"Customer Behavior Analysis and Sales Probability Modeling for Peak Shopping Months"

Report Overview

This report analyzes online shopper behavior for a newly launched e-commerce platform over the past year, with a specific focus on November and December, the busiest shopping months. The analysis aims to: Identify and profile two key customer segments — low purchase rate customers and returning customers. Explore browsing patterns and engagement behaviors for these groups. Evaluate the probability of achieving specific sales targets under a new marketing campaign scenario, to guide strategic planning for next year.

Key Results

The findings reveal significant differences between customer segments in terms of purchase rates, seasonal behavior, and engagement patterns. Correlation analysis highlights how certain browsing activities are linked, offering opportunities to optimize user journeys. A binomial probability model estimates the likelihood of achieving ambitious sales targets for returning customers, providing a realistic view of campaign potential.

Data description:

Column	Description
`SessionID`	unique session ID
`Administrative`	number of pages visited related to the customer account
`Administrative_Duration`	total amount of time spent (in seconds) on administrative pages
`Informational`	number of pages visited related to the website and the company
`Informational_Duration`	total amount of time spent (in seconds) on informational pages
`ProductRelated`	number of pages visited related to available products
`ProductRelated_Duration`	total amount of time spent (in seconds) on product-related pages
`BounceRates`	average bounce rate of pages visited by the customer
`ExitRates`	average exit rate of pages visited by the customer
`PageValues`	average page value of pages visited by the customer
`SpecialDay`	closeness of the site visiting time to a specific special day
`Weekend`	indicator whether the session is on a weekend
`Month`	month of the session date
`CustomerType`	customer type
`Purchase`	class label whether the customer make a purchase

```python
# Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

# Load and view your data
shopping_data = pd.read_csv("online_shopping_session_data.csv")
shopping_data.head()
```

...	S. ...	Adminis... ...	Administrative_Duration ...	Inform... ...	Informatio
0	1	0	0	0	
1	2	0	0	0	
2	3	0	0	0	
3	4	0	0	0	
4	5	0	0	0	

