Lancaster University School of Computing and Communications

Group Name: Team Kappa

Project Name: Promotion Price Sensitivity

Company: N Brown

Group Spokesperson:

Ben Grime – 36601557 (Presentation, Voiceover and report collation.)

Group Members:

Akash Marar – 3665057 (Department, seasonality insights and exploratory data analysis)

Donald (Tze Thoe) – 36575827 (Price elasticity findings and exploratory data analysis)

Anam Khan – 36577898 (Regression/OLS modelling)

Pratamesh Takale – 36561140 (ARIMA, SARIMA modelling)

Zeerak Khan -36641751 (Presentation collation/production and introduction)

Abstract

Using Ordinary Least Squares (OLS) regression, the mean price elasticity of products in N Brown's department 0 is 0.20 with a slight right skew. This provided more nuanced and reliable results than a chronological mean-based average. SARIMA time series analysis was trained on 2018 data to forecast for Dec 2018 and Jan 2019. For department 0, the model predicted an average of 10,600 daily orders over the month of Dec 2018, and 6,400 over Jan 2019 (MAE = 5,666, MAE = 3,886, for the sum of orders over the months, respectively).

Introduction

In the ever-evolving landscape of the retail industry, understanding consumer sensitivity to price changes is essential for maintaining a competitive advantage. This report focuses on N Brown, a prominent company in the online and catalogue clothing retail sector.

Price elasticity of demand (PED) is crucial in making informed pricing decisions. It is the ratio between the percentage change in demand and the percentage change in price. This measures how the quantity demanded of a good responds to a change in its price, with most consumer goods and services having a PED between 0.5 and 1.5 [1]. That is, for a 1% increase in price, demand decreases by 0.5% to 1.5% (and vice versa). For N Brown, assessing price elasticity is far more important than just adjusting prices. It's about understanding market trends, consumer behaviour, and the impact of various other economic factors.

This report delves into a detailed analysis and interplay between price elasticity, discounts, and sales. The following primary research questions are answered: (i) determine the price elasticity of products; (ii) predict the number of orders N Brown will receive in a subsequent month if no discounts are applied. Both questions unearth different aspects of N Brown's pricing strategy and its influence on consumer behaviour and sales performance.

The first question, involving elasticity, is crucial for understanding the degree to which pricing adjustments can influence customer demand. The extent to which N Brown's discounting strategies impact the volume of orders and how this can be optimised for maximum profitability and market penetration is quantified. The second question, involving forecasting, is pivotal in assessing baseline customer demand and forms a critical part of strategic decision-making regarding pricing and promotions.

Additionally, seasonal trends are examined. In an industry where consumer preferences and purchasing behaviours shift in tandem with seasons, understanding this aspect is essential for tailoring inventory to leverage seasonal opportunities. Lastly, the department yielding the most sales for N Brown is identified and analysed in the context of the previous research questions.

To answer these questions, time series forecasting techniques were utilised, notably the Autoregressive Integrated Moving Average (ARIMA) and seasonal ARIMA (SARIMA) models. These models are adept at capturing the datasets sequential nature and the underlying patterns, trends, and seasonal variations. The decision to use only data from 2018 is driven by its consistency in comparison to 2019 data. Its regular chronology is essential for the integrity of predictive models.

To summarise, this report aims to dissect N Browns current market position, as well as equipping the company with the robust forecasting tools to navigate future market dynamics efficiently and effectively.

Methodology

Four sections are discussed in the following methodology: preprocessing, exploration, elasticity, and time series modelling.

Preprocessing

In the preprocessing phase of our project, minimal adjustments were necessary due to the relatively clean state of the dataset we were provided. We removed any rows with missing values, converted the date of order to datetime format and transformed product number and department number into categorical data types. These steps were important for optimising the dataset for the subsequent analysis.

Exploration – 2018 Data

The data spans 2018 and 2019. However, only 2018 data was used for the following reasons which can be observed in Figure 1:

- Consistency – the 2018 data contains daily entries throughout the year, providing a consistent time series for analysis.
- Sporadic 2019 data – the 2019 data is sporadic, with several entries only in early January and then just monthly entries at the start of each month. This could introduce bias.
- Data integrity for modelling – to ensure the integrity of the predictive model, dataset with a regular pattern of data points is required. This is provided by 2018 data but not 2019 data.

