Predicting and Optimizing Transportation Costs for a National Retailer Under Tiered Rebate Constraints

Anurag Soni, Adarsh Goyal, Shan Lin, Surya Gundavarapu, Matthew A. Lanham
Purdue University Krannert School of Management
soni16@purdue.edu; goyal45@purdue.edu; lin882@purdue.edu; sgundava@purdue.edu; lanhamm@purdue.edu

Abstract

In this study we developed a transportation delivery decision-support system in collaboration with a high-end national retailer. After understanding the constraints of the business problem and the primary transportation providers tiered rebate policy, we framed this problem into an analytics problem. Our analytical solution predicts expected rebate rate at a week-level by framing it as a time-forecasting problem. Traditional ARIMA model are used to build features that served as inputs to machine learning models (e.g. Random Forest, Artificial Neural Network, and Deep Learning model) that we found led to even better forecasts of costs. As agreed upon with the partner, we focused on obtaining a model that achieved the lowest cross-validated Root Mean Squared Error (RMSE). Lastly, we developed a tool that would simulate various transportation scenarios to demonstrate how much of the delivery business could be allocated to other smaller carriers, while still ensuring the retailer would remain in a certain strategic rebate tier with their major vendor to minimize their overall yearly transportation costs.

Introduction

A national retailer transports millions of dollars' worth of their products via a popular package delivery company. The retailer gets rewarded with a significant annual rebate from the carrier based on the total shipment dollars over the course of a year. However, the decision to choose a delivery vendor is made on a weekly basis by the retailer. For every shipment, there is a trade-off between increasing the total sales with the popular carrier (and thus achieving a higher tiered discount later) or choosing an alternative lower-cost transportation vendor today. Therefore, we helped the retailer to figure out how much of their current and future deliveries should be done by each carrier, in order to reduce the other transportation costs for a specified planning horizon. Thus, knowing how to calculate applicable rebate levels to take advantage of reduced logistical costs, and determining whom will deliver what and when, is an important problem for supply chain departments.

Methodology

Data Source
The database included two tables, one detailing retailer's record of transactions and the other detailing the package company's record of transactions. The retailer's record of transactions had all of their shipments while only a subset of them were rebatable. The business rules applicable to determination of the eligibility of rebate helped us isolate the necessary records.

Analysis Pipeline
As with any dataset, data cleanup was essential for validating and building models on this dataset since the retailer's record of transactions had several missing values that were to be imputed based on factors like Tracking Number, Shipment Number and the state of origin for the package. These were communicated by the retailer since the eligibility of rebate depends on the origin country of shipment and the type of shipment.

Once reasonable completeness was achieved in the dataset the two datasets were validated against each other to ensure they were brought to the same level of accuracy. We computed the rebate amount from both datasets to see if they matched. We found that the retailer considered both US and Non-US sender in their calculation and thus overestimated the rebate amount. Judging from the skewed distribution of the US and Non-US, we speculated that 95% of the missing values belonged to the US Sender, while the remaining went to Non-US sender, which would tally the rebate amount of the delivery company.

Amount Distribution by Sender Country

Amount by Sender Country (With Null)
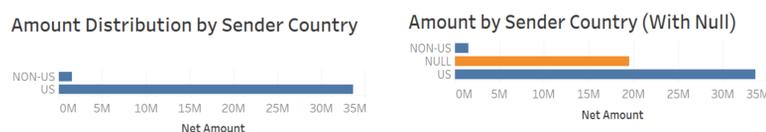
Exploratory Data Analysis
The delivery service requested by the retailer was mostly for Transportation (SHP) while Return (RTN) and Post Delivery Adjustments (ADJ) service charge remained less than 120,000 USD. Among SHP service, the trend of cargo manifest parallels with that of net charge, weight and quantity. Import, World Wide Service and other SHP did not fluctuated as the way of cargo manifest. Delivery Amount soared in January and December around Christmas season. Because the delivery amounts were in the millions of dollars, we took logarithms of the values before proceeding to build models.

Model Building and Comparison
To forecast delivery throughout a year, we tested several models from Moving Average method and Time Series Analysis and compared their fitness to the trend.
I. **Simple Moving Average** is sum of the data points over a given period and then divided by the number of periods to eliminate strong fluctuations.
II. **Exponential Weighted Moving Average** is a hybrid approach between exponential smoothing and moving average. Data points from older data are weighted differently from data points near the beginning of the data set.
III. **Autoregressive Integrated Moving Average (ARIMA)** model is a generalization of an autoregressive moving average model which is applied to non-stationary data. Differencing step can be applied one or more times to eliminate the non-stationarity.

Optimization
We need to acknowledge that the delivery company that is in use might not always be the cheapest option and that the volume based shipment rebates justified the increased costs. But now that the delivery company is negotiating for a flat rebate amount, it might be wise to direct some of our shipments to alternate cheaper carriers if possible. We assumed the availability of a x% cheaper alternative and built an optimization model that determines what percentage of the shipments on average can be diverted for the alternative shipping company.