Rows: 5 Expand

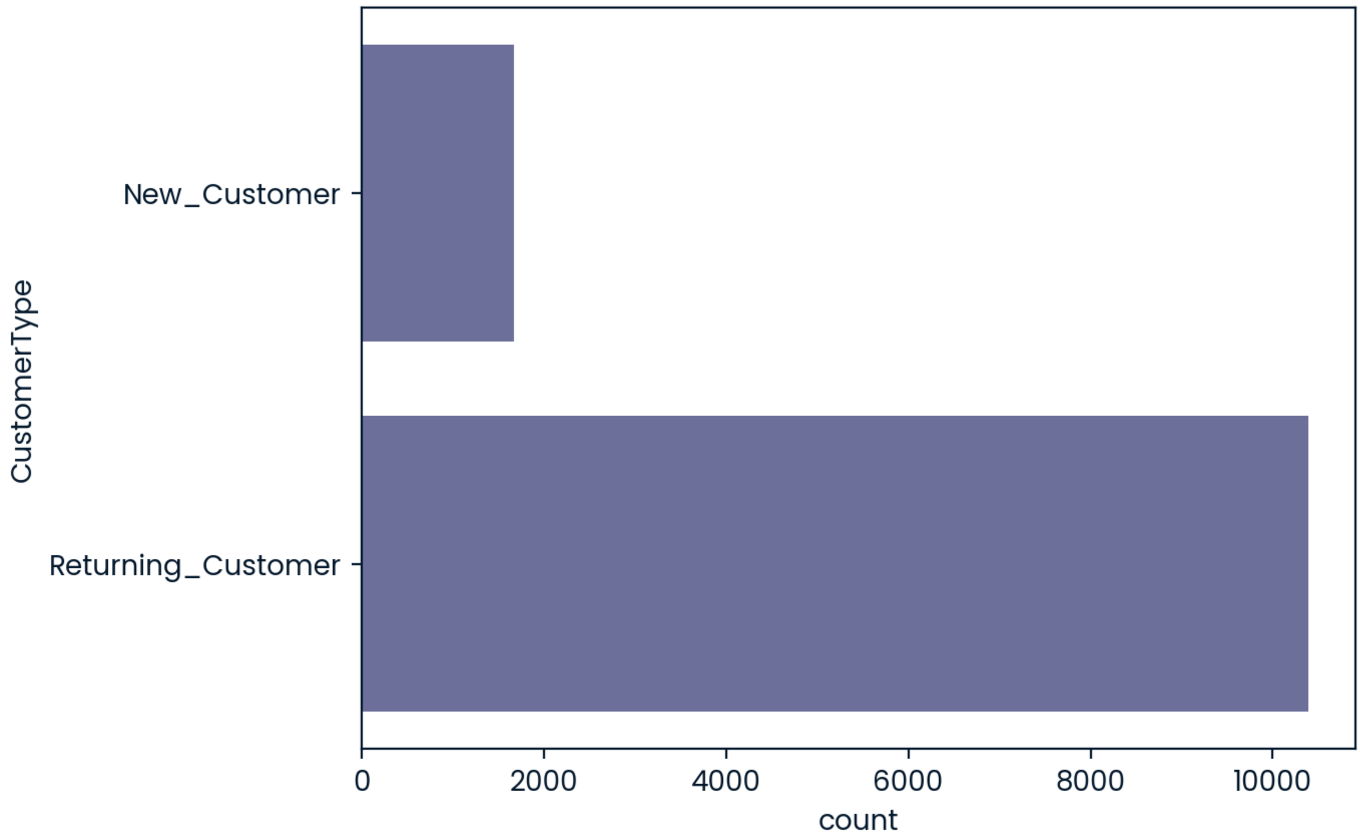
```python
print(shopping_data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12055 entries, 0 to 12054
Data columns (total 15 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   SessionID                12055 non-null  int64
 1   Administrative           12055 non-null  int64
 2   Administrative_Duration  12055 non-null  float64
 3   Informational            12055 non-null  int64
 4   Informational_Duration   12055 non-null  float64
 5   ProductRelated           12055 non-null  int64
 6   ProductRelated_Duration  12055 non-null  float64
 7   BounceRates              12055 non-null  float64
 8   ExitRates                12055 non-null  float64
 9   PageValues               12055 non-null  float64
 10  SpecialDay               12055 non-null  float64
 11  Weekend                  12055 non-null  bool
 12  Month                    12055 non-null  object
 13  CustomerType             12054 non-null  object
 14  Purchase                 12054 non-null  float64
dtypes: bool(1), float64(8), int64(4), object(2)
memory usage: 1.3+ MB
None
```

```
# Convert month names to numbers, set customer type as category, and plot customer type distribution

shopping_data['Month'] = pd.to_datetime(shopping_data['Month'], format='%b',errors='coerce').dt.month
shopping_data['CustomerType'] = shopping_data['CustomerType'].astype('category')
def bar(xa):
    sns.countplot(y=xa,data=shopping_data)
bar('CustomerType')

plt.show()
```



Overall distribution of customer types

This chart shows that Returning_Customers significantly outnumber New_Customers.

