Forecasting Sales Based on Promo

Erdi Olmezogullari

Introduction

- In this task, we figure out the market demand with respect to promos and sales.

- During that process, we follow up those steps, below:

  1)Loading data,

  2)Data cleansing and formating

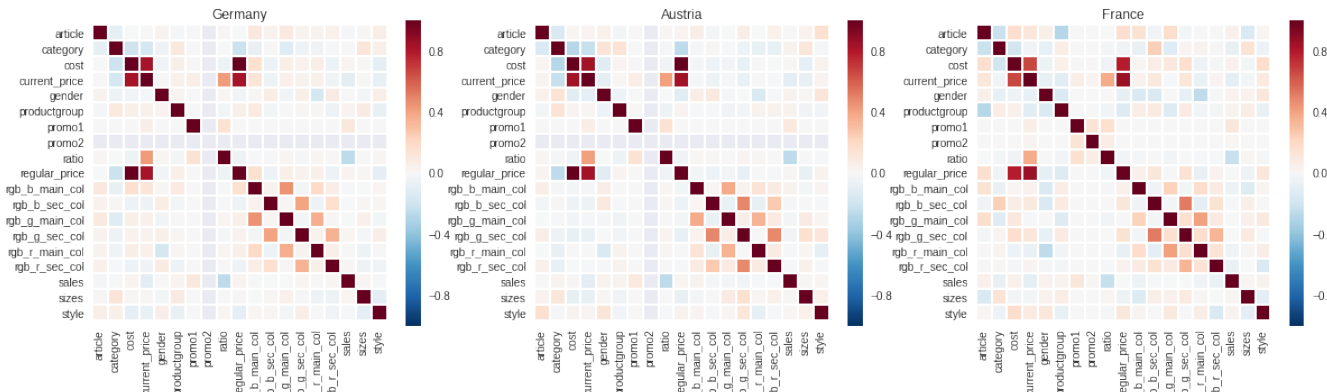 3)Building a model

  4)Forecasting

Summary

- To explore valuable insight related to our promos from data, we used mainly two different techniques, respectively.

  - Time Series Analysis (SARIMAX)

  - Regression Tree and Random Forest Regression

- We got valuable results from our sales data by using to our statistical and ML based data processing like above.

- Our insights related promos about data;

  - Promos are working in Germany and Austria well. The both countries have same patterns.

  - However, Promos are not working in France. Promo1 s better than promo2 in France because the returning is more high.

  - The number of promos by promo are almost same in all country. There is no promo2 in Germany and Austria, properly

  - Although all countries have similar trend on the sales, their trends gains are different. The order of gain of trend from highest to lowest is Germany, Austria, France, respectively

- In time series analysis, we can focus on only five features (date, country, promo1, promo2, sales) to create a time series model fast and easy. However, the rest of features may play important roles on the sales. To cover the rest of them, we referred to use Regression Tree and Random Forest Regression as a complementary approaching

- In addition, the modern approach can be Deep Learning. RNN and LSTM have suitable structure to process time series analysis.

Country	Promo1	Promo2
Germany	+	N/A
Austria	+	N/A
France	+/-	-

+ : Successful
- : No successful
+/ : Moderate Successful
N/A : No promo

A quick glance to data

Feature Correlation Matrices by Country

Germany Austria France

Promo1, Promo2 vs Sales by Month (Normalized)

