

Using Analance to Predict Optimal Selling Price

The selling price of an item can be unpredictable given the variety of economic factors influencing this number. However, predicting the selling price of an item can be a big opportunity for retail organizations to capitalize on findings and improve sales. Further, using predictive models to ascertain product pricing can also help organizations move from an absolute pricing model to a dynamic pricing model.

Explore insights Analance found while analyzing data with selling price as the target variable using multiple econometric predictors.



BUSINESS CHALLENGE

The need to move from an absolute pricing strategy to a dynamic pricing strategy.



SOLUTIONING PROCESS AT A GLANCE

The process of statistical consulting and solutioning starts with a thorough understanding of the business challenge, its impact, and the data available for analysis. With this information, we arrive at a solution to mitigate or control the challenge, offer continued client support, and adjust models over time.



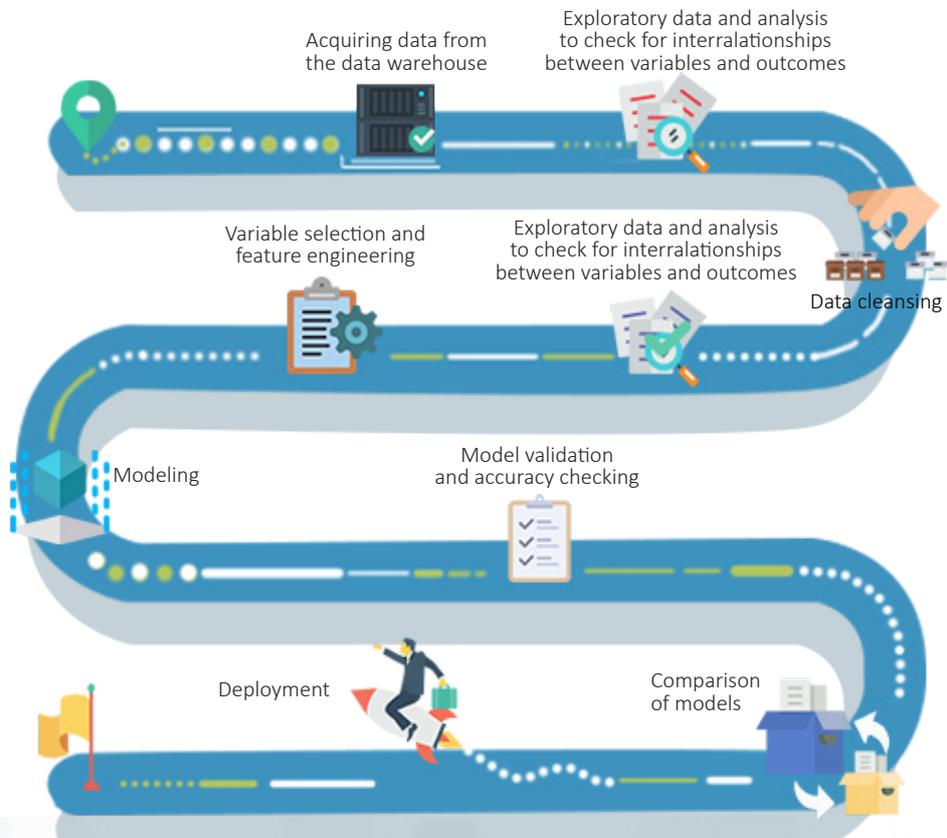
OUR PROCESS

A data dump was acquired and put through a stringent exploratory process before trying to correlate what information was available to solve the challenge at hand.

The following variables were available for analysis:

- **Supplier** – Supplier from who product was bought
- **Loc** – Location product was shipped from
- **Category** – Primary categorization of product
- **Sub category** – Secondary categorization
- **Stock on hand** – Number of products currently available in warehouse

PREDICTIVE MODELING STAGES



- **Stock on order** – Number of products to be bought to replenish inventory
- **Back order** – Number of items to be purchased for fulfilling customer request
- **Buy Price** – Cost to buy one item
- **MOQ** – Minimum order quantity
- **Lot Multiplier** – Nearest 10th round off of back order.
- **Sell Price** – Price one item of product was sold for at the point of sale



THE MODELING PROCESS

A Mathematical/Statistical/Econometric predictive model is a method of representing a variable of interest (outcome) as a function of other variables (predictors) with an assumed relationship between the outcome and predictors.

All variables available were studied to understand distributions. Data was cleaned by the means of handling outlying values, missing values, and looking for inter-relationships between predictors before looking to see if any data had a significant relationship with the outcome.

Outlying values were removed from the data after which missing values were replaced with column means (average value) for numeric data and column modes (most frequent label) for categorical or ordered data. The industry standard Box Plot was used to check for outliers and a tabulation or summary of data was used to check for missing values. If two or more predictors were inter-related (for instance height in cm, height in inches, height in feet), only one was used.

The industry standard metric VIF (Variance Inflation Factor) was used as a deciding factor. Predictors with VIF values above 6, indicating a significant inter-relationship with another predictor/s, were removed from the analysis.

The industry standard statistical test used to check for relationships between predictors and the outcome of interest is the Chi-Squared test of association. A p-value less than .05 indicates that the probability of association occurring in the population is less than 5% and is of no significance. Using this test for all predictor-outcome combinations helped in restricting the analysis for only those predictors that majorly influenced readmissions.

A further exhaustive exploratory analysis of data using graphs and advanced visualizations found that:

- Minimum order quantity is significantly related to selling price.
- Lot multiplier is significantly related to selling price.
- Supplier variety is significantly related to selling price.
- Product categorization is significantly related to selling price.
- Stock on hand is significantly related to selling price.
- Back order quantity is significantly related to selling price.
- Buying price is the strongest predictor of selling price.



CONCLUSION AND RECOMMENDATIONS

The buying price of the product came out to be the strongest predictor of selling price. The multiple linear regression model was able to predict 92.6% of data variation and the model was a statistically significant fit for the data. It was the overall winner of predictive ability as measured by the Root Mean Squared Error. Multiple linear regression also came out as the best, as validated by the RMSE, among the models chosen for the analysis and prediction.

Using models similar to those developed in this project, businesses have the ability to predict the optimal selling price of a product. They can transform from using an absolute pricing model to a dynamic pricing model where product prices are determined in real time. Dynamic pricing is a win-win for customers and sellers as the price sold or bought for is optimal for both parties. Further, the addition of other significant econometric factors can enhance and improve the predictability and accuracy of the model/s involved, further increasing the benefits for both the parties.



MODEL USED

The modeling process began with using the industry preferred choice (namely, Multiple Linear Regression). Other models, such as Decision Trees, Random Forest, Bayesian Regression, and Neural Networks, were also considered and results from them were analyzed to find the top performing model.

We decided which model was most effective using the following tabulation of the model against the industry standard metric RMSE.

MODEL	RMSE
Multiple linear regression	7.58
Bayesian regression	45.64
Decision tree	7.79
Random Forest	8.29
Neutral network	7.69

BEST MODEL