International Journal of Statistics and Applied Mathematics 2017; 2(6): 85-91

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2017; 2(6): 85-91

www.mathsjournal.com
Received: 11-09-2017
Accepted: 12-10-2017

Rajasekhar Kalla
Research Scholar,
Department of Statistics,
Sri Venkateswara University,
Tirupati, Andhra Pradesh, India

Dr. Saravana Murikinjeri
Sr. Data Scientist, Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, India

R Abbaiah
Prof. Department of Statistics
Sri Venkateswara University,
Tirupati, Andhra Pradesh, India

Hemasekhar Palla
Research Scholar, Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, India

Correspondence
Rajasekhar Kalla
Research Scholar,
Department of Statistics,
Sri Venkateswara University,
Tirupati, Andhra Pradesh, India

Data challenges and remedies of demand forecast in fashion business

Rajasekhar Kalla, Dr. Saravana Murikinjeri, R Abbaiah and Hemasekhar Palla

Abstract
The sales forecast is an essential aspect in the area of fashion business. In particular, forecasting the future demand for fashion goods is much harder due to some reasons such as short life span of the products, seasonal effects, available sales history etc. There are various traditional and advanced forecasting techniques available for forecast the demand. But all those models perform well only if we consider data in a proper way. The present study explains the different types of data challenges which occurs very frequently while forecasting the demand for fashion goods. Along with the issues, it addresses their remedies as well.

Keywords: Data challenges in demand forecasting, issues in demand forecasting, challenges and solution in fashion business

Introduction

In the apparel and fashion industry, decisions can be taken based on future demand. The future demand can be estimated by using various forecasting techniques. In the process of forecasting future demand, practitioners used to face so many challenges with the data such as data insufficiency and inconsistency. Here we address some of the very frequent data challenges.

In the retail business, data available on three different dimensions. They are product, geography and date. Each dimension may have different hierarchies. At geography, there may be some hierarchies like Country, Region, State, District and Store. At Product, there may be some hierarchies like Company, Department, Category, Sub-category and SKU. A forecasting model can outperform only when we consider the above-mentioned dimensions in proper order as in fashion industry there may be a huge number of products and most of the products may sell at different locations (stores) and different time periods. The above levels may change depends on retailer business strategies. Date dimension can be available on different levels such as Yearly, Half yearly, Quarterly, Monthly, Weekly and Daily.

Geography Hierarchy
Product Hierarchy

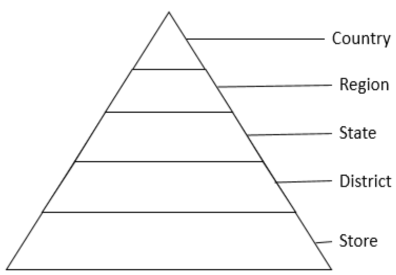


Fig 1.1

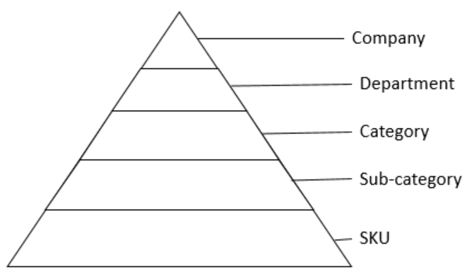


Fig 1.2

Fig 1: Illustrates the hierarchies of geography, product and Date.

In the dataset which we use to forecast the demand must be the combination of above two-dimension data. If we do not consider the data in proper order, we may face so many data challenges.

Issues

In the retail business, there are so many challenges. Here we address very frequent issues. They are

1. Standard Discount
2. Product Status
3. Product Life Cycle
4. Price Change
5. Historical Data Duration
6. Product Line Extension
7. Segmentation
8. Aggregation Levels
9. Missing values
10. Outliers
11. Demand Factors

Standard Discount

In retail business, most of the retailors used to implement various sales promotion techniques to increase sales, attract new customer and to retain current customers. Consumers would almost always choose an item that cost less even if two products are practically similar in terms of features. In generally customers always expect minimum discount on most of the product. On some products, there may not be any discounts as they are in high-demand. This discount may change from product to product based on their demand. There is no need to considered all the geo-prods sales history into the demand modelling. Because the geo-prod which do not have discounts or very less discounts may spurious the effect estimates and the estimated coefficients are not reliable. So, a modeler need to be very careful about the geo-prod which need to be included into the model data. For example, a geo-prod has 10 weeks of sales history. In these 10 weeks of sales history, first 3 weeks it doesn’t has discounts. From week 4 to 5 it has 5% discount. From 6 weeks on wards it sold for 20% discount. Now the question is which weeks information has to consider into the model data?

