भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

IMB 562

DEEPAK GEORGE, KARTHIK KURAM, RAMALAKSHMI SUBRAMANIAN, SUMAD SINGH AND U. DINESH KUMAR

TEACHING NOTE

MARKDOWN OPTIMIZATION FOR AN INDIAN APPAREL RETAILER

CASE SYNOPSIS

Siddharth Sinha is the CEO of an apparel retailer WE SELL STYLE (WSS). The retail chain was set up in 2008, housing more than 100 brands. In 2015, they operated over 200 stores in all four regions of the country. They primarily focused on providing good quality fashion at a remarkably low price.

Markdown planning has been an important aspect of the apparel business. It is important to understand that the demand for fashion apparel is seasonal – affected by current fashion, variations in the seasons, festivals and hence difficult to estimate. An apparel retailer could go off target – either by overestimating or underestimating the demand, with overestimating being prevalent.

The ordering–manufacturing–stocking cycle is easily a 6-month cycle before the selling actually starts; with an expectation to improve sales year on year, the procurement team buys more, making an increase in the variety of colors and styles to offer more to the consumer. However, not all styles sell as expected, leaving higher than expected stocked inventory, which requires an impetus to sell. The impetus in the industry comes in the form of ‘‘end of season markdown (EOSS)’’.

Decision on the percentage of markdown for EOSS is one of the most critical tasks for an apparel retailer. This activity starts months ahead of the EOSS. The product team and the planning team sit together and come up with an EOSS plan at the style level. In the decision process, procurement and planning team use their domain expertise and judge the performance of style using metrics such as rate of sales, full price sell-through, inventory left, and more. The key decision is to quantify the degree of non-performance of styles that did not sell as forecasted and by how much to markdown for the EOSS.

This note was prepared by Deepak George, Karthik Kuram, Ramalakshmi Subramanian, Sumad Singh and U Dinesh Kumar, Professor of DS&IS, for the sole purpose of aiding classroom instructors in the use of Markdown Optimization for an Indian Apparel Retailer, Case No. IMB 561. It provides analysis and questions that are intended to present alternative approaches to deepening students’ comprehension of the business issues presented in the case and to energize classroom discussion.

TEACHING OBJECTIVES

This case can be used in the Business Analytics/Data Science course of MBA and executive MBA programs. The case is ideal for teaching markdown business strategy used by retailers and introductory/advanced-level concepts in clustering, forecasting, and optimization. The case helps the instructor to demonstrate applications of hierarchical clustering, time series forecasting using regression, and non-linear optimization. The instructor may download all the data sets and the corresponding R codes from the following link:

http://hrm.iimb.ernet.in/iimb/Download/index.htm

REFERENCE MATERIAL

The following textbooks may be used as references.

1. Tudor Bodea and Mark Ferguson, *Pricing: Segmentation and analytics*, Business Expert Press, 2012, New York.
2. Rob J Hyndman and George Athanasopoulos, *Forecasting: principles and practice, University of Western Australia Report, 2014.*
3. D N Gujarati, D C Porter and S Gunasekar, *Basic Econometrics*, 5th Edition, Tata McGraw Hill Education Private Limited, 2012.

LEARNING OBJECTIVES

The primary objective of the case is to demonstrate how analytical models such as clustering, forecasting, and non-linear optimization can be leveraged by retailers to maximize revenue during end of season sale (EOSS) by arriving at the optimal markdown discount for the merchandise.

Other learning objectives include the following.

1. Develop clustering model to segment stores so that appropriate pricing strategy can be devised for each of the store clusters using *R*.
    a. Validating cluster results
    b. Visualizing and profiling clusters
2. Develop a multivariate time series forecasting model using regression.
    a. How do we perform exploratory data analysis (EDA)?
    b. How can regression forecasting model be built and checks for autocorrelation in residuals, heteroscedasticity, etc. conducted?
    c. How can non-linear relationship be accounted in the model?
    d. How do we quantify price elasticity, promotion/aging effect, etc.?
    e. Generate future forecasts and perform out-of-time validation of model.

3. Develop a non-linear optimization model to identify optimal discount which maximizes profit.
    a. How can demand function from forecasting model be used in optimization model so that optimization model can iterate with different values of discount % and finally arrive at the optimal discount which maximizes revenue?

SUGGESTED CASE QUESTIONS

PART A – CLUSTERING

1. List and derive the metrics that can be used in ‘‘hierarchical clustering’’ and ‘‘partition around medoids’’ clustering algorithms. (Note: Use the data in sheet “Clustering_Raw_data.xlsx”).
2. Do you find outliers in the derived data from Q1? If yes, how can the same be treated for use in cluster modeling?
3. Develop a hierarchical clustering model with the modified data from Q2. How many clusters seem appropriate? Justify.
4. Develop a partition around medoids clustering model with the modified data from Q2. What are the advantages of using partitioning around medoids (PAM) over *K*-means? How do you decide on the appropriate number of clusters in this scenario?
5. Validate the goodness of resulting clusters from hierarchical and PAM models obtained in Q3 and Q4. Which is a better model as per validation measures?
6. Having selected the most appropriate clustering model from Q5, perform cluster profiling and list the unique characteristics of each cluster.

PART B – TIME SERIES FORECASTING

7. Conduct exploratory data analysis on “Cluster=1, Brand=CRESCENT SET, Brick=CKD” combination to identify the following relationships. Give a short description about the relationships observed.
    a. Relationship between Sales units(sales_units) & Discount % (discount_per)
    b. Relationship between Sales units & Net Price (per_unit_netprice)
    c. Relationship between Sales units & Age (age)

   Note: 1) Use data in the following csv files “1. CRESCENT SET.CKD.csv” to answer this question.
   2) Variables names to be used are mentioned in brackets.
8. What is over-fitting and under-fitting in the context of regression models? What are the consequences of over-fitting?
9. Explain why we have to partition the time series data before building the forecasting model. Use data for “Cluster=2, Brand=BLINK, Brick=HAREMS” and partition this time series data as explained below.
    i. Consider all weeks until 51st week (including) of 2014 as training data.
    ii. Consider weeks from 52nd week of 2014 to 3rd week of 2015 as test data.

   Note:
    - These 4 weeks are winter EOSS weeks.
    - User data in “2.BLINK.HAREMS.csv” only for answering questions from now onwards.
10. Develop a time series forecast model using regression on the training data to forecast sales units for “Cluster=2, Brand=BLINK, Brick=HAREMS” combination using the below variables as predictors.
    a. Lag 1 (i.e. immediate previous week) of sales units
    b. Discount %

c. Lag 1 (i.e. immediate previous week) of discount %
d. Promotion week flag
e. Age

Apply appropriate transformations and evaluate the model fit.

11. Perform checks to ensure that the model is valid and assumptions of regression are met. Conduct appropriate statistical test and back the findings by visual examination of relevant plots.
12. Based on the model result, explain the following:
    a. How is this forecasting model able to account for trend and seasonality?
    b. What is price elasticity and determine the price elasticity value (a proxy representing price elasticity is enough) from the model output?
    c. How do you interpret the coefficient of promotion week flag variable?
13. How do you check if the forecasting model is able to explain most of the important features of the time series? Explain white noise in the context of time series.
14. Using the forecast model built, generate sales units forecast for test period (52nd week of 2014 to 3rd week of 2015).
    a. Assess the forecast model accuracy on the test time period which is not used for modeling by calculating MAPE for the test period.

    Note: In "Forecast_test_week_predictor_input.xlsx" sheet "Input data for forecasting" contains the data required for forecasting sales and sheet "Actual Sales" contains the actual sales generated.

