Sales forecasting for the European drug store Rossman

Table of contents

1 Introduction

Predicting sales is a vital part for any business across all sectors, from manufacturing, retail, logistics, to wholesale. However, this is one of the most difficult tasks a business can undertake due to the complexities involved. Sales are driven by a great deal of different factors such as the store location, proximity to competition, macro scales of yearly seasonality, to the micro scales of the time of day and the day of the week, whether there is a promotion or what the weather is doing (Hasan, 2024). All of these things influence sales in different ways, so as you can see, this makes forecasting sales the ultimate challenge for a business.

1.1 Historic trend in sales

As a company Rossmann, a part of the A S Watson group, is the market leader for health and beauty retail in Germany with around 100 stores. It also has over 4,500 stores across Europe, from Poland, Turkey to Spain, employing over 60,000 people (Group, 2024). We have been asked to

2 Methodology

2.0.1 Store data cleaning

Have a look at the structures of the stores data
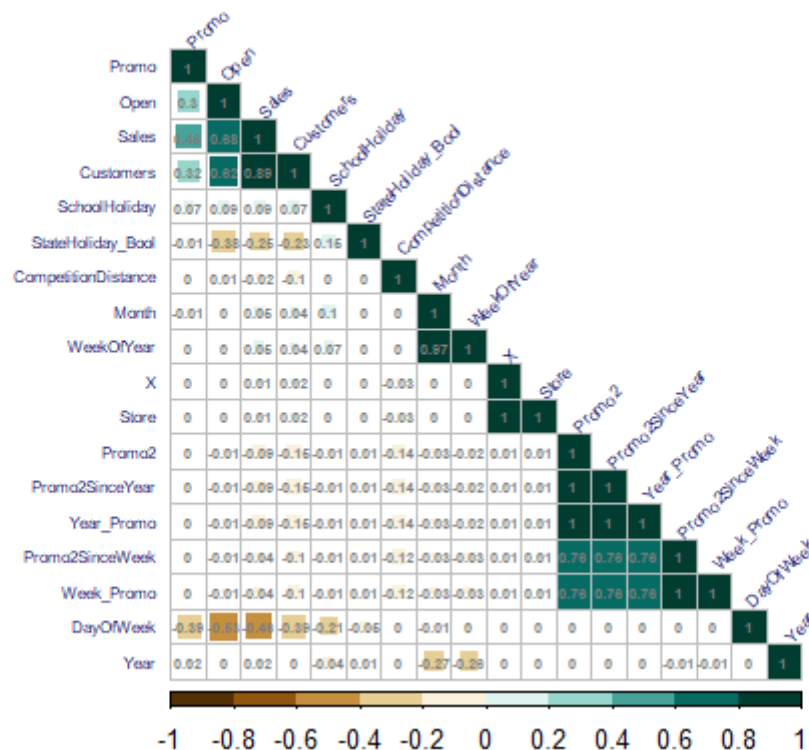
Describe the steps used to