Remedy

To overcome this issue, we need to define a threshold value. It specifies the minimum discount percentage that regular shoppers expect before they purchase products. This specification may change from product to product. For apparels, we can specify 15% of minimum standard discount and for footwear it may be 20%.

Based on this threshold value, we filter the data. It means, the products which have more discount then we specified for those products can be considered as the modelling data. The products which do not satisfy (i.e., products which have a lesser discount than the specified discount) the observations can be ignored from the modelling data. In the above example, if we specify the threshold value is .10(10%) then we can consider observations from week 6 to week 10.

Product Status

Retail businesses must balance inventory, price and promotions to maintain profit levels. In the process of maintaining inventory, most of the products goes through two stages. They are active stage and clearance stage. When a product is in-demand then that product considered under the active stage. When a product is not in-demand or end of its life cycle, then that product considered under clearance stage. In particular, clearance sales are used as a promotional strategy by many companies as an inventory management. The main objective of clearance sale is to reduce excess inventory levels in a store. By minimizing inventory levels, we can maximize shelf space for in-demand products helps maintain sales levels and profits. At end of the product life cycle, sales are unstable due to so many reasons such as broken assortments, incorrect inventory levels, season changes, product discontinue, etc. So, that product gets moved to clearance stage from active stage. Once a product moved to clearance stage, no need to consider all the clearance stage sales history into the demand model.

Remedy

To overcome this issue, we need to specify a threshold value which determine how many weeks of clearance history need to be consider from its initial clearance week. For example, we specified that the threshold value 25 weeks. It means the products which exceed 25 weeks of initial clearance sales are unstable and can be ignored. So, we need to consider the clearance sale products which do not exceed the threshold value.

Product Life Cycle

There are four stages in every product life cycle. They are Introduction stage, Growth stage, Maturity stage and Decline Stage.

Introduction Stage

At the introduction stage, business seeks to build product awareness and develop a market for the product. In this stage, product introduces with high promotions and lesser prices to build market share rapidly.

Growth Stage

At this stage, company’s sales and revenue start s increasing also retailer seeks to increase brand value and market share. I this stage, retails have to face competition along with market share. Here retailer has to maintain product quality and additional features may be added. Price is stable or may increase as increasing demand with little competition.

Maturity Stage

At this stage, the retailer has to maintain brand importance and sales continued to grow but not as compared to the past. The main objective to maintain market share while maximizing profits. Prices may be lower because of the competition.

Decline Stage

At this stage, a product may not have good importance as customer taste has been changed. Here retailer has to maintain the product, reduce cost and finding the new users. If there is no profit from the product, discontinue it and sell the remaining inventory to another firm who is willing to sell that product.

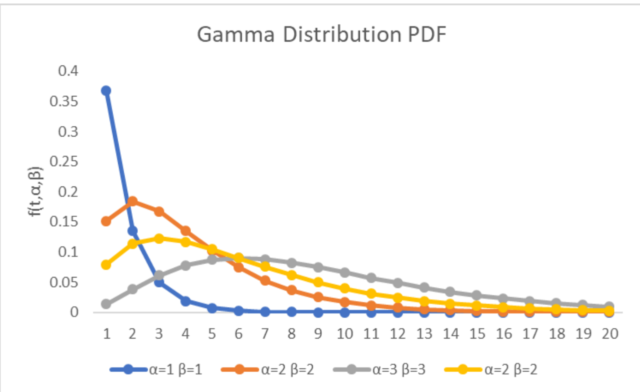
Every product passes through these stages. The product life cycle applies to all types of products like branded and category. But their life span changes from product to product. Some type of products may have long life span and some products may be very less. For example, foot wear may have long life span and swim suits may have very less life span. Every retailer always attempts to maximize the profit and revenue over the entire life cycle of the product. In order to

achieve the desired levels of profit, an introduction of the new product is crucial. After a certain number of periods from the initial introduction of products, it may face some issues such as broken assortments, shelf placement at the stage of decline. If a product is selling and has broken assortment, the result is low sales units. If we use sales history of these product complete life span, results may not be reliable because at end of the product life span sales pattern is unstable.
Also, demand might be too unstable for modelling at beginning of the product life cycle. This is because products can be stocked in middle of the week or the inventory data not reliable.

