**QBUS 6600 Assignment 2**

**Big W Project**

Hongjin Gou 510332280

Word count: 5193 words

15/10/2023

Contents

[Executive Summary 1](#_Toc148178043)

[Introduction 2](#_Toc148178044)

[1.1 Business context and problem formulation 2](#_Toc148178045)

[Data processing, EDA, and feature engineering 3](#_Toc148178046)

[2.1 Data Processing 3](#_Toc148178047)

[2.2 Exploratory Data Analysis (EDA) 3](#_Toc148178048)

[2.3 Feature Engineering 6](#_Toc148178049)

[Model Building 6](#_Toc148178050)

[3.1 Ordinary Least Squares Model 6](#_Toc148178051)

[3.2 Elastic Net Model & Lasso Model 7](#_Toc148178052)

[3.3 Decision Tree 9](#_Toc148178053)

[3.4 Bagging 10](#_Toc148178054)

[3.5 Random Forest 12](#_Toc148178055)

[Recommendations and Conclusions 14](#_Toc148178056)

[4.1 Recommendation 14](#_Toc148178057)

# Executive Summary

To provide a comprehensive analysis of the key factors influencing the overall sales value and suggests strategies for enhancing future sale performance. The initial step is to import external data and merge it with the data given by BigW. Rather than focus on certain state, we have chosen to evaluate all states collectively to ensure that our further analysis remains inclusive of the entire Australia market since final decision would be made based on the whole market.

Throughout the EDA process, we depict the in-depth visualization of the dataset. Subsequently, in the model building process, we proceed with the ordinary least square (OLS), lasso, elastic net, xgboosting, and random forest models, using a selection of features to estimate their performance on the test set. Upon identifying the most effective model, we utilize insights gained from the model to provide recommendations for improvement and supplementary information as well. There are some key findings during the analysis.

**Key findings:**

* Sales in the NSW and QLD stores collectively account for more than 50% of the total sales in all of Australia.,
* Approximately 10% of the total sales are attributed to online channels, suggesting that online shopping and delivery may need further enhancements to strengthen its performance compared to K-mart.
* Incorporating external data sources with existing datasets is necessary since the original one is not sufficiently qualified to build models for prediction due to the presence of data leakage issues.
* The feature selection process has identified certain factors with notably higher correlations to total sales. These key features include consumer lifestage, geographical distance, weekly income levels, and sales channels.
* Grouping customers by region (postcode) based on their weekly income levels reveals distinct differences in purchasing power.
* The lifestage group categorized as "Budget" contributes to roughly 50% of the total sales.
* While most models perform similarly, it is worth noting that the complicated models demonstrate relatively superior performance compared to other simpler modeling techniques.
* Our recommendation is to prioritize efforts aimed at improving the online sales channel, brand loyalty and better customer experience in the regions where the majority of Budget and Mainstream lifestage group members are located, particularly in areas with medium weekly income levels, because these regions collectively represent 80% of total sale value.
* Catering to Premium and high-income demographics, we consider enhancing customer services, the standard of product quality, and the overall shopping experience, with a special focus on the physical stores in the high-income regions.

# 1. Introduction

## 1.1 Business context and problem formulation

BigW is one of the largest retail companies in Australia, owned by WoolWorths, offering a wide range of common and everyday necessities. In the past, its primary competitors were K-mart and Target. However, with the development of ecommerce, online shopping companies such as Amazon and Ebay have attracted core customers from BigW. Recognizing this as a threat, BigW aims to expand its online business channel, while maintaining its strong in-store sale performance. Additionally, BigW invests approximately $50m annually in various social media platforms to attract more customers.

To address the competitive and tough situation, the primary objective of the project is to identify data-driven strategies for optimizing revenue generation and marketing investment. To achieve this, we collected some data from external resources like ABS, including information on income, holiday, and population density. Our dataset includes a range of features, including gender, sales channels, average household size, median weekly household income, distance\_to\_target, co\_location\_flag, customer\_postcode, store\_id, store\_postcode, lifestage, customer\_price, people, distance\_to\_kmart, financial\_week\_end\_date, and store\_state, and some of which have notable correlation with weekly-ending total sale value.

Furthermore, we have leveraged all available data, including training set, media investment set, and store information set to construct our foundational dataset. Consequently, we have then thoughtfully selected certain variables from the extensive features set and have combined these subsets as key features that will be shown in the EDA part for our further analysis.

As proceeding with the model building process, we would conduct various machine learning approaches to predict total store sales, such as decision tree, random forest, and elastic net. Additionally, RMSE and R square would be the estimator to assess the performance of models on the test set. The ultimate purpose of the best performed model is to identify the most relevant features that significantly contribute to maximizing the revenue. By uncovering these critical factors, the company can generate targeted strategies to enhance the overall total sale value and therefore strengthen its competitive position in the market.

