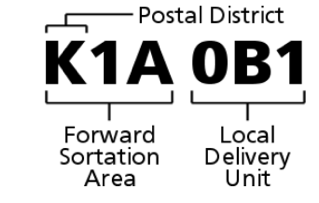
**Overview**

One of the responsibilities of the Canada Transportation long and medium-term planning team is to determine the geospatial distribution of carrier coverage. Canada transportation team currently uses 1 year of historical data and average it per postal code to generate their forecasts each quarter, and seeks to improve their long and medium planning forecasts of customer demand in different geographical areas through automation. To meet these needs, the MOP team has developed 5 year-out demand forecasts for about 1,652 postal codes, which is currently at Forward Sortation Area (FSA) granularity level (Zip3) in Canada. This project will implement a system that will provide forecasts of the proportion of Amazon shipments in Canada per postal code region.

**Data Description and Exploratory Analysis**

Fig.1 shown below, describes the basic structure of Canada postal codes.



The Zip6 postal codes data source is not readily available. We decided to use FSA data for our analysis and to generate forecasts. The FSA granularity describes a geographical region in which all the FSA start with the same three characters. The Canada FSA data was extracted from the d\_perfectmile\_pkg\_attributes\_v2\_na and d\_outbound\_shipment\_package tables, and it includes the sum of the packages shipped at the weekly level per FSA.

Four years of data were extracted for this project (June 2017 – July 2021), and are based on weekly historical CA outbound shipment packages data at the FSA level. We explored the geographical information by calculating the percentage shipments, and the average distribution of the FSA at the weekly level. FSAs were divided into 8 percentile buckets (1, 5, 25, 50, 75, 95, 99, 100) based on mean weekly shipments. To have a stratified sample, one FSA was randomly selected from each bucket, and we computed the average number of packages per FSA (see appendix A1). To obtain the number of packages forecasted for an FSA, it is important that the final output of the forecasts is presented as proportions of total network shipments. This is because the proportion of network shipment per FSA is multiplied by the total Network shipments, which come from the Topline Forecast and Unit per box (UPB) as shown in Fig.2 below.

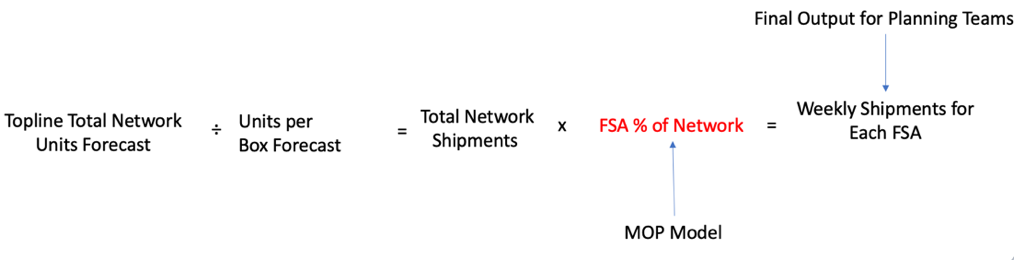


Fig. 2. Final output demand forecasts

The fig. 3 below shows the time series graphs of how the FSA reacted differently from the onset of Covid – 19 (see appendix B for the rest of the plots). From the V4B and R0K FSA plots, we see a consistent seasonality pattern between July 2017 and December 2019, but from March 2020 to December 2020, we see an irregular seasonality pattern due to a sharp increase in the shipment packages. The plots further showed a decrease in shipment packages between January 2021 and July 2021. Generally, we see that the growth rate in the shipment values continued when compared to the weeks before the onset of the Covid-19 (from July 2017 to November 2019).

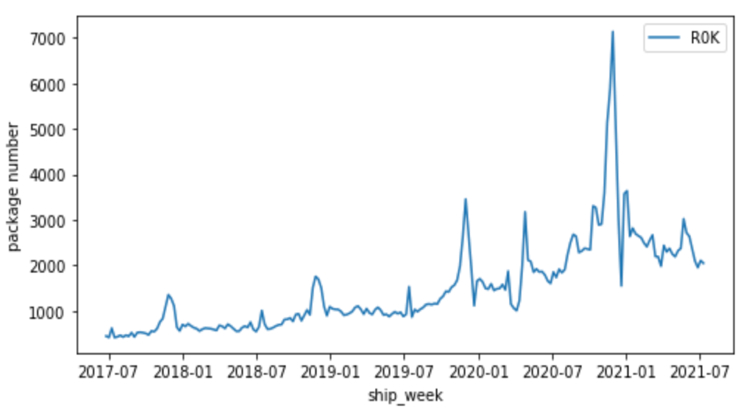
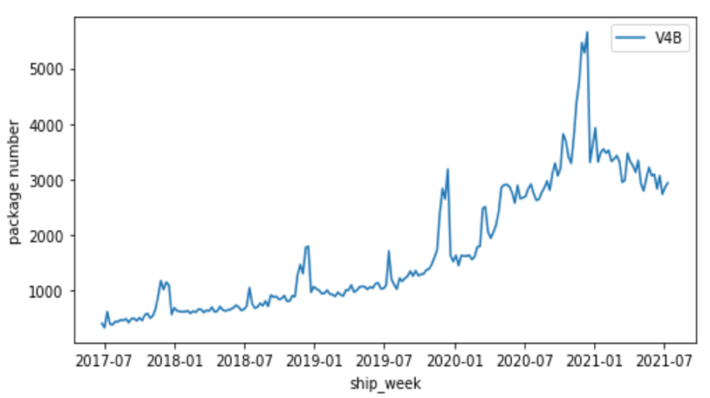


Fig.3 Percentage shipments and Package numbers of V4B and R0K FSA

**Pre-processing:**

The FSA postal codes data was divided into two subsets for training and testing purposes. Four years (212 weeks) of data were extracted for the modeling phase. We use 3 years (156 weeks) of data for the training dataset, and 1 year (52 weeks) of data for the testing dataset. We split our data as follows: the training dataset: 2017 – 06 – 11 to 2020 – 06 – 28, and the testing dataset: 2020 – 07 – 05 to 2021 – 06 – 27.

