# Search and guest behavior

After scaling the two highest search terms "old spaghetti factory" and "spaghetti factory" we can see that there the relationship of guest counts and search activity is not significant.

Looking at the initial visualizing of guest behavior and search traffic for the keywords, google searchs do not seem to help predict guest activity.

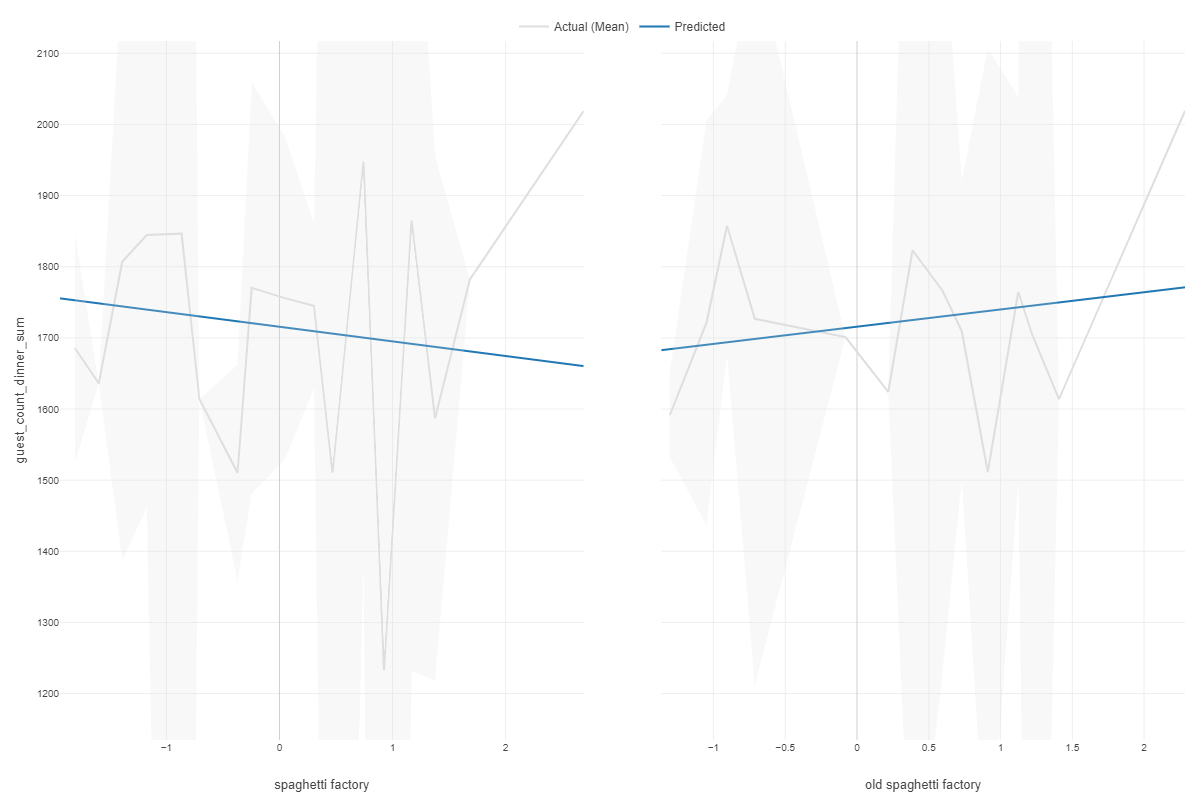


# Linear Regression with Keyword behavior - Prediction

After fitting the model with the scaled keywords, we created a linear model to predict the dinner guest counts.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| R Squared | Adj R Squared | RMSE | F Ratio | P Value |  |
| 0.0069654 | -0.034411 | 237.4285 | 0.1683426 | 0.8455611 | 51 |

Unfortunately, the summary for this model is not looking promising, even with the most direct keywords to predict guest counts, with a high p-value and low R squared value there is no significant evidence to say that keywords predict guest behavior.

