GROUP ASSESSMENT COVER SHEET
Faculty of Design and Creative Technologies

AUT
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

Paper Name	**Industrial and Business Analytics**	**Paper Code:**	**STAT702**	**Assignment Due Date**	26/05/2021
Lecturer:	**Sarah and Patricio**	**Tutorial Day**	**Friday**	**Date Submitted**	26/05/2021
Tutor:	**Sarah and Patricio**	**Tutorial Time**	**2pm – 4pm**	**No. Words/Pages**	

In order to ensure fair and honest assessment results for all students, it is a requirement that the work that you hand in for assessment is your own work. If you are uncertain about any of these matters then please discuss them with your lecturer.

Plagiarism and Dishonesty are methods of cheating for the purposes of General Academic Regulations (GAR) http://www.aut.ac.nz/calendar

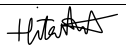
Assignments will not be accepted if this section is not completed and signed.

Please read the following and **tick** ✓ to indicate your understanding:

1. I understand it is my responsibility to keep a copy of my assignment.	☒ **Yes**	☐ **No**
2. I have signed and read the **Student's Statement below**.	☒ **Yes**	☐ **No**
3. I understand that a software programme (Turnitin) that detects plagiarism and copying may be used on my assignment.	☒ **Yes**	☐ **No**

Student's Statement:
This assessment is entirely my own work and has not been submitted in any other course of study. I have submitted a copy of this assessment to Turnitin, if required.
In this assessment I have acknowledged, to the best of my ability:

- The source of direct quotes from the work of others.
- The ideas of others (includes work from private or professional services, past assessments, other students, books, journals, cut/paste from internet sites and/or other materials).
- The source of diagrams or visual images.

Student ID	Name	Signature	Date
14869551	**Genevieve Connell**		**26/05/2021**
17989070	**Hitarth Asrani**		**26/05/2021**

The information on this form is collected for the primary purpose of submitting your assignment for assessment. Other purposes of collection include receiving your acknowledgement of plagiarism polices and attending to administrative matters. If you choose not to complete all questions on this form, it may not be possible for the Faculty of Design and Creative Technologies to accept your assignment.

STAT702 Industrial and Business Analytics Project

Genevieve Connell and Hitarth Asrani

26 May 2021

1. Analysis of Sales Data

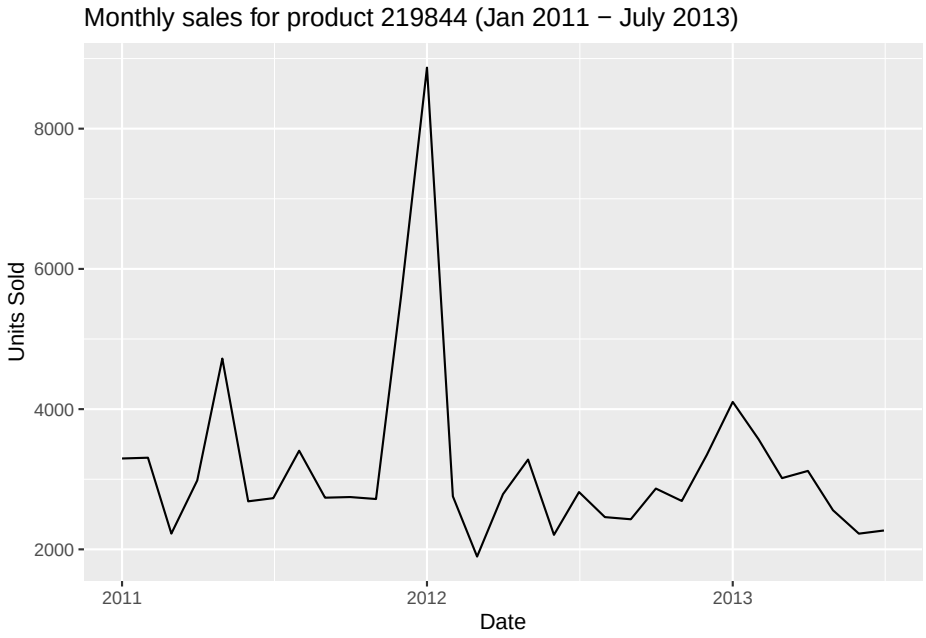
Product name: BIC Round Stic Xtra Life Ballpoint Pen, Medium Point (1.0mm), Red, 12-Count

