Chapter 15
Data Analysis of Retailer Orders to Improve Order Distribution

Michelle L.F. Cheong and Murphy Choy

Abstract Our paper attempts to improve the order distribution for a logistics service provider who accepts order from retailers for fast moving consumer goods. Due to the fluctuations in orders on a day to day basis, the logistics provider will need the maximum number of trucks to cater for the maximum order day, resulting in idle trucks on other days. By performing data analysis of the orders from the retailers, the inventory ordering policy of these retailers can be inferred and new order intervals proposed to smooth out the number of orders, so as to reduce the total number of trucks needed. An average of 20 % reduction of the total number of trips made can be achieved. Complementing the proposed order intervals, the corresponding new proposed order size is computed using moving average from historical order sizes, and shown to satisfy the retailers' capacity constraints within reasonable limits. We have successfully demonstrated how insights can be obtained and new solutions can be proposed by integrating data analytics with decision analytics, to reduce distribution cost for a logistics company.

Keywords Data analytics • Decision analytics • Order distribution • Inventory policy inference

15.1 Introduction

Third party logistics companies (3PL) are often faced with the challenges of managing the supply chain efficiency for their clients. For a 3PL who acts as the middle man for the distribution of goods for the brand owner to the retailers, several key performance indices (KPIs) are tracked as part of the service level agreement with their clients. One such KPI is the on-time delivery of orders to the retailers. Late deliveries will affect the sales of the products and may even affect market share of the product.

M.L.F. Cheong (✉) • M. Choy (✉)
School of Information Systems, Singapore Management University,
80 Stamford Road, Singapore 178902, Singapore
e-mail: michcheong@smu.edu.sg; goladin@gmail.com

© Springer International Publishing Switzerland 2015

L.S. Iyer, D.J. Power (eds.), *Reshaping Society through Analytics, Collaboration, and Decision Support*, Annals of Information Systems 18,
DOI 10.1007/978-3-319-11575-7_15

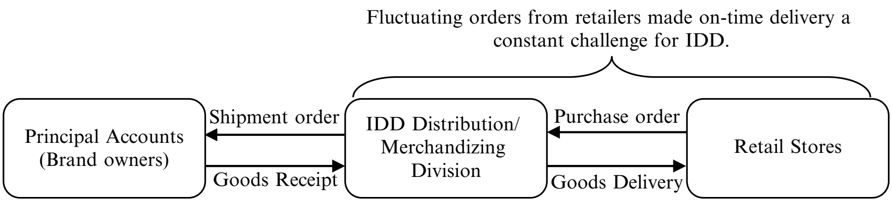

Fig. 15.1 IDD as middle-man for sales & distribution of goods

IDD is a leading integrated distribution and logistics services provider with its headquarter in Hong Kong. IDD provides a full suite of integrated distribution services covering Logistics, Distribution, Manufacturing and International Transportation. The Distribution/Merchandising division plays the middle man role (see Fig. 15.1) in distributing products for their principal accounts (brand owners) to retail stores. Products include food items such as corn flakes and chocolates, and health and beauty items such as toothpaste and shampoo.

The division was often faced with fluctuating orders from the retailers and it did not know how to best manage these fluctuations except to try its best to deliver the orders on time, and face possible penalties from the clients in case of underperforming the contracted KPI. The division wished to understand the fluctuations in orders through analysis of data captured in their IT systems. Through proper data analysis, the division hoped to gain insights on the order behavior of the retailers and propose alternative solution to achieve a win-win situation for the retailers and itself.

15.2 Literature Review

Previous work done on the fulfillment of orders from the upstream supplier or manufacturer to the downstream retailers in a two-stage supply chain under stochastic demand are often focused on sharing of Point-of-Sales (POS) information and implementing Vendor Management Inventory (VMI) so that the supplier can supply the right quantity at the right time to the retailers.

Many papers have highlighted the benefits of information sharing including reduced inventory, daily administration costs and delivery costs. Lee et al. (2000) modeled a two-stage supply chain with one manufacturer and one retailer, to quantify the benefits of information sharing and to identify the drivers that have significant impacts. They showed that manufacturer can obtain larger reductions in average inventory and average cost when the underlying demand is highly correlated over time, highly variable, or when the lead time is long. However, Raghunathan (2001) showed that sharing of demand information is of limited value when the parameters of the demand process are known to both parties, under AR(1) demand with a nonnegative autocorrelation coefficient. The reason is that the manufacturer can forecast the demand information shared by the retailer with a high degree of accuracy using retailer order history, rather than using only the most recent order from the retailer to forecast the future orders. The accuracy increases monotonically with each subsequent time period. Consequently, the value of information shared by the retailer decreases monotonically with each time period, converging to zero in the limit. Thus, if the manufacturer uses its available information intelligently, there is no need to invest in inter-organizational systems for information sharing purposes.

Yu et al. (2002) also modeled the two-stage supply chain of a beauty product supplier and a retail store. They found that increasing information sharing will lead to Pareto improvement (at least one member in the supply chain is better off and no one is worst off) in the performance of the entire supply chain. Cheng and Wu (2005) extended the two-stage supply chain to consider multiple retailers and allowed correlation of orders to be negative, an extension from Yu et al. (2002). They introduced three different levels of information sharing from level 1 with only knowing retailers' order information; to level 2 with knowing both the retailers' order and customer demand information; and finally to level 3 with real-time information of customer demand through EDI. The optimal inventory policy under each of them was derived. Finally, they showed that both the inventory level and expected cost of the manufacturer decrease with an increase in the level of information sharing. However, they also showed that there was no difference between the inventory level and expected cost of the manufacturer for levels 2 and 3 of information sharing. This implied that there was no need for real-time sharing of demand information or VMI implementation for a two-stage supply chain.

Steckel et al. (2004) stated that whether the sharing of POS information is beneficial or not depends on the nature of the demand pattern represented by the POS information. If the demand pattern conveys continual change in ultimate downstream customer demand, the POS information can in fact distract the upstream decision maker from the more relevant information available from the orders placed by the downstream agent and the supply line. Gaur et al. (2005) extended the results of Raghunathan (2001) to cases in which demand is (AR(p), $p>1$) or (ARMA(p, q), $p>1$, $q>1$). They found that the value of sharing demand information in a supply chain depends on the time series structure of the demand process. When both the demand process and the resulting order process are invertible, demand can be inferred by the manufacturer without requiring further information from the retailer. When demand is invertible but the resulting order process is not, sharing demand information is necessary. They proposed that the demand process is inferable from retailer's order quantity, if the upstream manufacturer's forecast of demand obtained by observing retailer's order quantity, converges almost surely to the actual realization of the demand as time t tends to infinity.

Williams and Waller (2010) compared the order forecasts for the highest echelon in a three-stage supply chain, using POS data versus using order history for cereal, canned soup and yogurt. Their results show that order forecast accuracy depends largely on the product characteristics (seasonal or not) and forecast horizon. In general, POS data produces a better forecast. However, for canned soup which is a seasonal product, POS data did not outperform order history for short term forecasting; whereas and for yogurt which is a short-life span product, POS data performs almost the same as order history.

In our case, IDD did not have any Point-of-Sales (POS) data or shared demand information from the retailers, thus IDD was unable to know or infer the actual demand. Instead, we hope to perform data analysis on historical order information to infer the inventory policies of downstream retailers, and to propose new order intervals and order sizes from historical order data to reduce distribution cost. By playing a proactive role in recommending order interval and the corresponding order size, the retailers need not place order actively, and IDD can better plan distribution to reduce cost.

