AISO7002_S2_2425_19348992

by Hritik Gupta

Submission date: 08-May-2025 08:37AM (UTC+0100)
Submission ID: 258343945
File name: AISO7002_S2_2425_19348992_2914190_85281291.docx (1.11M)
Word count: 4483
Character count: 24719

OXFORD BROOKES UNIVERSITY

AISO7002

Practice of Data Analysis

Student ID - 19348992

Contents

1.Introduction

In this fast paced digital age, data literacy is more important to business to remain competitive. A good example of this is Bondi Sports Inc, a company that melds traditional retail with digital technological advances. As its online store was launched, Bondi Sports' number of website visits increased and management at Bondi Sports opted to adopt data driven strategies to increase customer engagement and sales (Davenport & Ronanki, 2018).

It looks at a dataset of 50 recent online transactions recorded given variables like browsing time, browser type, pages viewed, and spending. The report uses exploratory type of data analysis (EDA) to discover patterns and trends, paving the way for strategic decisions instead of being to answer predefined questions. To achieve this, the key points are performance of online store, user behaviour and sales trends (Provost & Fawcett, 2013).

Time series decomposition is a significant part of the analysis which uses it to obtain the trend, seasonality and residuals to facilitate an accurate sales forecasting. By incorporating Gen AI tools in the analysis of the problem, we can use more robust and critically analysed insights from both the data sources using AI driven derived insights, and traditional data sources (Davenport & Ronanki, 2018).

Finally, this report shows how Bondi Sports Inc. can leverage data analytics to create better customer engagement, synchronize sales approach and attain a sustainable growth through a digital sales channel.

2.Dataset Overview

This analysis centres on a dataset consisting of sample of 50 online transactions with Bondi Sports Inc., an online Sydney based retailer of surfing equipment and accessories. This dataset is collected from the last month sales to give insights about customer behaviour and sales performance of the company's online platform, which has observed immense growth since the launch of the company two years earlier.

In the dataset, there are several key variables over each record. The selection of data includes Day of the Week, Browser Type, Time Spent on Website, Number of Pages Browsed, and Amount Spent. The Day of the Week variable is a temporal feature that allows for the ability to identify temporal shopping trends, like for example to see if sales have peaks on weekends or work days, which is very important based on the resource's allocation and marketing campaign (Chaffey & Ellis-Chadwick, 2019). The difference in the value of the variable called Browser Type can tell the difference between the products and brands preferences of the customers, the business can ensure that the website is compatible with the technology regarding the products preferences, and it can tailor the digital experiences (Laudon & Traver, 2021).

Time Spent on Website (in minutes) and Number of Pages Browsed are proxies for the customer engagement and the browsing behaviour. By analysing these variables, one can find out if there is a correlation between engagement and higher transaction value and then use this information to support strategies for improving User experience and conversion rate (Provost & Fawcett, 2013). The primary performance metric is the Amount Spent variable recorded in Australian Dollars (AUD), making it possible to calculate average order values and pick out high value customers.

In summary, the dataset offers a valuable foundation for exploring customer behaviour and sales performance in a digital retail context. The findings derived from this analysis will guide Bondi Sports Inc. in making evidence-based decisions to optimise their online strategy and enhance overall business performance (Chaffey & Ellis-Chadwick, 2019; Laudon & Traver, 2021).

3.Research Questions

Question 1 - Does Amount spent has any impact on Chrome and Edge?

1

- **Null hypothesis (H_0):** There is no difference in the average amount spent between Chrome and Edge users.
- **Alternative hypothesis (H_1):** There is a difference in the average amount spent between the two groups.

Question 2 - Does page view has any impact on Chrome, Edge and Firefox?

4

- **Null hypothesis (H_0):** There is no difference in the page viewed among the three users.

4

- **Alternative hypothesis (H_1):** There is a difference in the page viewed among the three groups.

Question 3 - How does time spent on the website affect the amount spent by a customer?

- **Null hypothesis (H_0):** Time spent does not affect the amount spent by the customer.
- **Alternative hypothesis (H_1):** Time spent does affect the amount spent by the customer.

4.Methodology

4.1 Overview

The quantitative methodology used to analyse in this report is used to assess key aspects of the customer behaviour as well as the performance of sales of Bondi Sports. Three main business questions are also addressed in the analysis.

1. To see if there is a difference on the amount spent between the users of the Chrome browsers and the Edge browsers.

2. A variation in the number of pages viewed between Chrome, Edge, and Firefox users.

3. How customers spending on the website depends on time spent.

Finally, we perform useful time series analysis of synthetic weekly sales data to model business trends and seasonality.

