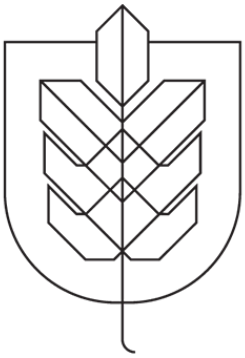


HULT
INTERNATIONAL
BUSINESS SCHOOL

FORECASTING AND PREDICTING THE FUTURE USING DATA

A3: BUSINESS INSIGHT REPORT

Professor: Thomas Kurnicki
Author: Gabriel Martinelli

Abstract

This report analyses a dataset emulating the sales of an e-commerce store managed by a popular American influencer over the course of a year. The dataset includes various variables, ranging from the buyer's age and gender to shipping expenses charges to the customer and the ratings received by the site.

The analysis develops along three main lines: predictive, qualitative, and forecasting. Predictive models were built to identify which drivers led consumers to leave a 5-star review or lower. However, it emerged that none of the available variables proved significant in predicting the business outcome.

This result highlights that the collected variables are not at all adequate to predict what rating the consumer might give. Consequently, if the available variables are inconclusive, it becomes necessary to investigate and understand which might be relevant.

The qualitative analysis therefore aimed to define which variables future models should focus on. From an analysis of a random sample of 50 reviews rated 1 or 2 stars, and a sample of 50 5-star reviews, it emerged that quality, design, and delivery times are the factors that most affect customer satisfaction.

To complete the dataset exploration, forecasting models were built to estimate future growth rates of overall sales and sales by category. Once again, the available data proved insignificant. No seasonality was detected, and the confidence intervals were wide and cantered around zero.

Despite the limited relevance of the data, the analysis still generated valuable business insights to be considered in future planning. Data always carries a message, and even when results differ from expectations, an in-depth analysis will always prove useful.

Table of Contents

Introduction

This paper presents an analysis of a dataset covering the global merchandise sales of a well-known American influencer from November 2023 to November 2024.

The main focus of the analysis is to predict, based on order characteristics, when a buyer is likely to leave a 5-star review, with the goal of identifying the most impactful aspects and enhancing customer satisfaction.

At the same time, forecasting models were developed to estimate future sales growth, both overall and for each product category.

The analysis faced major setbacks related to the significance of the available data, which proved unable to capture either the drivers of customer satisfaction or the volatility of sales. Nevertheless, the study still provides insights into how the challenges encountered by the predictive and forecasting models can be addressed by fundamentally redefining the type of data that should be collected.

Predictive Analysis

Before starting any analysis, cleaning the data is essential. For the dataset examined, the cleaning process focused on removing empty rows or those lacking an order date, eliminating duplicates based on the order ID, and converting categorical variables (Product Category, Buyer Gender, and International Shipping) into numerical variables. Converting these variables was necessary to normalize them alongside the others, making them comparable in subsequent analyses.

Once the cleaning was completed, the business outcome was defined: a successful outcome corresponded to an order that received a 5-star rating. Several predictive models were then built to determine whether, based on order characteristics, it was possible to predict the ratings consumers would give.

The first of these models was a Gini decision tree (figure 1) , designed to uncover splits that separate 5-star ratings from others based on variables such as shipping charges, sales price, and buyer demographics. However, to visualize the branches, the cp parameter had to be set extremely low, a clear sign that these variables could not meaningfully discriminate the target.

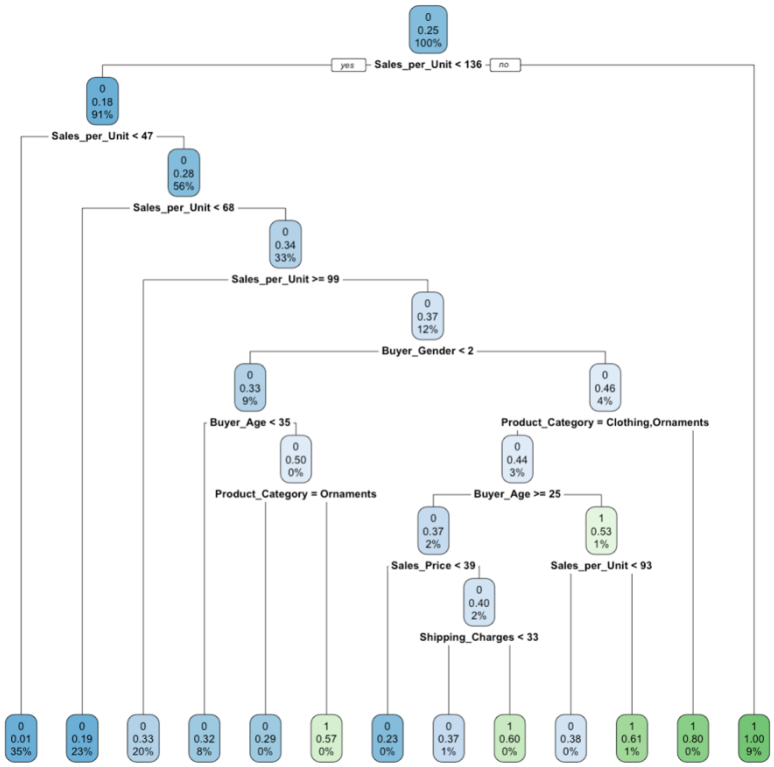


Figure 1: Gini Decision Tree, source: own elaboration

This inability was confirmed by the logistic regression (figure 2) fitted afterwards, which revealed that none of the variables considered in the model were significant. It is enough to note that the lowest p-value was 0.360, well above the typical threshold for significance.

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)            -1.1267988  0.2566997  -4.390 1.14e-05 ***
Product_Category        0.0504113  0.0701931   0.718    0.473
Buyer_Gender            0.0568188  0.0620955   0.915    0.360
Buyer_Age               0.0006603  0.0054504   0.121    0.904
International_Shipping -0.0081101  0.1337385  -0.061    0.952
Sales_Price             0.0009332  0.0013697   0.681    0.496
Shipping_Charges        0.0015383  0.0024672   0.623    0.533
Quantity                0.0174753  0.0260965   0.670    0.503
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Logistic Regression, source: own elaboration

To investigate more deeply the interactions between variables and the target, a random forest (Figure 3) was then constructed. This approach undeniably reduced the error compared to the single decision tree, but it also quickly plateaued (around 70-80 trees) and settled at approximately 0.215, which remains relatively high.

