Delivery Miss Root Cause Analysis Using Tableau

Background

This case begins from the underperforming supply chain KPI - IS DEA, which refers to the package delivery on-time rate. Comparing the missed units of each city, we find that Beijing and Shanghai are major contributors to total missed units. Therefore, to seek the causes of high missed rate of packages in Beijing and Shanghai, we use Tableau to generate a story interpreting the relationships among missed units, date, classes of carriers and detail reasons.

Root-cause Analysis

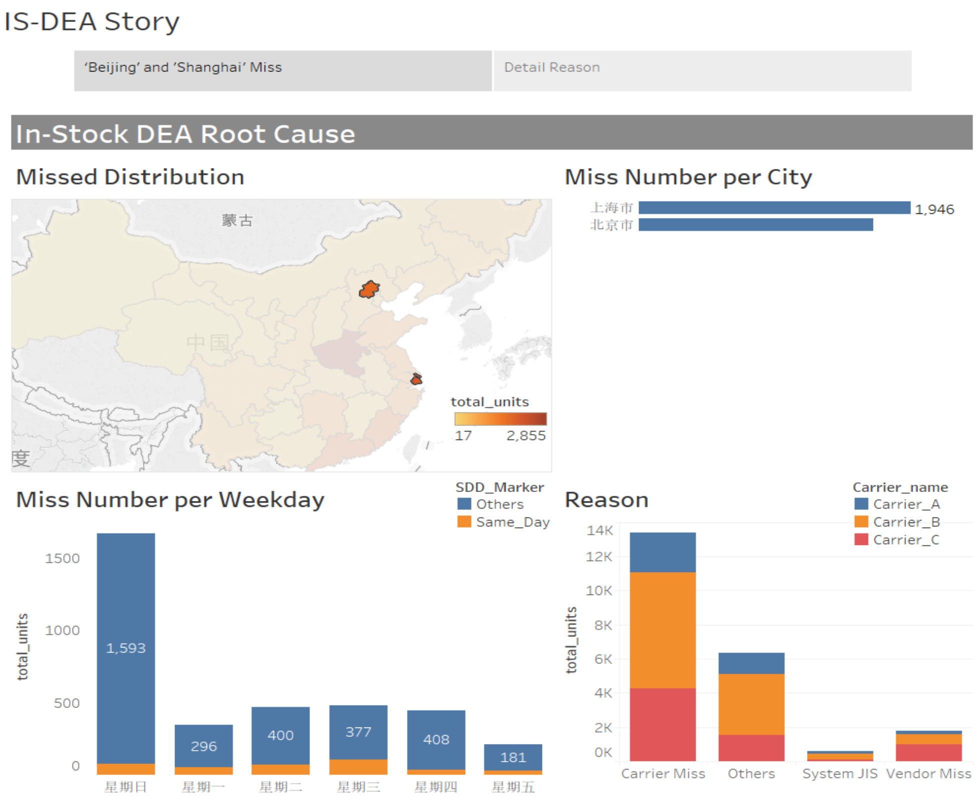
According to the "Beijing and Shanghai miss" dashboard, we find that there is a prominent peak of missed units of both cities on weekend, especially on Sunday, throughout the whole week. This might be because the number of orders placed on weekends are times of that on weekdays, putting excessive pressure on package delivery on Sunday. This leads to lots of missed units as a result, therefore driving down the on-time delivery rate (DEA). Moreover, DEA Root Cause Analysis Dashboard (See Dashboard 1) shows that the major reason of missed units is "Carrier Miss", which means that carriers did not delivery packages on-time. Carrier-miss Reason Distribution (Dashboard 2) indicates that "Carrier Miss" is completely due to the carrier fault. This further proves that the limit capacity of last mile carriers is the root cause of the delivery miss.

Business Proposal

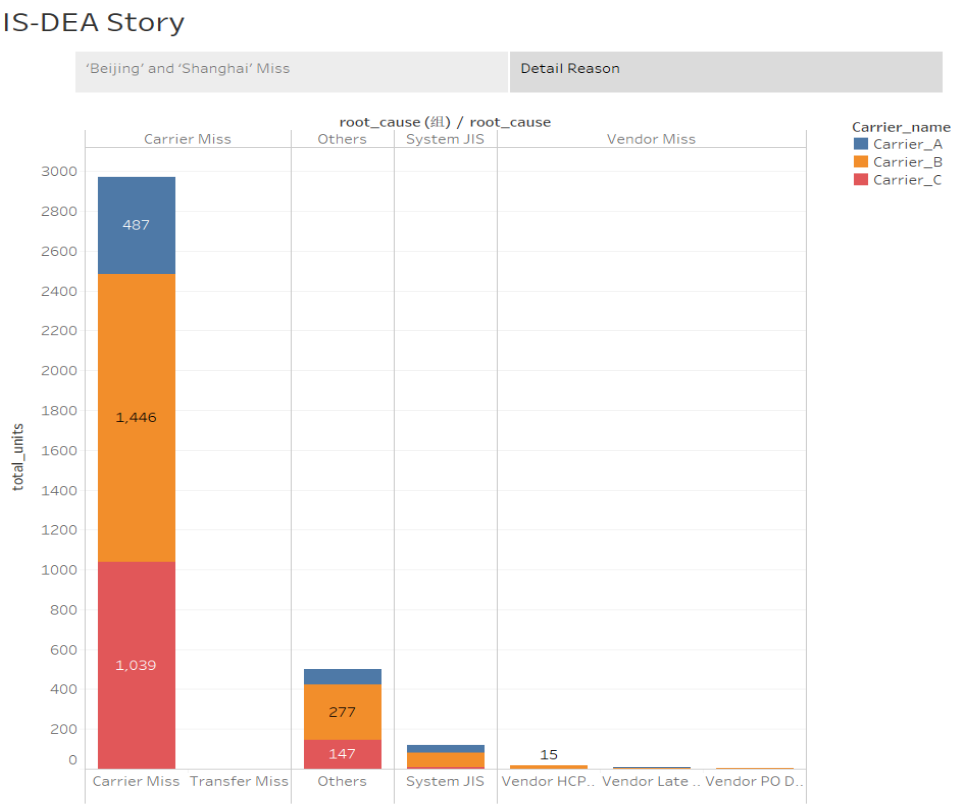
Based on this findings, we developed several proposals to increase the delivery on-time rate:

(1) Develop last mile delivery capacity planning plan and daily monitor mechanism to obtain the visibility of carrier performance;

(2) Adjust delivery labor between weekdays and weekend and add more carrier delivery labors on weekends to ensure the on-time delivery.



Dashboard 1: DEA Root Cause Analysis



Dashboard 2: Carrier-miss Reason Distribution

Product Attribute Forecast Using Machine Learning

Background

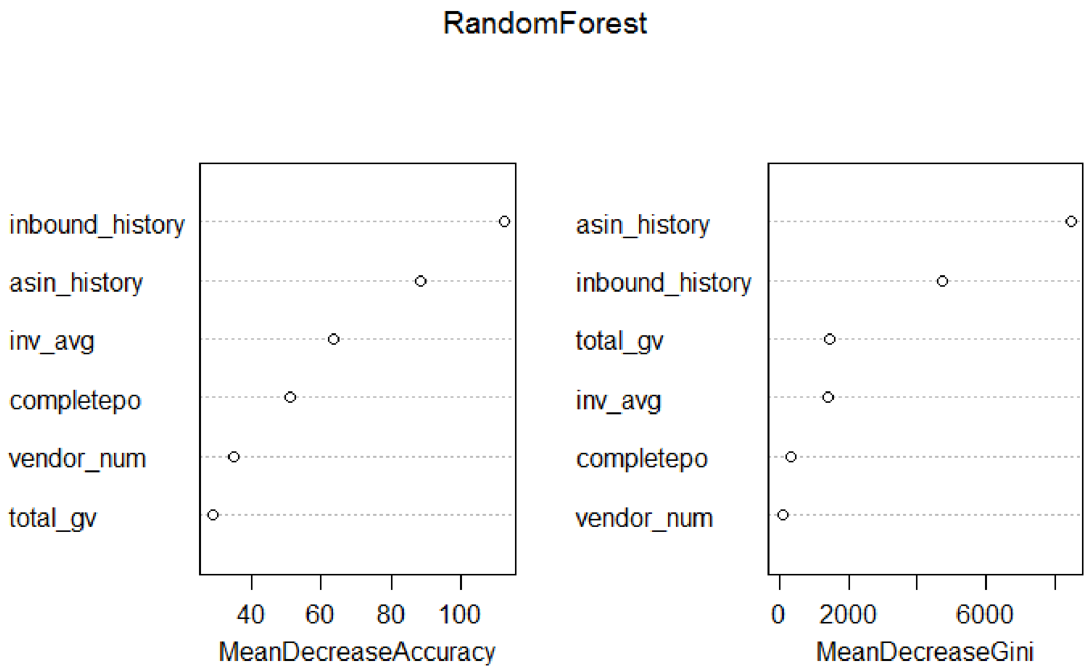
This case results from the underperforming NIS DEA. During the daily KPI monitor, we find that the Not In-stock Delivery Estimated Accuracy (NIS DEA) missed the goal (93.5%) very frequently. After diving deep into this problem, we found that in most of the cases, the attribute of the ASINs should have been set to “OB” (obsolete), which means the ASIN would never be purchased from the vendors, while the system still shows the ASIN are procurable and keeps buy-boxes on website. Once customers place orders on these ASIN, the order will never be fulfilled, leading the orders to miss the promises. Therefore, the NIS DEA is underperforming. In general, the buyable attribute of the ASINs are updated by in-stock managers manually, which is time-consuming and inefficient. With the development of big data, machine learning modelling technics are widely utilized in the IT industry to improve business performance and work efficiency. This project aims to use the machine learning methodologies to solve the misclassification of the ASINs automatically and improve the NIS DEA performance.

Data Preparation

Firstly, using the methods of descriptive statistic and hypothesis testing, we find some plausible significant differences of “label” in “asin_history”, “inbound_history”, “completepo”, “inv_avg”, “total_gv”, “vendor_num” and “confirmpo”, which are features originally selected. After removing the correlations between features, we finally chose six features in the modelling. Detailed data processing is referred to addition file on my github website https://github.com/GYHenryTT/NIS-DEA-Model.

Modelling

We choose Random-Forest as our model to make prediction. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.



Results

After parameter testing and feature reselecting, the prediction of our model reaches the sensitivity of 86.85% and specificity of 89.43%, which are much more accurate comparing to manual work. Repetitions of the same data provide the same result. By utilizing this machine learning model, delivery on-time rate could increase by 312 bps.

Certainly, our model has lots of room for improvement in reducing overfitting and balancing sensitivity and specificity. Some other method can be used like data amplification and PCA.