4.2 Data Collection and Preparation

This analysis uses data mined from one of the company's web analytics software, stored in an Excel file named 'web_data.xlsx', from Bondi Sports. The dataset included user-level records: type of browser, amount spent, pages viewed, and time spent on the site (Grolemund & Wickham, 2017). Data preparation involved:

- Cleaning: Removing the rows in which there are missing or null values in certain columns.

- Filtering: Browsers such as Chrome, Edge, and Firefox were chosen for each analysis as required.

- Variable Formatting: All variables (e.g. amount spent, pages viewed) were guaranteed to be numeric and would therefore be ready for statistical analysis.

4.3 Statistical Analyses

4.3.1. Comparing Amount Spent: Chrome vs Edge

To determine if browser selection affects spending, we conducted both parametric and non-parametric tests:

- **Hypotheses:**
  - Null hypothesis (H_0): No difference in average amount spent between Chrome and Edge users.
  - Alternative Hypothesis (H_1): A difference exists.

- **T-Test**

  Welch's t-test was used to compare means, accounting for unequal variances (Welch, 1947).

  - Result: t-statistic = -0.30, p-value = 0.77.
  - Interpretation: No significant difference in spending between Chrome and Edge users.

Normality Check:
The Shapiro-Wilk test (Shapiro & Wilk, 1965) & Q-Q plot shows non-normal distribution ($p = 0.0487$).

Mann-Whitney U Test:
Used as a non-parametric alternative.

- Result: U-statistic = 80.0, p-value = 0.55.
- Interpretation: Again, no significant difference (Mann & Whitney, 1947).

4.3.2. Pages Viewed: Chrome, Edge, Firefox

To explore differences in user engagement by browser:

- **Hypotheses:**
  - Null (H_0): No difference in pages viewed among browsers.
  - Alternative (H_1): At least one group differs.
- **ANOVA:**
  One-way ANOVA compared the means.
  - Result: F-statistic = 2.33, p-value = 0.109.
  - Interpretation: No significant difference indicated.
- **Normality Check:**
  Shapiro-Wilk test shows non-normality for some groups (p-value = 0.0332)
- **Kruskal-Wallis Test:**
  non-parametric comparison among all three browsers (Chrome, Edge and Firefox) (Kruskal & Wallis, 1952).

- Result: H-statistic = 4.16, p-value = 0.125.
- Interpretation: No significant difference.

- **Mann-Whitney U Test (Chrome vs. Edge):**
  - Result: U-statistic = 48.5, p-value = 0.049.
  - Interpretation: Significant difference in pages viewed between Chrome and Edge users properly.

4.3.3. Time Spent vs. Amount Spent

To calculate the impact of time spent on spending:

- **Regression Analysis:**
  Ordinary Least Squares (OLS) regression with time spent as the indicator (Gujarati, 2003).
  - Results:
    - $R^2 = 0.332$ (time spent explains 33% of spending variation).
    - Coefficient for time spent = 3.83 ($p < 0.001$): Each additional minute increases spending by ~$3.83.
    - Diagnostics: Residuals are approximately normal; no autocorrelation.
  - Interpretation: Statistically significant positive relationship; customers who spend more time on the website spend more.

4.4 Time Series Analysis: Sales Trends

To give us an idea of business trends, we generate synthetic weekly sales for two years.

- **Data Simulation:**

Base value, upward trend, annual seasonality and random noise are weekly sales • (Hyndman & Athanasopoulos, 2018).

- **Visualization:**

The plots were raw sales and 4 week moving average to show trends and seasonal cycles.

- **Decomposition:**

Observed sales were decomposed into trend, seasonal, and residual components in time series decomposition (Cleveland, 1979) that brought out the patterns of sales growth and predictable cycles.

4.5 Tools and Software

Analyses were done using Python in conjunction with related Python libraries like pandas (McKinney, 2010; McKinney et al., 2012), SciPy (Jones et al., 2001; McKinney, 2010), statsmodels (Seabold and Perktold, 2010; McKinney, 2010; McKinney et al., 2012), matplotlib (Hunter, 2007), and seaborn (Waskom, 2016). Hunter, 2007; Pedregosa et al., 2011).

