Time Series Forecasting for Aerospace Defense Exports

Problem Statement

The Aerospace Company currently produces turbojet, turbo propeller and gas turbine engines and parts for the civil aircraft industry but is looking to add a defense exports department to their business. They want their data analyst to verify that expanding from civil to defense exports is a safe business move and make a recommendation on who to market their products to. The company only wants to focus on marketing to countries that already buy these items from US companies. These countries should be able to afford the engines and parts, and have a need for them.

Data Cleaning

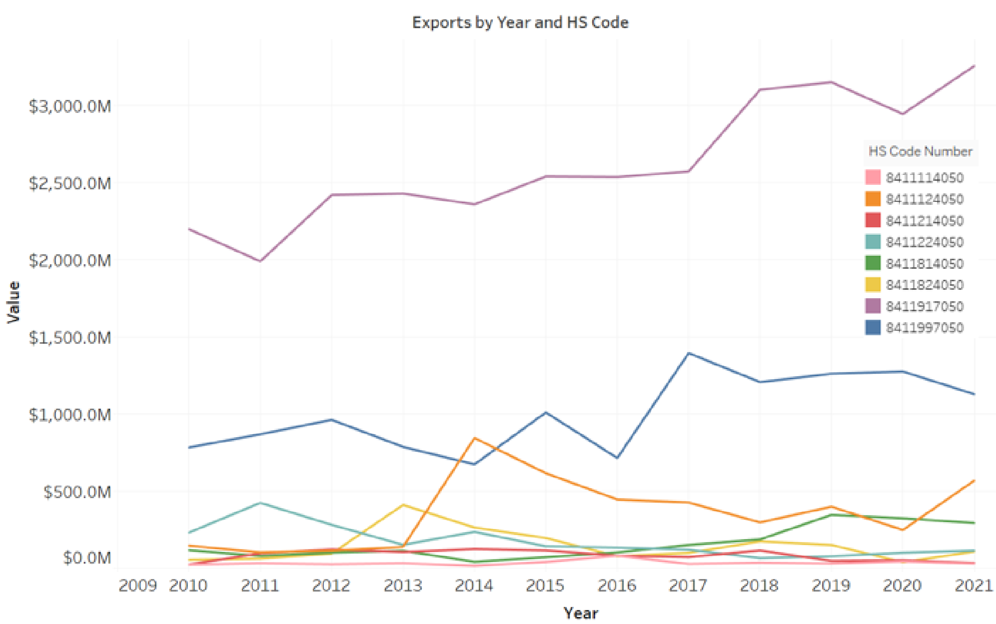
Historical export data is available from the International Trade Administration (ITA) of the U.S. Department of Commerce, where two different data sets were pulled. The first being total exports for Defense Aircraft and Parts (in millions of dollars) for each month from 1992 - 2021, which will be referred to as yearly_shipments. The second dataset is exports by year and country for all eight HS Codes that the company products fall under, which will be referred to as exports. This second dataset contains data from 2010 to 2021 for all countries that the US exports any of the eight categories to, as well as a sum of those categories worldwide exports designated with a Country equal to '.WORLD '.

The data pulled from ITAs resource is already very clean, not missing any data points and had only the data necessary to the problem. Export value information was in an object format with '$' included in both datasets; to prepare the data for analysis, dollar signs were removed and column types were changed to integers. For yearly_shipments, the exports values were in a large table formatted with years as rows and export amount for each month of that year as the columns. This was changed to a single column format with the year-month combined for easier plotting and modeling. The only other changes were in the exports data, with the first being to make the Country column data type a string instead of an object. The second was removing all rows for countries where the US exported $0 to, since the Company only wants to focus on known consumers. Again, the source for this data was already in a great format, so very little cleaning needed to be done

