

# online shoppers purchasing intention ~~prediction~~

@Shubham SAINI  
@Manisha RAWLA  
@Alina TENNE

# Dataset Preview

	Administrative	Administrative_Duration	Informational	Informational_Duration	ProductRelated	ProductRelated_Duration	BounceRates	ExitRates	PageValues	SpecialDay	M
0	0	0.0	0	0.0	1	0.000000	0.20	0.20	0.0	0.0	0.0
1	0	0.0	0	0.0	2	64.000000	0.00	0.10	0.0	0.0	0.0
2	0	0.0	0	0.0	1	0.000000	0.20	0.20	0.0	0.0	0.0
3	0	0.0	0	0.0	2	2.666667	0.05	0.14	0.0	0.0	0.0
4	0	0.0	0	0.0	10	627.500000	0.02	0.05	0.0	0.0	0.0

	Month	OperatingSystems	Browser	Region	TrafficType	VisitorType	Weekend	Revenue
0	Feb	1	1	1	1	Returning_Visitor	False	False
0	Feb	2	2	1	2	Returning_Visitor	False	False
0	Feb	4	1	9	3	Returning_Visitor	False	False
0	Feb	3	2	2	4	Returning_Visitor	False	False
0	Feb	3	3	1	4	Returning_Visitor	True	False