**Modeling**

We created two ARIMA (Auto-Regressive Integrated Moving Average) models to compare against their current methodology, which is based on 1-year averages. One of the ARIMA models is non – seasonal, and the other is seasonal, also called the SARIMA model (Seasonal Autoregressive Integrated Moving Average). We fit the ARIMA and the Seasonal ARIMA models for each FSA and then generated the forecasts across all the FSA. To evaluate the model performance, we compared the ARIMA Model and the SARIMA Model forecast errors to the baseline forecast errors. For clarification, we denoted: V0:1-year averages (baseline methodology), V1: as the ARIMA model, and V2: as the SARIMA Model

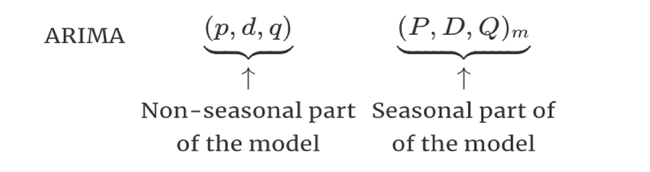
ARIMA model, also known as the non-seasonal ARIMA model, uses past errors in a regression-like model and states that the next observation is the mean of every past observation. In other words, the ARIMA algorithm is based on the idea that the information in the past values of the time series can be used to predict future values. ARIMA model has 3 order characteristics, ARIMA (p, d, q) model.

p = order of the autoregressive part;

d = degree of first differencing involved;

q = order of the moving average part.

We use auto.arima() function to generate the order (lowest Akaike information criterion, AIC) and the order generated was (1,1,3) order. We went further to fit the data with the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The SARIMA model supports the time series data with a seasonal component. It is formed by including additional seasonal terms in the ARIMA model as shown below:

****

Where P is the seasonal autoregressive order, D is the seasonal difference order, Q is the seasonal moving average, and m represents the number of observations per year. For the SARIMA model, we used (1,1,1) order for the non-seasonal part, and (1,0,1) order for the seasonal part of the model with a frequency of 52. The parameters are generated automatically from the **pmdarima** package. The package chooses the parameters with the lowest AIC (Akaike information criterion) and then fits the model with the chosen parameters. For further comparison of the model performance and selection, we decide to have a third version model, which is the Prophet model (V3). The prophet model works best with time series that have strong seasonal effects and several seasons of historical data. Prophet model library utilizes the additive regression model y(t) comprising the following components:

t

Where g(t): models non – periodic changes, s(t): represent periodic changes (e.g., weekly/yearly seasonality), h(t): effects of holidays and is the error term that accounts for any unusual changes not accommodated by the model. The forecast errors of the ARIMA, SARIMA, and Prophet models were compared to the current forecast error, to know how these models performed.

**Model Performance**

Percentage errors are generated to evaluate model performance. We considered MAPE (Mean Absolute Percentage Error), and WAPE (Weighted Average Percentage Error) to measure the percentage errors of the forecasts to actual values. We used WAPE to select the best model because it ensures that the small differences in FSA that have lower demand don’t skew the evaluation metric. Both MAPE and WAPE are derived from the Absolute Percentage error (APE). APE is calculated as shown below:

The MAPE is calculated as follow:

Also, below is how WAPE is computed:

Where Wt is the weighted actual values, and t represents FSA

We calculated the MAPE and WAPE error metrics on all the models. The same test dataset was used to compare the performance of the baseline model and our model. The line graph and the table shown below (Fig.4) summarizes the comparison of the current forecast methodology (V0) and the models.

**MAPE Comparison**

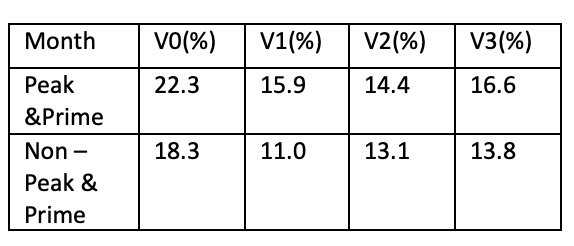
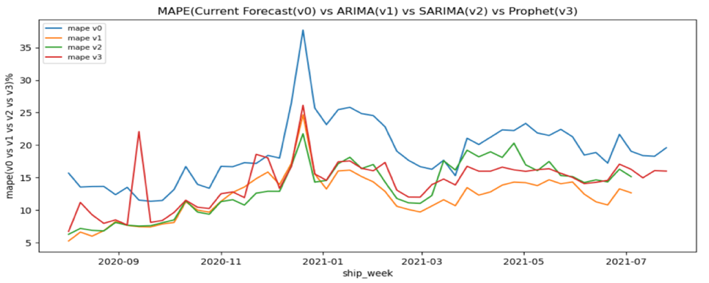
****

Fig.4 Test (out-of-sample) MAPE (Mean Absolute Percentage Error)

As shown above, we can see that the forecast errors of the ARIMA, SARIMA, and Prophet model performed better than the current forecast error, except that in September 2020, the Prophet model forecast error was the highest compared to the current forecast and the other models. In the peak season weeks, December 2020, we see a high MAPE of the models. The current forecast error was the highest followed by the ARIMA and the Prophet model. This is as a result of the expected high-volume shipment packages during the peak season. In addition, the table showed that the SARIMA (V2) model had the best model performance in peak seasons while the ARIMA (V1) model was better in off-seasons.

**WAPE Comparison**

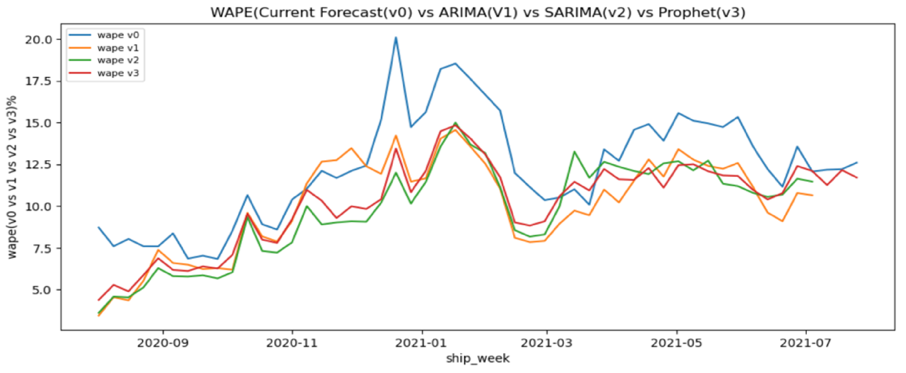
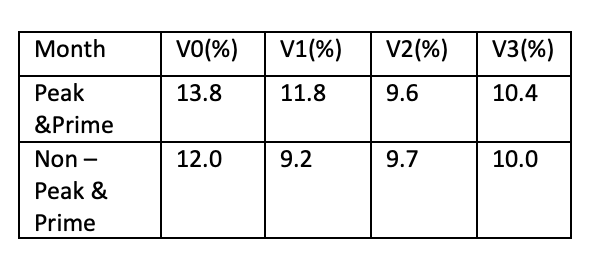
** **

Fig.5 Test (out-of-sample) WAPE (Weighted Average Percentage Error)

The WAPE of the models as shown above revealed that the models’ forecast errors are very close and keep increasing from March 2020 to November 2020. As expected in December 2020, we see a high spike in their forecast errors respectively. In addition, we can see a sharp drop in January 2021, but from February 2021 to July 2021, the forecast errors continue to increase. We noticed from our investigations that these fluctuations were a result of an increase in shipment packages during covid-19 (see appendix B). With 9.6% peak and 9.7% non-peak seasons, we see that the WAPE of the SARIMA (V2) model performed better than the rest of the other models (see Table in fig.5 above).

