Business Forecasting

Group 6
ANVAY GOTHIVAREKAR
MIHIR SHAH
TRAVIS MCDOUGAL
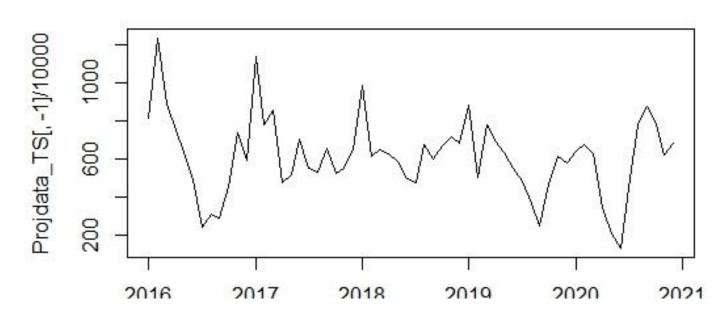
Agenda

Introduction

Methodology

Recommendation

Q&A

Introduction

The data set being used for the project is a monthly sales data of a supply chain organization starting from 2016 to 2020. In this project, we are forecasting the demand for a subset of hair care products that are sourced from a distribution center in Whittier, CA. We will be looking at the sales figure of last 5 years in order to determine the demand. The long term purpose of this forecast is to help the distribution leads to adequately and effectively plan labor for the upcoming months. The simulation data was provided by the marketing team in order to complete the case study. The data was used to produce accurate forecasts using the R software. The data was then converted into a Time series in order to generate a number of forecast models and determine which is the best model to implement to produce accurate forecast results for this project.
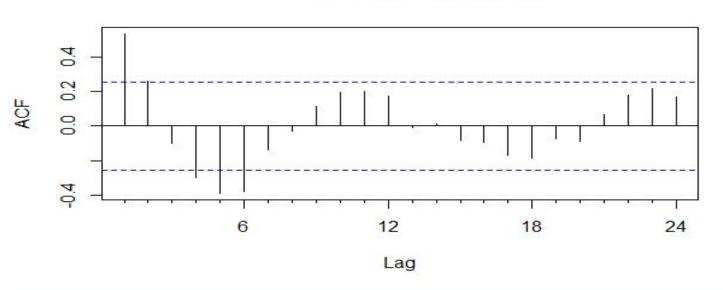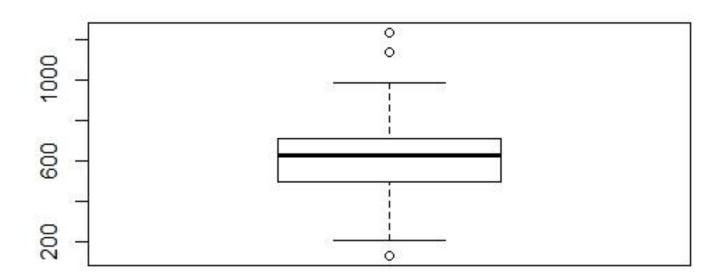
Introduction

- **Time Series Plot:**

- **ACF:**

Series Projdata_TS[, -1]

● **Boxplot for the Time Series:**

- **Time Series Summary:**

```
  Min. 1st Qu.   Median    Mean 3rd Qu.     Max.
 131.1   496.1    624.6   612.9   705.4   1235.0
```

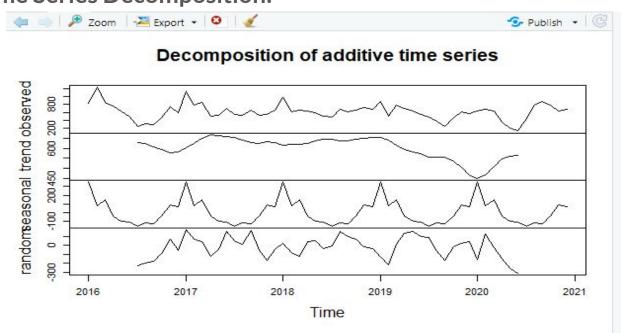- **Time Series Inference:**
  From the above time series plot, we can state that there has been an abrupt change at the start of the year 2018, due to which we can observe a downward trend moving forward.
  The Acf is a graph that is used to determine if a Time Series is dependent on its past observations. The above Acf plot shows that the most recent point of the observation is the most influential point in the time series.