Remedy
To overcome product life cycle difficulty, we need to define minimum and maximum threshold values. If a product life span exceeds those thresholds (below to the minimum threshold and more than maximum threshold), then we can exclude those records from the model data. We can fix these thresholds based on product purchase behaviour. It may change from one product to another product.
There are so many statistical distributions that can be used to model reliability data. In particular, two distributions are most commonly used and very simple distributions. They are exponential distribution and gamma distribution.

Exponential Distribution: Most practicioners used to prefer exponential distribution as it is very simple to use. The probalitity density function of the exponential distribution is defined as
$f(t) = \lambda e^{-\lambda t}$

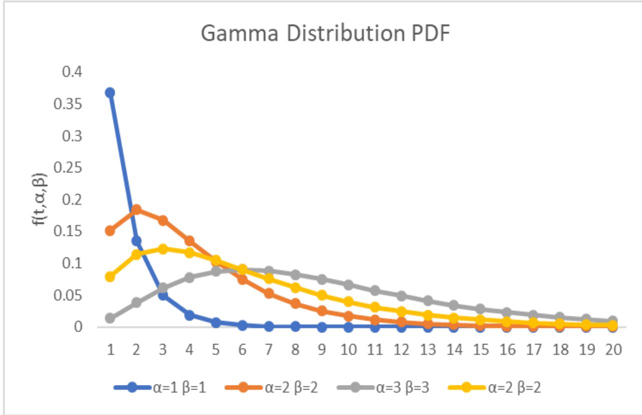
Where t: is the random variable which represents time.
λ: is the parameter of the distribution.
λ=1/m; where m is the mean time between the random variable t.
Depending on λ, f(t) will be scaled differently.



Gamma Distribution: It is a flexible life distribution model that offer a good fit to some sets of product life data. A random variable t that is gamma-distributed with shape α and rate β is denoted
t~Γ(α,β)≡Gamma(α,β)
The general formula for the probability density function of the gamma distribution is

$f(t;\alpha,\beta) = \frac{\beta^{\alpha} t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}$ for t>0 and α, β>0

Where Γ(α): is a complete gamma distribution.
α : is shape parameter
β : is scale parameter



Price Change
A business can use different types of pricing strategies when selling a product. The of each product can be set to maximize the revenue and margins. Pricing is very important component in the demand estimation. Based on the price of a product, consumer decides whether to purchase a product or not. Hence, correct setting price decisions are key to success in retail business. Business gives discounts on products to sell in high quantities. These discounts may be in the form of markdowns or promotions or etc. In retail business, prices are two types. They are standard prices and current prices. Standard prices are nothing but the price of a product when it introduced into the market. Current price is the price which is after removing markdowns or promotional cost from the standard price. When product introduced into the market, standard price and current price may be the same for most of the products as there won't be markdowns or promotions at initial period. But most of the product get price reduction in their life cycle. They get varied once business start giving discounts on products. These variations may change from product to product. Some products may be varied very low and some products may be varied very high. A demand model need to be consider only required price discounted observations. If we consider all the price values (with discounts and without discount), the results may not be reliable and inconsistent. So first, we need to select only required prices differenced observation from the product sales history.

Remedy
This filtering can be done based on the price ratio. Price ratio is the ratio of current price to standard price. If price ratio is close to 0 (i.e., current and regular prices both are same) then those records can be ignored from the model data. We should specify a threshold value and the geo-prod which has price ratio is lesser than the specified threshold value can be ignored. For example, the specified value is .2 (i.e., a geo-prod price ratio value is 20%) then the observations with price ratio is lesser than .20 are excluded from the modelling. Also, remove the observation with price ratio greater than 1 as actual price exceeds regular price.

Historical Data Duration
Demand forecasts would always depend on historical data. An accurate, statistically generated forecast has several components such as seasonality effect, cyclical effect and trend effect. There must be sufficient historical data to extract all of these components to generate a forecast for future period. In most of the cases there should be at least 2 years of previous data. But even when there is no sufficient historical data, forecast is still possible.