2



## Datasets informations and features

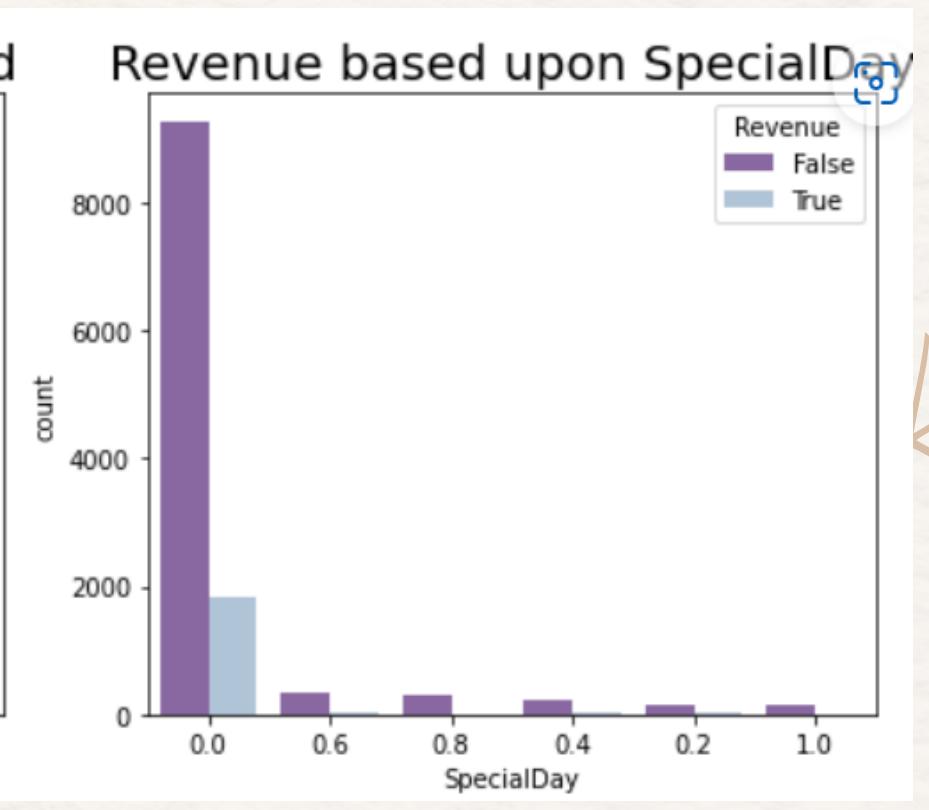
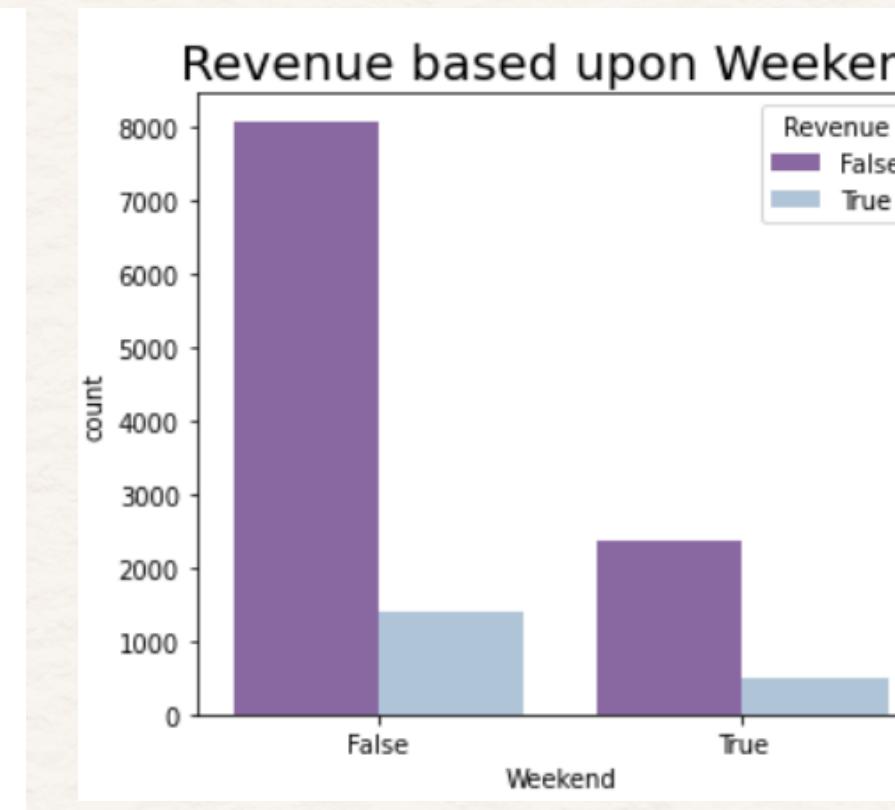
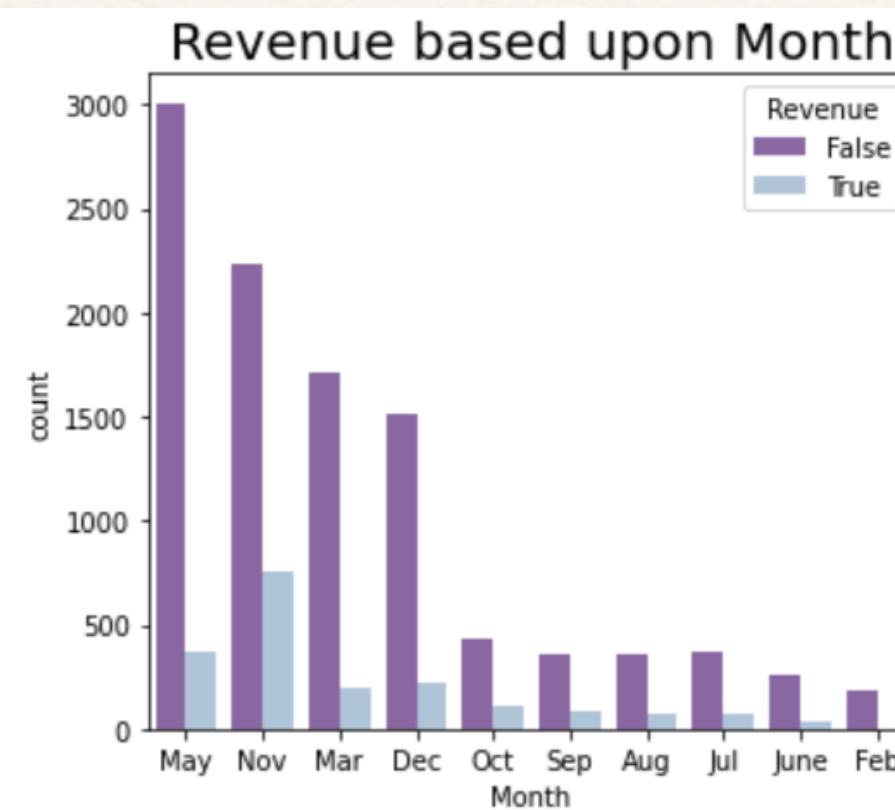