**Deep Dive of Model Errors and Pre – Pandemic Analysis**

As a result of the MAPE and WAPE outputs, we investigating the FSA that have high errors. The actual number of packages and their corresponding proportions of each FSA were plotted against the predictions of the ARIMA and SARIMA models as shown in Appendix C and D. As we can see in Appendix C, the ARIMA model (V1) generated flat lines across the weeks, and are below the actual values of packages in the FSA. The flat lines are a result of weekly data used for the Moving Average model, and since there is no recent data added to it, the forecasts stayed the same in most of the weeks. Also, fig.6 plots showed the SARIMA model (V2) version against the actual number of packages and their proportions. From the plots, we see that the model predictions are lower than the actual packages, and an increase in the volume of packages distributed in the prime and peak seasons (17500 packages/week in October 2020 and about 31000 packages/week in November and December 2020). This pattern is reflected in the rest of the plots found in Appendix D.

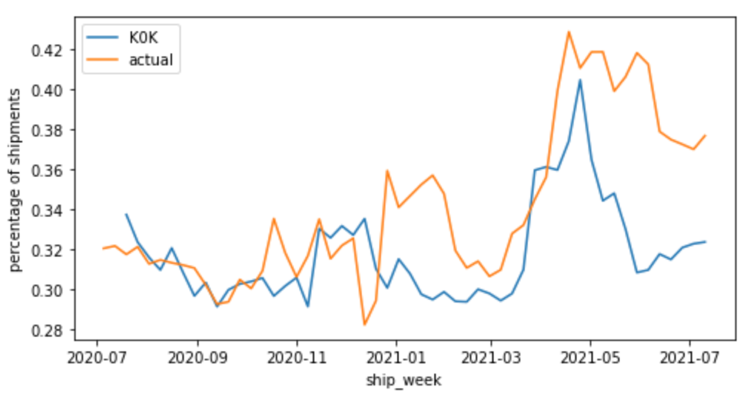
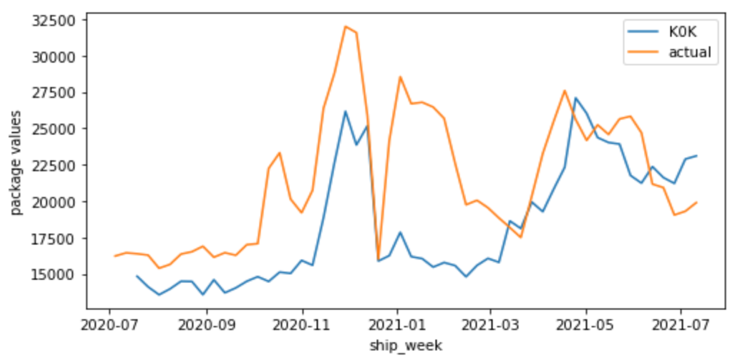
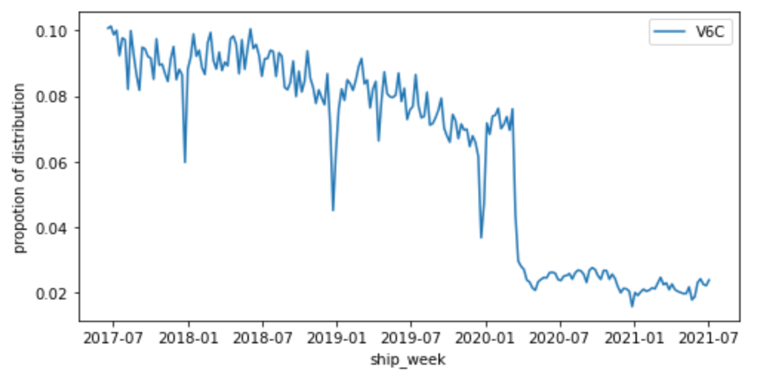
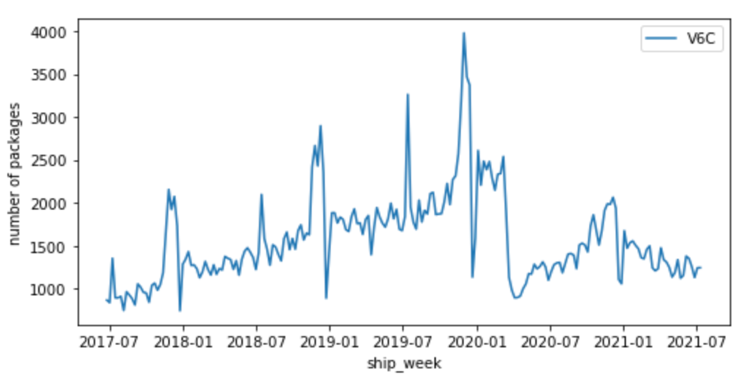


Fig.6 SARIMA Model (V2) Versus Actual values (K0K FSA)

We further our analysis by looking into the behaviors of all the FSA with high forecast errors. Appendix A2 showed the first twenty FSA with high forecast errors. The investigation revealed that most of the FSA postal codes with high forecast errors are found in the 1st, 5th, and 25th percentiles. As shown in Fig.7, V6C and H3B FSA found in the 50th and 70th percentiles reflected a sharp drop in shipments proportion distribution during the peak of Covid-19 periods (from March 2020 to July 2021). We discovered that these areas are the downtown commercial and university areas. This helps us to understand why the models’ errors are high in these FSA.