Germany Austria France

Correlation of Promo1, Promo2, Sales by Country
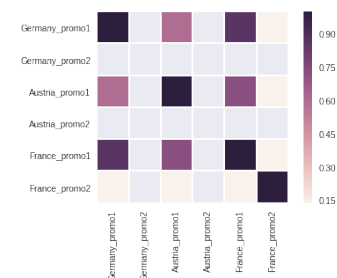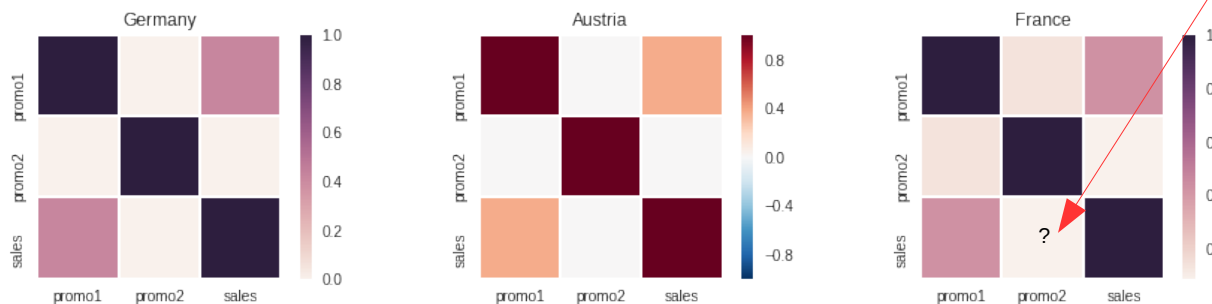
Germany Austria France

- All countries have similar correlation matrix with respect to the features of product.
- According to those matrices, promo1 and promo2 don't have neither positive neither negative correlation with any kind of features.
- There are significant high correlated features. However, we skipped them in this step.

- According to the normalized graphs related to Promo1, Promo2 and Sales, we can see the effects of Promo1 and Promo2 on top of Sales
- The promos based on Germany and Austria may have same patterns on the sales. However, France has more different pattern than the rest two countries.
- Promo2 may have more positive effects than Promo1 on the sales in France.
- Promo2 may less contribution on the sales some points on the graph as we pointed out with "?". That proposal is also improved by the correlation matrices (bottom left).
- Promo2 is not available for German and Austria.
- Promo1 may have similar pattern on the sales in the both Germany and Austria
- After this first introduction, we need to verify these patterns on the sales.
- In all countries, Promo1 has positive correlations according to the correlation matrices (bottom left).