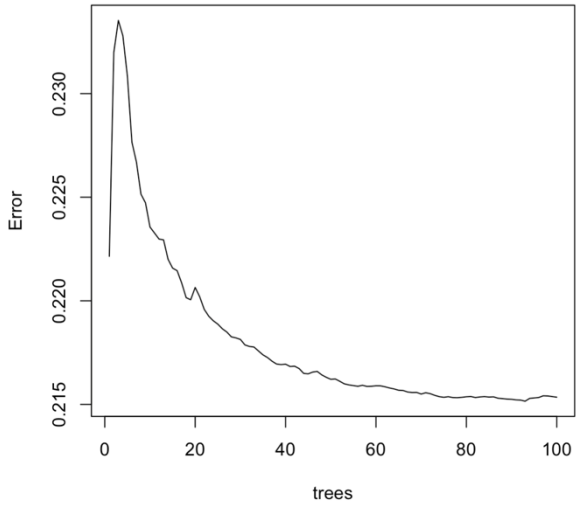


Figure 3: Random Forest, source: own elaboration

The confusion matrix (Figure 4) further confirmed the models' inability to successfully predict the target, showing an accuracy of 70.25%, which is actually lower than simply always predicting the dominant class (71.47%).

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1025  408
         1   32   14

               Accuracy : 0.7025
                 95% CI : (0.6785, 0.7257)
    No Information Rate : 0.7147
    P-Value [Acc > NIR] : 0.8565
```

Figure 4: Confusion Matrix, source: own elaboration

Given the models' failure to predict customer ratings, it became necessary to understand which variables should instead be considered. For this purpose, a qualitative analysis was performed on two samples of 50 customer reviews each: the first containing 1-2 star reviews, and the second containing only 5-star reviews. This step was crucial to identify which variables had been overlooked during data collection. The blend of predictive and qualitative analysis thus provided a deeper understanding of what drives customer satisfaction, and what does not, and consequently, where future models should focus.

Predictive Analysis Insights

The combination of the predictive analysis and qualitative analysis certainly provided important insights.

The models clearly demonstrated how the variables collected in the dataset, such as buyer age, shipping charges, product category, and sales price, were irrelevant at predicting 5-star reviews. The decision tree struggled to create meaningful branches, the logistic regression revealed that none of the variables had a significant p-value, and while the random forest did reduce error, it did so insufficiently and quickly reached a plateau.

However, these initially discouraging results prompted a deeper qualitative analysis. By examining two samples, one containing negative reviews and the other positive, it became possible to identify which variables truly mattered for customer satisfaction.

In the sample containing the negative reviews, 32% of customers complained about delivery, 16% about product quality, and another 16% about design. On the other hand, in the sample containing the positive reviews, 32% were satisfied with quality, 26% with delivery, and 16% with design.

This suggests that future data collection should focus on variables like perceived quality and delivery performance to better train models to predict the target outcome.

This first phase of analysis teaches us that sometimes understanding what does not drive outcomes is just as valuable. Indeed, knowing what fails to predict satisfaction provides a clear guide for where to redirect further analysis.

Forecast Analysis

The forecasting analysis focused on estimating future sales trends by applying ARIMA models to the available data.

The first step was to aggregate monthly sales to create time series, which were then examined through ACF and PACF plots (figures 5 and 6).

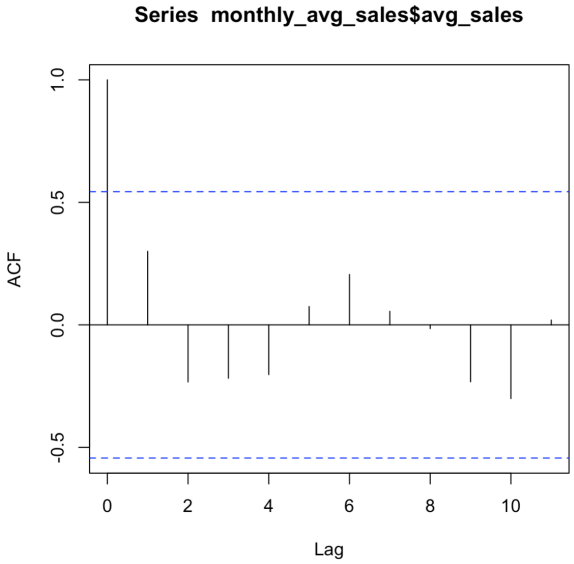


Figure 5: ACF Plot, source: own elaboration

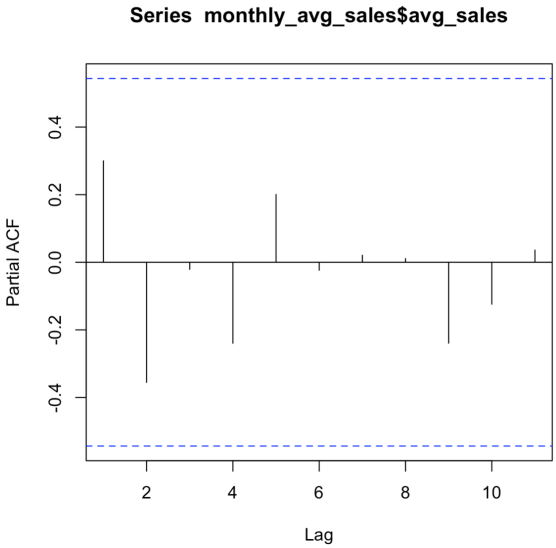


Figure 6: PACF Plot, source: own elaboration

These plots showed weak autocorrelation and an almost complete absence of seasonal patterns, indicating that sales fluctuated randomly around their mean.

As a result, the auto.arima function employed selected an ARIMA(0,0,0), effectively treating the series as white noise. The resulting forecast (Figure 7) thus settled around the historical mean, with extremely wide confidence intervals.

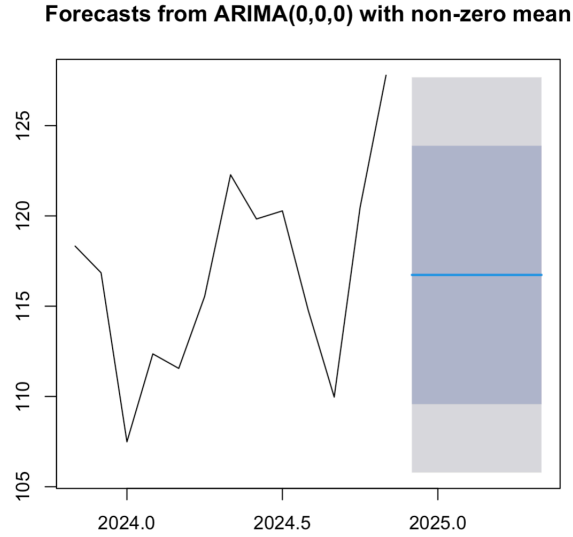
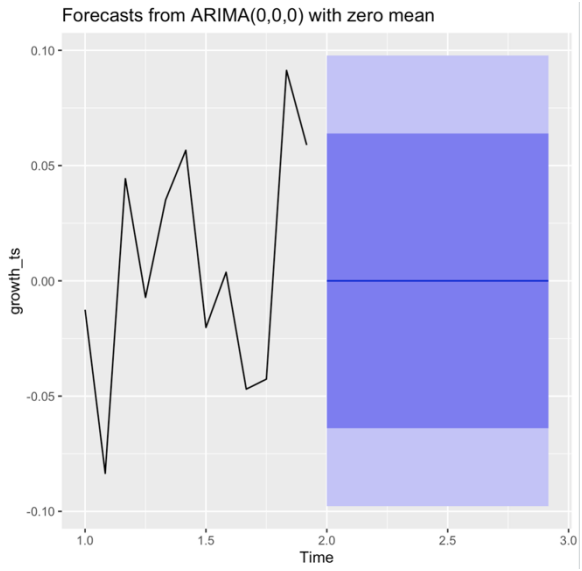


Figure 7: Sales Forecast, source: own elaboration

Quoting Professor Thomas Kurnicki from one of his lectures in the "Forecasting and Predicting the Future using Data" course at HULT International Business School: "*If your boss asks you to forecast sales, you should reply that, after statistical adjustments, you can forecast the growth rates of sales*."

Consequently, the next phase of the analysis focused on forecasting sales growth rates (Figure 8).



However, once again, auto.arima picked the (0,0,0) specification, forecasting future growth at zero, with wide intervals that indicated an equal likelihood of positive or negative shifts.