PART C – OPTIMIZATION

15. Formulate an optimization model and solve it to determine the optimal discount % to be given for "Cluster=2, Brand=BLINK, Brick=HAREMS" combination for each of the 4 weeks of EOSS.
    a. Objective function: Maximize the total revenue generated during EOSS.
        i. Total revenue is defined as the revenue generated during 4 weeks of EOSS and revenue from the left over inventory after EOSS. Assume the residual inventory left over even after EOSS is liquidated by giving a flat discount of 60%.
    b. Constraints
        i. Sales units in a week should be less than equal to starting inventory of the same week.
        ii. Relationship between demand (sales unit) for a week and predictors should follow the relationship identified in the forecasting regression model.
        iii. Discount for a week should be greater than equal to previous week's discount and discount should vary between 10% and 60% only.
        iv. Sales & inventory values for a week should be ≥ 0.
        v. Inventory should be updated on basis of the sales for a week.
    c. Inputs required
        i. Inventory at the beginning of EOSS = 2,476
        ii. Age of the brick at the beginning of EOSS = 96 weeks
        iii. Previous week discount = 57.9%
        iv. Previous week sale = 48
        v. MRP of the brick = INR 606

16. What are the weekly forecasted sales units if the optimal discounts identified are implemented for the EOSS?
17. We know that actual revenue realized by the retailer for "Cluster=2, Brand=BLINK, Brick=HAREMS" combination during the 4 weeks of EOSS is INR[1] 41,320. Then, what is the incremental lift in revenue the retailer would have achieved in these 4 weeks if he/she implemented our analytics solution instead?

RECOMMENDED SOLUTION TO THE QUESTIONS

PART A – CLUSTERING

1. List and derive the metrics that can be used in ''hierarchical clustering'' and ''partition around medoids'' clustering algorithms. Note: Use the data in "Clustering_Raw_data.xlsx".

Recommended Solution

The raw data (refer to Clustering_Raw_data.xlsx) has the following variables:

a. Store ID (Categorical Variable)
b. Store Name (Categorical Variable)
c. Store Area (Continuous Variable)
d. Zone (Categorical Variable)
e. Net Sales in INR for various brands in women's ethnic wear (Continuous Variable)
f. Total Discount in INR for various brands in women's ethnic wear (Continuous Variable)
g. Total Cost of Goods Sold in INR for various brands in women's ethnic wear (Continuous Variable)

Since Store ID and Store Name are just identifier variables, the same cannot be used as an input metric in clustering algorithms. Further, since the retailer wants to maximize the revenue by optimizing the markdown strategy during EOSS, the following metrics can be derived at store level by using the continuous variables:

1) **Markdown Sensitivity:**
Derived as follows:

$$\left(\frac{\text{Net sales for each brand in a store}}{\text{Discount given for the respective brand in the same store}} \right)$$

This metric would tell us how much sales was generated for every rupee of discount given on a particular brand during the festive winter EOSS by a particular store. So, higher markdown sensitivity implies lower discount was given and vice versa.

2) **Net Profit per square foot**
Derived as follows:

[1] 1 USD = INR 65.8 in May 2016

$$\left(\frac{\text{Net sales for each brand – Total cost of goods sold for the same brand}}{\text{Respective store area}} \right)$$

Note: Both the metrics derived above are at store level. Also, the net profit has been normalized by the respective store area for an appropriate comparison.

3) **Location**
Clustering algorithms usually require the use of continuous variables as input metric. Categorical variables such as location can also be used as an input metric in hierarchical and PAM algorithms, provided specialized distance measure such as Gower's distance (not Euclidean) is used for the computation of distance matrix. Hence, along with markdown sensitivity and net profit per square foot, location can also be used as an input metric in the hierarchical and PAM algorithms.

Refer to Clustering_Steps.xlsx [Festive Winter Derived Data] for the computation of derived input metrics.

2. Do you find outliers in the derived data from Q1? If yes, how can the same be treated for use in cluster modeling?

Recommended Solution

Presence of outliers may result in the formation of isolated clusters. To avoid the same, outliers are usually identified and treated before the input metrics are supplied to the clustering algorithms. Outliers may be identified by looking at the mean of each input metric and finding any unusually high measurement. In the derived data, we can see that there are few stores, across different brands showing unusually high measurement on markdown sensitivity and net profit per square foot. These outliers need to be treated. One of the popular methods of outlier treatment in clustering is to replace the unusually high measurements with a suitable percentile value (example: 97^{th} or 99^{th}) of the respective input metric. The advantage of replacing the actual measurement with percentile value is that, even by replacing the actual measurement, the essence of the store performance is not lost and store would still remain a high performance store while curtailing the formation of isolated clusters

Refer to Clustering_Steps.xlsx [Outlier Detection] for the identification of outliers

Refer to Clustering_Steps.xlsx [Model Data] for data to be used in clustering after outlier treatment.

3. Develop a hierarchical clustering model with the modified data from Q2. How many clusters seem appropriate? Justify.

Recommended Solution

In order to run hierarchical clustering, the following parameters are required.

a) Distance matrix as calculated from the model data.
b) Agglomeration method to be used.

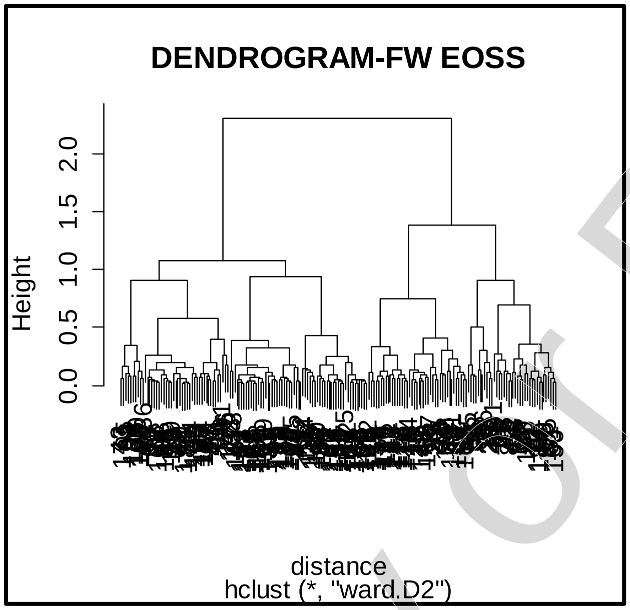
Distance matrix in the present case is a proximity measure between the stores on the input measurements. We can use distance measures such as Euclidian, Maximum, Manhattan, Canberra, Minkowski, Gower, etc. Before the calculation of distance matrix, it is advisable to center or scale the input metrics. Scaling is performed by subtracting the mean of every input metric from each element (store measurement) and dividing by the standard deviation of the input metric.

Euclidian distance is the most popular distance measure used in clustering. However, Euclidian distance would not be meaningful in the presence of categorical variables. Since, we have location as one of the input metrics, we cannot use Euclidian distance. Gower's distance is a very popular distance measure used when the input metrics are a mix of categorical and continuous variables. Hence, in the present scenario, we would compute distance matrix using Gower's distance.

There are many agglomeration methods such as single, complete, average, Ward's, McQuitty, centroid, etc. The choice of agglomeration method would depend on a particular case. For the present scenario, we can use Ward's method which tries to minimize the variance within a cluster.

It is important to note that hierarchical clustering treats each store as a singleton cluster at the outset and then successively merges pairs of clusters based on the method used until all clusters have been merged into a single cluster that contains all the stores. Hierarchical clustering can be graphically visualized by using dendrograms.

Hierarchical clustering does not require the analyst to pre-specify the number of clusters to be formed. It produces a hierarchy of clusters starting from singleton store to a single cluster containing all the stores. The analyst has to decide to cut the hierarchy at some point to derive the requisite number of clusters. This would depend on factors such as the business objectives and dissimilarities in clusters.