There are so many challenges associated with demand forecasting when data is limited. There are numerous statistical methods available to forecast the demand. To make more accurate forecasts, we need to combine historical data with forecast models. In the retail business, there are some geo-prods which may have more than 5 years of historical data. If we consider the entire available historical data, it would lead to inefficient estimates (i.e., under forecast or over forecast). Because there is a significant difference between current prices response to very beginning (ex: before 5 years) pricing responsiveness.

Remedy

To make the more accurate forecast, we need to be considered only last 3 years of historical data. If we go beyond the 3 years, we can't predict better estimates as the huge variability in the sales and price pattern. If we are using weekly data, then recent 156 weeks information is enough for a GP.

Product Line Extension

Forecasting new product introductions is becoming important as product life cycle is short and assortment turnover the increases. In particular, it is very difficult in retail sector such as fashion, electronics, books, etc. Forecast demand for new product is very hard as we don't get historical data for that product. But forecast the demand is still possible by considering information from the similar types of products under the same category.

In the retail business, each store is classified into two categories. They are comparable store and non-comparable store.

Comparable store: Comparable stores are also called as well-established stores. The store which operates from long enough to establish a stable sales patterns is known as comparable stores. If stores are comparable, then sales and revenue can be compared with previous year.

Non-Comparable store: Noncomparable stores are also called as newly opened stores. The stores which do not have good enough sales history to establish a stable sales patterns are called as non-comparable stores. If stores are non-comparable, it doesn't support to compare one-year sales or revenue values with previous years values.

If a store is non-comparable, then forecasting demand for that store is very difficult. To overcome this difficulty, we need to look for the similar type of stores. The similar type of store means the store which opened in the same location, sells a similar type of products and it should be comparable then those stores are called as like stores. Using that store's information, we can forecast future demand for the non-comparable stores.

We can classify products also into two categories. They are established products and new products. Established products are the products that are analogous to one have similar time-series patterns because they appeal to similar customer taste, have comparable levels of competition in the market. and new products are the products which introduced recently into the market. If we want to forecast demand for the newly introduced product, we need to use established products or like products. Once we have identified the data related to the item in the group, you can use that data to build a model to simulate the product or forecast how it performs in the market.

Remedy

To address this type of problems, we need to fix a threshold value. If a product life cycle is lesser than the specified threshold value, then we can use like products to forecast the demand.

Data Segmentation

In figure 1, there are some hierarchies for geography, product and date. From these hierarchies, we need to form homogeneous groups. Because one category products price responsiveness may change to another category. For example, Men's apparel price responsiveness may vary from Kid's apparel price responsiveness. To overcome this issue, data needs to be grouped/segmented based on business rules.

A segment is defined as the subset of geography and product. To create a segment, select a parent level anywhere in the modelling hierarchy for geography and product. All the child levels under the parent level are included in the same segment. A segment can be created at any level of the geography and product. All the nodes within the segment are homogeneous. For example, if a segment is created for men's apparel, then all the men's apparel related SKU's should come under this segment.

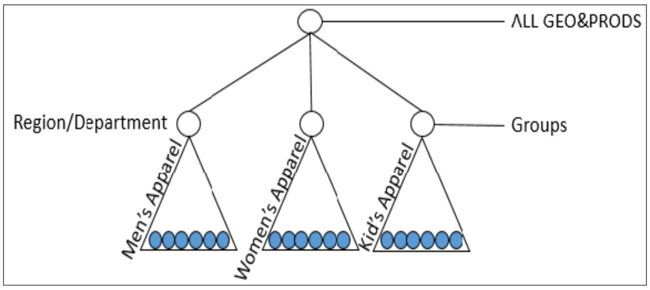


Fig 2: Illustrate of taxonomy the levels.

Taxonomy choices determine scope, scale and complexity of the development effort. The main objective of the problem is to avoid the over or under engineering the design. When making taxonomy and granularity choices in the hierarchies, the natural segment for the retailer is preferred. In figure2, a region from the geography and Department from the Product hierarchy is preferred as under those levels there won't be a big difference in price response (i.e., similar characteristic). From the above hierarchy, the selected levels are the natural segment levels.

Aggregation Levels

Once data taxonomized, each segment needs to be aggregated to required level and build models separately on each segment. Because at Store Keeping Unit (SKU) level data is unstable, inconsistent and insufficient. No model can predict better estimates at SKU level. We can determine the best level by visualize the data. In our hierarchy, we created segments at Region in Geography and Department in Product hierarchy.

If we aggregate a segment to Region and Category level, plot looks as shown below.