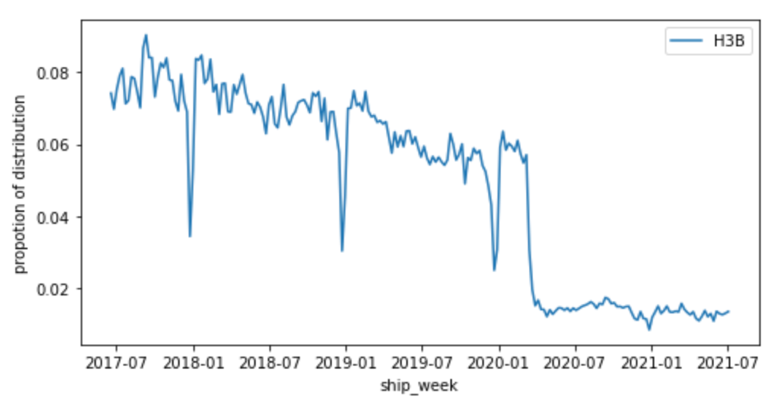


Fig.7 High Error Postal Codes in the 50th and 75th Percentiles (V6C and H3B FSA)

Furthermore, we investigated what happened before Covid-19 started. Carrying out the pre-pandemic analysis will give us more insights into how data would have performed. Our belief is that, when we compare the model performance of our models before and during the pandemic, it will provide more insights on the model selection because we had stable data before the pandemic. A similar analysis was carried out as we did from 2017 to 2021. Four years of data were also extracted for this analysis (2015 – 2019), and we explored the data using the same FSA that were randomly selected from the 8 percentile buckets as shown in table 1 above. The pre-processing was carried out, and we fitted the models with the pre-pandemic data. We compared the ARIMA model, the SARIMA, and Prophet models to the current forecasts before the pandemic. Since the current forecast uses a Moving Average of 1 year of historical data, we extracted data from 2017 – 12 – 31 to 2018 – 12 – 23. Training dataset; 2015-12-06 - 2018-12-23, and testing dataset; 2018-12-30 - 2019-12-22.

**Pre – Pandemic Model Evaluation**

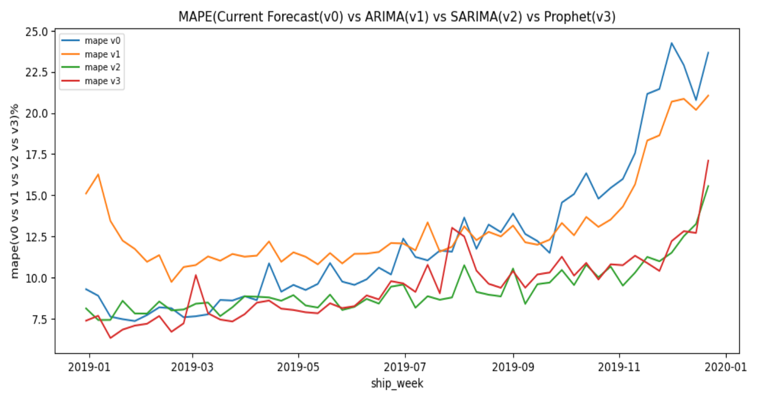
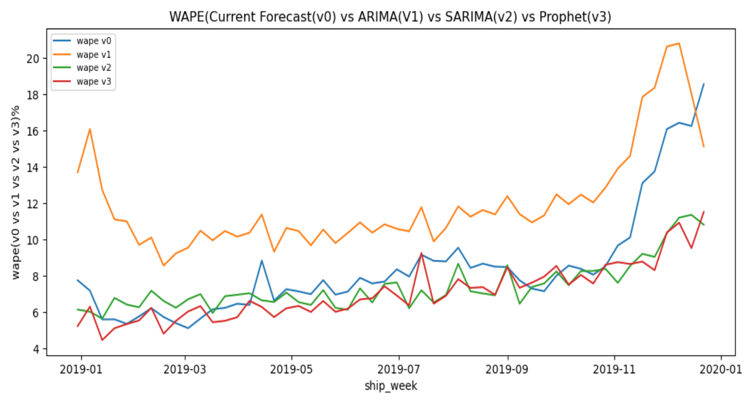
 

Fig.8 Test MAPE and WAPE of the Pre-pandemic Forecasts

MAPE WAPE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Month | V0(%) | V1(%) | V2(%) | V3(%) |
| Peak and Prime | 19.0 | 18.5 | 11.5 | 11.9 |
| Non-Peak and Prime | 11.2 | 12.2 | 8.9 | 9.0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Month | V0(%) | V1(%) | V2(%) | V3(%) |
| Peak and Prime | 13.9 | 16.9 | 9.4 | 9.3 |
| Non-Peak and Prime | 7.5 | 11.2 | 7.1 | 6.7 |

Table.3 Test MAPE and WAPE of the Peak and Non – Peak seasons

As shown in Fig.8, investigations revealed that the MAPE of the models performed similarly, except that the ARIMA model was higher in January 2019. As expected, in November and December 2019, where we see the high volume of packages, the CA forecast methodology and ARIMA model errors were higher compared to the SARIMA and Prophet models. On the other hand, we see that the ARIMA model performed poorly compared to the other models. Also, we see high errors metrics of the current methodology and ARIMA model in Peak and Prime months.

In conclusion, from the summary table in Fig.5, we see that the SARIMA model performed better because it has the lowest error metrics when compared to other models in peak and off-peak seasons. Before the pandemic (WAPE, Table 3), the Prophet model differ from the SARIMA model by 0.1% and 0.4% in peak and off-peak seasons (small difference between the two models). Therefore, in terms of the model performance evaluation, model processing reliability, consistency, and interpretability, we decided to select the SARIMA model for generating the Canada forecasts.

**Next steps**

We created the DasBoard Application for Canada forecasts, and the codes have been reviewed and approved by the team members. Currently, we are finalizing the gamma and production stages of the application and expect to complete the process before the end of the internship. In addition to the Canada forecasts, the MOP team has received a request for Mexico geo-forecasting. For Mexico geo – forecasting project, the same analysis for Canada geo-forecasting project can be replicated to generate forecasts for Mexico. Before the Mexico geo – forecasting project, we will recommend that Mexico's current forecast methodology is known earlier before analyzing the data. Also, it is important to know the postal codes granularity level for the data analysis and forecast generation.

In addition, we will recommend either a process called Seasonal Adjustment (removing seasonal component) or creating an additional advanced model that can capture complex seasonality in data. For instance, the use of a Triple Exponential Smoothing model also called Holt-Winters Exponential Smoothing. This model is carried out by adding a parameter (called gamma) that can control the influence on the seasonal component. It can be modeled as either an additive or multiplicative process for a linear or exponential change in seasonality. For our data, the best option is to use multiplicative seasonality. Comparing the error metrics of these models and choosing the best model will help to generate the forecasts.