# 2. Data processing, EDA, and feature engineering

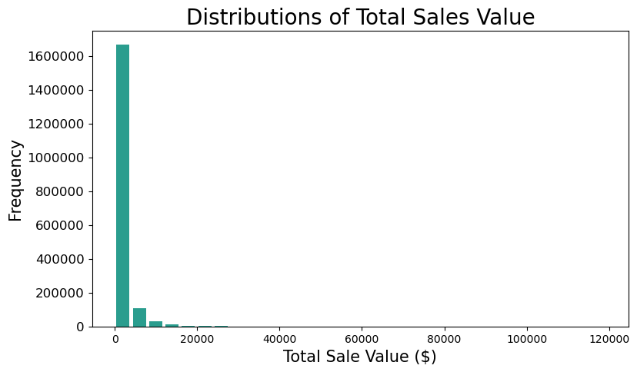
## 2.1 Data Processing

Before conducting the Exploratory Data Analysis (EDA), we merged multiple original datasets and the external data we intend to use, based on store id and store postcode. Following this, we eliminated irrelevant features for the upcoming study, such as duplicate features generated during the data merging. In addition, we removed some features that could cause data leakage, such as the transaction count, total promotional sales value for each store, and customer data. Data leakage means allowing the model to learn or know something it otherwise would not know from the training dataset and invalidating the estimated performance of the constructed model (Brownlee, 2020). These pieces of data should be removed due to data leakage, we would need access to when predicting the weekly total sales amount for each Big W store in the future.

Regarding the 'price\_lifestage\_segment' feature, we split it into two new features named 'Customer\_price' and 'Lifestage' to further explore their impact on the target value in the subsequent analysis. Then, we filled the missing values for 'Customer\_price', 'Lifestage', and the categorical features in the added external data with the string 'NA'. For the missing values in the weekly media amount spent, we filled them with the number zero.

## 2.2 Exploratory Data Analysis (EDA)

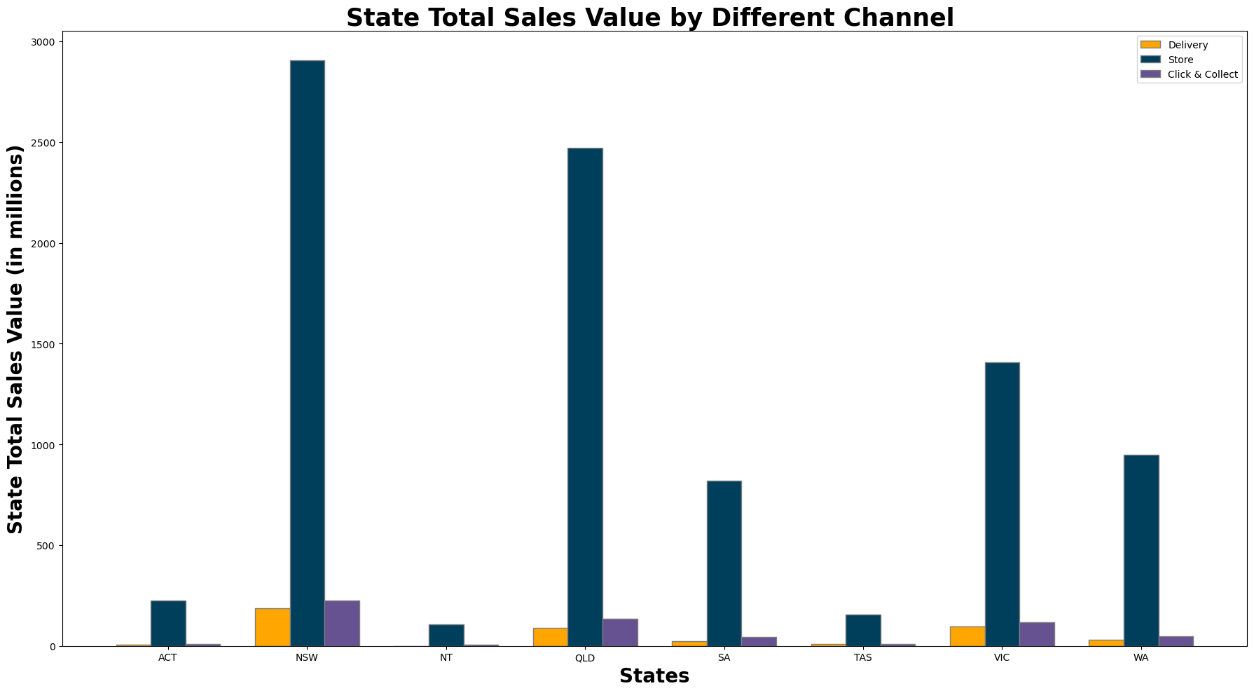
Exploratory Data Analysis (EDA) is an essential step before modelling the total weekly sales of each Big W store. Furthermore, EDA acts as a precursor to building models by visualizing and understanding the distribution of our target variable (total sales value) and its relationship with other features. We can take the right action regarding feature engineering, outlier processing, and feature transformation. In addition, we also added some primary data corresponding to each postcode area, such as people, families, and median weekly household income. We will analyze these data to find the specific factors that each customer group affects each Big W store's total weekly sales volume.