Exploratory Data Analysis and Pre-Processing

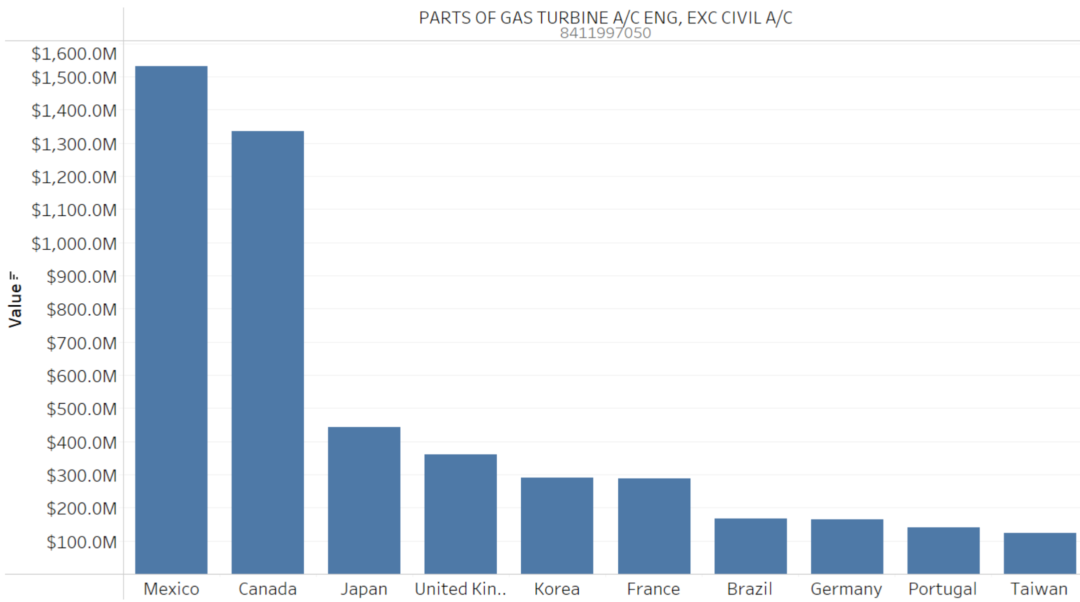
The first part of this problem is to find if moving into defense exports is a smart business move, so we first visualized the trend for total exports for defense aircraft and parts (in millions of dollars) from 1992 - 2021. The yearly_shipments data is broken into exports by month, which contains noise that makes the trend harder to visualize. So for overall trend visualization, the exports were summed for each year and plotted.

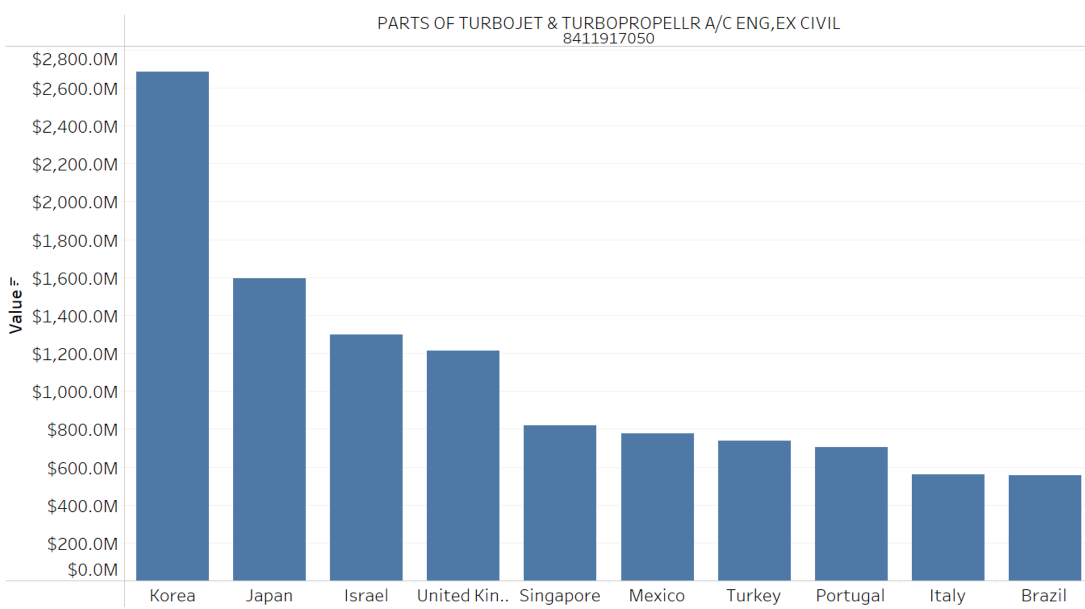
The value of yearly shipments seem to be increasing but also heavily impacted by factors outside of predictive means. For example, you can see the increase in exports in the years following 2001, when the US was attacked and went to war. These types of events are unpredictable but have a large effect on defense exports. The overall trend looks good, but the Company is focused on the products that they produce, being turbojet, turbo propeller and gas turbine engines and parts. The comparison of these eight categories over the last 12 years is shown below.



There are eight different HS Codes in this dataframe which is a lot to look at. This graph shows that HS Code 8411917050, parts of turbojet and turbopropeller engines, and HS Code 8411997050, parts of gas turbine engines, were the top two exports by a wide margin, at $3.2B and $1.1B in 2021 respectively. The company would be smart to start into the export business gradually, and focusing on these top two categories is the best business move. The next two highest HS codes would be 8411124050 at $0.57B and 8411814050 at $0.29B in 2021. These two categories could be explored in a future iteration of this project.