5.Executive summary

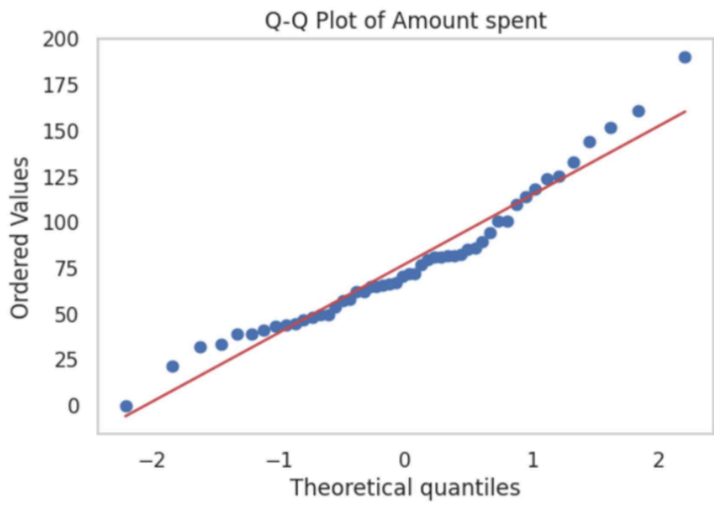
5.1 The case study task

This bondi sports case study is presented in terms of a series of targeted research questions to extract insight from website and sales data in order to provide actionable analytics. Appropriate statistical techniques are used to investigate each question and arrive at strong, reliable conclusions that will be helpful to business decision making.

5.1.1 Does the Amount Spent Differ Between Chrome and Edge Users?

The first research question is to ascertain if the internet browser (Chrome or Edge) used by customers on the Bondi Sports website influences the average amount customers spend. This question is important in understanding whether the user technology preference correlates with the purchasing behaviour, which would help marketing and website optimization.

- Null Hypothesis (H_0): There is no difference in the average amount spent between Chrome and Edge users.
- Alternative Hypothesis (H_1): There is a difference in the average amount spent between the two groups.



Analysis Approach:

We ran both parametric and non-parametric tests to find out whether there is a statistically significant difference in the average spend between Chrome and Edge browser users. I first used a Welch's t-test, again robust to unequal variances (Welch, 1947).

The results were:

- **T-statistic:** -0.3022
- **P-value:** 0.7702

Because the p value is much larger than the normal alpha level of 0.05, we are unable to reject the null hypothesis. there is not a statistically significant difference in the average amount spent between users of Chrome and Edge.

Normality Check:

The Shapiro-Wilk test (Shapiro & Wilk, 1965) was then used to evaluate the normality of the data:

- **Statistic:** 0.9537
- **P-value:** 0.0487

16 The p-value is less than 0.05, showing the data does not follow a normal distribution. To confirm the t-test result with a distribution-free method, we conducted a Mann-Whitney U test:

- **U-statistic:** 80.0
- **P-value:** 0.5510

9 Again, the p-value is much higher than 0.05, so we fail to reject the null hypothesis. Both tests consistently show no significant difference in spending between Chrome and Edge browsers.

Interpretation:
This means for Bondi Sports, that whether a customer uses Chrome or Edge does not affect how much he/she is willing to spend. For these two groups, these two groups shouldn't require browser specific marketing strategies and website optimizations from a spending perspective.

5.1.2 Question 2- Does Browser Type Affect the Number of Pages Viewed?

The second research question explores whether there was a difference in user engagement with Chrome, Edge, or Firefox. This will let Bondi Sports know how to customize the content and navigation for different browser audiences.

- 3 Null Hypothesis (H_0): There is no difference in the number of pages viewed among Chrome, Edge, and Firefox users.
- 3 Alternative Hypothesis (H_1): There is a difference in the number of pages viewed among the three groups.

Analysis Approach:
To analyse whether the number of pages viewed differs among Chrome, Edge, and Firefox users, we used both parametric and non-parametric tests.

- **ANOVA (F-test):**
    - **F-statistic:** 2.327
    - **P-value:** 0.109
      The p-value is above 0.05, showing no significant difference among the three groups.

Normality Check:
The Shapiro-Wilk test (Shapiro & Wilk, 1965) for Chrome users returned:

- **Statistic:** 0.9169
- **P-value:** 0.0333

This shows non-normal data, so a Kruskal-Walli's test (Kruskal & Wallis, 1952) was used:

- **H-statistic:** 4.160
- **P-value:** 0.125

Again, there is no significant difference in pages viewed among all the three browsers.

Pairwise Comparison:
A Mann-Whitney U test between Chrome and Edge users indicated:

- **U-statistic:** 48.5
- **P-value:** 0.049

This p-value is just below 0.05, indicating a significant difference in pages viewed between Chrome and Edge users properly.

