Final Project

Mohammed Padghawala.

2023-11-24

```
library(tidyverse)
library(MASS)
library(MLmetrics)
library(ggplot2)
```

Data Cleaning and Preparation

```
ECB <- read.csv("E-commerce Customer Behavior.csv")
```

```
cat("Number of observations (rows):", nrow(ECB), "\n")
```

```
## Number of observations (rows): 350
```

```
cat("Number of attributes (columns):", ncol(ECB), "\n")
```

```
## Number of attributes (columns): 11
```

> Five out of eleven attributes are qualitative variables, so first, we convert those categorical variables into factors. We obtained the same output in the regression model even before converting them to factors, as R is a smart language that can handle qualitative variables well. However, it is good practice to convert them to factors for a better understanding of the dataset.

```
ECB$Gender <- as.factor(ECB$Gender)
ECB$City <- as.factor(ECB$City)
ECB$Membership.Type <- as.factor(ECB$Membership.Type)
ECB$Satisfaction.Level <- as.factor(ECB$Satisfaction.Level)
ECB$Discount.Applied <- as.numeric(ECB$Discount.Applied)

str(ECB)
```

```
## 'data.frame': 350 obs. of 11 variables:
## $ Customer.ID          : int 101 102 103 104 105 106 107 108    ...
## $ Gender               : Factor w/ 2 levels "Female","Male": 1 2 1 2 2 1 1 2 1 2 ...
## $ Age                  : int 29 34 43 30 27 37 31 35 41 28 ...
## $ City                 : Factor w/ 6 levels "Chicago","Houston",..: 5 3 1 6 4 2 5
## $ Membership.Type      : Factor w/ 3 levels "Bronze","Gold",..: 2 3 1 2 3
```

```
## $ Total.Spend : num 1120 780 511 1480 720 ...
## $ Items.Purchased  : int 14 11 9 19 13 8 15 12 10 21 ...
## $ Average.Rating       : num 4.6 4.1 3.4 4.7 4 3.1 4.5 4.2     ...
## $ Discount.Applied : num 0 1 0 0 1 0 1 0 1 0 ...
## $ Days.Since.Last.Purchase: int 25 18 42 12 55 22 28 14 40 9 ...
## $ Satisfaction.Level : Factor w/ 3 levels "Neutral","Satisfied",..: 2 1 3 2 3 1 2 1 3 2
...
```

We can observe in this screenshot that if we don't convert them into factors, we will not know how many males and females or customers are from which cities in the dataset summary. Now, after converting them to factors, we can see the proper picture of these qualitative variables too.

```
    Gender              Age           City
 Length:350         Min.   :26.0   Length:350
 Class :character   1st Qu.:30.0   Class :character
 Mode  :character   Median :32.5   Mode  :character
                    Mean   :33.6
                    3rd Qu.:37.0
                    Max.   :43.0
```

```
summary(ECB)
```

```
## Customer.ID      Gender       Age                City   Membership.Type
## Min.   :101.0  Female:175  Min.   :26.0  Chicago     :58  Bronze:116
## 1st            Male :175  1st Qu.:30.0  Houston     :58  Gold :117
## Median :201.5             Median :32.5  Los Angeles :59  Silver:117
## Mean   :275.5             Mean   :33.6  Miami       :58
## 3rd Qu.:362.8             3rd Qu.:37.0  New York    :59
## Max.   :450.0             Max.   :43.0  San Francisco:58
## Total.Spend    Items.Purchased  Average.Rating  Discount.Applied
## Min.   : 410.8  Min.   : 7.0   Min.   :3.000  Min.   :0.0
## 1st Qu.:       1st Qu.: 9.0   1st Qu.:3.500  1st Qu.:0.0
## Median :       Median :12.0   Median         Median :0.5
## Mean   : 845.4  Mean   :12.6   Mean   :4.019  Mean   :0.5
## 3rd            3rd Qu.:15.0   3rd Qu.:4.500  3rd Qu.:1.0
## Max.   :1520.1  Max.   :21.0   Max.   :4.900  Max.   :1.0
##                        Satisfaction.Level
## Days Since Last Purchase  Neutral   :108
## 1st Qu.:15.00             Satisfied :125
## Median :23.00             Unsatisfied:117
## Mean :26.59
## 3rd Qu.:38.00
## Max. :63.00
```

The dataset contains the following columns: 1) Customer ID: Numeric identifier for each customer. 2) Gender: Gender of the customer. 3) Age: Age of the customer. 4) City: City where the customer is located. 5) Membership Type: Type of membership the customer holds. 6) Total Spend: Total amount spent by the customer. 7) Items Purchased: Number of items purchased. 8) Average Rating: Average rating given by the customer. 9) Discount Applied: Whether a discount was applied to the customer's purchases. 10) Days Since Last Purchase: Number of days since the customer's last purchase. 11) Satisfaction Level: Customer's level of satisfaction.