We could only find two pieces of prior work which have similar objectives like ours to use data analysis to improve supply chain performance. Hausman et al. (1973) analyzed the demand data for 126 women's sportswear over 18 months to obtain three different data-generating processes, (1) ratios of successive forecasts are distributed lognormally; (2) ratios of successive forecasts are distributed as t (Student); and (3) actual demands during unequal time periods are distributed as negative binomial. They concluded that negative binomial was most closely representing the underlying process and simple to adapt to a decision model. Johnston et al. (2003) examined the order size of customers to improve the supply chain. The specific activity mentioned in the paper was that items with intermittent demand, the size of customer orders is required to produce an unbiased estimate of the demand. Also the knowledge of the distribution of demand is important for setting the maximum and minimum stock levels. Both works did not continue to use results of the analysis to make further supply chain related decisions. We think that we are the first to integrate data analytics and decision analytics, where historical data was analyzed to obtain insights to support decision making to improve the supply chain.

Our paper is organized as follow. Section 15.3 will describe the data analysis process to infer the inventory policy of the retailers. Based on the results obtained in Sect. 15.3, we propose a distribution strategy in Sect. 15.4. Based on the proposed distribution strategy in Sects. 15.4, 15.5 and 15.6 will compute the new proposed order interval and order sizes respectively. Section 15.7 aims to assess if the new proposed order sizes will violate retailers' capacity constraint. Section 15.8 compares the number of delivery trips based on the proposed strategy with historical data. Finally, Sect. 15.9 provides the conclusions.

15.3 Data Analysis of Retailer Orders to Infer Inventory Policy

The two sets of data (see Appendix) used for analysis were *Logistic data* and *Store Location data* for a cornflakes product (with each different packaging of the same product represented as a different SKU Code). *Logistic data* provided information on Retailer (identified by CustomerNo), SKU Code, SKU Description, Order Date, Order Quantity, Delivery Date, Delivery Status, Shipped Date, and Shipped Quantity; while *Store Location data* provided the Store Code (identified by Shiptocode), Store Name and Location in geo-information format. In total, there are 326 unique retailers, 191 unique SKU Codes, and 2,681 order records.

With only the historical purchase order information, the initial analysis aimed to categorize the retailers into two possible inventory policies namely, Periodic Review (PR) and Continuous Review (CR). The following assumptions were made:

1. The raw *Logistic data* was reconfigured into a new table with the number of orders for each day of the week (Monday, Tuesday, etc.) for each retailer using Order Date, regardless of the SKU item and order size.
2. Since the objective was to understand the ordering behavior of the retailers, the actual SKU item ordered is immaterial. The analysis result in the appendix supported that the ordering behavior of the retailer was independent of the SKU item ordered.
3. The order size is determined when the retailer has decided to place an order, so it is not the cause for placing order, but rather the result of placing order. Thus, when analyzing the ordering behavior, the order size was not considered. However, the order size would be computed after the order policy and order interval were determined.
4. Without loss of generality, we assumed zero delivery lead time, that is, Delivery Date is the same as Order Date. From the actual data, Delivery Date could be different from Order Date due to planned or unplanned delays.

   (a) Planned delay is usually represented by a fixed delivery lead time T days. As we are only concerned with the delivery of the orders instead of the inventory levels of IDD and the retailers, we can apply the analysis results to positive lead time T by simply shifting the results by T days.
   (b) Unplanned delay is usually due to operational inefficiencies with too many causes, and will not be included as part of the analysis.

5. Only retailers with at least ten orders were included in the analysis to ensure validity of the data analysis.

Based on the assumptions, the data were reconfigured according to day of week j. To explain the data analysis performed, we define the following notations:

- i = Retailer index number, $i = 1$ to I
- j = Day of week corresponding to the calendar date. $j = 1$ to 7, where 1 = Monday, 2 = Tuesday and so on. Note that there might be several orders by the same retailer i on different calendar dates which correspond to the same day of week j.
- $\bar{O}_{ij}$ = Set of orders by retailer i on day of week j
- M_{ij} = Number of orders by retailer i on day of week j, where $M_{ij} = |\bar{O}_{ij}| \geq 0$
- $\bar{R}_i$ = Set of all the orders placed by retailer i.

$$\bar{R}_i = \bar{O}_{i1} \cup \bar{O}_{i2} \cup \bar{O}_{i3} \cdots \cup \bar{O}_{i7}$$

- N_i = Number of orders by retailer i, where $N_i = |\bar{R}_i| \geq 0$
- X_{ij} = Ratio of the number of orders placed by retailer i on day of week j and the total number of orders placed by retailer i.

$$X_{ij} = \frac{|\bar{O}_{ij}|}{|\bar{R}_i|} = \frac{M_{ij}}{N_i}$$

- Y_{iw} = Sum of any two ratios X_{ij} of retailer i for any 2 days of week j, where $w = 1$ to ${}^7C_2$ represents the combination index number and there are ${}^7C_2 = 21$ unique combinations.

The two possible inventory policies considered are:

1. *Periodic Review (PR)* – This policy refers to reviewing the inventory level after a fixed interval period and placing the order quantity sufficient to fill up to the order-up-to level. Usually, small retailers who cannot afford the time and effort to review their inventory on a continuous basis will adopt the Periodic Review Policy. By analyzing the percentage of orders on each day of the week, we could infer the day which the retailer usually placed order.

Rule 1: Periodic Review with Single Dominant Day
If there exist a $Max_j(X_{ij}) > X_{cut}$*, then retailer i is assumed to employ the periodic review policy on the dominant order day j, with a confidence interval of* $(1-\alpha)\%$ *and level of significance of* $\alpha\%$.

In our paper, we have selected X_{cut} = 40 % and state that if there exist a $\text{Max}_j(X_{ij}) > 40\%$, then retailer i is assumed to employ the periodic review policy on the dominant order day j, with more than 93.48 % confidence that the observation did not occur by chance with level of significance less than 6.52 %. Refer to Appendix for proof.

Rule 2: Periodic Review on 2 Days, But with Single Dominant Day
If there exist a $Max(Y_{iw}) > Y_{cut}$*, then retailer i is assumed to employ the periodic review policy on 2 days of the week represented by the combination index w, with a confidence interval of* $(1-\alpha)\%$ *and level of significance of* $\alpha\%$*. For this combination w, if* $X_{iq} > X_{ir}$ *where q and r are the days of week represented by combination w, then q will be the dominant order day.*

In our paper, we have selected Y_{cut} = 60 % and state that if there exist a $Max(\text{Y}_{\text{iw}}) > 60\%$, then retailer i is assumed to employ the periodic review policy on 2 days of the week represented by the combination index w, with more than 97.67 % confidence that the observation did not occur by chance with level of significance less than 2.33 %. For this combination w, if $\text{X}_{\text{iq}} > \text{X}_{\text{ir}}$ where q and r are the days of week represented by combination w, then q will be the dominant order day.

2. *Continuous Review (CR)* – This policy refers to continuously reviewing the inventory level and order only when the inventory level reaches the reorder point, regardless of the day of week. Usually, larger retailers who have a warehouse and inventory management team can afford to continuously review their inventory and adopt the Continuous Review policy. Similarly, by analyzing the percentage of the total number of orders on each day of the week, we could infer that the retailers who adopted the Continuous Review policy did not have a specific day to place order, so their orders were evenly spread over 7 days.

Figure 15.2 below shows two typical retailers. The blue histogram shows a Periodic Review retailer who placed about 90 % of his orders on Monday, while the red histogram shows a Continuous Review retailer who placed orders evenly on every day of the week.