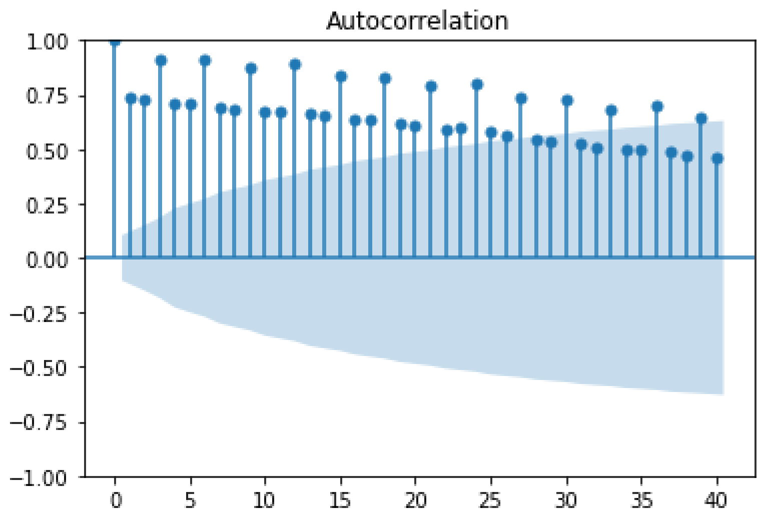
The data for both HS Codes identified above were then separated into their own dataframes to see what countries were the top importers of goods. Parts for turbo engines were bought by 147 different countries, while parts for gas engines were bought by 153 countries. This is too many countries to focus on, and the company only wants to look at the top ten for each category. The dataset with HS Codes only has data from 2010-2021, to get the top ten countries over this time period, those 11 years were totaled and graphed visually. Results are below.



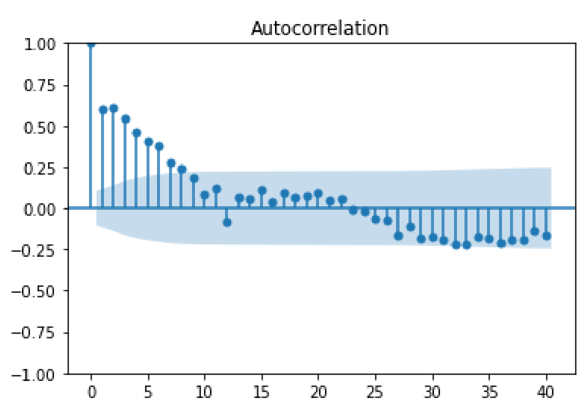


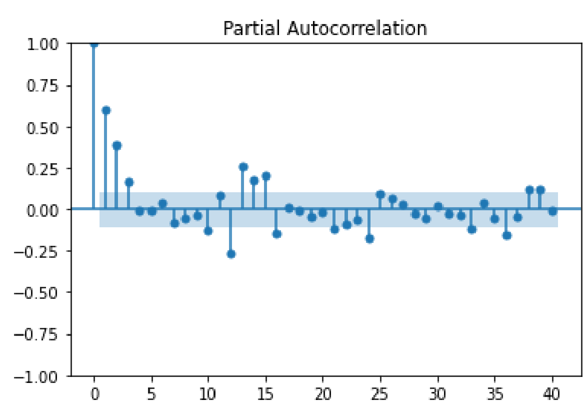
All of these countries would be smart to market to, as they are already buying these products. It helps that there is some overlap in the groups, with Korea, Japan, United Kingdom, Mexico, Portugal and Brazil in both top tens lists, making them good potential buyer options.

The above information is valuable, but time series forecasting is needed for the company to be sure that the move into defense exports will be profitable for them. For time series forecasting, the algorithms like for data to be stationary, which can be tested for with a Dickey-Fuller test on the Export Value column of the yearly_shipments dataframe. The results showed a test statistic (p-value) of 0.73, which is much larger that the critical value of 0.05, meaning the data needs to be transformed to make it more stationary. Both autocorrelation and partial autocorrelation plots were made to see if the data showed any seasonality, the autocorrelation plot is shown below.



You can see the spikes every three lags, with the 12th lag being larger than the one before it, which shows a quarterly and yearly trend to the data. Yearly is the larger trend, so the yearly difference of the data was taken and the autocorrelation and partial autocorrelation plots were again produced.