Summary of Key Statistics - Customer ID: Ranges from 101 to 450 (350 total customers) -Age: Ranges from 26 to 43 years, with an average of approximately 33.6 years. -Total Spend: Varies from $410.80 to $1520.10, with an average spend of around $845.38. -Items Purchased: Customers purchased between 7 to 21 items, averaging around 12.6 items. -Average Rating: Ranges from 3.0 to 4.9, with an average rating of about 4.02. -Days Since Last Purchase: Ranges from 9 to 63 days, with an average of approximately 26.6 days.

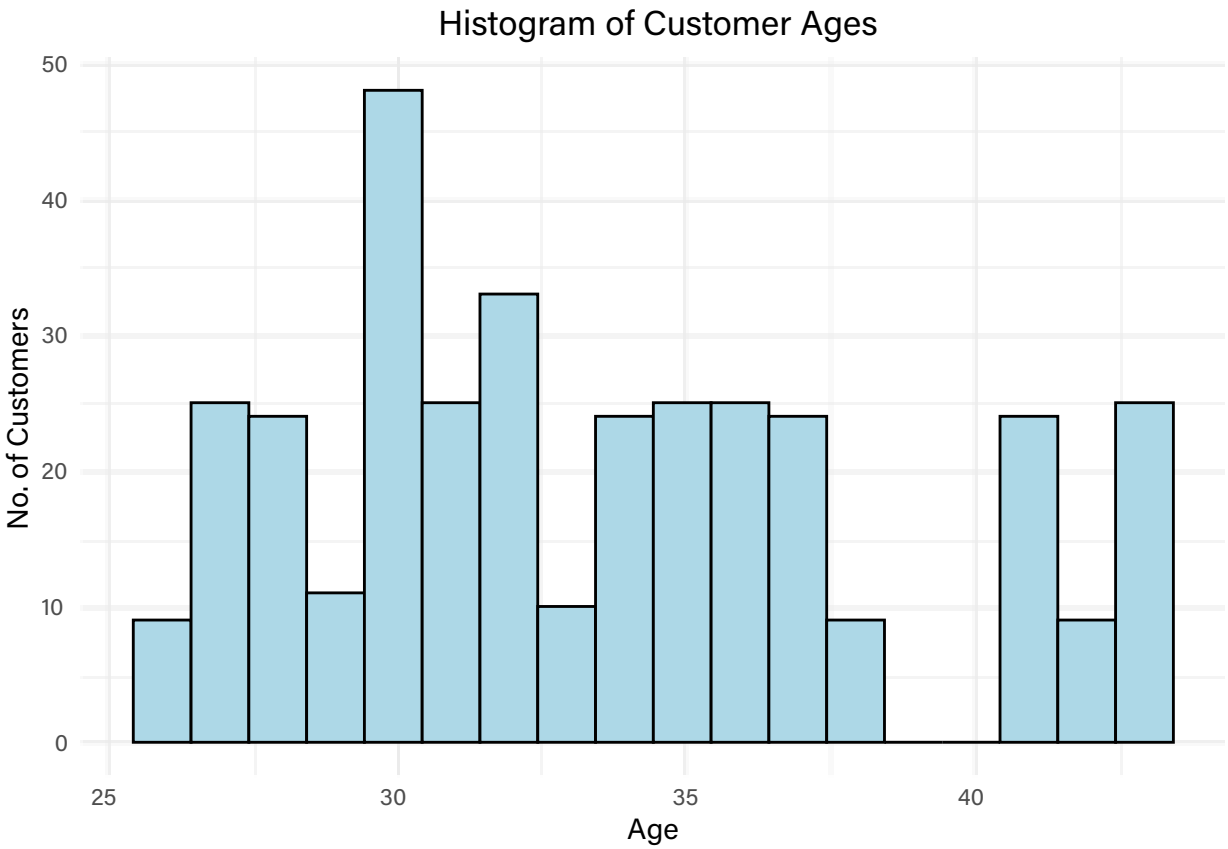
```
missing_values <- sapply(ECB, function(x) sum(is.na(x)))
missing_values
```

```
##             Customer.ID                  Gender                     Age
##                       0                       0                       0
##                    City       Membership.Type            Total.Spend
##                       0                       0                       0
##         Items.Purchased      Average.Rating      Discount.Applied
##                       0                       0                       0
## Days.Since.Last.PurchaseSatisfaction.Level
##                       0                       0
```

There are no missing values in any of the columns in this dataset.