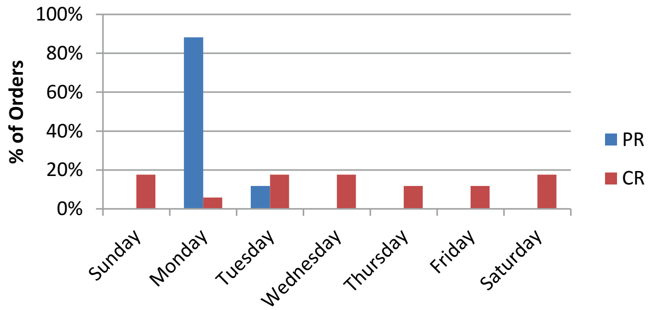

Fig. 15.2 Example of periodic review and continuous review policy retailers

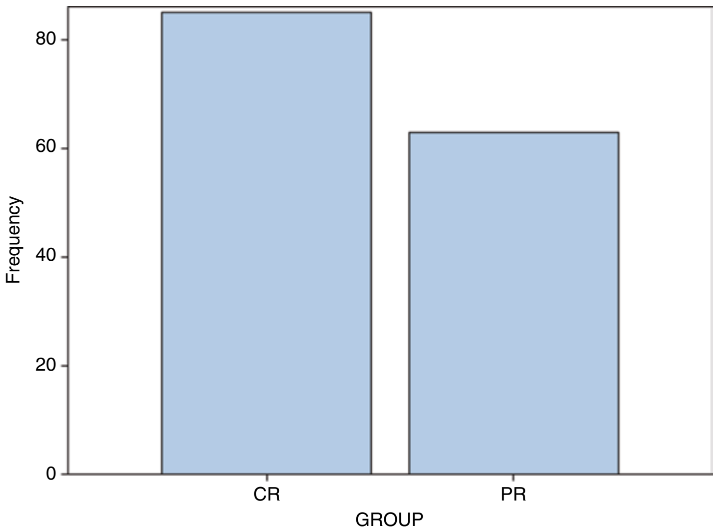

Fig. 15.3 Frequency count of retailers for different inventory policies

To compute the frequency counts for the different number of retailers for each inventory ordering policy, we adopt the following notations:

- $\bar{P}$ = Set of retailers i who employed the periodic review policy based on Rule 1 and Rule 2

$$\bar{P} = \left\{ i \,|\, \exists\, Max_j \left(X_{ij} \right) > 0.4 \, or \, Y_{iw} > 0.6 \right\} = \bar{C}'$$

- $\bar{C}$ = Set of retailers i who employed the continuous review policy

$$\bar{C} = \left\{ i \,|\, \mathrm{v}\, Max_j \left(X_{ij} \right) > 0.4 \, or \, Y_{iw} > 0.6 \right\}$$

Our result in Fig. 15.3 shows that most of the retailers employed the Continuous Review policy, that is, $|\bar{C}| > |\bar{P}|$. Since these Continuous Review policy retailers accounted for the bigger portion of the business and orders from them are rather even, they will form the base load of orders for distribution requiring an almost fixed number of trucks, while the orders from the Periodic Review policy customers will be added on top of the base load, needing the additional trucks.

15.4 Distribution Planning Strategy

After establishing the number of retailer adopting either the Continuous Review (CR) or Periodic Review (PR) policy, we continue to understand how the orders from these retailers distribute across the different days of the week. As every retailer can place order for more than one product, we will define a retailer-product combination since we are only interested to know on which day of the week the retailers place their orders and not what products they order. Each retailer-product combination refers to a particular retailer ordering a particular product. By splitting these retailer-product combination by retailers, Fig. 15.4 shows the distribution for Continuous Review policy retailers (blue bars) which appears to be evenly spread out from Monday to Friday, while the distribution for Periodic Review policy retailers (red bars) has highs and lows from Monday to Friday. This prompted that the fluctuations in orders were caused primarily by the Periodic Review policy retailers. Such fluctuations of orders day to day, will result in needing different number of trucks for each day.

Focusing only on those retailers who adopt the Periodic Review policy, and based on their top order day, Fig. 15.5 shows that the maximum number of orders occurred on Monday, and this number was about twice that of Tuesday, the second highest order day. To ensure on time deliveries on Monday, IDD had no choice but to maintain a large fleet of trucks. However, on the other days of the week (Tuesday, Wednesday, etc.), a smaller number of trucks will be sufficient to complete all deliveries. This will result in excessive number of idle trucks on the other days of the week.

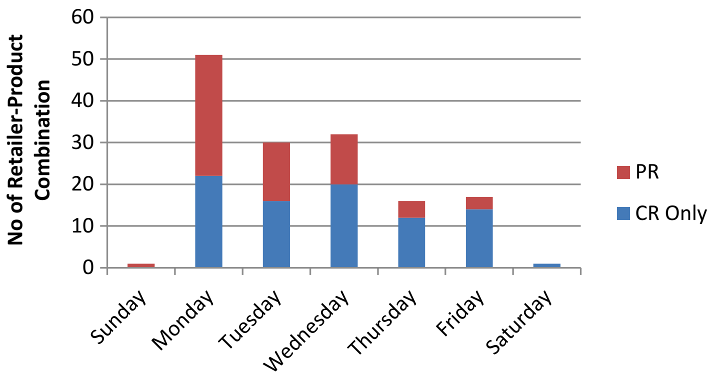

Fig. 15.4 Distribution of retailer-product combination for different day of the week

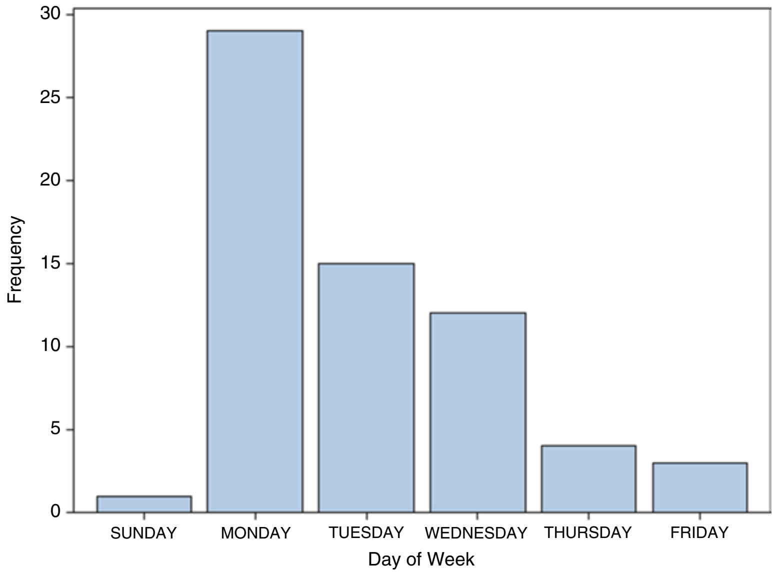

Fig. 15.5 Order frequency for each day of the week for periodic review policy retailers

For the retailers i in set $\bar{P}$,

- $\bar{P}_j$ = Set of retailers i who employed the periodic review policy on dominant day j
- $\bar{P}_1 > \bar{P}_2 > \bar{P}_3 > \bar{P}_4 > \bar{P}_5 > \bar{P}_7$. Note that there are no retailers who employed periodic review policy on Saturday.
- $\bar{P}_1 \sim 2\bar{P}_2$

IDD hoped to even out the distribution for every day of the week, so that the number of trucks used for distribution could be reduced. Since the fluctuations were caused by the Periodic Review policy retailers, the improved distribution plan would only consider smoothing out the orders from these retailers.

IDD can propose to split the retailers for Monday into two groups, each with an order interval of 14 days, instead of 7 days. Group 1 will receive goods on every 1st and 3rd Monday, while Group 2 will receive goods on every 2nd and 4th Monday. The cycle then repeats for 52 weeks in a year. For the other days of the weeks, the retailers will receive goods once a week only on their dominant day.

By carefully allocating retailers belonging to Monday into two groups, IDD can reduce the number of deliveries required for Monday, and thus reducing the total number of trucks required for the entire delivery operations. The allocation of retailers into the two groups (ideally about 50 % of Monday retailers in each group) will depend on their geographical location to minimize the travel distances. Based on the geographical location of the Monday retailers in Fig. 15.6, the Monday PR retailers are divided into five groups in (i) Kowloon, (ii) New World territory region, (iii) Yuen Long & Tuen Mun, (iv) Tung Chung, and (v) the biggest group is in the Hongkong island region. We recommend to split them into two groups,

Fig. 15.6 Geographical location of periodic review policy retailers on Monday

where the biggest group in Hongkong island will be in the first group, while the others will be in the second group, and each group will receive their orders on alternate Monday. Such a split will ensure delivery efficiency.