The test statistic (p-value) of this year differenced data was 0.027, smaller than the critical value of 0.05, so this should be enough of a transformation for the time series models to run properly.5

Algorithms & Modeling

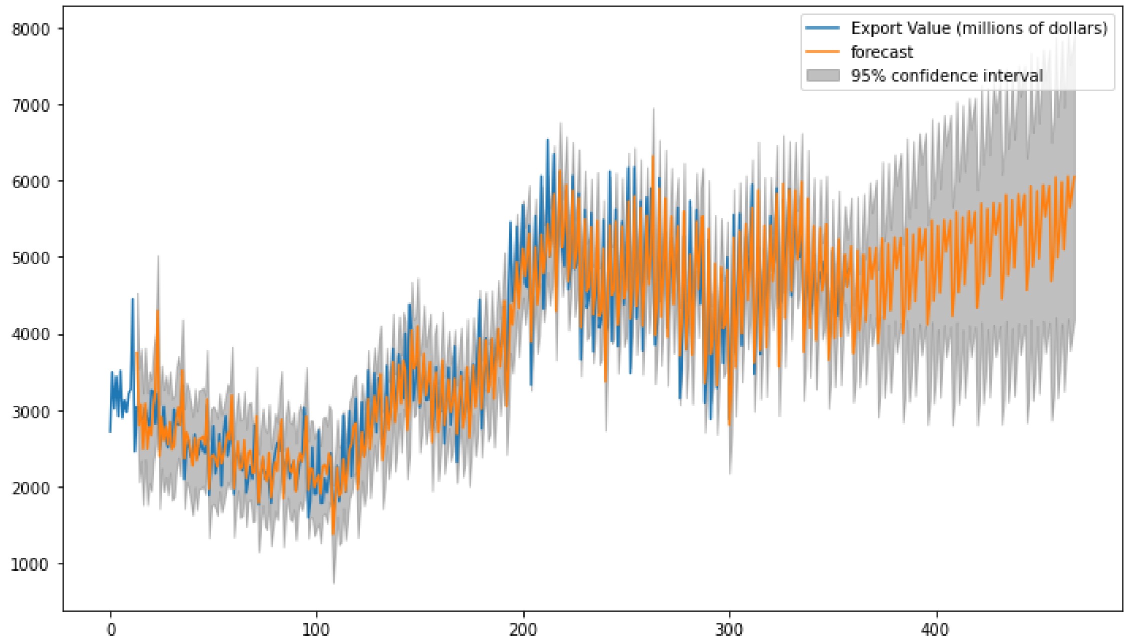
Four models were created for the yearly_shipments data. Three of those were autoregressive integrated moving average (ARIMA) models, and one was a seasonal ARIMA (SARIMA) model. All three ARIMA models were run with the year differenced data, while the SARIMA model was run with the original Export Value data since the model itself can account for the seasonality of the data. Every model was compared against each other based on the resulting Akaike information criterion (AIC) and the Bayesian information criterion (BIC) values.

The first ARIMA model was a pure autoregressive model, with parameters of (3,0,0). The second was a pure moving average model, with parameters of (0,0,7). Neither of these models were performing as well as desired, so a combined ARIMA model was produced with order of (2,0,1). This model performed closer to that of the AR, but I still felt like more improvement could be made. Knowing that there was a yearly and quarterly seasonality to the data, it made sense to try a Seasonal ARIMA model that took into account the yearly trend. Many iterations were made on the order and seasonal order parameters to find the best performing ones, which ended up being an order of (1,0,2) and a seasonal order of (0,1,1,12). The results are listed in the table below.

Model	AIC	BIC
AR	5116.5	5135.9
MA	5127.1	5162.1
ARIMA	5117.9	5137.3
SARIMA	5039.3	5062.4

Results

The Seasonal ARIMA model had the best fit for the data based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), and the plot of the model prediction versus actual fit very well.



Based on this graph, the model predicts that yearly shipments of aerospace defense goods will increase over the next decade. So it is a safe business move for the company to move into this export sector, focussing specifically on the two top categories first, then expanding further based on the success of that move.

Future Steps

This project covers a good first step for the Aerospace Company to make a move into defense exports, but further steps could be taken. Depending on the profitability of the first two export categories, the company may want further analysis on other defense sectors. Since this dataset does exhibit periodic behavior, another future part of this project could be to integrate spectral analysis to better understand the underlying periodic behavior. Which in turn could lead to a better predictive model, but it is outside the scope of this project.

Additionally, data could be collected that would help model some of the outside forces affecting the value of exports. Such as who the US has partnerships with, geographical location or types of aircraft in a specific country's arsenal that the company produces parts for. All of these could factor into the actual exports this company can count on, but is not information that is currently included available for this project.