Exploratory data analysis

```
ggplot(ECB, aes(x = Age)) + geom_histogram(binwidth = 1, fill =
"lightblue", color = "black") + theme_minimal() + labs(title =
"Histogram of Customer Ages", x = "Age", y = "No. of Customers") +
theme(plot.title = element_text(hjust = 0.5))
```



1) Age Range: The histogram shows customer ages ranging from around 25 to a bit over 40 years. This suggests that the dataset does not include younger teenagers or older adults beyond this range, or possibly that such age groups are not frequent customers. 4

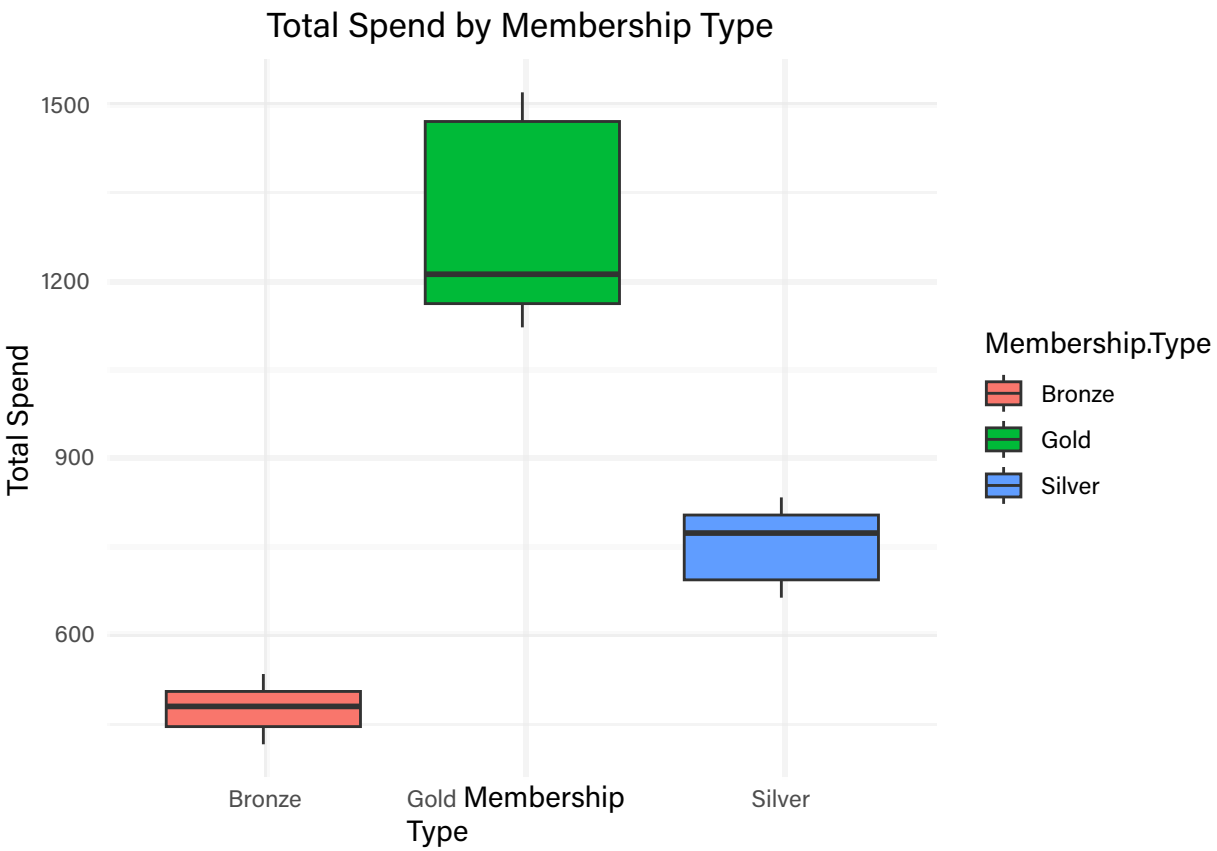
2) Most Common Age Group: There is a peak in frequency at around the age of 30. This indicates that customers around 30 years old are the most common within this e-commerce platform's dataset.