# Time Series Analysis – Decomposition - Approach#1

After applying time series decomposition, we got three different components, such as Trend, Seasonal, Residual as shown above.
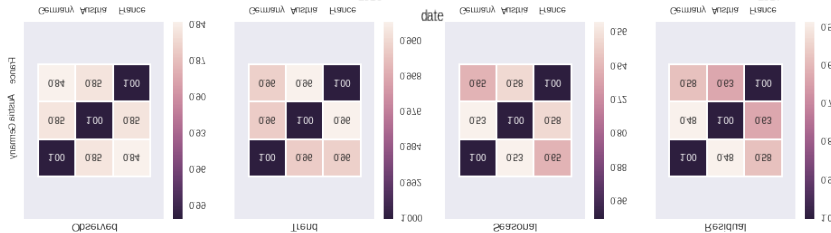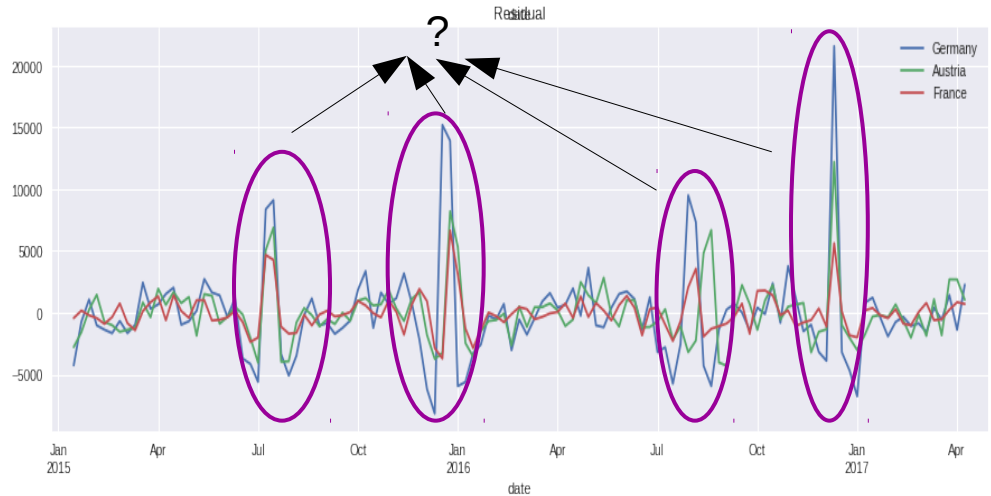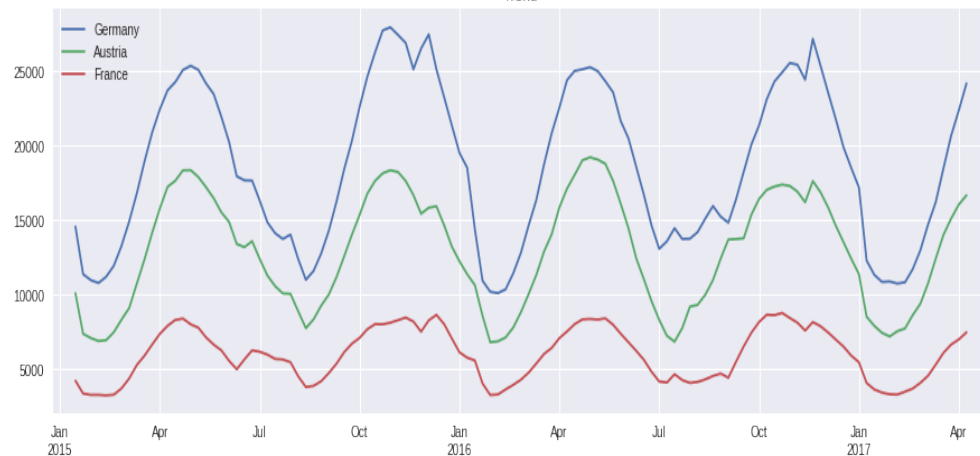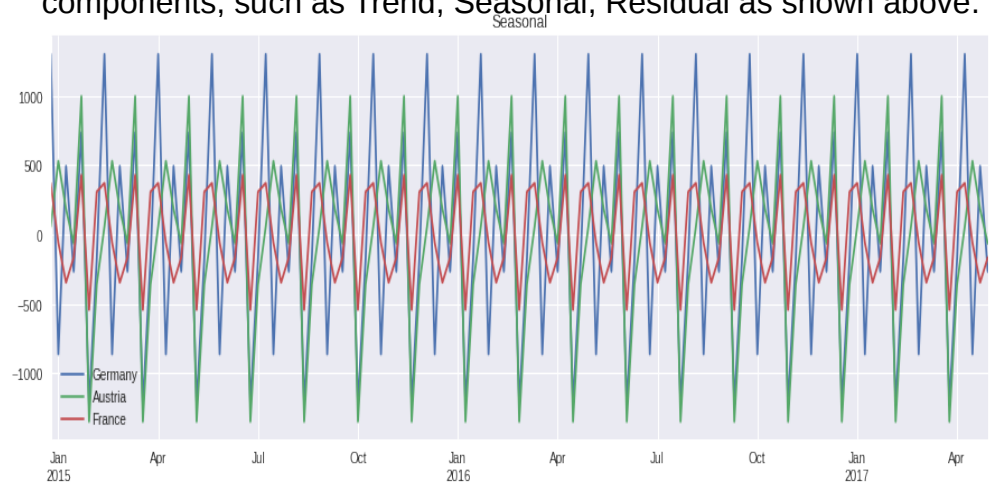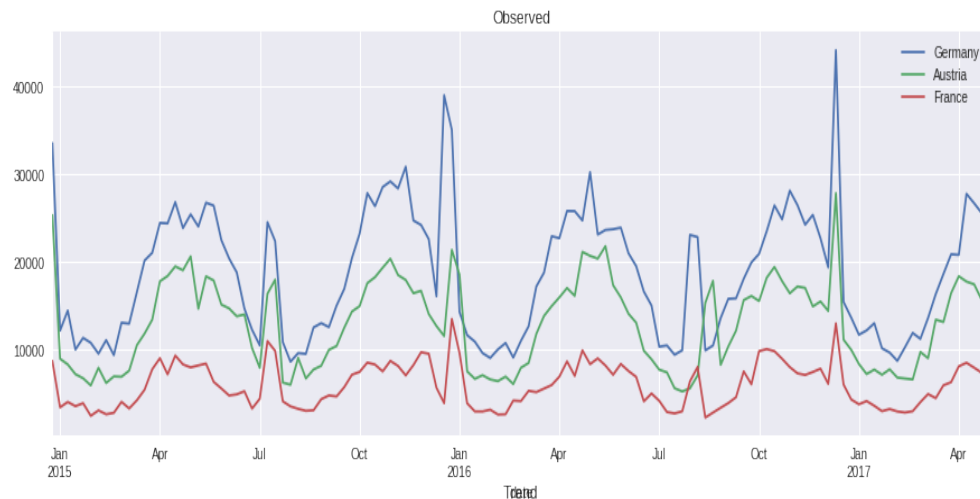
After decomposition, our expectation is that residual component has to be stationary. It seems stationary, overall. However, a few critical effects rather than noise affects the curve where we point out. Mostlikely, those effects may be related to promos since promos don't have neither seasonal nor threading patterns.