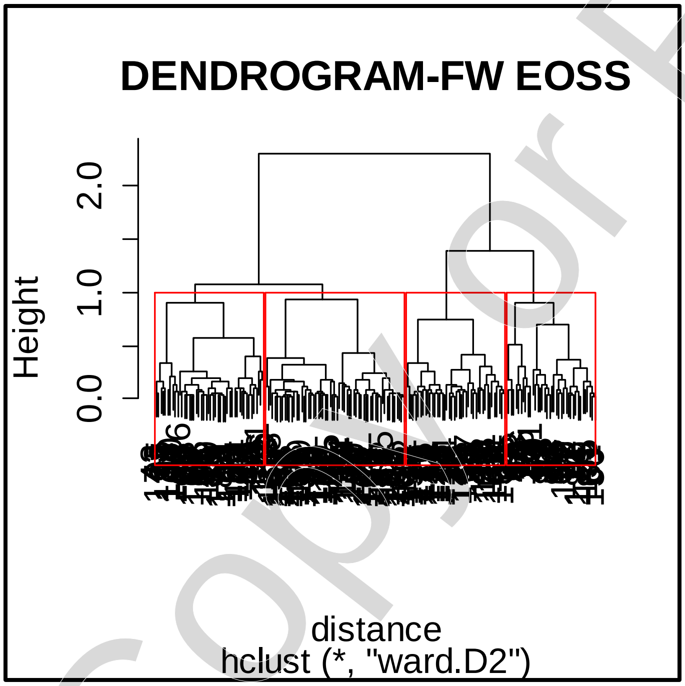
In the present scenario, we can cut the dendrogram at a pre-specified level of similarity or where the gap between two successive combination similarities is large. An extract of the gap between successive combination similarities (from 10 to 1 cluster) is shown below:

Iteration	Cluster No.	Merging with Cluster No.	Merge Height	From_to_Next _Cluster	% Change in Height
175	147	166	0.5078	10	14.33
176	159	172	0.5806	9	19.09
177	163	170	0.6914	8	8.3
178	168	173	0.7488	7	20.85
179	169	176	0.9049	6	0.28
180	175	177	0.9074	5	3.18
181	171	174	0.9363	4	14.75
182	179	181	1.0744	3	29.01
183	178	180	1.3861	2	66.7
184	182	183	2.3106	1	-100

The table can be read as follows:

At iteration 175 of the hierarchical algorithm, cluster number 147 merges with cluster number 166 at a height of 0.5078. At this juncture, there are 10 clusters in the dataset. If the number of clusters is reduced from 10 to 9, then there would be a 14.33% change in the homogeneity (height being a proxy measure for homogeneity).

Since the objective of clustering is to group stores based on its similarity of chosen performance metric, so that differing markdown strategy can be applied for every identified cluster of stores, it is important to pay particular attention to the number of clusters that can be formed. Too few clusters would not yield the desired result and too many clusters would lead to non-implementable strategy. Considering the zones in which the stores operate, business objectives, implementation ease, and change in homogeneity (14.75% is a large change if we choose 3 clusters instead of 4), we can see that 4 would be the appropriate number of clusters We cut the dendrogram so as to arrive at 4 clusters.



Refer to Hierarchical Clustering.R for all the codes required to reproduce the hierarchical cluster modeling.

4. Develop a partition around medoids clustering model with the modified data from Q2. What are the advantages of using PAM over *K*-means? How do you decide on the appropriate number of clusters in this scenario?

Recommended Solution

Partition around medoids (PAM) is a more robust version of *K*-means. It is less sensitive to outliers as compared to *K*-means. Another advantage of using PAM as compared to *K*-means is the use of categorical variable as input metric. So in the present scenario, we can also use location as an input metric which cannot be used with *K*-means clustering owing to its restrictions on use of categorical variables.

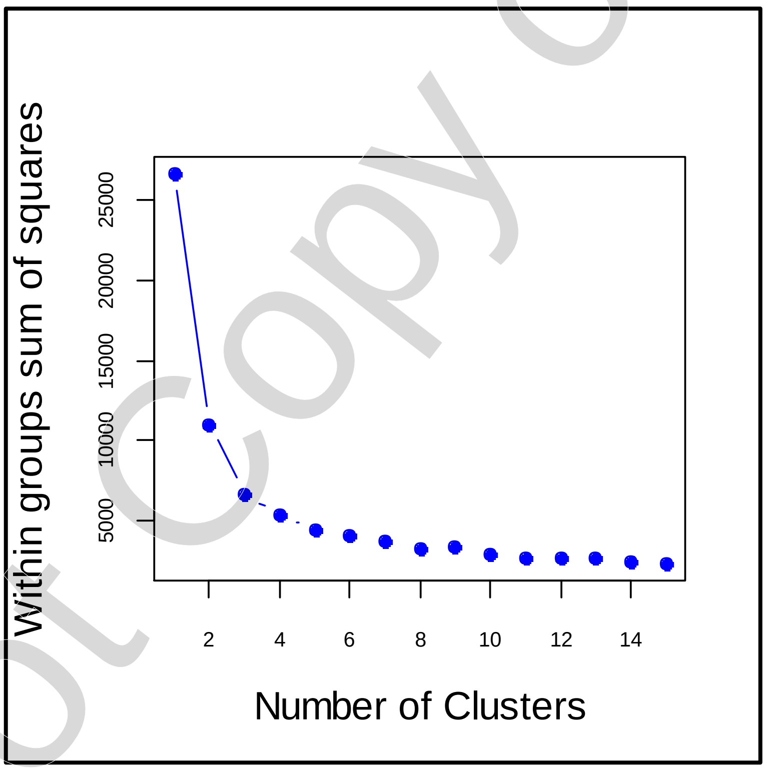
In order to run PAM, the following parameters are required:

a) Distance matrix as calculated from the model data.

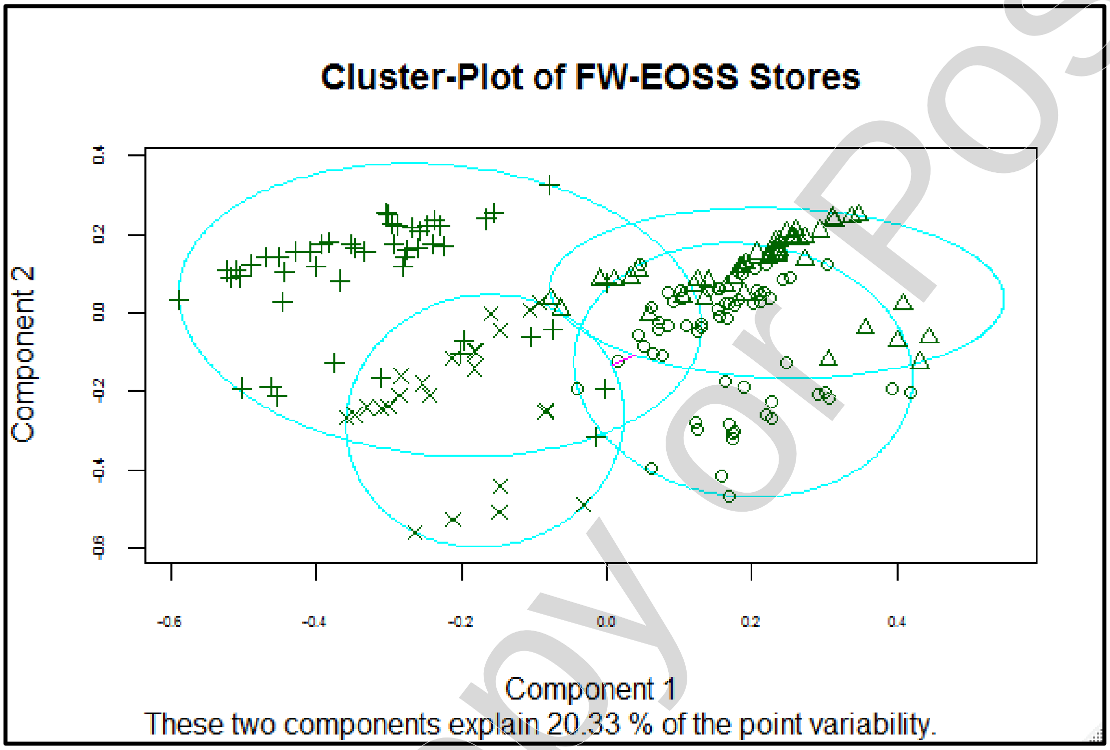
b) Pre-specified number of clusters to be formed.

