Forecasting BASF's Custom Material Demand Using ABC-XYZ Analysis

Team 7 Final Report
Business Analytics Using Forecasting
Professor Galit Shmueli
Fall 2018

Project Members
Li-Cheng Pan | Travis Greene | Patrizia Mach
Triguna Ashin Wijaya

January 8, 2019

Executive Summary

The project goal was to optimize BASF's global supply chain and production line for custom materials. Due to the varied nature of parts used in the automobile industry, different models of cars require unique pieces. As car models are discontinued, demand for their parts decreases. Further, the intermittent nature of custom material demand and spare parts meant that BASF must constantly worry about potential capacity breaches. Considering a lead time from order to ship-out of two months, our aim was to give BASF's automobile division a monthly demand forecast for the next two months for each unique material.

The original data extracted through the BASF ERP system contained the number of daily shipped units for 826 unique materials from its automotive division. Given our forecasting goal, we aggregated the irregularly-spaced demand units into regular monthly intervals. Materials with negative and zero demand volumes in the test set period have been filtered out. The training period was 59 months long (10/12-8/17); the test period 12 months (09/17-08/18). RMSE in the test period was used as the performance metric as per BASF company standard evaluation.

In order to provide a simple heuristic to understand both the behavior and importance of different materials, the material series were grouped using ABC-XYZ analysis. For all series in each cell of the ABC-XYZ matrix, we tested commonly-used forecasting methods and selected the method producing the lowest average RMSE, while also taking into account computing time. The chosen model was then used to forecast the future demand for each material in that group. In nearly all groups, a naive forecast was found to be best, meaning past order quantities were often the best indicator for future orders. Exponential Smoothing (ETS) performed better than the naive in groups AZ and BZ, which represent the series with lowest forecastability and high to medium importance. In the low forecastability, high-importance group AY, an ensemble of methods represented a nearly 19% improvement in RMSE over the naive. Given the global scale of BASF's operations, these relatively minor improvements could potentially save millions of dollars in excess inventory costs.

As the majority of factory and production managers are not data analysts, we built an easy-to-use interactive forecasting platform using R Shiny to allow simple but immediate forecasts of quantities per series (appendix figure 9). The platform makes forecasts (both point and interval) based on our recommended method for each group of materials and shows these forecasts in relation to an estimated maximum capacity. The system further provides a historical view of a material's demand behavior and an interface for finding materials with similar historical demand characteristics to give an idea of future behavior.

Going forward, we recommend that BASF consider the unit cost of materials when conducting the ABC-XYZ analysis. We also suggest that BASF provide more information related to the types of materials to more easily identify correlated time series. Lastly, we advise future teams to explore the performance of forecasting methods using a multi-year, roll-forward validation scheme.

Detailed Report

1. Problem Description

BASF is a global chemical company that specializes in the manufacturing of a variety of custom-made materials. Our project primarily relates to parts used in the automobile industry. One of BASF's key competitive advantages is its *Verbund* concept, by which its factories and production partners' infrastructure are intelligently linked, allowing for the efficient reuse of steam energy, excess materials and manufacturing byproducts. Currently, BASF relies on a combination of demand forecasts from in-house experts and various manufacturing partners in order to estimate production demand. BASF's in-house business intelligence team creates high-level macroeconomic strategic forecasts as well as single customer behavior forecasts aimed at advanced demand planning and sensing. As such, the company seeks a more data-driven approach to these customer-level forecasts for factory managers. Our proposal is therefore aimed at augmenting these low-level single customer forecasts with material-level demand forecasts. This solution is designed to give production managers ample time to implement appropriate policy and scheduling changes, such as increasing or reducing factory production levels. In our study, material demand is operationalized as the number of monthly shipped units of a given material.