3) Symmetry: The distribution does not appear to be symmetric.

4) Business perspective: The distribution of customer ages is concentrated around the early 30s, with a peak at age 31. Marketing strategies may benefit from focusing on the age demographics most represented in the histogram.

```
ggplot(ECB, aes(x = Membership.Type, y = Total.Spend, fill =
Membership.Type)) + geom_boxplot() + theme_minimal() +

  labs(title = "Total Spend by Membership Type", x = "Membership Type", y =
  "Total Spend") + theme(plot.title = element_text(hjust = 0.5))
```



The boxplot displays the distribution of 'Total Spend' for different 'Membership Type' categories, which appear to be 'Bronze', 'Gold', and 'Silver'.

1) Spread:
   - 'Gold' members tend to spend the most, with a median spend significantly higher than both 'Silver' and 'Bronze' members.
   - 'Silver' members have a median spend that is lower than 'Gold' but higher than 'Bronze'.
   - 'Bronze' members spend the least, with the lowest median spend of the three groups.
2) Variability:
   - The 'Gold' membership boxplot shows a larger interquartile range (IQR) compared to the 'Bronze' and 'Silver', indicating more variability in spending within the 'Gold' membership.
   - The 'Silver' membership shows a smaller IQR than 'Gold', suggesting less variability in spending among 'Silver' members.
   - The 'Bronze' membership has the smallest IQR, indicating the least variability in spending among its members.
3) Outliers: There are no visible outliers in theese categories, suggesting that spends are relatively consistent within these IQRs.
4) Comparisons:
   - The difference in median spend between 'Gold' and the other two types suggests that 'Gold' members are likely a more lucrative segment for the e-commerce business.

- Business strategies could include targeting Gold members with premium offers and incen-tivizing Silver and Bronze members to upgrade.

```
ggplot(ECB, aes(x = Days.Since.Last.Purchase, fill =
Satisfaction.Level)) + geom_histogram(position = "identity", alpha
= 0.5, bins = 30) + scale_fill_manual(values = c("Satisfied" = "skyblue", "Neutral" = "orange",
    "Unsatisfied" = "purple" theme_minimal() +
    labs(title = "Days Since Last Purchase by Satisfaction
    Level", x = "Days Since Last Purchase",
         y = "No. of Customers") +
  theme(plot.title = element_text(hjust = 0.5))
```

Satisfied Customers (Sky Blue): This group has the highest frequency of customers with a smaller number of days since their last purchase, peaking at around 10-25 days. This indicates that customers who are satisfied with the service or product tend to make repeat purchases relatively quickly.

Neutral Customers (Orange): The distribution for neutral customers is centered around 15-30 days, with a moderate frequency. The trend suggests that while these customers are not entirely dissatisfied, they do not exhibit the same promptness in making another purchase as satisfied customers.

Unsatisfied Customers (Purple): This group has a peak later than the other two groups, around 40-60 days since the last purchase. The spread of the distribution is also wider, indicating that unsatisfied customers are the least consistent in their purchasing patterns and take longer to return for subsequent purchases.

Potential Business Actions : Targeted Interventions-The data could inform targeted interventions to increase purchase frequency. For satisfied customers, reinforcing positive behavior with loyalty programs could be effective. For neutral customers, promotional offers or feedback requests might stimulate more frequent engagement. For unsatisfied customers, addressing their concerns and improving their experience is crucial.

Hypothesis Test

Null Hypothesis (H0): The average number of days since the last purchase for satisfied customers is equal to that of unsatisfied customers. In other words, customer satisfaction does not affect the frequency of purchases. Alternative Hypothesis (H1): The average number of days since the last purchase for satisfied customers is not equal to that of unsatisfied customers, indicating that customer satisfaction does affect the frequency of purchases.

```
data_satisfied <- ECB[ECB$Satisfaction.Level == "Satisfied",
"Days.Since.Last.Purchase"] data_unsatisfied <- ECB[ECB$Satisfaction.Level ==
"Unsatisfied", "Days.Since.Last.Purchase"]
t_test_result <- t.test(data_satisfied, data_unsatisfied, alternative = "two.sided",
var.equal = TRUE)
print(t_test_result)
```

```
##
## Two Sample t-test
##
## data: data_satisfied and data_unsatisfied
## t = -25.251, df = 240, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -27.05688 -23.14087
## sample estimates:
## mean of x mean of y
## 17.69600 42.79487
```

t-value: -25.251 : The negative t-value indicates that the mean of the satisfied customers is less than the mean of the unsatisfied customers. This suggests that satisfied customers may make repeat purchases sooner than unsatisfied customers.

p-value: $<$ 2.2e-16 : The p-value is extremely small (less than the standard alpha level of 0.05), which provides very strong evidence against the null hypothesis. It suggests that the observed difference in means is highly unlikely to have occurred by random chance.

