which ports have the cheapest land and labor costs?
* Does a dataset currently exist that has compiled this information?
* Are there any relationships or correlations between port growth rates?

**Dataset:** The dataset was comprised of the top ~90 port locations throughout the world in terms of annual TEU (20ft equivalent Shipping Containers) volume ranging from 1.67 million to 50 million TEU per annum from 2020 to 2024. As a proxy for labor cost and property values local wages in USD and median home values were used. The intention of these proxies is to approximate repair labor costs and facilities costs. All values are in USD. Additionally, this data includes TEU growth rates for each year as well as the CAGR (Cumulative Annual Growth Rate).

**Methodology:**

**Data Collection:** Most of the top ports in the world list their annual TEU publicly although there are some sources that collate/ aggregate this information specifically worldshipping.org/top-50-ports

**EDA:** Initial review of the data includes computing medians,

and skew as well as producing some kernel density estimation.

graphs to determine the relative shape of the data as well as

accommodate for any skewing which was evident.

**Analysis:** Used python’s in-built descriptive statistics to obtain median, skew and kurtosis. As well as producing a variety of kernel density plots to understand the distribution of shipping container TEU, in addition I also computed a correlation Matrix, where I was surprised to see that the year over year correlation for all four years is nearly 1.0 with the minimum being .99. As a result of the high correlation, ports maintain their TEU rank and magnitude consistently year over year. It also means that recent history will be an excellent predictor of next year’s TEU. See Exhibit 1 for more information.

A chart with numbers and symbols

AI-generated content may be incorrect.

**Models Used:**

**K-means clustering (K=4)**

**Principal Component Analysis Isolation Forest**

**Linear Regression**

**Polynomial Regression**

**Piecewise Regression**

**Spline (Regularized using s=1)**

**Results:** Regression(s) on historic port volume yielded an average R2 ranging between .68 and .81 in the aggregate.

* Individual R2 scores were upwards of .99 indicating a high degree of predictability and stability especially with our outlier classified groups which including multiple ports in China and one in Singapore.
* Quanzhou Port in China is notable for three reasons.
  + Its cumulative annual growth rate is 23%, the highest of the observed population.
  + Its monthly local wage cost is the lowest of the observed population.
  + Its median home value is among the lowest in the observed population.

Principal component analysis identified four distinct groups (Accelerating, Moderate Growth, Flat and Declining)

A chart of different colored dots

AI-generated content may be incorrect.

A graph with a line and a line

AI-generated content may be incorrect.

**Performance Metrics:** PC1: 0.6493, PC2: 0.2142, PC3: 0.1109, PC4: 0.0241, PC5: 0.0011, PC6: 0.0002, PC7: 0.0001, PC8: 0.0001

PC1 & PC2 > 85%

A graph of a graph

AI-generated content may be incorrect.

Bibliography

The anomaly classification was not surprising, as nearly all anomaly locations were in Asia, specifically China (7), Singapore (1) and South Korea (1). Which anecdotally is known to be where the most of the world’s goods are produced.

A chart with different colored squares

AI-generated content may be incorrect.

**Conclusions**

**Limitations:** The observation period ended up being a 4-year period which included volumes during covid (an unprecedented time period), in many publications the data does not appear to go back further on publicly available data, which should be considered a limitation of this analysis.

**Conclusion:** This study sought to determine the feasibility of forecasting container traffic using publicly available port data. The analysis determined that there is a tremendous amount of volume stability over time with year over year correlation coefficients at or approaching 1.0 making it both feasible and reliable.

The clustering and principal component analysis enabled the segmentation of ports into categories based on growth behavior. This analysis also identified a single strategic port which has both a high growth TEU growth rate and low proxied operational cost. These classifications also identified multiple Asia based locations (specifically China) as high performing outliers / TEU locations. Which is in-line with expectations as China produces much of the world’s goods.

**Project Location:**

[**kjregan120/cs668\_capstone: Data and Analytics research related to annual Port Growth**](https://github.com/kjregan120/cs668_capstone/tree/main)

**Input Dataset:**

[cs668\_capstone/Top\_90\_Ports\_uploaded.xlsx at main · kjregan120/cs668\_capstone](https://github.com/kjregan120/cs668_capstone/blob/main/Top_90_Ports_uploaded.xlsx)

**Python EDA, and Modeling:**

[cs668\_capstone/TEU\_Growth\_Rates.ipynb at main · kjregan120/cs668\_capstone](https://github.com/kjregan120/cs668_capstone/blob/main/TEU_Growth_Rates.ipynb)

**Output File (with Classifications) and redundant growth rate columns**

[**cs668\_capstone/Port\_TEU\_Output.csv at main · kjregan120/cs668\_capstone**](https://github.com/kjregan120/cs668_capstone/blob/main/Port_TEU_Output.csv)