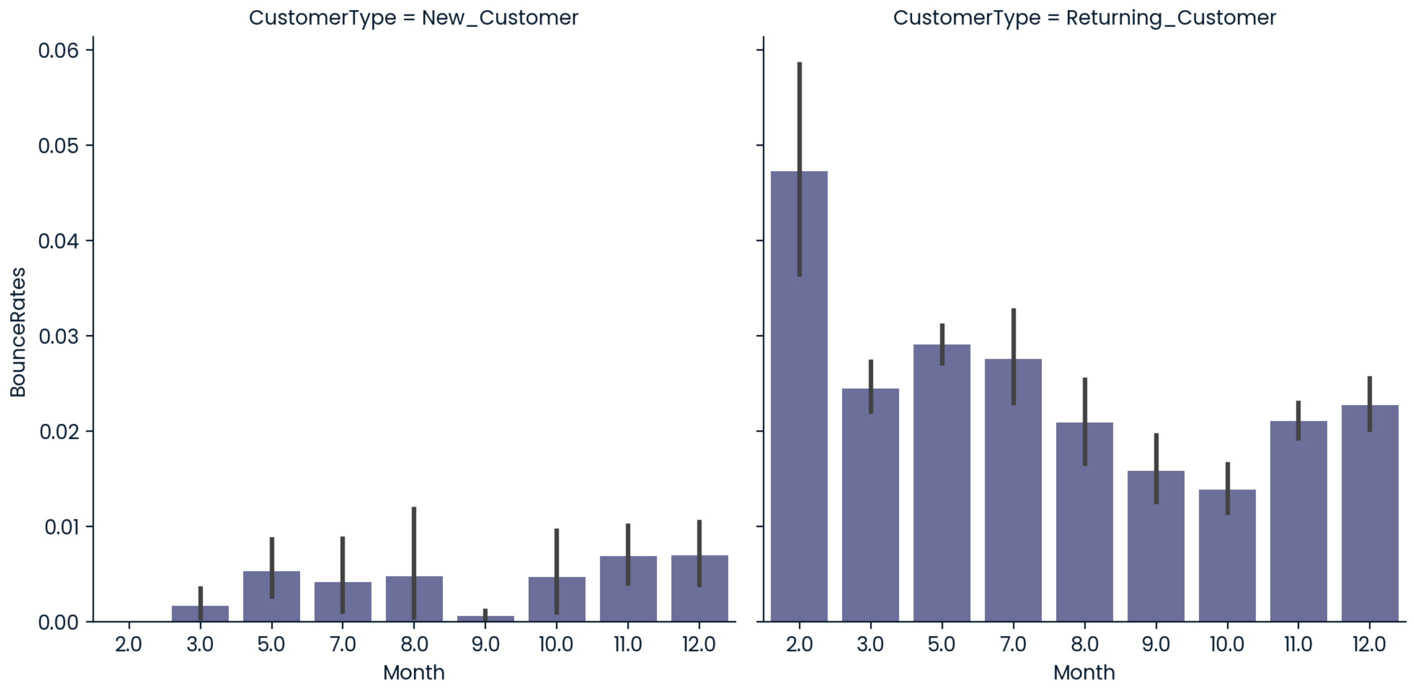
This means the business is much stronger at retaining existing customers than attracting new ones.

While this indicates strong customer loyalty, it also highlights an opportunity to improve new customer acquisition strategies.

```python
# Create a categorical plot (bar, box, etc.) with optional column/row facets

def cat(xa,ya,type,var=None,ver=None):
    sns.catplot(x=xa,y=ya,data=shopping_data,kind=type,col=var,row=ver)
    plt.show()
# Example: Bar plot of BounceRates by Month, split by CustomerType

cat('Month','BounceRates','bar','CustomerType')
```



1. New_Customers: Bounce rates are consistently very low throughout the year. This suggests new visitors are generally engaged and do not leave the site quickly. No month stands out as problematic, indicating a stable user experience for newcomers.
2. Returning_Customers: There is a clear spike in February (Month = 2) with a bounce rate around 6%. After February, the bounce rate gradually declines, hitting the lowest points in September and October (Months 9 and 10). The February spike could be due to marketing campaigns, promotions that don't resonate with returning users, or even technical issues during that period.

```python
# Calculate the average purchase rate in November and December for each customer type

purchase_rates= shopping_data[shopping_data['Month'].isin([11,
12])].groupby('CustomerType')['Purchase'].mean().to_dict()
print(purchase_rates)
```

```
{'New_Customer': 0.2733516483516483, 'Returning_Customer': 0.1955937667920473}
```

In November and December, New_Customers have a noticeably higher purchase rate (27.3%) compared to Returning_Customers (19.6%). This suggests that during the holiday season, marketing campaigns or seasonal promotions are particularly effective at converting new visitors into buyers.

