

E-commerce Customer Purchase Intentions

Sidney Kung and Jeffrey Muñoz



Overview

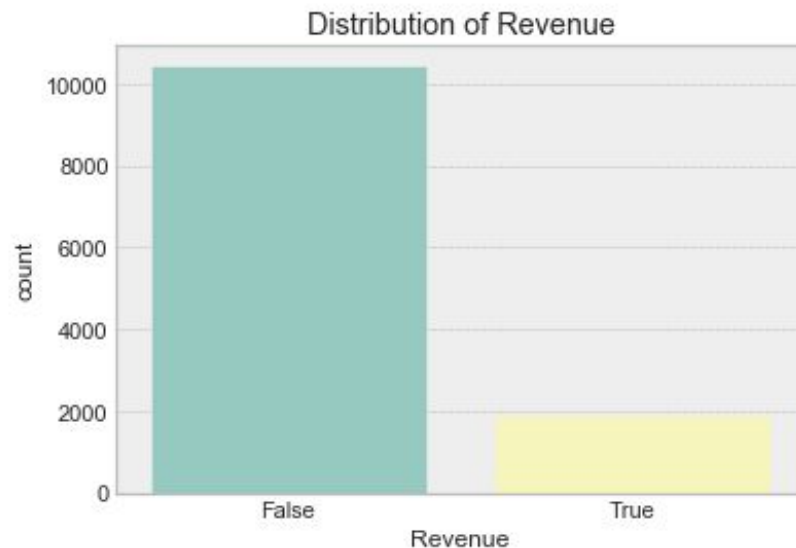
- As **e-commerce grows**, we want increase purchase likelihood
- Being able to **predict sales** would be a **beneficial tool** for any company
- We can create a **classification model** to determine these predictions
- Looking at our model results, there are clear **areas of opportunity**



Business Problem

What are the areas of opportunity to increase conversion among people who visit the website?

- To investigate this, we built a **classification model** predicting whether a site visitor will **purchase the product** or not.
- Evaluation Metrics: **Recall**, F1 Score, and AUC Score



Business Insights

1. Which features have the highest impact on predicting revenue?
2. How does the number of visitors differ each month?
3. Do returning visitors have a higher revenue than new visitors?





Dataset

- **Online Shoppers Purchasing Intention Dataset** from UCI Machine Learning Repository
- Website traffic over **one-year period**
- 12,330 Sessions
- Target Variable: **'Revenue'**
- Features of interest:
 - **'PageValues'**
 - **'BounceRates'**
 - **'VisitorType'**



A decorative graphic of a winding road with dashed white lines, curving from the bottom left towards the top right. Several colorful location pins (red, purple, yellow, and blue) are placed along the road. In the top left corner, there is a grey quarter-circle graphic.

Process

1. Gather data from UCI Repository
2. Clean data
3. Model Iterations
4. Evaluate Each Model with Metrics
5. Final Model Selection
6. Analysis