BASF faces two major obstacles in optimizing its supply chain and production line for these custom materials. First, because specific models of cars require unique parts, parts ordered from BASF are only demanded throughout the lifecycle of the relevant car model. Second, the intermittent nature of custom material demand means that BASF must deal with potential capacity breaches. For example, a customer may suddenly place an order for a number of units that BASF's factories cannot possibly produce given the lead time. In these cases, the forecast should warn production managers that capacity is likely to be reached, so that necessary changes to cope with rapid shifts in demand can be made. According to BASF, the required lead time to take action and adjust requires a two-months-ahead forecast. This way managers know what quantities are expected to be shipped in the next two months.

For the purpose of monthly planning, our forecasts will be generated on a monthly basis, which means that in January, for example, forecasts for February and March will be made. Then, in February, forecasts for March and April will be made, and so on.

2. Data Description

The data provided to us by BASF were the result of collected and extracted through the ERP (Enterprise Resource Planning) software SAP Hana. In total there were 108,000 rows representing orders of 826 unique materials. We were given columns such as the date when units were shipped (*date*), the production company (*company*), the product SKU (*desc*), the number of shipped units (*demand*), the pseudonymized receiving company of the shipment (*ship2*), and the pseudonymized type of material (*material2*). Data quality issues are to be expected when obtaining data from an automated system in a company as large as BASF. In this case, a small fraction (roughly 50 cases) of demand values were negative. On the advice of BASF domain experts, these were assumed to be errors generated by the ERP system. See figure 2 in the appendix for an example of the original data. Although the data exhibited no clear seasonal patterns, evidence of a Chinese New Year effect remains (appendix figure 3.1).

3. Data Preparation

The data obtained from BASF originally contained the irregularly- timed number of shipped units for 826 unique materials. We first cleaned the data by removing observations with negative demand, and then aggregated shipped orders to regularly-spaced monthly orders emulating the monthly demand for each material. Materials which did not receive any orders during the 12-month test set period (9/17-8/18) were excluded from analysis.[1] One material was removed because it did not contain any order with positive values. The training period was 59 months (10/12-8/17). As per company standard, RMSE in the test period was used as the evaluation metric for each forecasting method. We decided on a 12-month testing period in order to evaluate model performance throughout a year and in order to detect seasonal patterns. See figure 3 in the appendix for a visual of the series divided into training/testing periods.

4. Forecasting Solution

ABC-XYZ analysis was used to group the series. This technique is commonly employed by consulting companies and is recommended by some forecasting experts as a practical alternative to time series clustering.[2] It has the advantage of providing analysts with a simple heuristic for understanding both the behavior and importance of different materials. ABC-XYZ analysis groups series into three levels of importance: A (high importance), B (medium importance), and C (low importance). "Importance" is quantified by that particular material's proportion of total sales among all unique materials. In other words, if one material makes up a disproportionately large amount of the total demand among all materials, that material will be deemed "important."[3] In this case, in confirmation of the Pareto Principle, roughly 75% of demand for all materials came from just 20% of the unique materials. See figure 4.

The XYZ portion of the analysis attempts to group series by quantifying how "forecastable" each series is. This metric is computed by using both a one-step-ahead naive and seasonal forecast, measuring the RMSE of the better of these forecasts, then dividing by the mean value of the series. The resulting metric is similar to a coefficient of variation and allows us to better compare series of varying levels and helps mitigate the dominating effects of series with very high demand. We found that the top 20% of materials (roughly 165 series) make up about 53% of all errors over all series. See figure 5 in the appendix for the XYZ results.

The results of the analysis were promising. There was only one series that was highly difficult to forecast and also very important (AZ group). Conversely, there were over 330 series (BX, CY, CX) that were both easy to forecast and relatively low volume. For the series in these groups, we adhere to parsimony and apply a naive approach to forecasting[4].

The results of our ABC-XYZ analysis revealed that a naive forecast is surprisingly difficult to beat (see figure 7). Only two groups showed a better performance under a different method than the naive forecast. For example, in the AY group (high importance and medium forecastability) an ensemble method

[1] We recommend that these materials receive a naive forecast until more information can be obtained.

[2] http://kourentzes.com/forecasting/2016/10/15/abc-xyz-analysis-for-forecasting/

[3] The logic behind this comes from the *Pareto Principle,* which states that many economic phenomena exhibit a pattern in which 80% of the outcome can be attributed to just 20% of the input.