Figure 8: Sales' Growth Rate Forecast, source: own elaboration

The analysis then turned to forecasting both sales and their growth rates for each individual product category. Yet the pattern remained unchanged. None of the categories exhibited significant autocorrelation or seasonality (Figures 9,10 and 11).

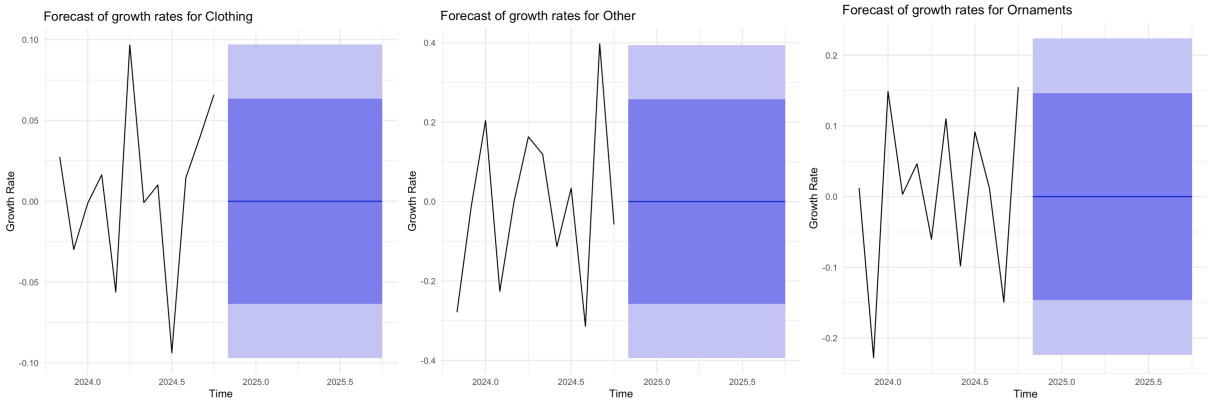


Figure 9, 10 and 11: Clothing, Other, Ornaments Growth Rates Forecast, source: own elaboration

Overall, the forecasting analysis highlighted the volatility and unpredictability of sales. Much like the predictive analysis, this suggested the need to reconsider both the approach and the type of data being collected to enable more meaningful forecasting in the future.

Forecast Analysis Insights

The forecasting analysis highlighted several important aspects regarding the sales dynamics captured by the dataset.

The forecasts consistently settled around the historical mean with wide confidence intervals, indicating pronounced volatility where future performance is equally likely to fluctuate above or below this average. The near-total absence of trends, seasonality, and autocorrelation, both in overall sales and in the sales of each individual category, suggests not only that the time series is too short but also that key variables may have been overlooked.

The same interpretation applies to the forecasts of growth rates, leaving the e-commerce business exposed to uncertainty. However, despite these discouraging findings, two key insights stand out.

The first is that to gain a clearer picture of the e-commerce trajectory, more time is needed, implying a longer observation period to build a sufficiently rich time series. The second is that, given this is an e-commerce operation selling an influencer's merchandise, it is misleading to estimate its performance without taking into account the influencer's social media activity.

These forecasts, when analysed alone, lose much of their meaning if not compared with the social campaigns being run. In fact, linking sales trends to the number of promotional posts could reveal patterns that would otherwise remain invisible when only considering the transactional data contained in this dataset.

Conclusion

In this paper, the combination of predictive, qualitative, and forecasting analyses was used to examine the annual sales performance of an e-commerce business selling merchandise for an American influencer.

All predictive models consistently demonstrated that the variables present in the dataset were unable to predict customer satisfaction, represented by the ratings customers gave after placing their orders. Indeed, the p-values for these variables were all insignificant, and the confusion matrix showed that the model could not even outperform naive guesses.

The forecasting analysis further revealed the absence of trends, seasonality, or autocorrelation. Forecast models repeatedly settled around historical means with wide confidence intervals, highlighting the volatility of the business and their inability to produce accurate estimates given the available data.

However, the qualitative analysis of customer reviews pointed to critical aspects that the structured data missed: quality, design, and delivery emerged as the real drivers of customer satisfaction. These factors should therefore be prioritized in future data collection and incorporated into upcoming predictive and forecasting models, as they have the potential to uncover hidden patterns.

Although the existing data proved incapable of predicting the business outcome, it still revealed a key message. It is essential to reshape the data collection strategy by integrating insights on sentiment, product quality and delivery satisfaction, and the influencer's social activity. Doing so will enable the building of more accurate models in the future.

Appendix

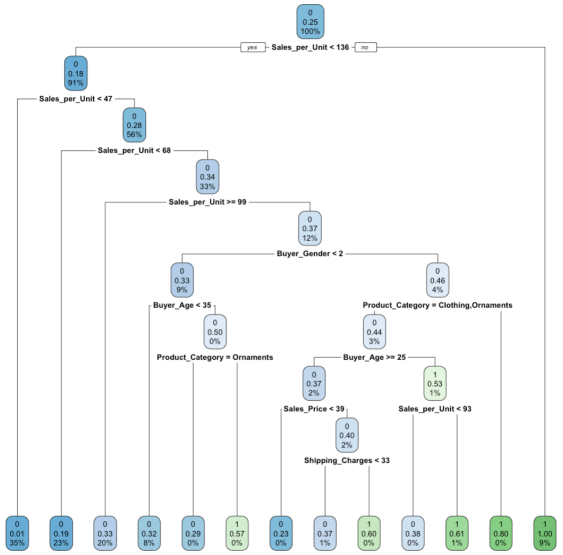
```r
library(readr)
library(dplyr)
library(lubridate)
library(stringr)
library(rpart)
library(rpart.plot)
library(randomForest)
library(caret)
library(dplyr)
library(caret)
library(forecast)
library(ggplot2)
library(stringr)
library(tseries)
library(rugarch)

