Group members
20192029 Bayarmaa Tumur-Ochir
20192010 Mubarrat Tajoar Chowdhury
20192024 MD Khalequzzaman Chowdhury Sayem

Times series data
https://www.kaggle.com/c/store-sales-time-series-forecasting/overview

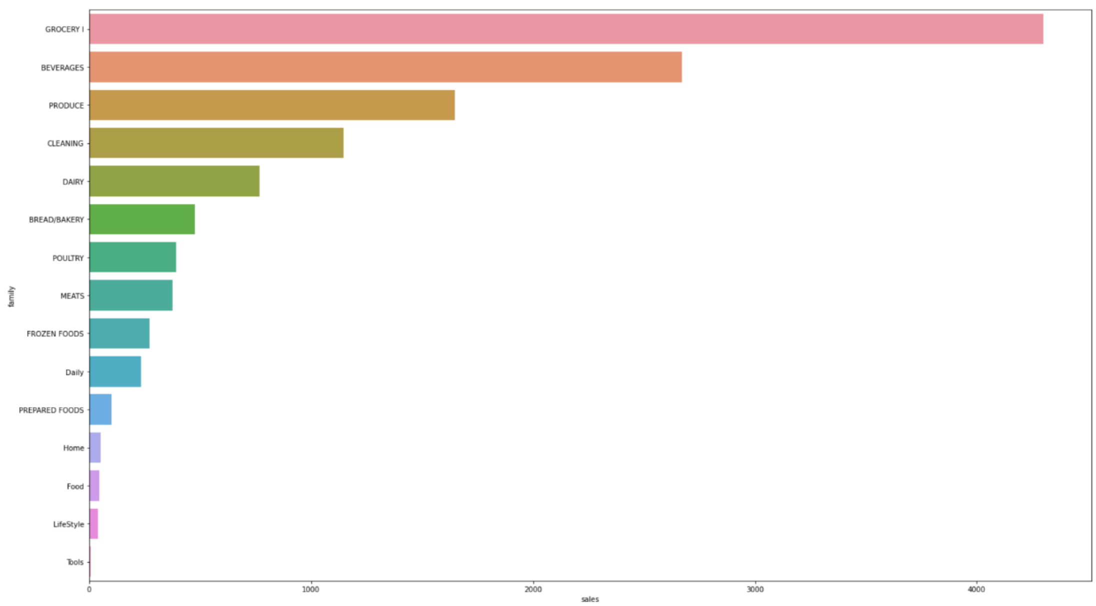
Data description
The data we are using in this project is the store sales on data from Corporación Favorita, a large Ecuadorian-based grocery retailer. The data includes dates, store and product information, whether that item was being promoted or not, as well as the sales numbers. Duration Covered: 2013 to 2017

Problem
Forecasting product demand is very important for brick-and-mortar grocery store businesses which have to decide how much inventory to buy to meet its demand on a monthly basis. If prediction is an overestimate, the warehouse is overstocked, not to mention about the holding costs for durable products and potential profit loss for perishable food as well. If prediction is an underestimate, there is revenue loss for failed selling opportunity. With the help of accurate forecasting, retailers can satisfy their customers while reaching their potential revenue.

We consider traditional time-series models, such linear/seasonal regression, ARIMA models, and then consider Deep Learning based models to fit into the sales data. And then finally, a model that fit the training data best is chosen for prediction.

Exploratory analysis

The sales record includes sales volume information on 33 family of products across all stores in Ecuador. Below is barplot for sales volume of all products across all stores.
Data upto 2016 was chosen as training set and data for the year 2017 as test set.



Since multiple models in both “Traditional” and “AI” class of models had to be considered before reaching a conclusion, only one from thirty-three family of products were chosen as a vast number of time series models in both categories would have had to be considered.
“Grocery I” sales data were taken, as this product is the most sold .

Fig: Beverage Sales From 2013-2016

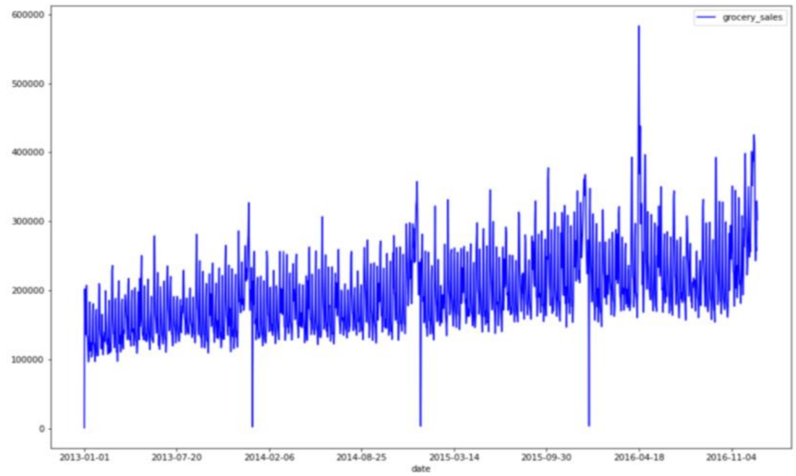


Fig: Grocery Sales From 2013-2016

.

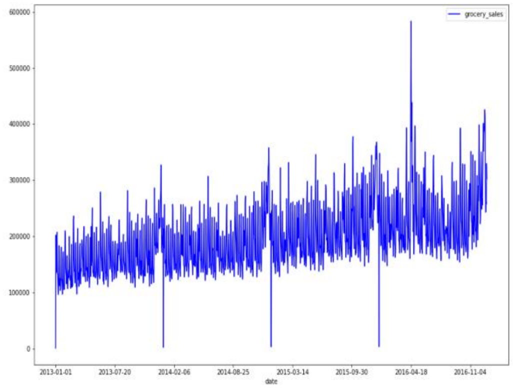
This graph as well most of the other product graphs has sharp downward spikes at around 1 year interval. This hints at presence of seasonality,

A magnitude 7.8 earthquake struck Ecuador on April 16, 2016. People rallied in relief efforts donating water and other first need products which greatly affected supermarket sales for several weeks after the earthquake. So, the graphs have upward sharp upward spikes during those days of heightened sale.

Also in Ecuador, wages are paid at 15th of every month and at the last day of the month. This too is expected to impact sales pattern

Section 2: Traditional Time Series models

Starting off with the grocery sales data. Gradual upward trend hints at suitability of a Linear-Seasonal hybrid deterministic model



To decide which direction to head towards to decide on a final model, some basic parameters are used to judge the model:

R-squared value: Gives the proportion of variance in the time series data the model accounts for. The higher the better. Value close to 0.4 are considered bad, whereas value close to 0.7 is considered to indicate to good fitting model.
AIC: This quantity considers maximum likelihood of the function and number of parameters in the model. This quantity penalizes having too many parameters in the model and rewards for having a high likelihood estimate indicating that the model is true. The lower the better. Lower value rewards simple models that have high likelihood
BIC: Penalizes and rewards for same reasons as AIC but also considers the size of the sample used. Lower sample number is penalized. The lower the value for BIC the better.
MSE: Sums the difference between prediction and actual data, squares them and averages all the errors for all the predictions. The lower the value, the better the fit.
Augmented Dickey Fuller Test : To explain briefly, this is a hypothesis test where if the required critical value indicating a high significance level is exceed but the test statistic, it is concluded that the series of data being tested is stationary.

To start off simple, a general linear regression was performed based on current time:

Model: $y_t = \beta_0 + \beta_1$

```
mean_squared_error: 2845330211.047445
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.196
Model:                            OLS   Adj. R-squared:                  0.196
Method:                 Least Squares   F-statistic:                     266.8
Date:                Sun, 19 Dec 2021   Prob (F-statistic):           8.17e-54
Time:                        07:51:34   Log-Likelihood:                -13448.
No. Observations:                1093   AIC:                         2.690e+04
Df Residuals:                    1091   BIC:                         2.691e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.324e+05   4934.504     26.837      0.000    1.23e+05    1.42e+05
day           83.5967      5.118     16.333      0.000      73.554      93.639
==============================================================================
```

With an MSE of 2.8 billion and a very large AIC, BIC score and and R-square value of just 0.196, the model is expectedly bad as it’s too simplistic.

Model 2:

Based on the observed spikes at around a year interval and the expectation that 15-day wage pay interval will affect sales, a Linear-Seasonal Hybrid deterministic model is implemented , accounting for both seasonality frequencies.

Yearly seasonality value was chosen to be 364 instead of 365 or 366 because those values showed a massive drop is R-Squared value to below 0.1.

ols for linear-seasonal model for grocery:

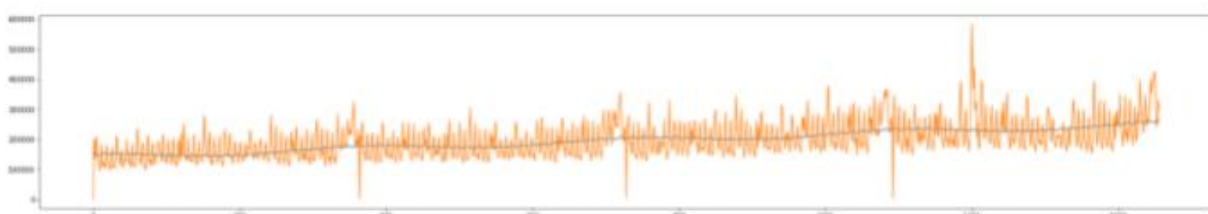
$$y_t = \beta_0 + \beta_1 t + \beta_3 cos(2\pi f_1 t) + \beta_4 sin(2\pi f_1 t) + \beta_5 cos(2\pi f_2 t)$$
$$+ \beta_6 sin(2\pi f_2 t)$$

set f_1 = 15days to account for bi-weekly wage in-flow

set f_2 = 364 days to account for yearly seasonality

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.294
Model:                            OLS   Adj. R-squared:                  0.292
Method:                 Least Squares   F-statistic:                     120.9
Date:                Sun, 19 Dec 2021   Prob (F-statistic):          4.14e-107
Time:                        07:51:34   Log-Likelihood:                -17828.
No. Observations:                1457   AIC:                         3.567e+04
Df Residuals:                    1451   BIC:                         3.570e+04
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.417e+05   2658.300     53.318      0.000    1.37e+05    1.47e+05
day           75.1986      3.174     23.694      0.000      68.973      81.424
cos1        1333.0734   1850.818      0.720      0.471   -2297.493    4963.639
sin1        2426.5249   1851.549      1.311      0.190   -1205.473    6058.523
cos2        8998.5808   1855.727      4.849      0.000    5358.386    1.26e+04
sin2         478.0746   1883.180      0.254      0.800   -3215.971    4172.121
==============================================================================
```

Again a poor R-squared value and huge AIC BIC scores are seen. However ,MSE score decreased by over 300 million so the model was pursued. The plot showing green prediction line again orange actual data also clear shows the poor fit.