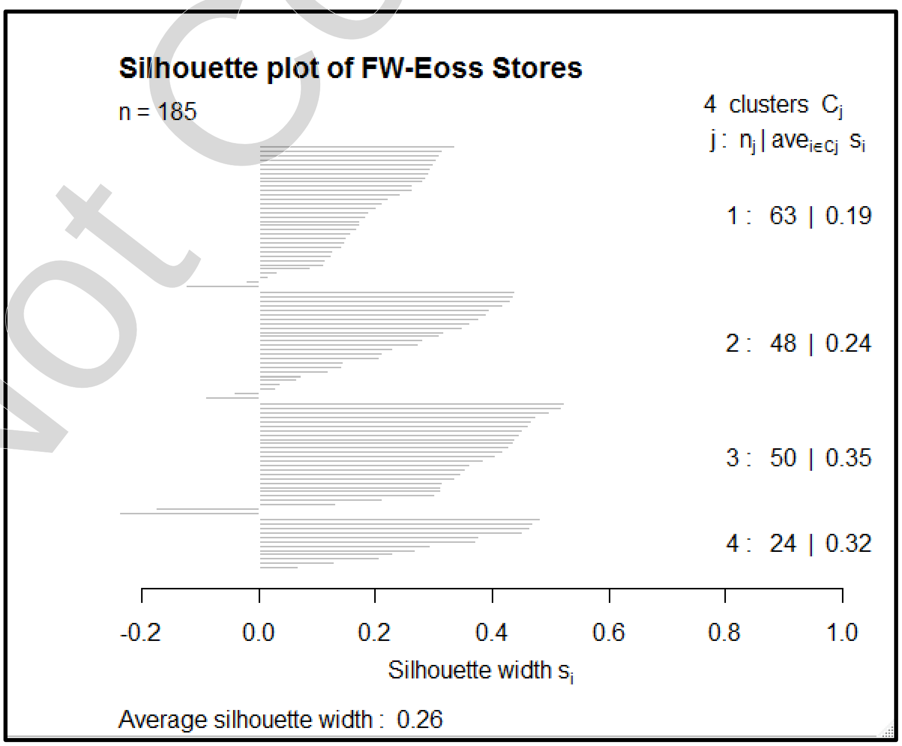
For distance matrix, we can use the same matrix generated for hierarchical clustering using Gower's distance. For deciding on the appropriate number of clusters, we can compare the sum of squared errors (SSEs) for a number of cluster solutions. A plot of SSE against a series of sequential cluster levels can provide a useful graphical way to choose an appropriate cluster level. An appropriate cluster number could be the number at which the reduction in SSE slows down drastically. Therefore, we look for the elbow in such a plot and decide on the appropriate number of clusters.

The following plot was generated using the model data to decide on the appropriate number of clusters.



By looking at the above plot, we see that 4 would be a reasonable number of clusters even under PAM. We run the PAM algorithm with distance matrix calculated from Gower's distance and 4 as the number of clusters. We can visualize the clusters formed and its separation using cluster plots and silhouette plot.





Refer to PAM Clustering.R for all the codes required to reproduce the PAM cluster modeling.

5. Validate the goodness of resulting clusters from hierarchical and PAM models obtained in Q3 and Q4. Which is a better model as per validation measures?

Recommended Solution

The resulting cluster solutions are usually validated with some of the popular validation measures to assess their goodness. These are as follows:

a) Average silhouette width
b) Within clusters sum of squares
c) Minimum cluster separation
d) Maximum cluster diameter
e) Average distance within clusters
f) Average distance between clusters
g) Dunn Index

The following table shows the ideal values for these validation measures:

Validation Measure	**Ideal Value**
Average silhouette width	Higher the better
Within clusters sum of squares	Lower the better
Minimum cluster separation	Higher the better
Maximum cluster diameter	Lower the better
Average distance within clusters	Lower the better
Average distance between clusters	Higher the better
Dunn Index	Higher the better

We can calculate the above measures for both hierarchical and PAM models developed in Q3 and Q4 using cluster.stats function of fpc package in *R*. The results are tabulated as follows:

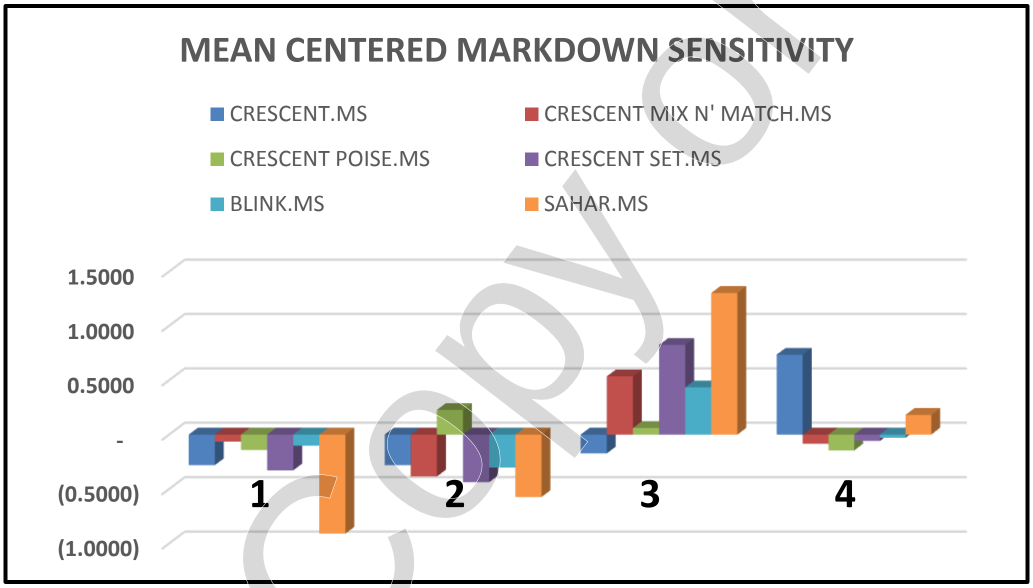
	Hierarchical	PAM
avg.silwidth	0.2671866	0.2629391
within.cluster.ss	3.728455	3.60799
min.separation	0.08496225	0.01994104
max.diameter	0.6377485	0.4799404
average.within	0.1786602	0.180527
average.between	0.3035302	0.3055732
dunn	0.1332222	0.04154899