95% Confidence Interval: [-27.05688, -23.14087] : This interval does not contain zero, which further indicates that the difference in means is statistically significant. It provides a range of plausible values for the true mean difference between the two groups.

Sample Estimates: Mean of x (satisfied customers): 17.69600 days Mean of y (unsatisfied customers): 42.79487 days

Conclusion from the Test : The results of the t-test lead us to reject the null hypothesis in favor of the alternative hypothesis. This means there is statistically significant evidence to suggest that the number of days since the last purchase is different between satisfied and unsatisfied customers, with satisfied customers making repeat purchases in fewer days on average than unsatisfied customers.

Business Implications: The business should prioritize customer satisfaction strategies since they have a significant impact on the frequency of purchases. Understanding and addressing the factors contributing to customer dissatisfaction can potentially lead to increased purchase frequency and customer retention. The results could justify investments in customer service improvement, user experience enhancements, or satisfaction surveys to keep customers engaged and satisfied.

```
var.test(data_satisfied, data_unsatisfied)
```

```
##
## F test to compare two variances
##
## data: data_satisfied and data_unsatisfied
## F = 0.82069, num df = 124, denom df = 116, p-value = 0.2793
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.5720982 1.1747444
## sample estimates:
## ratio of variances
## 0.8206938
```

In previous chunk "var.equal = TRUE indicates the assumption that both groups have the same variance". So here we conduct the F test to ensure that variances are equal. An F-value close to 1 suggests that the variances are similar.

Linear regression model

Split the E-commerce customer behavior dataset data set into two parts ECBTraining and ECBTest.

```
set.seed(100)
EB <- sample(2, nrow(ECB), replace=TRUE, prob=c(0.8, 0.2))
ECBTraining <- ECB[EB == 1, ]
ECBTest <- ECB[EB == 2, ]
```

Here the dataset is randomly split into training (80%) and testing (20%) sets using the sample function. This is to ensure that the model can be trained on one subset of data and tested on a separate subset for validation purposes. First we execute code "set.seed(100)" just to ensures that the random process of splitting the data can be replicated for consistency in results.

Forward Stepwise Selection

```
null_model <- lm(Total.Spend ~ 1, data = ECBTraining[,2:11])

full_model <- lm(Total.Spend ~ .,data = ECBTraining[,2:11])
```

First we create the "null model." This model assumes that the total amount spent can be predicted by intercept only (no predictors), irrespective of any other variables. It serves as a baseline, providing a comparison point for more complex models.

Then we build full model with all available predictors included except cutomer ID as it is a unique code for each customer. It shouldn't help predict how much a customer will spend. Adding Customer.ID as a predictor in models can cause issues. And include rest of the variables as we they think might affect the total amount spent.