#Importing the dataset
merch_sales <- read_csv("/Users/gab/Desktop/FORECASTING AND PREDICTING THE FUTURE USING DATA/INDIVIDUAL ASSIGNMENT/merch_sales.c

##########################################
###########CLEANING/WRANGLING##############
##########################################

#removing rows that have no order date
rows_before <- nrow(merch_sales)
merch_sales_clean <- merch_sales %>%
  filter(!is.na("Order Date"))
rows_after <- nrow(merch_sales_clean)
merch_sales <- merch_sales_clean

#checking NA
anyNA(merch_sales)
#changing Order Date to date obj
merch_sales$`Order Date` <- as.Date(merch_sales$`Order Date`)
#removing duplicates rows
merch_sales <- merch_sales[!duplicated(merch_sales$`Order ID`), ]

#changing chr columns to num in order to normalise and compare them later
merch_sales <- merch_sales %>%
  mutate(`Product Category` = case_when(
    `Product Category` == "Clothing" ~ 1,
    `Product Category` == "Ornaments" ~ 2,
    `Product Category` == "Other" ~ 3
  ))

merch_sales <- merch_sales %>%
  mutate(`Buyer Gender` = case_when(
    `Buyer Gender` == "Male" ~ 1,
    `Buyer Gender` == "Female" ~ 2,
```

```r
  ))

merch_sales <- merch_sales %>%
  mutate(`International Shipping` = case_when(
    `International Shipping` == "No" ~ 0,
    `International Shipping` == "Yes" ~ 1,
  ))

#removing the spaces and changing them with "_"
names(merch_sales) <- str_replace_all(names(merch_sales), " ", "_")

#normalise the variables to compare them

rescale <- function(x){
  min_max <- (x-min(x))/(max(x)-min(x))
  return(min_max)
}

merch_sales$Product_Category_norm <- rescale(x=merch_sales$Product_Category)
merch_sales$Buyer_Gender_norm <- rescale(x=merch_sales$Buyer_Gender)
merch_sales$Buyer_Age_norm <- rescale(x=merch_sales$Buyer_Age)
merch_sales$Sales_Price_norm <- rescale(x=merch_sales$Sales_Price)
merch_sales$Shipping_Charges_norm <- rescale(x=merch_sales$Shipping_Charges)
merch_sales$Sales_per_Unit_norm <- rescale(x=merch_sales$Sales_per_Unit)
merch_sales$Quantity_norm <- rescale(x=merch_sales$Quantity)
merch_sales$Total_Sales_norm <- rescale(x=merch_sales$Total_Sales)

#########################################################
###########PREDICTIVE ANALYSIS ON RATINGS###############
#########################################################

#define business outcome
business_suc <- merch_sales[which(merch_sales$Rating == 5), ]
business_fail <- merch_sales[which(merch_sales$Rating < 5),]

merch_sales$businessoutcome <- c()
merch_sales[which(merch_sales$Rating == 5), c('businessoutcome')] <- 1
merch_sales[which(merch_sales$Rating < 5), c('businessoutcome')] <- 0

#define random sample
indx <- sample(x=1:nrow(merch_sales), size=0.8*nrow(merch_sales))

merch_sales_train <- merch_sales[ indx, ]
merch_sales_test <- merch_sales[-indx,]

#build the decision tree
my_tree <- rpart(businessoutcome~Product_Category + Buyer_Gender + Buyer_Age + International_Shipping +
                   Sales_Price + Shipping_Charges + Quantity,
                 data=merch_sales_train, method="class", cp=0.0007252)
```

```r
rpart.plot(my_tree)

#building a Logistic regression
my_logit_norm <- glm(businessoutcome~Product_Category + Buyer_Gender + Buyer_Age + International_Shipping +
                       Sales_Price + Shipping_Charges + Quantity,
                     data=merch_sales_train, family = "binomial")
summary(my_logit_norm)