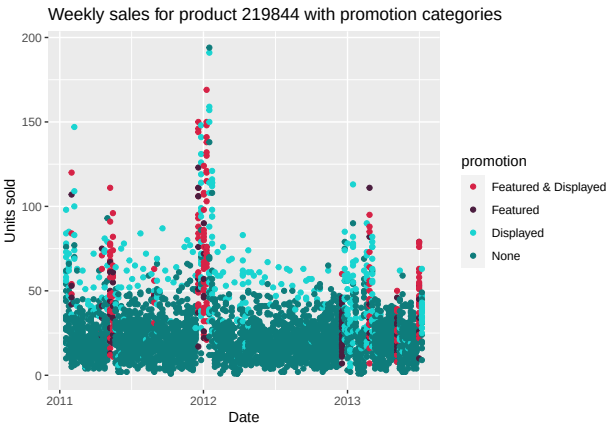
Sales sku_id: 219884

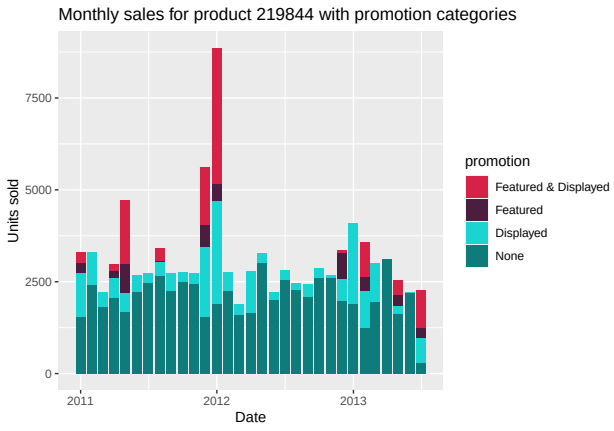
Reviews asin: B00006IE7J

1.a

From Jan 2011 - July 2013, 98434 units of sku 219844 were sold with a monthly mean of 3175.3. In the plots below monthly sales do not appear to follow a trend or seasonal pattern. Three months stand out with high sales, these are May 2011, December 2011 and January 2012. The most significant high sales occurred in January 2012 when 8871 units were sold. As shown in the plots below these high monthly sales correspond with a high proportion of stores featuring and/or displaying the product, resulting in unusually high sales.

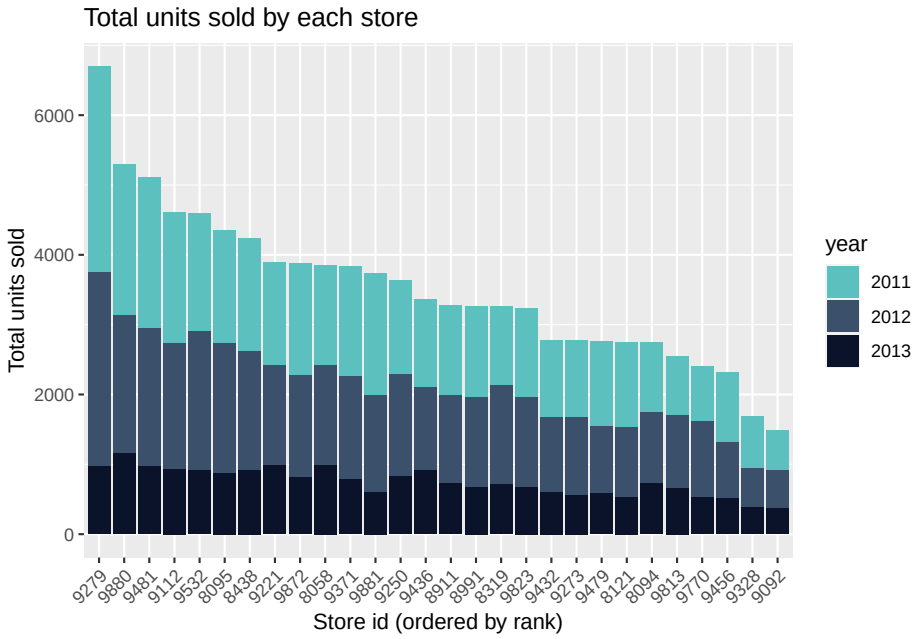






1.b

All stores have been ranked based on their total sales from Jan 2011 to July 2013. The store with the highest total units sold is ranked '1', this is store 9279 with 6698 units. Below is a bar chart showing the total units sold with stores ordered by rank. Sales for 2013 are much lower than 2011 and 2012 as only half the years data is included.



For a more detailed look at store performance monthly sales have been plotted for each store and categorised as 'High', 'Average' or 'Low' performing. These categories are based on the interquartile range for monthly sales. This range is 71 - 138 and captures 50% of all monthly sales. Monthly sales within the range are classified as 'Average', sales greater than 138 are classified as 'High' performing and those below 71 are classified as 'Low' performing. Summary statistics for store monthly sales are below.