15.5 Implications of New Proposed Order Interval

Figure 15.7 shows the historical average order interval of Periodic Review policy retailers belonging to Monday. Note that the historical average order intervals are not in multiples of 7 days because these retailers only ordered predominantly on Mondays, but may still order on other days. Our proposed solution was to ‘force’ them to order only on alternate Mondays, which will make their order interval 14 days. The same principle will apply to retailers who predominantly order on other days of the week, where their average order interval will be ‘forced’ to be 7 days. This is known as the Power-of-Two principle where by approximating optimal order intervals to the nearest power-of-2 order interval, the total cost is guaranteed to increase not more than 6 %.

Although the total cost to the retailers will not increase by more than 6 %, there are other implications when ‘forcing’ them to order on alternate Mondays:

- For retailers whose historical average order interval is less than 14 days, they will be receiving orders less frequently than before, and the order size received will be larger. The main concern here would be whether the retailers would have

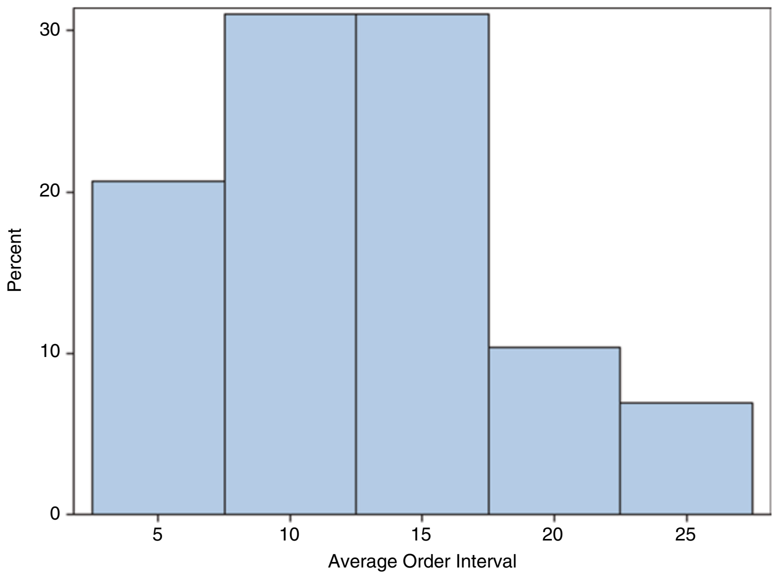

Fig. 15.7 Average order interval for periodic review policy retailers on Monday

sufficient capacity to receive the larger orders. This issue will be addressed in the next two sections.

- For retailers whose historical average order interval is more than 14 days, they will be receiving orders more frequently than before, and the order size received will be smaller. The main concern here would be whether the retailers would have the manpower to receive the orders more frequently. We will not address this issue in this paper.

15.6 Computation for Corresponding Proposed Order Size

The corresponding proposed order sizes can be computed using a moving averaging method, where the averages are computed using historical orders. Assuming historical orders in a particular period will represent future orders in the same period, the proposed order sizes are pre-computed based on historical order data for each retailer, using the proposed order interval of 7 or 14 days.

As defined previously,

- $\bar{O}_{ij}$ = Set of orders by retailer i on day of week j
- M_{ij} = Number of orders by retailer i on day of week j, where $M_{ij} = |\bar{O}_{ij}| \geq 0$
- $\bar{R}_i$ = Set of all the orders placed by retailer i
- N_i = Number of orders by retailer i, where $N_i = |\bar{R}_i| \geq 0$

So, for every retailer i in set $\overline{P}_j$, where j is the dominant order day,

- T_j = Proposed order interval.

$$T_1 = 14,\ T_2 = T_3 = T_4 = T_5 = T_6 = T_7 = 7$$

- k = Order index number of historical orders where $k = 1$ will be the first order. $k = 1$ *to* N_i
- k' = Order index number of proposed orders where $k' = 1$ will be the first order
- O_{ijk} = Order size of historical order k by retailer i for dominant order day j
- t_{ijk} = Time interval between historical order k and order k + 1, by retailer i on dominant order day j. For N_i historical order, there will be ($N_i - 1$) time intervals.
- $Q_{ijk'}$ = Proposed order size for order k' for retailer i for dominant order day j

The computation method has five main steps for any retailer i with dominant order day j, and proposed order interval T_j.

1. For initialization,

   (a) Compute the first historical average daily demand based on historical order $k = 1$

   $$Q_{ij1} = D_{ij1} * T_j$$

   (b) Compute the first proposed order size for order $k' = 1$,

   $$Q_{ij1} = D_{ij1} * T_j$$

   This proposed order size Q_{ij1} should cater adequately to demand for the first T_j days.

   (c) Let $D_{ij1} = D_{ijp}$ where the subscript p in D_{ijp} denotes previous average daily demand.

2. Compute a new average daily demand based on the closest equivalent order interval.

   $$D_{ijn} = \sum_{k=s}^{s+K} O_{ijk} \Big/ \sum_{k=s}^{s+K} t_{ijk}$$

   Where,

   - s is the starting order index number
   - K is the number of historical orders whose sum of the historical order interval matches closest the proposed time interval T_j K changes for every computation of D_{ijn}.
   - n in D_{ijn} denotes new average daily demand
   - For initialization, s = 1. For subsequent iterations, $s = K + 1$.

3. Compute the applied average daily demand by averaging the new average daily demand obtained in step 2, with the previous average daily demand. In case where

the actual demand is known, the actual demand for the past T_j days can replace D_{ijp} for a more accurate average demand to be applied for the next T_j days.

$$D_{ija} = \left(D_{ijp} + D_{ijn}\right) / 2$$

4. Compute the adjusted proposed order size for the next T_j days

$$Q_{ija} = D_{ija} * T_j$$

   By actively adjusting the proposed order size based on historical value on a moving average, the order size will be able to cater to demand changes.
5. Let $D_{ijp} = D_{ijn}$ and repeat Steps 2, 3 and 4 until the all the proposed order sizes for the entire year of 52 weeks are computed.

Example Computation Based on Table 15.1

1. Initialization
   (a) The first average daily demand was computed from the first order quantity and order interval (i.e. average daily demand = order quantity/order interval). First average daily demand = 10/5 = 2.0
   (b) Using this average daily demand, the proposed order quantity = 14 days* average daily demand = 14* 2.0 = 28. This order quantity should cater adequately to demand for the next 14 days.
2. Compute the new average daily demand based on the closest equivalent order interval. New average daily demand for the closest equivalent order interval of 15 days = (10 + 8 + 9 + 8)/(5 + 4 + 3 + 3) = 2.33
3. Compute the applied average daily demand by averaging the new average daily demand with the previous average daily demand of 2.0. The applied average daily demand = (2.33 + 2.0)/2 = 2.17
4. Adjusted order quantity for the next 14 days interval = 14* 2.17 = 30 (to nearest integer). This new order size of 30 should cater adequately to demand for the next 14 days.
5. Steps 2, 3 and 4 are repeated until the all the proposed order sizes for the entire year of 52 weeks are computed.

Table 15.1 Computation of proposed order size & adjusted order size for 14-day interval

Historical data				14 days order interval	
Order #	Order quantity	Order interval	Average daily demand	Proposed order quantity	Adjusted order quantity
1	10				
2	8	5	2.0	28	
3	9	4	2.0		
4	8	3	3.0		
5	10	3	2.7		30

15.7 Retailers' Capacity Constraint Check

Proposing a longer order interval will result in a larger order size, which may violate the storage capacities at the retail stores. However, the storage capacity at each of the retail stores was not captured in the raw data. We could however infer from the historical purchase order data, assuming that retailers who placed large order in the past would have a large storage capacity.