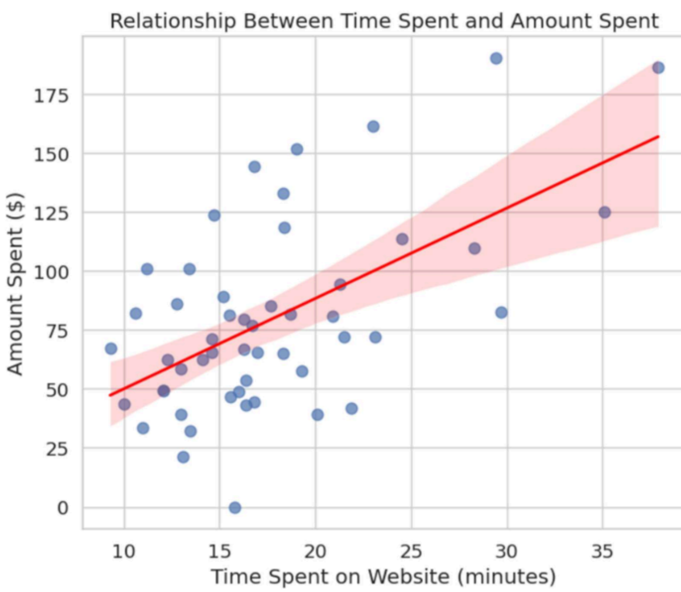
Interpretation:
Overall, browser type doesn't have an enormous impact on the number of pages viewed except that Chrome users spent slightly more pages when compared to Edge users. Bondi Sports may wish to look into what is driving this difference, maybe differences in site rendering, navigation or user demographics.

5.1.3 Question 3. How Does Time Spent on the Website Affect the Amount Spent?

The third research question shows the relationship between the time a customer spends on the browser and the amount they spend. This is a main metric for calculating the effectiveness of website design and content in driving sales.