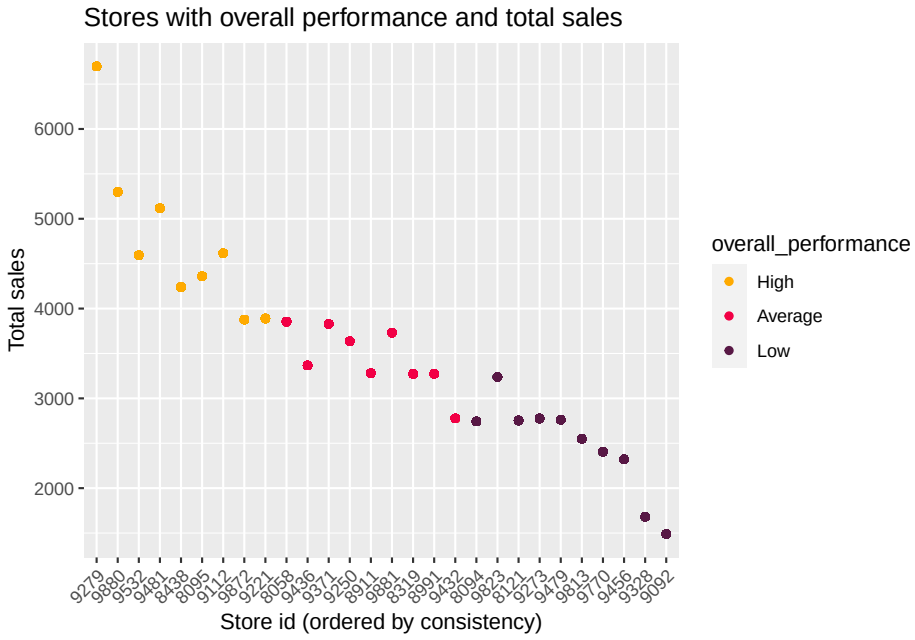
Min	1st Qu.	Median	Mean	3rd Qu.	Max
19	71	95.5	113.4	138	571

An interesting feature of these plots is that some stores have more consistent performance than others. For example store 9112 had 5 low performing months but was ranked at 4, much higher than store 9872 which had only 1 low performing month but was ranked 9. Stores with high occurrences of low sales are not reliable and are a source of risk to an organisation.

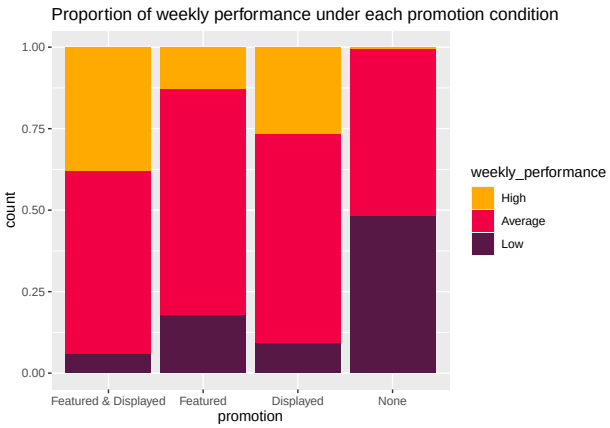
As an alternative to ranking stores based on their overall sales, stores have been given a ranking

based on their consistency. For every high performing month stores are given a score of 3, for every average month a score of 2 and for every low performing month a score of 1. The plot below shows the stores ordered by this new inconsistency ranking and their overall sales. Under the new consistency ranking scheme stores with unreliable performance such as 9112 drop in rank whilst stores with reliable performance such as 9532 move up in rank.

Overall consistency ranking seems most appropriate for sales performance as it rewards stores with reliable monthly sales. Consistency ranking has been used to split stores into three groups of overall performance. The top third of stores (consistency rank 1-9) are evaluated as 'High' performing. The middle third of stores (rank 10-18) are evaluated as having 'Average' performance and stores in the bottom third (rank 19-28) are evaluated as 'Low' performing.



Below is a plot showing that when a product is featured and/or displayed weekly sales are more likely to be high. Weekly performance has been categorised into 'High', 'Average' and 'Low' based on the interquartile range of weekly sales. When the product is featured and/or displayed a higher proportion of weekly sales have been 'High' or 'Average' compared to no promotion. Low performing stores could boost their sales by increasing the number of weeks they promote the product and unreliable stores could ensure high sales by holding regular promotions.



Question 2

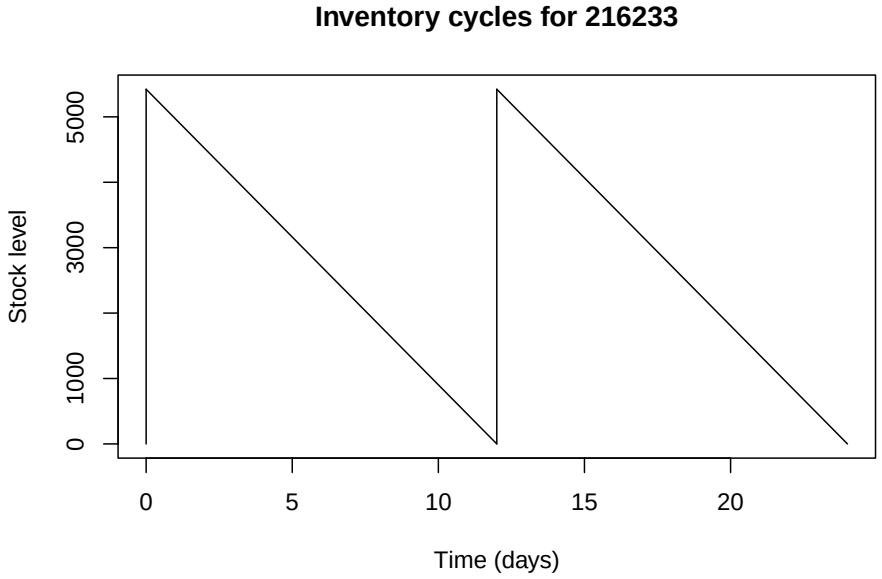
2.a.i

The Economic Order Quantity (EOQ) model is used to find the best order quantity so that total costs are minimised. Key assumptions to this model are that demand is constant and known, there is no lead time, orders arrive instantaneously and back orders are not allowed. Another assumption is that stock levels are under continuous review.