A graph of a log of sales

Description automatically generated with medium confidence

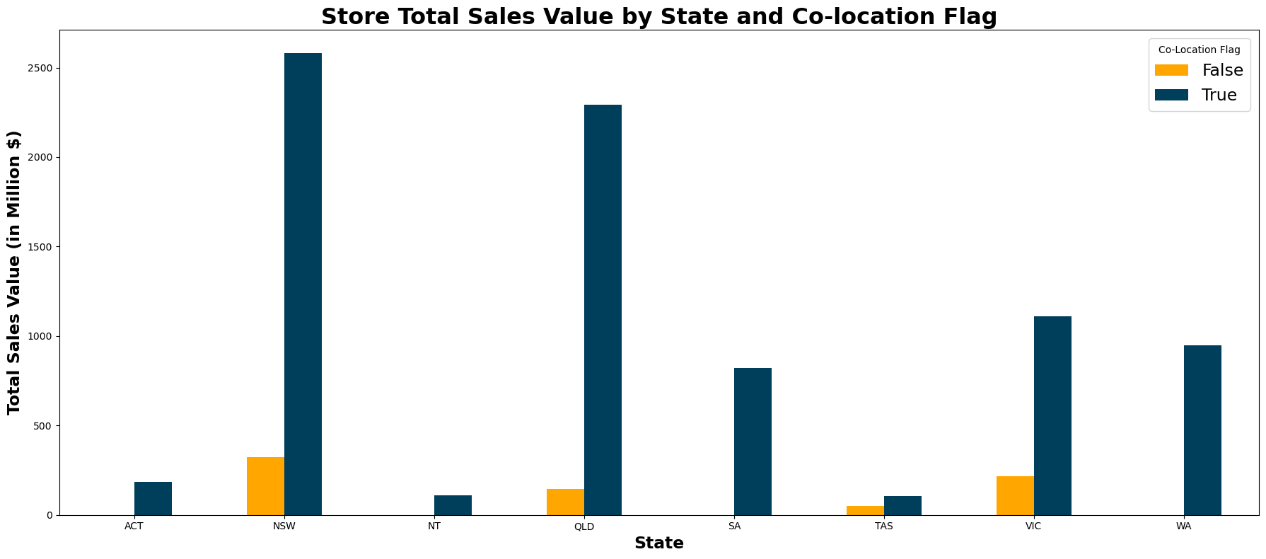
*Figure 1. Distribution of total sales value*

The analysis begins with an examination of weekly total sales value distribution. In figure 1, most sales values are spread across 0 to AU$4000, with a few spikes indicating weeks with exceptionally high sales. It is important to identify the tail, which contains relatively high values. However, these values ​​are less frequent, as they could be linked to promotional events, holidays, or other external factors affecting sales. Then, we examined the logarithmic transformation of the sales values as shown in the second image, "Distributions of Log Total Sales Value". The log-transformed sales value data becomes more bell-shaped than before, which means it is a normal distribution. Data transformation is beneficial when preparing data for machine learning models or statistical analysis, as many algorithms and techniques assume data is the normal distribution. Using a logarithmic scale often helps reduce the impact of extreme values and improves our model's prediction performance.



*Figure 2. Bar chart of total sales value from different channel*

The chart above shows that walk-in sale is the dominant sales channel across all states, significantly outperforming delivery and click-collect. The main reason for this phenomenon is that when customers shop in stores, it provides an opportunity for consumers to engage with products, assessing aspects like fit, feel, and efficiency. This situation is particularly vital for products such as apparel, shoes, or gadgets. Additionally, in-store purchases offer instant gratification, eliminating the waiting time tied to delivery. Moreover, the in-store shop can also bypass potential shipping costs, which can be significant, especially for faster delivery methods. Furthermore, direct delivery sometimes requires the recipient's signature, and in its absence, the parcel might be redirected to a pickup point, degrading the shopping experience.



*Figure 3. Bar chart of total sales value and co-location flag*

The figure 3 shows that when Big W and Woolworths are located in the same geographical location, sales are 8 to 10 times that of other Big W stores. This integrated shopping model means customers can conveniently purchase various goods at the same location, from daily groceries to household necessities. Furthermore, this model will attract more potential customers. When Big W and Woolworths are located at the same location, they will share the customer traffic brought by each other. Customers of Woolworths may be attracted to browse the products of Big W, and vice versa. Shared marketing strategies or joint promotions might exist between Big W and the nearby Woolworths store, which would attract more customers and promote sales.

A pie chart with text on it

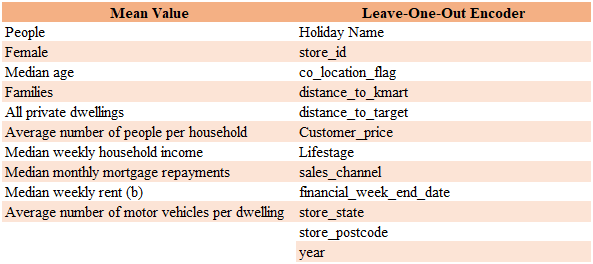
Description automatically generated

*Figure 4. Bar chart of sales percentage by customer price*

Big W's two primary customer price segments are 'budget' and 'mainstream', constituting 44.77% and 38.04% of its market share, respectively. This phenomenon indicates that a large portion of Big W's customers are cost-conscious and often seek value for money. 'Mainstream' follows closely behind, implying that these customers balance affordability with product quality.

