Trevor Maxwell

DSC 680

2024-05-05

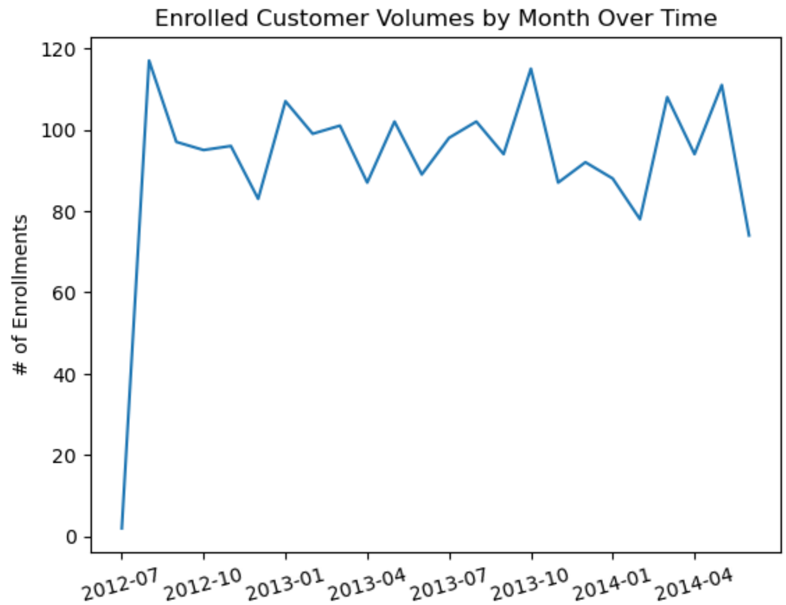
Marketing Campaign Offer Acceptance Prediction

Marketing strategy is a key component of a company's success. With the right marketing campaign aimed toward the right demographic, companies can engage with more customers, increase the number of services and products offered, increase revenue, and achieve faster growth for the company. Machine learning techniques can strengthen the effectiveness of a company's marketing strategy in different capacities. These methods can be used to develop different campaigns aimed at different markets. Wang, Tong, et al used techniques that pushed different campaigns towards the same demographic and found that campaigns with experience incentives yielded better results than campaigns with only sales incentives (2021). Machine learning can aid campaigns aimed at internal customers as well as prospective customers. A banking company used these techniques to determine which prospective customers are more likely to subscribe to a new product offering (Alexandra and Sinaga, 2021). This paper discusses using supervised machine learning techniques to predict which internal customers are more likely to subscribe to a new product after receiving 6 marketing campaigns.

The dataset used in this analysis contains attributes of customers who have purchased products from the company in the past. This company had a wide variety of different products from different foods and wines to gold products, and the attributes contain the amount spent on each of these products. There are also demographic attributes such as year of birth, income, education, and number of kids and teens living in the home. In terms of engagement with the company, some attributes that include if the customer ever had a complaint, when the customer enrolled with the company, when the customer last bought a product, the number of website visits by the customer in the last month, and the number of products purchased via the website, catalog, or in-store. All these attributes are used to determine if the customer accepts an offer by the last marketing campaign.

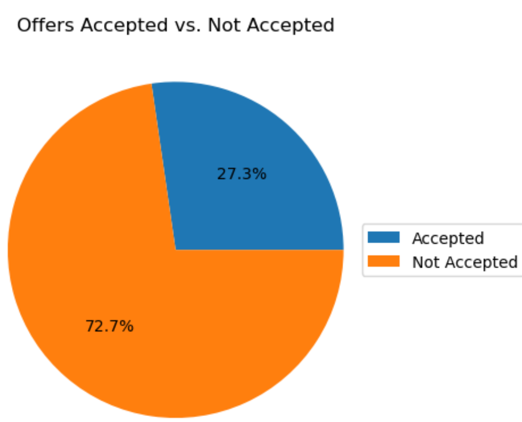
A few attributes needed cleaning to optimize the exploratory data analysis phase and perform the machine learning algorithms. Firstly, the ID column was dropped. The date the customer enrolled was transformed into the month of the year when the customer enrolled. The education feature contained two different values for a Master's education which was cleaned as well as changing the value from *Graduation* to *Bachelor's* for a better definition of the value. There were also variations within the marital status feature, and each of these values was transformed into either *In Relationship* or *Single*. Additionally, outliers were found in the income feature during the exploratory data analysis phase. These outliers were handled by removing the observations that had an income value greater than or equal to the interquartile range threshold. There were 24 observations where the income was missing, therefore these observations were also dropped which left 2,208 observations in the dataset. Lastly, there was a feature identifying at which campaign the customer accepted the offer (6 total features). These features were combined into a single attribute identifying if the customer accepted the offer at any point during the 6 marketing campaigns. This new feature became the target.