A measure of reasonableness will be computed as,

$$\text{Ratio Z} = \text{Maximum}\left(\text{Proposed Order Size}\right) / \text{Maximum}\left(\text{Historical Order Sizes}\right)$$

As defined previously,

- $\overline{P}$ = Set of retailers i who employed the periodic review policy
- $\overline{R}_i$ = Set of all the orders placed by retailer i
- N_i = Number of orders by retailer i, where $N_i = |\overline{R}_i| \geq 0$
- k = Order index number of historical orders for retailer i where $k = 1$ to N_i
- O_{ijk} = Order size of historical order k by retailer i for dominant order day j
- k' = Order index number of proposed orders where $k' = 1$ will be the first order
- $Q_{ijk'}$ = Proposed order size for order k' for retailer i for dominant order day j

For every retailer i in set $\overline{P}$, we determine the ratio of $Z_{k'}$ as,

$$Z_{k'} = Max_{k'}\left(Q_{ijk'}\right) / Max_k\left(O_{ijk}\right)$$

Table 15.2 shows the percentage of Periodic Review policy retailers with their respective ratio X. Ratio Group 1 has 47 % of the retailers who have Ratio $Z_{k'} < 1$, which means that the proposed order size will not exceed their storage capacity. Ratio Group 2 has 34 % of the retailers who have Ratio $Z_{k'}$ between 1 and 2, which means that the proposed order will be within 1–2 times their maximum order size, which is still reasonable. Ratio Group 3 has the remaining 19 % of the retailers who have Ratio $Z_{k'}$ above 2, which means that the proposed order size have a high chance of exceeding their storage capacity. Cost savings derived from the new distribution strategy can be passed on to these retailers to entice them to accept the new order interval and order size, especially for those in Ratio Group 3.

Table 15.2 Ratio $Z_{k'}$ of proposed order size/maximum historical order size

Ratio group	%	Ratio $Z_{k'}$
1	47	$Z_k' <= 1$
2	34	$1 < Z_k' <= 2$
3	19	$Z_k' > 2$

15.8 Comparing Number of Delivery Trips

For retailers who employed the Continuous Review policy, there will be no change to the number of orders and thus no change to the number of delivery trips required. For retailers who employed the Periodic Review policy, the number of orders will be changed according to the proposed order intervals (14 days for Monday, and 7 days for other days of the week). The total number of delivery trips made for both policies, was compared with the original number of trips for two groups on Mondays, and 1 group each for Tuesday to Sunday, in Table 15.3.

The number of trips made based on fixed delivery day and fixed interval is reduced by about 20 % and up to 47.3 % for Sunday. The biggest improvement comes from the split of the Monday group into two groups, so that the number of trips needed on any Monday is around 1,100 trips, instead of 2,900 trips in total. This will reduce the total number of trucks required for the entire delivery operations, and in turn reduce the cost of distribution.

15.9 Conclusions

In this paper, we have demonstrated how a logistics company can make use of the data they have captured in their order system to infer the ordering behavior of their retailers. By performing data analysis to categorize the retailers into Periodic Review or Continuous Review policy groups, we could identify that the fluctuations in the number of orders were primarily caused by retailers who employed the Periodic Review policy. These Periodic Review policy retailers were then classified according to their dominant order day and the result showed that the Monday group had double the number of orders than other days of the week. The proposed solution

Table 15.3 Comparison between number of delivery trips

	Monday group 1 (14 day)	Monday group 2 (14 day)	Tuesday (7 day)	Wednesday (7 day)	Thursday (7 day)	Friday (7 day)	Saturday (7 day)	Sunday (7 day)
Original number of trips	1,433	1,455	1,469	1,271	1,074	1,007	67	165
Number of trips based on fixed delivery day and fixed interval	1,148	1,147	1,188	1,034	956	996	67	87
Reduction percentage	19.9 %	21.2 %	19.1 %	18.7 %	11.0 %	1.1 %	0 %	47.3 %

was to split the Monday retailers into two groups with order interval of 14 days, while the other retailers will have order interval of 7 days. The overall reduction in the number of trips made was about 20 % to as high as 47.3 %. The largest savings would be derived from the reduction in the number of trucks to support the entire delivery operations. We have successfully demonstrated how new solutions can be proposed by integrating data analytics with decision analytics, to reduce distribution cost for a logistics company.

15.10 Teaching Note

15.10.1 Overview

Many operations management problem ranging from demand forecasting, inventory management, distribution management, capacity planning, workforce scheduling, and queue management are usually solved using known OM/OR techniques such as algorithms, heuristics, and optimization techniques. However, such a typical OM/OR solution methodology often assumes that the actual cause of the problem is known and the problem objective is well defined.

Practitioners like us would know that real business problems do not present themselves clearly, often resulting in people solving the wrong problem. Thus, in this course, the students will be exposed to the Data and Decision Analytics Framework (Fig. 15.8) which helps the analyst to first identify the actual cause of

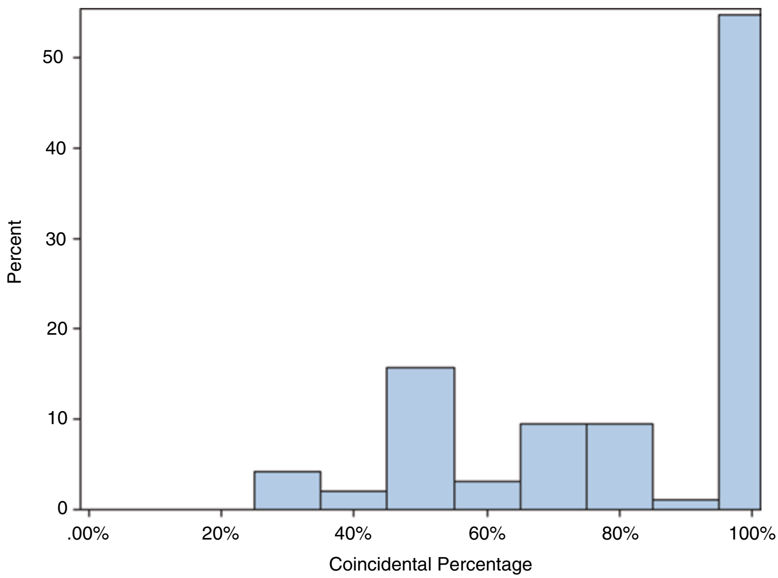

Fig. 15.8 Data & decision analytics framework

business problems by collecting, preparing, and exploring data to gain business insights, before proposing what objectives and solutions can and should be done to solve the problems.

These steps are missing in most problem solving frameworks, particularly in solving operations management problems, where the actual cause of the problem is assumed to be known and the problem objective is assumed to be well defined. However, we advocate that careful data analysis needs to be performed to identify the actual cause of business problems, before embarking on finding the solution.

15.11 Typical Flow of Classroom Activities

A typical flow of classroom activities is depicted in the flow chart in Fig. 15.9. A case usually covers multiple perspectives of operations management topics and the instructor will first cover the topics in terms of the theories and applications. When there are mathematical calculations involved, the instructor can use class activities to supplement and enhance the students' understanding.

After that, the instructor will present the case and facilitate the discussion so that the students can appreciate the case problem and think about the solution methodology according to the Data and Decision Analytics Framework. Once the students understand the intent of the case and what they are supposed to do, the instructor can facilitate the hands-on laboratory session using the step-by-step lab guide. At the end of the lab session, the instructor can instruct the students to complete assignment questions related to the case.

15.12 Introduce Operations Management Topics

For this case, the two topics to be covered include inventory management and distribution management. For inventory management, the understanding of the Periodic Review (PR) policy and Continuous Review (CR) policy should be highlighted. The instructor can ask the students the following questions to facilitate discussions:

- Give examples of goods which the periodic review policy will be more applicable
- Similarly, give examples of goods which the continuous review policy will be more applicable
- What are the advantages and disadvantages of each policy?