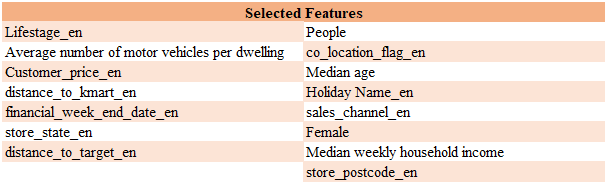
## 2.3 Feature Engineering

Before conducting Feature Engineering, we divide the test data into the train and validation sets. This operation is because we can use the validation set to evaluate the performance of multiple different models and, based on the results, select the best model to train again and apply to the final test set. In this way, we can also know whether our model has overfitting or underfitting issues while also avoiding data leakage.



*Figure 5. Table of mean values*

After splitting the original dataset, we use the mean values of the train set and validation set to fill the features shown on the left side of the table above. In addition, we applied a leave-one-out encoder to the features on the right side of the table. This processing can convert categorical data into numerical data types, which facilitates including these features in a linear model. Moreover, it can enhance the performance of these features in tree models.



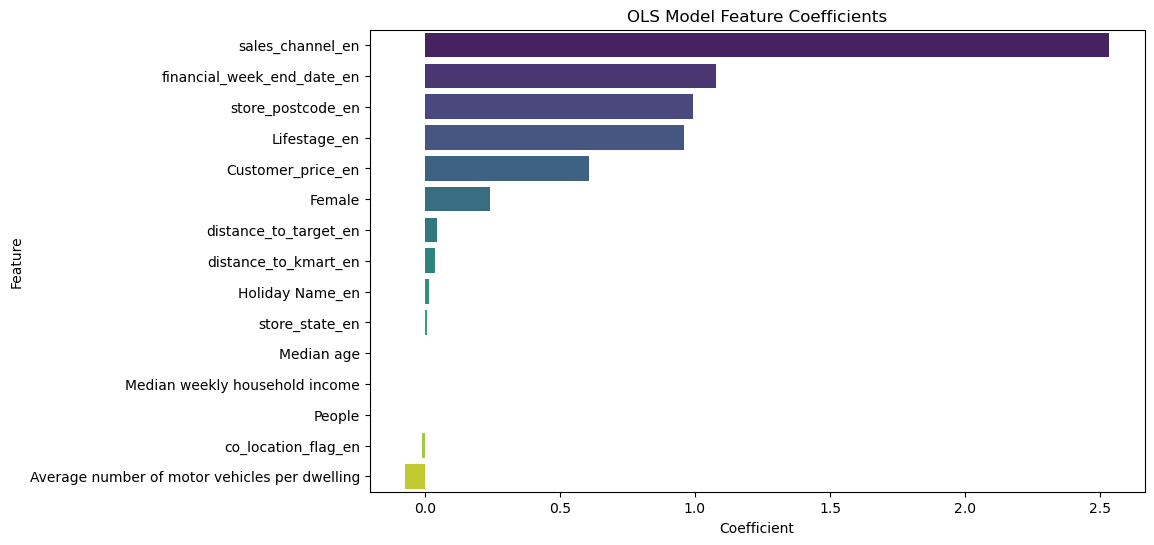
*Figure 6. Table of selected features*

Before building the model, we need to select suitable features. Our basis is the correlation among features to ensure that multicollinearity does not occur, as well as the correlation between features, our target value, and the Recursive Feature Elimination (RFE) filter. Combining these factors, we have selected the features shown in the figure 6 for our model.

# 3. Model Building

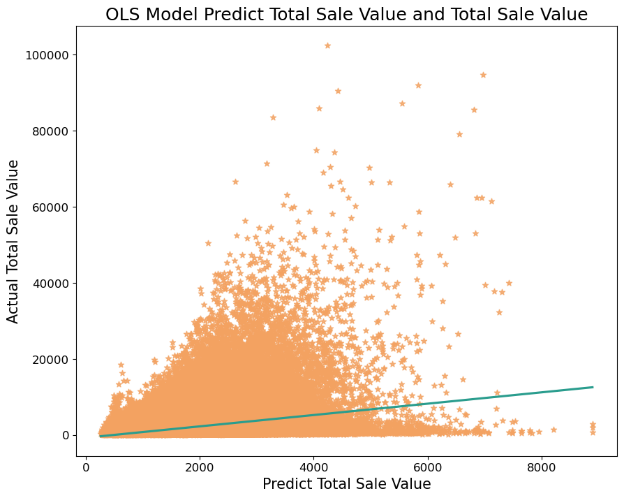
## 3.1 Ordinary Least Squares Model

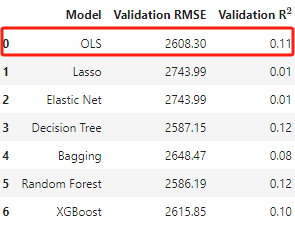
Ordinary Least Squares (OLS) models assume that the analysis establishes a model to elucidate the relationship between one or more explanatory variables and a continuous variable, intending to minimize the sum of square errors (Zdaniuk, 2014). We chose OLS as the first model due to its high interpretability and low computational cost while also using it as our benchmark model.



*Figure 7. OLS model feature coefficients*

Besides, we can further understand the contribution of the features we selected to the model by looking at the coefficients of each feature. As shown in figure 7, the linear relationship between the features (‘dictance\_to\_target\_en‘ ‘dictance\_to\_target\_en‘Median age, Median weekly household income, People) and the target value is essentially zero. The feature with the most vital linear relationship with the target value is sales channel.





