Build Model:
I divide the dataset into 80% train and 20% test data. Models are trained using train dataset and validated using test dataset.
As our target is to predict sales given grps of different market channels(tv, radio, digital) and temperature, the next step is to pick suitable models.
In this case, the factors we need to take into account for predicting sales are adstocks from different channels and temperature. First I assumed they have linear interactions between each other and apply OLS regression to predict the linear relations between them.
Temperature is already given while adstocks are not. So it is also important to find a suitable adstock model. There are several types of adstock models, Simple Decay-Effect Model, Log-Decay Model, Negative-Exponential Model and Logistic (S-Curve) Decay Model. But since diminish and decay effects are considered in this case, I decide to choose only Negative-Exponential Model and S-Curve Decay Model.

Adstock model:
This is implemented in adsmodel function. Different model will be applied by changing the 'method' parameter('neg_exp' for Negative-Exponential Model, 's_curve' for S-Curve Decay Model) for this function.

Sales model:
As is mentioned before, linear model is tried first by applying OLS regression.

Tuning Parameters:
After model selection, its time to tune parameters and find the most suitable parameters. As adstock = function of (GRPs, diminishing, decay), and we have different ranges of diminishing and decay values for different channels. I tuned six parameters(tv_dim, tv_decay, radio_dim, radio_decay, digital_dim, digital_decay). I set interval for each parameter, e.g. tv_dim=list(range(120, 151, 30)) means 120, 150 will be tried, tv_dim=list(range(120, 151, 10)) means 120, 130, 140, 150 will be tried. The performance of these parameters are estimated using r squared value, which is the higher the better. And the parameters set which maximizes r squared value will be selected for modeling test dataset.

Validate Model:
After selecting the optimal set of parameters, they are applied in computing adstocks given test dataset, and then OLS regression is still for predicting the sales given the computed adstocks and temperature.

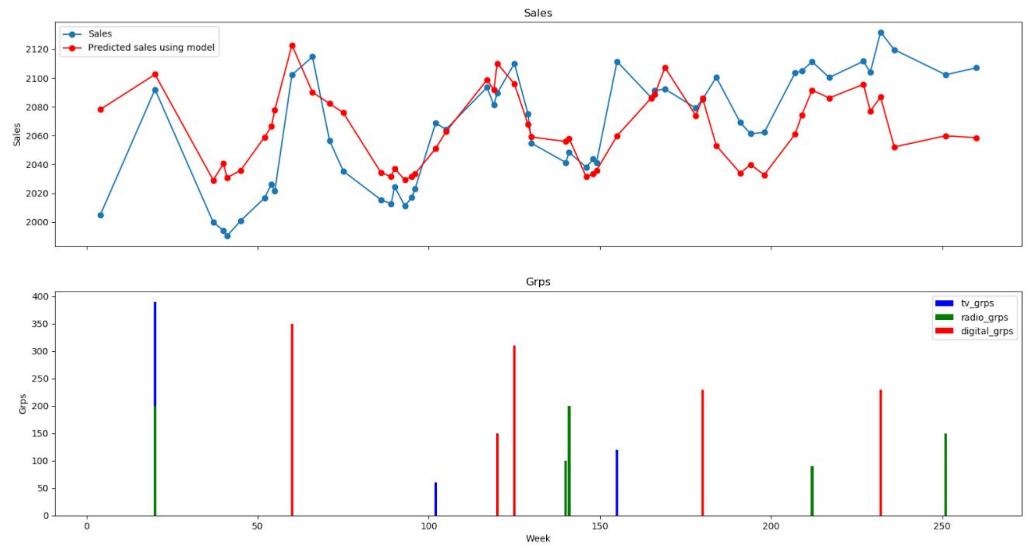
Test Result and Interpretation:
The following is the output of the model.
Negative exponential Model:
train max r squared value: 0.202932131482
tv_dim, tv_decay, radio_dim, radio_decay, digital_dim, digital_decay: 150 0.6 180 0.3 70 0.6