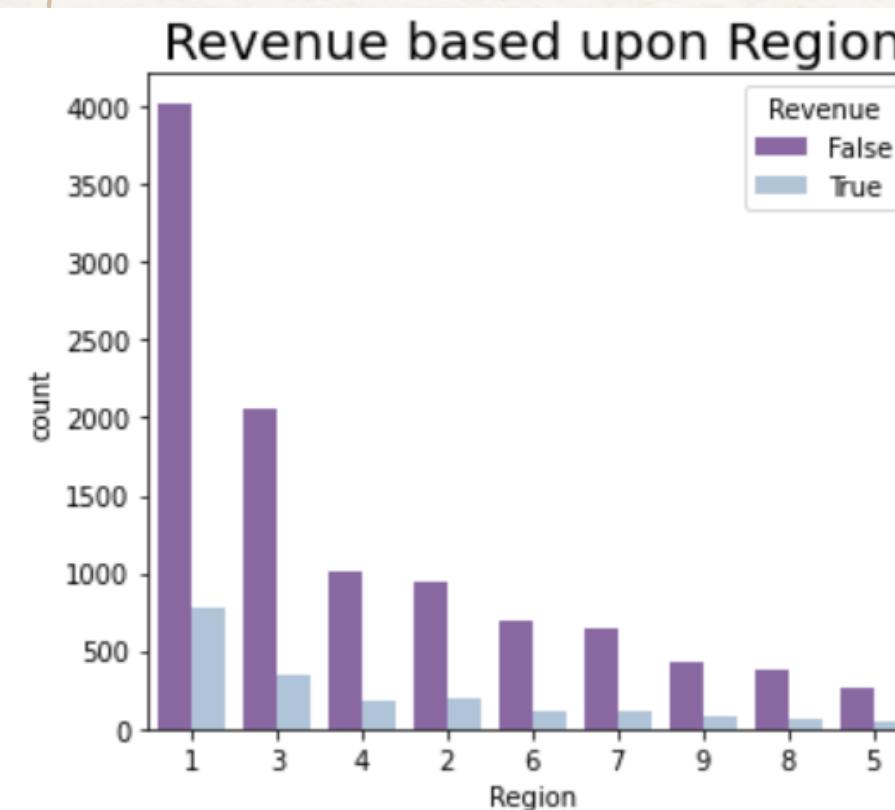
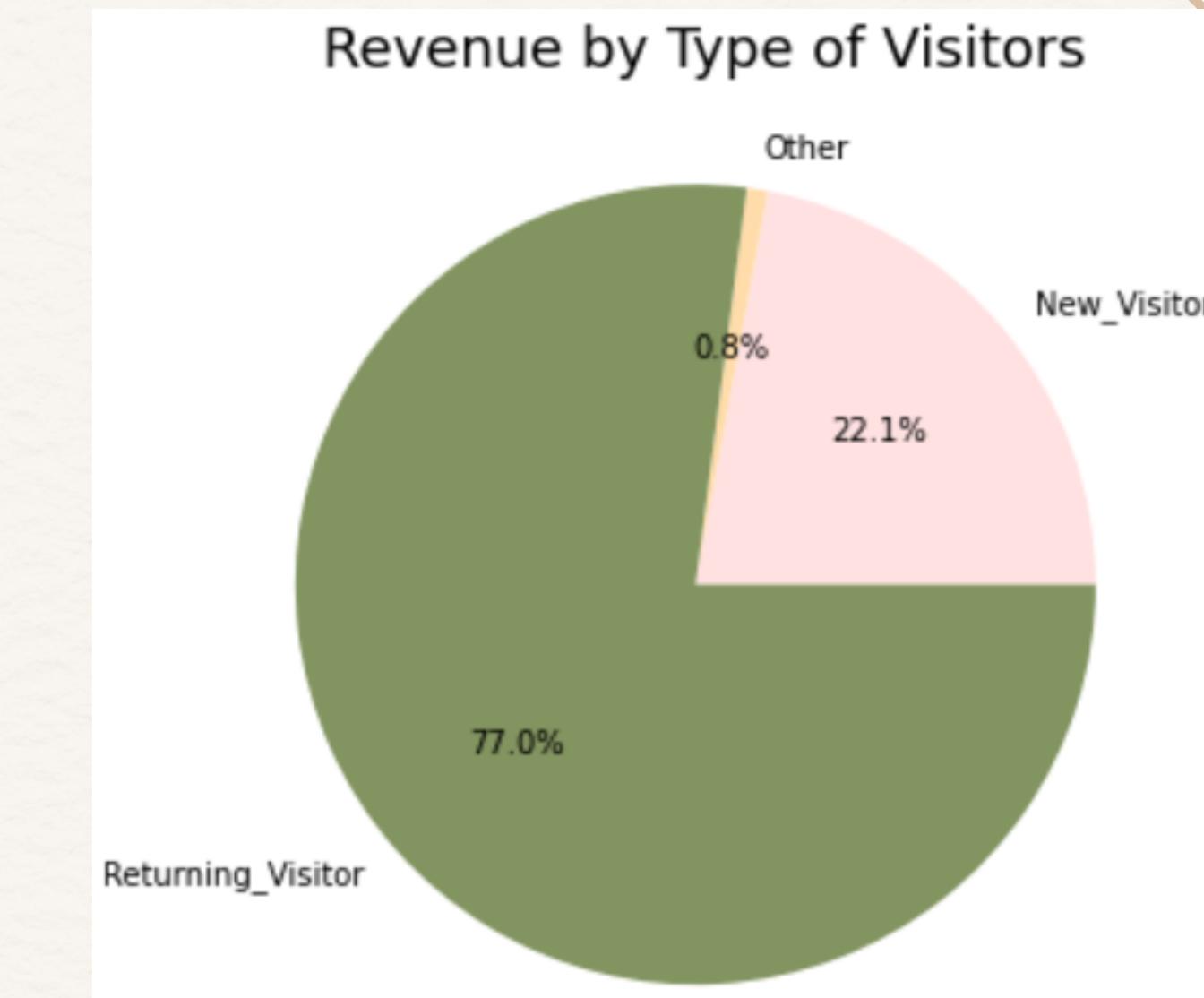
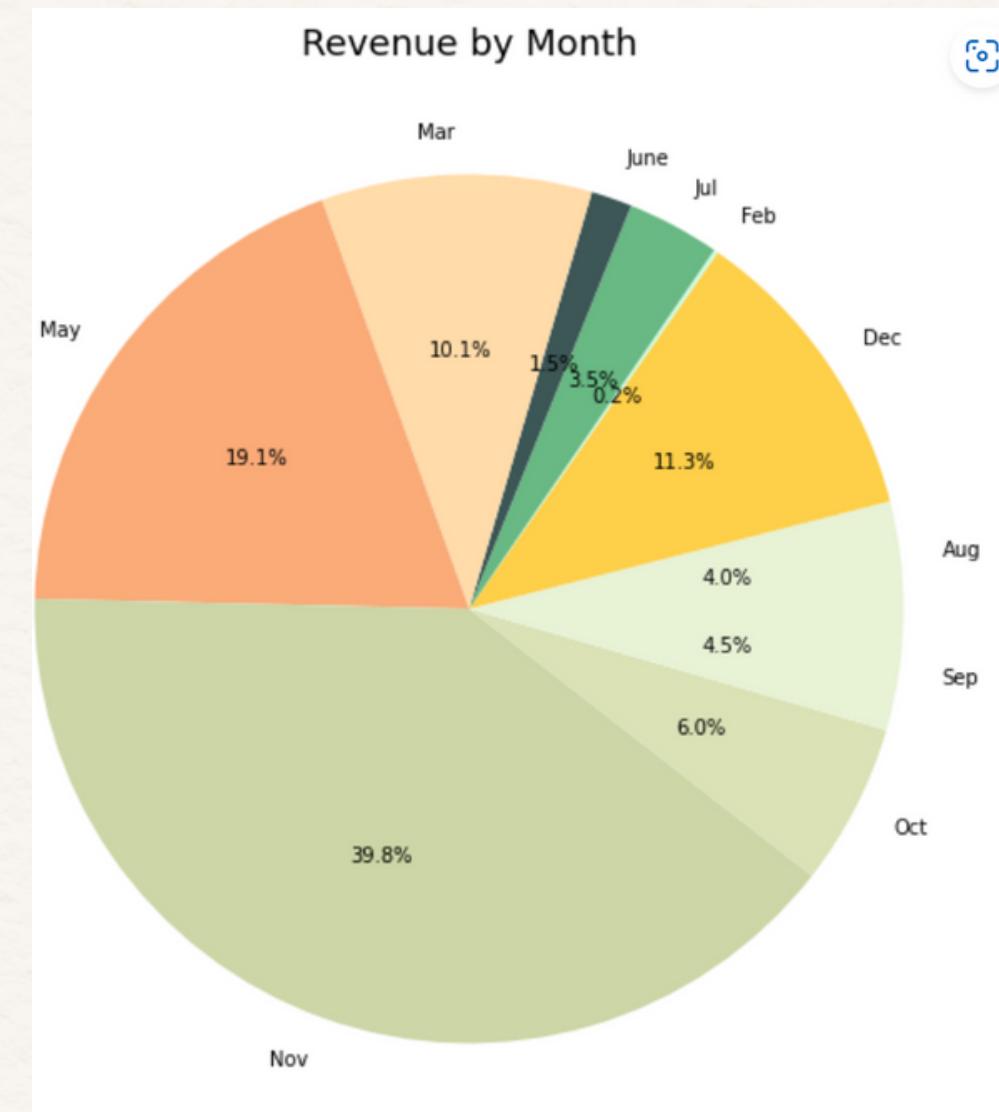
---