```
mean_squared_error: 2486115295.642784
```

Model 3:

ols for linear-seasonal model for grocery:

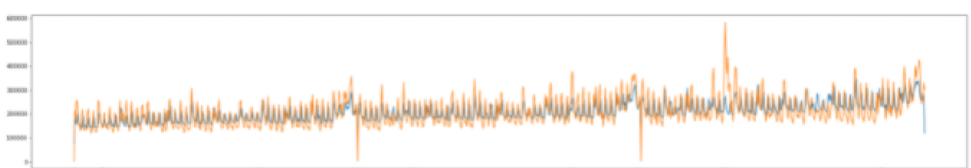
$$y_t = \beta_0 + \beta_1 t + \beta_3 cos(2\pi f_1 t) + \beta_4 sin(2\pi f_1 t) + \beta_5 y_{t-364}$$

set f = 15days to account for bi-weekly wage in-flow

shift tried 365,366 but ended up on 364

The seasonality expression for annuity was replaced with an AR term with a lag of 364 and the result were as follows:

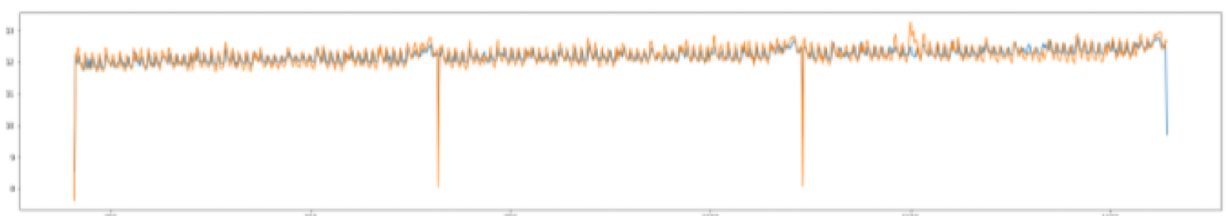
```
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.421
Model:                            OLS   Adj. R-squared:                  0.419
Method:                 Least Squares   F-statistic:                     197.7
Date:                Sun, 19 Dec 2021   Prob (F-statistic):          2.01e-127
Time:                        07:51:35   Log-Likelihood:                -13269.
No. Observations:                1093   AIC:                         2.655e+04
Df Residuals:                    1088   BIC:                         2.657e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       6.224e+04   5419.175     11.486      0.000    5.16e+04    7.29e+04
day           36.4318      4.922      7.402      0.000      26.774      46.089
cos         1994.8548   1940.635      1.028      0.304   -1812.956    5802.666
sin         1034.5084   1943.489      0.532      0.595   -2778.901    4847.918
y365           0.6176      0.030     20.476      0.000       0.558       0.677
==============================================================================
Omnibus:                      220.372   Durbin-Watson:                   1.237
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              572.906
Skew:                           1.052   Prob(JB):                    3.94e-125
Kurtosis:                       5.855   Cond. No.                     7.51e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctl
y specified.
[2] The condition number is large, 7.51e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

<Figure size 2160x360 with 0 Axes>
```



```
mean_squared_error: 2050657475.1886632
```

Significant drop in MSE is observed at 2 billion. Around 4480 million drop. R squared is much better at above 40%, however AIC BIC values are still huge despite a significant drop. Significant changes in model coefficient is observed with one less Beta coefficient to be calculated. While model could still be improve the switch to AR process leading to positive changes in the metrics motivated the replacement of the remaining trigonometric expression with another AR term instead. Before that was implemented, it was observed that the prediction plots are more closely following the trend but falling short in terms of height, as the green prediction line above shows, Hence, to reduce the gap, log transformation was applied to the actual data and the same model was retrained. The results are:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.573
Model:                            OLS   Adj. R-squared:                  0.572
Method:                 Least Squares   F-statistic:                     487.8
Date:                Sun, 19 Dec 2021   Prob (F-statistic):          7.77e-201
Time:                        07:51:35   Log-Likelihood:                 78.773
No. Observations:                1093   AIC:                            -149.5
Df Residuals:                    1089   BIC:                            -129.6
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7829      0.303     15.811      0.000       4.189       5.376
day            0.0002   2.38e-05      6.680      0.000       0.000       0.000
y15           -0.0427      0.022     -1.925      0.055      -0.086       0.001
y365           0.6459      0.020     32.373      0.000       0.607       0.685
==============================================================================
Omnibus:                      612.784   Durbin-Watson:                   1.314
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            34754.659
Skew:                           1.810   Prob(JB):                         0.00
Kurtosis:                      30.387   Cond. No.                     4.28e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctl
y specified.
[2] The condition number is large, 4.28e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

<Figure size 2160x360 with 0 Axes>
```



```
mean_squared_error: 0.05069042601777997
```

Drastic improvement in MSE but it’s just because of the scaling. R-squared is now at a decent level of 0.572, AIC BIC’s dropped to negative indicating a better likelihood

function. The plot also looks significantly better, with the gaps reduced. The standard errors of the coefficients are lower.
Next, the remaining trigonometric expressions are replaced as expressed earlier.

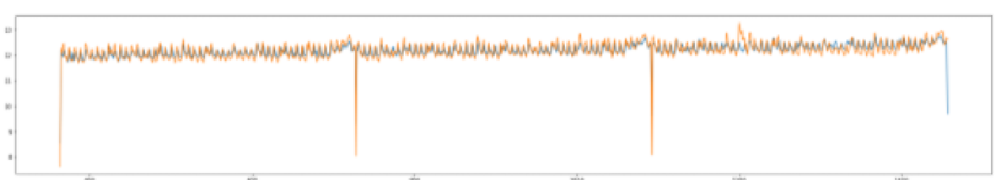
Model 4

ols for linear-seasonal model for grocery:
$y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-15} + \beta_3 y_{t-364}$

shift 15days to account for bi-weekly wage in-flow

shift tried 365,366 but ended up on 364

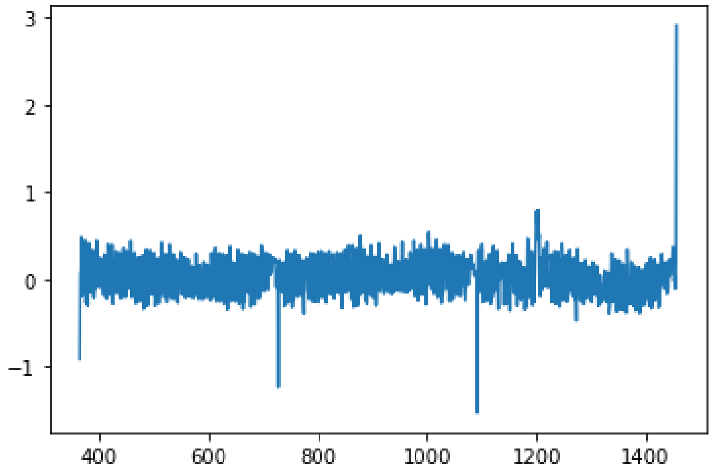
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.573
Model:                            OLS   Adj. R-squared:                  0.572
Method:                 Least Squares   F-statistic:                     487.8
Date:                Sun, 19 Dec 2021   Prob (F-statistic):          7.77e-201
Time:                        07:51:35   Log-Likelihood:                 78.773
No. Observations:                1093   AIC:                            -149.5
Df Residuals:                    1089   BIC:                            -129.6
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7829      0.303     15.811      0.000       4.189       5.376
day            0.0002   2.38e-05      6.680      0.000       0.000       0.000
y15           -0.0427      0.022     -1.925      0.055      -0.086       0.001
y365           0.6459      0.020     32.373      0.000       0.607       0.685
==============================================================================
Omnibus:                      612.784   Durbin-Watson:                   1.314
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            34754.659
Skew:                           1.810   Prob(JB):                         0.00
Kurtosis:                      30.387   Cond. No.                     4.28e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctl
y specified.
[2] The condition number is large, 4.28e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

<Figure size 2160x360 with 0 Axes>
```



```
mean_squared_error: 0.05069042601777997
```

This improvement did not lead significant improvement in any of the metrics. Even the unchanged beta coefficients are similar. Surprisingly, lag-15 AR term has a negative coefficient in both of the last models despite last payday 15 days ago or earlier. Fresh cash in-flow was expected to boost sales. In this hybrid class model dealt thus far, this is by far the best for it having best fit and having fewer parameters than the last model which offered similar fit.

So, for being the best thus far, the residual analysis was done. Below is plot for residuals followed by the result of Dickey-Fuller test to see if the residual is stationary. This is to confirm iid property of residuals, which would indicate the model being a very good fit.



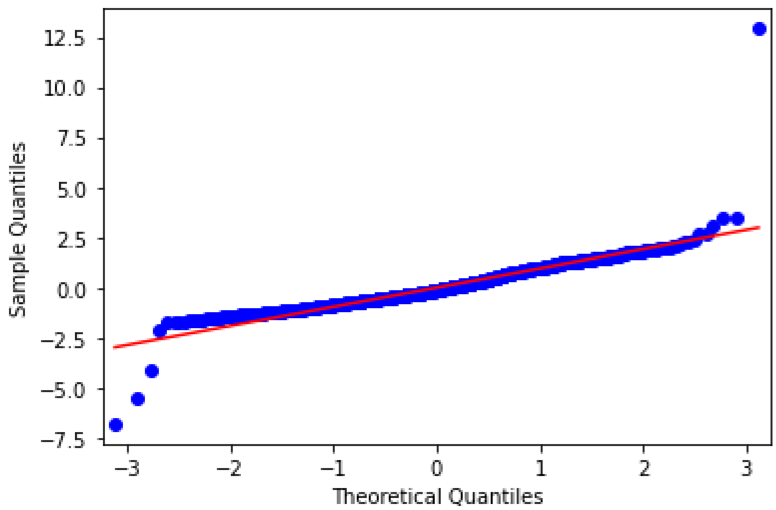