*Figure 8. OLS model performance*

The figure 8 shows that the performance of the OLS model is not ideal due to many of the features we selected being of categorical type. Therefore, in the following model selection, we will consider more complex models as well as decision tree models.

## 3.2 Elastic Net Model & Lasso Model

In order to explore whether using other linear models could yield better prediction results, we decided to use the Elastic Net Model for further research. Then, through the l1\_ratio of the Elastic Net Model, we selected the Lasso Model to compare its performance with that of the Elastic Net Model.

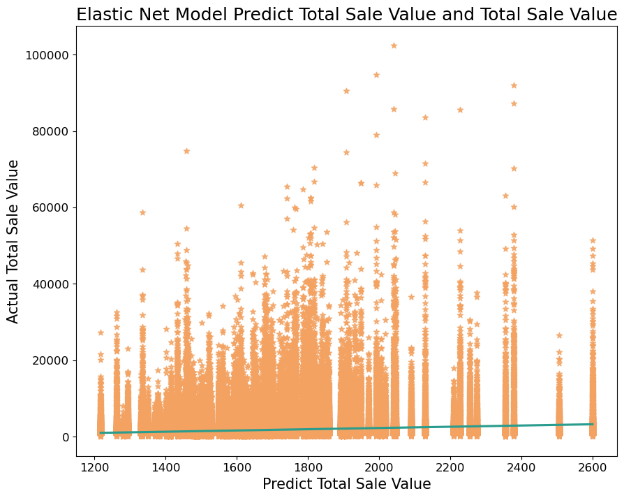
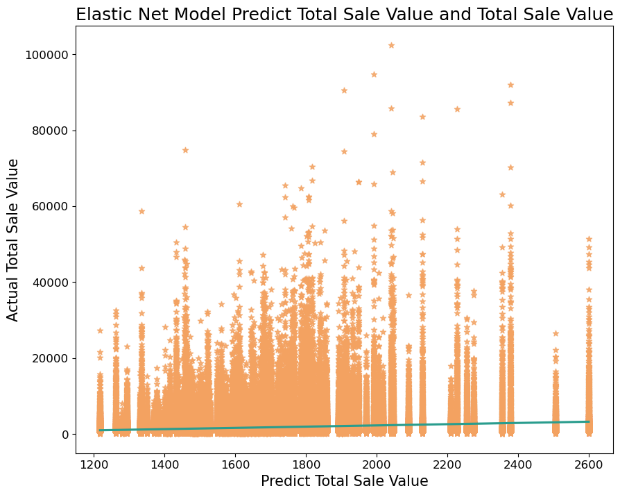
A diagram of a circular object with arrows

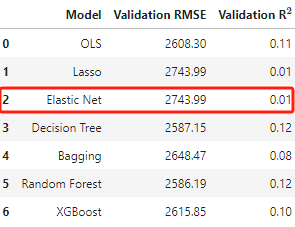
Description automatically generated

*Figure 9. Model comparison*

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve the regularization of statistical models (CFI Team, n.d.). Compared to basic linear regression (OLS), the main advantage of using Elastic Net is that it can apply regularization, which helps to avoid overfitting and improves the model's generalization ability. Due to L1 regularization, Elastic Net selects features by pushing the coefficients of less critical features towards zero.

In the modelling phase, the Elastic Net Regression model is utilized to predict the target variable based on the selected features. Elastic Net Regression is instrumental when there is a correlation among features, as it combines the properties of Ridge and Lasso regression, allowing for learning sparse models where few feature weights are non-zero while maintaining model complexity. We employ the Elastic Net model with cross-validation to select the best hyperparameters for the model, thus reducing the risk of overfitting the training data. We defined a range of L1 ratios, creating 20 values between 0.01 to 0.99. The L1 ratio corresponds to the trade-off between Lasso (L1) and Ridge (L2) regularization, where a ratio of 1 corresponds to Lasso, and 0 corresponds to Ridge. After training the model on the train set, we obtained an l1\_ratio of 0.99. This value indicates a strong preference for Lasso regularization. Hence, we proceeded to use the Lasso model to predict the dataset.





*Figure 10. Elastic net model performance*

The figure 10 shows that the Elastic Net and Lasso Model's prediction performance is consistent. Nevertheless, their prediction outcomes are worse than the OLS model. This is because the Elastic Net and Lasso models may generate a simpler model, where many feature, coefficients are shrunk to zero, reducing the model's complexity. In some cases, this might lead to a decrease in the model's prediction accuracy.

## 3.3 Decision Tree

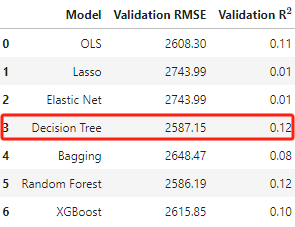
The decision tree is a supervised learning method that predicts outcomes by splitting features at various levels. Each internal node of the tree represents a test on an attribute, and each branch represents the outcome of that test, making the decision tree particularly suitable for understanding the multifaceted decisions involved in a sales process. Given BigW's dataset, applying decision trees to continuous variables provides an attribute importance score, aiding BigW in identifying key features for sales prediction. Moreover, compared to the previously mentioned OLS model, decision trees can capture non-linear patterns in data, which can be crucial for sales forecasting. Furthermore, unlike other intricate models such as Lasso and Elastic Net, decision trees offer more interpretability. This clarity is beneficial when the sales team, managers, or stakeholders of BigW seek insights behind the predictions.