The datasets have 10 numerical & 8 categorical  
features total of 18 attributes.

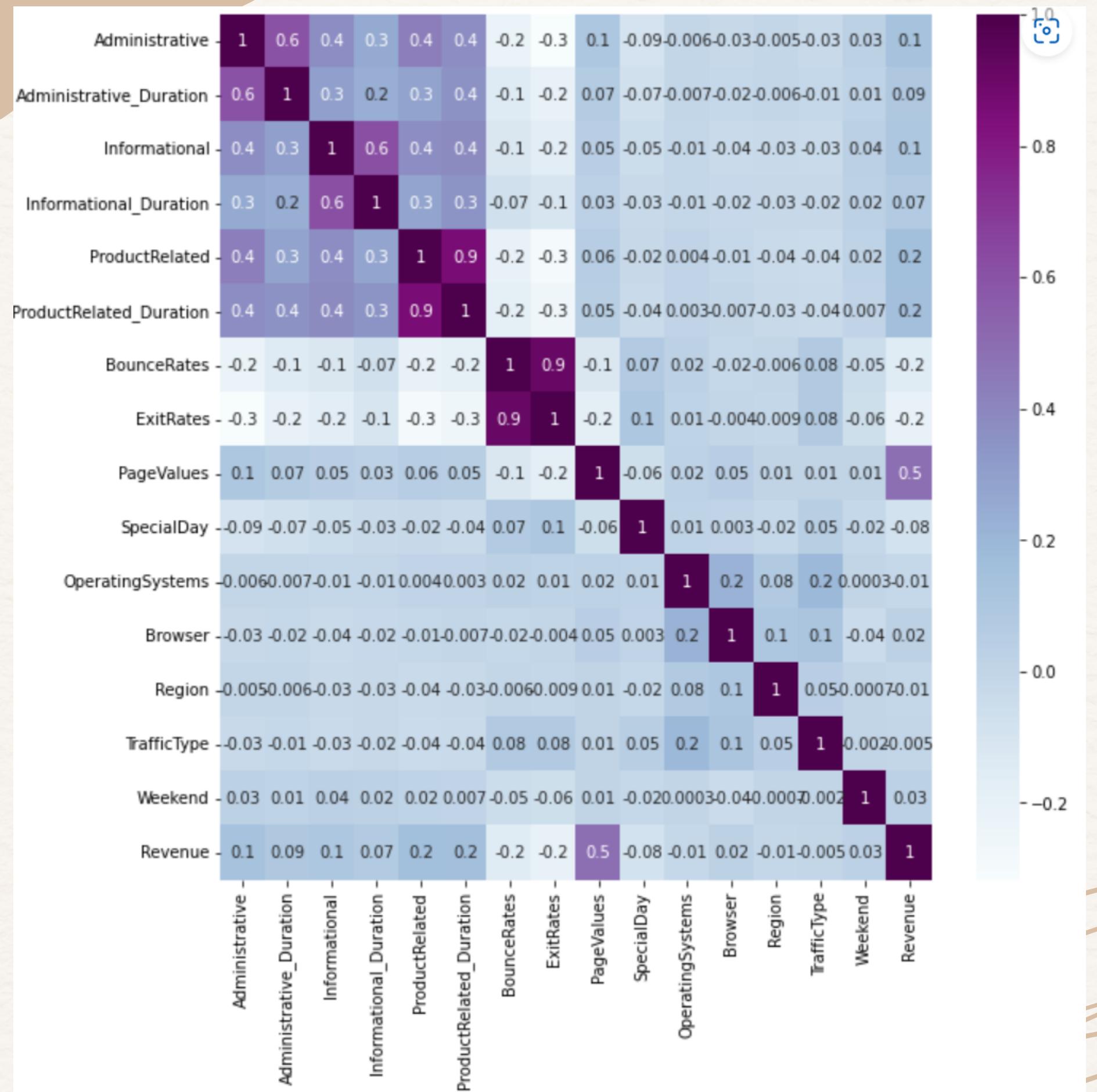
Datasets don't have any missing values.

The dataset has no mission values but has  
unbiased.