Figure 1: Correlation plot show collinearity

Table 1: Variance Inflation Factors for the linear model

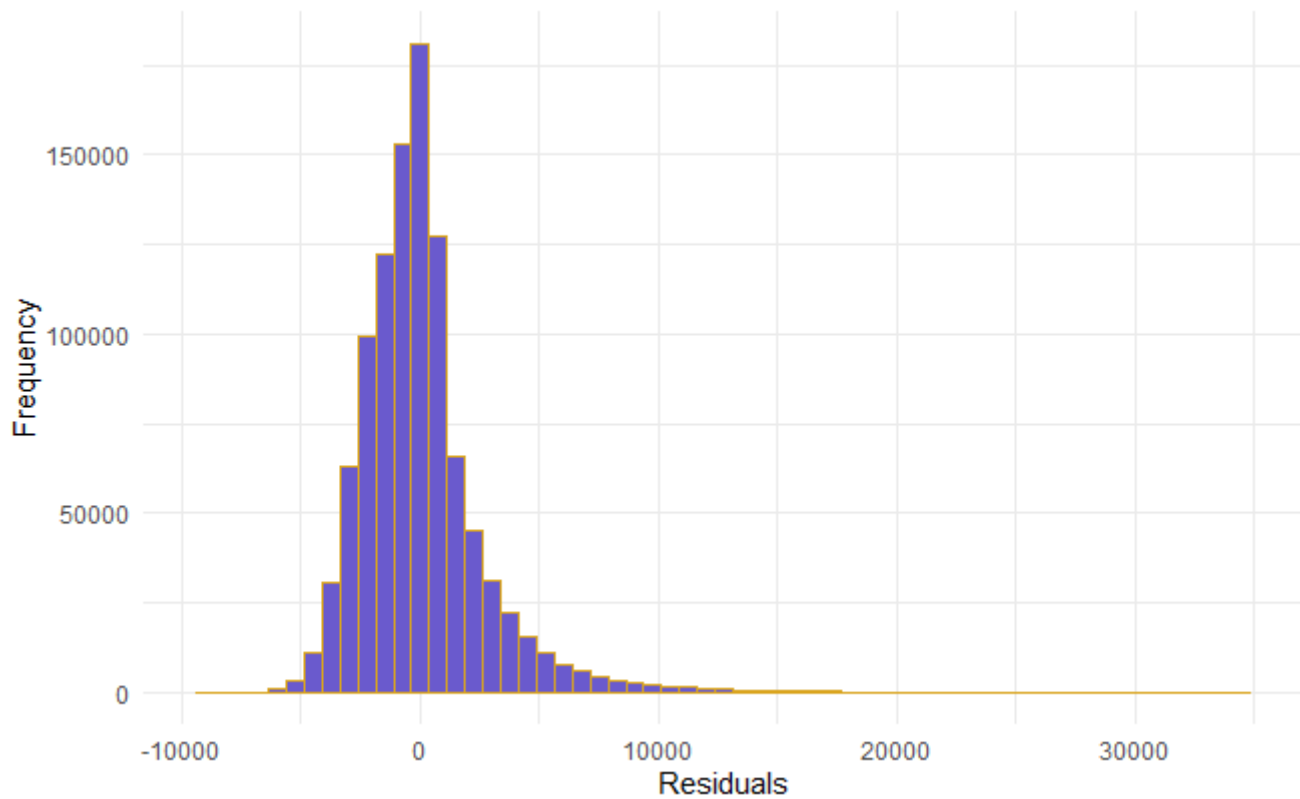|	GVIF	Df	GVIF^(1/(2*Df))
DayOfWeek	1.677217	1	1.295074
Open	1.834113	1	1.354294
Promo	1.201859	1	1.096293
StateHoliday_Bool	1.304303	1	1.142061
Year	1.079366	1	1.038925
Month	15.586181	1	3.947934
StoreType	2.310017	3	1.149748
Assortment	2.256575	2	1.225639
CompetitionDistance	1.069460	1	1.034147
Promo2	1.039179	1	1.019402
WeekOfYear	15.472296	1	3.933484
Store	1.006353	1	1.003172

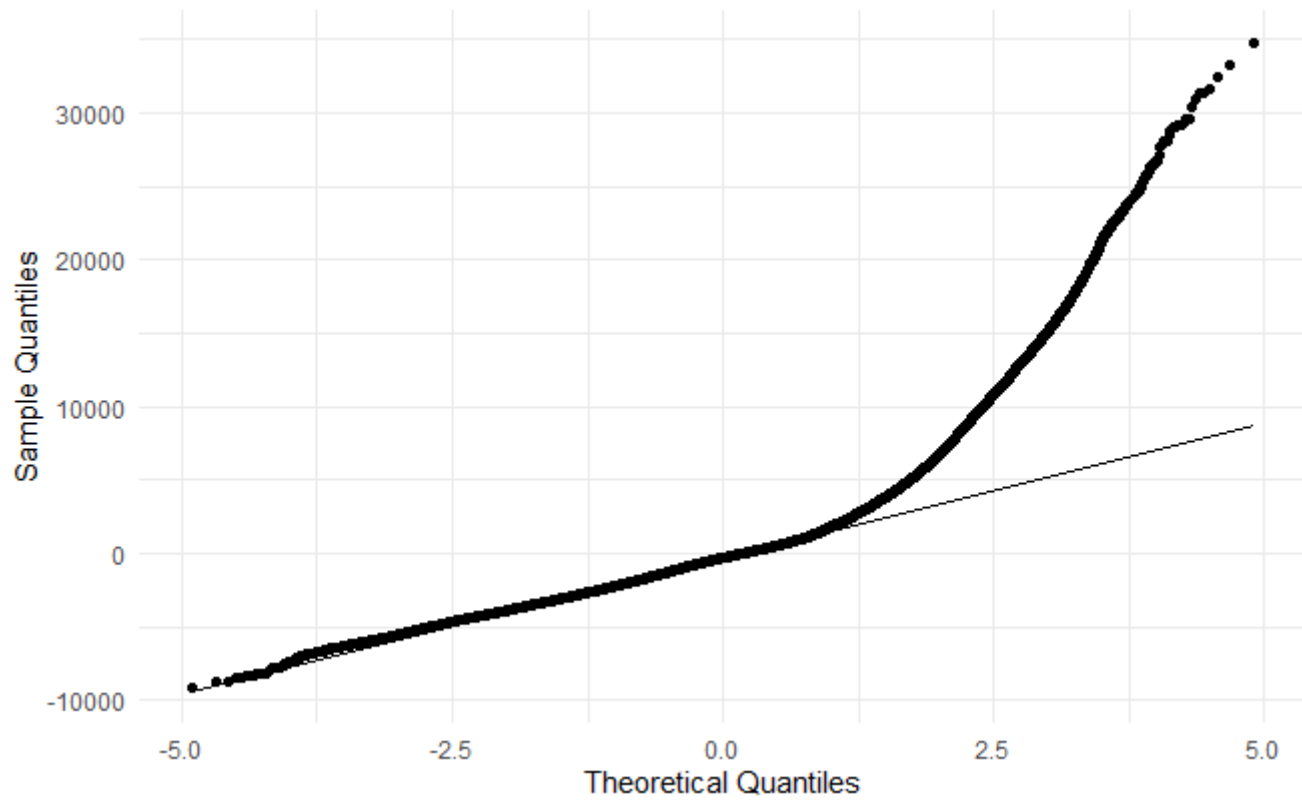Figure 2: Plot of residuals for the linear model

