**21l-1806**

**21l-1834**

**21l-1883**

**PROJECT-3**

**REPORT**

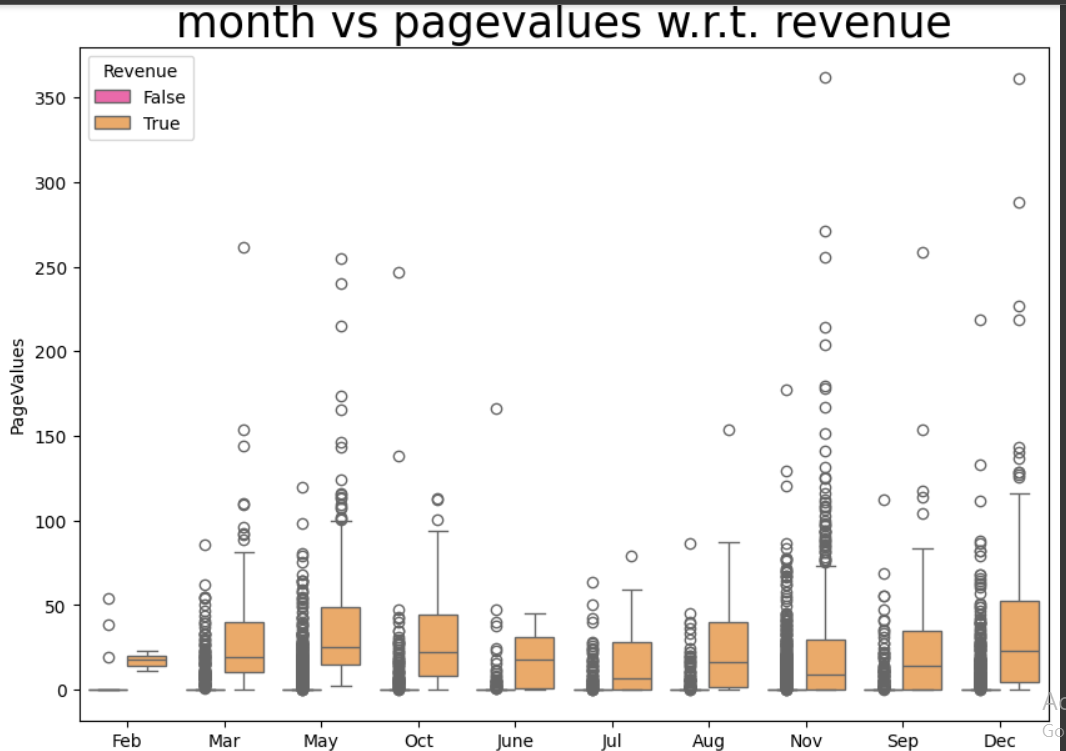
This report summarizes the findings of a comprehensive data analysis conducted on a dataset pertaining to user behavior on a website. The primary objective of the analysis was to identify patterns and insights that could potentially enhance the website's revenue generation.

**Data Exploration:**

**1. Descriptive Statistics:**

- The dataset contained various features related to user behavior, such as page views, bounce rates, exit rates, and revenue.

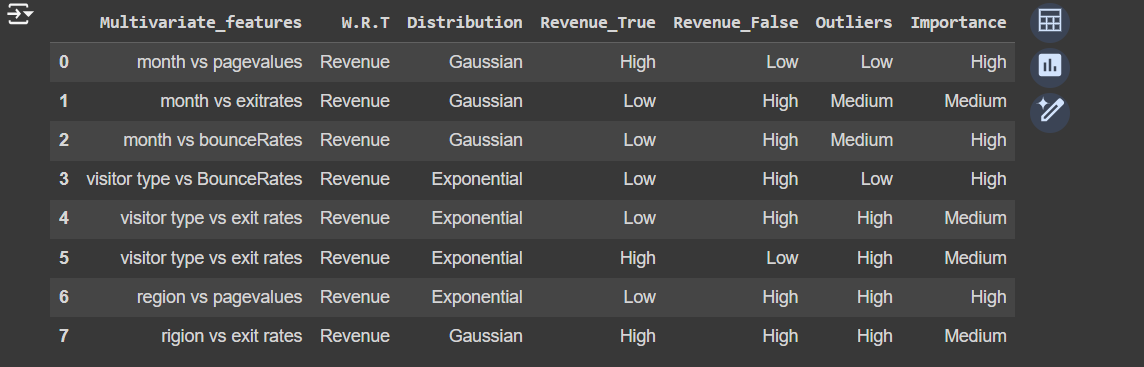
- Exploratory data analysis revealed interesting trends, such as a positive correlation between page views and revenue.