Exploration – Department 0

The order distribution chart shown in Figure 2 indicates significant variance in orders received across departments. Specifically, department 0 exhibits the highest volume of orders. In contrast, the departments 15, 10, and 14 record minimal activity. Whilst less generalisable, minimising the scope of the predictions made by the model produces more reliable results. There are less covariates (in this case, by removing departments) to produce noise and variation in the data. As such, daily order volume was more accurately forecast for department 0 alone, instead of less accurately forecast across all departments.

Exploration – Seasonality

The graph show in Figure 3 presents a compelling narrative on the influence of discounts on consumer behaviour. Discounts are a catalyst for increased orders, particularly as one approached the end of October as can be seen in Figure 3. Thus, the model must account for the seasonal peaks observed to predict the trend in future orders accurately. This observation was instrumental in highlighting the necessity to embed seasonality into the predictive models, in turn enhancing their precision.

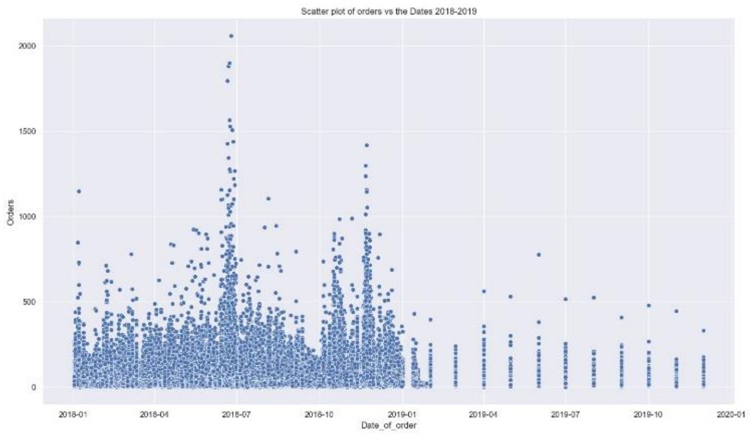

Figure 1 – distribution of 2018 and 2019 data

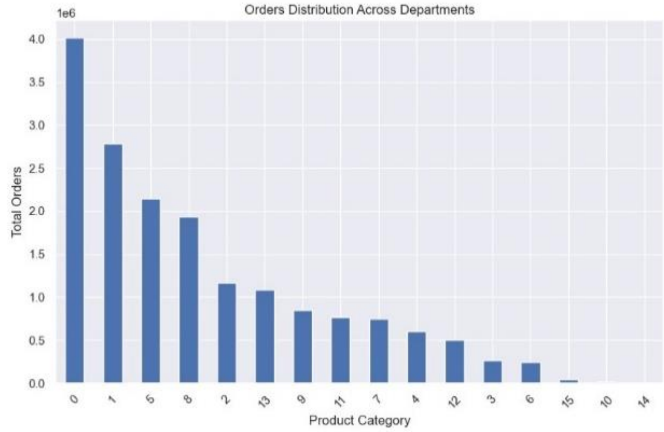

Figure 2 – distribution of orders across departments

Figure 3 – seasonal trends in orders

Elasticity – Averages, Regression & OLS

A simple way to determine the PED of products would be to calculate chronological changes to price and demand and take the mean. However, this method is particularly susceptible to extreme samples and lacks nuance. In the data provided by N Brown, there exist samples with sales in the order of thousands. This skews mean-based averages significantly. Further, per product, there are not sufficient samples for a median-based average of PED to be reliable.

Linear regression is a statistical method used to inform the user about linear relationships between variables, making it particularly suitable for the calculation of price elasticity. Ordinary Least Squares (OLS) was employed to model the relationship between discounts and sales. OLS aims to minimise the sum of the squared differences between the observed values and the values predicted by the model (residuals). This enabled a more nuanced measurement of alterations in discounts and its impact on sales over simply using mean-based averages over time. Ensuring the statistical significance of the OLS model's coefficients and the model's overall goodness of fit is imperative.

For the regression analysis of price elasticity, some key assumptions were made:

1. A linear relationship between the discount (as a proxy for price change) and the number of orders.
2. Only considered regression results with a p-value < 0.05.
3. The elasticity was calculated using the mean price (average discount) and mean quantity (average orders).
4. The discount was used as a proxy for price changes. This assumes that the base price of 100 remains constant and that discounts are the primary factor affecting price variations.
5. Only data from department 0 was analysed.