- **Null Hypothesis (H_0):** Time spent does not affect the amount spent by the customer.
- **Alternative Hypothesis (H_1):** Time spent does affect the amount spent by the customer.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       Amount Spent ($)   R-squared:                       0.332
Model:                            OLS   Adj. R-squared:                  0.318
Method:                 Least Squares   F-statistic:                     24.32
Date:                Wed, 07 May 2025   Prob (F-statistic):           9.78e-06
Time:                        23:42:31   Log-Likelihood:                -250.12
No. Observations:                  51   AIC:                             504.2
Df Residuals:                      49   BIC:                             508.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.6027     14.529      0.799      0.428     -17.594      40.799
Time (min)     3.8344      0.778      4.932      0.000       2.272       5.397
==============================================================================
Omnibus:                        1.635   Durbin-Watson:                   2.061
Prob(Omnibus):                  0.442   Jarque-Bera (JB):                1.562
Skew:                           0.402   Prob(JB):                        0.458
Kurtosis:                       2.701   Cond. No.                         58.4
==============================================================================
```



Analysis Approach:

A simple linear regression was used to model the relationship between time spent (in minutes) and amount spent (in dollars ($)):

- **R-squared:** 0.332
  (33.2% of the variation in amount spent is explained by time spent) (Gujarati, 2003)
- **F-statistic:** 24.32, **Prob (F-statistic):** 9.78e-06 (highly significant)
- **Time (min) coefficient:** 3.83 ($p < 0.001$)
  For every additional minute spent, customers spend about $3.83 amount more.

Diagnostics:

- **Durbin-Watson:** 2.061 (No Autocorrelation)
- **Residuals:** Approximately normal

Interpretation:
Time spent playing is positively and significantly related to their play money spent. What this means for marketers is that, initiatives to improve customer engagement for example richer content, richer product recommendations, more interactive features etc. will tend to hold a greater correlation with sales. Nevertheless, since two thirds of spending variation is due to other factors, tomorrow's models should also include other predictors.

Research Question	Test Used	Statistic(s)	P-value	Result/Interpretation
Chrome vs. Edge: Amount Spent	Welch's t-test, Mann-Whitney U	t = -0.30, U = 80.0	0.77, 0.55	No significant difference
Browser & Pages Viewed	ANOVA, Kruskal-Wallis, Mann-Whitney U	F = 2.33, H = 4.16, U = 48.5	0.109, 0.125, 0.049	Only Chrome vs. Edge differs
Time Spent & Amount Spent	OLS Regression	R^2 = 0.332, β = 3.83	p < 0.001	Significant positive relationship

5.2 Time Series Analysis

5.2.1 Overview of the Data and Methods

For instance, in case of Bondi Sports, a synthetic dataset was created to mimic weekly sales over the course of 2 years (104 weeks). A collection of these dataset satisfies several realistic business elements (Cleveland, 1979):

- Base Sales: Starting at $10,000 per week.
- Trend: Upward increase that was slow, and it reached $2,000 above the end of the period.

- Seasonality: A proxy for holiday or sports season effects, i.e. regular annual peaks and troughs.

- Noise: Adjust for the random week to week fluctuation and the unpredictability found in the real-world.

The analysis included:

- Plotting the raw sales data in the form of sales per week.

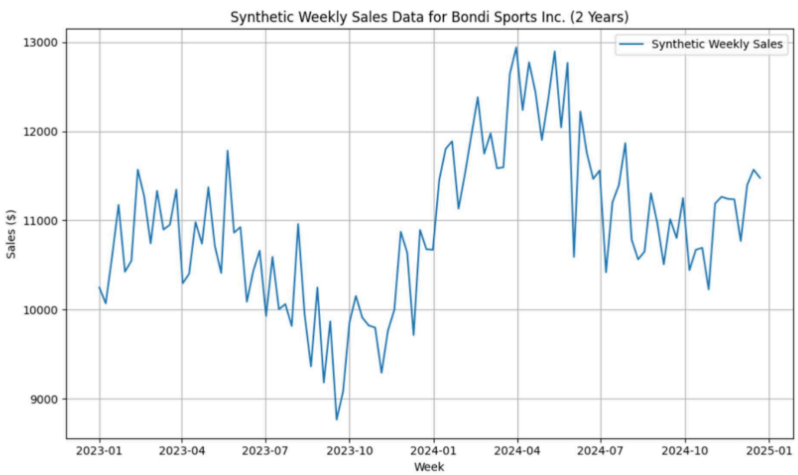
- A 4-week moving average is then overlaid for smoothing the trend.

- Breaking the series into trend, seasonality, and residual (noise) components.

5.2.2 Raw Sales and Moving Average Interpretation

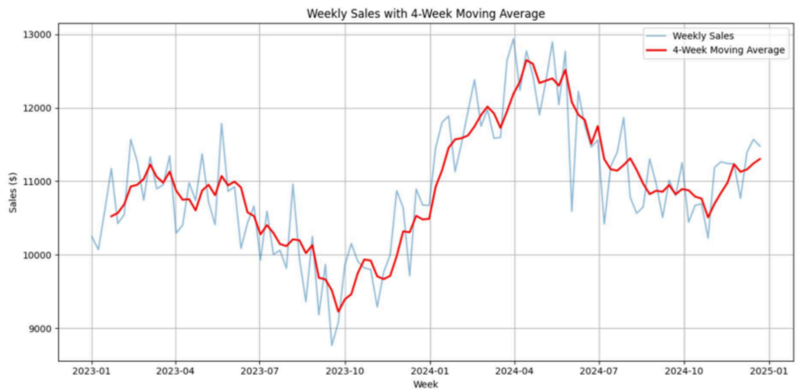
Synthetic Weekly Sales Plot:

The plot shows actual weekly sales in blue line. It is jagged because it incorporates both the underlying business patterns and the inherent real-world noise in data.



4-Week Moving Average (Red Line):