**2. Multivariate Feature Analysis:**

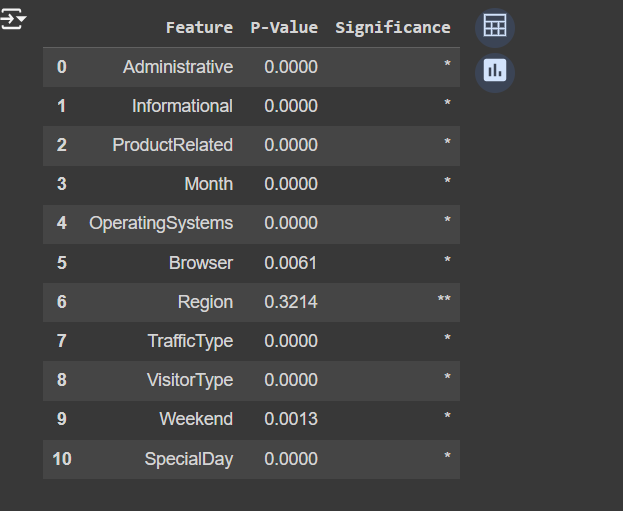
- Multivariate analysis identified several significant relationships between features and revenue.

- For instance, visitors spending more time on product-related pages were more likely to generate revenue.



**3. Categorical Feature Analysis:**

- Chi-square tests revealed that certain categorical features, such as visitor type and region, significantly influenced revenue.



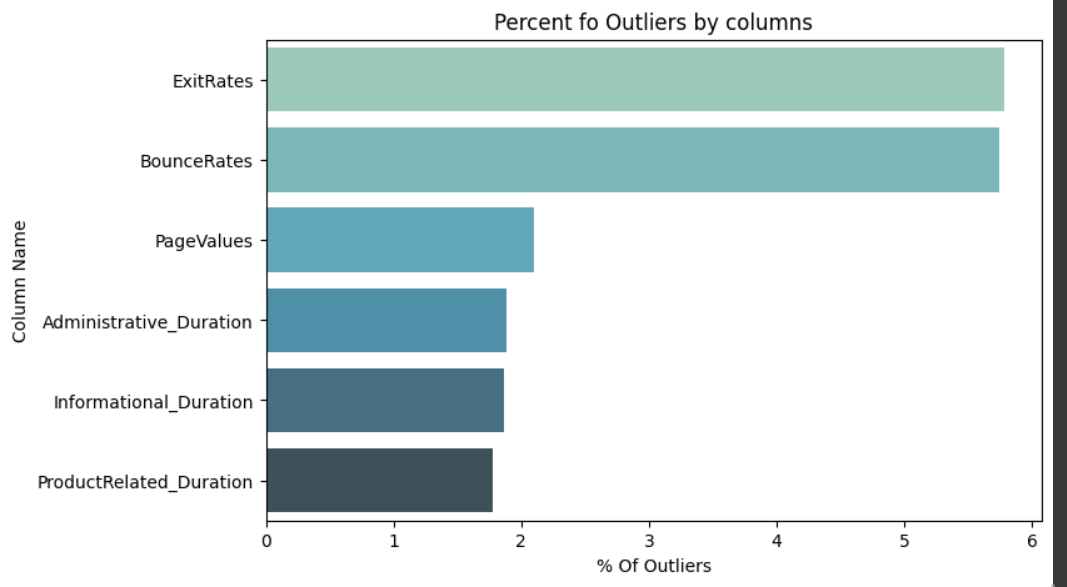
**4. Numerical Feature Analysis:**

- Two-sample t-tests and non-parametric tests indicated that several numerical features, such as administrative duration and bounce rates, differed significantly between revenue-generating and non-revenue-generating user groups.

**5. Outlier Detection:**

- Outliers were identified in several features using standard deviation calculations.

- Imputation methods were employed to address missing values.