For distribution management, the instructor can cover the travelling salesman problem, multiple traveling salesman problem, and vehicle routing problem, introducing the different heuristics which are used to obtain good feasible solution in each problem. The main objective of distribution management is to design tours that will reduce the number of trips made when delivering goods, so as to reduce

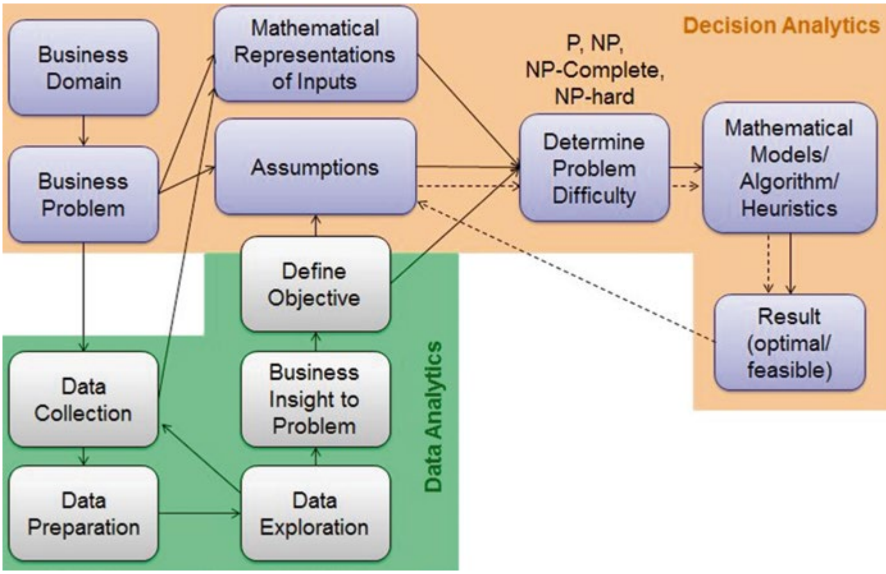

Fig. 15.9 Typical flow of classroom activities

distribution cost. The instructor can ask the students the following questions to facilitate further discussions:

- What other constraints will affect the design of the tour (time window, delivery trucks capacity constraints, client's preferences, traffic conditions)?
- What practical considerations should the vehicle routing planner consider when planning route for a particular driver (familiarity with road, ability to handle different truck size)?
- What practical considerations should the vehicle routing planner consider when planning route for a particular truck (types of goods – refrigerated or not, size of truck, maximum tonnage, door types – open at the back or at the sides)?

15.13 Conduct Case Discussion

15.13.1 Introduce the Case

The case is about IDD which is a leading integrated distribution and logistics services provider with its headquarters in Hong Kong. IDD provides a full suite of integrated distribution services covering Logistics, Distribution, Manufacturing and International Transportation.

The Distribution/Merchandising division of IDD plays the middle man role (see Fig. 15.10) in distributing products for their principal accounts (brand owners) to retail stores. Products include food items such as corn flakes and chocolates,

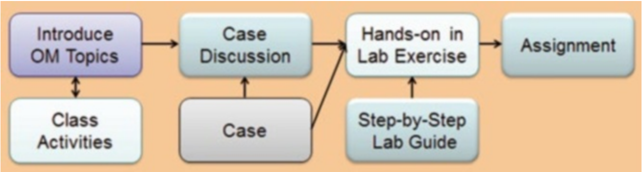

Fig. 15.10 IDD facing problem in distributing fluctuating orders to retailers

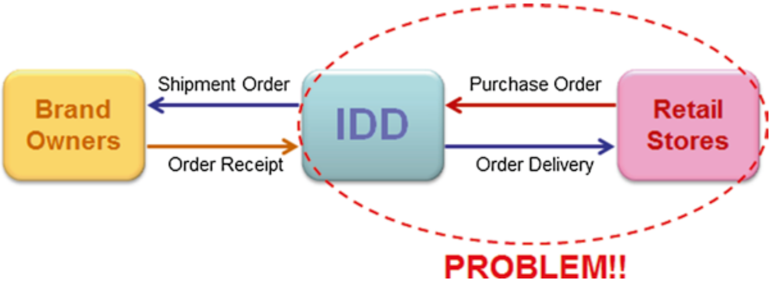

Fig. 15.11 Bullwhip effect experienced in IDD's supply chain

and health and beauty items such as toothpaste and shampoo. The division faces distribution challenges from IDD to the retailers. Orders from retailers fluctuates daily and these fluctuations resulted in the distribution team working very hard with delivery trucks rushing to deliver orders on every Monday, while on other days of the week, the team sees idle trucks parking at the warehouse un-utilized. Playing the passive middle-man role, IDD can only prepare the maximum resource capacities (e.g. drivers and trucks) in order to handle such uncertainties.

The instructor can go further to explain the bullwhip effect in supply chains which is caused by factors such as long lead time, batch ordering and demand variation. In this case, the fluctuations in the retailers orders are likely to be caused by batch ordering behavior of the retailers since demand variation on fast moving consumer goods like cornflakes and toothpaste are relatively low, as shown in Fig. 15.11.

15.13.2 *Possible Solutions and Data Provided*

After the case introduction, the instruction will ask the students to suggest possible solutions to solve the problem and for each viable suggestion, the students can discuss the pros and cons. One possible suggestion would be to implement Vendor Managed Inventory (VMI) where IDD will deliver the required quantity of products just in time, and the retailers need not place orders actively. For this suggestion, the instructor can ask the students to discuss about the pros and cons of Vendor Managed Inventory.

The Pros include:

- VMI solution will be win-win for both IDD and the retailers
- IDD can plan the deliveries better and reduce the overall cost of deliveries
- The retailers can eliminate manpower to do inventory checks and place orders

The Cons include:

- VMI implementation will require that the retailers share their Point-of-Sales (POS) data with IDD
- Due to confidentiality and trust, most retailers will not be willing to share their POS data

At this point, the instructor can highlight that IDD's IT system stores historical records of the orders from the retailers as well as the store location of each retailer provided in the Appendix of the main paper. With the order data provided (consisting data of 326 retailers and 2,681 orders), the instructor can direct the students to focus on the following four fields:

- Customer No – this is the unique customer ID
- Order Date – this is the order date
- Original Qty – this is the order quantity
- StorerClientCode – this is the store code

With the store location data provided, the instructor can direct the students to focus on the following three fields:

- Latitude – this is the latitude of the store location in geo-information format
- Longitude – this is the longitude of the store location in geo-information format
- Shiptocode – this is the store code which corresponds to StorerClientCode in the Order Data table

15.13.3 Classification Rule

After understanding the data provided, the instructor will lead the discussion on how to infer the retailers' inventory ordering behavior from using the order date. To perform the inference, the instructor needs to explain the Classification Rule (Rule 1 provided in the main paper) which is used to classify the retailers according to Continuous Review (CR) policy or Periodic Review (PR) policy.

At this point, the instructor can ask the students what if X_{cut} is chosen to be say, 60 %? Will the number of retailers categorized into PR retailers be more or fewer?

Upon using the classification rule to categorize the retailers into PR and CR policy, the instructor can explain that by plotting simple bar charts to visualize how

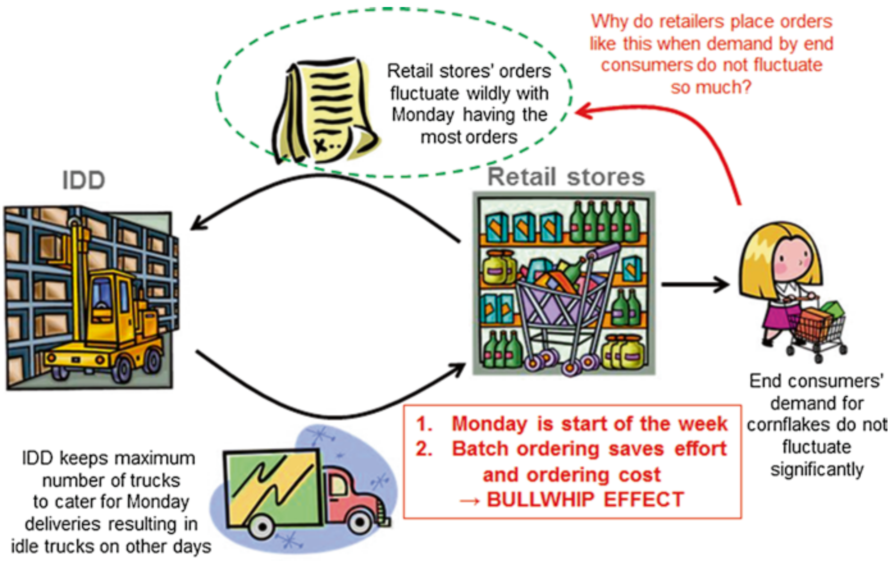

Fig. 15.12 Laboratory exercise activities

many PR and CR retailers place their orders on their dominant order day, students will be able to identify the cause of the order fluctuations and proceed to recommend a new distribution strategy.