Demand for product 216233 has been estimated based on the annual demand from 2012, this is 169591. The optimum order quantity is calculated based on this demand and the annual order and holding costs. The following EOQ formula is used to determine optimal order quantity where k is order cost and h is holding cost.

$$Q^* = \sqrt{\frac{2kA}{h}}$$

The optimal order quantity is calculated to be 5422 with an inventory cycle of 12 days. This means that every 12 days 5422 units are ordered, resulting in 31 annual orders. This model results in the smallest possible annual inventory cost of 8132.6803762 with an annual order cost of 4066.1803762 and an annual holding cost of 4066.5. This EOQ model is plotted for two cycles below.



2.a.ii

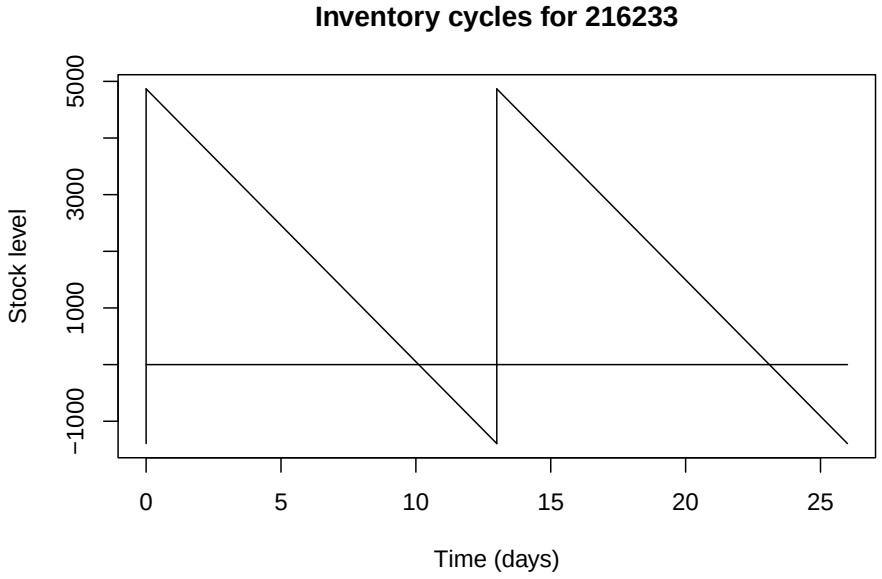
The Optimum Backorder Model is used to find the best order quantity so that total costs are minimised when backorders are allowed. Similar to the EOQ model assumptions are made that demand is constant and known, there is no lead time, orders arrive instantaneously and stock levels are under continuous review.

Annual demand is estimated as 169591 using 2012 data. Backorders cost is approximately 5% of the total price per unit. In the 2012 sales data total price for product 216233 varies from 78.4 to 134.7. For evaluating the backorder cost the mean total price 78.4 has been used, resulting in a backorder cost (p) of 6.22 per unit. The optimum quantity (Q^*) and optimum maximum inventory level (S^*) are calculated using the following forumlas.

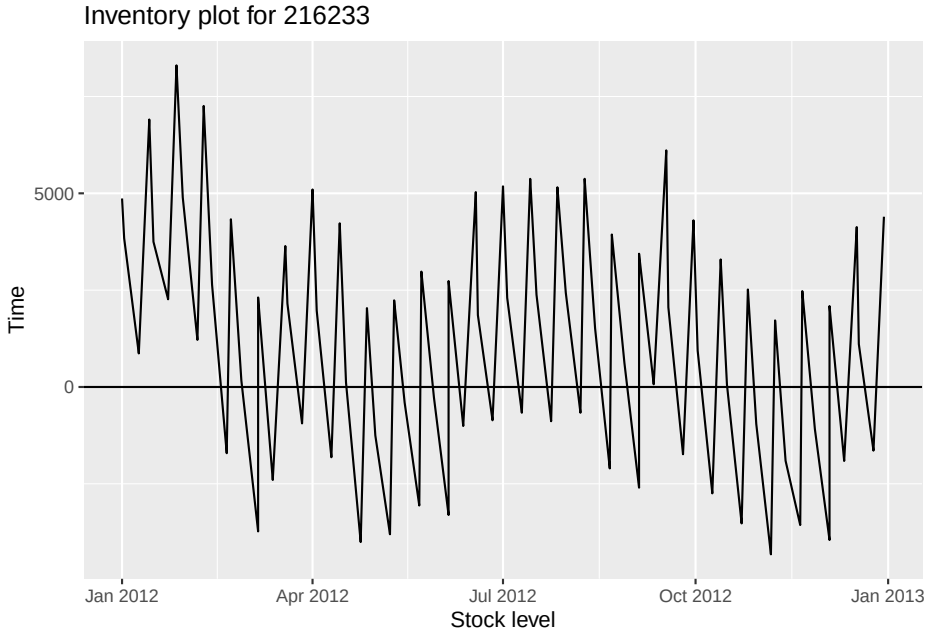
$$Q^* = \sqrt{\frac{2kA}{h}}\sqrt{\frac{p+h}{p}}$$

$$S^* = \sqrt{\frac{2kA}{h}}\sqrt{\frac{p}{p+h}}$$

Optimal order quantity is calculated to be 6040 with an inventory cycle of 13 days. This means that every 13 days 6040 units are ordered, resulting in 28 annual orders. The optimum inventory level is 4867 and the proportion of time taking back orders is 23%. This model results in the smallest possible annual inventory cost of 7300.44 with an annual order cost of 3650.14 and an annual holding cost of 2941.35 and annual backorder cost of 709. This EOQ model is plotted for two cycles below.