Comparing Promos and Residual

Promo1, Promo2 vs Residual by Country (Normalized)
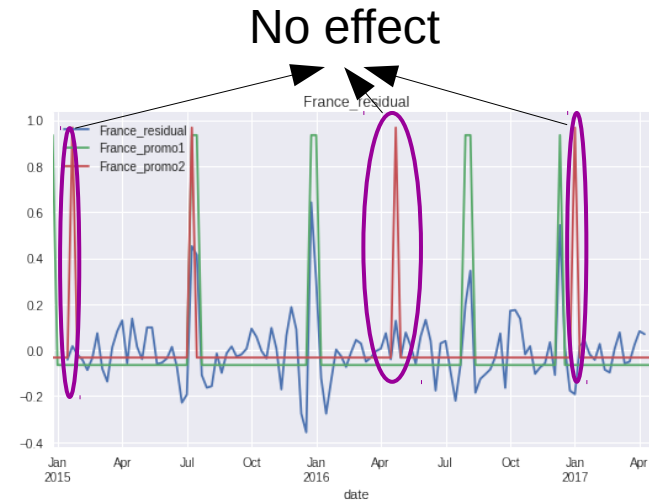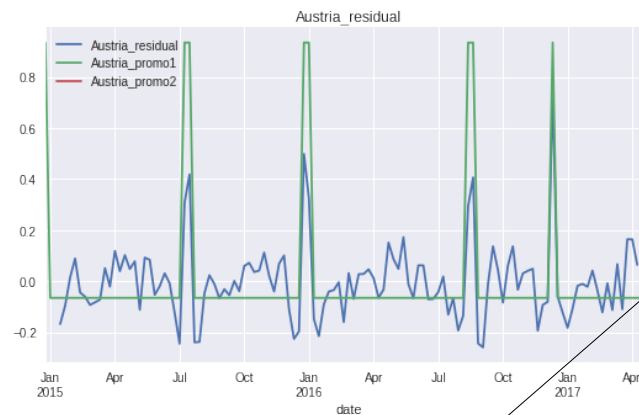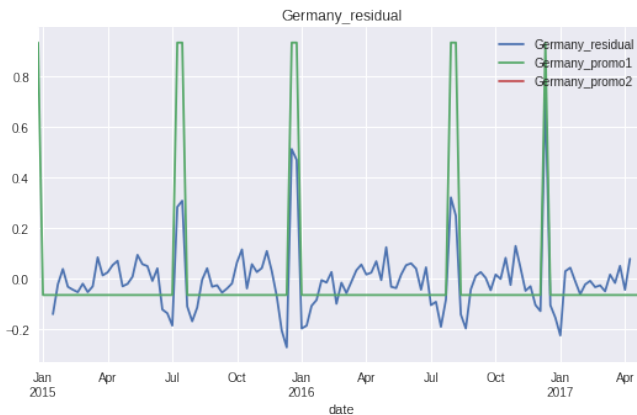
No effect

- According to the graphs, above, our residual components have promos' effects as we discussed in the previous slide.
- Germany and Austria have same effect in terms of promos on the residual.
- However, France has more different pattern than the rest country. As we proposed at the beginning, Promos may not work for France market well. In this graph, we can obviously see.
- Promo2 don't have positive effect on the sales mostly.