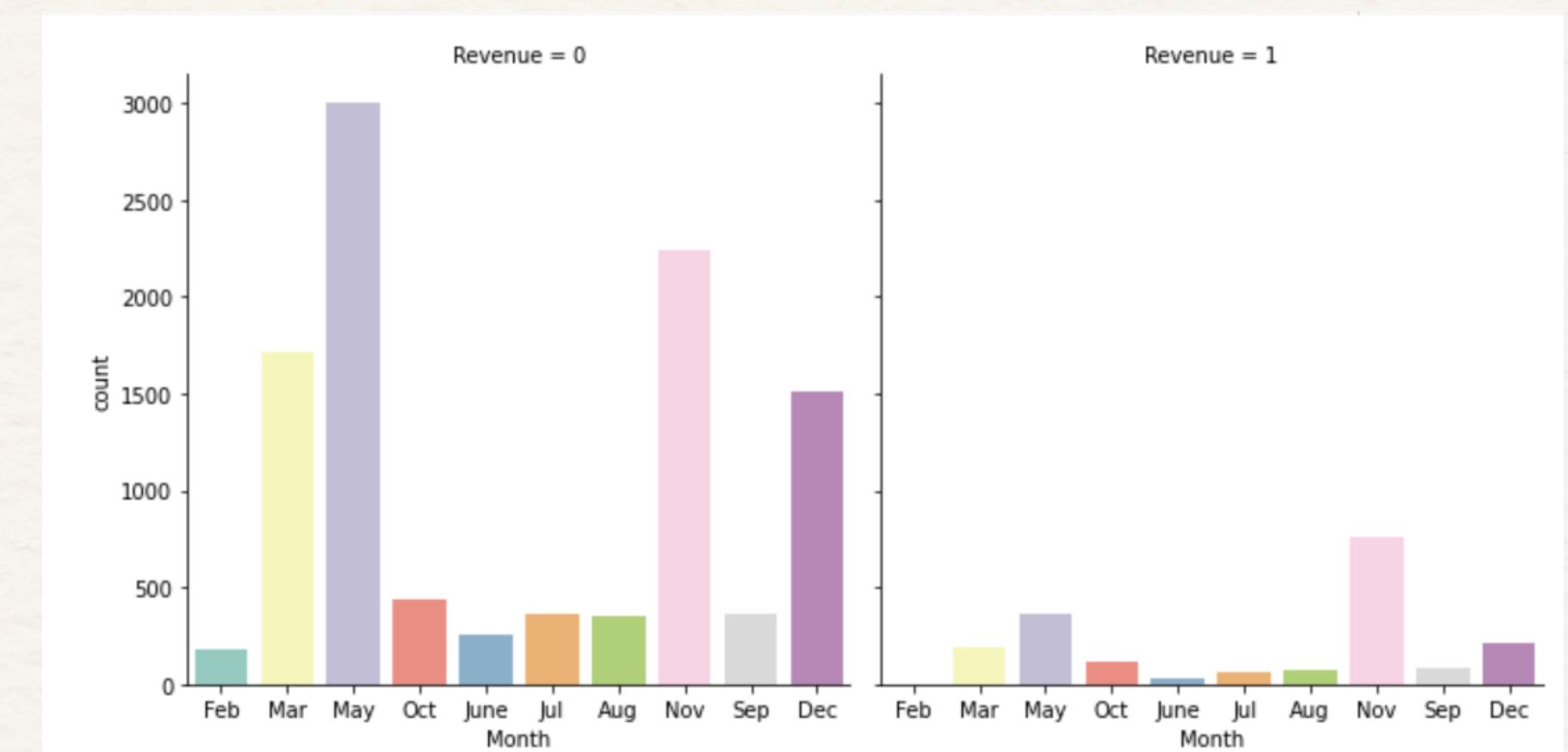
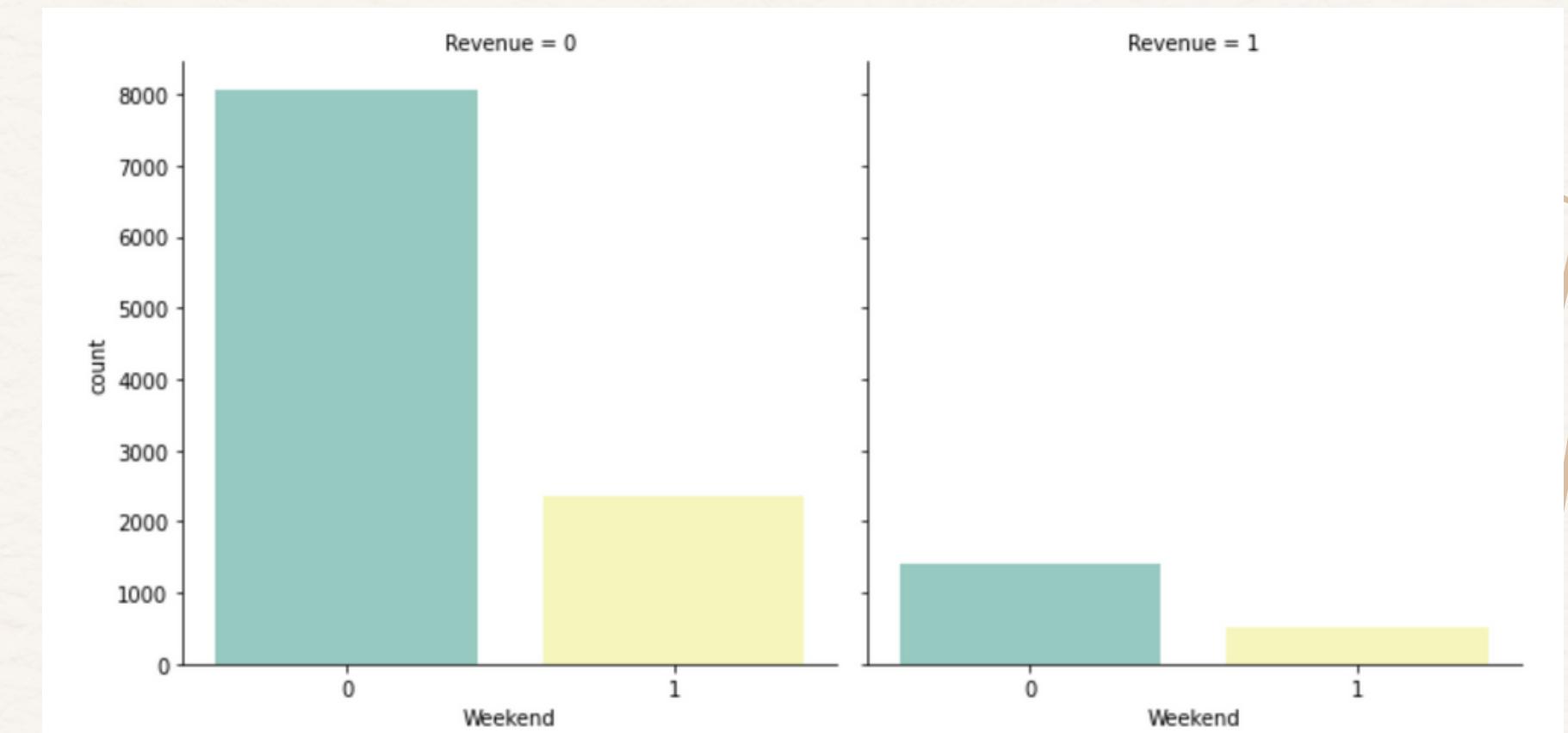