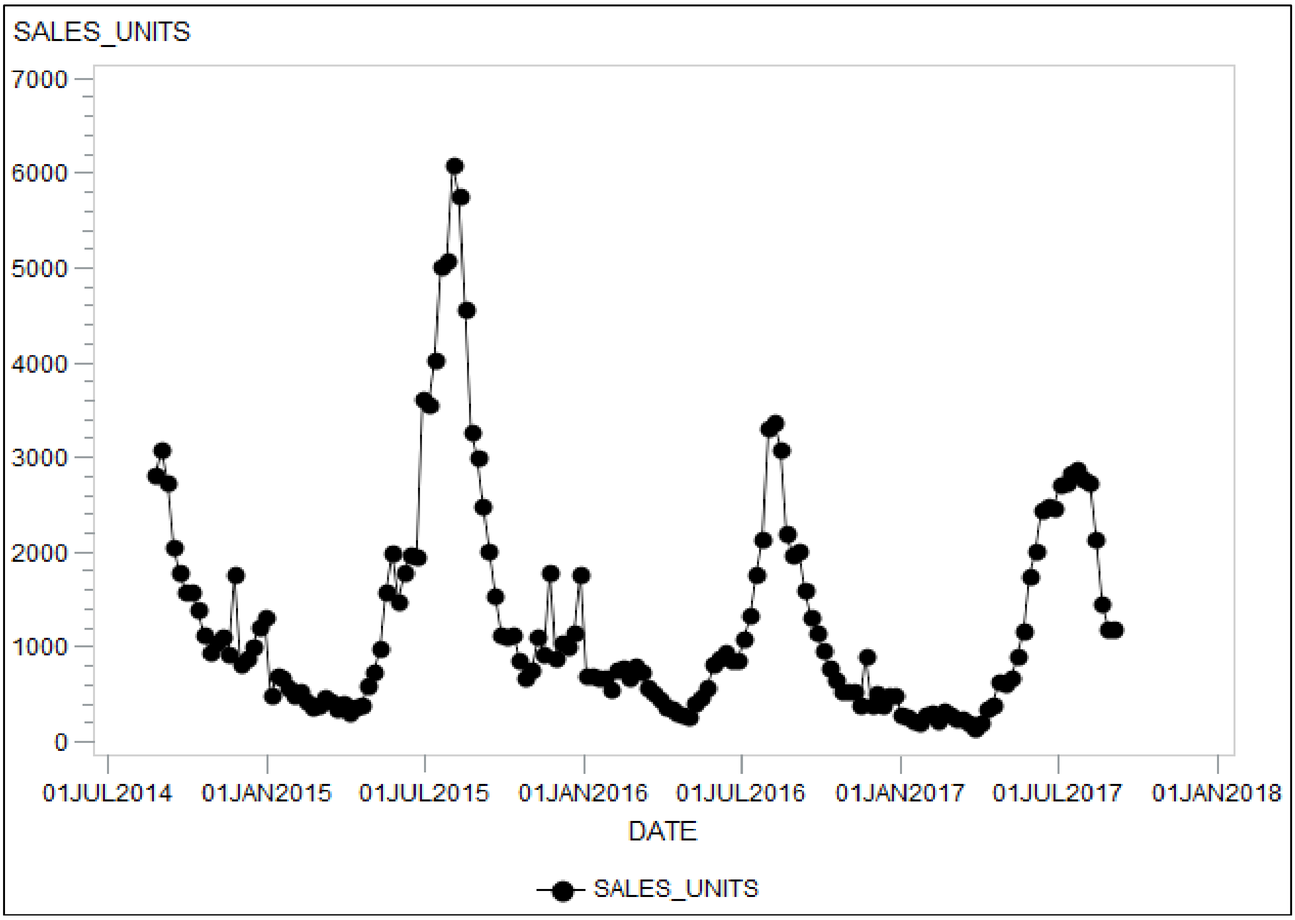


Fig 3: It illustrates the sales pattern of a geo-prod at Region and Category level.

If we aggregate a segment to State and Sub category level, plot looks as shown below.