[4] For our purposes, a naive forecast is defined as using the most recently observed value of a series as a prediction for the next value in the series.

consisting of a seasonal naive, STLF,[5] auto arima, and ETS performed better than a naive forecast. The improvement over the naive RMSE for this group was about 19%. In the AZ and BZ groups (high and mid importance, and low forecastability), which in total contain five series, the ETS method performed the best, with only a narrow improvement on the naive RMSE by 1.3%. These minor gains may nevertheless represent significant cost savings for a global corporation as large as BASF. The additional computation time required for the ETS and ensemble forecasts is negligible. For these reasons, for series in the AZ, BZ, and AY groups we recommend using ETS and ensemble methods, respectively.

After evaluating test set performance, we then graphed forecasting errors and noted that the January, April, and August forecasts appear to show a pattern. Specifically, the January and April forecasts tended to be too *low*, while August's forecasts tended to be too *high*. In practical terms, this means that BASF may tend to underforecast customer demand in January in April and may need to acquire buffer stock to avoid losing out on important business. In contrast, we might caution BASF about overpredictions in August in order to avoid excess inventory. According to our discussions with BASF's domain experts, overpredictions should be favored over underpredictions, all else equal. In this case, this was not a deciding factor because all methods tested systematically tended to overpredict, on average.

5. Deployment and Maintenance

For BASF to take full advantage of our analysis, we created a Shiny app showcasing the functionality of our approach. The app contains three main functions: the first is the recommended forecasting method for each material with a customizable forecast horizon. In addition to displaying the ABC-XYZ grouping and suggested forecasted method, the interface includes a Dicky-Fuller test evaluating the probability of the series being a random walk. This test assuages any doubts about the value of recommending a naive forecast for many of the groups.[6] We also include point and prediction intervals. Prediction intervals can be useful for estimating when potential capacity breaches are likely to occur. Lastly, we included an estimated maximum capacity value based on the historical maximum order value for each series.

The second main function of the platform is to help users understand the historical behavior of each series. For example, users may want to know in which month and years this material received the most orders to anticipate for a new, but similar product. Analysts might also be interested in the frequency and size of most recent orders to make informed management decisions.

Finally, the third tab in the app allows users to visually find materials whose previous order behavior is closest to the material of interest. This is computed through a distance matrix of all series based on the autocorrelation function of each material. Once we obtained this distance matrix, a simple K-nearest neighbors' algorithm was used to find the top *k* nearest neighbors. Users view a *k* of five similar series as default. The rationale follows that users may want to use these series to estimate the future behavior of a similar material of interest.

[5] The STLF method first deseasonalizes a series, fits an ETS model to this deaseasonalized series, then adds back the estimated seasonal component and makes a forecast.
[6] If the series is not meaningfully different from a random walk, then we conclude there is simply no signal in the data and thus the best we can do is to predict the previous value plus some Gaussian random noise.

Since the BASF Business Analytics Department already uses R Shiny Server for their forecasting and visualization needs, our solution is easily integrated into the BASF workflow. Analysts and factory production managers simply log in to the R Shiny Server platform remotely. They are then able to make production and staffing decisions based on our forecasts. We suggest that the ABC-XYZ analysis be re-run every six to twelve months as new materials are introduced and as business conditions change. After this regrouping, an analyst should then retest the groups to find the optimal method per group.

6. Limitations and Recommendations

We note some limitations to our solution. Firstly, the current ABC-XYZ analysis does not include cost of production due to insufficient data. This information could greatly change the importance rankings. We recommend that BASF includes this information in future analyses.

A second limitation is the 12-month testing period. Given further time and scope, perhaps a better way of measuring performance would be to explore a roll forward validation strategy that allows us to look at monthly performance over several years. It is likely that using different training/test set time periods, would lead to different recommendations on optimal forecasting methods in each group.

