BUAN 6356.502(Mon 7-9.45pm)
Group Project Report

GroupNo:5
Members:

1. Shaili Gunda
2. Nikhil Sanjay Shelke
3. Parthipan Raj Shanmugam Sakthivel
4. Srikar Brahmadevara
5. Sparshitha Bandi

Information about the dataset:
The dataset includes feature vectors from 12,220 different sessions. To avoid biases toward a marketing initiative, special day, user information, or time frame, the dataset was constructed so that each session belonged to a different user over a one-year period.

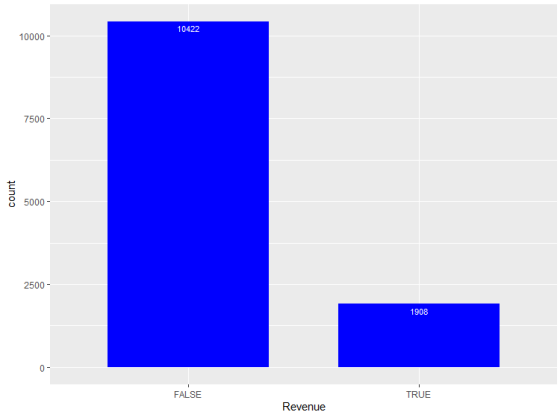
Objective:
Our goal is to boost the company's profitability by generating insights about customers who did and did not shop from the website and determining why.

Data Sources:
We are going to use the Online shopping dataset for this project and it is available on the link: https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset

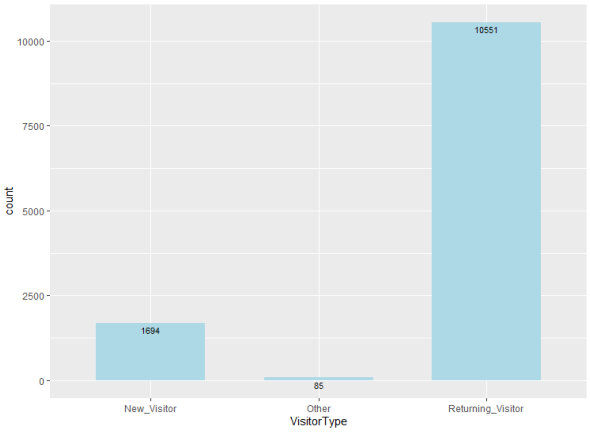
Hypothesis:
- Affiliation to determine the relationship between existing or new client patterns and their weekend, weekday, or special day shopping habits.
- There are ten numerical and eight categorical attributes in the dataset.
- The class label can be set to the 'Revenue' attribute.
- The "Product Related" and "Product-Related Duration" columns show how many distinct pages the customer visited during that session.
- The parameters measured by "Google Analytics" for each page on the e-commerce site are shown in the columns "Bounce Rate," "Exit Rate," and "Page Value."
- The value of the "Exit Rate" field for a certain web page is calculated as the percentage of all pageviews to that page that occurred last in the session.
- The "Page Value" column shows the average value of a website page visited by a user prior to completing an e-commerce transaction.

Insights:

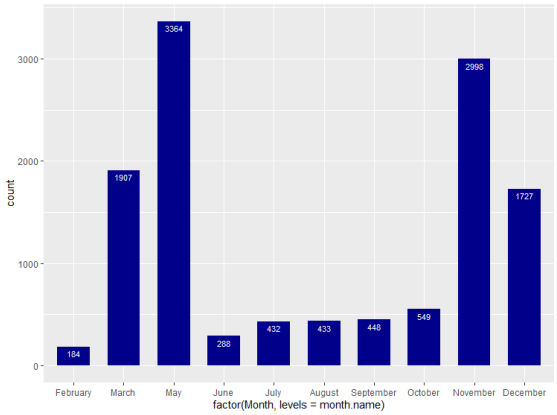
Revenue Wise Count



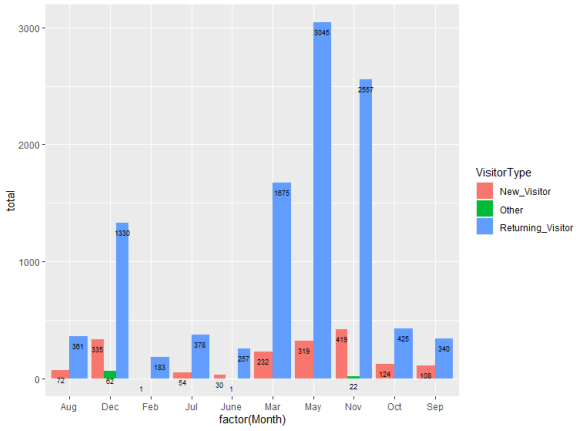
Visitor Wise Count



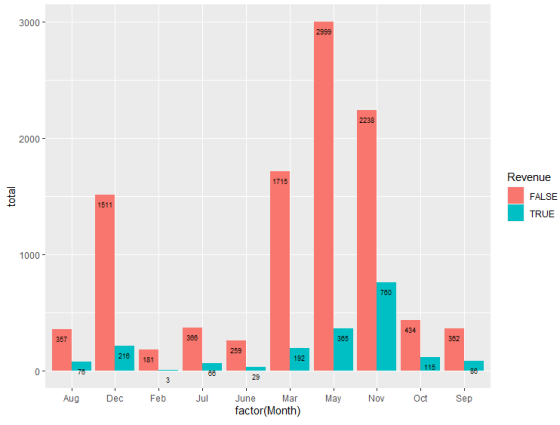
Month Wise Count



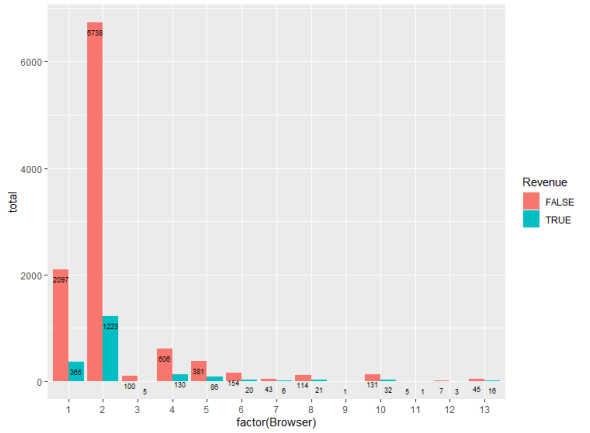
Month Wise Visitors Count



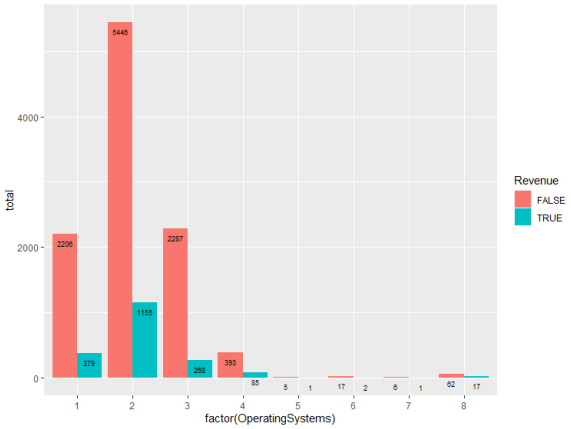
Month Wise Revenue Count



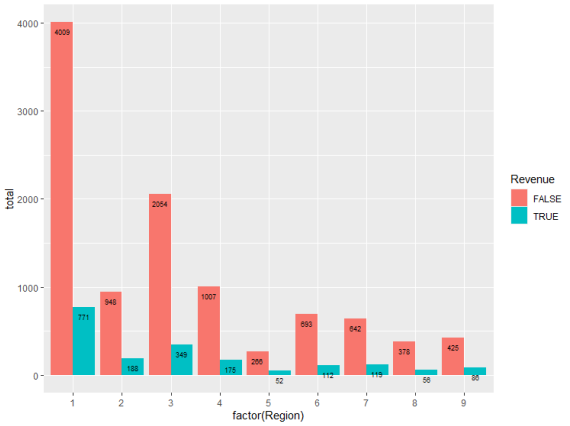
Browser Wise Revenue Count



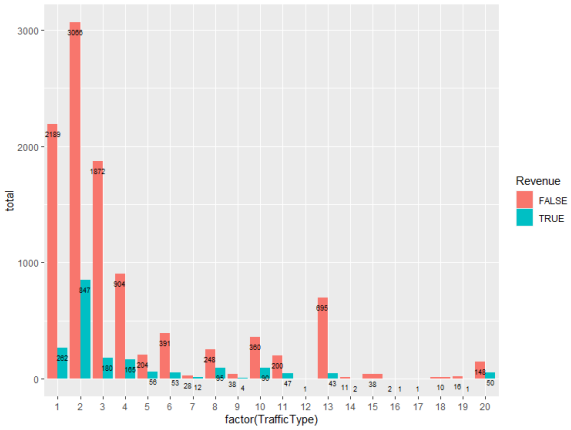
Operating Systems Wise Revenue Count



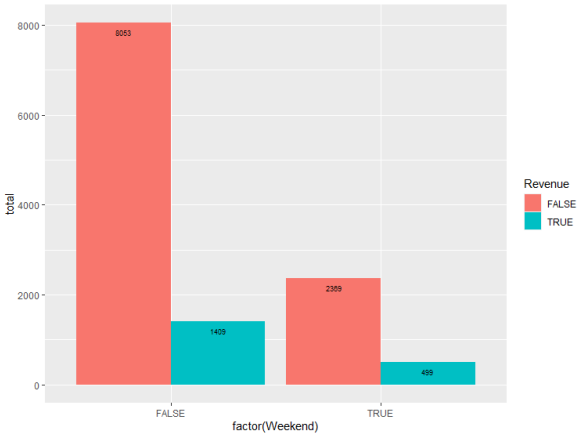
Region Wise Revenue Count