```
AUGMENTED DICKEY-FULLER TEST for: residuals of SARIMAX(0, 1, 2)x(0, 1, 2, 7) fo
r bread sales
Test statistic = -3.072
P-value = 0.029
Critical values :
        1%: -3.4364762217105844 - The data is not stationary with 99% confidenc
e
        5%: -2.8642449457774135 - The data is  stationary with 95% confidence
        10%: -2.5682102104987337 - The data is  stationary with 90% confidence
```

Residual plot shows fairly regular variations except for at the outlier earthquake-induced spikes and seasonal spikes.

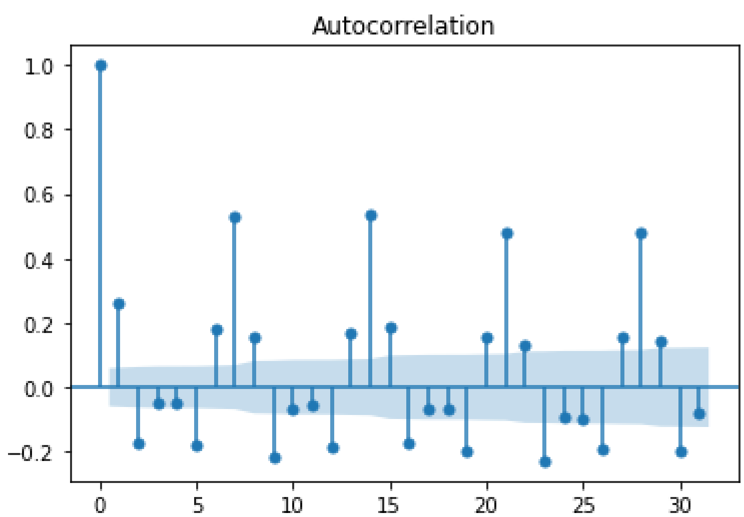
The dickey fuller test confirms this with the stationarity confirmation test failing only at 1% significance level.

However, as the plots show, the model fails to capture some all the seasonality within the data. Given the need for a better seasonal model, and considering the improvement that this model has brought.

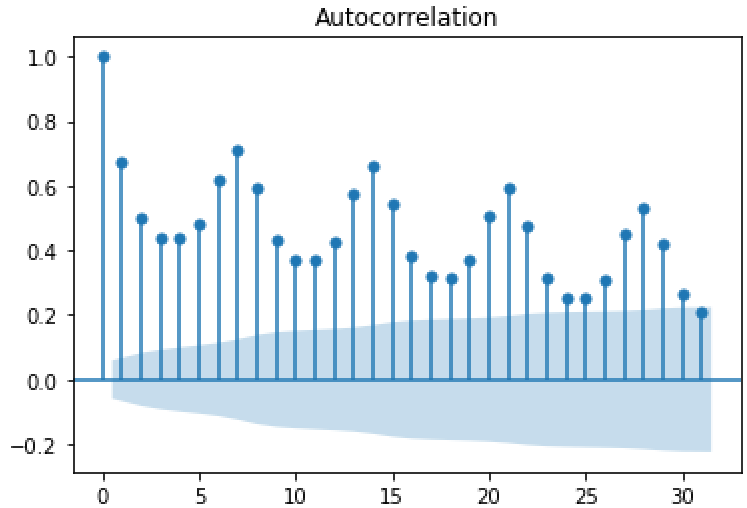
The qq-plot below is decent with slight curvature towards the middle and leaving the linear track only at the far edges. The residuals could have been considered to be normally distributed with iid properties, had it not been for the Autocorrelation plot showing otherwise.

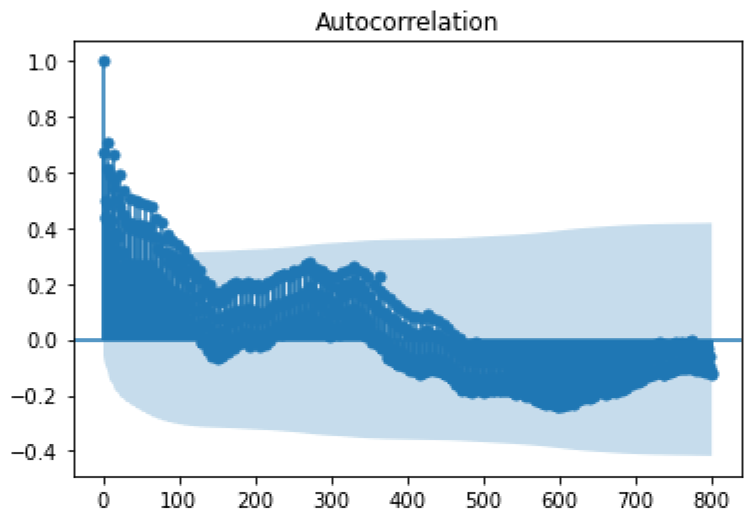


Below is the ACF plot of residuals which showsstrong evidence of presence of seasonality within the residuals. There are large non-zero correlation spikes at regular interval centering other, shorter regularly spaced spikes with significant non-zero values, indicating such presence in the original data that the model could not eliminate. Besides the ACF plot, even the past prediction plot clearly shows the model's failure to capture annual seasonal spikes. So, the next model considered is the Seasonal Arima model



Model 6: Seasonal Arima.





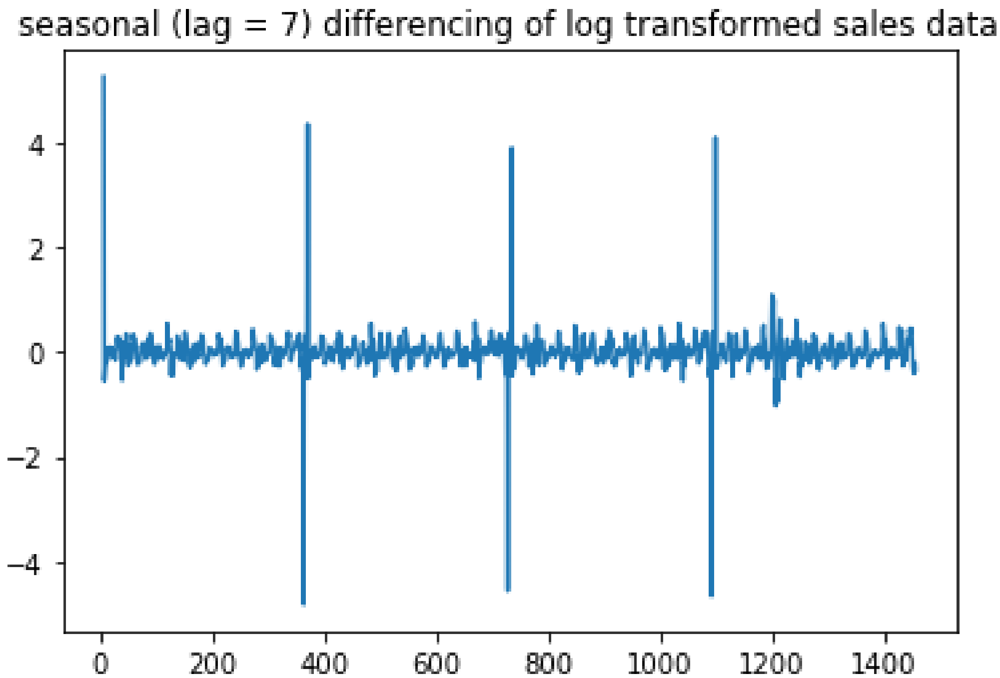
The first acf plot of the sales data shows regularly a dominant correlation spike at lag 1 and then at every lag 7 seven terms, centering around shorter, yet significant spikes. However, the spike for annuity could not be captured even with a plot for broader lag range. This could perhaps be because of the fact that, while the spikes occur at a certain point during the year, perhaps because of some holiday, autocorrelation for lag of 364 does not occur in other dates in the year.

In an attempt to determine appropriate SARIMA lag values, differencing was performed for various combinations of seasonal differencing lags combined with or without general first differencing at lag 1. This was also done to make the data stationary. The values considered are, combinations of lag 1 , 7 (suggested by acf plot), 15(suggested by wage payment interval), 364(based on annual spikes).

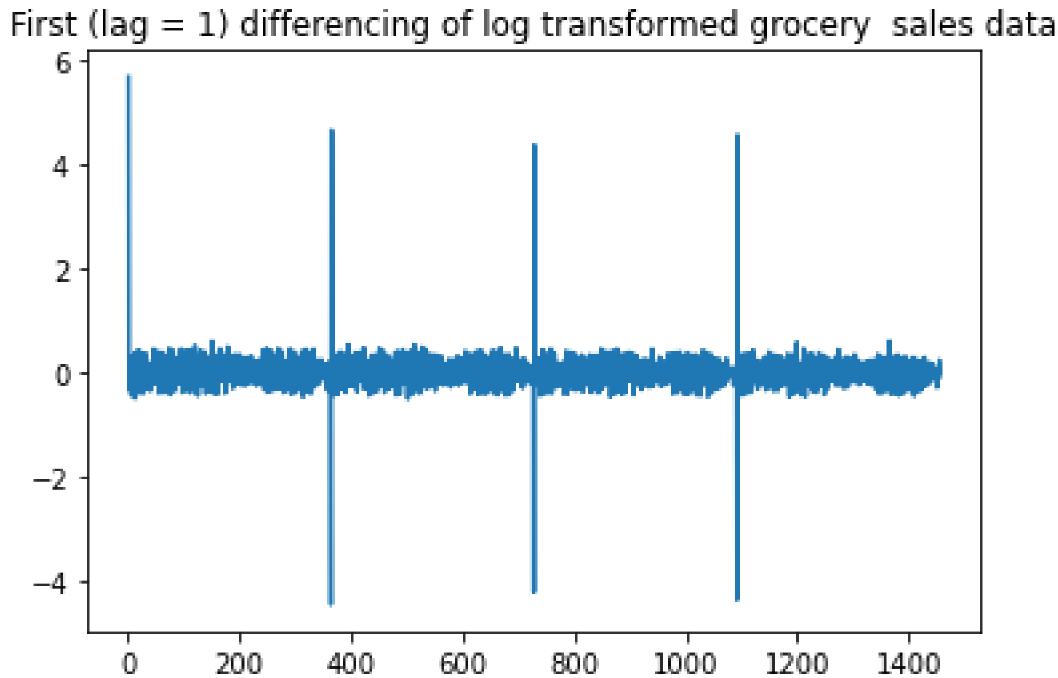
The plots for all the differencing combinations followed by their dickey-fuller test results (to judge success of differencing intended to make data stationary) are shown in order:


```
AUGMENTED DICKEY-FULLER TEST for: original beverages sales data
Test statistic  = -2.968
P-value = 0.038
Critical values :
        1%: -3.434921564946909 - The data is not stationary with 99% confidenc
e
        5%: -2.8635590328954197 - The data is  stationary with 95% confidence
        10%: -2.567844919904611 - The data is  stationary with 90% confidence
```


```
AUGMENTED DICKEY-FULLER TEST for: log transformed grocery  sales data
Test statistic  = -2.844
P-value = 0.052
Critical values :
        1%: -3.434918371231736 - The data is not stationary with 99% confidenc
e
        5%: -2.8635576234668982 - The data is not stationary with 95% confiden
ce
        10%: -2.5678441693558898 - The data is  stationary with 90% confidence
```

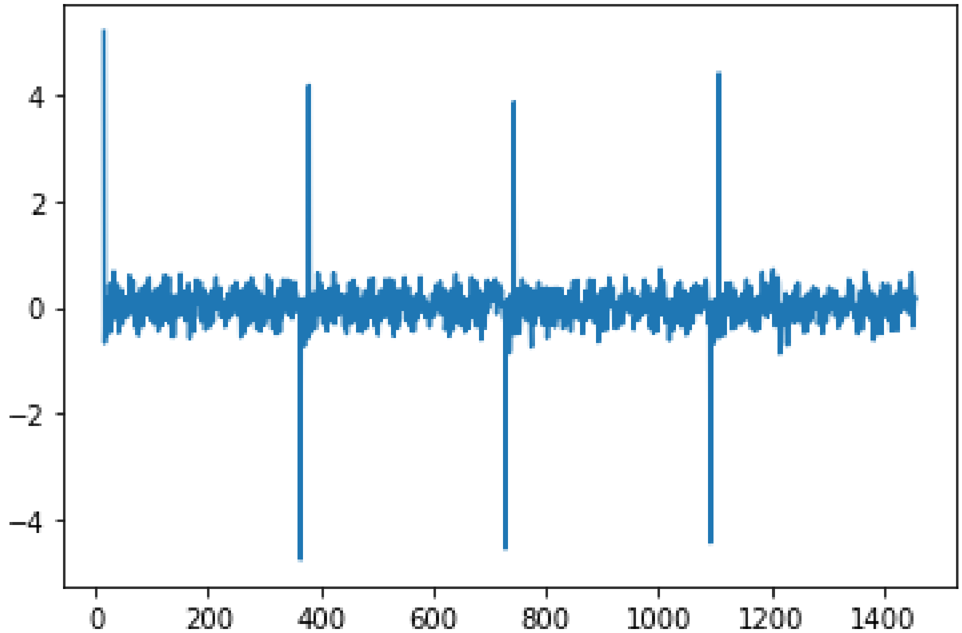



```
AUGMENTED DICKEY-FULLER TEST for: seasonal (lag = 7) differencing of log tran
sformed sales data
Test statistic = -13.715
P-value = 0.000
Critical values :
        1%: -3.4349343845870006 - The data is  stationary with 99% confidence
        5%: -2.8635646903561653 - The data is  stationary with 95% confidence
        10%: -2.5678479326174157 - The data is  stationary with 90% confidence
```

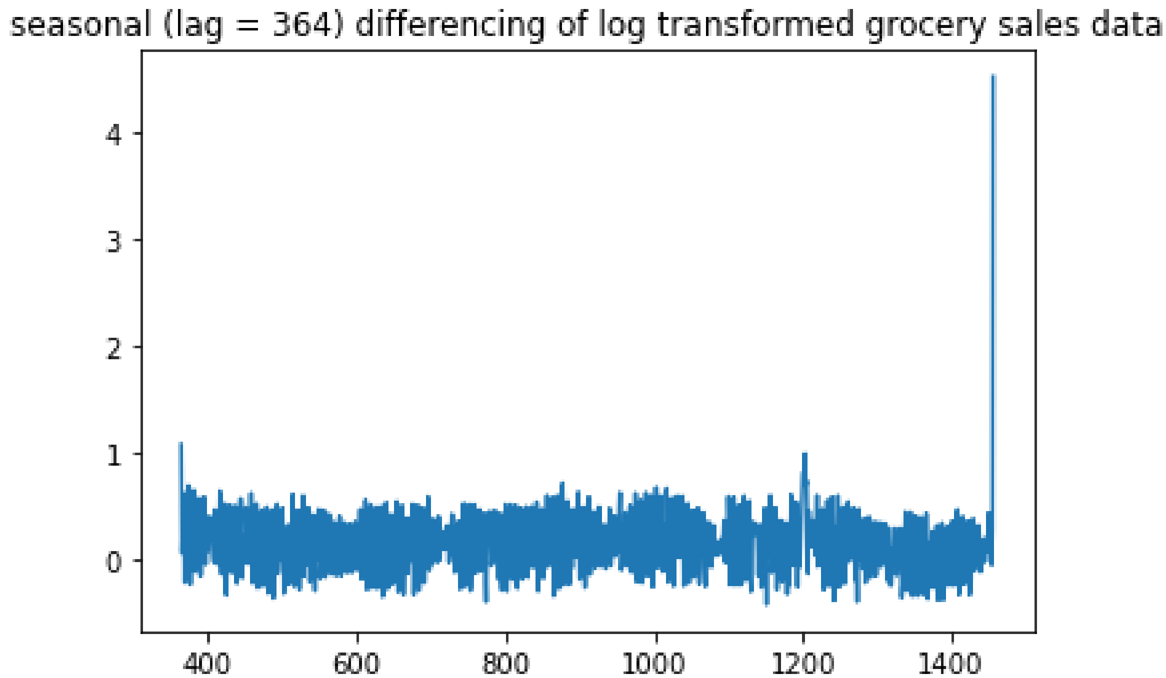



```
AUGMENTED DICKEY-FULLER TEST for: First (lag = 1) differencing of log transfo
rmed grocery  sales data
Test statistic = -12.844
P-value = 0.000
Critical values :
        1%: -3.434918371231736 - The data is  stationary with 99% confidence
        5%: -2.8635576234668982 - The data is  stationary with 95% confidence
        10%: -2.5678441693558898 - The data is  stationary with 90% confidence
```

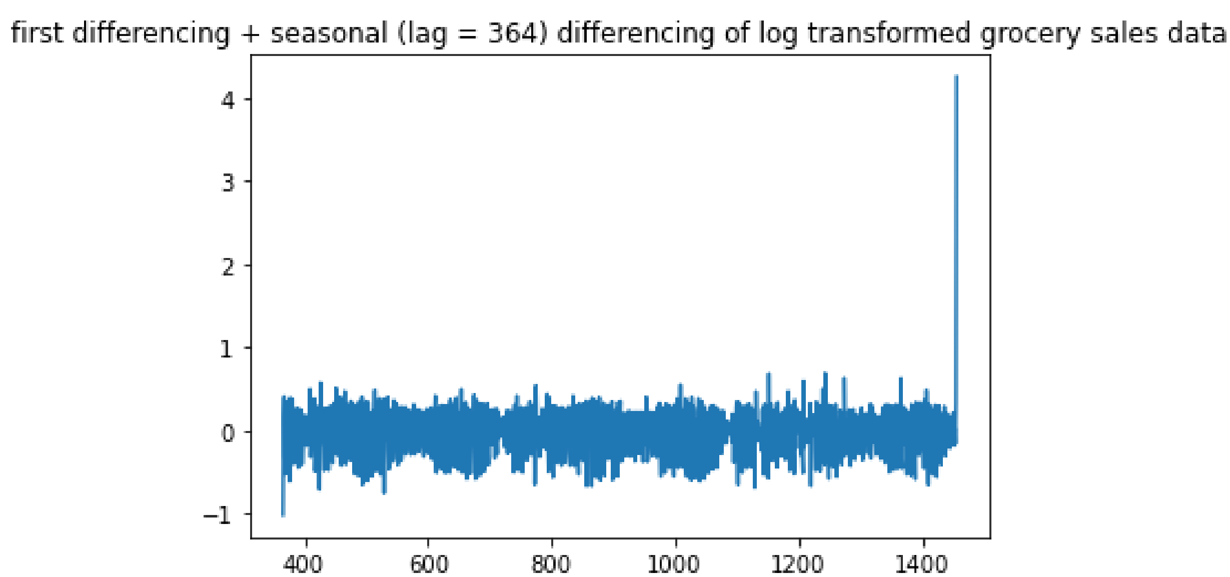
seasonal (lag = 15) differencing of log transformed grocery sales data