A graph with orange and green lines

Description automatically generated

A graph with multiple colored bars

Description automatically generated



*Figure 11. Decision tree model performance and results*

Figure 11 shows that feature importance offers a metric indicating the significance and proportion of each feature in the construction of the decision tree model. In other words, the higher the score, the more crucial the feature is in the decision-making process. The store's postcode has the highest importance score at 54%, almost 2.5 times that of the second-ranking feature. This suggests that the store's postal code has a significant impact on the total sales value. This could be because, in Australia, postcodes are typically assigned based on geographical locations. Remote areas might have postal codes that cover broader territories, making the code a representation of location, which can easily affect store sales. Consequently, BigW might consider location-specific marketing campaigns or strategies tailored to the unique characteristics of each store's location in the future.

Lifestage ranks as the second most crucial feature, accounting for 21%. Based on its encoding, it indicates the life stage of customers, including young adults, retirees, and families with children. Different age stages and experiences play a pivotal role in influencing customers' buying patterns and, subsequently, the total sales value. For instance, younger individuals might be inclined to spend on trendy items, whereas retirees might have a preference for vintage styles. Given the importance of the customer's life stage, BigW could design marketing campaigns targeted at specific life stages, potentially boosting sales.

Ranked third is the financial week end date, suggesting that certain financial weeks or seasons can have a significant effect on sales. This might be due to holidays, promotions, or other cyclical market dynamics. Based on this, BigW might strategize sales activities at specific times of the week, such as weekend promotions, to attract potential customers and boost sales figures.

Subsequent features like the sales channel, median weekly household income, and median age have moderate importance scores. This suggests that while they do influence the model's predictions, their effect isn't as pronounced as the top features. For instance, the median age and median weekly household income might reflect the spending power of the customer base. On the lower end, features like distance to Kmart have relatively lower importance scores. In essence, while they do contribute to the model's decision-making process, they aren't primary drivers.

Upon predicting with the optimal tree model on the validation set, the previously applied logarithmic transformations were reverted to the original scale. As displayed in Figure 11, the Decision Tree's performance outshines earlier models, notably Lasso and Elastic Net. With a relatively low RMSE, the model's predictions, on average, deviate by 2587.15 from the actual sales figures. Its explanatory power, with R square of 0.12, far surpasses the 0.01 of preceding models. However, despite its comparative superiority, the model is still underfitting. This underperformance might stem from the inherent complexity or non-linearities in the data, causing it not to capture all variability.

Tree models efficiently allow BigW to discern influential features to aid decision-making, but they fall short in analyzing how each variable impacts the target variable. For example, it's challenging to determine if the relationship is linear, exponential, etc., or which direction it follows.

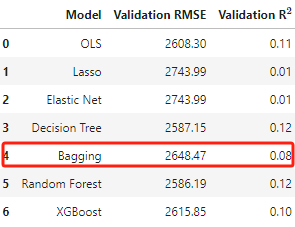
## 3.4 Bagging

Following the trial of the Decision Tree model, the Bagging Regressor was our next choice for exploring ensemble techniques. Bagging, which stands for Bootstrap Aggregating, involves training a decision tree, multiple times on different subsets of the training data. By averaging or voting on the predictions from these multiple models, Bagging tends to reduce variance, potentially offering an improvement over a single decision tree model.

In the initial phase, we harnessed the BaggingRegressor from scikit-learn and undertook hyperparameter tuning via RandomizedSearchCV. This approach, akin to the method employed with the tree model, employs random sampling of parameter combinations to ascertain the optimal setup. From this process, it was deduced that the model should employ 100 decision trees, taking advantage of averaging the predictions to diminish variance, thereby counteracting overfitting. Each decision tree in this ensemble is trained on a random subset comprising 50% of the original training data to ensure diversity among the trees. Furthermore, each tree uses only 50% of the original features. The bootstrap parameter being set to True indicates that samples are drawn with replacement, allowing a sample to be selected more than once for training a single tree. Such parameter adjustments are aimed at ensuring each base model in the Bagging ensemble is trained on varied data subsets, enhancing model diversity.

A graph with orange dots

Description automatically generated



*Figure 12. Bagging model performance*

The performance of the Bagging model is depicted in Figure 12, showcasing its prowess in predicting the total sales value. The chart accentuates the correlation between the predicted and actual sales values. The closer the data points are to the diagonal line; the more accurate the model's predictions are. From the performance metrics table provided, the Bagging Regressor achieved a Validation RMSE of 2648.47, signifying an average deviation of 2648.47 from the actual sales figures in the predictions. Additionally, with an R^2 score of 0.08, the model explains 8% of the variability in the target variable, indicating a moderate capability in capturing underlying patterns in the dataset. Compared with previous models, the Bagging Regressor outperforms both Lasso and Elastic Net models in terms of RMSE and R^2. Its RMSE is slightly greater than that of the Decision Tree model, suggesting its predictions have a marginally larger average error. Overall, even though its R^2 value lags behind that of the Decision Tree model, it presents a considerable enhancement over the Lasso and Elastic Net models.