For the business, this is an opportunity to:

1. Intensify acquisition campaigns in these months to attract even more new customers.
2. Analyze why returning customers purchase less in this period—perhaps they already bought earlier in the year or the offers feel less appealing to them.

```
# Find the strongest correlation with Administrative_Duration for returning customers in Nov & Dec

df_corr = shopping_data[(shopping_data['Month'].isin([11, 12])) &
(shopping_data['CustomerType'] == 'Returning_Customer')]
[['Administrative_Duration','Informational_Duration','ProductRelated_Duration']].corr().sort_values('Administrative_Duration',ascending=False)
top_correlation = {'pair':(df_corr.index[1],
df_corr.columns[0]),'correlation':df_corr.iloc[1,0]}
print(top_correlation)
```

```
{'pair': ('ProductRelated_Duration', 'Administrative_Duration'), 'correlation':
0.41689293883422823}
```

Among Returning_Customers in November and December, the strongest positive relationship with Administrative_Duration is observed for ProductRelated_Duration (correlation ≈ 0.42). This indicates that customers who spend more time on administrative pages (e.g., account, policy, or checkout info) also tend to spend more time browsing product-related pages.

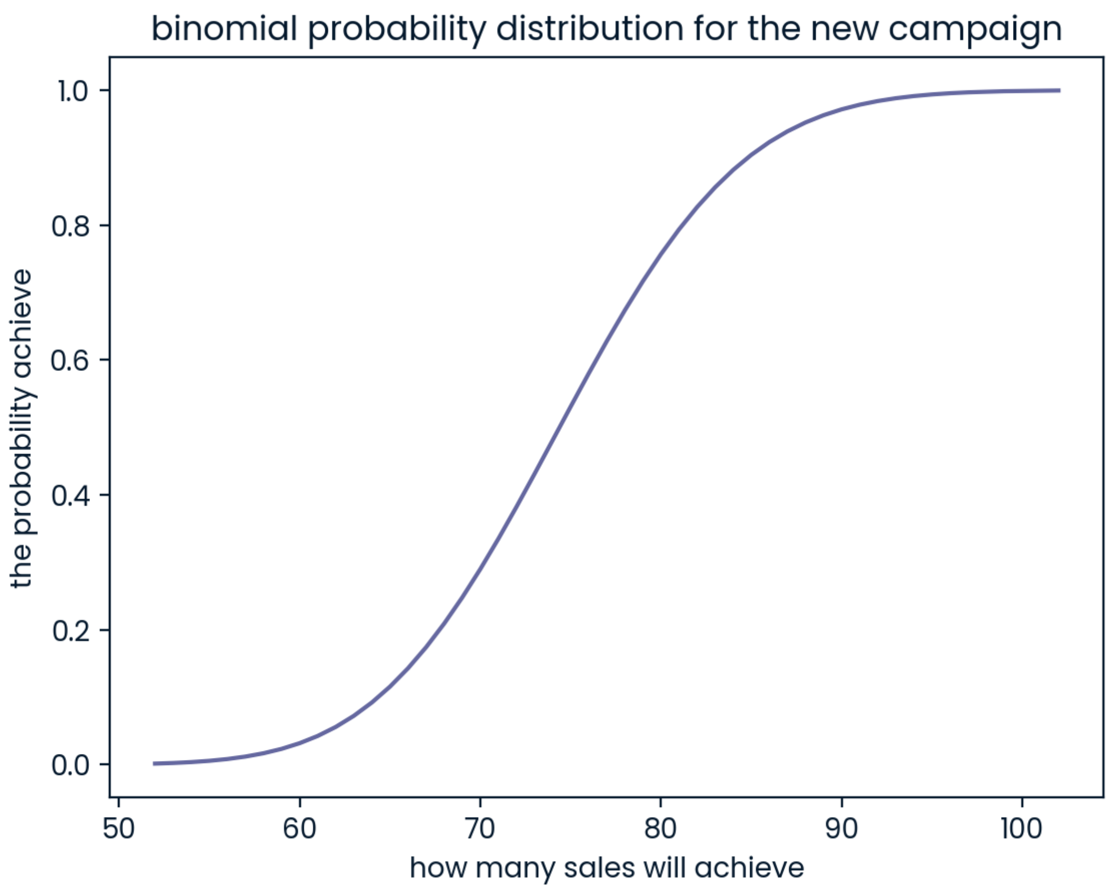
- Implication: Improving the usability and relevance of administrative sections could indirectly encourage deeper product exploration, potentially boosting purchases for returning customers during the holiday season.

```python
# Calculate and plot the binomial probability distribution for achieving sales in the new campaign

prob_at_least_100_sales = stats.binom.cdf(100,500,0.15)
print(prob_at_least_100_sales)
values=stats.binom.rvs(500,0.15,size=1000)
prob= [stats.binom.cdf(_,500,0.15) for _ in values]
g= sns.lineplot(x=values, y= prob)
g.set_title('binomial probability distribution for the new campaign')
g.set_xlabel('how many sales will achieve')
g.set_ylabel('the probability achieve')
```

```
0.9989564851276144
```

```
Text(0, 0.5, 'the probability achieve')
```