Figure 3: QQ Plot of Residuals
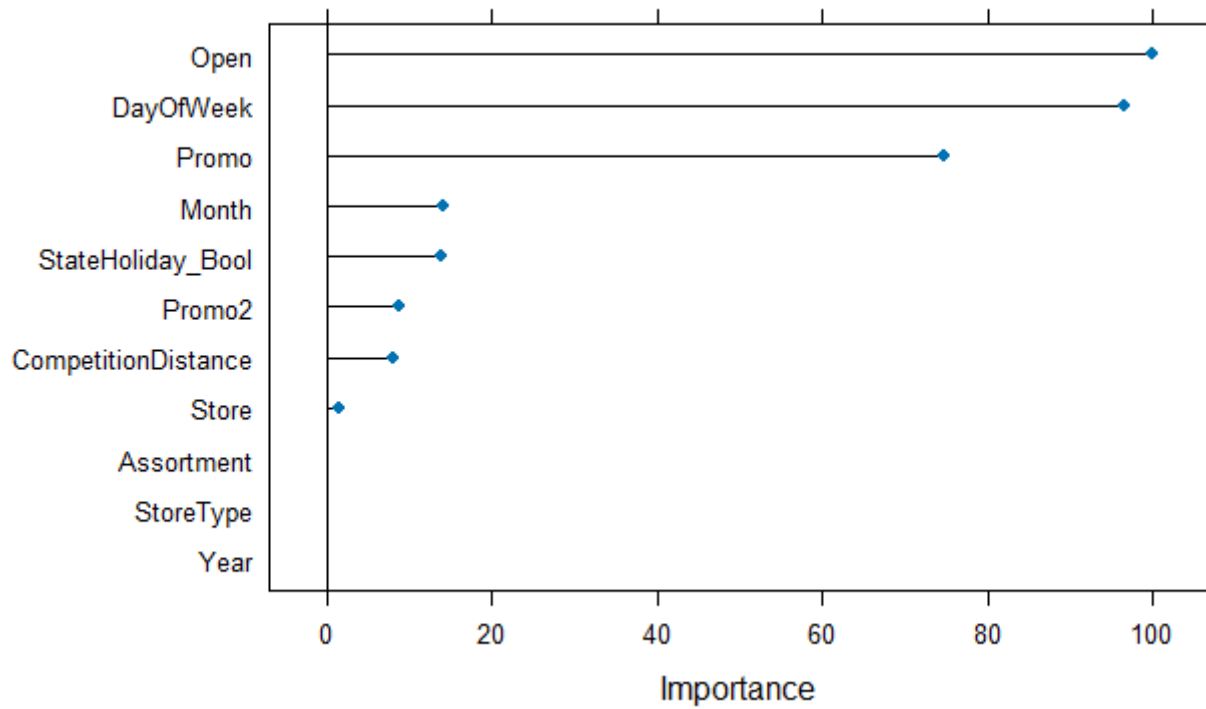
2.0.2 Decision Tree