```
AUGMENTED DICKEY-FULLER TEST for: seasonal (lag = 15) differencing of log tra
nsformed grocery  sales data
Test statistic = -9.633
P-value = 0.000
Critical values :
        1%: -3.434973278556316 - The data is  stationary with 99% confidence
        5%: -2.8635818545830376 - The data is  stationary with 95% confidence
        10%: -2.567857072940785 - The data is  stationary with 90% confidence
```

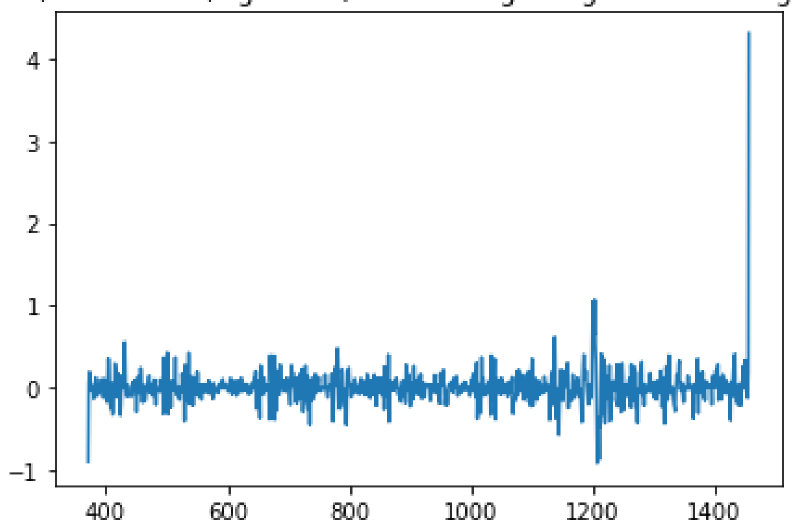



```
AUGMENTED DICKEY-FULLER TEST for: seasonal (lag = 364) differencing of log tr
ansformed grocery sales data
Test statistic = -2.993
P-value = 0.036
Critical values :
        1%: -3.436425000208065 - The data is not stationary with 99% confidenc
e
        5%: -2.864222352544219 - The data is  stationary with 95% confidence
        10%: -2.5681981773275466 - The data is  stationary with 90% confidence
```

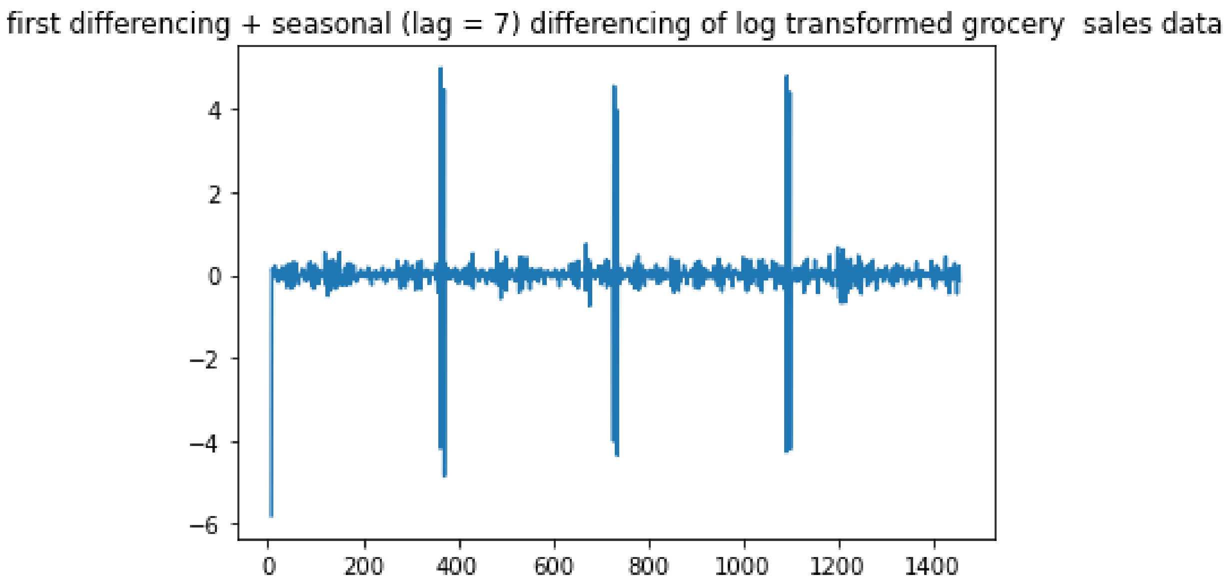


```
AUGMENTED DICKEY-FULLER TEST for: first differencing + seasonal (lag = 364) d
ifferencing of log transformed grocery sales data
Test statistic = -11.941
P-value = 0.000
Critical values :
	1%: -3.436425000208065 - The data is  stationary with 99% confidence
	5%: -2.864222352544219 - The data is  stationary with 95% confidence
	10%: -2.5681981773275466 - The data is  stationary with 90% confidence
```

seasonal (lag = 7)+ seasonal (lag = 364) differencing of log transformed grocery sales data




```
AUGMENTED DICKEY-FULLER TEST for: seasonal (lag = 7)+ seasonal (lag = 364) di
fferencing of log transformed grocery  sales data
Test statistic = -5.538
P-value = 0.000
Critical values :
	1%: -3.436510851955201 - The data is  stationary with 99% confidence
	5%: -2.864260220574562 - The data is  stationary with 95% confidence
	10%: -2.5682183458999943 - The data is  stationary with 90% confidence
```

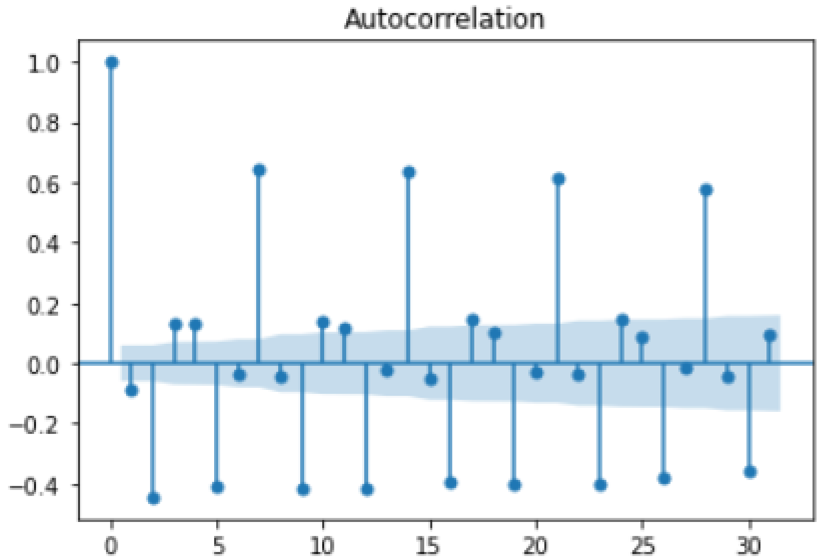



```
AUGMENTED DICKEY-FULLER TEST for: first differencing + seasonal (lag = 7) dif
ferencing of log transformed grocery  sales data
Test statistic = -14.913
P-value = 0.000
Critical values :
        1%: -3.434950510599593 - The data is  stationary with 99% confidence
        5%: -2.8635718069244227 - The data is  stationary with 95% confidence
        10%: -2.5678517223401403 - The data is  stationary with 90% confidence
```

Having passed all significance levels of DF test,to form sarima model, first differencing with seasonal lag = 7 was chosen for showing a regular mean with constant noise with very stable variance, other than the yearly spikes.

Since the only other model that managed to account for the yearly spikes with an acceptable noticeable stable variance was first differencing with seasonal differencing lag value 364, this was also chosen. However, the relevant sarima model could not be trained because of insufficient system resources. The training kept stopping in the middle after running out ram.

The acf plot of the log differenced data of seasonal differencing lag 7 and first differencing is shown below

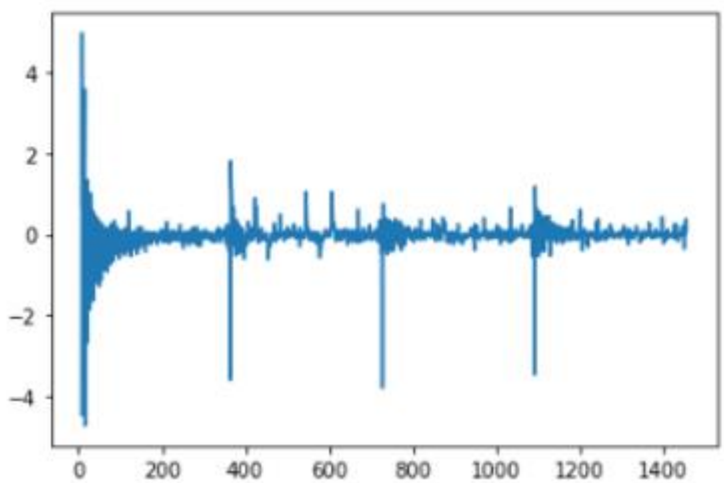


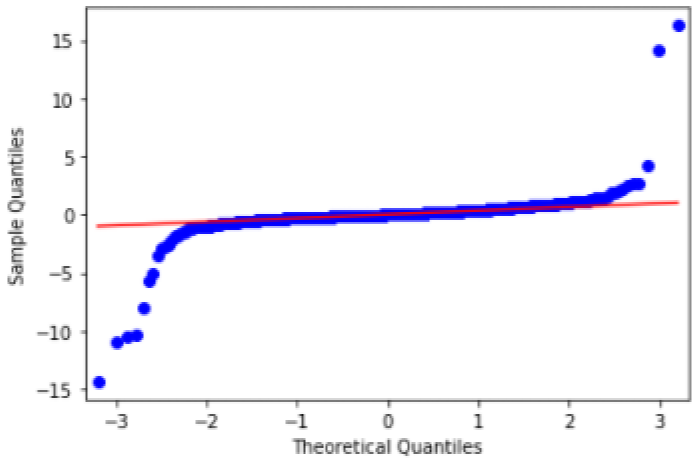
The seems to be alternating spikes of correlation starting at lag 2 other than the regular spike at lag 7. Besides that, many shorter significant spikes can be observed at further lags. There is no clear sign of tail-off or cut-off. Going by the mindset of parsimony, AR order of 2 with differencing of 1 in both the general and seasonal models is considered, with the seasonal lag order being 7. MA order was keo

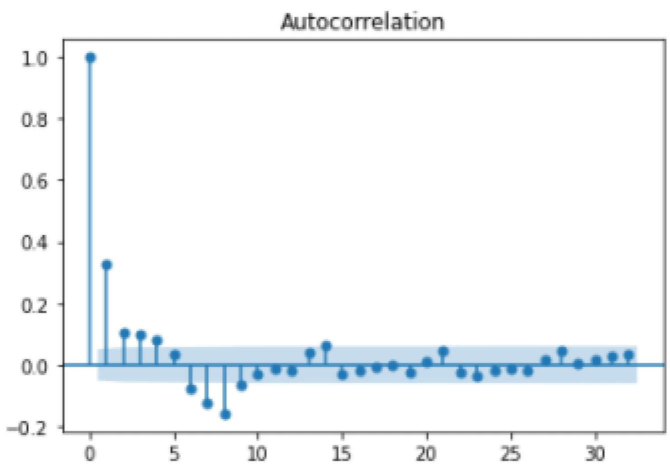
The trained SARIMAX(0, 1, 2)x(0, 1, 2, 7) model is as follows:

```
                                     SARIMAX Results                                      
==========================================================================================
Dep. Variable:                        differenced   No. Observations:                 1457
Model:             SARIMAX(0, 1, 2)x(0, 1, 2, 7)   Log Likelihood                -216.913
Date:                            Sun, 19 Dec 2021   AIC                            443.826
Time:                                    07:52:05   BIC                            470.219
Sample:                                         0   HQIC                           453.675
                                           - 1457                                         
Covariance Type:                              opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -1.7451      0.009   -191.605      0.000      -1.763      -1.727
ma.L2          0.7465      0.008     95.764      0.000       0.731       0.762
ma.S.L7       -1.8630      0.014   -130.114      0.000      -1.891      -1.835
ma.S.L14       0.8644      0.014     62.652      0.000       0.837       0.891
sigma2         0.0685      0.000    155.737      0.000       0.068       0.069
===================================================================================
Ljung-Box (L1) (Q):                   6.38   Jarque-Bera (JB):            379649.93
Prob(Q):                              0.01   Prob(JB):                         0.00
Heteroskedasticity (H):               0.53   Skew:                            -5.14
Prob(H) (two-sided):                  0.00   Kurtosis:                        81.63
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

This model too passes the Augmented dickey fuller test in all significance levels
Below are the plots of residuals, qq-plot of residuals, Acf plot of residuals in order:







As, seen in the case of plot of differenced data, this model also fails to account for the seasonal spikes. However, the qq-plot being straight in the entire middle region and the passing of ADF test points towards the residual being stationary. However, the acf plot indicates there is still some autocorrelation within the residual. The AIC and BIC are significantly larger than the best linear-seaonal hybrid deterministic models encountered before. Hence, this model too is not a perfect fit,

Section 3: Deep Learning

For deep learning, we considered two types of models, first MLP & then moved to 1D-CNN to forecast this time series.

Data Preparation for deep learning:

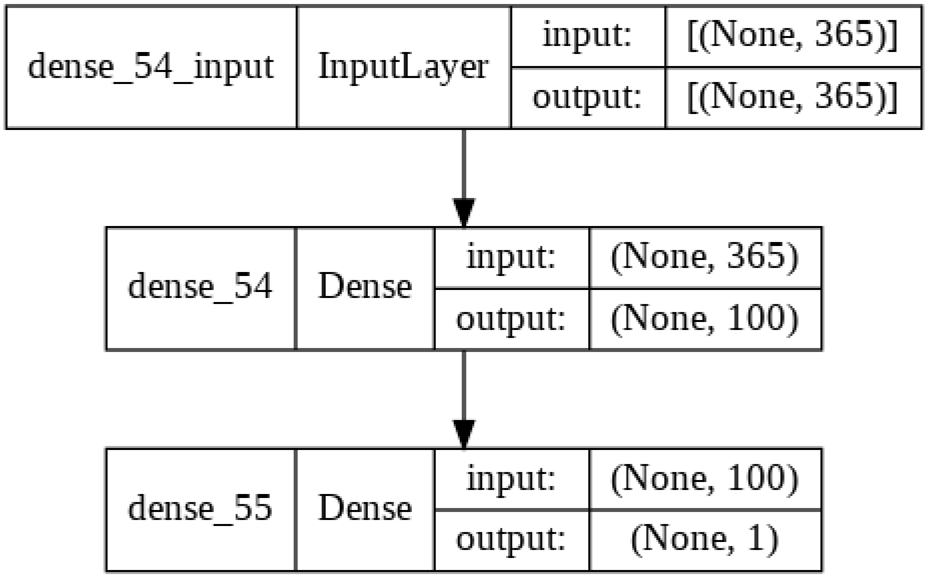
```
def series_to_supervised(data, window=1, lag=1, dropnan=True):
    cols, names = list(), list()
    # Input sequence (t-n, ... t-1)
    for i in range(window, 0, -1):
        cols.append(data.shift(i))
        names += [('%s(t-%d)' % (col, i)) for col in data.columns]
    # Current timestep (t=0)
    cols.append(data)
    names += [('%s(t)' % (col)) for col in data.columns]
    # Target timestep (t=lag)
    cols.append(data.shift(-lag))
    names += [('%s(t+%d)' % (col, lag)) for col in data.columns]
    # Put it all together
    agg = pd.concat(cols, axis=1)
    agg.columns = names
    # Drop rows with NaN values
    if dropnan:
        agg.dropna(inplace=True)
    return agg
```

To make data ready for deep learning models, a function was made to convert series data to supervised data where “window” refers to the size of training data for a prediction & “lag” means after how many days prediction(t + lag) will be made based on the previous window.

As shown in the previous plots of daily sales data, downward spikes were cycling around in about a year. Widow size was taken 365 throughout the experiments in the deep learning part & different values of lag were checked for that, which will be explained later.

Grocery

For the starter, we started with MLP with a one hidden layer. Architecture of the model is given below,



Two lag differences have been tested in this architecture for this series.
Lag 15: We have taken 15 since salary is paid in every 15 days. It is assumed to have some effect on sale.
Lag 7: Autocorrelation function plots of log-transformed sales data showed sharp-regular non-zero peaks at an interval of 7, indicating seasonality.

Results Analysis for MLP for Lag 15 & 7:

Lag/ Error Measures	Lag = 15	Lag = 7
R2 Score	0.8268891185250363	0.792905996090861
Mean Squared Error	624456192.2201644	745239831.94775

Here, we got almost similar accuracy for both lags with a slight edge to Lag 15. Both of the models were quite close to the original results as R^2 score & Fig 3.1 & 3.2 suggest but in the plot of ACF(of residuals[all ACF plots in section 3 are for residuals]) (Fig 3.3 & 3.4) quite a few numbers of values exceed the boundary of insignificance for both lags. Hence, the residuals are not independent. To improve that we moved to a different architecture. A CNN layer is added to the architecture since CNN has an advantage to extract the features better than a fully connected layer.

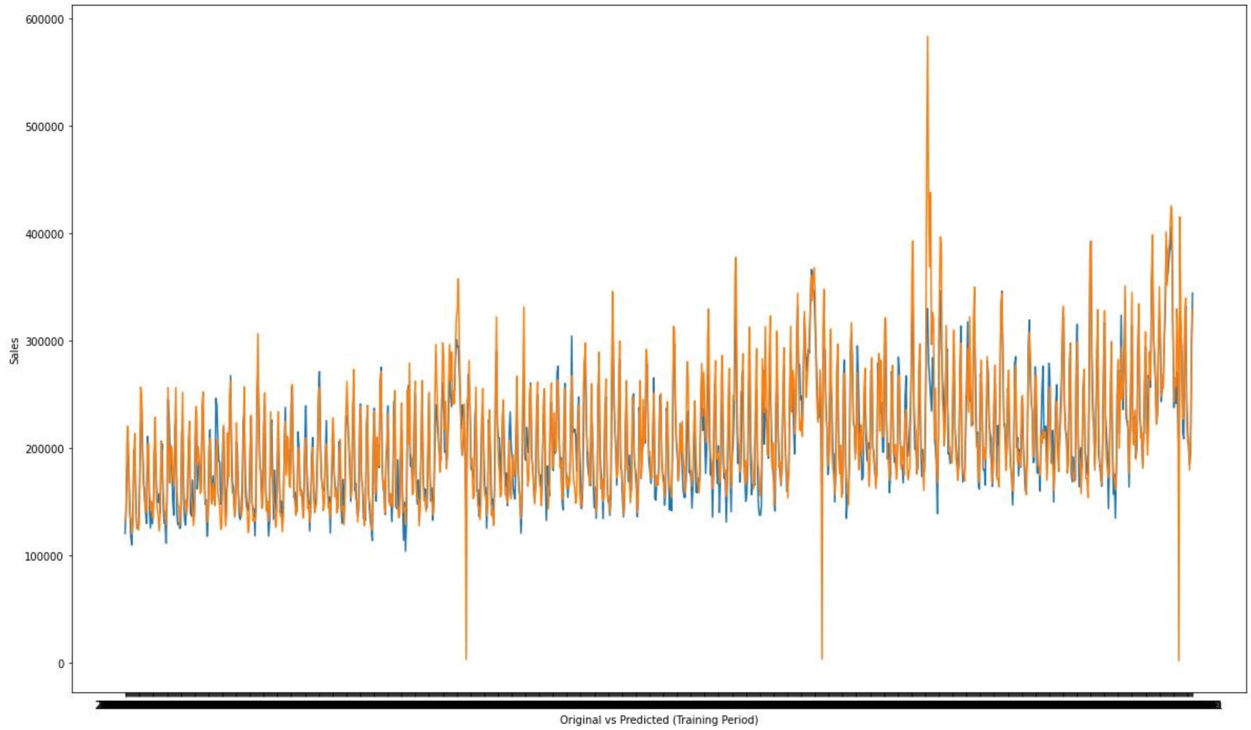


Fig 3.1: Prediction & Original Data (Training Period) [Blue=Prediction] for lag 15