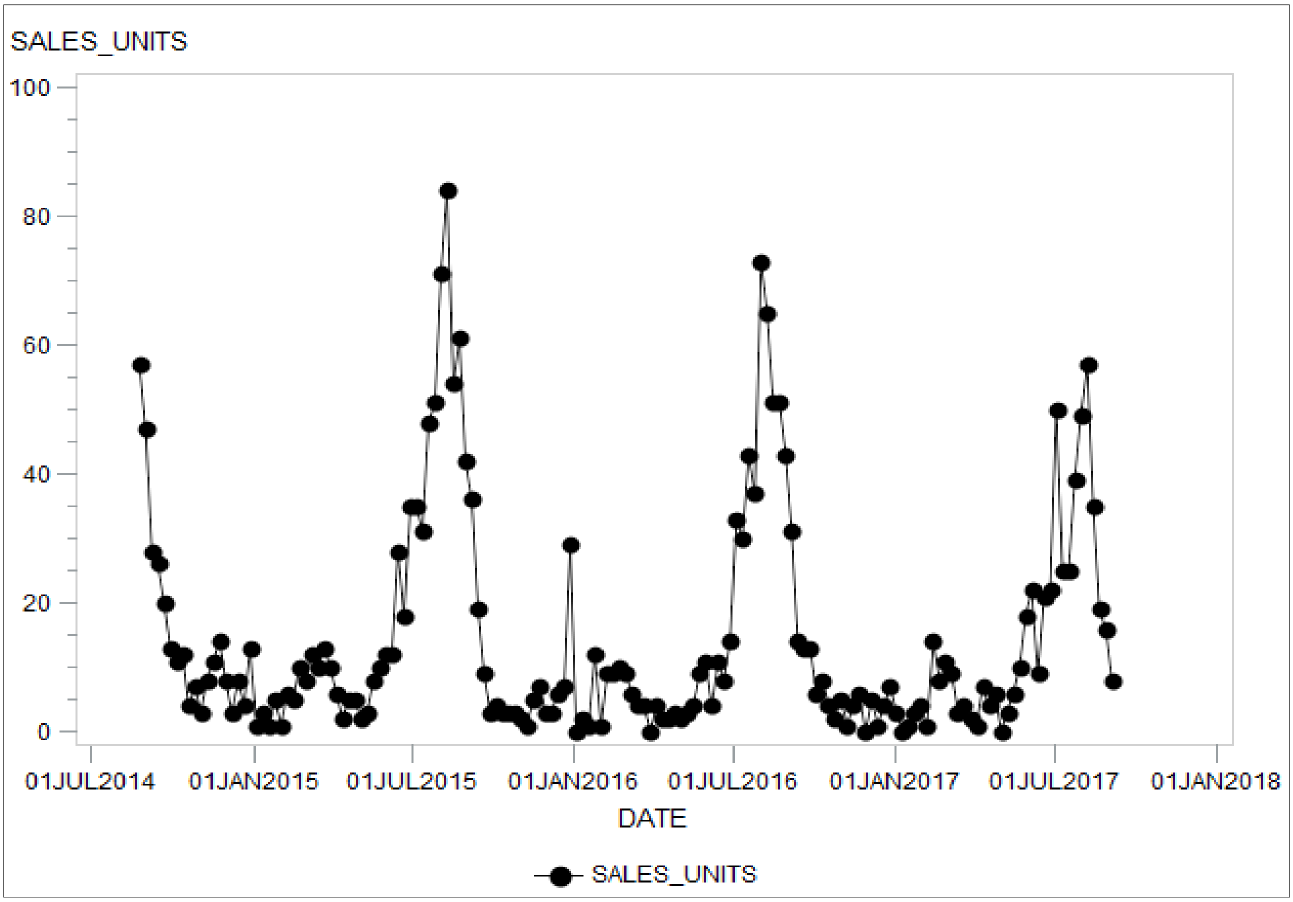


Fig 4: It illustrates the sales pattern of a geo-prod at State and Sub-Category level.

If we aggregate a segment to Store and SKU level, plot looks as shown below.