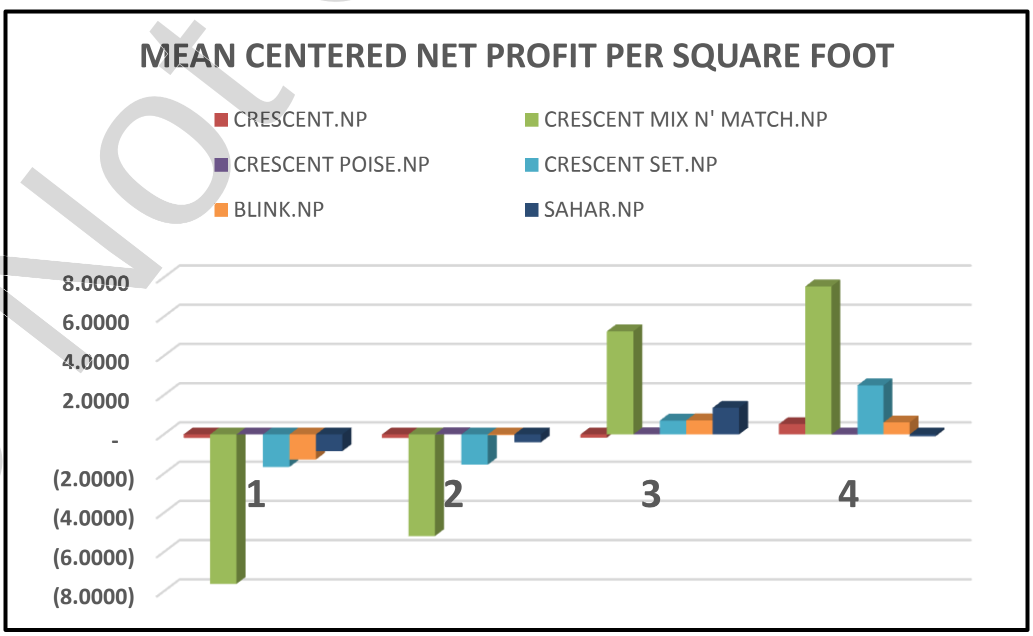
From the above validation measures, it is clear that hierarchical clustering scores marginally better than PAM and hence we select hierarchical clustering as the most appropriate method for the given data. Refer to Cluster Validation.R for codes required to reproduce the above table.

6. Having selected the most appropriate clustering model from Q5, perform cluster profiling and list the unique characteristics of each cluster.

Recommended Solution

Cluster profiling is the last step in the process of clustering. After the appropriate clustering algorithm has been identified and clusters obtained, we proceed to profile the stores in each cluster by determining the unique characteristics. For this purpose, we can plot the mean centered markdown sensitivity and mean centered net profit per square foot as shown here.





The insights generated from cluster profiling can be effectively used by the business to devise appropriate markdown strategy in order to maximize their revenue. From the graphs, the following unique features can be deduced:

a) Cluster 1:
   1. Mostly dominated by stores located in north and east zones.
   2. Brands Crescent and Sahar have the least markdown sensitivity in these sets of stores, that is, high discounts were given on these brands to generate sales.
   3. Brands Crescent Mix and Match, Crescent Set, Blink, and Sahar have the least net profit per square foot in these sets of stores.

b) Cluster 2:
   1. Mostly dominated by stores located in west zone.
   2. Brands Crescent Mix and Match, Crescent Set, and Blink have the least markdown sensitivity in these sets of stores.
   3. Brand Crescent Poise has the highest markdown sensitivity in these sets of stores, that is, less discounts were given on this brand to generate sales.

c) Cluster 3:
   1. Dominated by stores located in south zone.
   2. Brands Crescent Mix and Match, Crescent Set, Blink, and Sahar have the highest markdown sensitivity in these sets of stores.
   3. Brands Blink and Sahar have the highest net profit per square foot in these sets of stores.

d) Cluster 4:
   1. Dominated by stores located in south zone.
   2. Brand Crescent has the highest markdown sensitivity in these sets of stores.
   3. Brands Crescent, Crescent Mix and Match, and Crescent Set have the highest net profit per square foot in these sets of stores.

Refer to Clustering_output.csv for the output of hierarchical clustering and Cluster_Profiling.xlsx for all the computations required to reproduce the mean centered markdown sensitivity and net profit per square foot graphs.

Conclusion: We can see from cluster profiling that a certain set of stores performs very well in certain brands on chosen input metrics and a certain set of stores does not perform well in certain brands on the chosen input metrics. Clustering is a scientific way of segmenting stores based on chosen input metrics. The insights generated from clustering can be used by the business to devise differing markdown strategy across clusters rather than a uniform markdown strategy across all the stores in the country.

PART B – TIME SERIES FORECASTING

Important Note

Data for "Cluster=1, Brand=CRESCENT SET, and Brick=CKD" is present in "1.CRESCENT.SET.CKD.csv" file included as part of this case.
Data for "Cluster=2, Brand=BLINK, and Brick=HAREMS" is present in "2.BLINK.HAREMS.csv" file included as part of this case.

7. Conduct exploratory data analysis on "Cluster=1, Brand=CRESCENT SET, and Brick=CKD" combination to identify the below relationships. Give a short description about the relationships observed.
   a. Relationship between Sales units(sales_units) & Discount % (discount_per)
   b. Relationship between Sales units & Net Price (per_unit_netprice)
   c. Relationship between Sales units & Age (age)

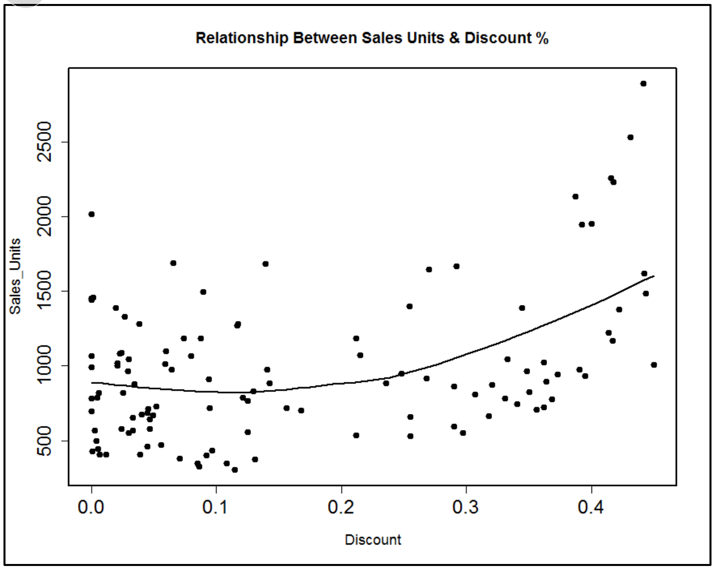
   Note: 1) Use data in the following csv files "1. CRESCENT.SET.CKD.csv" to answer this question.
   2) Variables names to be used are mentioned in brackets.

Recommended Solution

Data visualization as a part of exploratory data analysis is an important step to perform before diving into modeling. Visualizing the target variables relationship with the predictors reveals patterns in the data. It will help us identify presence of non-linearity, type of correlation, presence of outliers, etc.

Plot the target variable (dependent variable) Sales Units in the *Y*-axis and the independent variables in the *X*-axis.

Q.7.a

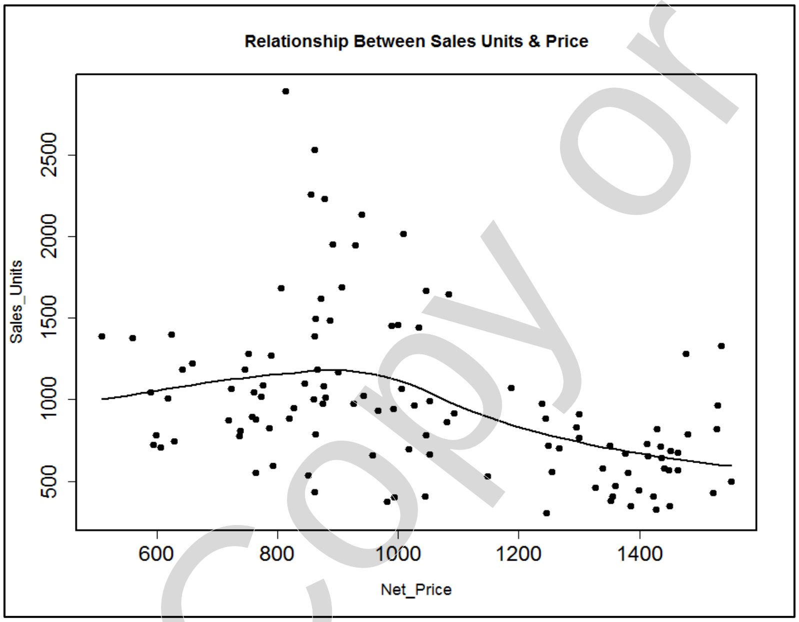


The scatterplot reveals the relationship between discount and sales.

