# Introduction

Name: Ba Huy Hoang Le

Student ID: s 224309594

This report aims to leverage different predictive models in different business scenarios, including retail and industrial production. This report is compiled and reserved for Method9. The following items will be discussed throughout this report:

1. Predict sales revenue of a GrocceryPlus store
2. Evaluate the interaction effect of promotional campaigns on the relationship between advertising expenditure and bike sales at BikeMart
3. Predict the likelihood of buying headphones at Gadget4U
4. Predict the resignation likelihood of a manager at CosmeticChain
5. Forecasting pale ale production volume for MoonlightAle

The report will also provide insights and managerial recommendations based on the analysis results for each section.

# Modelling Sales of a GrocceryPlus Store

A multiple linear regression model was developed to estimate the Sales (dependent variable) of a GroceryPlus store because ‘Sales’ is a continuous numerical variable.

## Pre-modelling analysis

The following analysis was conducted before developing the sales predictive model:

* Scatter plots confirm linear relationships between all independent and dependent variables.
* The correlation matrix showed 'Wages' (0.81), 'No. Staff' (0.74), 'Advertising' (0.84), and 'Car Spaces' (0.57) had moderate to significant positive relationships with Sales.

## Model Iteration

The initial model includes the four variables above. However, No. Staff and Car Spaces are statistically insignificant (p-values from all iterations> 0.05) and excluded from the final model.

## Model result

The final model’s equation

Estimated Sales= 1.74 + 2.44\*Wages + 0.028\*Adv

* Intercept = 1.74: On average, total sales revenue without any expenditure on wages and advertising is $1.74 million.
* Wages coefficient = 2.44: For every one million-dollar spent on wages, assuming no change in advertising and promotions expenditures, a supermarket's total sales revenue would grow by an average of $2.44 million.
* Advertising coefficient = 0.028: For every 1000-dollar increase in advertising and promotion expenditure, assuming no change in wages, total sales revenue would increase by an average of $28,000.

## Model Evaluation

R2 = 0.764: The model explains 76.4% of the variation of the independent variables, meaning it has a strong predictive power.

The standard error (1.74) is lower than the standard deviation (3.57), showing that the model provides relatively precise predictions compared to the natural variability of the sales revenue data.

## Recommendations

The report recommends:

1. Offer performance-based bonuses and salary increases to drive productivity and staff motivation to boost sales
2. Increase marketing budget and invest in advertising and provide promotion programs to attract customers

# Examining the Effect of Promotional Campaign and Advertising Expenses on Sales at BikeMart

## Model Development

A multilinear regression model was deployed with Advertising (x1), Promotions (x2), and Interaction (x1\*x2) as predictors. The model can explain 90.4% of the variation in total bike sales (R2 = 0.904). The high R2 value indicates that the model has a strong predictive power, confirmed by the ANOVA test with p-value <0.05. All independent variables (including the interaction term) are individually significant at p < 0.05.

## Model interpretation

Model Equation:

Estimated Bike Sales $ = 34466.61 +2.52\*Advertising 812.98\*Promotions + 2.38\*Interaction

Sales without advertising and promotions are $34,466.61. For every dollar spent on advertising, sales rise by $2.52, ceteris paribus. However, each additional promotional campaign reduces sales by $812.98, ceteris paribus. The interaction term, with a coefficient of 2.38, indicates that combining advertising and promotions increases sales by $2.38 per unit.

## Interaction interpretation

Figure 1 shows that when advertising spending is low, promotions can still drive some rise in bike sales, but their influence is limited. While promotions can increase sales, their effectiveness is limited without sufficient advertising spending and, hence, may not maximise revenue potential.

In contrast, high advertising spending leads to a significant increase in sales. The combination of high promotions and advertising leads to the largest sales increase. Therefore, a balanced marketing strategy that includes advertising and promotions is required.

A graph with a line and a point

Description automatically generated with medium confidence

Figure 1. Interaction of Promotion on the Relationship between Advertising and Bike Sales

# Modelling the Likelihood of Whether a Customer Will Buy Headphones After Purchasing a Mobile Phone at Gadget4U

This report deployed a logistic regression model to predict phone purchase likelihood at Gadget4U.

## Pre-modelling Analysis

Initially, the model comprises Annual Income (-0.17), Previous Purchases (0.21), Days Since Last Purchase (0.10), and Clicked on Ad (0.17). Although these independent variables have relatively low correlation coefficients, they intuitively influence customer purchasing behaviour.

## Model Development

Clicked on Ad (p-value = 0.08) and Days Since Last Purchase (p-value = 0.08) were removed from the final model because they were not statistically significant. The final model includes Annual Income (p-value = 0.03) -and Previous Purchase (p-value = 0.01).

## Model Results and Interpretation

Model equation:

Logit(p) = -1.13 - 0.01\*Annual Income + 0.22\*Previous Purchase

* Annual Income coefficient = -0.01: for every $1000 increase in annual income, the likelihood of buying headphones decreases by 0.86%, ceteris paribus
* Previous Purchases coefficient = 0.22: For each additional item purchased in the past year, the likelihood of purchasing headphones increases by 25%, ceteris paribus.

## Model Evaluation

Pseudo-R2 values are low. The Headphone Purchase model only explains 6.3-10.6% of the variation.

The model accurately classifies 72% of those who survey respondents. While the model is significantly better than a random process in classifying observation, compared with the PCC hit ratio of 59.1%, it falls short of the 73.9% standard hit ratio, which is not significantly better than random classification.