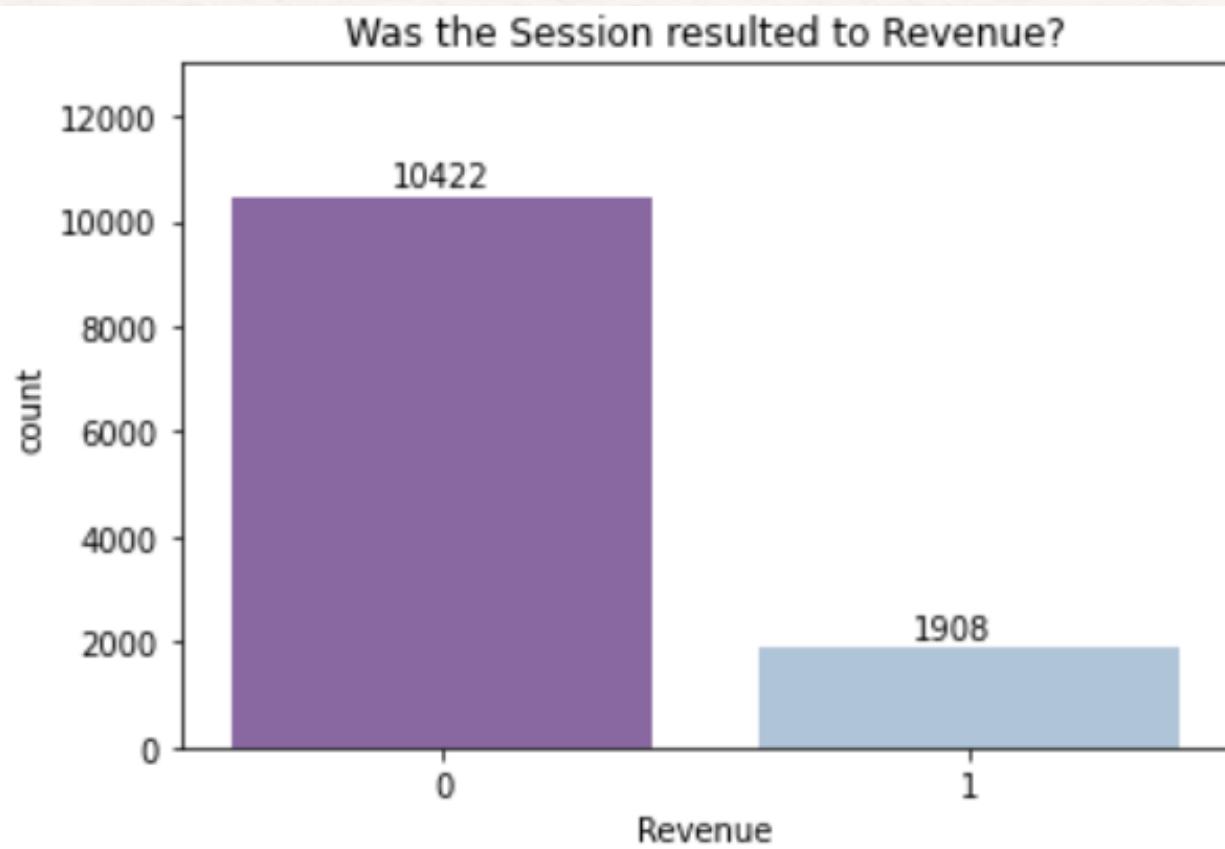
# Revenue Analysis