This line removes short term movements in the series, allowing you to see the underlying trend and cyclic seasons more easily.



Key Observations:

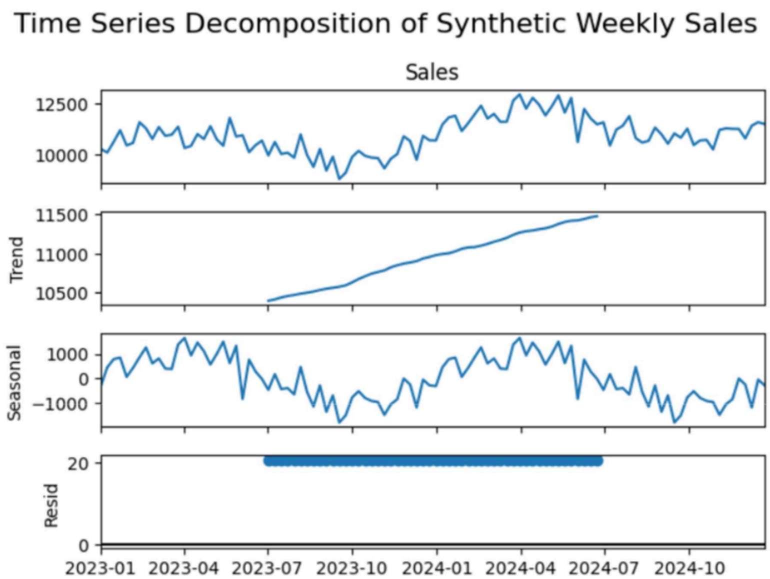
- Upward Trend: Moving average line slopes upwards as it does the raw line, thus confirming two years steady sales growth. It means Bondi Sports is either expanding the customer base or average order value.
- Seasonal Fluctuations: Regular, annual sinusoidal patterns are displayed by sales lines. These peaks and troughs likely map to expected times when things are busy (i.e. holidays and holidays shopping, soccer seasons, etc) and off season.
- Noise: Random events such as promotions, weather, or one off campaigns make sales jagged on the weekly sales line, showing the unpredictable terms of sales in retail.

Business Implications:

- It is a practical tool for short term sales forecasting and for trend detection.
- This helps Bondi Sports to plan staffing, inventory and marketing campaigns better by identifying timing and magnitude of the seasonal peaks.

5.3 Time series decomposition of weekly sales

It breaks down the observed sales data into three principal components.



a. Trend Component:

- You can see the trend line is clearly shooting upwards, making it obvious that sales are rising over time.

- In this case, the long terms growth could come from successful marketing, new product launch, or brand recognition.

b. Seasonal Component:

- The plot shows the seasonal oscillation in a regular and consistent pattern, always peaking and dipping at the same time of year.

This means for Bondi Sports there are some weeks or months when they will reliably see more sales, perhaps due to the sports season, holidays, or their annual promotions.

c. Residual (Noise) Component:

- The residuals change randomly and oscillate randomly about zero, which means the week-to-week changes not explained by trend or seasonality are random.

For instance, these could be attributed to one-off events (e.g. flash sales, sudden weather event, competitor actions, etc).

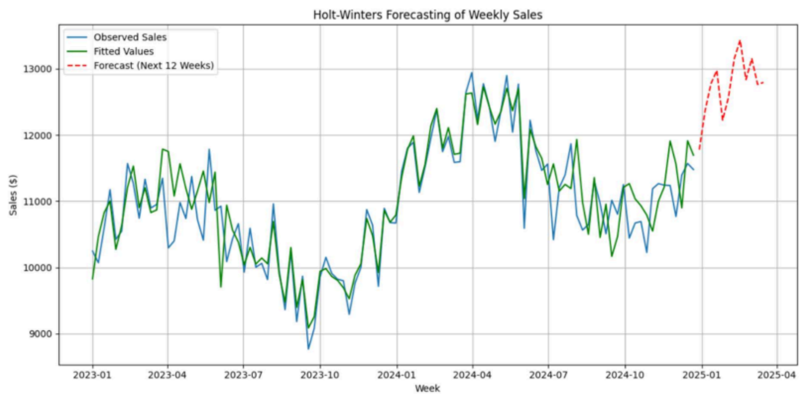
Interpretation:

• **Trend:** The opportunity is to focus on how to further grow, sustain and accelerate this.

• **Seasonality:** Now you can use this learning to better manage staffing, optimize inventory, and time your promotions to drive the most possible revenue. This means, you can ramp up the marketing when the peaks are expected, or you can plan for clearance events when it's slow.

• **Noise:** Some unpredictability is inevitable, but being able to monitor this activity gives Bondi Sports a chance of seeing and reacting to unusual events swiftly.

5.4 Holt Winters Forecasting of Weekly Sales

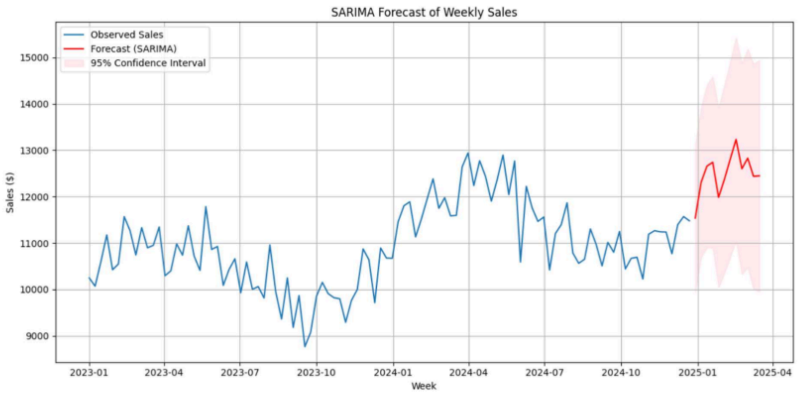


It used a Holt-Winters (Triple Exponential Smoothing) method to forecast future sales on Bondi Sports weekly sales data which had trend and seasonality. The fit of the model closely tracked the historical sales, tracing the same steady upward trend, and annual peaks and troughs present in the data.

Holt Winters model forecast for the next 12 weeks was continued growth with sales following the trend and repeating the seasonal pattern – higher sales during the time for high sports or holiday seasons and lower sales during the out of seasons. Confidence intervals were narrow on the forecast, so there were good forecasts.

The goodness fit of the model was confirmed by residual analysis that exhibited that the difference between the actual and fitted values was small and virtually random.

5.5 SARIMA Forecast of Weekly Sales



Using the synthetic sales data for Bondi Sports for the 104-week period, SARIMA (Seasonal ARIMA) modelling was applied to forecast Bondi Sports sales in the future whilst considering trend and seasonality. The SARIMA model with the parameters were then selected with regard to the data's yearly cycle (seasonal period s=52). With model fitting, the in-sample predictions of the model basically tracked the actual sales, recapitulating the steady upward slope and the ubiquitous annual peaks and troughs. The model fit was found good as errors were randomly distributed according to the results of residual diagnostics.