The focus on RMSE as the KPI to determine a group's suggested forecasting method may be a third limitation. The ABC-XYZ group RMSE values should be scaled so that individual materials with very high RMSE scores--i.e., outliers—do not disproportionately affect the group's overall RMSE. To solve this issue, we suggest including histograms or boxplots along with average RMSE since these show the distribution of RMSE scores. Individual outliers could then be analyzed, and specific methods could be used on just those outlier series.

Regarding recommendations for BASF, we suggest augmenting the data-driven forecasts with domain expertise of production managers for the more difficult-to-predict months of January, February, and August.

Lastly, we implore BASF to include more information related to the type of materials in the forecast analysis. We believe with further details on what type of material groups particular SKUs belong to, it would be possible to incorporate potentially correlated external series, such as raw petroleum, oil and steel indices to improve overall forecasts.

Appendix

Figure 1. A schematic representing various levels of forecasts conducted by BASF. Our forecasts will augment the existing forecasts in the bottom two levels of the hierarchy.

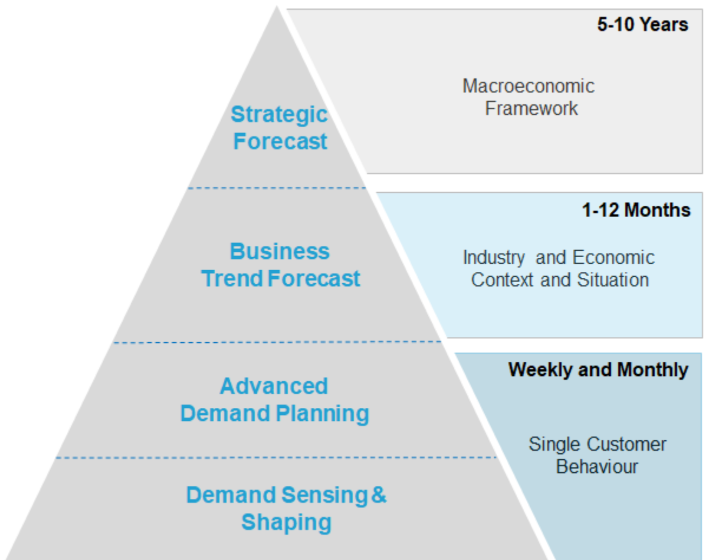

Figure 2. Excerpt of raw data provided by BASF. Note the row with negative demand.

No	date	company	desc	demand	ship2	material2
1	14/9/2016	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	2100	42	254
2	21/11/2016	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	4410	42	254
3	20/12/2016	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	1200	42	254
4	17/2/2017	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	6300	42	254
5	21/3/2017	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	5040	42	254
6	30/1/2018	BASF Polyurethane Sp	PRO-P.6-KE1-11-BUM-MH-30	-1500	42	254

Figure 3. Division of the 825 series into training and testing periods. The period to the left of the dashed line was used to train the various forecasting methods and the period to the right of the line was used to compute the group's average RMSE. Note that many series have values close to 0.

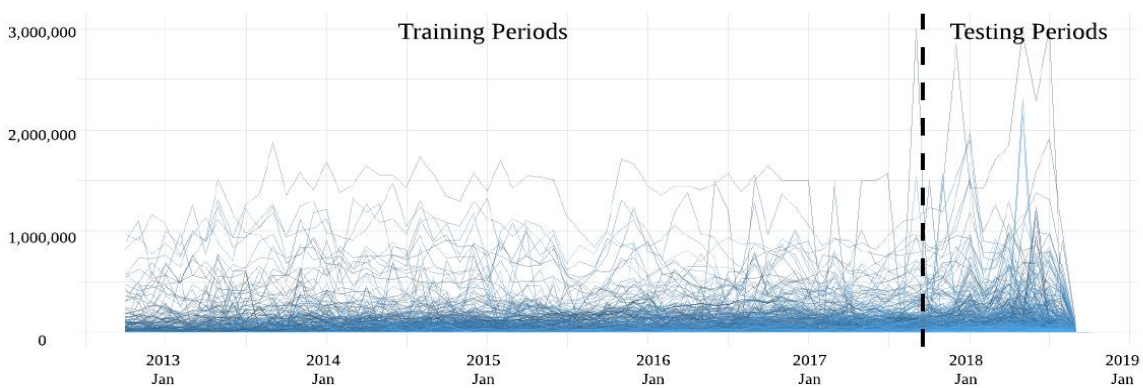

Figure 3.1 Mean order quantities per month in each year shows slight dip during the Chinese New Year month.