Time Series Modelling – ARIMA

A time series is a collection of data points arranged in a consistent time order [2]. The data is chronological, so using time series techniques for prediction was fitting. On a broader timescale, by analysing trends, seasonality, and sequential dependencies in the historical data, we can better predict future orders in the series. For daily demand forecasting, leveraging daily autocorrelations and smoothing effects makes our predictions more responsive to fluctuations.

To test the null hypothesis that a unit root exists in a time series sample, indicating non-stationarity, a popular statistical technique is the Augmented Dickey-Fuller (ADF) test [3]. The ADF test was utilised to verify if the time series was appropriate for ARIMA modelling. For the 2018 dataset, the ADF test yielded an ADF Statistic of -5.6657 (p = 9.16e-07). This confirms the stationarity of the series.

A popular statistical method for time series forecasting is the autoregressive integrated moving average (ARIMA) model. The autocorrelation between the present and lagged observations is modelled by the autoregressive (AR) component. The time series is made stable for modelling by the integrated (I) component using differencing of raw observations. The dependency between the present observation and residual errors from earlier time steps is accounted for by the moving average (MA) component. ARIMA(p,d,q), where p is the order of the AR term, d is the degree of differencing, and q is the order of the MA term, is a summary of the ARIMA model. During model training, these parameters are adjusted to best fit the properties of each time series. The capacity of ARIMA models to parsimoniously depict a broad range of serial correlation patterns in time series data with a small number of parameters is one of its main advantages. [4]

Time Series Modelling – SARIMA

For time series with seasonal cycles, seasonal ARIMA (SARIMA) models are used as an extension of ARIMA models. Seasonal AR and MA components are included in SARIMA in addition to the standard AR and MA components. This is expressed similarly with the addition of seasonal parameters; SARIMA(p,d,q)(P,D,Q)(m), where: P is the Seasonal autoregressive order, D is the seasonal differencing order, Q is the seasonal moving average order, and m is the periodicity of seasonality. In addition to normal differencing, seasonal differencing is necessary in SARIMA models so as to eliminate seasonal changes and keep the data stationary.

Accurate projections on time series with numerous seasonal cycles can be produced by SARIMA by modelling both short- and long-term seasonal dynamics [2].

Time Series Modelling – SARIMA Parameter Selection

To optimise the performance of the SARIMA model, a grid search was performed over the seasonal parameters. The base non-seasonal parameters were fixed at (1, 0, 1) from previous analyses. Considering the frequency of monthly data and the possible yearly seasonality, m was determined to be 12. The grid search aimed to minimise the Akaike Information Criterion (AIC), a widely-used statistic for model comparison that balances goodness of fit with model simplicity [5].

The grid search identified the SARIMA(1, 0, 1)x(0, 1, 1, 12) configuration as having the lowest AIC value among the tested models. This outcome confirmed the significance of seasonal integration and the contribution of first-order seasonal autoregressive and moving average components, affirming the data's underlying seasonal tendencies. Therefore, this tailored model holds the potential for more accurate and reliable forecasting of the time series in question.

Time Series Modelling – Training

Rather than using a train-test split like an 80-20 division, the model was trained on all of the 2018 data. This method was chosen considering the small data set (only one year's worth of data), which required that all available data points be used to adequately capture the innate seasonal patterns. Optimising the model's knowledge through training on the entire dataset is essential for precise forecasting in situations when data availability is limited [6]. A thorough grasp of seasonal trends, especially with Christmas, was made possible by this strategy. When the estimates for December 2018 and January 2019 were plotted against the actual data, it became evident how well the model captured the seasonal spike in December and how different the January predictions were from the actual data.

Results

Price elasticity – Chronological Mean

Using a mean-based PED per product across *all* departments, PED ranges from 1643.54 to -123.71, indicating a vast diversity in how different products respond to price changes. The mean PED across all 16 departments show a narrower range of PED values, from 7.38 to -0.35. Such extreme outliers seen using a mean-based average are expected to be accounted for using OLS. This highlights the nuance and subtlety OLS has over the chronological mean, as described in the methodology.

Price elasticity – OLS

Using OLS, the mean PED across all products in department 0 is about 0.20. This suggests that, on average, products exhibit a low sensitivity to price changes. The median elasticity is 0.16. This entails a slight right/positive skew for the distribution of PEDs. That is, some products are highly elastic (towards the right tail), causing a shift in the mean above the median where most products lie. The maximum price elasticity observed is around 3.85.

Products with a PED about this maximum is where one observes the cause for the slight right skew. The standard deviation of the price elasticity among products is approximately 0.28.