#predicting the test data
tree_pred <- predict(my_tree, merch_sales_test)

#building a confusion matrix
confusionMatrix(data =  as.factor(as.numeric(tree_pred[,2]>0.5)),
                reference = as.factor(as.numeric(merch_sales_test$businessoutcome)))

#building a random forest
my_forest <- randomForest(
  x = merch_sales_train[, c("Product_Category" , "Buyer_Gender" , "Buyer_Age" , "International_Shipping" ,
                            "Sales_Price" , "Shipping_Charges" , "Quantity")],
  y = merch_sales_train$businessoutcome,
  ntree = 100)
forest_pred <- predict(my_forest, merch_sales_test)
plot(my_forest)
varImpPlot(my_forest)

#checking the reviews, is there any relationship between 1 or 2 ratings and delivery issues?

filtered_rows <- merch_sales[merch_sales$Rating >= 1 & merch_sales$Rating <= 2, ]
sample_50 <- filtered_rows[sample(1:nrow(filtered_rows), 50), ]
rating_review_table <- sample_50[, c("Rating", "Review")]
View(rating_review_table)

#this table created with a sample of 50 orders that received a rating between 1 and 2 highlights that
#in the 32% of the orders there where, according to the reviews, problems with the quality, in the 16% of the orders
#problems with the delivery and in another 16% problems with an underwhelming design.
#these insights could be the groud for further reasoning.

#checking the reviews, is there any relationship between 5 ratings and quality, design e delivery?

filtered_rows_5stars <- merch_sales[merch_sales$Rating == 5 , ]
sample_50_5stars <- filtered_rows_5stars[sample(1:nrow(filtered_rows_5stars), 50), ]
rating_5_review_table <- sample_50_5stars[, c("Rating", "Review")]
View(rating_5_review_table)

#this table created with a sample of 50 orders that received a 5-star rating highlights that
#in the 32% of the orders the customers were satisfied about the delivery, in the 26% about the quality and in the
#16% about the design of the product.

#########################################################
```

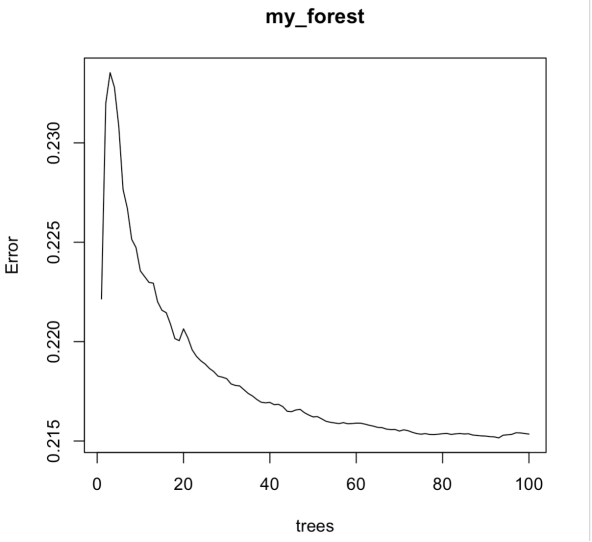
```r
###########PREDICTIVE ANALYSIS ON SALES#################
#########DO NOT INCLUDE IN THE ASSIGNMENT###############
#########################################################

summary(merch_sales$Total_Sales)

#define business outcome
business_suc_sales <- merch_sales[which(merch_sales$Total_Sales >= 137), ]
business_fail_sales <- merch_sales[which(merch_sales$Total_Sales < 137),]

merch_sales$businessoutcome_sales <- c()
merch_sales[which(merch_sales$Total_Sales >= 137), c('businessoutcome_sales')] <- 1
merch_sales[which(merch_sales$Total_Sales < 137), c('businessoutcome_sales')] <- 0

#define random sample
indx <- sample(x=1:nrow(merch_sales), size=0.8*nrow(merch_sales))

total_merch_sales_train <- merch_sales[ indx, ]
total_merch_sales_test <- merch_sales[-indx,]

#build the decision tree
my_tree_sales <- rpart(businessoutcome_sales~Product_Category + Buyer_Gender + Buyer_Age + International_Shipping +
                         Sales_Price + Shipping_Charges + Sales_per_Unit,
                       data=total_merch_sales_train, method="class", cp=0.001)
rpart.plot(my_tree_sales)

#building a Logistic regression
my_logit_sales <- glm(businessoutcome~Product_Category + Buyer_Gender + Buyer_Age + International_Shipping +
                        Sales_Price + Shipping_Charges + Sales_per_Unit,
                      data=total_merch_sales_train, family = "binomial")
summary(my_logit_sales)

#predicting the test data
tree_sales_pred <- predict(my_tree_sales, total_merch_sales_test)

#building a confusion matrix
confusionMatrix(data =  as.factor(as.numeric(tree_sales_pred[,2]>0.5)),
                reference = as.factor(as.numeric(total_merch_sales_test$businessoutcome_sales)))


#########################################################
#############FORECAST ANALYSIS ON SALES#################
#########################################################

converted_order_dates <- as.Date(merch_sales$Order_Date, format="%Y-%m-%d")
sum(is.na(converted_order_dates))
min(converted_order_dates)
max(converted_order_dates)

```

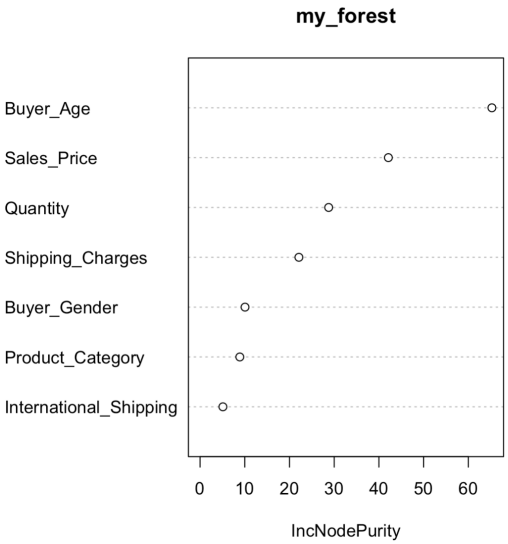
```r
#plot the average sales trend
merch_sales <- merch_sales %>%
  mutate(Month = floor_date(Order_Date, "month"))

monthly_avg_sales <- merch_sales %>%
  group_by(Month) %>%
  summarise(avg_sales = mean(Total_Sales, na.rm=TRUE))

ggplot(monthly_avg_sales, aes(x = Month, y = avg_sales)) +
  geom_line() +
  labs(title = "Monthly average sales trend",
       x = "Months",
       y = "Average sales per month") +
  theme_minimal()

adf.test(monthly_avg_sales$avg_sales)

acf(monthly_avg_sales$avg_sales)
pacf(monthly_avg_sales$avg_sales)