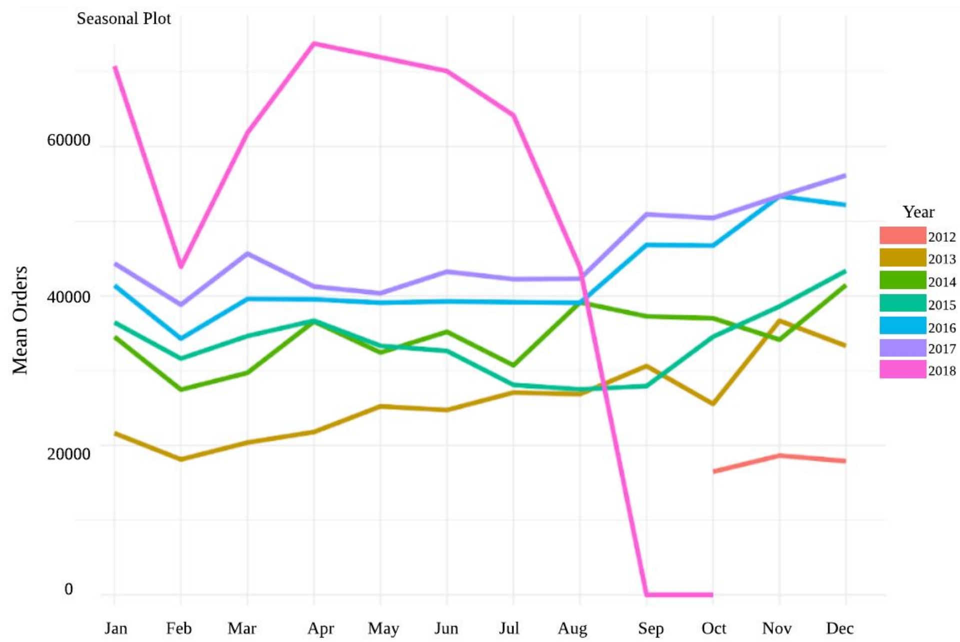

Figure 4. The bigger the "bump," the greater polarization of few materials to make up total demand. The straight line represents the scenario where materials are equally important.

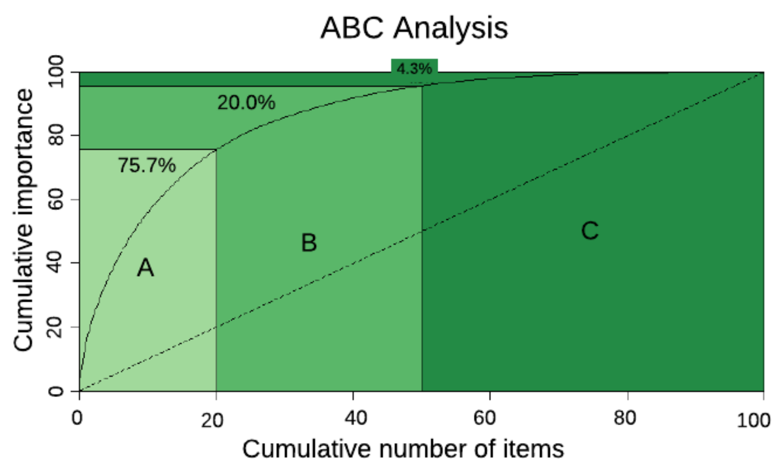

Figure 5. Just 20% of the series account for more than 52% of all naive/seasonal naive RMSE.