The cumulative binomial probability shows that reaching 100 or more sales out of 500 sessions for returning customers, even with the new campaign boosting the purchase rate to 15%, is highly unlikely. The curve rises steeply around the expected value of ~75 sales and flattens near 100%, indicating that most outcomes fall well below the 100-sales mark.

This means the target of 100+ sales sits at the far tail of the distribution. Achieving it would require exceptional performance, so strategies like increasing traffic or further improving the conversion rate would

Results

- Customer Type Distribution Returning customers make up a larger share of sessions compared to new customers. This indicates strong customer loyalty, but also suggests slower new customer acquisition.
- Bounce Rate Trends New customers show consistently low bounce rates year-round, meaning they are engaged once they visit. Returning customers show a notable spike in bounce rate in February (~6%), followed by a decline toward the end of the year.
- Purchase Rates in November–December New customers: 27.3% Returning customers: 19.6% New customers are more likely to make a purchase in the holiday season, suggesting that holiday promotions are more effective on them.
- Browsing Behavior Correlation (Returning Customers, Nov–Dec) The strongest positive correlation is between Administrative_Duration and ProductRelated_Duration ($r \approx 0.42$). This suggests that returning customers who spend more time on administrative pages also spend more time exploring products.
- Probability Analysis for New Campaign With a 15% purchase rate, the probability of achieving at least 100 sales out of 500 sessions for returning customers is very low. The binomial distribution centers around 75 sales, and probabilities drop sharply beyond that point.

Recommendations

- Boost Returning Customer Purchases in Holidays Since returning customers purchase less than new customers in Nov–Dec, holiday offers could be better tailored to their preferences (e.g., loyalty rewards, exclusive discounts).
- Leverage Administrative–Product Link Improve navigation from administrative pages (account, order history, etc.) directly to relevant products to capitalize on the positive browsing correlation.
- Address February Bounce Spike Investigate February campaigns or site changes to identify why returning customers disengage during that month.
- Increase Volume or Conversion for Ambitious Targets Given the low probability of reaching 100+ sales from 500 sessions at 15% conversion, either:
- Increase traffic volume (more sessions), or Increase conversion rate (above 15%) through stronger campaigns. Improve New Customer Acquisition
- Since new customers respond well during holiday months, expand acquisition campaigns during that period to grow overall revenue.