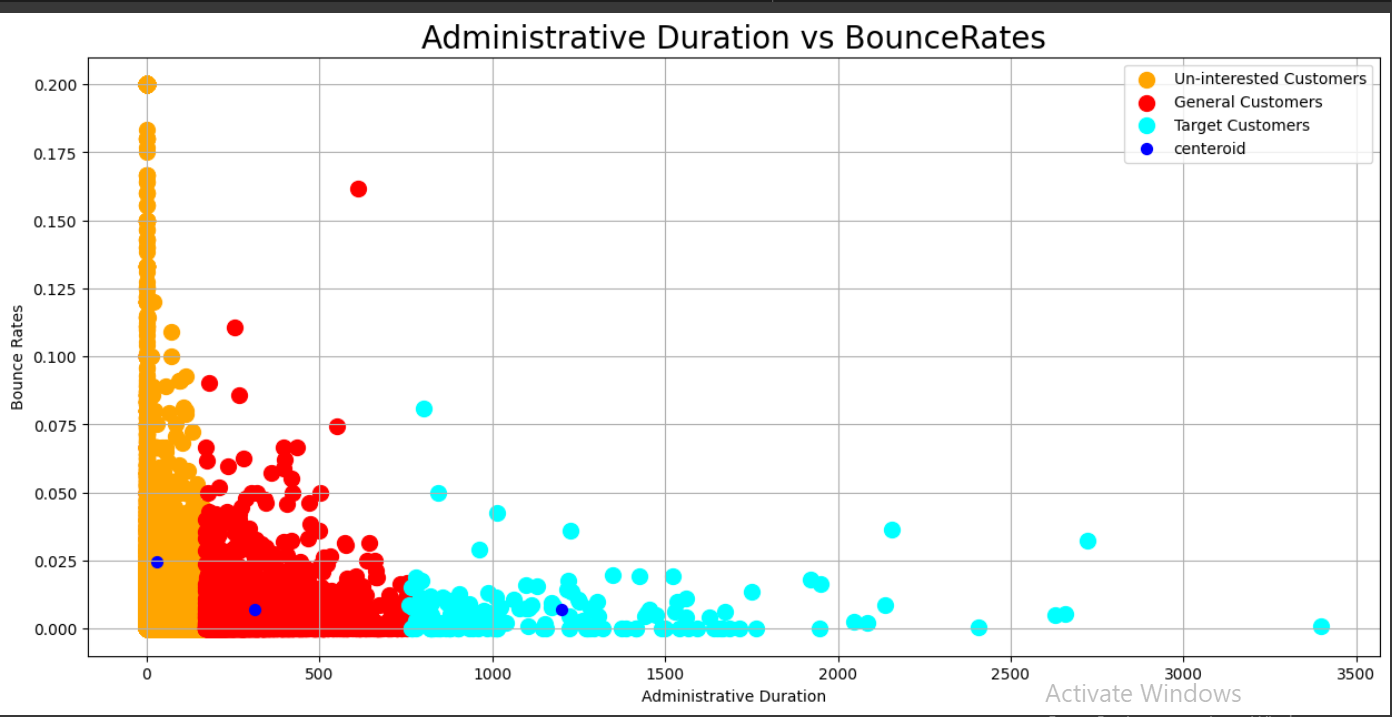


**Clustering Analysis:**

**1. Administrative Duration vs. Bounce Rates:**

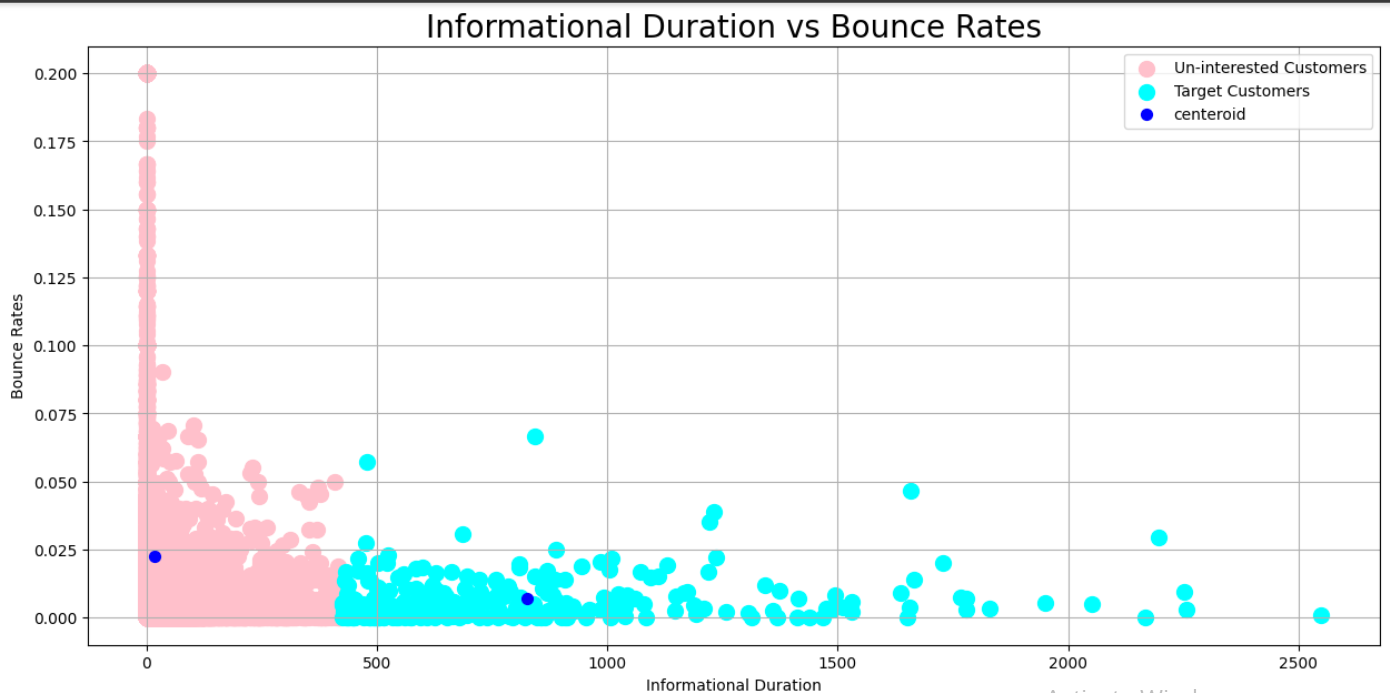
- K-means clustering identified three distinct clusters based