- Correlation between discount and sales is positive.
- Relationship is not linear but non-linear.

If sales generated by giving 20% discount is x, then sales generated by giving 40% discount is not $2x$ but much more than that.

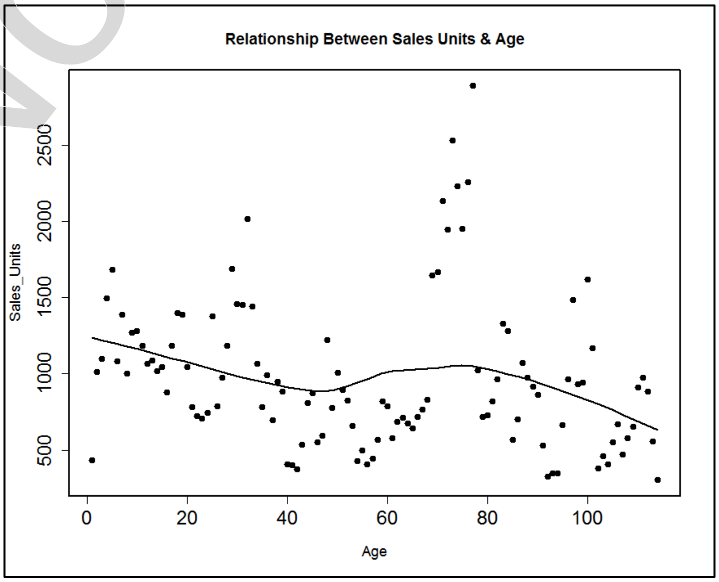
Q.7.b



The above scatterplot reveals the relationship between sales and net price.

- Correlation between discount and sales is negative.
- Relationship is approximately non-linear.

Q.7.c



The scatterplot reveals the relationship between sales and age.

- Correlation between age and sales is negative.
- Relationship is approximately linear.

Note: Run the following *R* codes in order to generate the above plots.

i. 00_data_import.R
ii. 01_time_series_creation_v1.R
iii. 02_exploratory_data_analysis_v1.R

8. What is over-fitting and under-fitting in the context of regression models? What are the consequences of over-fitting?

Recommended Solution

Over-Fitting

Intuitively, over-fitting occurs when a model fits the training data too well such that the model is not generalizable to a new unused test data. Over-fitting occurs when a statistical model starts capturing the noise of the data after it captures the entire signal in the training data. Specifically, over-fitting occurs if the model or algorithm shows low bias but high variance.

Over-fitting is often a result of an excessively complicated model.

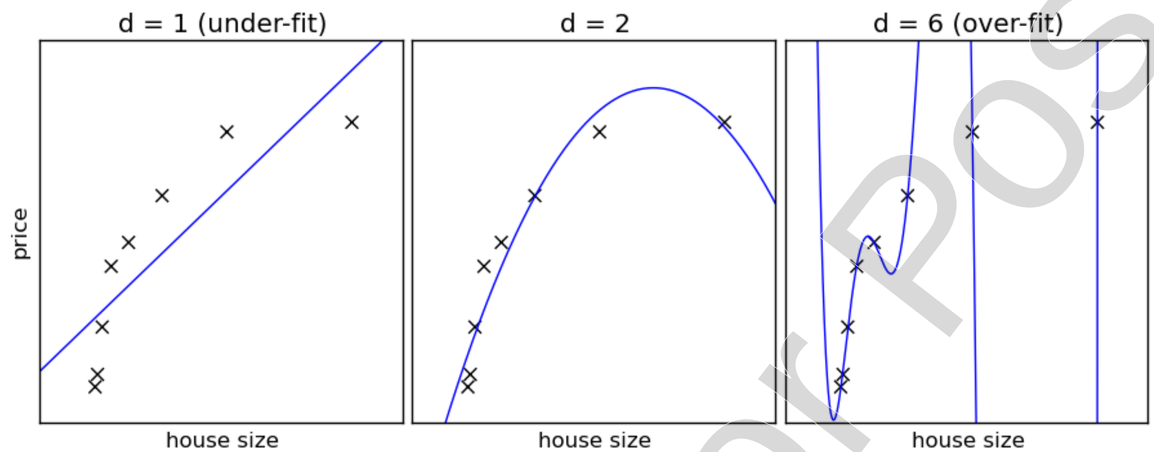
Under-Fitting

Under-fitting occurs when a statistical model or machine learning algorithm cannot capture the underlying trend of the data. Intuitively, under-fitting occurs when the model or the algorithm does not fit the data well enough. Specifically, under-fitting occurs if the model or algorithm shows low variance but high bias. Under-fitting is often a result of an excessively simple model.

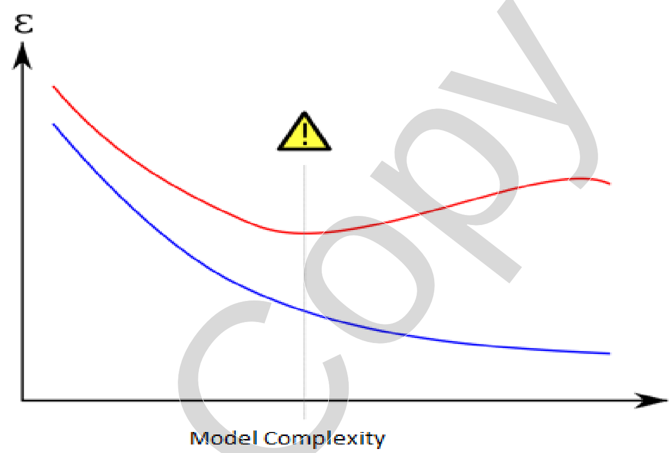
Both over-fitting and under-fitting lead to poor predictions on new data sets.

Illustration

d – Represents the complexity of the model
Here the target/dependent variable is price and independent variable is house size.



<u>Consequence of Over-Fitting</u>



The blue line represents training error and red line represents test error.

We can see that as model complexity increases beyond a particular threshold, the test error increases even though the training error decreases.

9. Explain why we have to partition the time series data before building the forecasting model. Use data for "Cluster=2, Brand=BLINK, Brick=HAREMS" and partition this time series data as explained below.
   i. Consider all weeks until 51st week (including) of 2014 as training data.
   ii. Consider weeks from 52nd week of 2014 to 3rd week of 2015 as test data.

Note:
- These 4 weeks are winter EOSS weeks.
- User data in "2.BLINK.HAREMS.csv" only for answering questions from now onwards.

Recommended Solution

We have to partition the time series data before building the model to avoid over-fitting. In context of time series forecasting problem, we should use out-of-time test data to ensure that the model is not over-fitting.

We are selecting the 4 EOSS weeks as out-of-time test weeks because we will be using this model to forecast the sales during EOSS weeks and not non-EOSS weeks; so that the test data used to measure accuracy of the model is replicating the real implementation scenario.

Note: Run the following R code to partition the time series data.

i. 03_data_partitioning_v1.R

10. Develop a time series forecast model using regression on the training data to forecast sales units for "Cluster=2, Brand=LINK, and Brick= HAREMS" combination using the following variables as predictors.

    a. Lag 1 (i.e. immediate previous week) of sales units
    b. Discount %
    c. Lag 1 (i.e. immediate previous week) of discount %
    d. Promotion week flag
    e. Age

    Apply appropriate transformations and evaluate the model fit.