**Appendix A1**

Table 1. Randomly Selected FSA in the Buckets

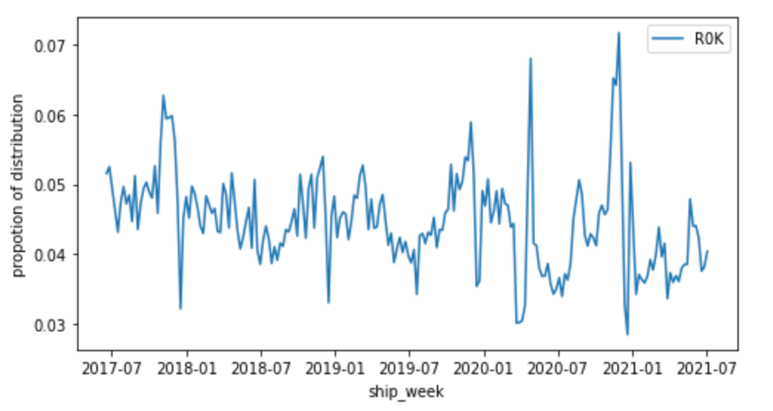
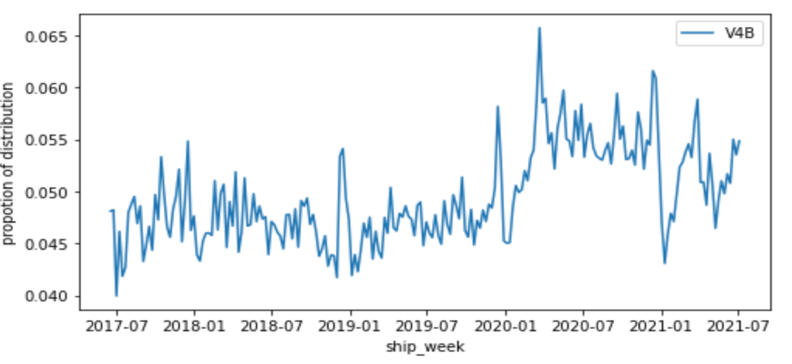
|  |  |  |  |
| --- | --- | --- | --- |
| **Percentile** | **Percentile Bucket** | **FSA Sample** | **Average Package Number** |
| 1 | (3.994, 29.486) | H0M | 10 |
| 5 | (29.486, 121.339) | G7Z | 112 |
| 25 | (121.339, 675.002) | G8J | 154 |
| 50 | (675.002, 1529.764) | R0K | 1481 |
| 75 | (1529.764, 2797.087) | V4B | 1720 |
| 95 | (2797.087, 5399.393) | H3C | 2933 |
| 99 | (5399.393, 9431.656) | N0B | 8634 |
| 100 | (9431.656, 16133.156) | K0K | 10689 |

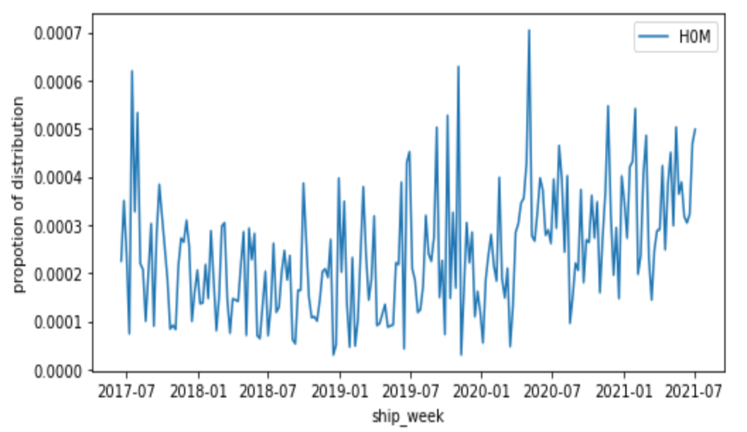
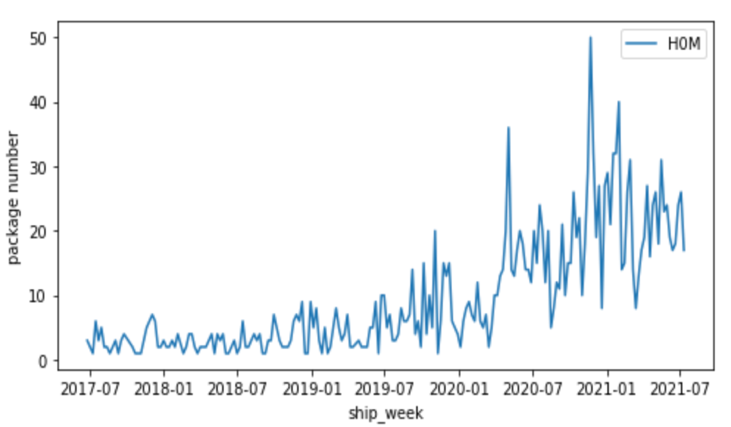
**Appendix A2**

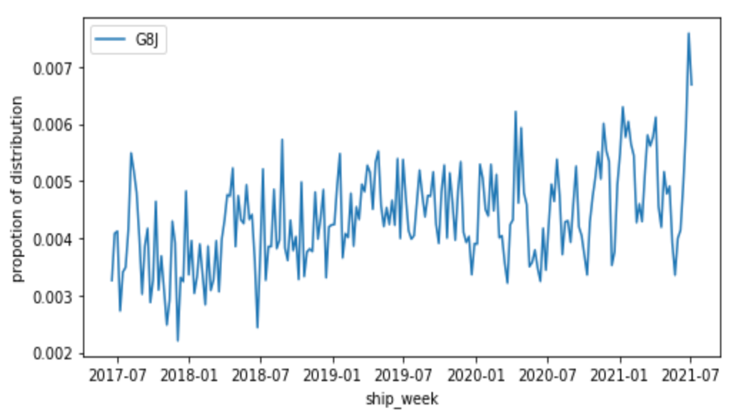
Table 2. FSA postal codes with High forecast errors