Modeling Time Series

Autocorrelation, Partial Autocorrelation, Histogram, Boxplot on Residual Component

Moving Average, Moving STD
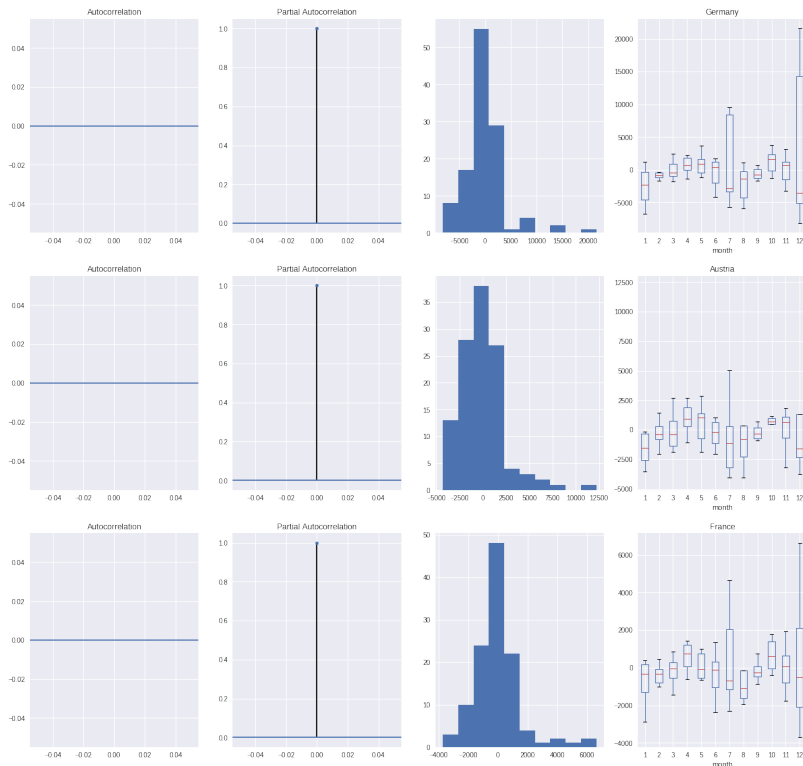
- To check the stationarity of residual component, we used those methods, above.
- According to boxplot, the variance is a little bit high due to promos effects.
- According to autocorrelation, there is no strongly correlate lags on this component.
- In addition these approaches, we can take difference (d = x(t) – t(t -1)) of actual series to obtain stationary components well.
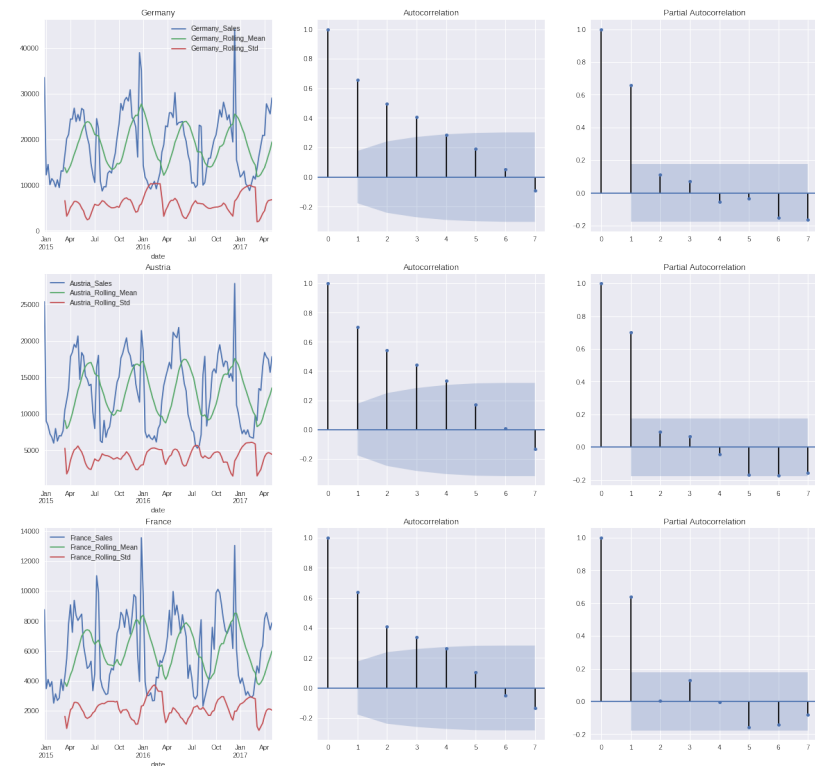