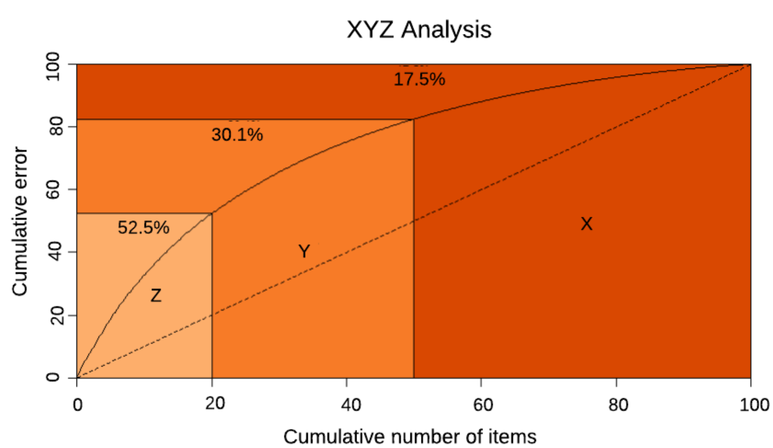

Figure 6. Grouping results.

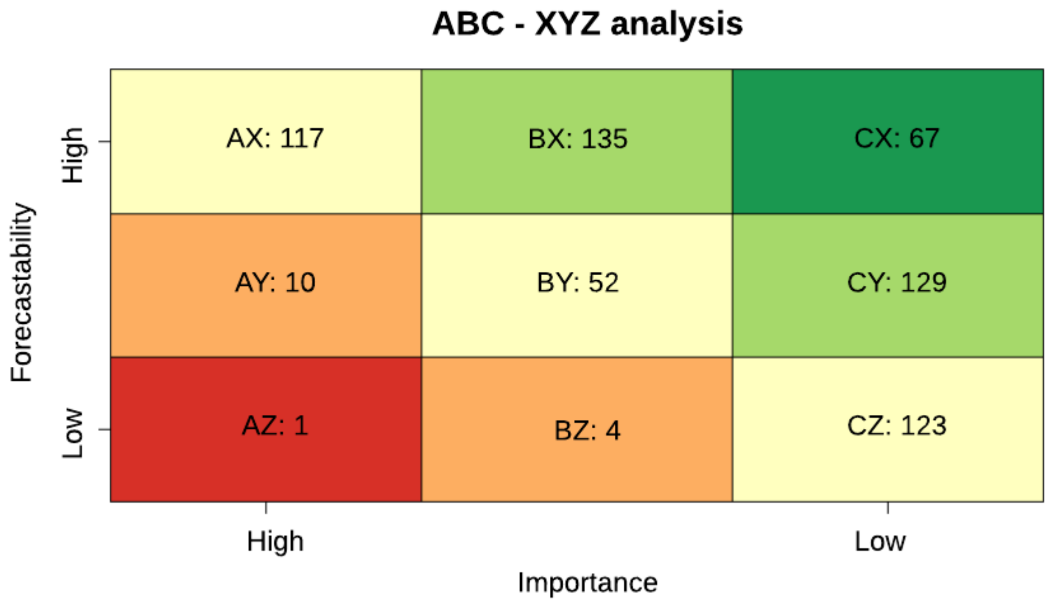

Figure 7. The best forecasting methods highlighted in red. Only three cells used non-naive forecasts.

Importance (% of total sales)

Ease of Forecasting

AX 117 series

Method	RMSE
auto.arima	136900
ets	135620
ensemble "aesz"	130710
naive	127509

BX 135 series

Method	RMSE
auto.arima	49974
ets	47859
ensemble "aesz"	45178
naive	42019

CX 67 series

Method	RMSE
auto.arima	11422
ets	12394
ensemble "aesz"	11105
naive	9887

AY 10 series

Method	RMSE
auto.arima	432475
ets	422401
ensemble "aesz"	420645
naive	517180
% better than naive	18.67%