Stepwise forward selection process starting from the null model, considering predictors specified in the full model. The stepAIC function chooses the best model by minimizing the AIC value. At each step, the AIC value is reported, with lower values indicating a potentially better model

```
forward_model <- stepAIC(null_model, direction = "forward", scope = formula(full_model))
```

```
## Start: AIC=3338.73
## Total.Spend ~
1 ## ##
                          Df Sum of Sq      RSS   AIC
## + City                    5 37189535167143 1817.9
## + Items.Purchased         1 35349326        2007351
## + Membership.Type         2 34721178        2613499
## + Average.Rating          1 32981630        2972047
## + Satisfaction.Level   2 30130450 7226227 2837.8
## + Age                          1 16559253 20797424
## +                              1 11789786 25567892
Days.Since.Last.Purchase          1 5972854 31383823
## + Discount.Applied     1 1546911 35809766 3329.8
##                                    37356677 3338.7
<none>
## Step: AIC=1817.87
## Total.Spend ~
City ##
##                           Df Sum of Sq RSS    AIC
## + Items.Purchased          1    92987 74155 1589.9
## + Age                      1 11590 155552 1799.5
## + Average.Rating           1  5109 162033 1811.1
## + Satisfaction.Level       2  4983 162160 1813.3
## <none>                              167143 1817.9
## +                            1 416 166726 1819.2
Days.Since.Last.Purchase
## Step: AIC=1589.88
## Total.Spend ~ City + Items.Purchased
##
##                           Df Sum of Sq RSS   AIC
## +                          1 13072.2 61083 1537.0
Days.Since.Last.Purchase   1 5157.5 68998 1571.5
## + Average.Rating           1 1094.5 73061 1587.7
## <none>                                 74155 1589.9
## + Satisfaction.Level       2 270.1 73885 1592.8
##
## Step: AIC=1537
## Total.Spend ~ City + Items.Purchased +
Days.Since.Last.Purchase ##
## Df Sum of Sq RSS AIC
## + Age               1 1597.96 59485 1531.5
## +                   2 933.06 60150 1536.6
Satisfaction.Level
```

```
## <none> ## +              61083 1537.0
Average.Rating        1     35.82 61047
                            1538.8
## ## Step: AIC=1531.49 ## Total.Spend ~ City +
Items.Purchased + Days.Since.Last.Purchase + ## Age ##
```

```
##                    Df Sum of    RSS   AIC
## + Satisfaction.Level Sq        837.88 58647
2 ## <none> ## +                 1531.5 59485
Average.Rating 1           1531.5 6.32 59479
                                  1533.5
## ## Step: AIC=1531.48 ## Total.Spend ~ City +
Items.Purchased + Days.Since.Last.Purchase + ## Age +
Satisfaction.Level ##
```

```
##                 Df Sum of    RSS   AIC
## <none> ## +     Sq         58647 1531.5
Average.Rating 1             7.7892 58639
                                  1533.4
```

At each step, the AIC value is reported, with lower values indicating a potentially better model.

Next we obtain Analysis of Deviance Table by calling anova on the forward model. This summary gives a concise overview of the model building process and the incremental contribution of each variable added to the model.

```
forward_model$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Total.Spend ~ 1
##
## Final Model:
## Total.Spend ~ City + Items.Purchased + Days.Since.Last.Purchase +
## Age + Satisfaction.Level
##
## ## Step Df Deviance Resid. Df Resid. Dev AIC ## 1 282 37356677.27
3338.733 ## 2 + City 5 3.718953e+07 277 167142.51 1817.867 ## 3 +
Items.Purchased 1 9.298736e+04 276 74155.15 1589.876 ## 4 +
Days.Since.Last.Purchase 1 1.307217e+04 275 61082.98 1536.995 ## 5 +
Age 1 1.597955e+03 274 59485.03 1531.493 ## 6 + Satisfaction.Level 2
8.378787e+02 272 58647.15 1531.479
```

A summary of the forward model is provides detailed output for the coefficients of the final model, including estimates, standard errors, t-values, and p-values.

```
summary(forward_model)
```

```
##
## Call:
## lm(formula = Total.Spend ~ City + Items.Purchased +
Days.Since.Last.Purchase + Age + Satisfaction.Level, data = ECBTraining[, 2:11])
##
## Residuals:
## Min 1Q Median 3Q Max ## -46.165
-8.756 -0.150 8.041 46.722
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 236.7877   44.1827  5.359 1.78e-07 ***
## CityHouston                 -33.7103   16.5217 -2.040 0.042178 *
## CityLos Angeles             211.0503   11.960 < 2e-16 ***
## CityMiami                   163.4946   9.3902 17.411 < 2e-16 ***
## CityNew York                505.0338   26.856 < 2e-16 ***
## CitySan Francisco 659.1396 22.0789 29.854          < 2e-16 ***
## Items.Purchased 26.5784 1.2811 20.747 < 2e-16 ***
## Days.Since.Last.Purchase -1.3944 0.2038 -6.842 5.15e-11 ***
## Age 2.0287 0.7684 2.640 0.00877 **
## Satisfaction.LevelSatisfied -11.4669 6.6699 -1.719 0.08671 .
## Satisfaction.LevelUnsatisfied -14.1059 14.8717 -0.949 0.34371
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.68 on 272 degrees of freedom
## Multiple R-squared: 0.9984, Adjusted R-squared: 0.9984
## F-statistic: 1.73e+04 on 10 and 272 DF, p-value: < 2.2e-16
```