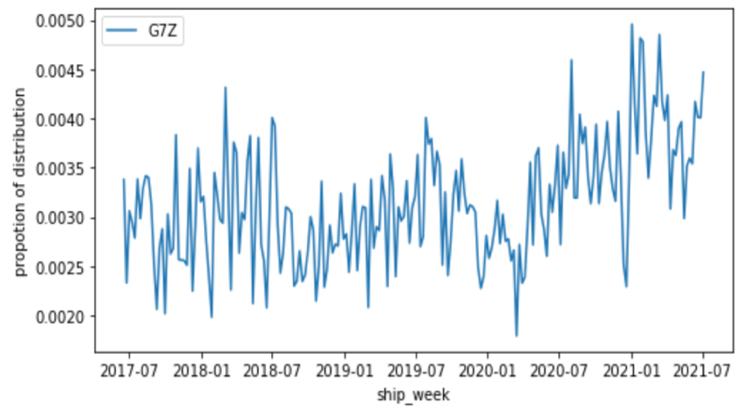
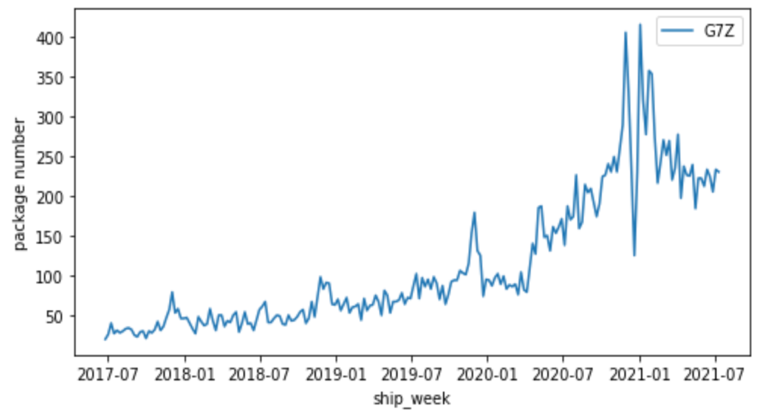
|  |  |  |
| --- | --- | --- |
| **Percentile** | **Percentile Bucket** | **FSA with High forecast errors** |
| 1 | (3.994, 29.486) | H4Y, L5P, T0P, Y0A |
| 5 | (29.486, 121.339) | B1M, H4Z, H5B, M5K, M5L |
| 25 | (121.339, 675.002) | B1W, H4T, K1A, K1P, M5X, V7X, V7Y, X0C |
| 50 | (675.002, 1529.764) | H3B |
| 75 | (1529.764, 2797.087) | H3A, V6C |

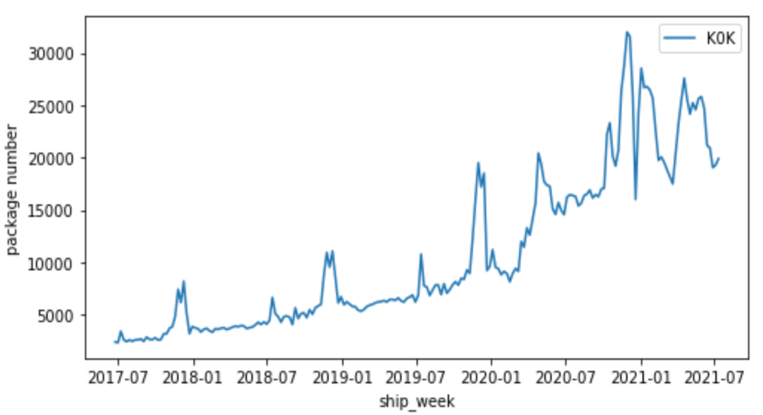
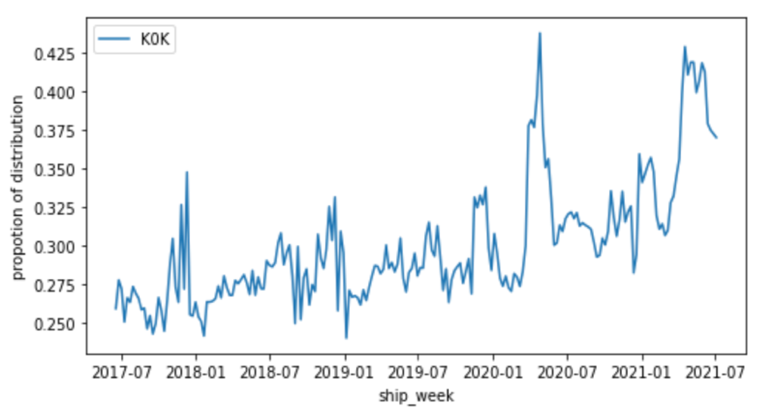
**Appendix B: *Percentage Shipments and Package Number of the Eight FSA***

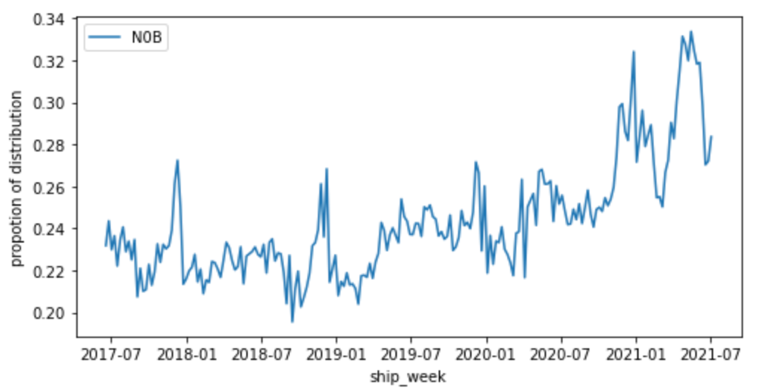
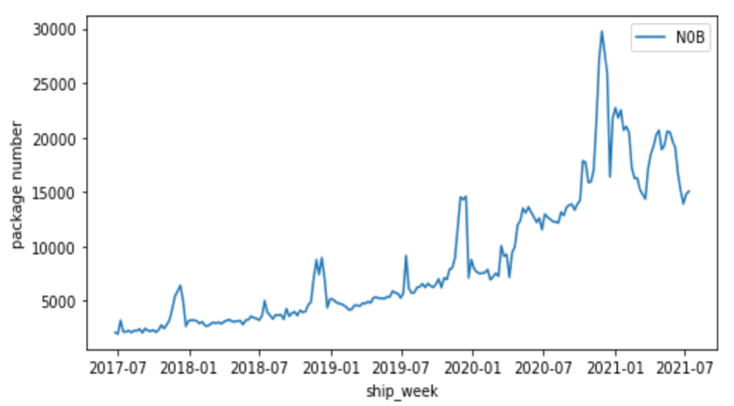