99.04% of products in department 0 are inelastic (PED < 1). This is a significant majority, reinforcing the inference that your customers are not highly price sensitive. Only 0.96% of products in department 0 are elastic (PED > 1), indicating that very few items experience significant demand changes with price adjustments. Given the overall inelastic nature of products, there's potential room for price increases without significantly harming demand. This could be beneficial for improving profit margins, especially on those products with lower PEDs.

Time Series Modelling – Metrics

MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) are favoured metrics for evaluating the performance of time series models. They measure the extent of prediction mistakes for continuous variables and meet the particular needs of forecasting with regard to predictive accuracy. The average size of prediction errors of for the sum of orders in each month is measured by MAE and the error magnitude is measured in the same unit as the prediction by RMSE.

Time Series Modelling – Dec 2018

Plots of the SARIMA model are shown in Fig 4 and Fig 5 (zoomed).

Order volume estimated by the model for Dec 2018 had an MAE of 5,666 and an RMSE of 7,379. Seasonal sales peaks, which are prevalent in December, can exacerbate inaccuracies when contrasted to less active months. This ought to be taken into consideration while evaluating these numbers. Incorporating additional datasets from adjacent years, such as 2017 or 2019, could significantly bolster predictions. Training with data spanning multiple years would provide the model with a richer context for identifying and learning from both long-term trends and seasonal patterns.

Time Series Modelling – Jan 2019

Orders for January 2019 were forecast under the scenario of no discounts being applied. However, it's important to note that the resulting comparison with actual data from January 2019 is somewhat constrained. The available actual data for this period is incomplete, encompassing only six specific dates. This limitation in the dataset might impact the relevance and accuracy of our comparison, as it doesn't provide a complete picture of the entire month's order trends.

An average divergence of 3,886 orders from the actual counts was indicated by the estimate for January 2019 (MAE = 3,886). This difference, albeit significant, is consistent with the anticipated difficulties associated with predicting in unusual sales circumstances. Moreover, the forecast appears to have significant flaws, as evidenced by the RMSE of 4,232, higher than the MAE. The intricacies of precisely forecasting orders in non-discount periods—a situation that may deviate from past patterns—are highlighted by these findings. In order to improve predicted accuracy, the analysis emphasises the necessity for model refinement and the incorporation of other departments to increase the size of the data.