15.14 Conduct the Laboratory Exercise

At this point, the students would have appreciated the case problem and understood that they need to perform the laboratory session with the following tasks (depicted in Fig. 15.12),

1. Infer the retailers' inventory ordering behavior by categorizing them into PR and CR according to the classification rule
2. Plot bar charts to visualize the distribution of the retailers according to their dominant order day, and use the bar charts to deduce the root cause of the order fluctuations
3. Propose new distribution strategy which can allow IDD to play a more active role to plan the delivery of the orders to the retailers on each day of the week, and propose the quantity to deliver
4. What constraints must IDD consider and how can IDD ensure that the new distribution strategy is practical (e.g. retailers' capacity challenge)?
5. Justify that the new distribution strategy will result in cost reduction.

15.15 Ensure Learning Outcomes Are Achieved

The entire case aims to achieve several learning outcomes:

1. *Exposure to supply chain business domain covering two major operations management topics including inventory management and distribution management*

   This learning outcome is achieved when the instructor covers the two operations management topics on the theories and the applications, together with class discussion and supplemented with class activities if needed.
2. *Ability to identify the actual cause of business problem by collecting, preparing, and exploring data to gain business insights, before proposing what objectives and solutions can and should be done to solve the problems using the Data and Decision Analytics Framework*

   This learning outcome is achieved when the students apply the steps in the Data & Decision Analytics Framework.
3. *Ability to propose solutions which are practical and provide cost justification*

The third learning outcome is achieved when the students perform the computations for the new proposed order size for the retailers' capacity constraint check and compute the reduction in the number of trips.

Finally, to further enhance the understanding of the case, the students can be asked to complete an assignment with the following question:

Assuming that you can dictate the type of data and information you can get from the business and you can propose a new "Order-to-Distribution-Process", propose an alternative solution to improve distribution and list the types of data needed from new the business process. Map the process flow for your proposed solution.

15.16 Appendix

1. *Logistic Data*

The logistic data contains information about the logistic transport of the goods to the retailer. Here, the retailer is identified by CustomerNo. Table 15.4 also contains some information about the expected delivery of the goods.

2. *Store Location Data*

The store location data contains the information of all the retailers' store location in geo-information format. Here in Table 15.5, the retailer is identified by Shiptocode.

3. *Proofs for Rules 1 & 2*

(a) *Proof for Rule 1: Periodic Review with Single Dominant Day*

If there exist a $Max_j(X_{ij}) > X_{cut}$, *then retailer i is assumed to employ the periodic review policy on the dominant order day j, with a confidence interval of* $(1-\alpha)\%$ *and level of significance of* $\alpha\%$.

Table 15.4 Comparison between number of delivery trips

Field name	Field description
CountryCode	Country Code
CustomerNo	Customer ID
ExpectedDeliveryDate	Expected Delivery Date of Good
OrderDate	Order Date
OrderKey	Order Key
OriginalQty	Original Order Quantity
PODDeliveryDate	Final Delivery Date
PODStatus	Final Delivery Status
PODStatusDescription	Final Delivery Status Description
PrincipalCode	Principal Code
PrincipalDescription	Principal Description
ShippedDate	Shipped Date
ShippedQty	Shipped Quantity
SkuCode	SKU Code
SkuDescription	SKU Description
StorerClientCode	Storer Code

Table 15.5 Comparison between number of delivery trips

Field name	Field description
Latitude	Latitude
Longitude	Longitude
Shiptoaddress1	Address 1
Shiptoaddress2	Address 2
Shiptocity	City
Shiptocode	Store Code
Shiptoname	Store Name
Storerkey	Storer ID
Storername	Storer Name

Consider an order from retailer *i* which can occur on any of the 7 days of the week.

- The probability of the order falling on a particular day of interest is $\frac{1}{7}$, and we call this the probability of success.
- Thus, the remaining probability of the order *not* falling on that particular day of interest is $\frac{6}{7}$, and we call this the probability of failure.
- This allows us to formulate a Binomial Test with $p = \frac{1}{7}$ and number of trials = 7, to determine the X_{cut} with the corresponding confidence interval $(1-\alpha)\,\%$ and level of significance $\alpha\,\%$.

From Table 15.6, it is observed that:

– If the percentage of occurrence of orders for a particular day of interest is 14.3 %, we are 73.65 % confident that the observation did not occur by chance with the level of significance of 26.35 %.

Table 15.6 PMF and CDF for binomial test for single day of interest

Number of occurrence on a particular day	% of occurrence	Probability Mass Function (PMF) of binomial distribution	Cumulative Distribution Function (CDF) of binomial distribution	$1-\text{CDF}=\alpha\%$
0	0 %	0.3399	0.3399	0.6601
1	$\frac{1}{7}=14.3\%$	0.3966	0.7365	0.2635
2	$\frac{2}{7}=\ 28.6\%$	0.1983	0.9348	0.0652
3	$\frac{3}{7}\ =42.9\%$	0.0551	0.9898	0.0102
4	$\frac{4}{7}\ =57.1\%$	0.0092	0.9990	0.0010
5	$\frac{5}{7}\ =71.4\%$	0.0009	0.9999	0.0001
6	$\frac{6}{7}=85.7\%$	0.0001	1.0000	0.0000
7	$\frac{7}{7}\ =100\%$	Approximately 0	1.0000	0.0000

  - If the percentage of occurrence of orders for a particular day of interest is 28.6 %, we are 93.48 % confident that the observation did not occur by chance with the level of significance of 6.52 %
  - If the percentage of occurrence of orders for a particular day of interest is 42.9 %, we are 98.98% confident that the observation did not occur by chance with the level of significance of 1.02 %
  - And so on.

- In our paper, we have selected $X_{cut}=40\%$ and state that if there exist a $\text{Max}_j(X_{ij})>40\%$, then retailer i is assumed to employ the periodic review policy on the dominant order day j, with more than 93.48 % confidence that the observation did not occur by chance with level of significance less than 6.52 %.

(b) *Proof for Rule 2: Periodic Review on 2 days, but with Single Dominant Day*

If there exist a $Max(Y_{iw})>Y_{cut}$, then retailer i is assumed to employ the periodic review policy on 2 days of the week represented by the combination index w, with a confidence interval of $(1-\alpha)\%$ and level of significance of $\alpha\%$. For this combination w, if $X_{iq}>X_{ir}$ where q and r are the days of week represented by combination w, then q will be the dominant order day.

We apply a similar Binomial Test here by grouping the 2 days of interest as 1 group, and the remaining 5 days as the other group.