Recommended Solution

This question requires students to develop a forecasting model using regression. We are using regression model instead of ARIMAX because regression gives more interpretable coefficients, which is required for the optimization model in the next stage.
After exploring different transformation, we found the below approximation of log–log model to be a good fit.

The corresponding regression model is given by:

$$\text{Log(Sales Units)} = \beta_0 + \beta_1 \text{Log(Sales Unit Lag1)} + \beta_2 \text{ Log(Discount \%)}$$
$$+\beta_3 \text{Log(Discount \% Lag 1)} + \beta_4 \text{Promo Flag} + \beta_5 \text{Age} .$$

```
Call:
lm(formula = ln_sales_x ~ ln_salesLag1_x + ln_discount_x + ln_discountLag1_x +
    sswin_promo_flg_x + age_x)

Residuals:
     Min       1Q   Median       3Q      Max
-1.27948 -0.17075 -0.00081  0.24749  0.97315

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.618675   0.543193   4.821 5.93e-06 ***
ln_salesLag1_x     0.509915   0.084620   6.026 3.83e-08 ***
ln_discount_x      0.357265   0.091333   3.912 0.000180 ***
ln_discountLag1_x -0.301669   0.090313  -3.340 0.001229 **
sswin_promo_flg_x  0.455792   0.127738   3.568 0.000585 ***
age_x             -0.014149   0.003408  -4.152 7.60e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3941 on 88 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.843,	Adjusted R-squared:  0.8341
F-statistic: 94.53 on 5 and 88 DF,  p-value: < 2.2e-16
```

$$\text{Log(Sales Units)} = 2.6186 + 0.5099\text{Log(Sales Unit Lag1)} + 0.3572\text{Log(Discount \%)}$$
$$-0.3016\text{Log(Discount \% Lag 1)} + 0.4557\text{Promo Flag} - 0.0141\text{Age.}$$

The *F*-test of the overall significance is a specific form of the *F*-test. It compares a model with no predictors to the model that you specify. A regression model that contains no predictors is also known as an intercept-only model.

***F*-test of the overall significance**

The hypotheses for the *F*-test of the overall significance are as follows:

- **Null hypothesis**: The fit of the intercept-only model and your model are equal.
- **Alternative hypothesis**: The fit of the intercept-only model is significantly reduced compared to your model.

The *F*-statistic *p*-value <2.2e-16 ensures that overall the model is significant.

***T*-test of predictor significance**

Each of the predictors is significant at 5%.

Adjusted R^2

Adjusted R^2 of 83.4% hints at good model fit.

11. Perform checks to ensure that the model is valid and assumptions of regression are met. Conduct appropriate statistical test and back the findings by visual examination of relevant plots.

Recommended Solution

Checking autocorrelation

Durbin–Watson test can be used test for auto-correlation of residuals.

```
> dwtest(lmfit)

        Durbin-Watson test

data:  lmfit
DW = 2.1786, p-value = 0.7052
alternative hypothesis: true autocorrelation is greater than 0
```

The high *p*-value implies that significant autocorrelation does not exist in the model.

Checking Homoscedasticity

Breusch–Pagan test can be used to test for homoscedasticity assumption. The high *p*-values imply that significant autocorrelation does not exist in the model.

Also the following plot of residual vs predicted *Y* does not reveal any pattern which also supports this claim.

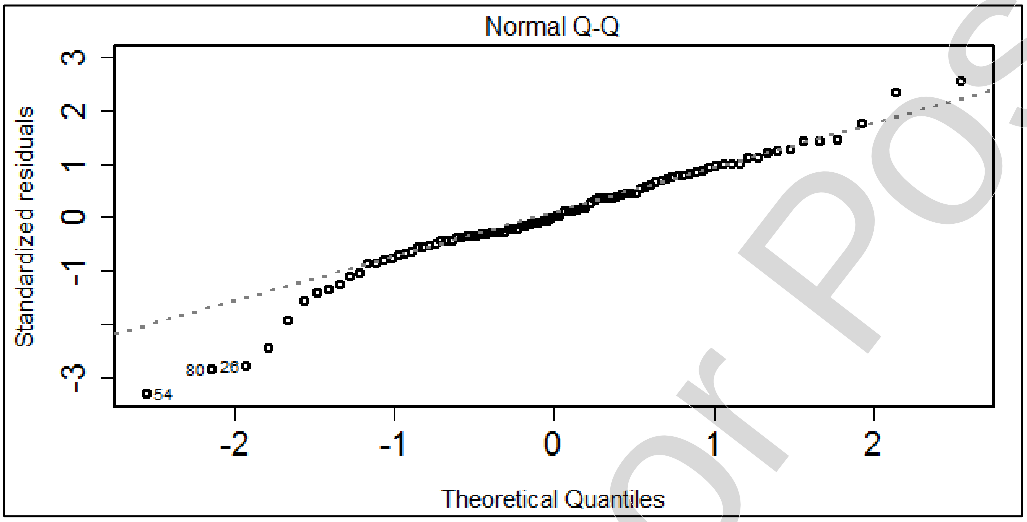
```
> bptest(lmfit)

        studentized Breusch-Pagan test

data:  lmfit
BP = 8.036, df = 5, p-value = 0.1543
```

Check for normality

The following *Q*–*Q* plot shows good approximation to a normal distribution; so, this assumption is not severely violated.



Note: Run the following R code

1) 05_forecast_model_building_v1.R

12. Based on the model result, explain the following:

   a. How is this forecasting model able to account for trend and seasonality?
   b. What is price elasticity? Determine the price elasticity value (a proxy representing price elasticity is enough) from the model output.
   c. How do you interpret the coefficient of promotion week flag variable?

Recommended Solution:

12 a.

The age variable is a straight line which represents the trend of the time series.
The promotion week flag is only 1 during the EOSS weeks; thus, capturing the seasonality factor.

12 b.

Price elasticity shows the relationship between price and quantity sold and provides a precise calculation of the effect of a change in price on quantity demanded.

Strictly speaking, price elasticity = $\frac{\%\text{ Change in Sales}}{\%\text{ Change in Price}}$.
However here, since we take log transformation, etc. the value from the model is not price elasticity in the strict sense.

The coefficient of discount % variable from the model represents the following.

$$\frac{\text{Change in log(Sales Units)}}{\%\text{ Change in log(Discount \%)}}$$

This is a good proxy for price elasticity in this business context and represents how much sales units will change when we increase discount by 1%.

The proxy for price elasticity from the model is 0.357.

12 c.

The coefficient of promotion week indicates how much the sales change in weeks with EOSS promotion compared to weeks without EOSS promotion.

13. How do you check if the forecasting model is able to explain most of the important features of the time series? Explain white noise in the context of time series.

Recommended Solution

If residual from the model is white noise, then it means that the forecasting model was able to ‘‘learn’’ most of the features of the time series. This indicates good model fit.
Ljung–Box test is used to check for white noise.

```
> Box.test(resid(lmfit),type="Ljung-Box")

        Box-Ljung test

data:  resid(lmfit)
X-squared = 1.5155, df = 1, p-value = 0.2183
```

Note: Run the following R code.

1) 05_forecast_model_building_v1.R

14. Using the forecast model built, generate sales units forecast for test period (52nd week of 2014 to 3rd week of 2015).

    a. Assess the forecast model accuracy on the test time period which is not used for modeling by calculating MAPE for the test period.

Note: In “Forecast_test_week_predictor_input.xlsx” sheet “Input data for forecasting” contains the data required for forecasting sales and sheet “Actual Sales” contains the actual sales generated.