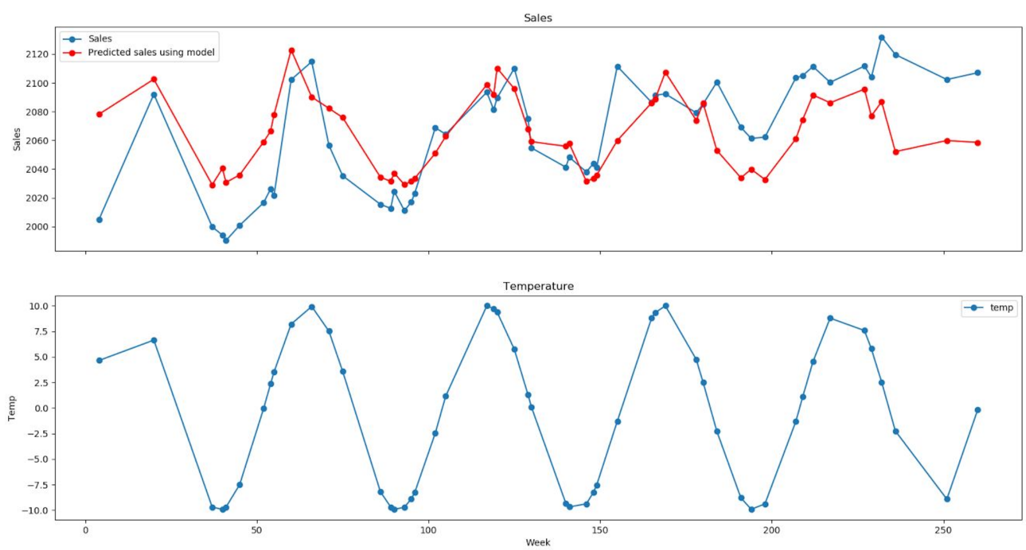
```
(52,)
model test r squared:  0.410469696405
model test summary:                             OLS Regression Results
==============================================================================
Dep. Variable:             sales_test   R-squared:                       0.410
Model:                            OLS   Adj. R-squared:                  0.360
Method:                 Least Squares   F-statistic:                     8.181
Date:                Tue, 26 Dec 2017   Prob (F-statistic):           4.31e-05
Time:                        22:50:47   Log-Likelihood:                -255.55
No. Observations:                  52   AIC:                             521.1
Df Residuals:                      47   BIC:                             530.9
Df Model:                           4
Covariance Type:            nonrobust
======================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept        2035.7627      8.432    241.446      0.000    2018.801    2052.725
tv_test_ads        35.0917     11.244      3.121      0.003      12.472      57.712
radio_test_ads      5.4615     16.453      0.332      0.741     -27.637      38.560
digital_test_ads   31.1602     12.080      2.579      0.013       6.858      55.463
temp_test           1.6763      0.707      2.371      0.022       0.254       3.098
==============================================================================
Omnibus:                        0.874   Durbin-Watson:                   2.431
Prob(Omnibus):                  0.646   Jarque-Bera (JB):                0.613
Skew:                           0.266   Prob(JB):                        0.736
Kurtosis:                       2.973   Cond. No.                         28.3
==============================================================================

tv_dim, tv_decay, radio_dim, radio_decay, digital_dim, digital_decay:  150 0.6 180 0.3 100 0.6
```

The r squared value is 0.410 in this example. I have tried several times and find this values varies a lot in different tests. I think this is due to the fact that the dataset is too small that the model cannot the fitted quite well and has uncertainty. It may also due to that the tuning parameter interval is not small enough(save running time) that the best parameters set is not chosen.

The effects of different channels can be also seen from the above result. In the column of coef, which denotes coefficient, tv_test_ads has a coefficient of 35, which means 1 tv adstock will lead to 35.1 additional sales. 1 radio adstock results in 5.5 additional sales and 1 digital ads results in 31.2 additional sales. 1 temperature increase leads to 1.7 additional sales.

I have also plot two figures to show the relation between sales and adstocks, sales and temperature.





From the first figure, we can tell that the increase of digital_grps leads to great increase of sales, the effect of radio is not obvious, which coincides with the previous observations from OLS regression. The effect of tv, however, is either not obvious in this plot.
From the second figure, I observed that the trend of sales seems to follow the trend of temperature, which hasn't been discovered by OLS.