2.a.iii



The plot above shows the weekly demand from the 2012 sales data plotted with the optimum order frequency, 13 days and optimum order quantity 6040 from the Optimum Backorder Model. Stock starts at the optimum inventory level 4867, decreases with every weekly sale quantity from 2012 data and increases with inventory added at the optimum frequency and quantity.

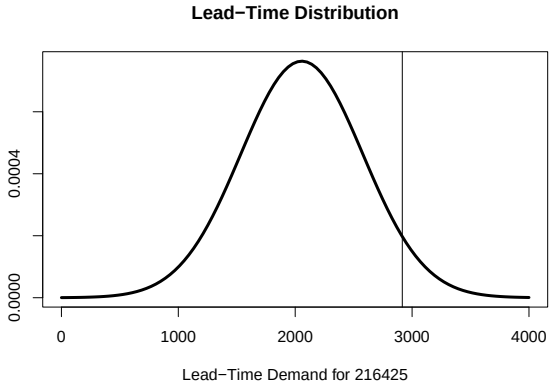
Actual demand is not constant and as a result the quantity and frequency found with the back order model is not always appropriate. The model performs best when demand is constant, for example in July.

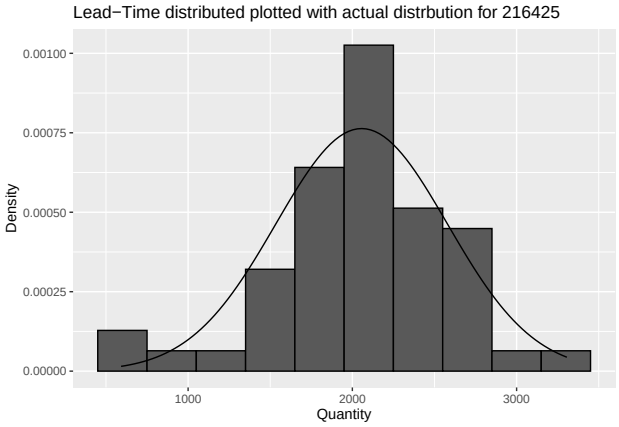
2.b.i.

For this multi-inventory model demand during a one week lead time has been estimated using the mean and standard deviation of observed data in 2012. Demand has been estimated as a normal distribution with a mean of 2057 and standard deviation of 523. As the plot below shows, whilst the actual demand for 2012 does not perfectly follow this distribution it is a adequate approximation.

The expected annual demand is estimated to be 106939. Given this annual demand and the costs of holding and reordering stock, the recommended multi-inventory model is to order 821 units whenever the order quantity reaches the reorder point of 2916 units. Approximately 130 orders will be placed per year and safety stock is 821. This approach ensures roughly 95% of the time stock will be sufficient for weekly demand. The expected annual costs are 10926.75 per year. If demand was certain the annual costs would only be 5338.47 so the additional cost of holding safety stock is 5588.28.

```
## integer(0)
```



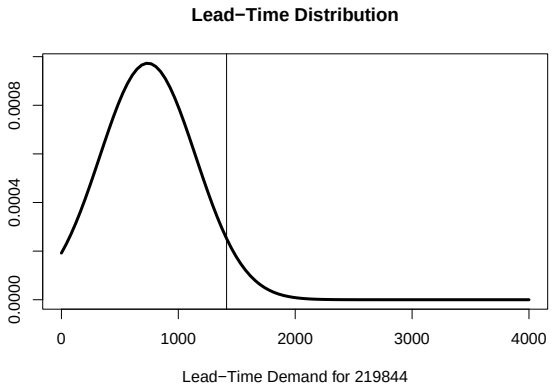


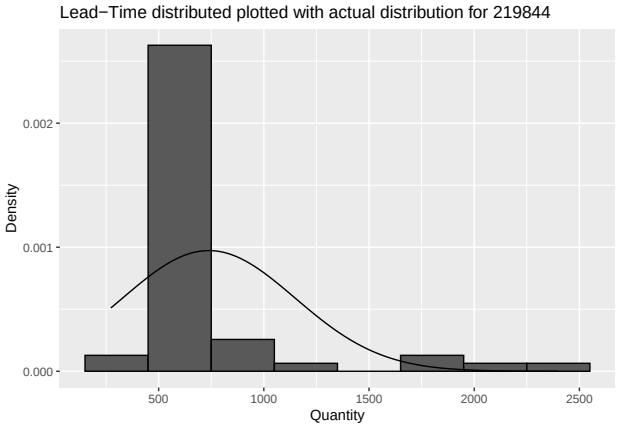
2.b.ii

For this multi-inventory model, demand during a one week lead time has been estimated using the mean and standard deviation of observed data in 2012. Demand has been estimated as a normal distribution with a mean of 739 and standard deviation of 410. As the plot below shows this distribution is a very poor fit for the observed data. It is recommended that a more accurate distribution is used for estimating the Lead-Time distribution and reorder point. It is likely that the reorder point calculated with this normal distribution is unnecessarily high.