- Before building a model for forecasting by using SARIMAX, we processed the data by using moving mean and std
- According to autocorrelation function, we have four powerful lags to use as a MA(4) order.
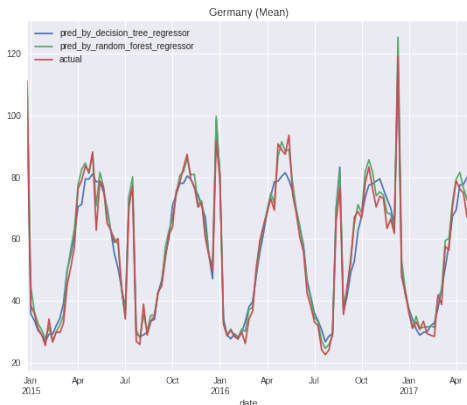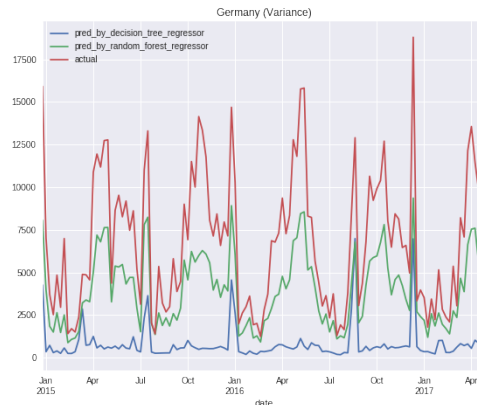
Forecasting
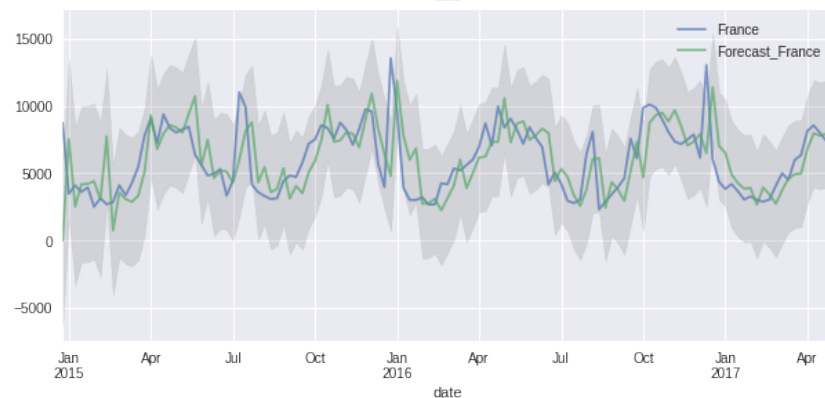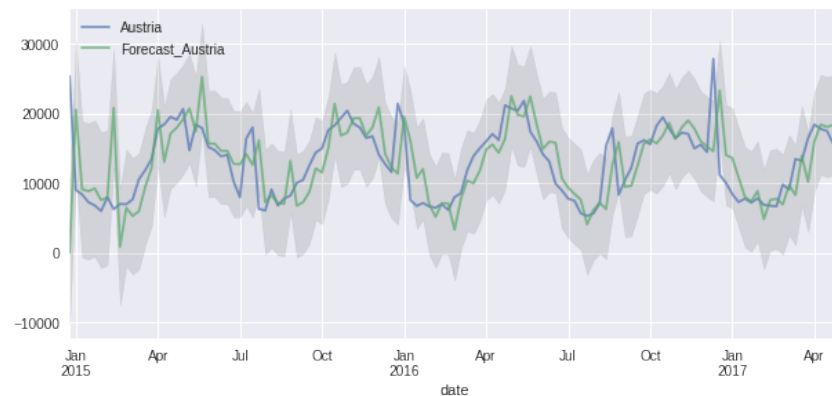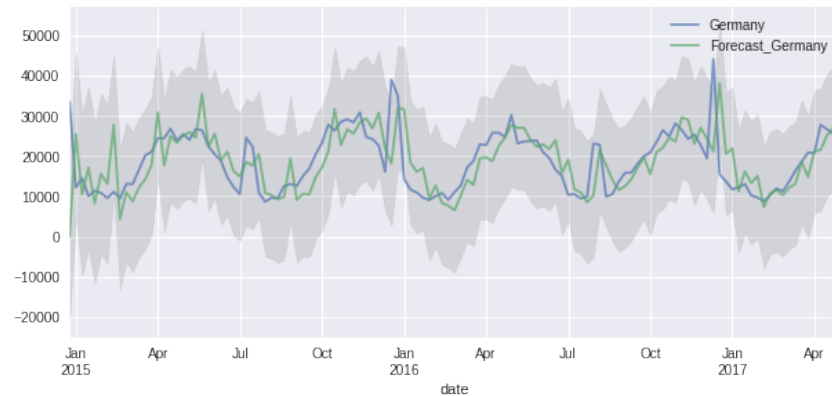
- We built three different models for each country to forecast the sales
- As you see in the graphs, there are two curves such as actual sales and forecasting sales.
- At this point any given date, we can make a forecasting about our future sales, roughly.