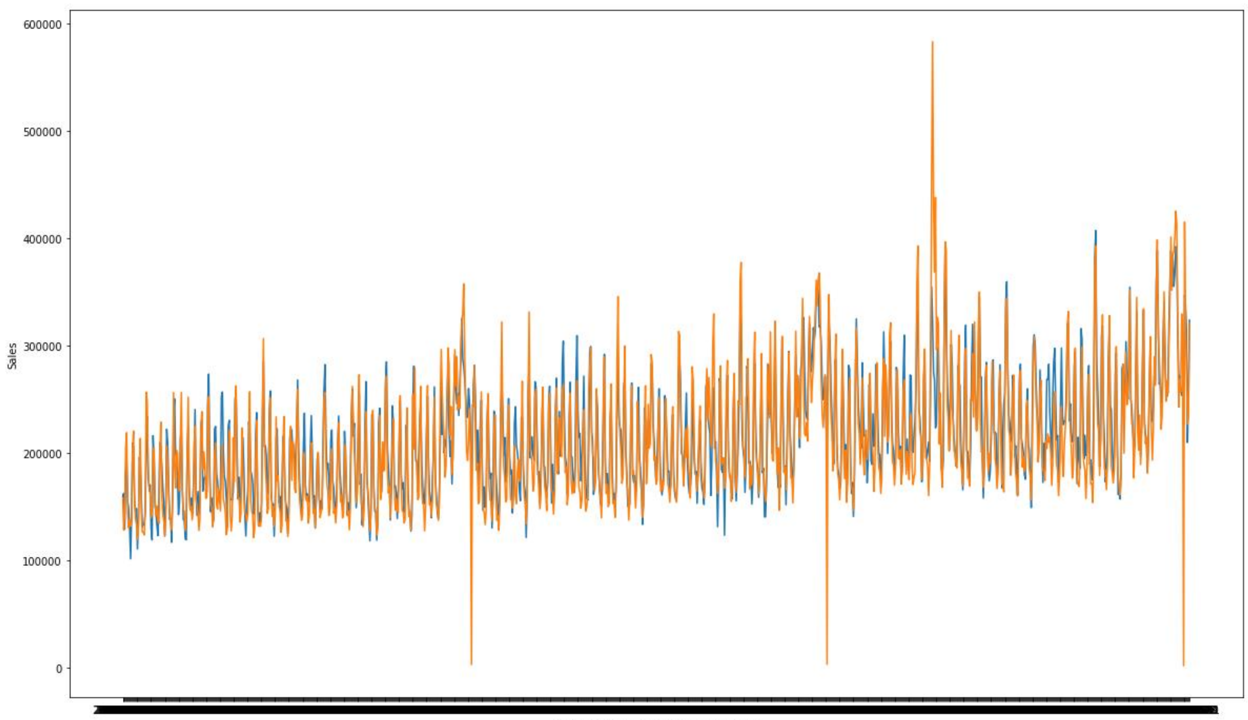


Fig 3.2: Prediction & Original Data (Training Period) [Blue=Prediction] for lag 7

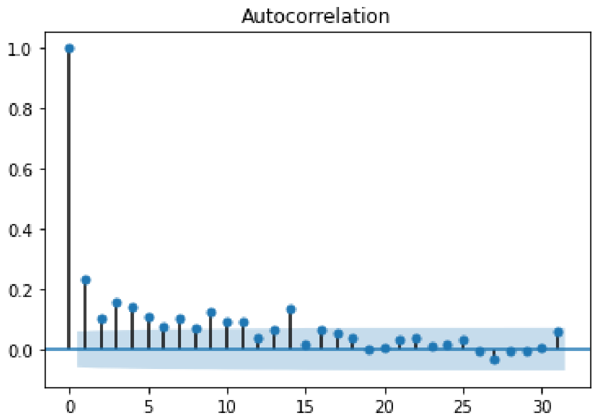


Fig: 3.3 ACF of MLP Lag 15

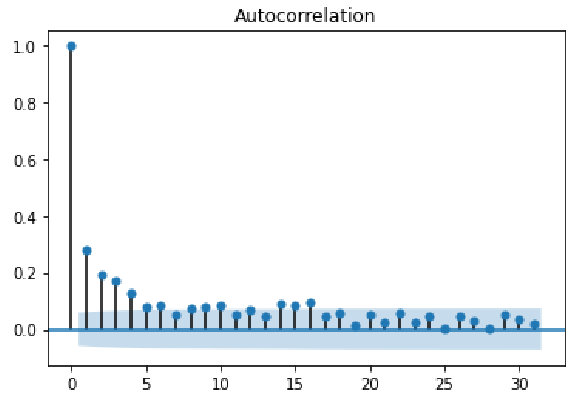


Fig: 3.4 ACF of MLP Lag 7

CNN Architecture:

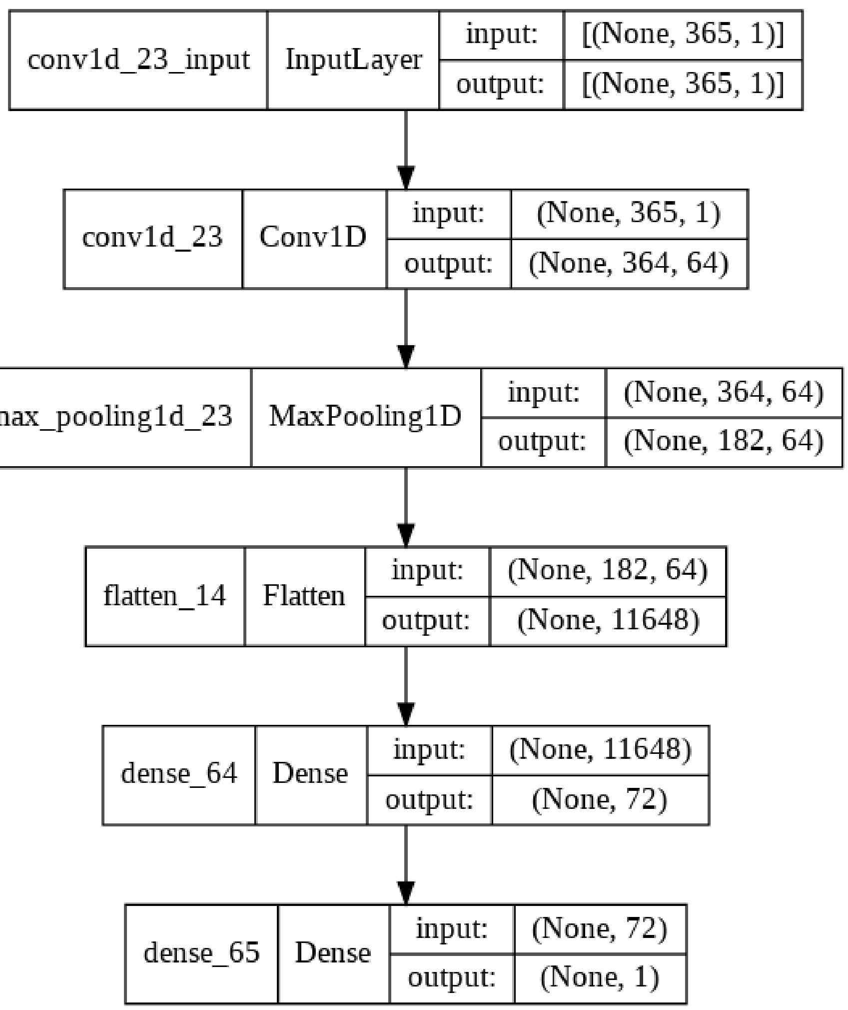


Figure 3.5 : Architecture of 1st CNN model used

We reduced the size of dense layer from 100 to 72 to reduce the number of parameters as our data is limited to train a large deep learning model.

Result of CNN

Lag/Error Measures	Lag = 15	Lag = 7
R2 Score	0.8999994734000627	0.901143926936365
Mean Squared Error(MSE)	360728034.70554453	355739335.21165884

Here, in terms of accuracy, we have got a huge boost compared to previous model but for Autocorrelation function shown in Figure 3.6 & 3.7, lag 15 one has

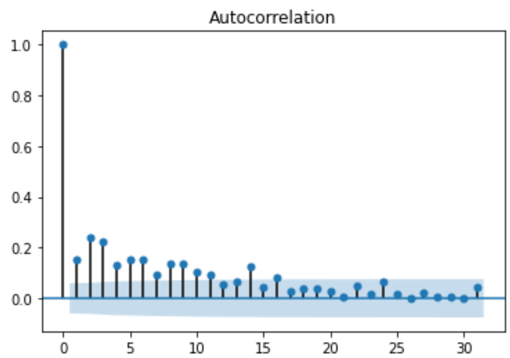


Fig 3.6: ACF CNN 1st Lag 15

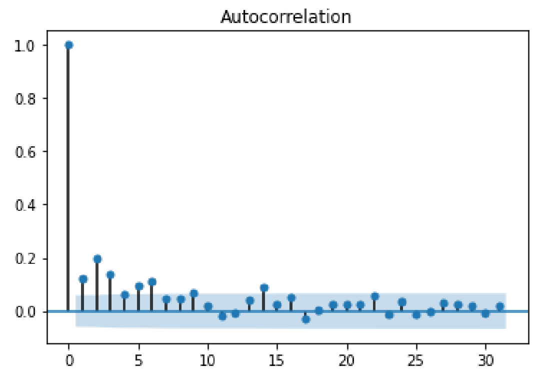


Fig 3.7: ACF CNN 1st Lag 7

more points out of the accepted region. Also lag 7 has higher accuracy according to R^2 & MSE. To see if we can reduce the number of point out of the boundary, we tried the same architecture with an added convolutional layer. Our assumption was, it might catch the small trends inside large trends.