Understanding customer data is important for the company as well as the modeling techniques. It is crucial to understand the volume of customers enrolling with the company to gauge the growth. This can be observed below.



<u>Fig 1:</u> Customer enrollment volumes over time.

One goal of using machine learning methods to enhance marketing campaigns is to increase the overall number of customers enrolling with the company. Better marketing campaigns increase revenue which can fund other projects to increase overall enrollment. In terms of the number of offers accepted within this dataset, the breakdown can be observed below.



<u>Fig. 2:</u> Percentage of customers who did or did not accept the marketing offer.

The majority (72.7%) of the internal customers did not accept the marketing offer. This is important since the distribution is not even throughout the dataset and will need to be handled accordingly during the modeling phase. It is also important to understand this breakdown in terms of customer demographics. Understanding how different customers interact with the campaign offers is an important detail for the business. Below is a percentage breakdown of each education level and whether the offer was accepted.

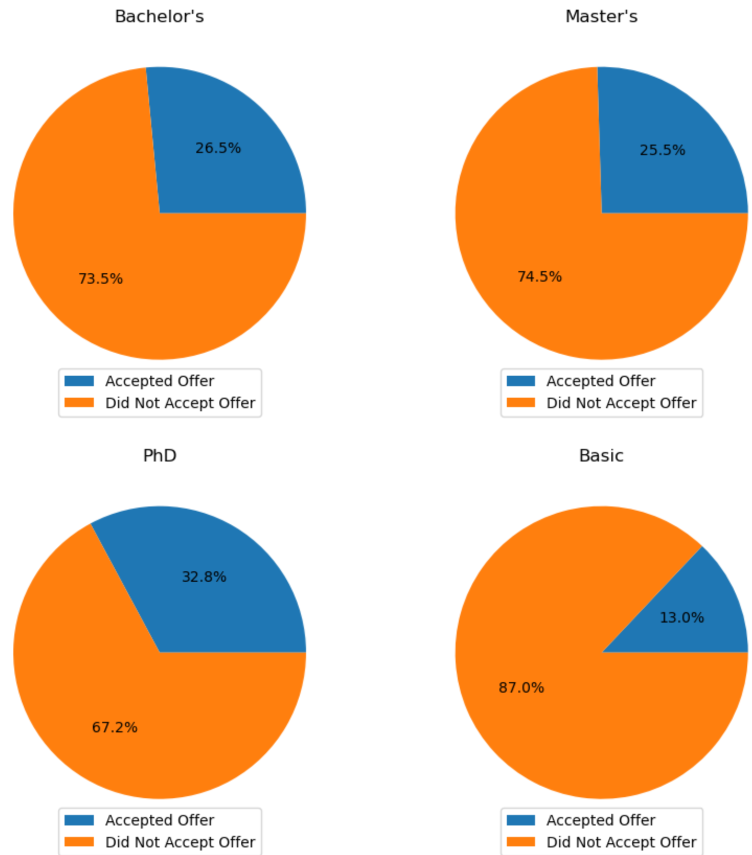
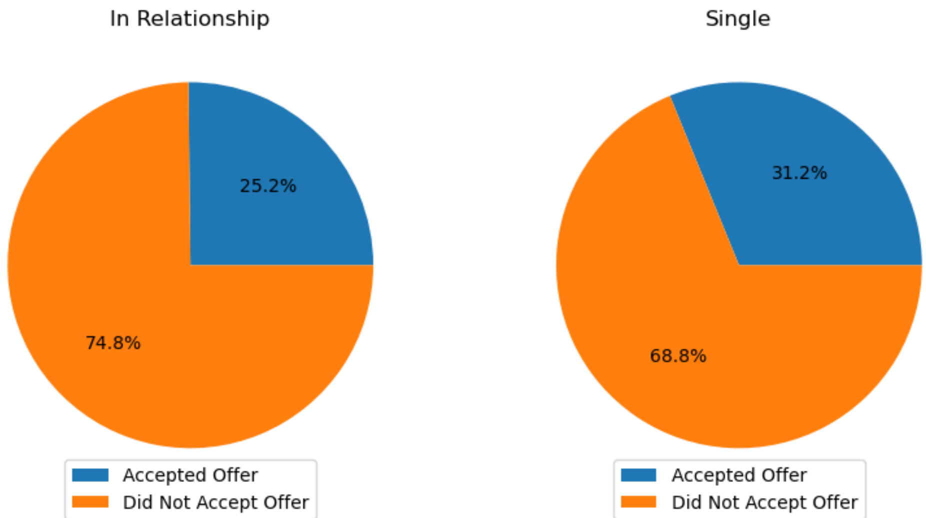