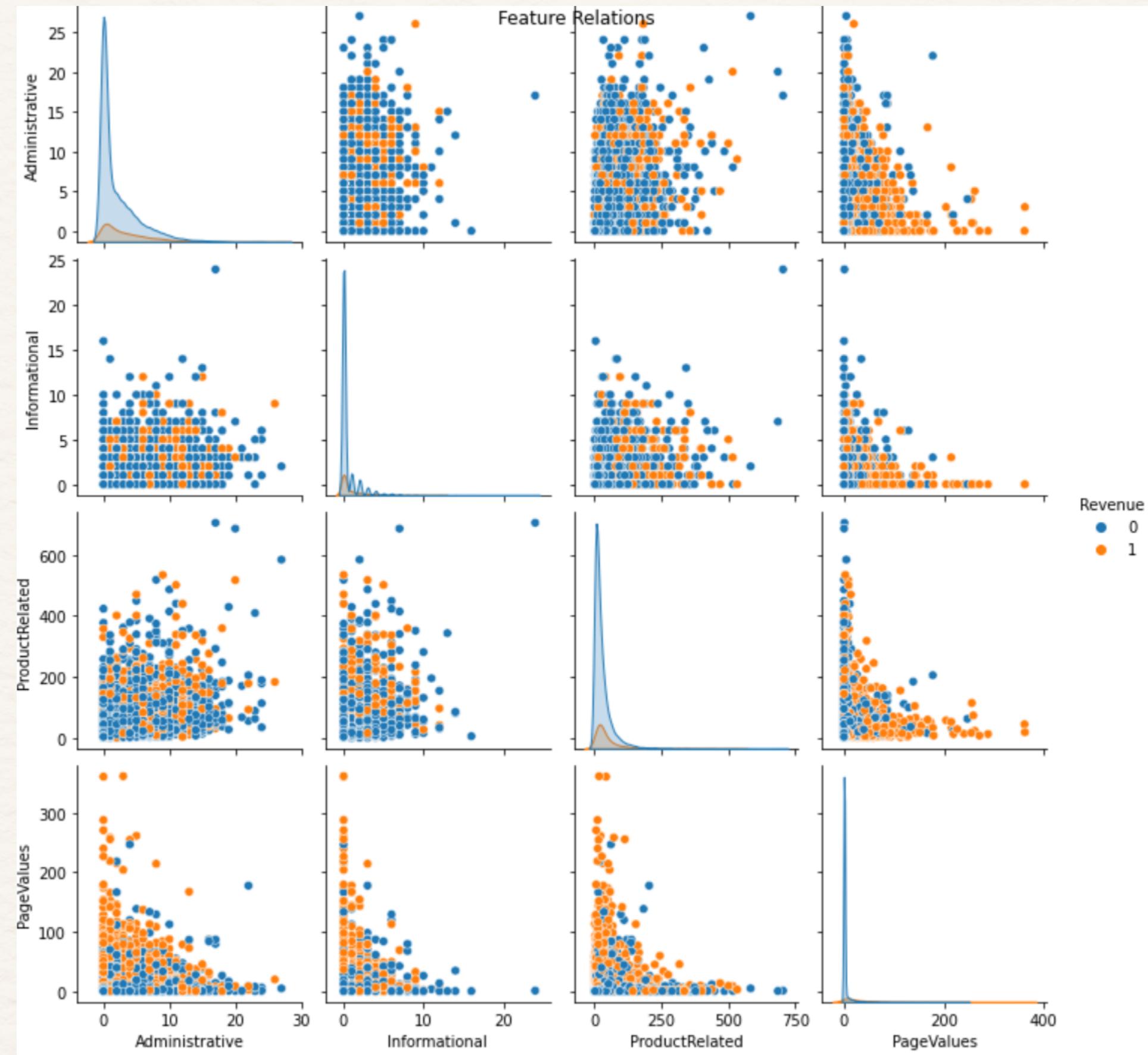
# Heatmap



# Uni-Variant Analysis

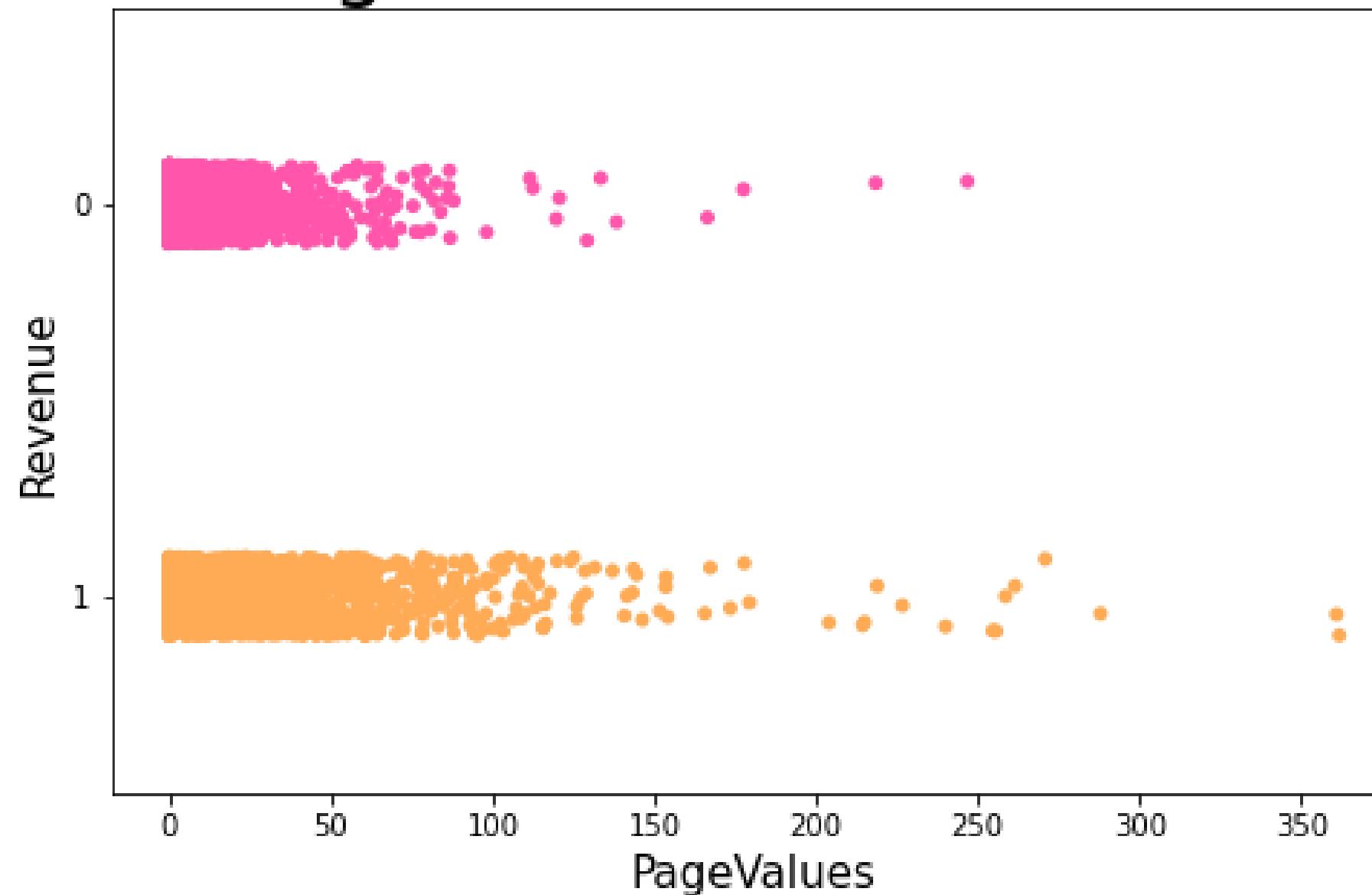


# Graphic Correlation



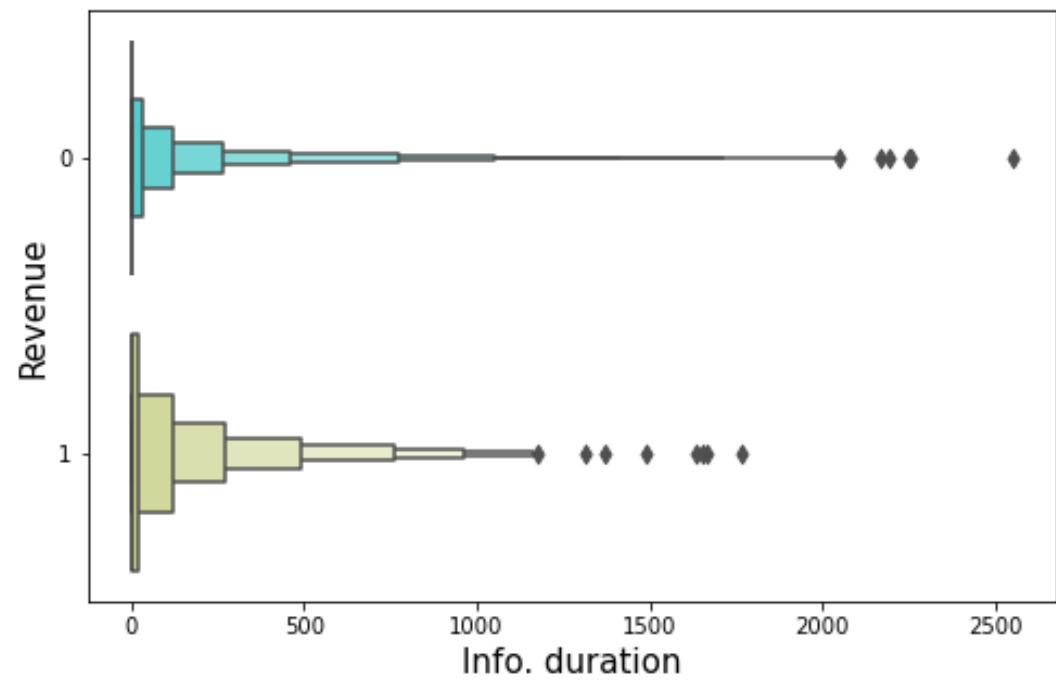
# Analysing Revenue & Page Values

Page Values vs Revenue

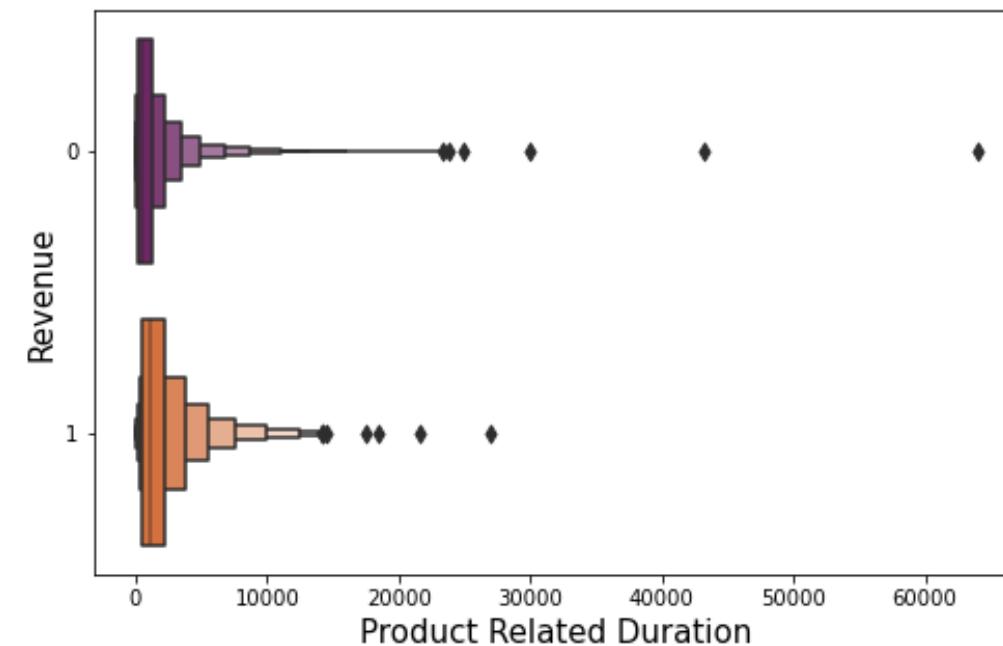


# A Closer Look on Revenue

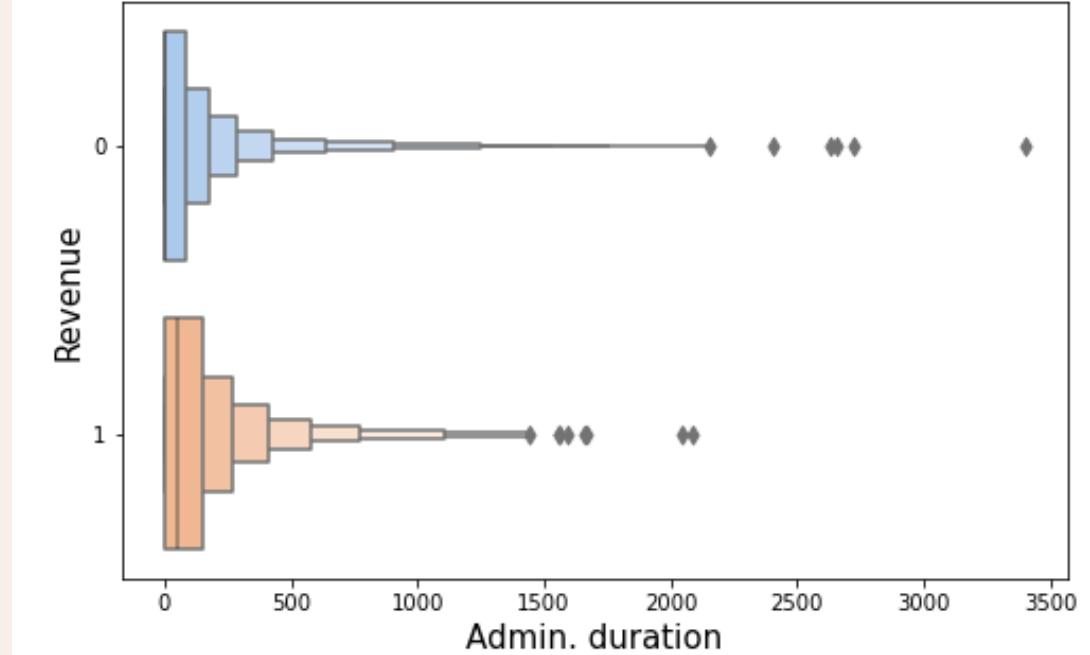
Info. duration vs Revenue



Product Related Duration vs Revenue



Admin. duration vs Revenue



# Super Vector Machine

→ Classifier and Regressor

- divides a set of objects into classes
- widest possible area around the class boundaries remains free of objects

SVM Initial Performance:

-----

Accuracy : 0.8875067604110329  
F1 Score : 0.5856573705179283  
Precision : 0.7577319587628866  
Recall : 0.4772727272727273

Confusion Matrix:

[[1494 47]
[ 161 147]]

# Logistic Regression



Estimator of a Logistic Model

- measures the relationship between categorical dependant variable and independant variables

Logistic Regression initial Performance:

```
-----  
Accuracy      : 0.8788534342888048  
F1 Score       : 0.5313807531380752  
Precision      : 0.7470588235294118  
Recall         : 0.41233766233766234  
Confusion Matrix:  
[[1498  43]  
 [ 181 127]]
```

# Results

- Prediction of Online Shopping Behaviour

Customers are more likely to:

- make purchases on special days
- make purchases in november (close to christmas)
- customer's purchases are highly connected to the webpages visited



## Unexpected Results:

- not more purchases during the weekends
- revenue has an inverse relationship to bounce and exit rates

Merci pour votre  
attention

# Chart Page

- Elaborate on what you want to discuss.