#ARIMA Model

sales_ts <- ts(monthly_avg_sales$avg_sales, start = c(2023, 11), frequency = 12)
arima_model <- auto.arima(sales_ts)
summary(arima_model)

forecast_arima <- forecast(arima_model, h = 6)
plot(forecast_arima)

#forecasting the growth rates
monthly_avg_sales_growth <- monthly_avg_sales %>%
  arrange(Month) %>%
  mutate(growth_rate = c(NA, diff(log(avg_sales))))

growth_ts <- ts(na.omit(monthly_avg_sales_growth$growth_rate), frequency=12)
growth_model <- auto.arima(growth_ts)
growth_forecast <- forecast(growth_model, h=12)
autoplot(growth_forecast)

##################################################
#########FORECAST EACH CATEGORY SEPARATELY#########
##################################################

merch_sales <- merch_sales %>%
  mutate(Product_Category = case_when(
    Product_Category == 1 ~ "Clothing",
    Product_Category == 2 ~ "Ornaments",
    Product_Category == 3 ~ "Other",
    TRUE ~ as.character(Product_Category)
  ))

categories <- unique(merch_sales$Product_Category)
for (cat in categories) {
  cat("\n\n--- Forecast for category:", cat, "---\n")
  category_data <- merch_sales %>%
    filter(Product_Category == cat) %>%
    mutate(Month = floor_date(Order_Date, "month")) %>%
    group_by(Month) %>%
    summarise(avg_sales = mean(Total_Sales, na.rm=TRUE)) %>%
    arrange(Month)


ts_data <- ts(category_data$avg_sales, frequency=12, start=c(2023,11))

model <- auto.arima(ts_data)
forecast_h <- forecast(model, h=6)

print(
  autoplot(forecast_h) +
    ggtitle(paste("Forecast of average monthly sales for", cat)) +
    ylab("Average Sales") + xlab("Time") +
    theme_minimal()
)
}

#forecasting the growth rates for each category

for (cat in unique(merch_sales$Product_Category)) {
  cat_data <- merch_sales %>%
    filter(Product_Category == cat) %>%
    mutate(Month = floor_date(Order_Date, "month")) %>%
    group_by(Month) %>%
    summarise(avg_sales = mean(Total_Sales, na.rm=TRUE)) %>%
    arrange(Month) %>%
    mutate(growth_rate = c(NA, diff(log(avg_sales))))

  growth_ts_category <- ts(na.omit(cat_data$growth_rate), frequency = 12, start = c(2023, 11))

  growth_model_category <- auto.arima(growth_ts_category)
  growth_forecast_category <- forecast(growth_model_category, h=12)

  print(
    autoplot(growth_forecast_category) +
      ggtitle(paste("Forecast of growth rates for", cat)) +
      ylab("Growth Rate") + xlab("Time") +
      theme_minimal()
  )
}
```



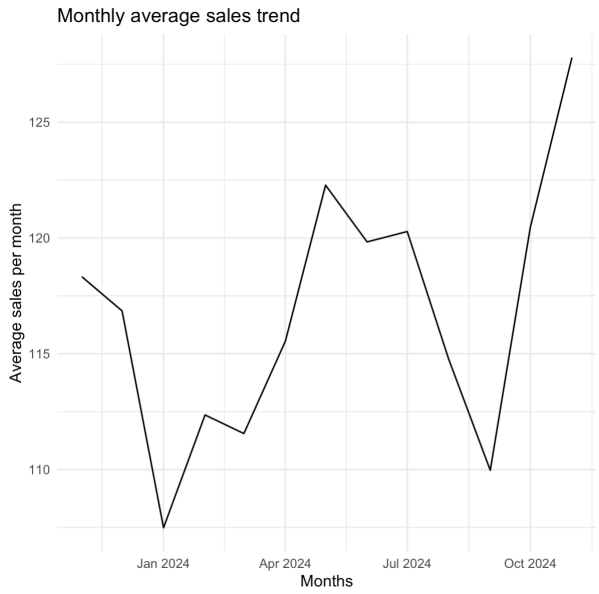

```
> summary(my_logit_norm)

Call:
glm(formula = businessoutcome ~ Product_Category + Buyer_Gender +
    Buyer_Age + International_Shipping + Sales_Price + Shipping_Charges +
    Quantity, family = "binomial", data = merch_sales_train)

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.1267988  0.2566997  -4.390 1.14e-05 ***
Product_Category      0.0504113  0.0701931   0.718    0.473
Buyer_Gender          0.0568188  0.0620955   0.915    0.360
Buyer_Age             0.0006603  0.0054504   0.121    0.904
International_Shipping -0.0081101  0.1337385  -0.061    0.952
Sales_Price           0.0009332  0.0013697   0.681    0.496
Shipping_Charges      0.0015383  0.0024672   0.623    0.533
Quantity              0.0174753  0.0260965   0.670    0.503
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7222.3  on 5914  degrees of freedom
Residual deviance: 7218.9  on 5907  degrees of freedom
AIC: 7234.9