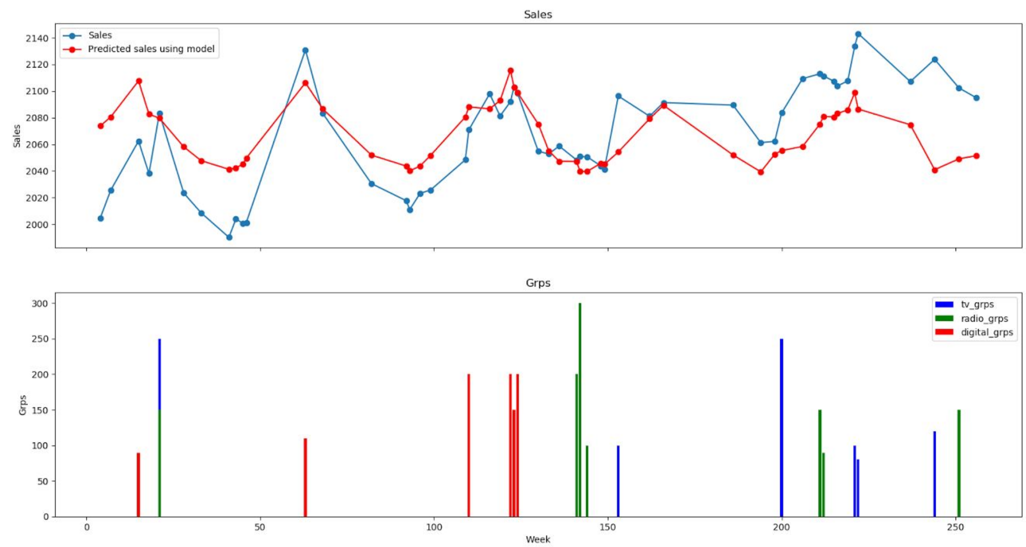
S-Curve Model:

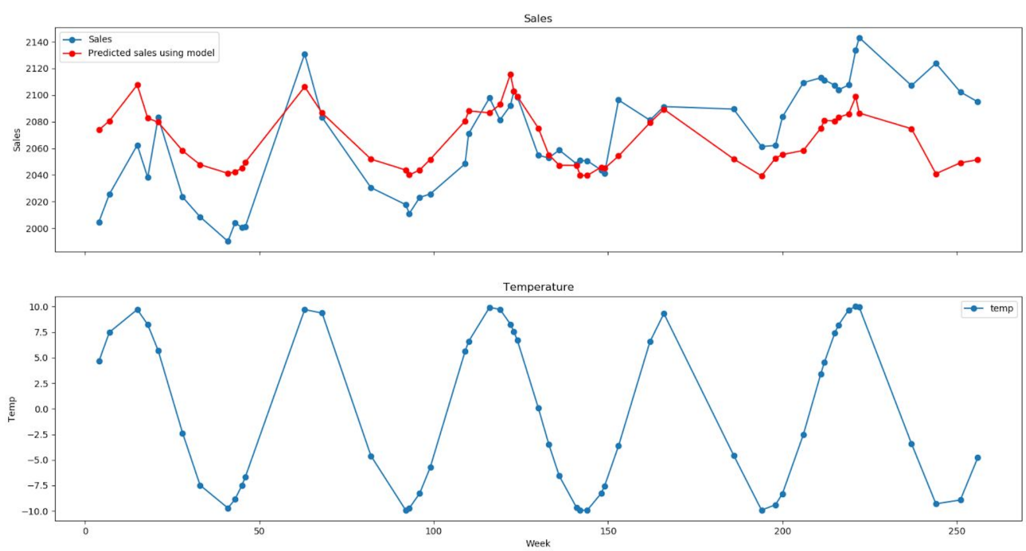
model test r squared: 0.299695246293

model test summary: OLS Regression Results

```
==============================================================================
======
Dep. Variable:           sales_test   R-squared:                    0.300
Model:                          OLS   Adj. R-squared:               0.240
Method:               Least Squares   F-statistic:                  5.028
Date:              Wed, 27 Dec 2017   Prob (F-statistic):         0.00186
Time:                      00:37:44   Log-Likelihood:             -256.14
No. Observations:                52   AIC:                          522.3
Df Residuals:                    47   BIC:                          532.0
Df Model:                         4
Covariance Type:          nonrobust
==============================================================================
============
                  coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept       2011.9694    39.249     51.261      0.000    1933.010    2090.929
tv_test_ads        0.0482     0.336      0.143      0.887      -0.628       0.725
radio_test_ads     0.2809     0.722      0.389      0.699      -1.171       1.733
digital_test_ads  40.4580    28.281      1.431      0.159     -16.436      97.352
temp_test          2.3706     0.689      3.439      0.001       0.984       3.757
==============================================================================
======
Omnibus:                      0.880   Durbin-Watson:                1.797
Prob(Omnibus):                0.644   Jarque-Bera (JB):             0.855
Skew:                         0.097   Prob(JB):                     0.652
Kurtosis:                     2.402   Cond. No.                      478.
==============================================================================
======
```

The r squared value for S_Curve is lower than negative exponential, but as mentioned before, this varies between tests. But it captures the effect of different marketing channels better. 1 tv_adstock, radio_adstock, digital_adstock and temp increase leads to 0.05, 0.28, 40.46, 2.37 sale increases respectively. This captures the trend better.





The above plot produced using s-curve model follows the same trend and same rule as the plot produced in the previous model.