Fig 3: Percentage breakdown of offer acceptance for each education level.

The PhD education level had the highest percentage of offer acceptance compared to the other 3 education levels, and the basic education level had a significantly smaller percentage of offer acceptance compared to the other 3 levels. This is an important insight into the customer population for the business to understand. Another important insight is the marital status of the customer and its breakdown for the same reasons as the education-level insight.

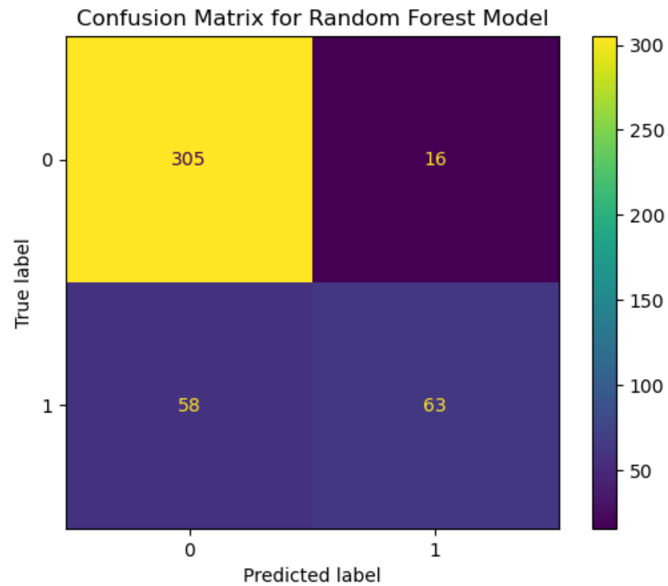


<u>Fig 4:</u> Percentage breakdown of offer acceptance for each marital status.

Customers with a *Single* marital status had a higher offer acceptance by 6% compared to those with an *In Relationship* marital status within the dataset.

Before inputting the data through the machine learning models, a min-max scaler was fitted to the data since distance-based algorithms were used. Additionally, a stratified shuffle split was used when splitting the data into training and test sets. This was performed since the breakdown of offer acceptances was strongly uneven and most of the customers did not accept the offer. This stratified shuffle split ensured the percentage of offers accepted and not accepted were the same between the training and test sets.