The final model suggests that all included cities, 'Items.Purchased', and 'Days.Since.Last.Purchase' are significant predictors of 'Total.Spend'. -The 'Age' variable is also significant, while 'Sat-isfaction.Level' categories are not significant at the 0.05 level. - The R-squared value is very high and close to 1 that suggests model is a good fit and explains almost all the variability in Total.Spend.

```
predictions <- predict(forward_model, newdata = ECBTest[,2:11])
MAE(y_pred = predictions, y_true = ECBTest$Total.Spend)
```

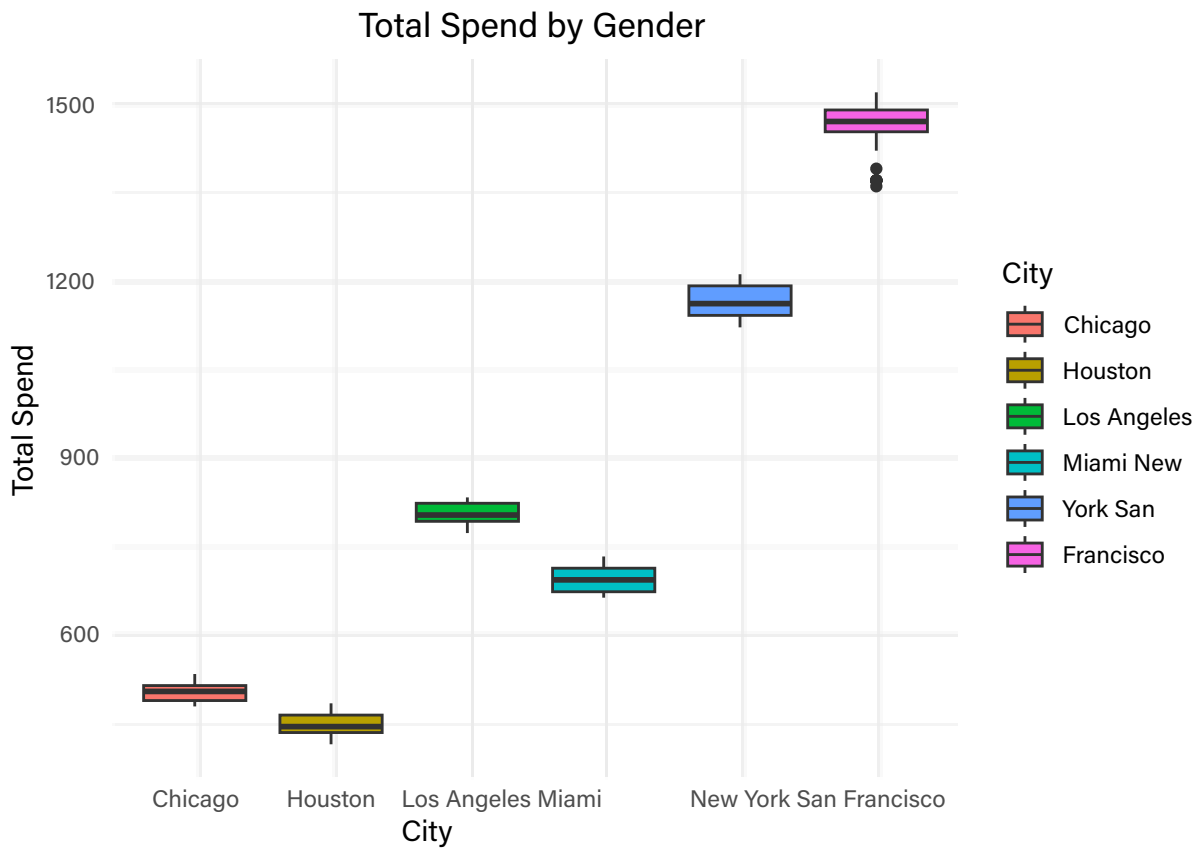
```
## [1] 11.24633
```

```
MSE(y_pred = predictions, y_true = ECBTest$Total.Spend)
```

```
## [1] 199.5523
```

In this context, it is observed that the variables "city of New York" and "Houston" possess the highest t-values, signifying their high level of significance. Additionally, their very low p-values suggest a significant association with Total Spend.

```
ggplot(ECB, aes(x = City, y = Total.Spend, fill = City)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Total Spend by Gender", x = "City", y = "Total Spend") +
  theme(plot.title = element_text(hjust = 0.5))
```