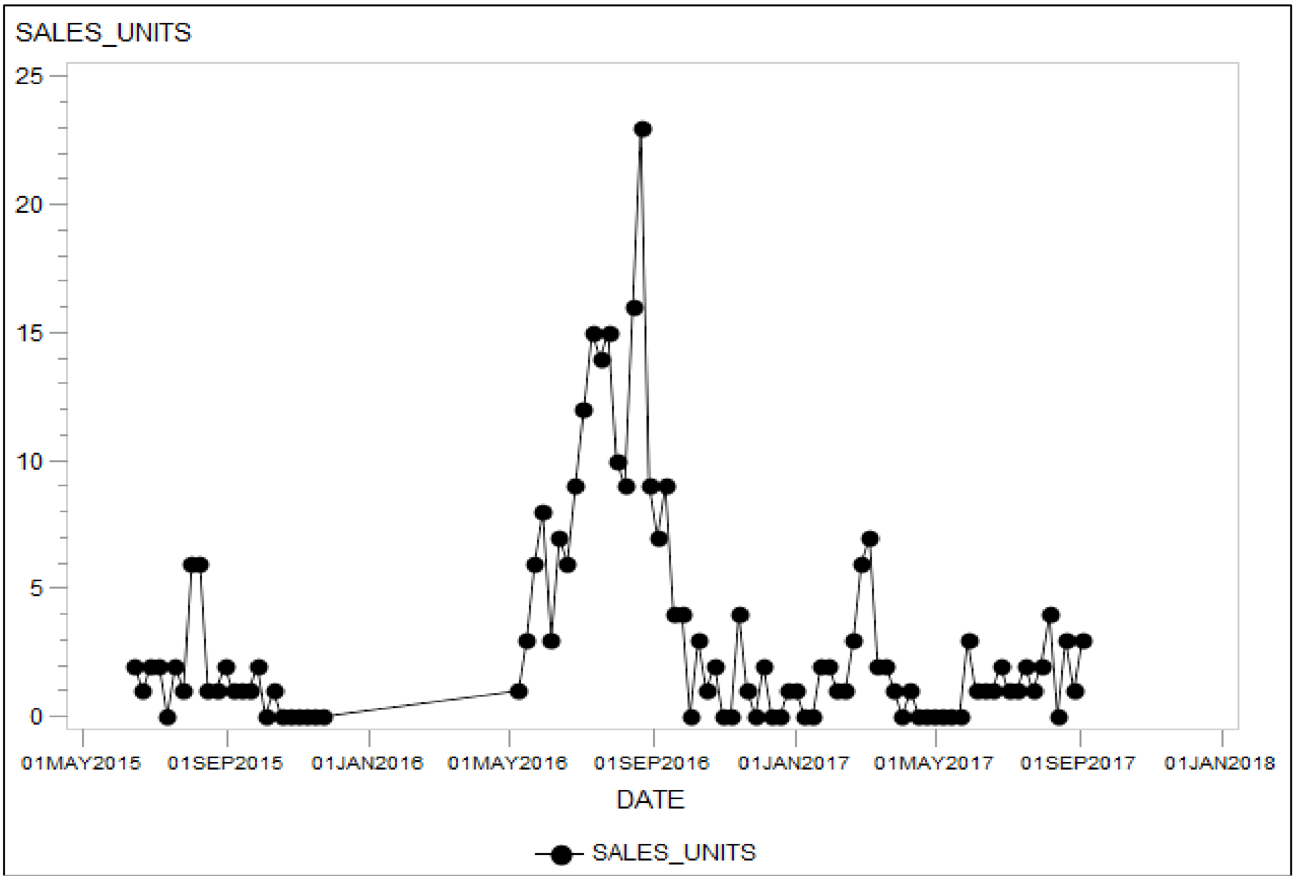


Fig 5: It illustrates the sales pattern of a geo-prod at Store and SKU level.

So, it is better to build model at top levels in the same segment instead of building models at very low level. The aggregation levels should satisfy the conditions of data sufficiency and consistency. Then only fitted models can outfit the results. If we observe above figures (3,4 and 5), Figure 3 has consistent and sufficient data compared to other two levels. A forecasting model can outperform on region and category level aggregated data as there is a specific pattern in the data. So, a forecasting model can extract timeseries components effectively.

Missing Values

Missing values are very common in any type of data. They may appear in time series for number of reasons including recording failure or previous attempts at cleaning the data. Missing values are presented in a timeseries data by a sequence gap in the time values. Before going to build statistical models on the data, need to replace missing values with proper values. There are so many ways to replace missing values. Here we have described about four methods. They are,

a. The missing observations are replaced with median value of the time series of observed values. But this method is outlier sensitive methods. If there are any outliers in the data, they may influence the median value.
b. Missing values may be replaced by using cubic spline interpolation method. This method is most popular of spline interpolation.
c. We can replace missing values by using AR (1). For this method, all the time series data prior to missing value are used to forecast the missing value. The first point of this forecast is used as the estimate for the first missing value.
d. We can replace current missing values by using average of same period(s) in previous years available information.