In light of its performance relative to the Decision Tree and Random Forest models, there might be merit in delving deeper into more robust ensemble methods like Random Forests. Such exploration would involve comparing feature importance to glean more insights from the data and potentially improve prediction accuracy.

## 3.5 Random Forest

Progressing from the Bagging Regressor, we ventured into the Random Forest model, another ensemble technique but with certain differences from the previously explored models. Decision Trees are foundational building blocks in both Bagging and Random Forest. A single tree is grown to its maximum depth and may sometimes lead to overfitting due to its high sensitivity to minor fluctuations in the training data.

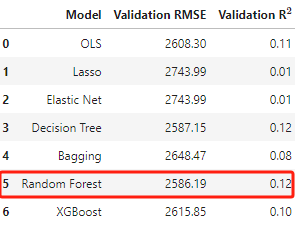
Bagging operates by training multiple Decision Trees on various subsets of the data. Each tree's prediction is averaged to produce the final output. Random Forests are an evolution of the Bagging method. While both use a collection of decision trees and sample data with replacement, Random Forests introduce an additional layer of randomness. At each split in the decision tree, only a random subset of the features is considered, which further diversifies individual trees and aids in decorating them.

A graph with different colored bars

Description automatically generated

A graph with orange dots

Description automatically generated



*Figure 12. Random forest model performance and results*

Figure 12 provides a visual representation of the importance of different features within the Random Forest model. This closely mirrors the results from the decision tree, reflecting the similarities in the logic between the two models. However, the Random Forest model introduces considerations based on the weight of features, thereby enhancing its robustness.

When compared to the tree model, the store postcode, Lifestage, and the financial week end date continue to occupy the top three positions. This reinforces the idea that the location of the store, the life stage of the customer, and the timing of promotional activities play crucial roles in influencing sales. On the other hand, features with lower importance, like distance to Kmart, have limited influence on the overall model. This could be due to a potentially low correlation between Big W and Kmart, which doesn't significantly impact sales. This is something to monitor continuously, as the importance might fluctuate based on the evolving market relationship between Kmart and Big W.

Same with tree models, the Random Forest model offers valuable insights into feature importance. Leveraging these insights can substantially aid BigW in refining its strategies to optimize sales and customer engagement. However, it retains the inherent limitations of tree-based models, wherein it doesn't explicitly showcase how each feature directly affects sales.

However, even though Random Forests aggregate results from many trees and give a sense of feature importance, it doesn't necessarily mean they always outperform a single decision tree or Bagging in terms of predictive accuracy. One reason is the focus of Random Forests on feature importance and model robustness, potentially at the expense of pinpoint prediction accuracy for specific datasets.

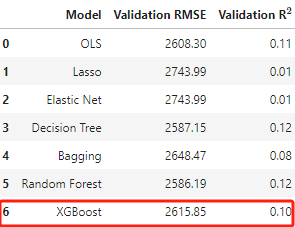
## 3.6 XGBoosting

XGBoost is a popular and efficient gradient boosting framework widely used in machine learning due to the strong predictive capabilities. Unlike the previously explored ensemble techniques, XGBoost works by sequentially adding trees to the model, attempting to correct the errors made by the preceding trees. This helps understand which features have a significant impact on the model's predictions.

In Figure 13, the median weekly household income is shown to be the fourth most important feature, accounting for 7%. This feature highlights the economic aspect of consumers. Areas with a higher median household income might be inclined to purchase pricier or premium products, thereby influencing sales. Sales channel and median age: These two features have notable importance, accounting for 4% and 3%, respectively. The sales channel underscores the platform of sale, while the median age reflects the age distribution of consumers.

A graph with different colored bars

Description automatically generated



*Figure 13. XGBoosting model performance and results*

In terms of performance, XGBoost has an RMSE of 2617.17, indicating that the model predicts with an average error of 2617.17. Compared to other models, XGBoost's RMSE is mid-range, lower than Lasso and Elastic Net, but slightly higher than Decision Tree and Random Forest. Its R² value is 0.10, implying that the model explains 10% of the data's variance. XGBoost performs relatively well but isn't the best. Its RMSE is moderate, while its R² is a bit lower than certain other models, like the Decision Tree and Random Forest. However, when compared to Lasso and Elastic Net, it significantly outperforms. It's worth noting that model selection isn't solely based on performance metrics; interpretability, runtime, and susceptibility to overfitting must also be considered. On the other hand, it's more complex and harder to explain than a single decision tree or linear model. Especially with a vast number of features or records, the training time could be relatively long, leading to increased computational costs.

In other words, XGBoost stands out as a potent contender in the list of models, showcasing robust performance and the ability to effectively handle non-linearities in data. For BigW, the choice of model might vary depending on the primary goal – be it sheer predictive capability, interpretability, or a mix of both. The accuracy and adaptability of XGBoost make it a top choice for predictive tasks, though BigW might need to invest resources to accommodate the longer training times and complexities in interpretation.

# 4. Recommendations and Conclusions

## 4.1 Recommendation

Based on the results from the random forest analysis, it's evident that sales values are closely tied to multiple features. Most notably, the store postal code, the life stage of consumers, the median weekly household income, sales channels, and weekly timing are paramount. In this context, a customer segmentation strategy is a commendable approach. When powered by data insights, they provide an unparalleled advantage for businesses aiming to resonate deeply with their target customers. For BigW, these data-driven insights regarding the significance of the weekly household income in shaping sales patterns present a golden opportunity. For instance, companies like Starbucks often adjust their menu prices based on the location of their outlets, understanding that areas with a higher median income might be willing to pay a little more.