Number of Fisher Scoring iterations: 4
```

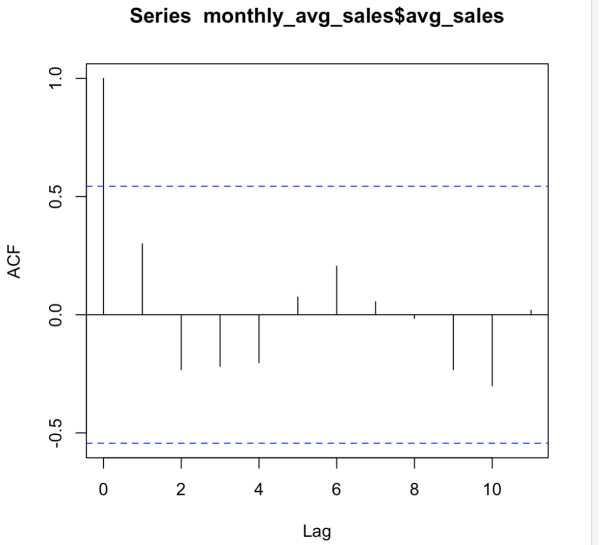
```
> confusionMatrix(data =  as.factor(as.numeric(tree_pred[,2]>0.5)),
+                 reference = as.factor(as.numeric(merch_sales_test$businessoutcome)))
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1025  408
         1   32   14

               Accuracy : 0.7025
                 95% CI : (0.6785, 0.7257)
    No Information Rate : 0.7147
    P-Value [Acc > NIR] : 0.8565

                  Kappa : 0.004

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.96973
            Specificity : 0.03318
         Pos Pred Value : 0.71528
         Neg Pred Value : 0.30435
             Prevalence : 0.71467
         Detection Rate : 0.69304
   Detection Prevalence : 0.96890
      Balanced Accuracy : 0.50145

       'Positive' Class : 0
```

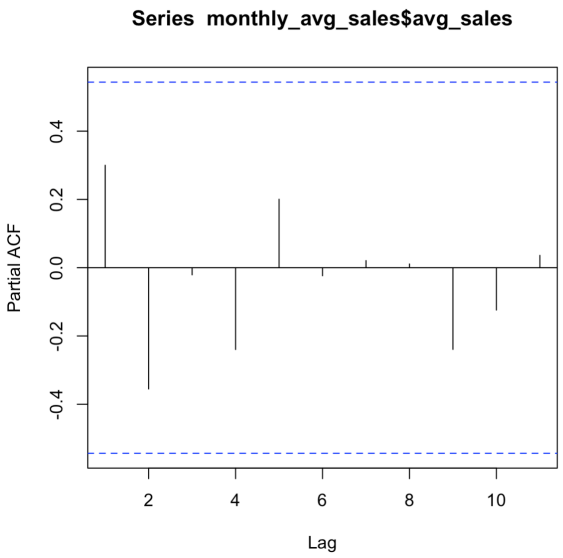


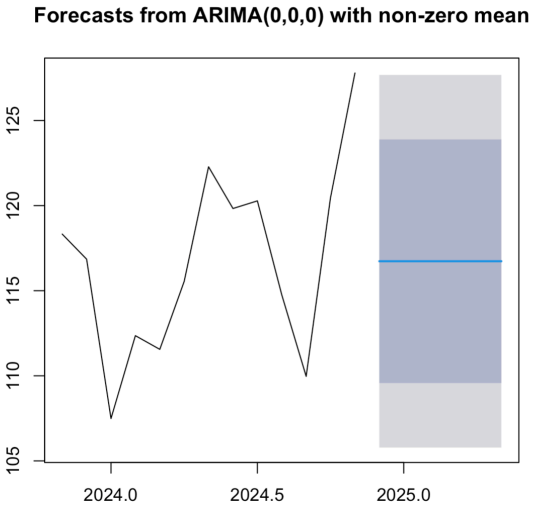


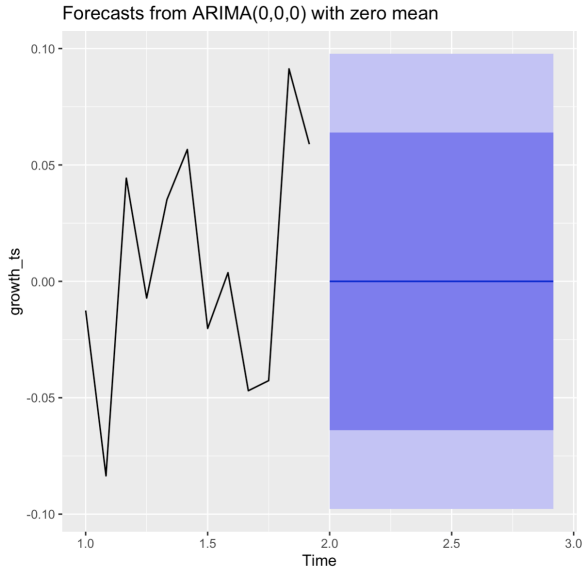