on administrative duration and bounce rates.

- These clusters represented uninterested, general, and target customer segments.



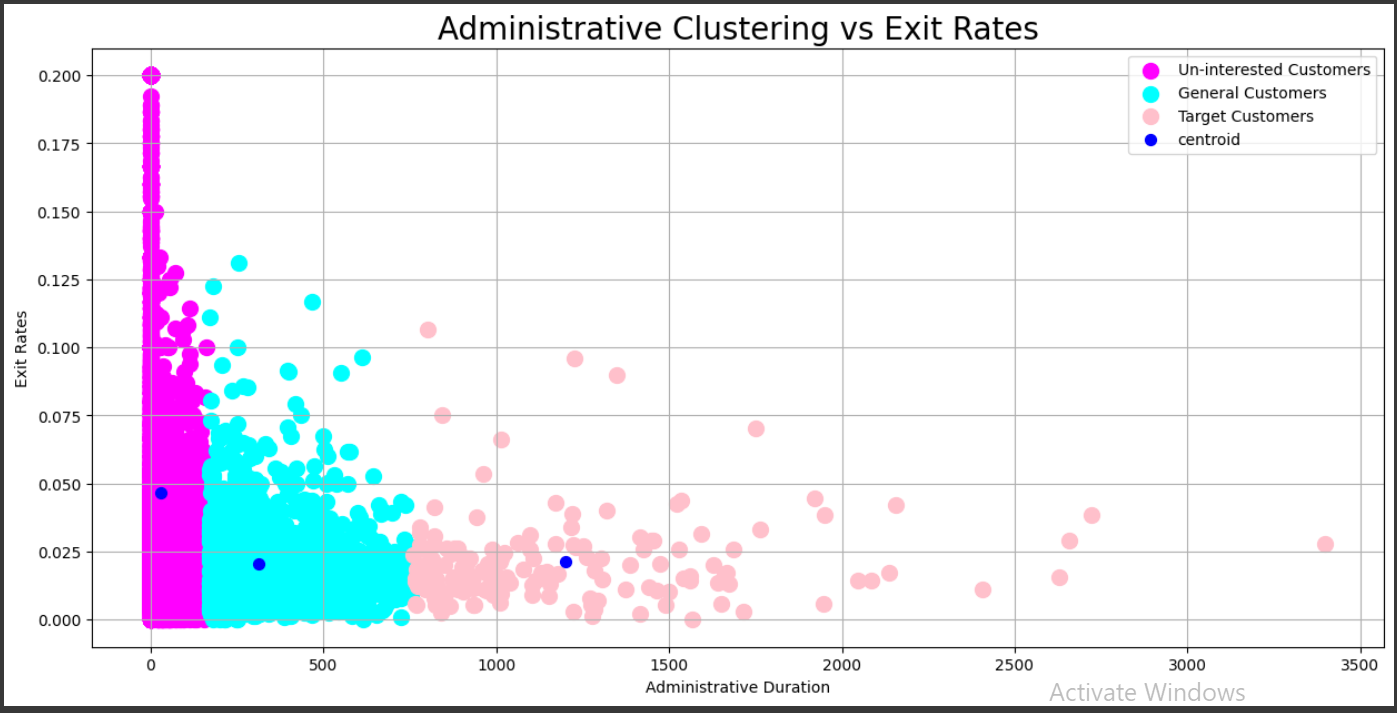
**2. Informational Duration vs. Bounce Rates:**