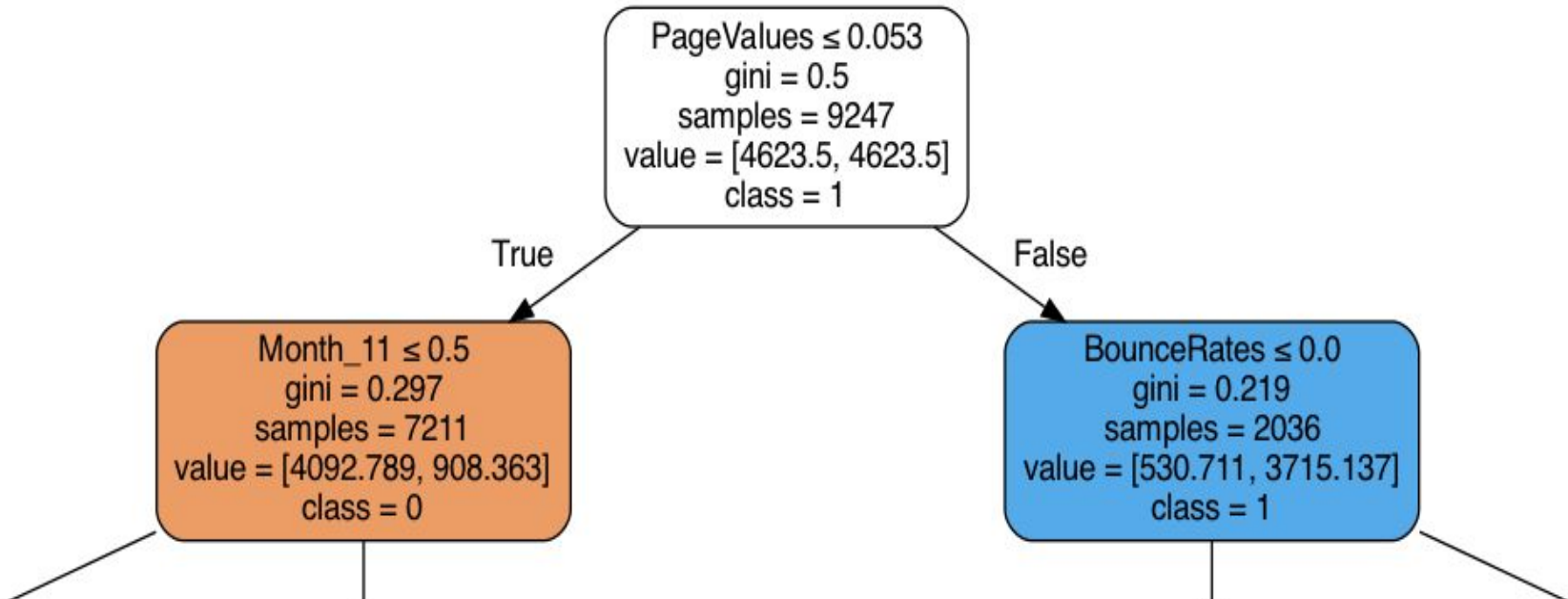


Final Model Results

	train F1 score	test F1 Score	train Recall	test Recall	test AUC
Baseline Decision Tree	1	0.57373	1	0.57606	0.74691
max_depth=3	0.65766	0.63685	0.61978	0.59229	0.77066
SMOTE on max_depth=3	0.63591	0.60802	0.4885	0.45907	0.70405
Class Balanced on max_depth=3	0.64129	0.66149	0.85724	0.86409	0.86081
RFECV on max_depth=3	0.81399	0.3538	0.896	0.8951	0.91924