Results

Simple Moving Average

Rolling Mean & Standard Deviation

Exponential Weighted Moving Average

Rolling Mean & Standard Deviation

Auto Regressive Models
On the adjoining graphs, we built three Auto Regressive Integrated Moving Average models with the following (p,d,q) parameters. The Residual Sum of Squares(RSS) is an important criterion to evaluate the performance of each model. The lower the number, the more accurate the model is.

ARIMA 1 (RSS: 4.0395) (2,1,0)   ARIMA 2 (RSS: 3.7260) (0,1,2)

Combined ARIMA (RSS: 3.4101)(2,1,2)

Model Comparison
ARIMA(2,1,2) is the best model with the lowest RSS. While the simple moving average appeared to track demand values closely than ARIMA did, the rolling standard deviation was not constant, making the model unreliable over test data.

Optimization Model

The variables that we were interested in knowing was what proportion of the transactions can our retailer afford to divert to a different cheaper delivery company in a way that would not downgrade their current account standing (and thereby any rebate they might qualify for) with their existing delivery company while saving on the annual shipping costs.

Objective Function
Min {TC} = TCORC – {TCORC * RP} + TCOCC

Constraints
POSTCC >= 0
POSTCC <= 1
POSTCC + POSTRC = 1

Parameters
TC = Total Annual Cost of Shipping
TCORC = Total Cost of Regular Carrier
RP = Rebate Percentage
TCOCC = Total Cost of Cheaper Carrier
POSTCC = Percentage of the shipments to Cheaper Carrier
POSTRC = Percentage of Shipments to Regular Carrier
PC = The factor by which the cheaper carrier is cheaper.

We also included other constraints in our model such as the maximum order quantity (O) that can be placed for every part at a time. This quantity will depend on the size of warehouse and the dimensions of the part.

Results
Assuming a constant savings factor, we notice that we are better off always diverting some of our shipments to a third party carrier. Further, if the factor by which the cheaper carrier is cheaper is changed into a random variable, we see that a policy of diverting about 9.82% of the shipments on average always is the most optimal method of bringing down total yearly costs. We note that the simulation ran for 15 minutes using 2000 iterations.

Algorithm
We used RiskOptimizer that was developed by Palisade software which uses a Monte Carlo simulation to identify the best input parameters to achieve the desired results.
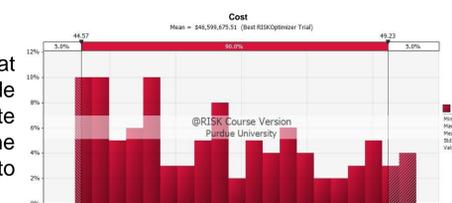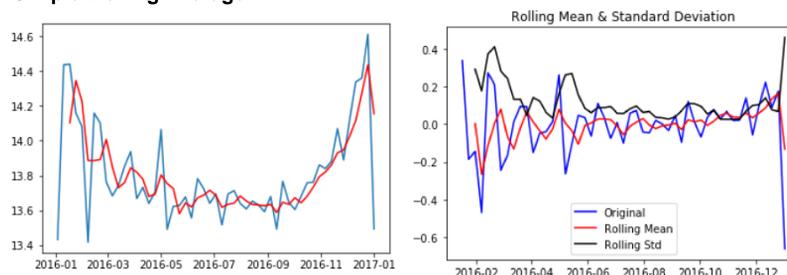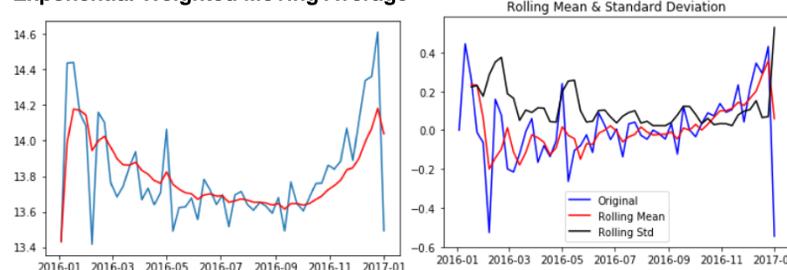

Conclusions

By applying our model, the retailer can have higher autonomy in requesting delivery while retaining rebate benefits from the major delivery company. According to the time series analysis, we found there was weekly seasonality throughout the year. Based on the trend, we ran the optimization model to determine the ratio for switching to cheaper delivery service. We suggested that the retailer should divest about 9.82% of the shipments to cheaper alternatives, which can lead to a saving up to $5 million yearly! To extend this analysis, we would like to incorporate more features such as delivery category, parcel details, geographic information, and historical transaction for future improvement and implementation.

Acknowledgements