Result of 2nd CNN Model

Lag/Error Measures	Lag = 15	Lag = 7
R2 Score	0.9144214848084599	0.9016556698970372
Mean Squared Error(MSE)	308704070.3451877	353897798.3693898

In this model, MSE has decreased significantly for lag 15 but the ACF plots shows more points outside of the blue region. In lag 7, accuracy remains almost same with more points getting away from the region.

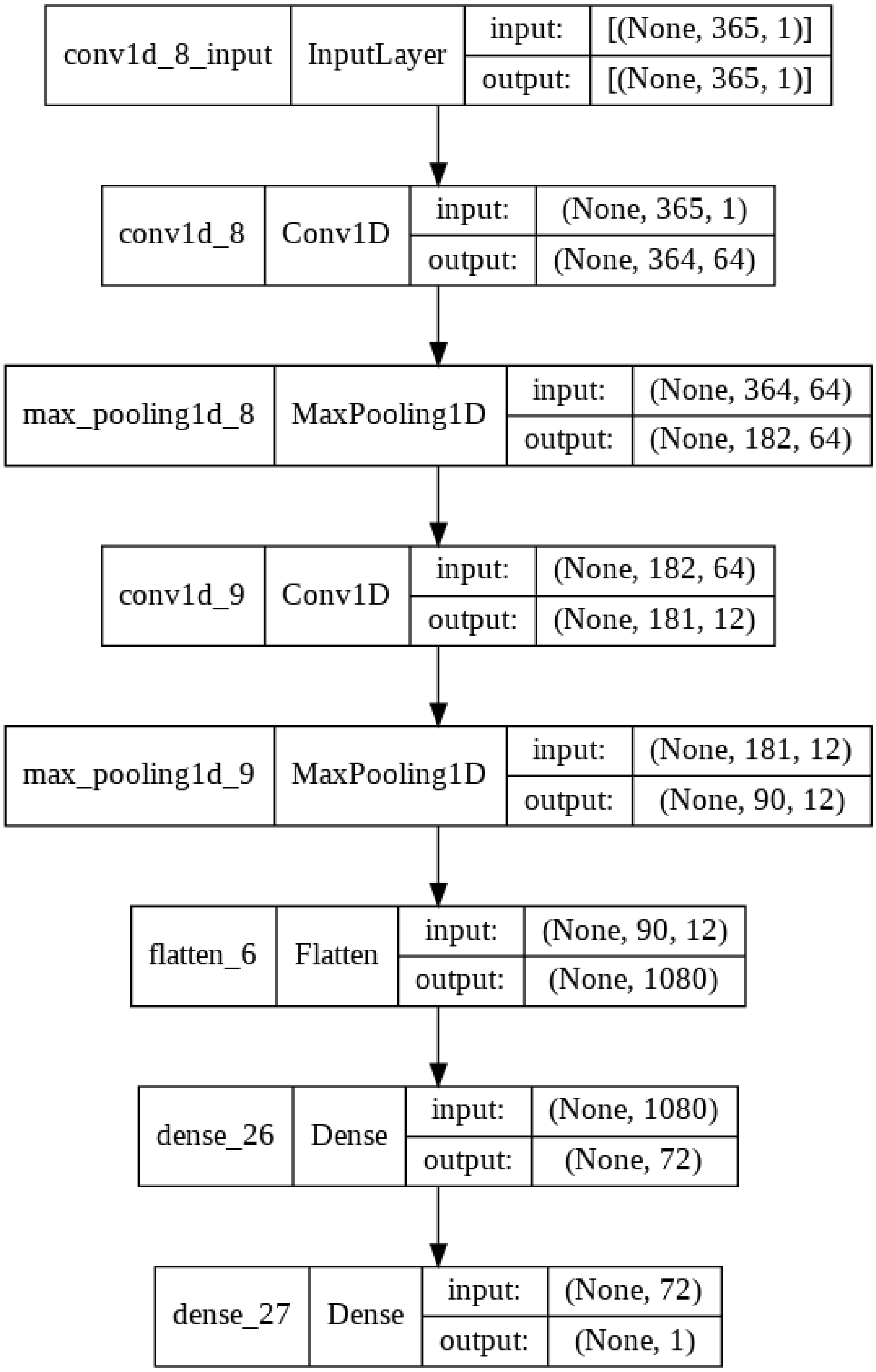


Fig 3.8: Architecture of the 2nd CNN Model

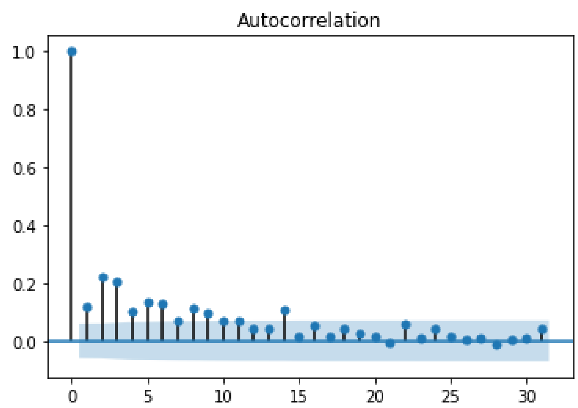


Fig 3.9: ACF CNN 2nd Lag 15

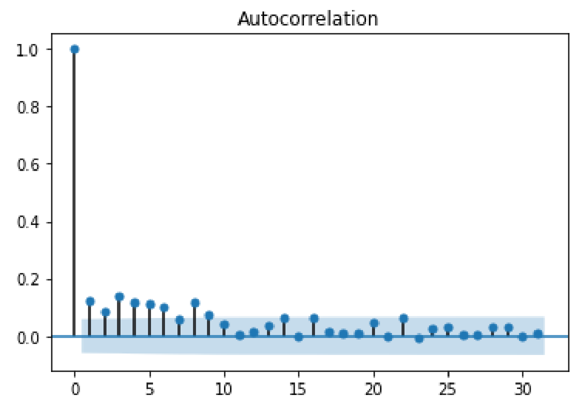


Fig 3.10: ACF CNN 2nd Lag 7

Section 4: Choosing Model

Among all the models considered so far, AI models show better fit by a large margin. The best deep learning models have R2-scores around 0.9, against around 0.5 being the norm in the traditional models encountered so far. On top of poorer fit, the traditional models fail to capture seasonality satisfactorily. So, deep learning models were chosen to be best for predicting sales data.

Among deep learning models, the CNN models have significantly better r2-score with MSE around half that of the MLPs. This is expected given the sophisticated structures of CNN involving filter and pooling layers on top of MLP structures which allow designers to craft them specifically to be able to capture the pronounced trends

Out of all the models discussed in this report, the lag 7 model of 1st CNN, has the accuracy almost as good as the highest one, while also having fewer significant non-zero autocorrelation spikes at difference lags in the acf plot. The mse was also decent among the four CNN models. This model could be argued to have been defeated in terms of MSE and R2-score by the second CNN model with lag15, however in terms of acf of residuals and having a smaller lag(making it more likely to be the true model) made the 1st CNN with lag 7, an overall good model.

Therefore, the first CNN model with lag 7 was chosen to be used for sales prediction. Below is plot showing it's performance on test set:

Testing the Chosen Model on test data (2017):

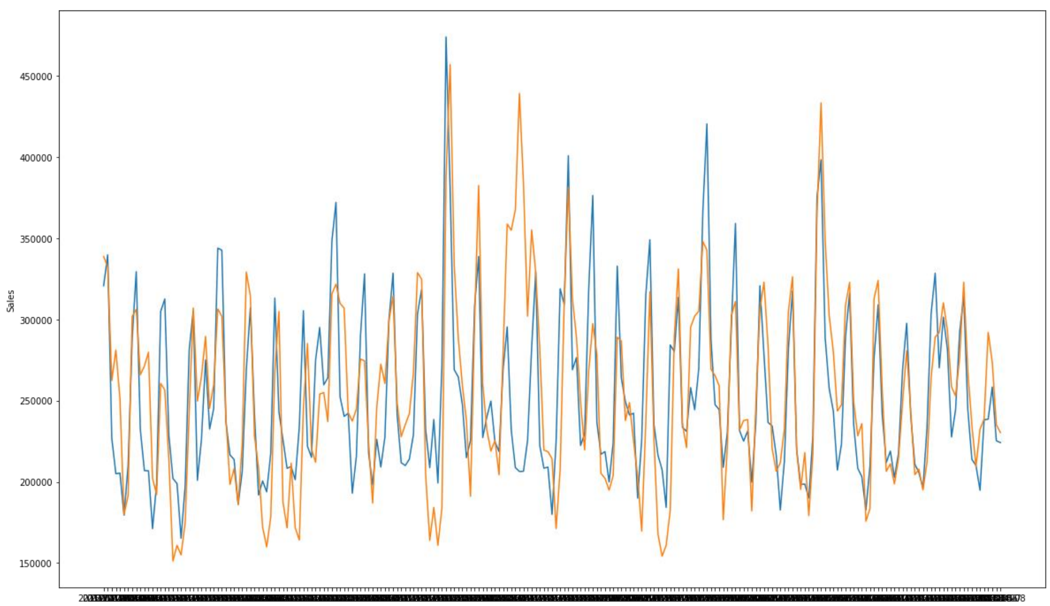


Fig 4.1: Original & Predicted Data [Blue: Original]

As can be seen, the test data predictions in this model are far better than what was seen for traditional models against their training data. The spikes are satisfactorily followed, despite there being some lag. Overall this model can be considered to be worthy of predicting sales data:

Secontion 6. Team member roles:

Sayem(20192024)

- Implementing Deep Learning models and their analysis.
- Writing report

Bayarmaa(20192029)

- Choosing models for prediction
- Checking theoretical soundness of the activities
- EDA presented in reports
- EDA presented within team (to guide project direction and help decide which models to implement and how to preprocess and use the available data)
- Writing report

Mubarrat(20192010)

- Implementing Traditional Models and their analysis.
- Writing report