```
> summary(arima_model)
Series: sales_ts
ARIMA(0,0,0) with non-zero mean

Coefficients:
         mean
      116.7311
s.e.    1.4881

sigma^2 = 31.19:  log likelihood = -40.29
AIC=84.57   AICc=85.77   BIC=85.7

Training set error measures:
                      ME     RMSE      MAE        MPE     MAPE      MASE      ACF1
Training set -2.691864e-14 5.365523 4.417826 -0.2108027 3.793569 0.4668477 0.3000907
```

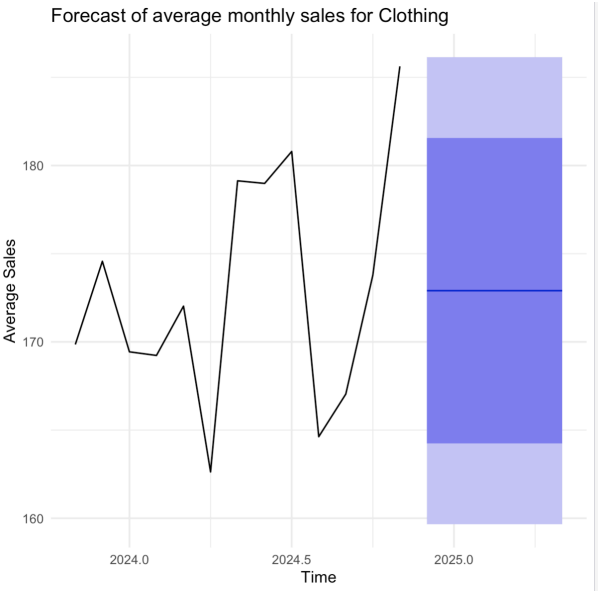

Monthly average sales trend
Average sales per month
125
120
115
110
Jan 2024
Apr 2024
Jul 2024
Oct 2024
Months

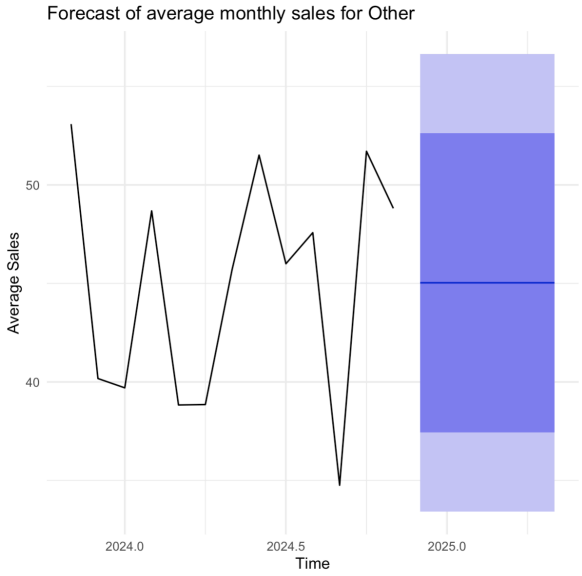

Series monthly_avg_sales$avg_sales
ACF
1.0
0.5
0.0
-0.5
0
2
4
6
8
10
Lag

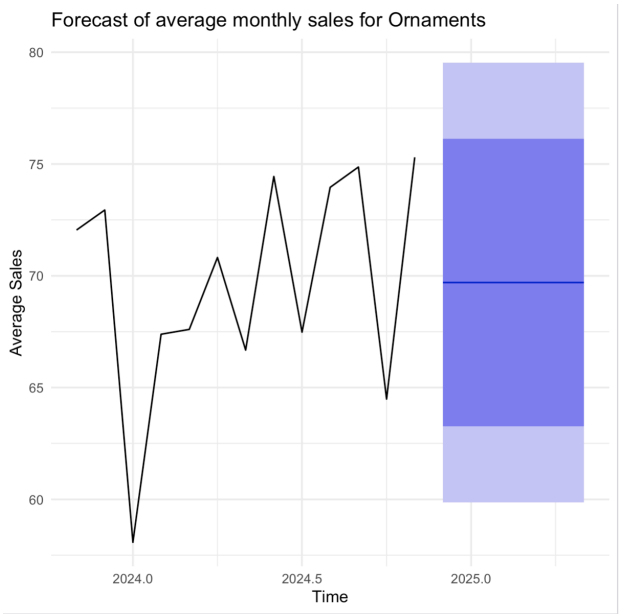

Series monthly_avg_sales$avg_sales
Partial ACF
0.4
0.2
0.0
-0.2
-0.4
2
4
6
8
10
Lag

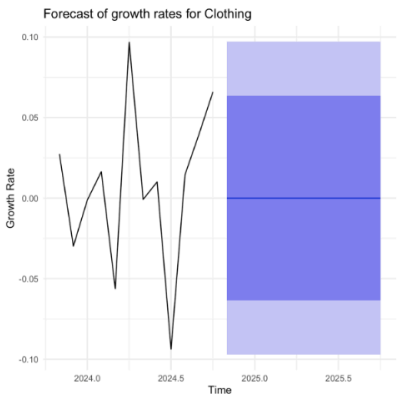

Forecasts from ARIMA(0,0,0) with non-zero mean
125
120
115
110
105
2024.0
2024.5
2025.0

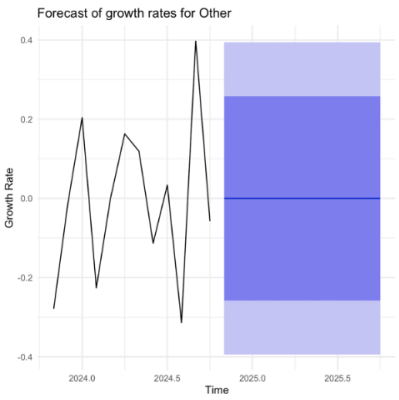

Forecasts from ARIMA(0,0,0) with zero mean
growth_ts
0.10
0.05
0.00
-0.05
-0.10
1.0
1.5
2.0
2.5
3.0
Time

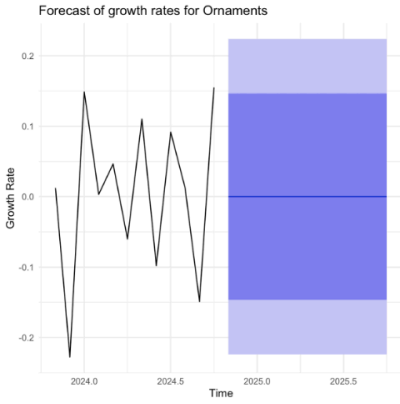

Forecast of average monthly sales for Clothing
Average Sales
180
170
160
2024.0
2024.5
2025.0
Time


Forecast of average monthly sales for Other
Average Sales
50
40
2024.0
2024.5
2025.0
Time


Forecast of average monthly sales for Ornaments
Average Sales
80
75
70
65
60
2024.0
2024.5
2025.0
Time


Forecast of growth rates for Clothing
Growth Rate
0.10
0.05
0.00
-0.05
-0.10
2024.0
2024.5
2025.0
2025.5
Time


Forecast of growth rates for Other
Growth Rate
0.4
0.2
0.0
-0.2
-0.4
2024.0
2024.5
2025.0
2025.5
Time


Forecast of growth rates for Ornaments
Growth Rate
0.2
0.1
0.0
-0.1
-0.2
2024.0
2024.5
2025.0
2025.5
Time

References

Global Retail sales Data: Orders, reviews & Trends. (2024, December 10). Kaggle. https://www.kaggle.com/datasets/adarsh0806/influencer-merchandise-sales