Recommended Solution

Students are required to forecast sales units for the next 4 weeks using the predictors indicated in Excel (shown below) and the model equation.

Week	Lag1_Sales_Units	Discount_per	Lag1_Discount _per	Promotion week_flag	Age (weeks)
52.2014	48	0.569002728	0.579304707	1	96
1.2015	Predicted sales in week 52	0.493558044	0.569002728	1	97
2.2015	Predicted sales in week 1	0.491279893	0.493558044	1	98
3.2015	Predicted sales in week 2	0.428838745	0.491279893	1	99

Note: Since we do not know the future test week sales, we should use the predicted sales for a week as the lag1 sales of the next week.

Model equation

$$\text{Log(Sales Units)} = 2.6186 + 0.5099\text{Log(Sales Unit Lag1)} + 0.3572\text{Log(Discount \%)}$$
$$-0.3016\text{Log(Discount \% Lag 1)} + 0.4557\text{Promo Flag} - 0.0141\text{Age}.$$

Substituting the predictors in the above equation, we get the following forecasted sales units.

Week	**Actual Sales Units**	**Forecasted Sales Units**	**Absolute Percentage Error**
52.2014	45	38.60479	14.2%
1.2015	44	32.54907	26.0%
2.2015	36	30.65673	14.8%
3.2015	30	27.96609	6.8%
			15.5%

MAPE=15.5%

Note: Run the following R code

1) 06_Forecasting_&_validation_v2.R

PART C – OPTIMIZATION

15. Formulate an optimization model and solve it to determine the optimal discount % to be given for "Cluster =2, Brand= BLINK, Brick= HAREMS" combination for each of the 4 weeks of EOSS.
    a. Objective function – Maximize the total revenue generated during EOSS.
        i. Total revenue is defined as the revenue generated during 4 weeks of EOSS and revenue from the left over inventory after EOSS. Assume the residual inventory left over even after EOSS is liquidated by giving a flat discount of 60%.

    b. Constraints

i. Sales units in a week should be less than equal to starting inventory of the same week.
ii. Relationship between demand (sales unit) for a week and predictors should follow the relationship identified in the forecasting regression model.
iii. Discount for a week should be greater than equal to previous week discount and discount should vary between 10% and 60% only.
iv. Sales and inventory values for a week should be ≥ 0.
v. Inventory should be updated on basis of the sales for a week.

c. Inputs required
   i. Inventory at the beginning of EOSS = 2,476
   ii. Age of the brick at the beginning of EOSS = 96 weeks
   iii. Previous week discount = 57.9%
   iv. Previous week sale = 48
   v. MRP of the brick = INR 606

Recommended Solution

The below optimization model is built in Lingo thus making it very readable.

Variables used in the optimization model

1) d1, d2, d3, d4 – Weekly optimal discounts that we want the optimization model to find.
2) s1, s2, s3, s4 – Weekly sales units
3) inv1, inv2, inv3, inv4 – Weekly inventory
4) promo_w1, promo_w2, promo_w3, promo_w4 – Flag variable indication of the week was under EOSS promotion.
5) mrp – Average MRP of the combination
6) age – Age of the combo in terms of weeks
7) preweek_discount – Previous week discount %
8) preweek_sale – Previous week sales units

<u>Code</u>

```
! Objective function = Revenue generated during 4 weeks of EOSS + Value of left over inventory;
max= (mrp*(1-d1)*s1 + mrp*(1-d2)*s2 + mrp*(1-d3)*s3+ mrp*(1-d4)*s4) +(mrp*(1-0.60)*inv_after_eoss);

!Inputs;
inv1= 2476;
age1=96;
preweek_discount=0.579;
preweek_sale=48;
mrp=606;
```

```
! Incrementing age for each week;
age2=age1+1;
age3=age2+1;
age4=age3+1;

! Assigning the flags to 1 since each of the 4 weeks are EOSS promotion weeks;
promo_w1=1;
promo_w2=1;
promo_w3=1;
promo_w4=1;

! Inputting coefficient from the forecasting model;
Intercept=2.618675;
sales_lag1_coef=0.509915*1;
disc_coef=0.357265*1;
disc_lag1=-0.301669*-1;
coeff_age=-0.014149*-1;
promo_coef= 0.455792*1;

! Constraints connecting sales to its predictors;
@log(s1) <= intercept + sales_lag1_coef*@log(preweek_sale) + disc_coef*@log(d1) - disc_lag1*@log(preweek_discount) -coeff_age*age1 +promo_coef*promo_w1 ;
@log(s2) <= intercept + sales_lag1_coef*@log(s1) + disc_coef*@log(d2) -disc_lag1*@log(d1) -coeff_age*age2 +promo_coef*promo_w2 ;
@log(s3) <= intercept + sales_lag1_coef*@log(s2) + disc_coef*@log(d3) -disc_lag1*@log(d2) -coeff_age*age3 +promo_coef*promo_w3 ;
@log(s4) <= intercept + sales_lag1_coef*@log(s3) + disc_coef*@log(d4) -disc_lag1*@log(d3) -coeff_age*age4 +promo_coef*promo_w4 ;

! Sales should be less than inventory;
s1 <= inv1;
s2 <= inv2;
s3 <= inv3;
s4 <= inv4;

! Updating inventory based on sales in the week;
inv2=inv1-s1;
inv3=inv2-s2;
inv4=inv3-s3;
inv_after_eoss=inv4-s4;
! Enforcing increasing discounts as a business rule;
d2 >= d1;
d3 >= d2;
d4 >= d3;
```

```
! Restricting upper limit & lower limit of discount as per business rule;
d1<= 0.6;
d1>=0.10;
! @GIN(d1/0.05);

d2<=0.6;
d2>=0.10;
!@GIN(d2/0.05);

d3<=0.6;
d3>=0.10;

!@GIN(d3/0.05);

d4<=0.6;
d4>=0.10;
!@GIN(d4/0.05);

! Positive constraint;
s1>=0;
s2>=0;
s3>=0;
s4>=0;
inv1>=0;
inv2>=0;
inv3>=0;
inv4>=0;
```

Final optimal discount % from the optimization model

Discount %	Optimal Values
d1	10.00%
d2	10%
d3	11.50%
d4	15.79%

16. What are the weekly forecasted sales units if the optimal discounts identified are implemented for the EOSS?

Recommended Solution

The final values of s1, s2, s3, and s4 are the forecasted weekly sales units if the optimal discount is implemented. Solving the above optimization model, we get the final values of s1, s2, s3, and s4 as shown below.

Weekly Sales Units	Forecasted Value
s1	20.74
s2	22.65
s3	24.55
s4	27.08

17. We know that actual revenue realized by the retailer for "Cluster=2, Brand=BLINK, Brick=HAREMS" combination during the 4 weeks of EOSS is INR 41,320. Then, what is the incremental lift in revenue the retailer would have achieved in these 4 weeks if he/she implemented our analytics solution instead?

Recommended Solution

The final values of the decision variable from the optimization model gives the optimal discount % to be implemented and the forecasted sales that will be generated if this optimal discount percent is implemented.

This table shows the values of decision variables from the model.

Decision Variables	Optimal Values
d1	10.0%
s1	20.74
d2	10%
s2	22.65
d3	11.5%
s3	24.55
d4	15.79%
s4	27.08

We also know MRP = INR 606.

Using this information, we can calculate the forecasted revenue generated during the 4 weeks of EOSS as shown in the equation.

(mrp*(1 – d1)*s1 + mrp*(1 – d2)*s2 + mrp*(1 – d3)*s3 + mrp*(1 – d4)*s4) = 50,659
Incremental lift in revenue = INR 50,659 – INR 41,320 = INR 9339.