- Clustering analysis on informational duration and bounce rates revealed two distinct customer segments: uninterested and target customers.



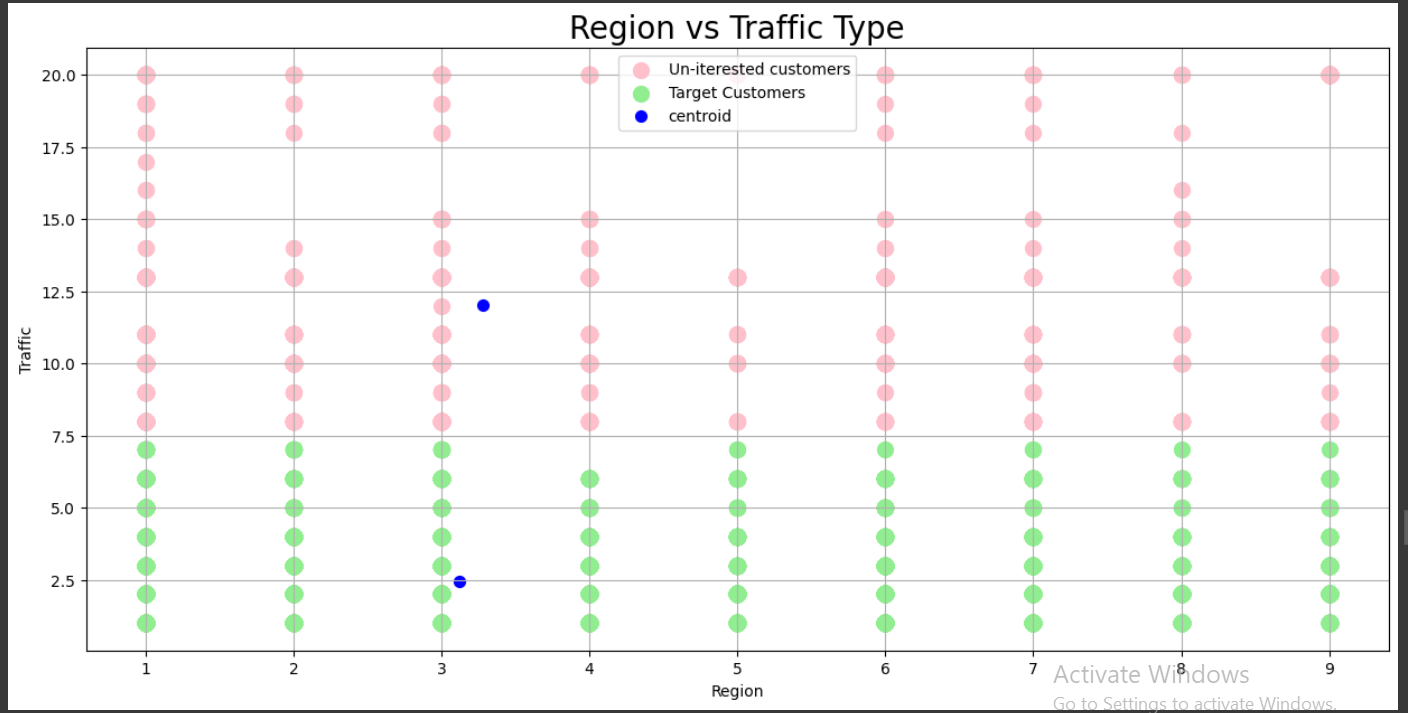
**3. Administrative Duration vs. Exit Rates:**

- Three clusters were identified based on administrative duration and exit rates, representing uninterested, general, and target customer segments.