The AUC of 0.676 indicates the model’s moderate to weak predictive ability to discriminate between success (buy headphones) and failure (not buy headphones).

## Recommendation

The report does not recommend using this model for future prediction. Further improvements can be made:

1. Increase the size of data collection.
2. Research and record more relevant variables with a stronger correlation with the dependent variable for a more reliable and optimal predictive model.
3. Investigate the Annual Income variable. Although its negative relationship with bike sales is expected (based on the negative correlation coefficient), the result is counterintuitive. One implication is that when customer incomes increase, they tend to gravitate towards more expensive or high-end audio products, which could result in less demand for the headphones in this analysis.

# Modelling the Likelihood of a Store Manager Resigning at CosmeticChain

A logistic regression model was developed to predict the likelihood of resignation at CosmeticChain.

## Store Manager Resigning Model

### Model Results

Logit(p) =0.42 - 0.13\*Age\_Manager + 0.38\*Experience\_Manager + 0.97\*Gender\_Manager\_Code

* Age\_manager coefficient = -0.13: Given the same experience and gender, being one year older decreases the likelihood of resigning by 12.1%.
* Experience\_Manager coefficient = 0.38: For every additional year of work experience, the likelihood of resigning increases by 45.7%, ceteris paribus.
* Gender\_Manager\_Code coefficient = 0.97: A male manager is more likely to resign than a female manager with the same age and work experience by 63.7%

### Model Evaluation

The model explains 33-46% of the variation in the dependent variables. Since pseudo-R2 values are relatively low, more relevant variables can be included to increase predictive power.

The model accurately classifies 76.7% of resignation cases. The model is practically significant, i.e., it is better than a random process in classifying observation, considering the PCC hit ratio (51.5%) and standard hit ratio (64.4%).

The AUC of 0.86 means the model fits the data well and is relatively strong at discriminating between success (resign) and failure (not resign).

Therefore, the model is suitable for future prediction usage, with consideration of more relevant predictors to improve predictive outcomes.

## Predicted Probability of Resigning

## Visualisation

Figure 2 shows that the likelihood that a male manager will leave their position is higher than that of a female manager, regardless of age or experience level. Both genders' resignation rates rise with experience, but male managers' rates rise more quickly. Compared to older managers, younger managers have a higher resignation rate. The likelihood of resigning increases with experience for both genders, but younger managers tend to reach this point earlier in their careers.

A graph of a graph showing the number of probabilities

Description automatically generated with medium confidence

Figure 2. Predicted probability of resignation for managers aged 34 to 36 at CosmeticChain

## Recommendations

This report recommends the following solutions to retain talents and reduce the chance of resigning:

* Incentives for younger managers through mentorship programs, career development opportunities, and clear promotion pathways.
* Initiatives for experienced managers through performance-based bonuses, recognition programs and leadership training
* Promote work-life balance and inclusive workplace

# Forecasting Pale Ale Production in the Upcoming Four Quarters for MoonlightAle

## General Observation and Model Selection

From 2014 to 2024, pale ale production has been on the upward trend. Production typically peaks in Q3 and dips in Q4 before rising again following Q1 and Q2. Given this upward trend, season variation, and cyclical pattern that typically occurs yearly, this report deploys a Multiplicative Time-series model to predict pale ale production.

## Handling data

The report conducts the following steps before making the forecast for pale ale production:

* Smooth the production figure using the 4-quarter centered Moving Average
* Generate deseasonalised pale ale production to the long-term upward trends by removing seasonal fluctuations
* Calculate the season index to capture seasonal variations across four quarters:
  + Q1: 0.965
  + Q2: 0.970
  + Q3: 1.183 (confirming the peak production volume for Q3)
  + Q4: 0.882 (confirming the dip in production volume for Q4)

## Time-series Model and Evaluation

Linear trend equation:

y = 20.512x + 1317

The model fits the data well, thanks to its high R2 value of 0.9027. The MAPE from this model is 3.66%, which is relatively low. On average, forecasted values deviate from observed production volume by just 3.66%. Therefore, this model is suitable for future forecasting purposes.

## Forecast Results

Based on the deseasonalised data and trend analysis, the forecast production volumes (litres) of pale ale for the upcoming four fiscal quarters are as follows:

* Q2, 2024: 2073.53
* Q3, 2024: 2553.60
* Q4, 2024: 1921.03
* Q1, 2025: 2121.57

# Conclusion

The findings in this report

1. Spending on wages and advertising significantly affects sales; investment in these variables will significantly boost sales. Further investigation of other variables is required for a more reliable model.
2. Promotional campaigns affect advertising spending and sales. Balanced investment in both variables is important to boost bike sales.
3. The logistic model developed for Gadget4U is not recommended. Further investigation on Annual income is recommended.
4. Male, more work-experienced, and younger managers are more likely to resign at CosmeticChain. Several incentives, from salary to training, are proposed to retain talent.
5. The time-series forecasting model has strong predictive power, and MoonlightAle can use it for future predictive purposes with close monitoring.

Limitations in this report:

1. Predictors have relatively high VIF values, but they are still included due to the lack of relevant predictors for the sales estimation model for a GrocceryPlus store.
2. Lack of relevant and strongly correlated for the classification model to predict purchase behaviour at Gadget4U.
3. Overall, the data provided is small, which can affect decision-making in the long term.