The life stage of consumers is not just a mere statistic but also an insightful gauge of the community's purchasing power, their lifestyle preferences, and a clue into the kinds of products they lean towards. In areas with higher numbers of young customers, consumers don't just have the means but often the inclination to align with products that represent a certain lifestyle aspiration. Fast fashion, exclusive collections, and even higher-priced eco-friendly products can find a receptive audience in these regions. Brands like H&M or Forever 21, with their rapid fashion cycles, appeal to this demographic by launching influencer collaborations and limited-edition lines (Neff 2014). In this customer segmentation, BigW would do well to prioritize upscale branding, roll out premium products, and collaborate with influential personalities on social media for endorsements. Hosting exclusive sale events in these areas can not only enhance sales figures but also amplify BigW's brand prestige.

On the other hand, for median families, the narrative should pivot. For BigW, the emphasis on value for money becomes imperative. However, this focus is not merely on affordability; it's about delivering quality within a budget. Walmart, for instance, with its 'Everyday Low Prices' commitment, taps into this sentiment (Thompson 2004). Bulk deals, value packs, and products that assure longevity should be the highlights. BigW can carve a trust-centric narrative in these segments, ensuring patrons that while the pricing is competitive, there's zero compromise on quality. Marketing promotions in these parts should spotlight affordability, quality, and the brand's unwavering commitment to serving the community.

For retirees, the strategy should oscillate between the two extremes. While affordability remains at the forefront, there's latitude to introduce occasional luxury products, perhaps as limited editions or during special promotional windows. A good example might be Costco, which offers a mix of budget-friendly deals and luxury brands under one roof, ensuring that retirees can access both everyday essentials and occasional splurges (Courtemanche & Carden 2014). Bundle offers, where discounts are granted on purchasing complementary products, might strike a chord here.

Nevertheless, while product assortment and promotional tactics are foundational, what truly augments BigW's distinction is a relentless feedback loop. Amazon, with its detailed review system and continuous engagement through personalized recommendations, epitomizes this approach (Chen 2015). Regular touchpoints with customers, discerning their shifting needs, and assessing the efficacy of rolled-out strategies will ensure BigW retains agility, aligning swiftly with market oscillations.

4.2 Conclusions

In summation, by tailoring its strategies in harmony with the fiscal landscapes of various regions, BigW isn't just optimizing for immediate sales; it's fostering enduring relationships. Through a more precise and data-driven lens, BigW is poised to elevate its market share. Over time, this methodology will not only drive sales but solidify BigW's positioning as a brand genuinely attuned to its community, forging a bond interlaced with corporate social responsibility.

In conclusion, this report provides a comprehensive analysis of the factors and models impacting total sale values. Various factors influence the total sale value to some extent, with two models demonstrating relatively better performance in prediction than others. By assessing a range of features and their impacts on the target variable, certain features have shown significant coefficients. Additionally, the complex models perform well on the test set, as indicated by the RMSE, marking them as preferred candidates for future revenue prediction and recommendation. Consequently, we have consolidated the key features, such as life stage, weekly income, and regions, to generate a systematic set of recommendations aimed at improving BigW’s overall performance in the future

**Reference**

Brownlee, J. (2020, August 15). *Data Leakage in Machine Learning*. *Machine Learning Mastery*.

<https://machinelearningmastery.com/data-leakage-machine-learning/>

CFI Team. (n.d.). *Elastic Net*. Corporate Finance Institute. <https://corporatefinanceinstitute.com/resources/data-science/elastic-net/>

Chen, H. M. (2015). Do online recommendations matter? - A multimodal investigation of Amazon’s co-purchase network. *Journal of Digital Information Management*, 13(3), 176–184.

<https://sydney.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_gale_infotracacademiconefile_A428095490&context=PC&vid=61USYD_INST:sydney&lang=en&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,Amazon%27s%20e-commerce%20recommendation&offset=0>

Courtemanche, C., & Carden, A. (2014). Competing with Costco and Sam’s Club: Warehouse Club Entry and Grocery Prices. *Southern Economic Journal*, 80(3), 565–585.

<https://doi.org/10.4284/0038-4038-2012.135>

Neff, J. (2014). WALMART GETS DATA WITHOUT DISCOUNTS; Consumers scan receipts in retailer’s digital price-match loyalty program*. Advertising Age*, 85(14), 4–2.

<https://sydney.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_proquest_reports_1535267840&context=PC&vid=61USYD_INST:sydney&lang=en&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,Walmart%20Everyday%20Low%20Prices&offset=0>

Thompson, C. J., & Arsel, Z. (2004). The Starbucks Brandscape and Consumers’ (Anticorporate) Experiences of Glocalization. The Journal of Consumer Research, 31(3), 631–642.

<https://doi.org/10.1086/425098>

Zdaniuk, B. (2014). Ordinary Least-Squares (OLS) Model. *Encyclopedia of Quality of Life and Well-Being Research*, 4515–4517.

<https://doi.org/10.1007/978-94-007-0753-5_2008>