Regression - Approach#2
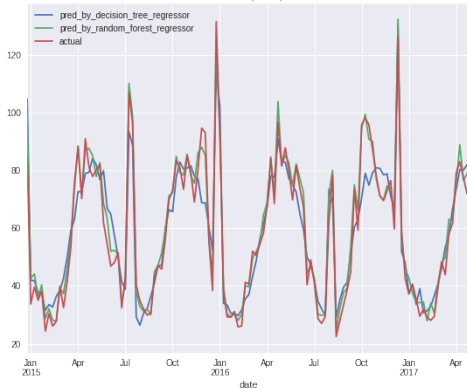
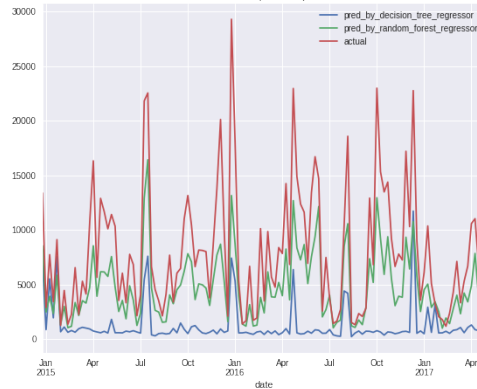**Actual vs Expected Sales (Mean)**   **Actual vs Expected Sales (Variance)**

- To see all features impact to sales, We used ML algorithm.
- Regression Tree is kind of regression method that takes categorical inputs and
  maps to continues value.
- In our case, sales is target feature. The rest of them is input features.
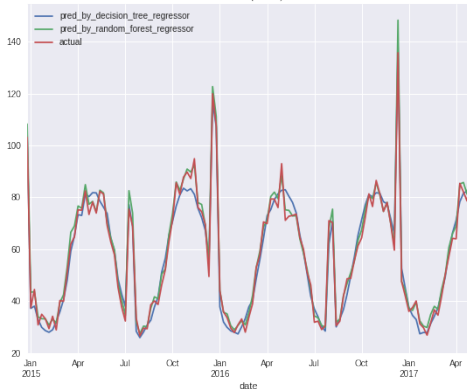- We also used Random Forest Regression. However, its performance worse than regression tree since it is over-fitting according to its variance graph.