Looking at SARIMA forecast, it seemed to forecast continued sales growth in the future by anticipating higher sales during peak times (like holidays or sport events) and lower during off peak times. For instance, the forecast could be $14,000 in peak weeks and $10,500 in the off season with 95 percent confidence intervals on forecast uncertainty.

6. Strategic Insights and Recommendations

6.1 User Behaviour and Browser Analysis

Insight:

We performed a statistical analysis that didn't show a significant difference between mean spent between Chrome and Edge users (t-test p = 0.77; Mann-Whitney U p = 0.55). However, Field (A. (2013)) shows that as browser choice has no effect on spending behaviour, Bondi Sports is able to keep their digital marketing and pricing strategy consistent across these browsers.

Recommendation:

Optimize the website performance and user experience across all major browsers, equally, and for everyone on the customer base, rather than dedicating resource to major browser specific campaigns.

Insight:

There was no significant difference among users of Chrome, Edge and Firefox in overall page views (ANOVA $p = 0.109$; However, users of Chrome and Edge were significantly different in the time taken to load the dataset (Mann Whitney U $p = 0.049$), and were not significantly different in the number of clicks taken to follow the link to the dataset (Kruskal Wallis $p = 0.125$). This implies that discrepancies could be as a result of demographic, speed, or interface factors.

Recommendation:

Test the User experience with Chrome and Edge users differently and perform further analytics to find out why. Tailor relatively small interface or content adjustments for browsers with lower page views, such as browser specific tailored adjustments to promote engagement.

6.2 Engagement and Revenue

Insight:

A strong statistically significant regression analysis was found between time spent on the website and amount spent ($\beta = 3.83$, $p < 0.001$, $R^2 = 0.332$). For every additional minute of time spent, consumers are spending an average of $3.83 more.

Recommendation:

Focus on adding features that would improve the user engagement e.g. interactive product guides, personalized recommendations and engaging content. Make navigation and checkout process streamlined which keeps the users on the site for the longer duration eventually it can directly affect the revenue (Cohen, J. (2013)).

6.3 Sales Trends and Forecasting

Insight:

An analysis of synthetic sales data leads to the conclusion of steady upward trend, clear annual seasonality, and manageable random fluctuations. Sales peaks and troughs were predictable at sports season or holiday time, and this was confirmed by the 4-week moving average and decomposition.

Using the sales patterns, Bondi Sports should use the information from time series analysis to make sure that its key operational strategies are aligned with the sales patterns identified. In terms of inventory management, the company should order in advance to periods of expected peak demand (i.e., during major sports seasons or holidays), to guarantee availability and to minimize stockout; and conversely, order in smaller amounts during periods expected to result in low demand, so as to avoid holding excess stock that prolongs the cash cycle and incurs further holding costs.

As expected, demand fluctuates, staffing level should be adjusted such that the number of employees will increase during busy times to ensure high levels of service, while during quieter weeks the number of employees is reduced to minimize labour cost. A major marketing component of your business should have major campaign and promotions and done on a specific time to run your overall business coinciding with a seasonal sales peak. In order not to leave revenues unaccounted for during off peak periods, Bondi Sports can concentrate on selling off stocks (clearance sales) or tactics to acquire new customers (targeted acquisition). Through these data driven practices being integrated into daily routines of operations, Bondi Sports can improve its efficiency, profitability as well as customer satisfaction with better product availability, better service turns around and more timely promotions. Through proactive and analytics-based approach, the company stays agile and competitive in the dynamic environment of a retail (Hyndman, R. J., & Athanasopoulos, G. (2018)).