- **Time Series Decomposition:**

Accuracy Metric

Determining the efficiency of forecast by using the **Mean Absolute Scaled Error (MASE)**

Why MASE?

MASE penalizes over and under forecasting.

MASE is suitable for data with trend/seasonal patterns

MASE can be used to compare forecast methods on a single series

Interpreting MASE?

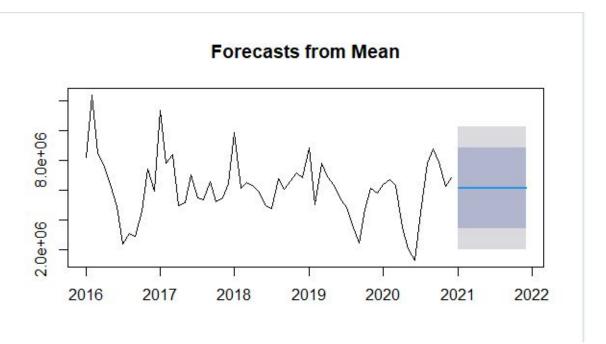A MASE score >1 needs a lot of work; the lower the better.

Methodology

In order to understand the data set better and provide accurate forecasts, we have performed a number of Forecasting and smoothing techniques which are as follows:
1. Mean Forecast.
2. Naive Forecast.
3. Seasonal Naive Forecast.
4. Random Walk Forecast.
5. ETS.
6. Holts-Winter.
7. Moving Averages.
8.ARIMA.
Let us have a look at each of the model forecast and accuracy to understand them better.

- **Mean Forecast:**

Forecasts from Mean

- **Output for Mean Forecast:**

```
          Point Forecast
Jan 2021        6122290
Feb 2021        6122290
Mar 2021        6122290
Apr 2021        6122290
May 2021        6122290
Jun 2021        6122290
Jul 2021        6122290
Aug 2021        6122290
Sep 2021        6122290
Oct 2021        6122290
Nov 2021        6122290
Dec 2021        6122290
```
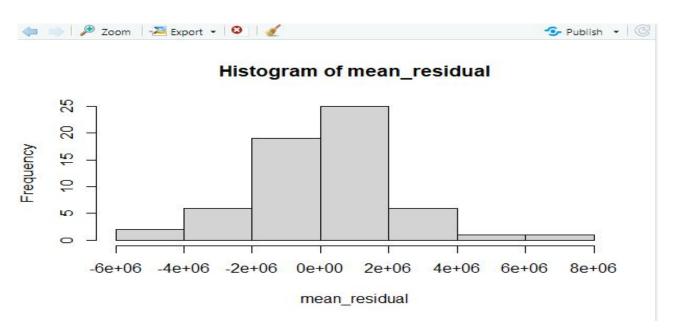
- **Accuracy for Mean Forecast:**

```
> accuracy(mean_fcst)
                      ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set -7.766706e-11  2033022  1493687  -17.43281  34.73084  0.8023893  0.532313
```

- **Mean Forecast Inference:**
  As the name suggests, Mean forecast returns the mean value of the data as the forecast for the future months. Mean forecast was performed in order to understand the data better to form further other models.
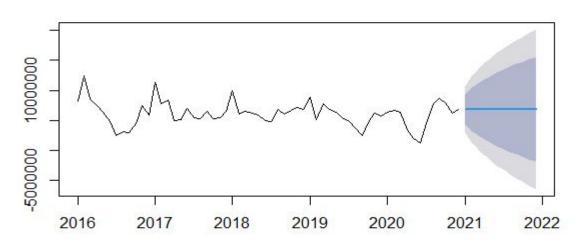
  MASE Score - 0.802