Outliers

Outliers are the observations that differ extremely from the rest of the observations in the timeseries data. These outliers may change the mean of the series resulting spurious results. These outliers may appear due to the following reasons.

a. Natural disaster or war.
b. Human errors: Mistakes at data entry or mistakes at data collection.
c. Intentionally reporting incorrect data.

Detecting outliers is important because they have an impact on the selection of the models, estimation of the parameters and forecasts. In timeseries data, outliers classified into 5 categories. They are

a. Additive Outliers
b. Temporal Changes
c. Level Shifts
d. Innovation Outliers
e. Seasonal Level Shift

There are number of techniques to identify the outliers in the data including STL decomposition, Classification and Regression Trees, ARIMA, Exponential smoothing, Neural networks, etc. We need to detect and adjust identified outliers by using suitable outlier detection methods.

Demand Factors

Three types of variables can be used in a model. They are dependent, independent and derived variables. Independent variables are the input to the model, dependent variable is the output of the model.

In demand forecasting, the most commonly used dependent variable is sales volume. Independent variables are Standard price, Current price, inventory volume, promotions, holidays, events and product life cycle.

From the above information, we can derive a variable called price ratio as the ratio of standard price and current price. We can use price ratio as an independent variable instead of standard and current price. If we miss any of these variables, forecasts may be not reliable.

Conclusion

Accurate sales forecast plays pivotal role in all the planning systems of any business type. Poor forecast accuracy results in making wrong assumption and leads to wrong decisions. It has proven that higher accuracy can be achieved with right data preparation from many studies. Data preparation process takes 80% of the time of any analytics project because we must make sure proper data fed into the model otherwise it is like “Garbage in and Garbage out”. We have described 10 major data challenges can face in the data related to fashion retail while building sales demand forecast models. This paper also talks about possible remedies to overcome the data issues. These remedies help to maintain data consistency and sufficiency which leads to higher accuracy. We attained 10%-15% accuracy improvement compares to base models by using these remedies.

The main objective of this paper is to list major data challenges and remedies at very high level at one place. The list is not limited to fashion retail, also applicable to other realms. We are working on complete details, empirical studies and the benefits for each data challenge mentioned in this paper as our future work.

References

1. Rajasekhar Kalla, Saravana Murikinjeri, Abbaiah R, Venkatarama Raju. Price elasticity model for fashion products. Global Journal of Pure and Applied Mathematics, 2017.
2. Minner S, Kiesmuller GP. Dynamic Product aquisition in Closed Loop Supply Chains. International Journal of Production Research. 2012; 50:2836-2851.
3. Chung Chen, Lon-Mu Liu. Joint Estimation of Model Parameters and Outlier Effects in Time Series. Journal of the American Statistical Association. 1993; 88(421):284-297.
4. Box GEP, GW Jenkins. Time Series Analysis, Forecasting and Control, Second Edition. Oakland, CA: Holden-Day, 1976.
5. Forecasting. Methods and Applications, 3rd edition by Spyros Makridakis; Steven C. Wheelwright; Rob J Hyndman.
6. Elements of Forecasting by Francis X. Diebold Review by: William M. Briggs Journal of the American Statistical Association. 1999; 94(445):345-346.
7. Regina Kaiser, Agustin Maravall. Seasonal Outliers in Time Series.
8. Maria Elena Nenni, Luca Giustiniano, Luca Pirolo S. Demand Forecasting in the Fashion Industry: A Review. International Journal of Engineering Business Management Special Issue on Innovation in Fashion Industry.
9. Master’s thesis, Johanna Kaperi; New Product Forecasting – Empirical evidence from Finnish textile companies. Alto Univeristy, School of Business.
10. Kenneth B, Kahn. An exploratory Investigation of new product forecasting practices; The university of Tennesse, Department of marketing, Logistics and Transportation; The journal of Product Innovation Management. 2002; 19:133-143.
11. Dr. Neethu Sharma. Marketing Strategy on different stages plc and its marketing implications of FMCG products; International Journal of Marketing, Finance Services & Management Research, 2013, 2(3). ISSN2277-3622.
12. Yazdan Mansuri, Sayed Reza Hashemi, Fakhroldin Maroofi. Factors Affecting Demand Management in the Supply Chain; International Journal of Agriculture and Crop Sciences.