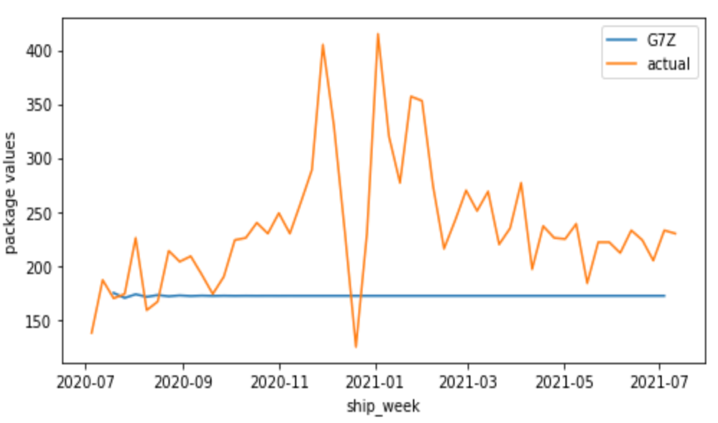
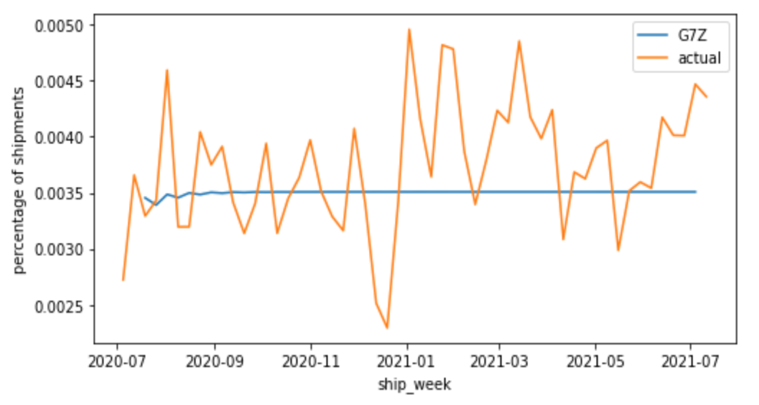
 

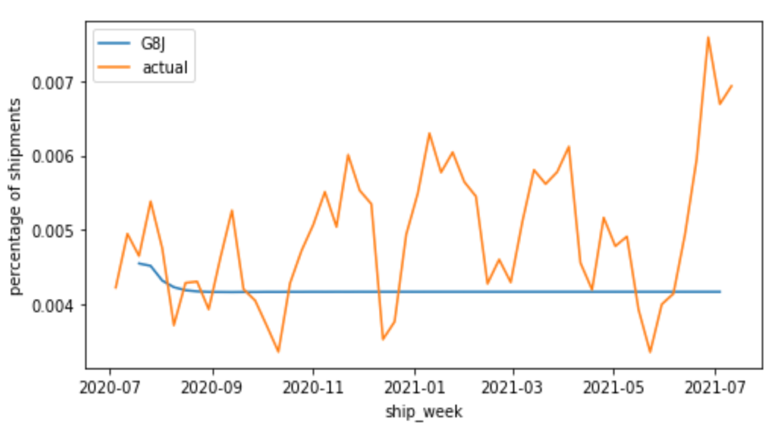
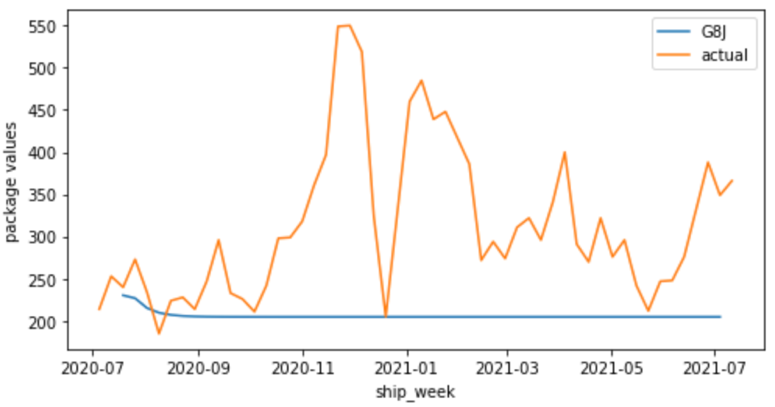
 

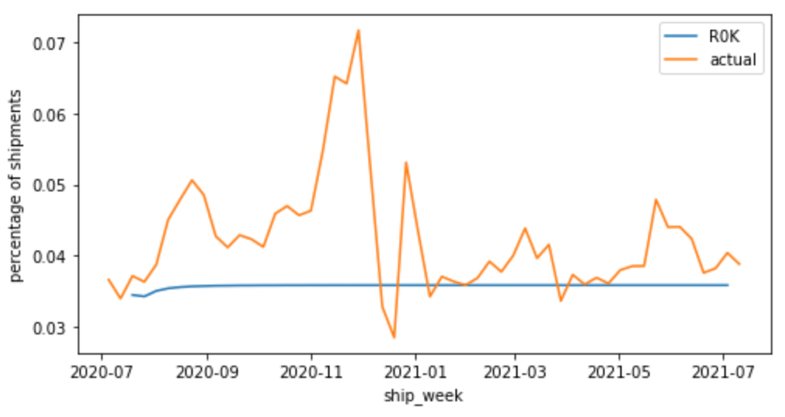
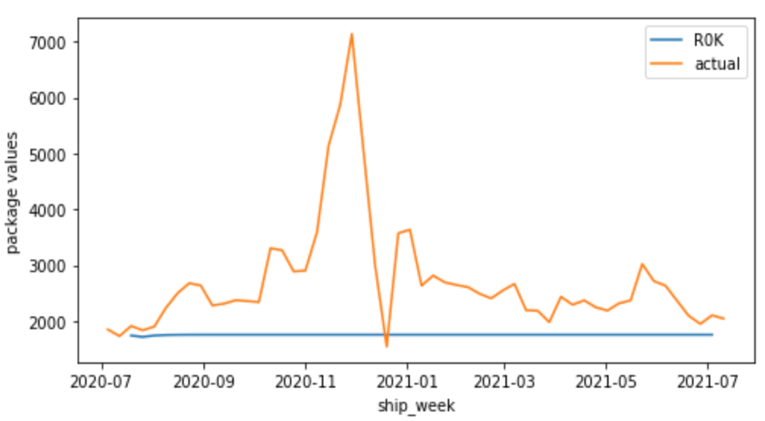