- **Histogram for Mean Residuals:**

Histogram of mean_residual

● **Naive Forecast:**

Forecasts from Naive method

- **Output for Naive Forecast:**

	Point Forecast
Jan 2021	6844852
Feb 2021	6844852
Mar 2021	6844852
Apr 2021	6844852
May 2021	6844852
Jun 2021	6844852
Jul 2021	6844852
Aug 2021	6844852
Sep 2021	6844852
Oct 2021	6844852
Nov 2021	6844852
Dec 2021	6844852

- **Accuracy for Naive Forecast:**

```
> accuracy(naive_fcst)
                     ME    RMSE     MAE       MPE    MAPE      MASE     ACF1
Training set -22286.89 1962749 1519964 -6.251836 27.16554 0.8165053 -0.24897
```
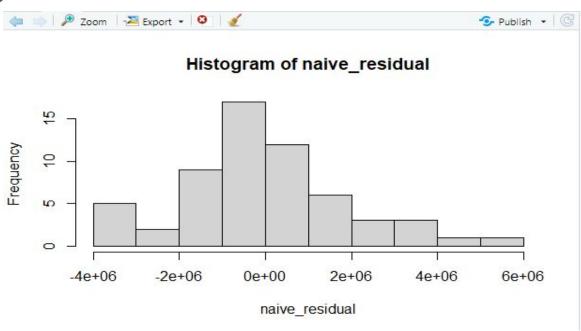
- **Naive Forecast Inference:**
  Naive method is mainly used as a benchmark to judge other models and when the process is fairly new. It is a model that uses the most recent data point to provide the future forecast. While forecasting, it is important to start with a Naive Model as it gives the forecaster an idea whether a more complex model is ideal or not for the forecasting in terms of performance.
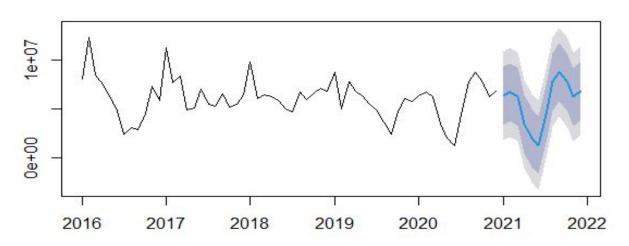
  MASE Score - 0.816

● **Histogram for Naive Residuals:**

- **Seasonal Naive Forecast:**

Forecasts from Seasonal naive method

- **Output for S-Naive:**

```
       Point Forecast
Jan 2021        6404143
Feb 2021        6727386
Mar 2021        6307139
Apr 2021        3487765
May 2021        2069570
Jun 2021        1311440
Jul 2021        4489529
Aug 2021        7776791
Sep 2021        8768251
Oct 2021        7882274
Nov 2021        6235568
Dec 2021        6844852
```
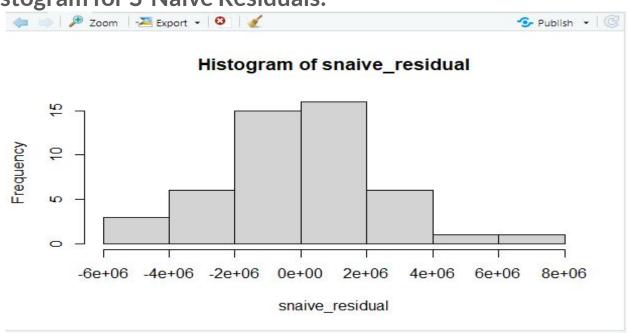
- **Accuracy for S-Naive Forecast:**

```
> accuracy(snaive_fcst)
                    ME      RMSE      MAE       MPE     MAPE MASE       ACF1
Training set -119252.9 2316953 1861549 -15.56629 40.08554    1 0.4372421
```
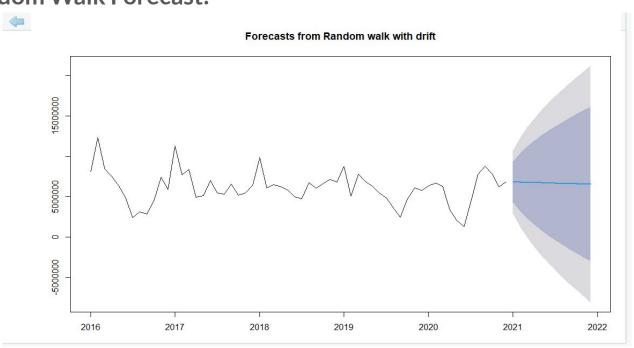
- **S-Naive Inference:**