The dataset was fitted to 3 different models: Logistic Regression, Random Forest, and K-Nearest Neighbors. For each model, a confusion matrix was created and the accuracy, precision, recall, and F1 score were calculated. Of the three models, the Random Forest model performed the best with an accuracy of 83.3%, a precision of 79.8%, a recall of 52.1%, and an F1 score of 63.0%. This model had the highest accuracy and had the highest recall score. Recall is important due to the nature of the business problem but also due to the uneven offer acceptance breakdown within the dataset. The confusion matrix can be observed below.



<u>Fig 5:</u> Random Forest model's confusion matrix.

If only accuracy was important, a model could have predicted every customer to not accept the offer and it would have been 72.7% accurate. The precision score depicts the number of true labels predicted correctly within the dataset. Of the 79 customers predicted to accept the offer, the Random Forest model predicted 63 correctly (79.75%). The Logistic Regression and K-Nearest Neighbor models had precision scores of 69.62% and 72.73%, respectively.

This predictive analysis with the Random Forest model allows the business to make more informed decisions on who to send the marketing campaign offers to. This will lead to better response rates, more engagement with customers, an increase in the number of services and products sold, increased revenue, and faster growth for the company.

Limitations/Challenges

There was no data on when the customer made the purchase, therefore there was no way to see the sales of products or overall sales over time. This would have been a key indicator to see the company's growth over time. Another challenge was observing each numeric variable with visualizations. This would have significantly increased the exploratory data analysis phase. Instead, correlated variables and variables of interest such as income were explored.

Ethical Assessment

There are ethical considerations within this analysis. This dataset contained 6 marketing campaigns for each customer, which can be aggressive to some customers. The number of times a customer receives marketing materials should be considered.

10 Questions an Audience Would Ask You

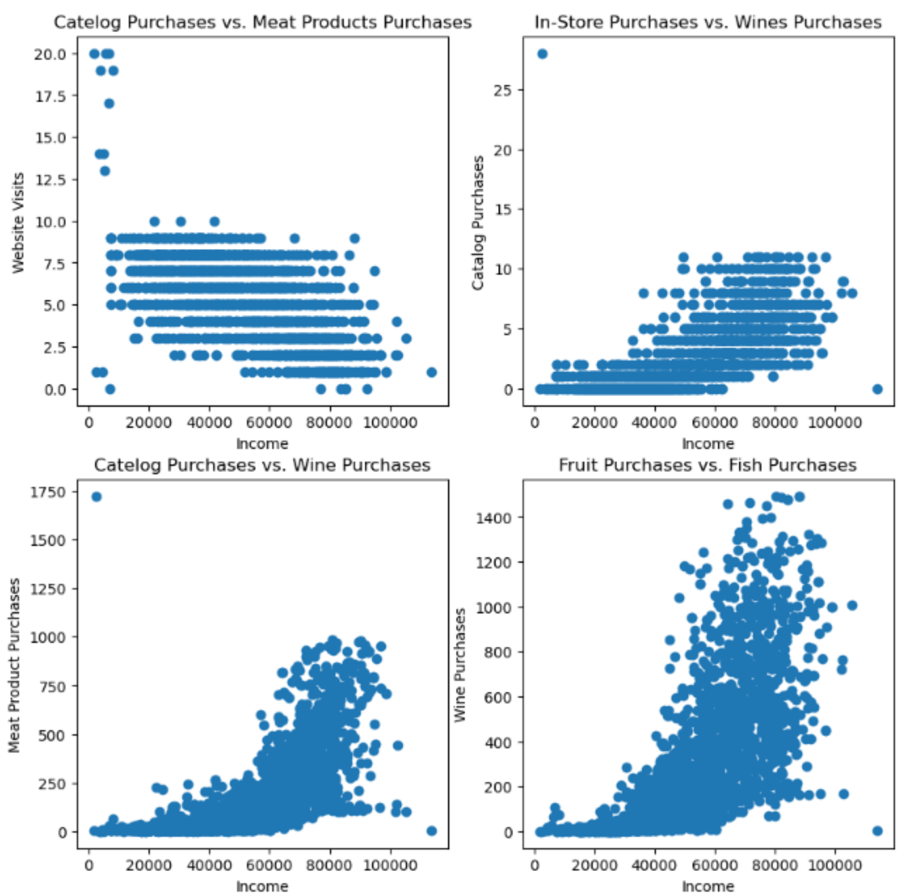
1. What are the sales volumes for the top products?