This is most likely attributed to the fact that seasonality is very important to predicting guest behavior this will be seen later.

# Correlation - Positive Pairs

After normalizing the data to ensure it is on the same level of scalability as the other variables, I ran an initial correlation to find the highly correlated variables.

As expected we found that last years monthly guest counts had a high positive relationship with the next year's guest counts. We know that seasonality is a strong predictor of guest counts.

The search volume for "Old spaghetti factory" and "spaghetti factory" were highly correlated after scaling the data to fit all searches within Oregon during that time frame.

We see that guest counts and total people in reservations have a positive weaker relationship.

We also see that there is a significant negative relationship between the search volume of old spaghetti factory and the average temperature. Signifying that as the the temperature decreases, the search volume decreases for OSF. and so do reservations.

|  |  |  |
| --- | --- | --- |
| Column 1 | Column 2 | Correlation |
| din\_ly\_sum | guest\_count\_dinner\_sum | 0.6615647 |
| old spaghetti factory | spaghetti factory | 0.6162000 |
| guest\_count\_dinner\_sum | ppl\_res\_din\_sum | 0.4445878 |
| din\_ly\_sum | ppl\_res\_din\_sum | 0.3787429 |
| month\_unq | season\_unq | 0.3066221 |
| italian food | spaghetti factory | 0.2696790 |
| din\_ly\_sum | temp\_avg\_mean | 0.2487840 |
|  |  |  |
|  |  |  |
|  |  |  |

# Important Predictor Variables

We can see as it is consistent in the correlation matrix that last years guest counts along with the reservations per day do help predict guest counts and have a strong relationship that is very seasonal.