  S-Naive is a seasonal naive model which is similar to the naive model to a lot of extent. However, in case of a S-Naive Model, we take the last seasonal observation instead of only the last observation. In the above S-Naive forecast plot also, we can see that the forecast is a repeating cycle of the last 12 months. S-Naive is basically an extension of the Naive method with some added complexity. When working on a data set that has very high seasonality, S-Naive is one of the efficient models to use.

  MASE Score - 1

- **Histogram for S-Naive Residuals:**

● Random Walk Forecast:

Forecasts from Random walk with drift

- **Output for RWF:**

```
          Point Forecast
Jan 2021        6822565
Feb 2021        6800278
Mar 2021        6777991
Apr 2021        6755704
May 2021        6733417
Jun 2021        6711130
Jul 2021        6688843
Aug 2021        6666556
Sep 2021        6644270
Oct 2021        6621983
Nov 2021        6599696
Dec 2021        6577409
>
```
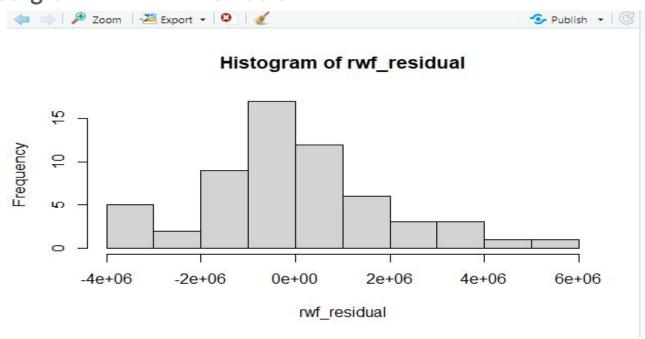
- **Accuracy for Random Walk Forecast:**

```
> accuracy(rwf_fcst)
                        ME     RMSE     MAE       MPE     MAPE      MASE      ACF1
Training set 2.051913e-10 1962622 1517320 -5.821731 27.03888 0.8150849 -0.24897
```
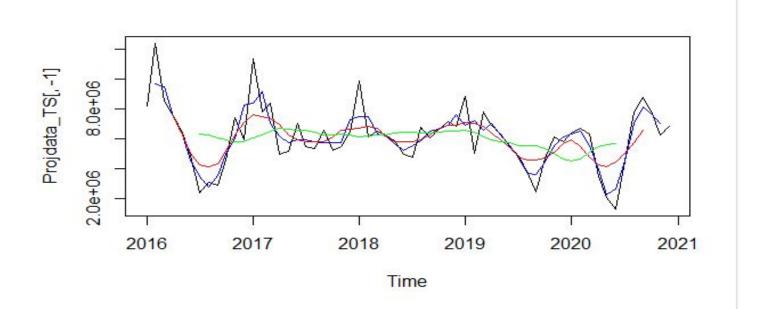
- **RWF Inference:**

  A Random Walk Forecast is similar to the Naive Model apart from the fact that it uses a drift model to return the forecast values. A RWF without drift model is a simple Naive model. A RWF model allows the forecast to increase or decrease over time based on the drift which is the average change seen in the data set.