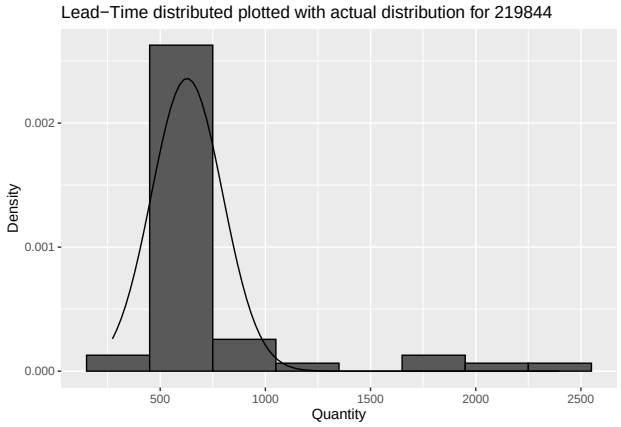
The expected annual demand is estimated to be 38410. Given this annual demand and the costs of holding and reordering stock, the recommended multi-inventory model is to order 492 units whenever the order quantity reaches the reorder point of 1413 units. Approximately 78 orders will be placed per year and safety stock is 492. This approach ensures roughly 95% of the time stock will be sufficient for weekly demand. The expected annual costs are 7582.37 per year. If demand was certain the annual costs would only be 3199.42 so the additional cost of holding safety stock is 4382.95.

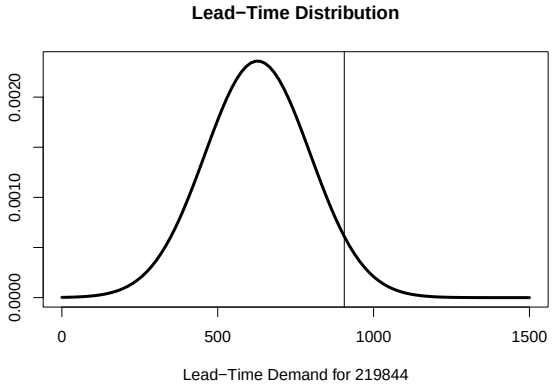
```
## integer(0)
```






```
## integer(0)
```





Question 3

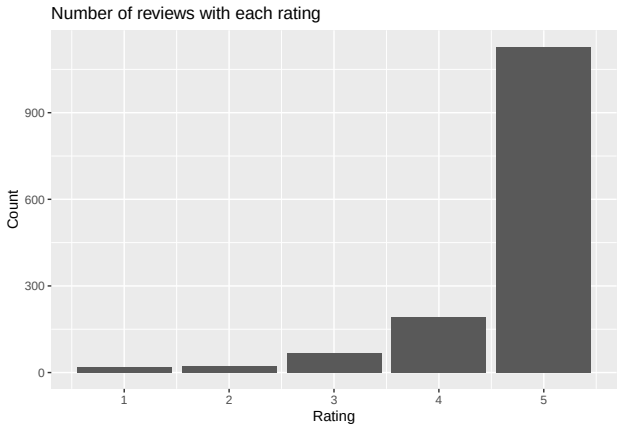
3.a

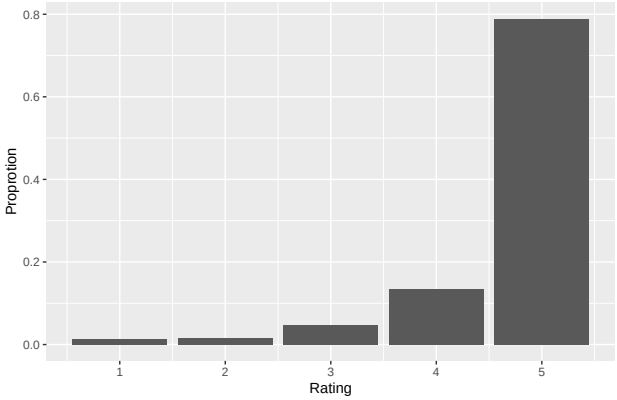
Reviews were rated from 1-5. The mean rating for product B00006IE7J is 4.7. Most reviews were rated highly with 79 given 5 and 13.4 given 4 stars.