Backward Stepwise Selection

```
backward <- stepAIC (full_model, direction='backward')
```

```
## Start: AIC=1533.44
## Total.Spend ~ Gender + Age + City + Membership.Type + Items.Purchased +
## Average.Rating + Discount.Applied + Days.Since.Last.Purchase +
## Satisfaction.Level
##
##
## Step: AIC=1533.44
## Total.Spend ~ Gender + Age + City + Membership.Type +
Items.Purchased + ## Average.Rating + Days.Since.Last.Purchase +
## Satisfaction.Level
##
## Step: AIC=1533.44
## Total.Spend ~ Gender + Age + City + Items.Purchased + Average.Rating +
##  Days.Since.Last.Purchase + Satisfaction.Level
##
##
## Step: AIC=1533.44
## Total.Spend ~ Age + City + Items.Purchased + Average.Rating +
## Days.Since.Last.Purchase + Satisfaction.Level
##
```

```
##                           Df Sum of     RSS    AIC
## - Average.Rating           1      8   58647 1531.5
## <none>                                 58639 1533.4
## - Satisfaction.Level       2    839   59479
## - Age                      1   1481 60120 1538.5
## -                          1   9880   68519
## - Days.Since.Last.Purchase 1  91574 150213 1797.6
## - City                     5 472393 531032 2147.0
##
## Step: AIC=1531.48
## Total.Spend ~ Age + City + Items.Purchased + Days.Since.Last.Purchase +
## Satisfaction.Level
##
##                           Df Sum of     RSS    AIC
## <none>                     Sq                58647
## - Satisfaction.Level       2    838   59485
## - Age                      1   1503 60150 1535.5
## -                          1  10094 68741 1574.4
## - Days.Since.Last.Purchase 1  92807 151454 1798.0
## - City                     5 702320 760967 2246.8
```

```
backward$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Total.Spend ~ Gender + Age + City + Membership.Type + Items.Purchased +
## Average.Rating + Discount.Applied + Days.Since.Last.Purchase +
## Satisfaction.Level
##
## Final Model:
## Total.Spend ~ Age + City + Items.Purchased + Days.Since.Last.Purchase +
## Satisfaction.Level
##
##
## Step Df Deviance Resid. Df Resid. Dev AIC ## 1 271
58639.36 1533.441 ## 2 - Discount.Applied 0 0.000000e+00 271
58639.36 1533.441 0 0.000000e+00 271 58639.36 1533.441 ## 4
##Gender Membership.Type 0 0.000000e+00 271
                                                58639.36 1533.441
## 5  -               1 7.789229e+00 272 58647.15 1531.479
      Average.Rating
```

```
summary(backward)
```

```
##
## Call:
## lm(formula = Total.Spend ~ Age + City + Items.Purchased + Days.Since.Last.Purchase +
## Satisfaction.Level, data = ECBTraining[, 2:11])
##
##
## Residuals:    1Q Median     3Q    Max
## -46.165 -8.756 -0.150 8.041 46.722
```

```
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                236.7877   44.1827   5.359 1.78e-07 ***
## Age                          2.0287    0.7684   2.640  0.00877**
##              ## CityHouston -33.7103 16.5217 -2.040 0.04228 *
## CityLos Angeles 211.0503 17.6462 11.960 < 2e-16 ***
## CityMiami                  163.4946    9.3902 17.411 <      ***
## CityNew York               505.0338   18.8053 26.856 2e-16 ***
## CitySan Francisco          659.1396   22.0789 29.854 <      ***
## Items.Purchased             26.5784    1.2811 20.747 2e-16  ***
## Days.Since.Last.Purchase -1.3944 0.2038 -6.842 5.15e-11***
##                            -11.4669    6.6699 -1.719 0.08671.
## Satisfaction Level Satisfied       14.8717 -0.949 0.34371
## Satisfaction Level Unsatisfied
##14.1059
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.68 on 272 degrees of freedom
## Multiple R-squared: 0.9984, Adjusted R-squared: 0.9984
## F-statistic: 1.73e+04 on 10 and 272 DF, p-value: < 2.2e-16
```

```
predictions_backward <-predict(backward, newdata = ECBTest[,2:11])
MAE(y_pred = predictions_backward, y_true = ECBTest$Total.Spend)
```

```
## [1] 11.24633
```

```
MSE(y_pred = predictions_backward, y_true = ECBTest$Total.Spend)
```

```
## [1] 199.5523
```