  MASE Score - 0.815

- **Histogram for RWF Residuals:**
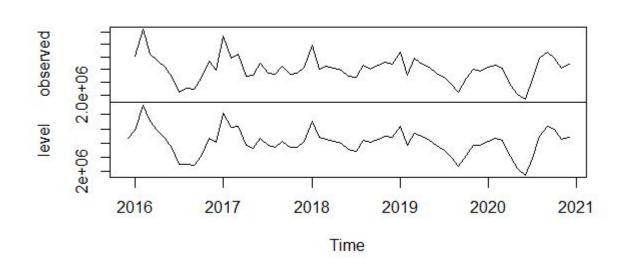
- **Moving Averages:**

- **Moving Averages Inference:**

  Moving Averages is mainly used when the most recent observations in a data set are more relevant and influential than the entire data set. It is upto the forecasters discretion to determine the order of the moving averages depending on the data set. A small order indicates more weightage to the most recent points. A Moving Average method prominently captures the main movements of the Time series. A small order Moving Average is used when the data has sudden shifts at irregular intervals. Similarly, a large order Moving Average is used when there are wide infrequent fluctuations. A large order MA will generally smoothen the Time series curve.
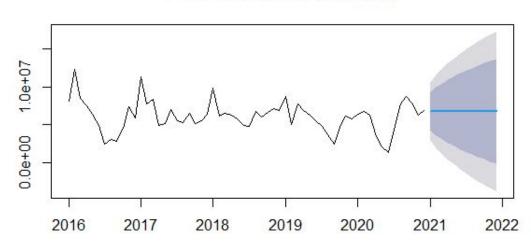
- **ETS Forecast:**

Decomposition by ETS(A,N,N) method

- **ETS Forecast for 12 Months:**

Forecasts from ETS(A,N,N)

- **Output for ETS:**

```
          Point Forecast
Jan 2021       6795209
Feb 2021       6795209
Mar 2021       6795209
Apr 2021       6795209
May 2021       6795209
Jun 2021       6795209
Jul 2021       6795209
Aug 2021       6795209
Sep 2021       6795209
Oct 2021       6795209
Nov 2021       6795209
Dec 2021       6795209
```

- **Accuracy for ETS Forecast:**

```
                  ME      RMSE      MAE        MPE      MAPE       MASE           ACF1
Training set 2457.991  1911109  1482103  -7.445058  27.39355  0.7961668  0.0001901067
```
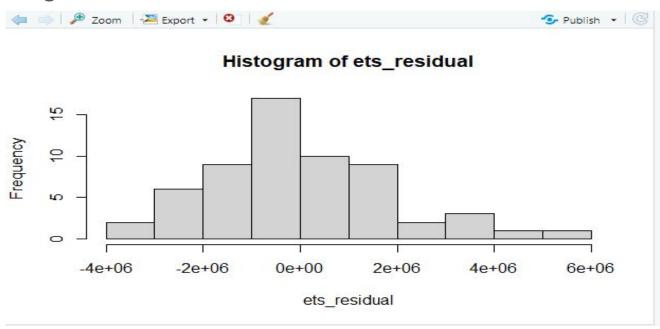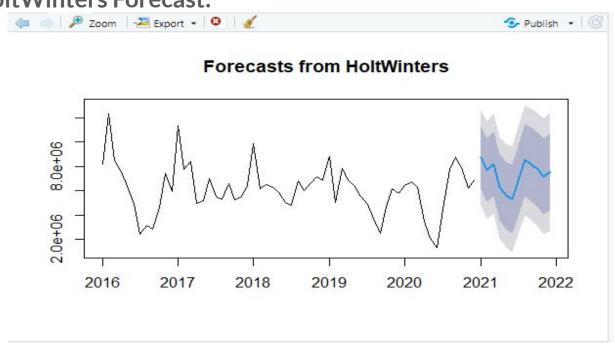
- **ETS Inference:**
  The ETS method focuses mainly on the Trend and Seasonal components of a Time series. A simple exponential smoothing ETS model is a (A,N,N) model with additive errors, No Trend and No Seasonality. If a particular model for ETS is not specified, a model which is best suited is automatically selected.