Product	Total Sales
mntwines	675871
mntmeatproducts	365088
mntgoldprods	97397
mntfishproducts	83371
mntsweetproducts	59886
mntfruits	58369

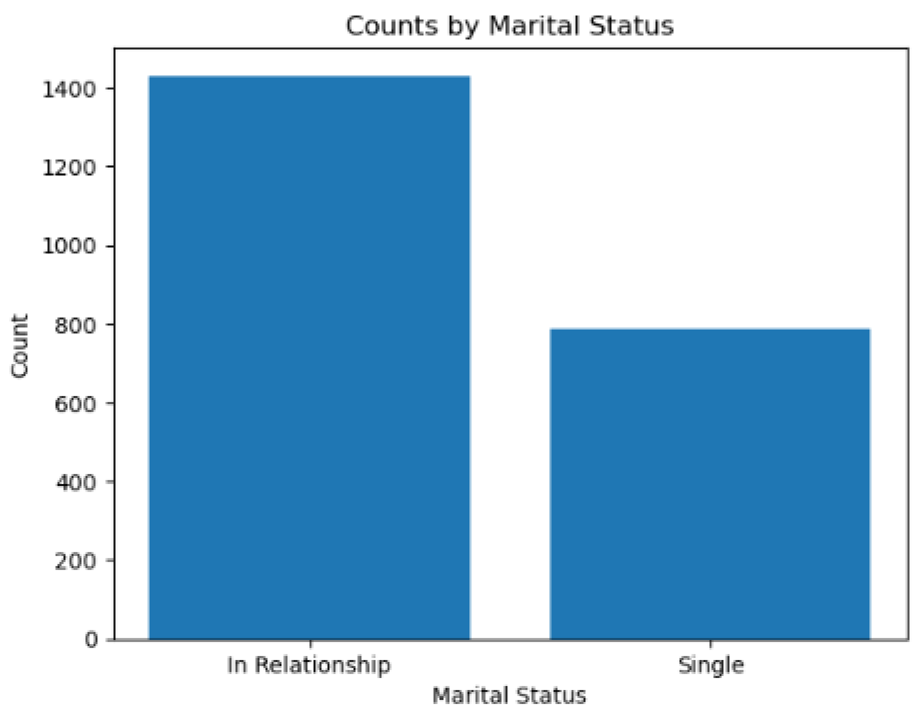
2. What are the most correlated features?

Variable 1	Variable 2	Absolute Correlation
Mntmeatproducts	Numcatalogpurchases	0.734127
Mntwines	Numstorepurchases	0.640012
Mntwines	Numcatalogpurchases	0.634753
Mntfruits	Mntfishproducts	0.594331