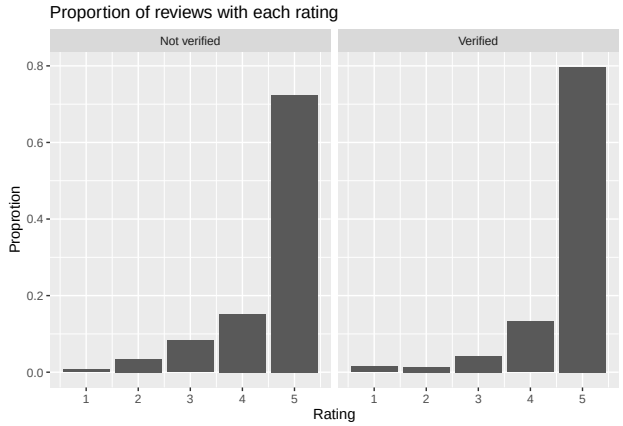
Min	1st Qu.	Median	Mean	3rd Qu.	Max
1	5	5	4.7	5	5

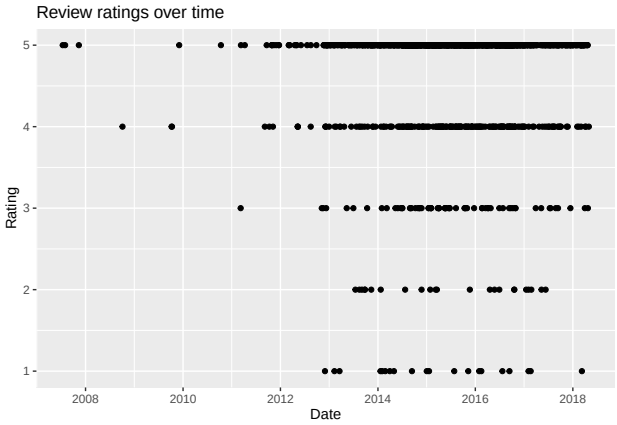
Some reviews were verified and others were not. The proportion of ratings given by these different groups are compared below. The unverified group gave a higher proportion of ratings 2-4 than the verified group.

Review ratings have been plotted below against time. Most reviews were given between 2014 and 2018. Rating does not appear to be strongly correlated with time.





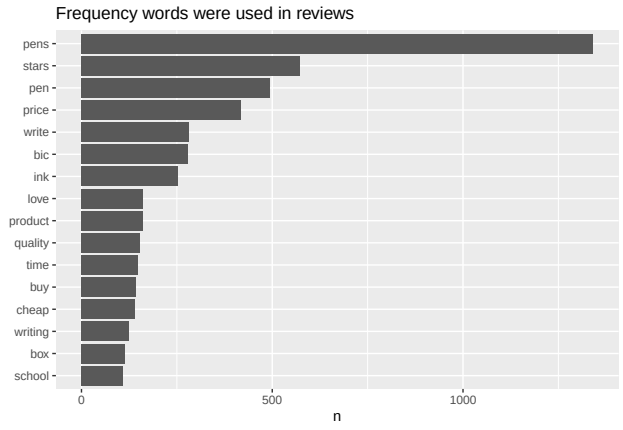




PLotting the number of reviews, we can confirm that the majority of reviews are rated an overall of 5 out of 5.

3b

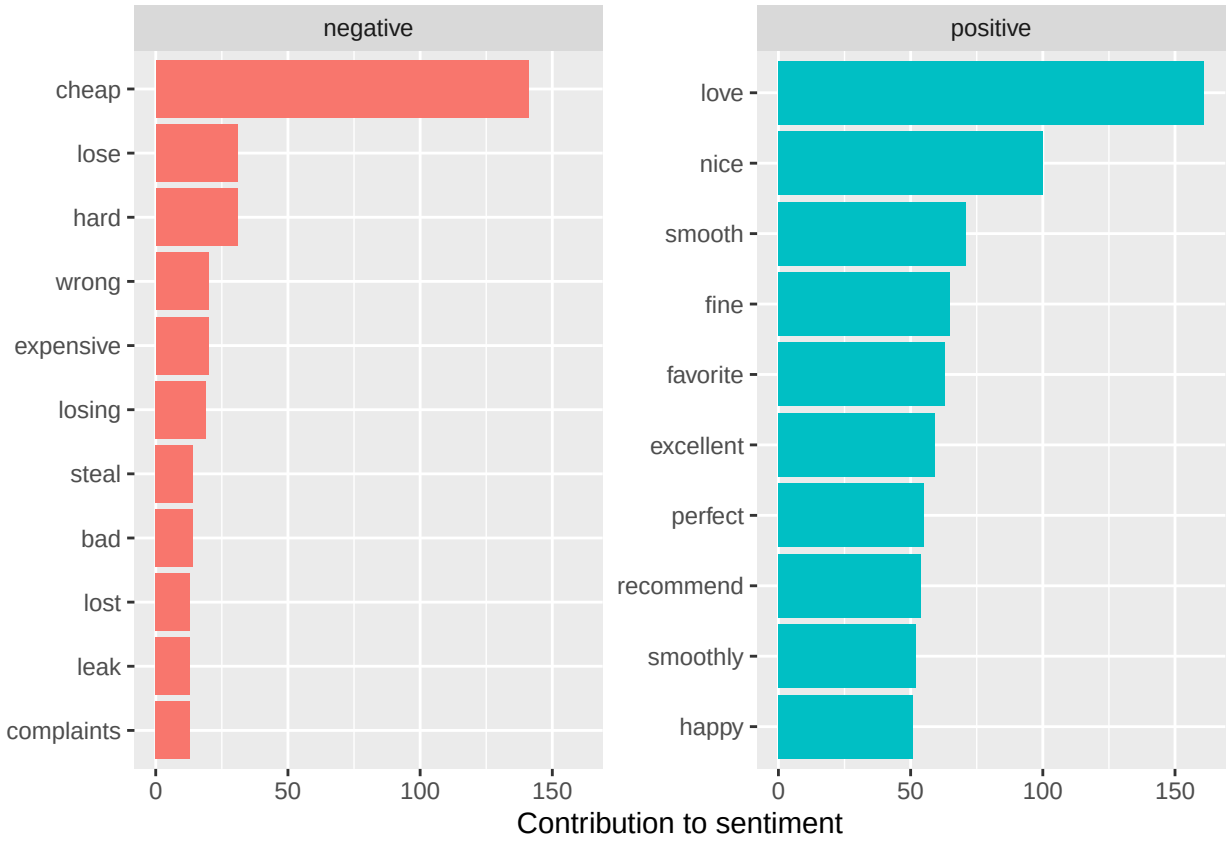
Review text has been tokenised and stop words have been removed. Stop words are words such as 'the' and 'it' that hold little information for sentiment analysis. The top 10 words in reviews are plotted below along with their frequency.