  MASE Score - 0.796

● **Histogram for ETS Residuals:**
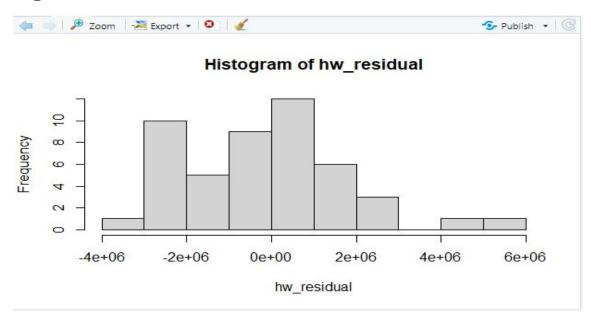
- **HoltWinters Forecast:**

- **Output for HoltWinters:**

	Point Forecast
Jan 2021	8717258
Feb 2021	7711330
Mar 2021	8220726
Apr 2021	6272805
May 2021	5555071
Jun 2021	5313043
Jul 2021	7035332
Aug 2021	8517464
Sep 2021	8185970
Oct 2021	7762297
Nov 2021	7163629
Dec 2021	7548539

```
> hw_fcst_seas
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = Projdata_TS[, -1])

Smoothing parameters:
 alpha: 0.2199965
 beta : 0
 gamma: 0.7043629
```

● **Histogram for HoltWinters Residuals:**

- **Accuracy for HoltWinters:**

```
> accuracy(hw_forecast_all)
                  ME      RMSE      MAE       MPE      MAPE      MASE      ACF1
Training set -118274.4 1974304  1517374  -10.19374  31.44212  0.8151136  0.461919
```

- **HoltWinters Inference:**
  HoltWinters is a triple exponential smoothing for the Level, Trend and Seasonal components in a Time Series. It has three smoothing parameters namely alpha, beta and gamma. There are two methods for a HoltWinters model which is an additive method or a multiplicative method.
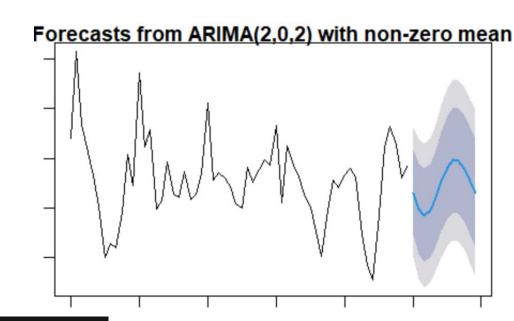
  MASE Score - 0.815

Box-Jenkins Model

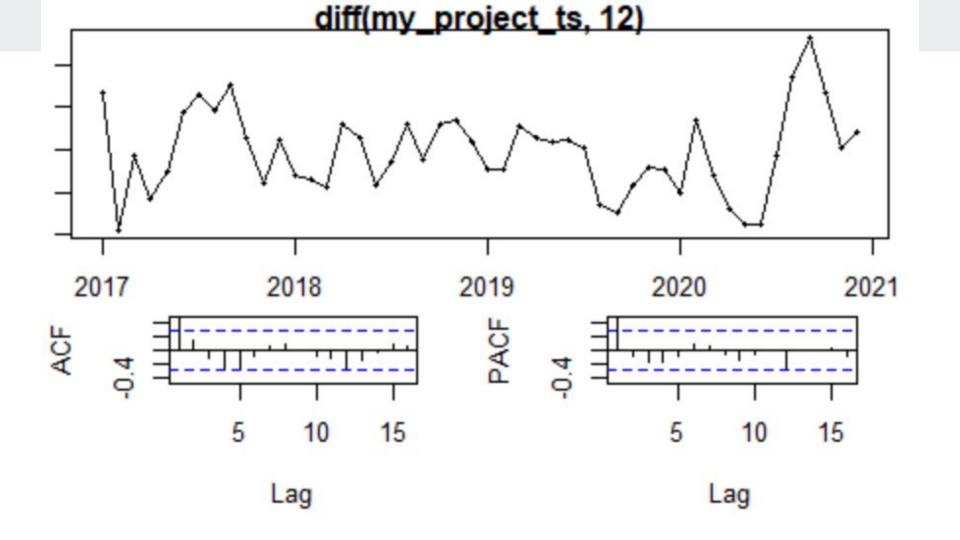ARIMA makes use of lagged moving averages to smooth time series data.