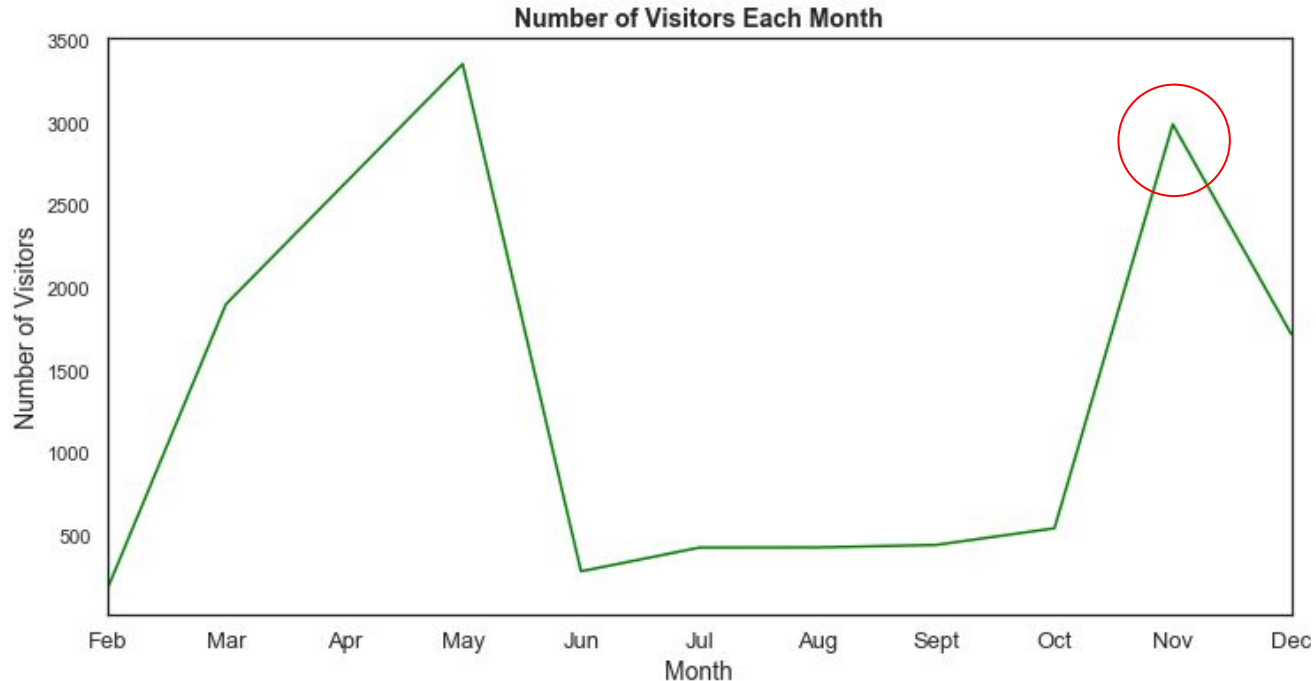
Business Insights

Question 1: Which features have the highest impact on predicting revenue?



Business Insights

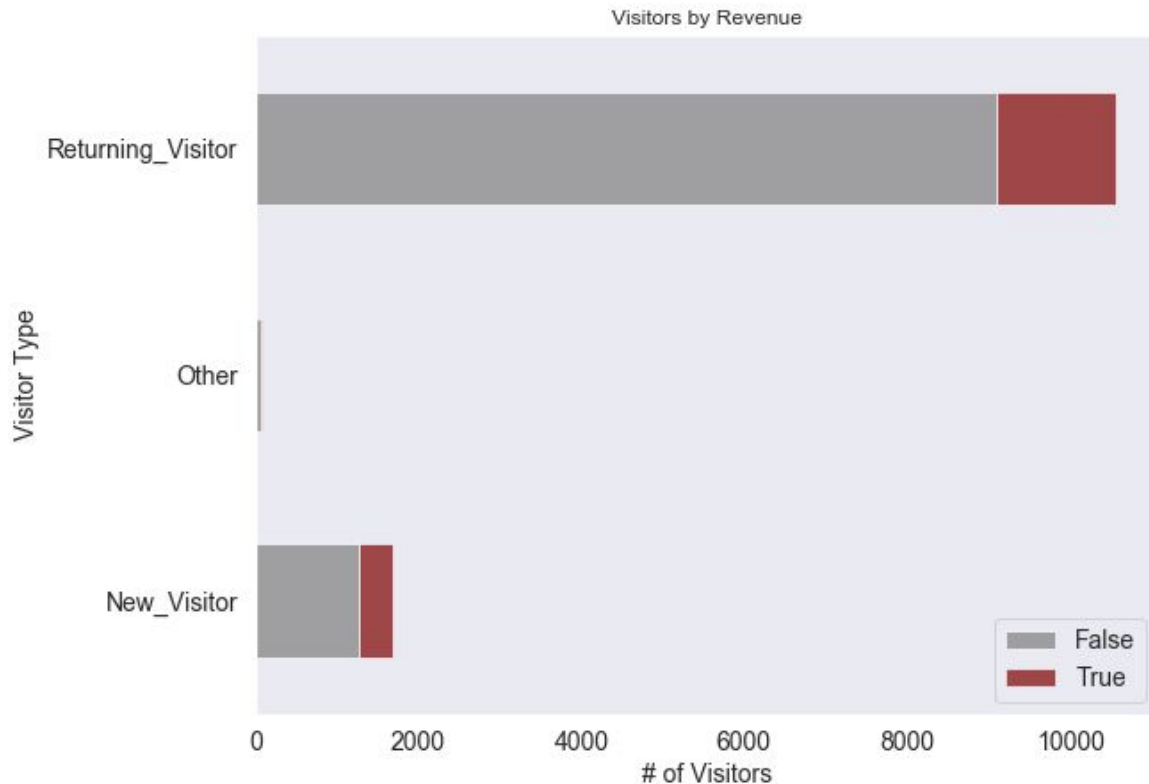
Question 2: How does the number of visitors differ each month?





Business Insights

Question 3: Do returning visitors have a higher revenue than one-time visitors?





Conclusion

Business Question: What are the areas of opportunity to increase conversion among people who visit the website?

- **Majority of site viewers** are **returning visitors** who are **not purchasing**
- Utilize personal viewer information to **target advertising** on those viewers
- Focus on **peak months** of **May** and **November**

Our classification model successfully increased the **true positive rate** to **86%**, **minimizing missed profits**.



Next Steps

- **Collect more data**, but include a legend for categorical variables to promote interpretability.
- **Webscrape** data from competing E-commerce stores to see how the model performs with that blind data.
- **A/B Testing** the web design of the prominence of Administrative and Informational page links to increase length of user session.