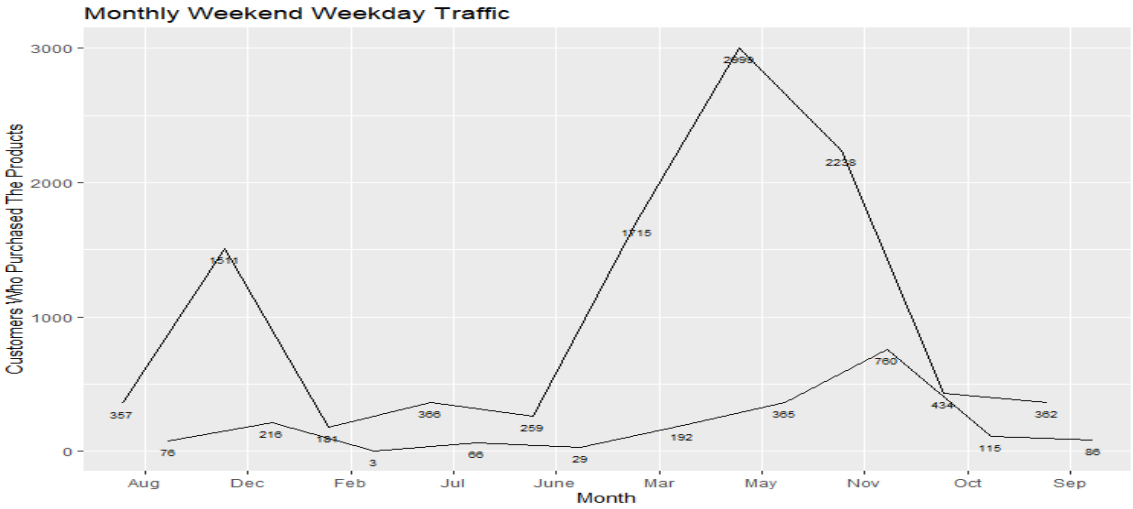
Traffic Wise Revenue Count



Weekend-Weekday Wise Revenue Count



Monthly Weekend-Weekday Traffic



Classifications:

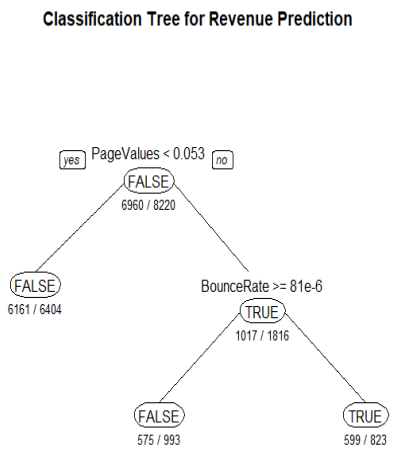
- **Modeling Data Pre-Processing Steps:**
  - The variables 'Traffic Type' and 'Browser' were re-coded since there were too many values at the factor level and the categories needed to be reduced.
- **Data for Training and Validation:**
  - The training and validation splits are retained at two-thirds and one-third, respectively.