Process

(1)   Observe TS
(2)   Examin ACF & PACF for time lags
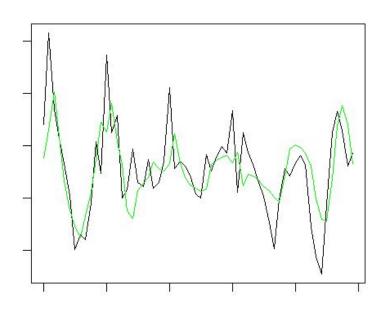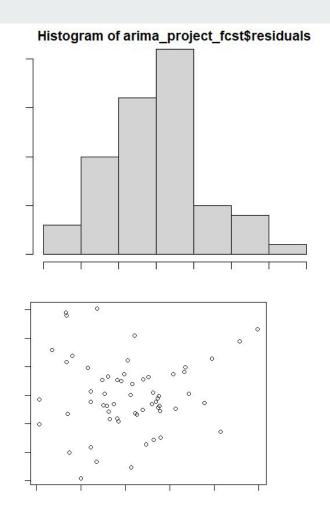(3)   Match pattern to determine best ARIMA model

MASE Score - 0.536

Forecasts from ARIMA(2,0,2) with non-zero mean

```
> accuracy(arima_project_fcst)
                     ME      RMSE      MAE       MPE
Training set  80953.05  1057897  839476.9  -4.616539
```

```
    MAPE       MASE         ACF1
19.09295  0.5363331  -0.01901179
```

diff(my_project_ts, 12)

Box-Jenkins Model

Histogram of arima_project_fcst$residuals

Recap

The Task

Help the Operations team make labor planning decisions for the upcoming year for a distribution center.

The DC in addition to picking, packing, and shipping, the DC also manufactures goods.

Based on the best model we can provide a guidance to the teams teams based on the forecasted demand.
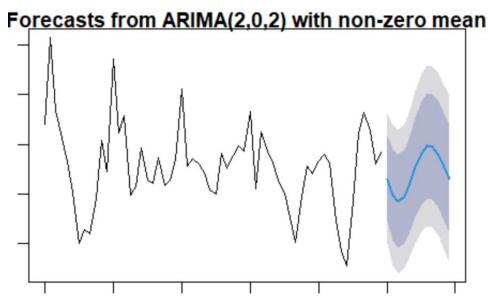
● **Accuracy Scores**

Model	MASE Score
Naive	.82
SNaive	1.00
ARIMA	**.53**
Holtwinters	.81
RWF	.82
Mean	.80
ETS	.79

Recommendation

Make Labor decisions based on the Arima model

Knowing that ARIMA model is intended for short term forecasts it is recommended that the is run on monthly basis

Additional inference about cyclical tendencies: January is seasonally high from a demand.

Forecasts from ARIMA(2,0,2) with non-zero mean

Q&A

Any further suggestions and feedback can also be emailed below:

Anvay Gothivarekar - ag1900@rutgers.edu
Mihir Shah - ms3073@rutgers.edu
Travis McDougal - travis.mcdougal@rutgers.edu