Figure 4: Importance of each variable to the model

```
RMSPE: 55.31 %
```

This is not such a good RMSPE as it is greater than 0.5, or 50%. General rule-of-thumb is that a good RMSPE is between 0.2-0.4.
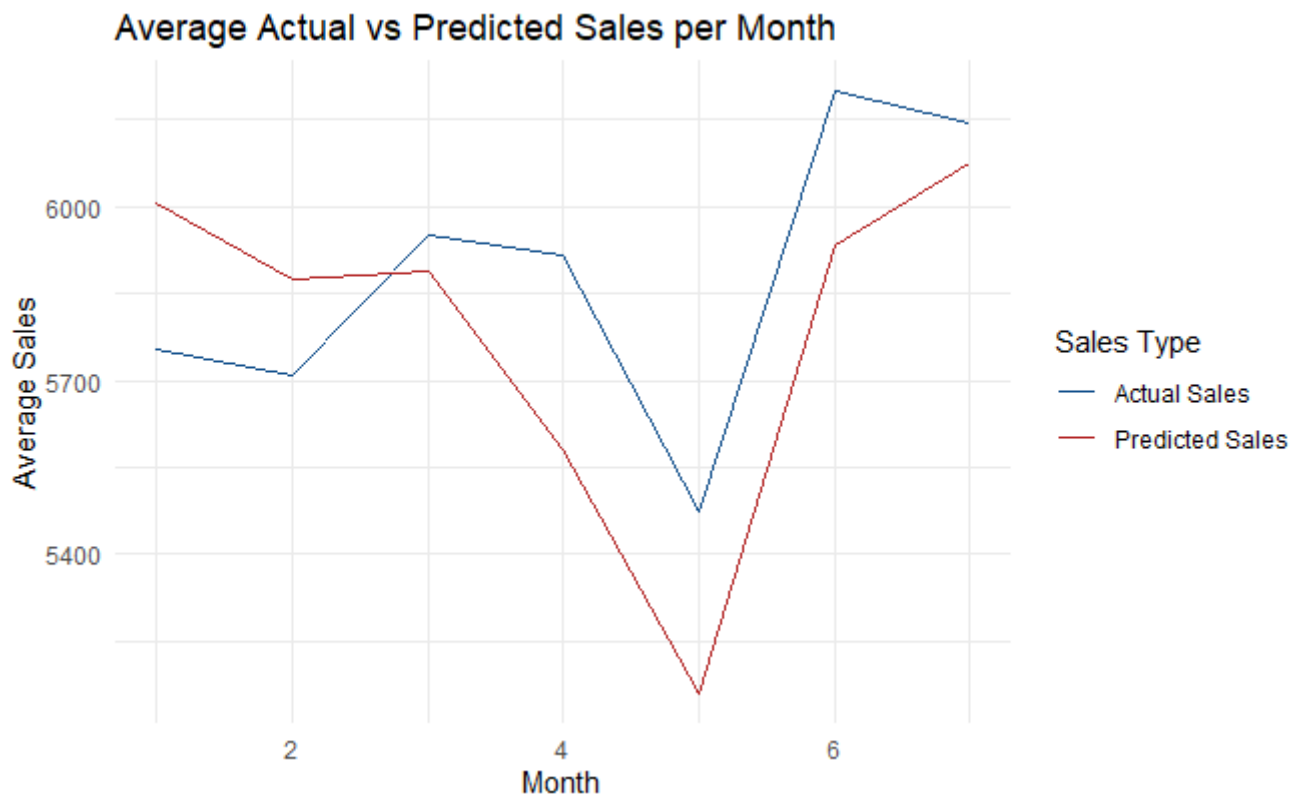
Figure 5: Average Actual vs Predicted Sales per Month

2.0.3 Random Forest

2.0.4 XGBoost

2.0.5 USE XGBOOST MODEL TO PREDICT SALES

3 Plotting the final sales results

Review the available data and describe it in terms of its variables, quality, and relevance to the sales forecasting

Link data sets together as appropriate

Pre-process the data as appropriate for further analytics, for example, you may want to encode any categorical data, create new variables, identify how many missing values there are and deal with them appropriately, etc.

Identify the key factors affecting sales, for example, you may want to check whether competition and promotions have an impact on sales, and how public holidays cause sales fluctuations.

Build a forecasting model (which can be a linear regression model, a neural network model or something else) using the variables you identified. Please make sure to justify the choice of your modelling approach.

Use the Root Mean Square Percentage Error (RMSPE) to forecast accuracy

4 Results

Interpret key results, assumptions and limitations of your analysis.

5 Conclusion

5.1 Limitations

5.2 Implications

5.3 Recommendations

References

Group, A.W. 2024. Rossmann. *AS Watson Group - A member of CK Hutchison Holdings*. [Online]. Available from: https://www.aswatson.com/our-brands/health-beauty/rossmann/.

Hasan, M.R. 2024. Addressing seasonality and trend detection in predictive sales forecasting: A machine learning perspective. *Journal of Business and Management Studies*. **6**(22), pp.100–109.