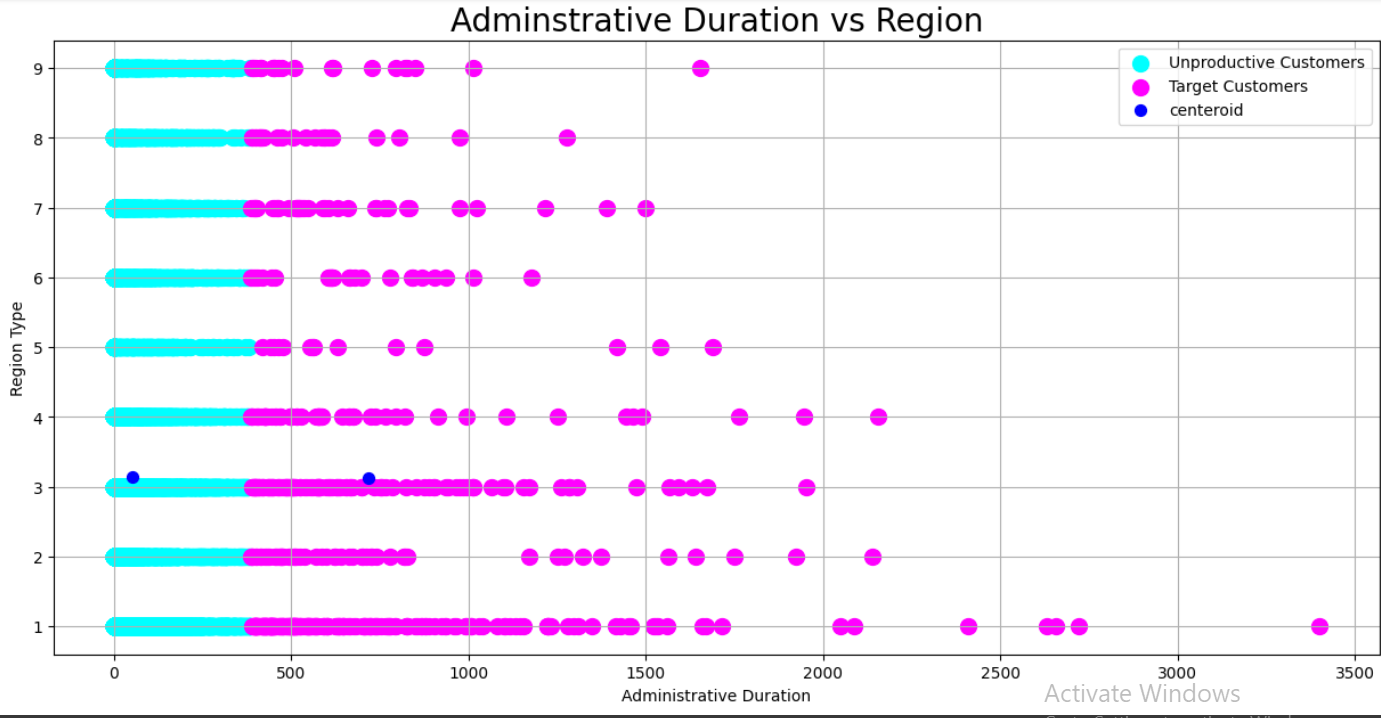
**4. Region vs. Traffic Type:**

- Clustering analysis on region and traffic type identified two customer segments: uninterested and target customers.



**5. Administrative Duration vs. Region:**

- Two distinct customer segments were identified based on administrative duration and region: unproductive and target customers.



**Modeling:**

**1. Support Vector Machine (SVM):**

- An SVM model was trained to predict revenue based on user behavior features.

- The model achieved a training accuracy of 98% and a testing accuracy of 85%.

**2. Naive Bayes:**

- A Naive Bayes model was trained to predict revenue based on user behavior features.

- The model achieved a training accuracy of 91% and a testing accuracy of 83%.

**3. Neural Network:**

- A neural network model was trained to predict revenue based on user behavior features.

- The model achieved a training accuracy of 95% and a testing accuracy of 87%.

**4. Logistic Regression:**

- A logistic regression model was trained to predict revenue based on user behavior features.

- The model achieved a training accuracy of 93% and a testing accuracy of 89%.

**Conclusion:**

The analysis provided valuable insights into user behavior patterns and their impact on revenue generation. The findings suggest that focusing on improving page views, reducing bounce rates, and targeting specific customer segments could potentially enhance the website's revenue performance. Further optimization efforts should be guided by these insights to maximize the website's revenue potential.