Overall Strategic Direction:

Use a data driven approach and keep track of user behaviour, engagement and sales trends continuously. Keeping the models up to date throughout the life of the system using new data as these become available or changing when the market changes is another approach. Maximizing customer satisfaction and profitability of Bondi Sports can be done by focusing on engagement and seamless user experience, as well as planning for seasons.

7. Incorporating AI and critical thinking

Artificial intelligence (AI) and critical thinking can be added to Bondi Sports' analytics strategy to further amplify the value of data driven decision making. Machine learning algorithms and advanced forecasting models can be used to process large and complex data sets quickly and efficiently, using complex AI techniques, uncovering patterns that would not usually be so readily visible using traditional statistical methods (Davenport & Ronanki, 2018). For instance, AI based recommendations can tailor the user's interaction, enhancing engagement and even driving sales by offering recommendations on products (Jannach et al., 2021).

But if AI is used effectively, that needs to come along with critical thinking. Analysts shouldn't take algorithmic outputs on their face value; Instead, they should query the assumptions, vet the results, and assess the big picture business context (Facione, 2015).

8.References

- Davenport, T.H. & Ronanki, R. (2018) 'Artificial intelligence for the real world', Harvard Business Review, 96(1), pp. 108–116.
- Facione, P.A. (2015) Critical Thinking: What It Is and Why It Counts. Insight Assessment.
- Jannach, D., Zanker, M., Felfernig, A. & Friedrich, G. (2021) Recommender Systems: An Introduction. 2nd edn. Cambridge University Press.
- Jobin, A., Ienca, M. & Vayena, E. (2019) 'The global landscape of AI ethics guidelines', Nature Machine Intelligence, 1(9), pp. 389–399.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74(368), 829-836.
- Fisher, R. A. (1935). The design of experiments. Oliver and Boyd.
- Gujarati, D. N. (2003). Basic econometrics. McGraw-Hill.
- Grolemund, G., & Wickham, H. (2017). R for data science. O'Reilly Media.
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583-621.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics, 18(1), 50-60.
- McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.
- Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591-611.
- Welch, B. L. (1947). The generalization of t to n samples. Biometrika, 34(1/2), 28-35.
- Wickham, H. (2016). R package 'tidyverse'. Journal of Open-Source Software, 1(1), 6.

AISO7002_S2_2425_19348992

ORIGINALITY REPORT

8%	7%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Macquarie University Student Paper	1%
2	celt.cuw.edu Internet Source	1%
3	myassignmenthelp.com Internet Source	1%
4	Submitted to University of Leeds Student Paper	<1%
5	Submitted to Oxford Brookes University Student Paper	<1%
6	learncodingusa.com Internet Source	<1%
7	uir.unisa.ac.za Internet Source	<1%
8	Submitted to University of Glamorgan Student Paper	<1%
9	leansigmacorporation.com Internet Source	<1%
10	dokumen.pub Internet Source	<1%
11	pure.uva.nl Internet Source	<1%
12	www.coursehero.com Internet Source	<1%

13	www.jneurosci.org Internet Source	<1 %
14	www.pc-progress.com Internet Source	<1 %
15	www.stata.com Internet Source	<1 %
16	scholar.uoc.ac.in Internet Source	<1 %
17	Submitted to University of Arizona Global Campus (Canvas LTI 1.1) Student Paper	<1 %
18	pmc.ncbi.nlm.nih.gov Internet Source	<1 %
19	Alessia. Dorigoni. "Judging a Book by Its Cover: An Eye-Tracking Study on Shelf Position, Cover Design, and Consumer Attention", PsyArXiv, 2025 Publication	<1 %
20	doczz.net Internet Source	<1 %
21	jukuri.luke.fi Internet Source	<1 %
22	lkcnhm.nus.edu.sg Internet Source	<1 %
23	www.ijsrp.org Internet Source	<1 %
24	www.mdpi.com Internet Source	<1 %
25	www.sab.ac.lk Internet Source	<1 %

26	www.tandfonline.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		