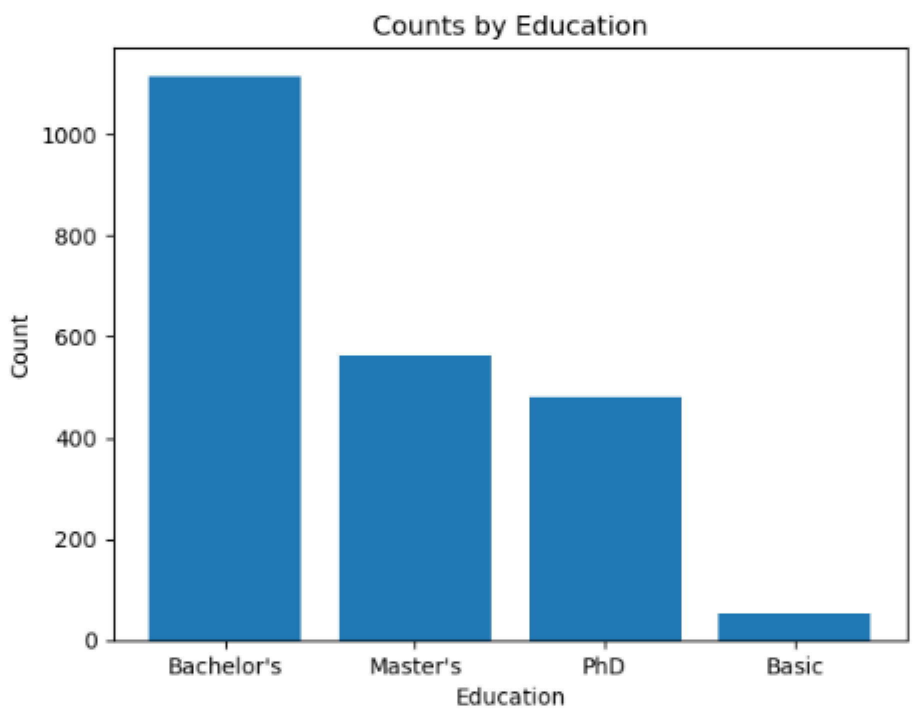
3. What features are most correlated with income?



4. What features are most correlated with the target variable?
    a. Numcatalogpurchases – 0.31
    b. Mntmeatproducts – 0.29
    c. Mntwines – 0.42
    d. Income – 0.23
5. What is the breakdown by marital status in terms of volumes?



6. What is the breakdown by education in terms of volume?



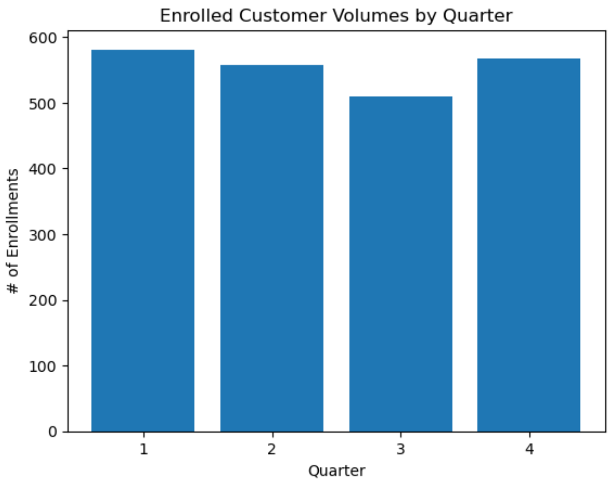
7. What is the breakdown of sales volumes for the top products by education?

education	mntwines	mntfruits	mntmeatproducts	mntfishproducts	mntsweetproducts	mntgoldprods
Bachelor's	318100	34426	199575	48442	34912	56549
Basic	391	600	618	921	654	1233
Master's	161668	13673	86473	21101	14540	24037
PhD	195712	9670	78422	12907	9780	15578

8. What is the breakdown of sales volumes for the top products by marital status?

marital_status	mntwines	mntfruits	mntmeatproducts	mntfishproducts	mntsweetproducts	mntgoldprods
In Relationship	433480	36558	228288	52745	37948	61444
Single	242391	21811	136800	30626	21938	35953

9. What are the enrollment volumes by quarter?



10. What are the purchase volumes for the methods of purchase?

method	Total Purchases
numstorepurchases	12849
numwebpurchases	9049
numcatalogpurchases	5840
numdealspurchases	5115

References

Alexandra, Jennifer, and Kristina Pestaria Sinaga. “Machine learning approaches for marketing campaign in portuguese banks.” *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 25 Oct. 2021, https://doi.org/10.1109/icoris52787.2021.9649623.

Wang, Tong, et al. “Evaluating the effectiveness of marketing campaigns for malls using a novel Interpretable Machine Learning Model.” *Information Systems Research,* vol. 33, no. 2, 21 Dec. 2021, https://doi.org/10.1287/isre.2021.1078.