BY 52 series

Method	RMSE
auto.arima	163618
ets	156720
ensemble "aesz"	153785
naive	150917

CY 129 series

Method	RMSE
auto.arima	21254
ets	22087
ensemble "aesz"	20455
naive	20074

AZ 1 series

Method	RMSE
auto.arima	727847
ets	727811
ensemble "aesz"	731814
naive	737292
% better than naive	1.29%

BZ 4 series

Method	RMSE
auto.arima	255848
ets	254859
ensemble "aesz"	257840
naive	257178
% better than naive	0.90%

CZ 123 series

Method	RMSE
auto.arima	20787
ets	21086
ensemble "aesz"	20615
naive	20018

Figure 8. Test performance using the ETS method. All forecasting methods produced similar boxplots for the test period, suggesting a systematic inability to predict demand in January, April, and August.

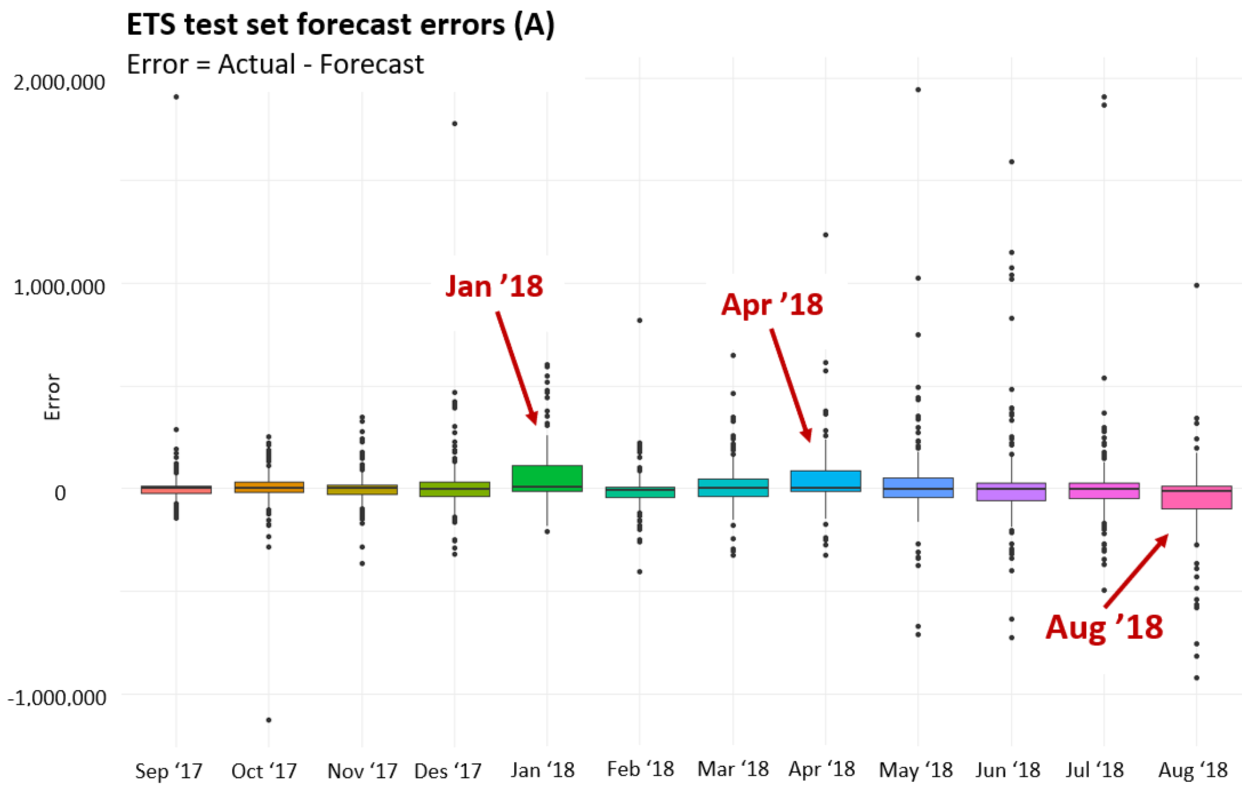

Figure 9. The Shiny-based Forecasting Platform.