In the plot above we can see a few words that can be considered as stop words. Words like `pen` and `stars` are not very valuable to the sentiment. These words are removed from the data set as well.

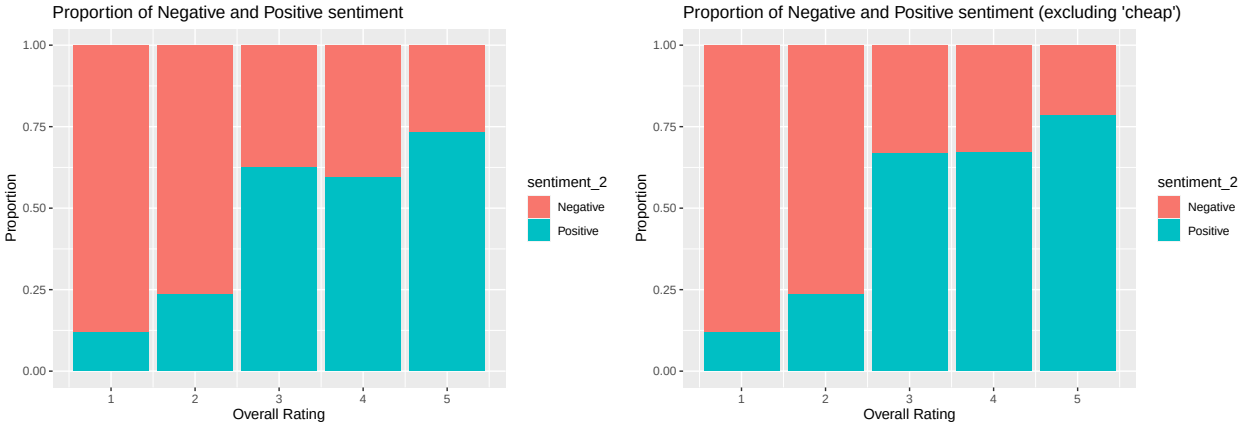
Notably the top words are 'pens', 'stars' and 'pen'. The product that is being analysed is bic pens, this explains why these words occur frequently. Users provide a 5 star rating, in their review users then refer to their rating and this explains why the word 'stars' occurs frequently. These words offer little information about how reviewers feel about the product so these are removed from reviews for the next part of analysis.

The remaining words are given a sentiment rating, high ratings indicate positive sentiment and low ratings indicate negative sentiment. The top 10 frequently used negative and positive words are presented below along with a word cloud of the top 100 words.


negative
cheap
lose
hard
wrong
expensive
losing
steal
bad
lost
leak
complaints
0
50
100
150
positive
love
nice
smooth
fine
favorite
excellent
perfect
recommend
smoothly
happy
0
50
100
150
Contribution to sentiment

Notably the word ‘cheap’ is the most commonly used negative word and is used significantly more frequently than any other negative word. ‘Cheap’ can be used to say something is low quality, however in the context of a pen ‘cheap’ is likely to be a positive descriptor as people often don’t want to spend too much money on pens. Overall 141 reviews included the word ‘cheap’, 0.7446809% were 5 star reviews and 0.1560284% were 4 star reviews. This indicates that ‘cheap’ is more likely to indicate a positive rather than negative sentiment.

For each review the word sentiments are added together giving the review an overall sentiment score. Below is a plot showing the proportion of reviews in each rating category were given an overall negative or positive sentiment rating. When the word ‘cheap’ is excluded there is a small change with a decreased proportion of 3-5 ratings classified as negative. This indicates better performance classification.


Proportion of Negative and Positive sentiment
Proportion
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
Overall Rating
sentiment_2
Negative
Positive
Proportion of Negative and Positive sentiment (excluding 'cheap')
Proportion
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
Overall Rating
sentiment_2
Negative
Positive

Appendix A