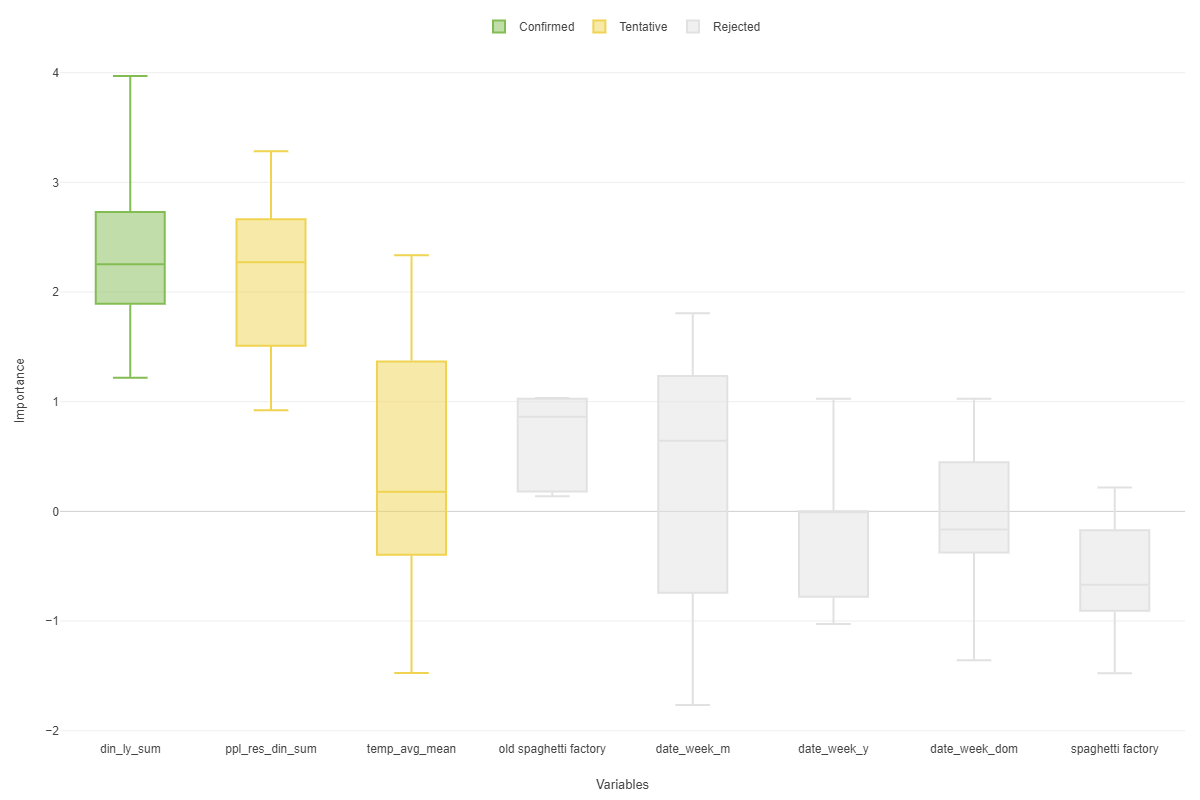
# Random Forest - Importance

A random forest serves as an exploratory analysis which indicates which variables are strongly associated with the increase or decrease of guest counts.

We can see that as consistent with our other correlation and previous results, that last years dinner guest counts is the strongest predictor of guest counts by far followed by the tentative variables of reservations and the weekly seasonality indicated by the two date week variables.

date\_week dom indicates the reoccurring dates such as the 25th or 11th each of each month

date\_week\_m indicates the day of the week in this case "Sunday"



# Linear model to find importance of variables to predict weekly guest counts for dinner - Coef. Table

It seems that the most important and significant variable to predict guest counts is last year's dinner guest counts.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Term | | | Coefficient | | Std Error | t Ratio | | P Value |  |  | |  |
| (Intercept) | | | -0.0704739 | | 0.7974009 | -0.0883796 | | 0.9299951 |  |  | |  |
| din\_ly\_sum | | | 0.4877747 | | 0.1259668 | 3.8722486 | | 0.0003708 |  |  | |  |
| italian food | | | 0.0052829 | | 0.1114829 | 0.0473877 | | 0.9624288 |  |  | |  |
| month\_unq | | | 0.1531184 | | 0.2961350 | 0.5170561 | | 0.6078294 |  |  | |  |
| old spaghetti factory | | | 0.0749770 | | 0.1496668 | 0.5009593 | | 0.6190152 |  |  | |  |
| ppl\_res\_din\_sum | | | 0.2271380 | | 0.1231019 | 1.8451215 | | 0.0720785 |  |  | |  |
| season\_unq | | | -0.0829952 | | 0.8115692 | -0.1022651 | | 0.9190330 |  |  | |  |
| spaghetti factory | | | -0.0472733 | | 0.1399751 | -0.3377269 | | 0.7372512 |  |  | |  |
| temp\_avg\_mean | | | 0.0435535 | | 0.1380409 | 0.3155116 | | 0.7539357 |  |  | |  |
| R Squared | Adj R Squared | RMSE | | F Ratio | P Value | Number of Rows | Degree of Freedom |  |  |  |  |  |
| 0.4467668 | 0.341389 | 0.6739541 | | 4.239668 | 0.0008557 | 51 | 8 |  |  |  |  |  |

# General Additive Model - Summary

By using a general additive model time series algorithm, we predicted the last 10 weeks of 2018's dinner guest counts using the previous 42 weeks of data. We can see that this does quite a better job at predicting weekly guest counts. and does it quite well with an average accuracy of 87% (MAPE = 12.9%). That is great!

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RMSE | MAE | MAPE | MASE | Number of Rows for Training | Number of Rows for Test |
| 253.0464 | 190.6524 | 0.1297463 | 0.8992028 | 43 | 10 |

This goes to show that seasonality is a very strong predictor of restaurant business. Although google searchs do indicate some activity online, it does not show the whole picture in regards to the Old Spaghetti Factory!