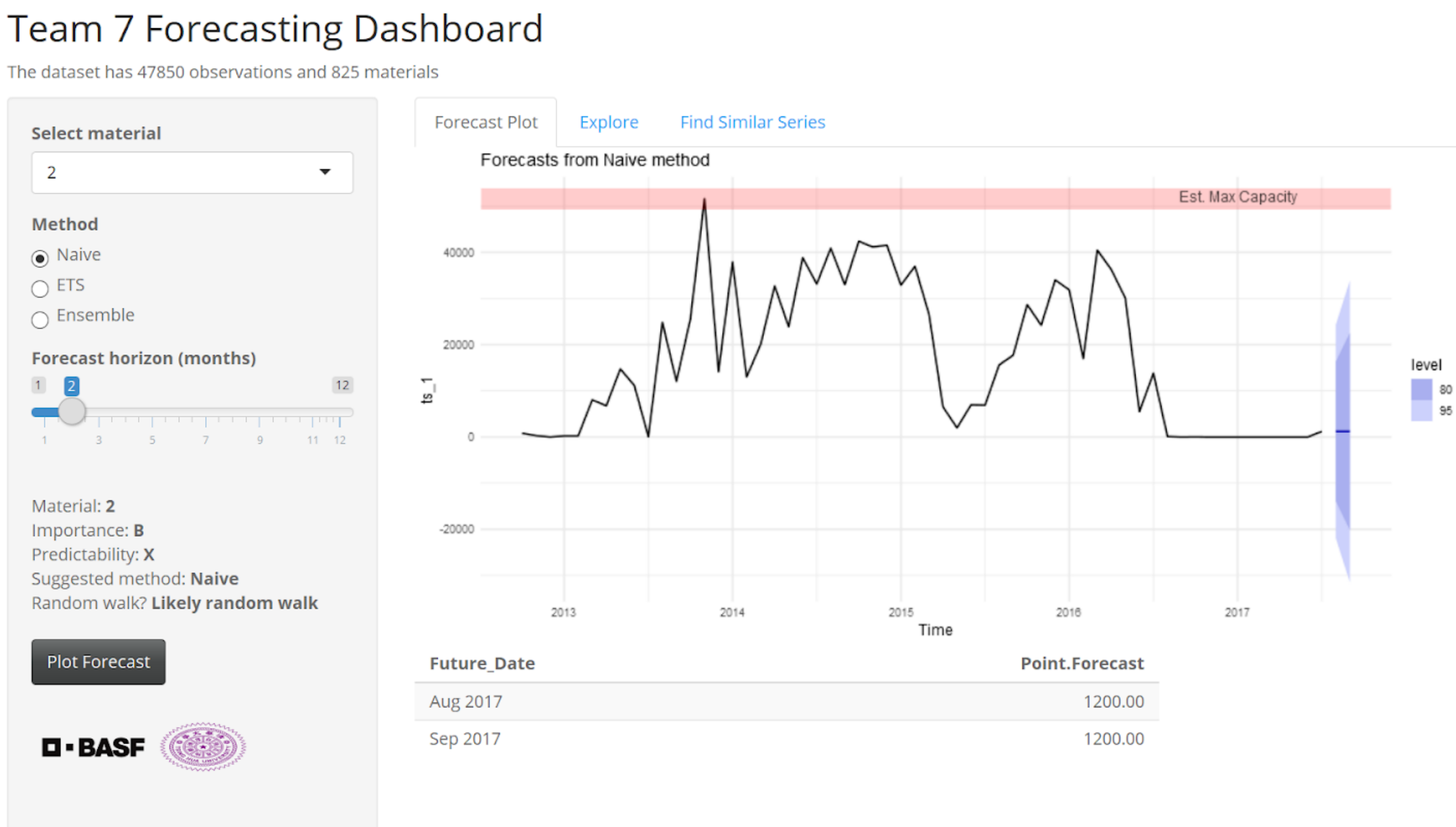

Figure 10. Proposed business process and solution application