Table 15.7 PMF and CDF for Binomial Test for 2 Days of Interest

Number of occurrence on a particular day	% of occurrence	Probability Mass Function (PMF) of binomial distribution	Cumulative Distribution Function (CDF) of binomial bistribution	1 − CDF = α%
0	0 %	0.0949	0.0949	0.9051
1	$\frac{1}{7}$=14.3%	0.2656	0.3605	0.6395
2	$\frac{2}{7}$= 28.6%	0.3187	0.6792	0.3208
3	$\frac{3}{7}$ =42.9%	0.2125	0.8917	0.1083
4	$\frac{4}{7}$ =57.1%	0.0850	0.9767	0.0233
5	$\frac{5}{7}$ =71.4%	0.0204	0.9971	0.0029
6	$\frac{6}{7}$=85.7%	0.0027	0.9998	0.0002
7	$\frac{7}{7}$ =100%	0.0002	1.0000	0.0000

- The probability of the order falling on two particular days of interest is $\frac{2}{7}$, and we call this the probability of success.
- Thus, the remaining probability of the order not falling on that two particular days of interest is $\frac{5}{7}$, and we call this the probability of failure.
- This allows us to formulate a Binomial Test with $p = \frac{2}{7}$ and number of trials = 7,
- to determine the Y_{cut} with the corresponding confidence interval $(1-\alpha)\%$ and level of significance $\alpha\%$.

From Table 15.7, it is observed that:

- If the percentage of occurrence of orders for any of the two particular days of interest is 14.3 %, we are 36.05 % confident that the observation did not occur by chance with the level of significance of 63.95 %.
- If the percentage of occurrence of orders for any of the two particular days of interest is 28.6 %, we are 67.92 % confident that the observation did not occur by chance with the level of significance of 32.08 %
- If the percentage of occurrence of orders for any of the two particular days of interest is 42.9 %, we are 89.17 % confident that the observation did not occur by chance with the level of significance of 10.83 %
- And so on.

- In our paper, we have selected $Y_{cut} = 60\%$ and state that if there exist a $Max(Y_{iw}) > 60\%$, then retailer i is assumed to employ the periodic review policy

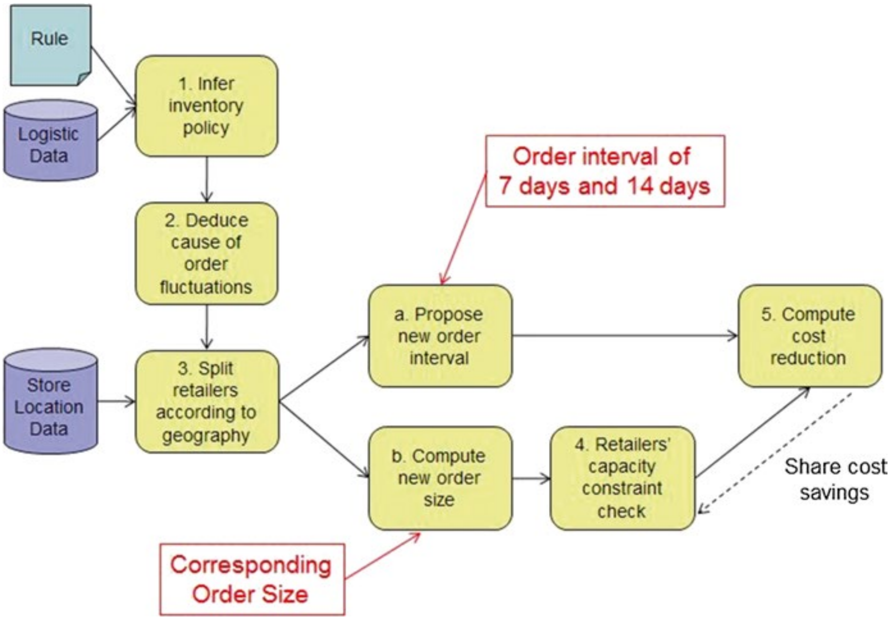

Fig. 15.13 Coincidental analysis of dominant day for periodic review policy retailers

on 2 days of the week represented by the combination index w, with more than 97.67 % confidence that the observation did not occur by chance with level of significance less than 2.33 %. For this combination w, if $X_{iq} > X_{ir}$ where q and r are the days of week represented by combination w, then q will be the dominant order day.

4. *Coincidental Analysis of Ordering Practice for Period*

Further analysis of the Periodic Review policy retailers in the Fig. 15.13 below shows the coincidental analysis of dominant day for retailers. About 60 % of them have 100 % of their orders fixed on the same day of the week. This further justified that the ordering pattern of the Periodic Review policy retailers is independent of the SKU item ordered.

Biography

Michelle L. F. Cheong is currently an Associate Professor of Information Systems (Practice) at the School of Information Systems (SIS) at Singapore Management University (SMU). Prior to joining SMU, she had 8 years of industry experience leading teams to develop complex IT systems which were implemented enterprise-wide covering business functions from sales to engineering, inventory management, planning, production, and distribution.

Upon obtaining her Ph.D. degree in Operations Management, she joined SMU in 2005 where she teaches the *Business Modeling with Spreadsheets* course at the undergraduate level and is the co-author of the book of the same name. She also teaches in three different master programmes at SMU on different variants of spreadsheet modeling courses covering different domains, including financial modeling, innovation modeling and IT project management. She recently designed and delivered an *Operations Focused Data Analytics* course for the Master of IT in Business (Analytics) programme at SIS. Apart from her teaching, Michelle is also the Director of Postgraduate Professional Programmes at SIS where she is in charge of two master programmes and Continuing Education & Training.

Michelle has won several awards including the *Most Promising Teacher Award* in 2007 and the *Postgraduate Professional Programme Development Award* in 2013, both from the SMU Center for Teaching Excellence. In addition, she has recently bagged the inaugural *Teradata University Network (TUN) Teaching Innovation Award* 2013, which recognizes excellence in the teaching of Business Intelligence and Business Analytics at the undergraduate, graduate and/or executive education levels.

Murphy Choy is currently a Data Analytics System and Learning Engineer at the School of Information Systems (SIS) at Singapore Management University (SMU). Prior to joining SMU, he had 4 years of industry experience in the area of risk analytics covering Asia Pacific, Middle East, Africa and Latin America. He has spearheaded several analytics initiatives in the bank and has done extensive research work for collections, recoveries and Basel II models. Murphy is especially competent in SAS software and many other analytics software, and also responsible for most of the laboratory exercises designed and taught in the Master of IT in Business (Analytics) programme.

He has served as Chairperson for the Singapore SAS user group and Section Chair for the SAS Global User group. He is pursuing his doctorate degree and his research interest is in the field of Operation Management and Text Mining. He has earned a MSc degree in Finance from University College Dublin and a BSc degree in Statistics from National University of Singapore. Murphy is also a co-author of the paper that won the inaugural *Teradata University Network (TUN) Teaching Innovation Award* 2013.

References

Cheng, T. C. E., & Wu, Y. N. (2005). The impact of information sharing in a two-level supply chain with multiple retailers. *The Journal of the Operational Research Society, 56*(10), 1159–1165.

Gaur, V., Giloni, A., & Seshadri, S. (2005). Information sharing in a supply chain under ARMA demand. *Management Science, 51*(6), 961–969.

Hausman, W. H., & Sides, R. S. G. (1973). Mail-order demands for style goods: Theory and data analysis. *Management Science, 20*(2, Application Series), 191–202.

Johnston, F. R., Boylan, J. E., & Shale, E. A. (2003). An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. *The Journal of the Operational Research Society, 54*(8), 833–837.

Lee, H. L., So, K. C., & Tang, C. S. (2000). The value of information sharing in a two-level supply chain. *Management Science, 46*(5), 626–643.
Raghunathan, S. (2001). Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Science, 47*(4), 605–610.
Steckel, J. H., Gupta, S., & Banerji, A. (2004). Supply chain decision making: Will shorter cycle times and shared point-of-sale information necessarily help? *Management Science, 50*(4), 458–464, Special issue on marketing and operations management interfaces and coordination.
Williams, B. D., & Waller, M. A. (2010). Creating order forecasts: Point-of-sale or order history? *Journal of Business Logistics, 31*(2), 231–251.
Yu, C. Z., Yan, H., & Cheng, T. C. E. (2002). Modelling the benefits of information sharing-based partnerships in a two-level supply. *The Journal of the Operational Research Society, 53*(4), 436–446.