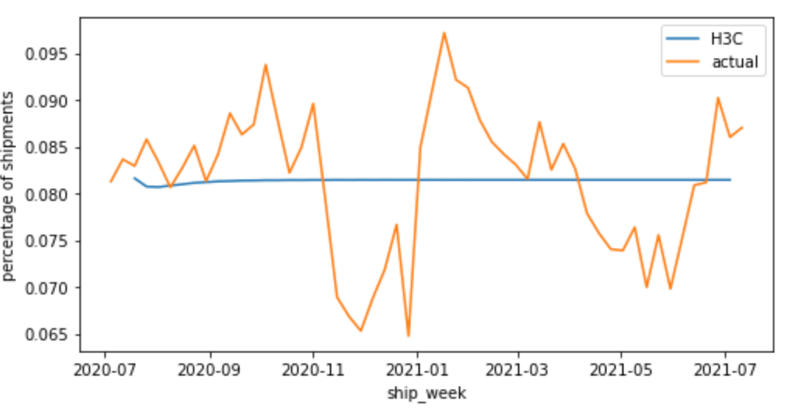
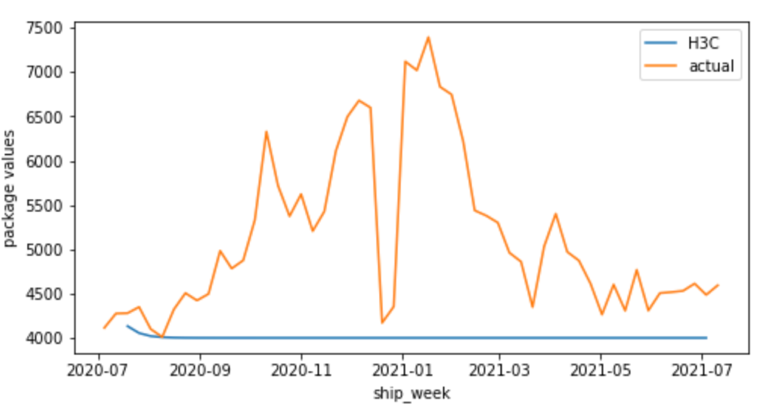
 

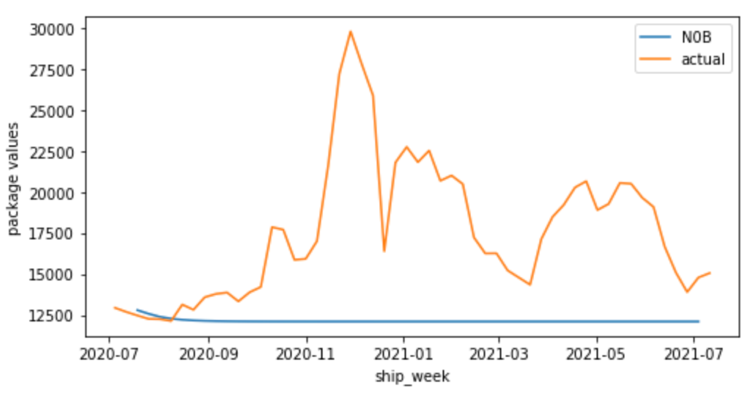
**Appendix C: *Moving Average Model (V1 version) Versus Actual Values***

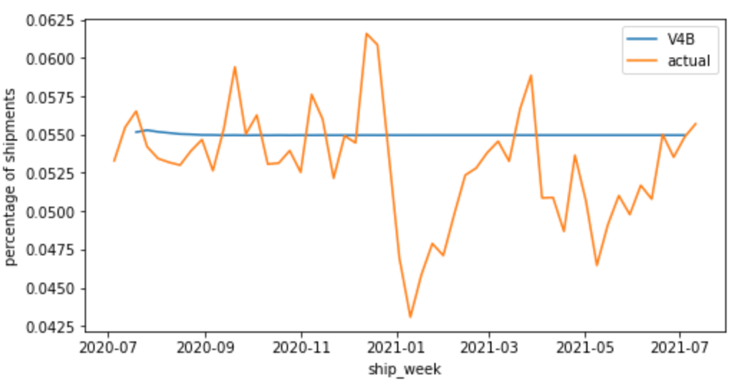
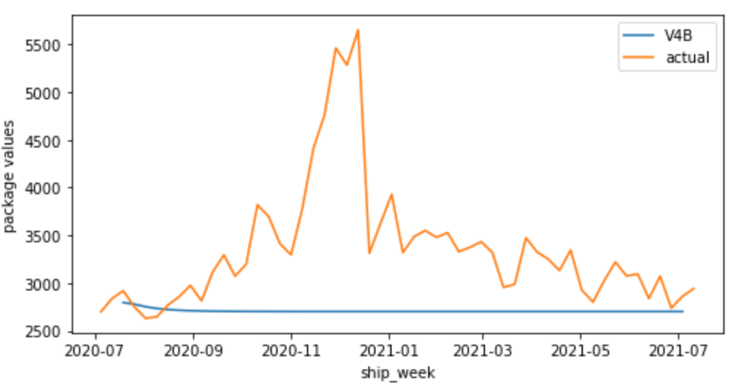
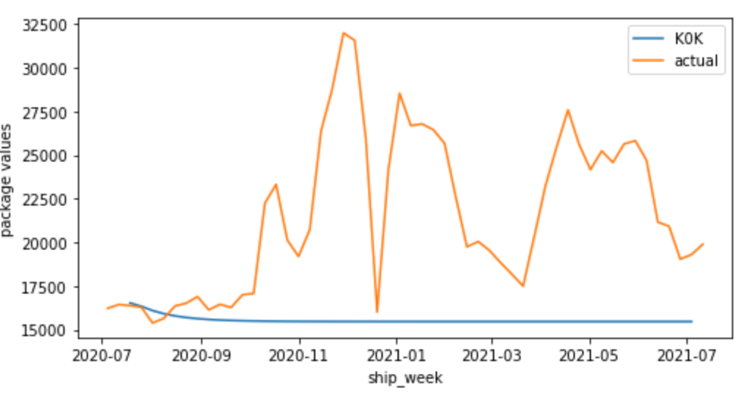
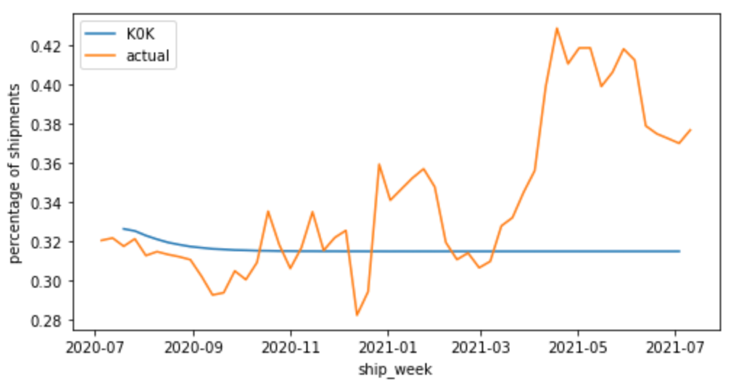


**Appendix D: *SARIMA Model (V2 Version) Versus Actual Values***