Scaling Data Points as Part of Pre-Processing

Decision Tree:

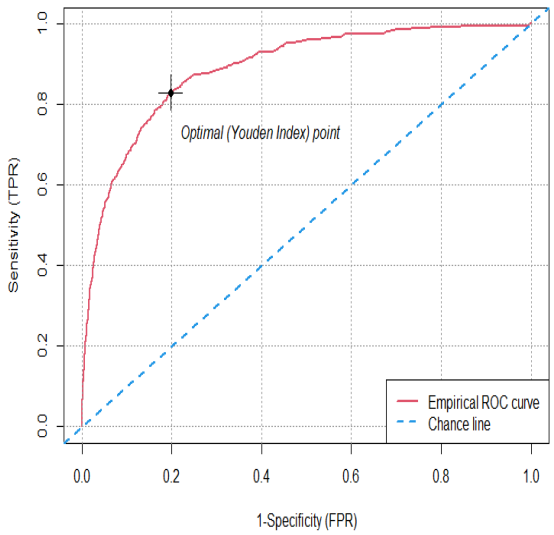
- Leaves of the Decision Tree: 3
- At the end of a visitor's visit, the tree shows whether income is earned or not.
- The tree is divided into sections based on the number of pages seen, the kind of visitor (returning or new), the month of purchase, the goods viewed, and the area from where the transaction was made.
- To bring all the data points on the same scale, the data is scaled using the 'center' and 'scale' techniques.

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  6736  661
     TRUE    224  599

               Accuracy : 0.8923
                 95% CI : (0.8854, 0.899)
    No Information Rate : 0.8467
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5166

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9678
            Specificity : 0.4754
         Pos Pred Value : 0.9106
         Neg Pred Value : 0.7278
             Prevalence : 0.8467
         Detection Rate : 0.8195
   Detection Prevalence : 0.8999
      Balanced Accuracy : 0.7216

       'Positive' Class : FALSE

>
```



Logistic Regression Model:

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  3386  413
     TRUE     76  235

               Accuracy : 0.881
                 95% CI : (0.8707, 0.8908)
    No Information Rate : 0.8423
    P-Value [Acc > NIR] : 1.013e-12

                  Kappa : 0.432

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9780
            Specificity : 0.3627
         Pos Pred Value : 0.8913
         Neg Pred Value : 0.7556
             Prevalence : 0.8423
         Detection Rate : 0.8238
   Detection Prevalence : 0.9243
      Balanced Accuracy : 0.6704

       'Positive' Class : FALSE
```



Conclusion based on Models:

- After assessing the decision tree and logistic regression, the Decision Tree model performed better in terms of model correctness, with an overall accuracy of 89.23 percent.
- We would utilize the Decision Tree model to determine the profitability of the online website because of its great accuracy. This profitability may be increased by concentrating on the following aspects of the data: website UI, exit rate, product-related information page, and bounce rate, as well as other factors that aid in better decision making.

Insights for Bettering Customer Experience through the website:

- The Remarkable Importance of Page Value encompasses clients who will compare various items and their suggestions.
- A significant improvement in recommendation tools and packages would result in increased website conversions. It comprises more items, which might increase income by utilizing the adverse effect of e-commerce in the longer run.
- A website's bounce rate may be lowered by using quicker refresh rates and building a visually appealing landing page with exceptionally low prices on items and offers available only to visitors.
- Personalizing messages for existing members and adding customer relationship management might also help with the engagement of customers.
- User experience pages with a simple and engaging design to keep people on the website pages longer.
- Providing users with product information and pricing along with special discounts.
- Increasing the number of visitors to the website using advertisements, coupons, and external promotions.