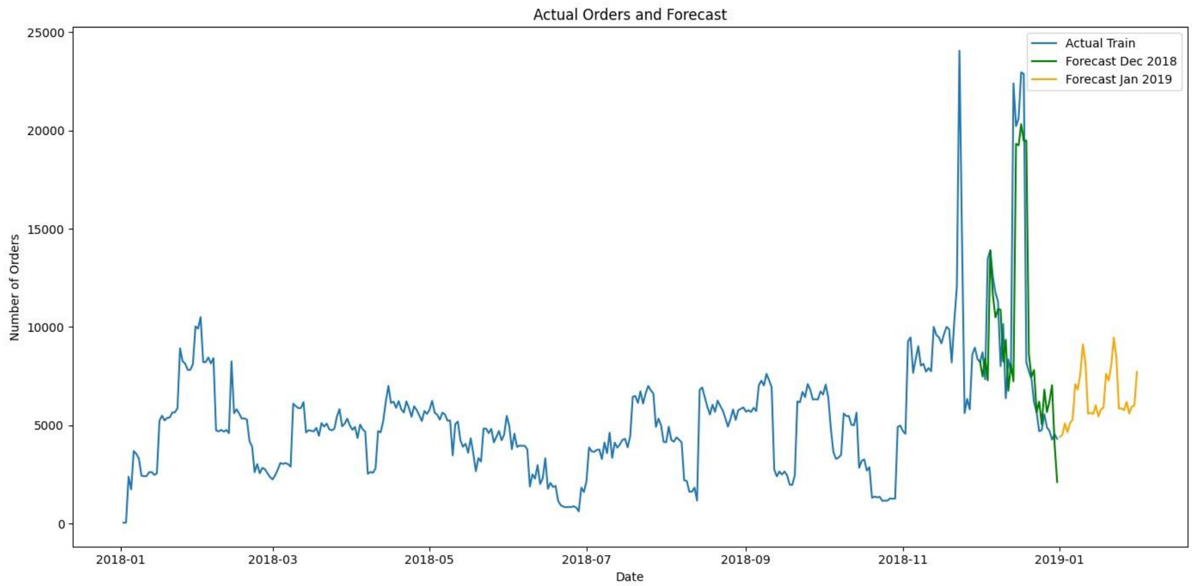

Figure 4 – forecast for Dec 2018 & Jan 2019

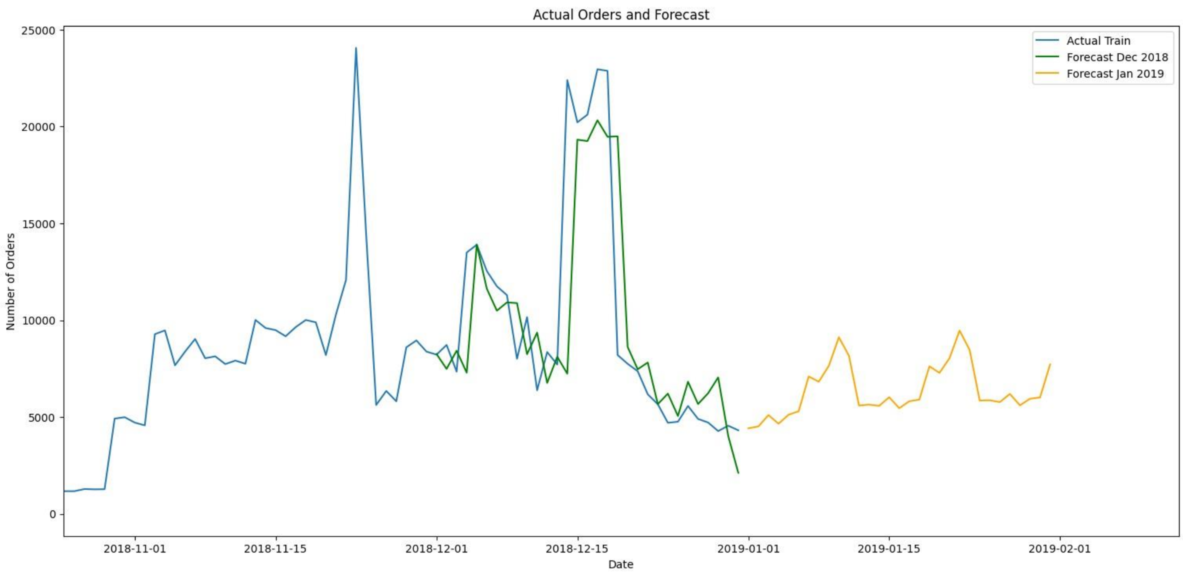

Figure 5 – forecast for Dec 2018 & Jan 2019 (zoomed)

Conclusions

Department 0 was identified as having the largest proportion of orders across all departments. Thus, the scope of findings was narrowed so as to produce more reliable results and reduce the interference of noise and variation between departments. 2019 data was, for the most part, removed due to its sporadic nature.

Using Ordinary Least Squares (OLS) regression, the mean PED of products in N Brown's department 0 is 0.20. The distribution had a slight right skew, with a median of 0.16. OLS seemed to provide more reliable results than a chronological mean-based average. This is due to its inherent ability to deal with outliers, of which there were significant ones in the data provided. This answers the first primary research question: determine the price elasticity of products.

SARIMA time series analysis was trained on 2018 data to forecast for Dec 2018 and Jan 2019. For department 0, the model predicted an average of ___ daily orders over the month of Dec 2018, and 6,379 over Jan 2019 (MAE = 5,666, MAE = 3,886, for the sum of orders over the month, respectively). This answers the second primary research question: predict the number of orders N Brown will receive in a subsequent month if no discounts are applied.

Bibliography

Anderson, McLellan, Overton, Wolfram – *Price Elasticity of Demand* (1997), Harvard University Press [1]

Hyndman, R.J & Athanasopoulos, G (2014) *Forecasting : principles and practice*. Heathmont, Vic: OTexts [2]

Dickey, D.A & Fuller, W.A – *Distribution of the Estimators for Autoregressive Time Series With a Unit Root* in *Journal of the American Statistical Association* (1979) (Vol. 74, No. 366, p. 427), Taylor & Francis [3]

Box, G.E.P – *Time series analysis: forecasting and control* (2016), Fifth edition, Hoboken, New Jersey: Wiley [4]

Akaike, H – *A new look at the statistical model identification* in *IEEE transactions on automatic control* (1974) (Vol. 19, No. 6, p. 716–723), IEEE [5]

Chatfield, C – *Time-series forecasting* (2001), Boca Raton: Chapman & Hall/CRC. [6]