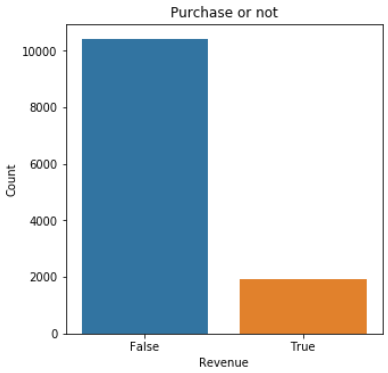
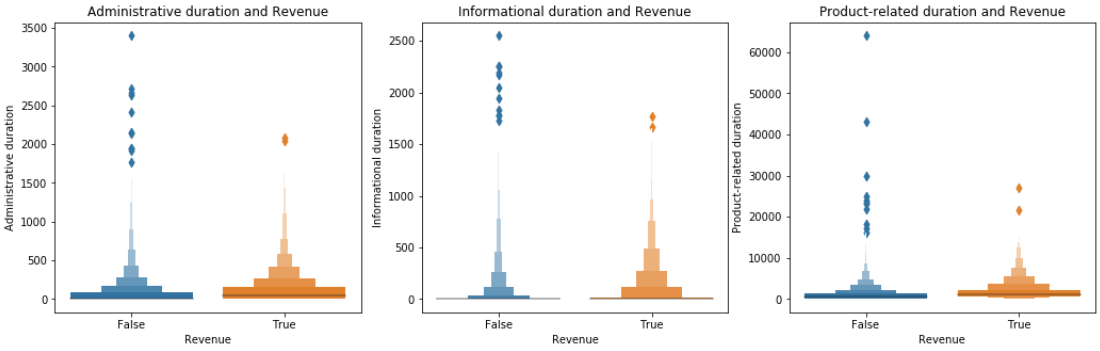
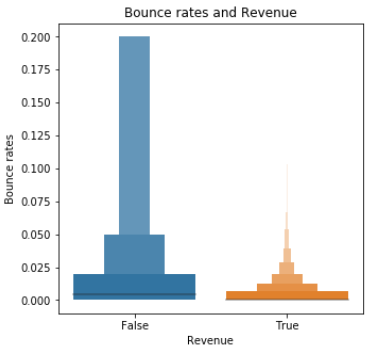
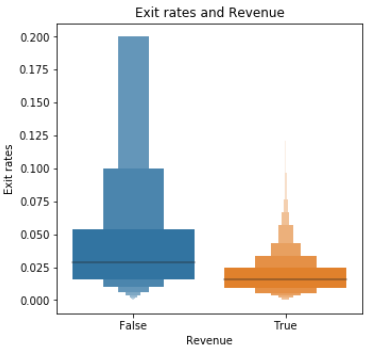
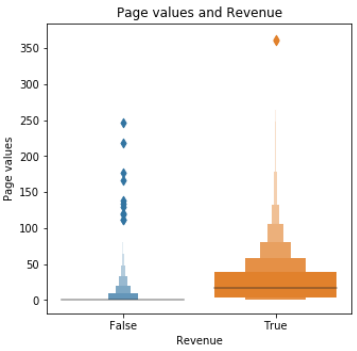
1. **Choose a dataset from next page.**
   Online Shoppers Purchasing Intention Dataset
2. **Analyze the data (statistics, correlation…).**

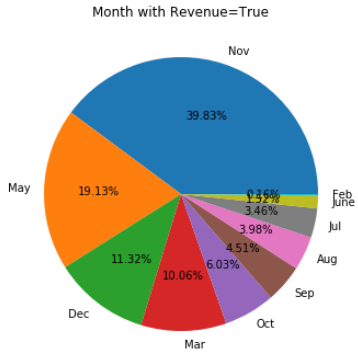


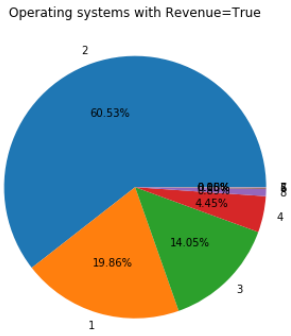


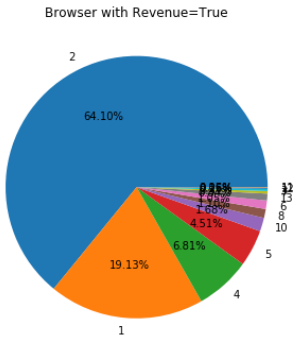


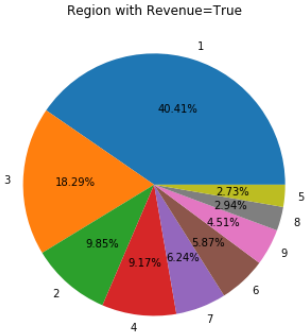


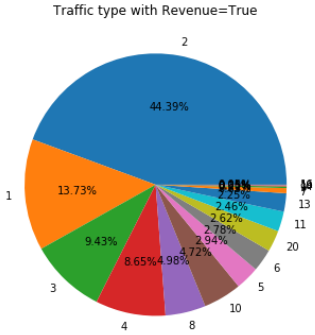


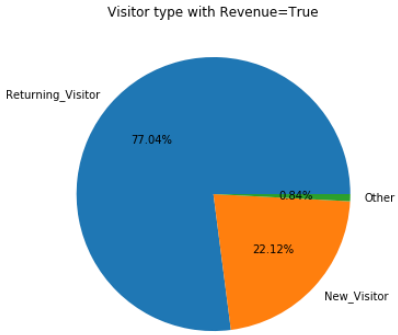


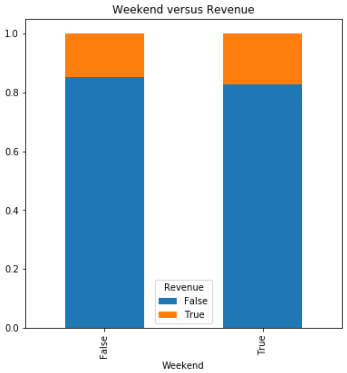

Operating systems with Revenue=True
2
60.53%
19.86%
14.05%
4.45%
3
4
1

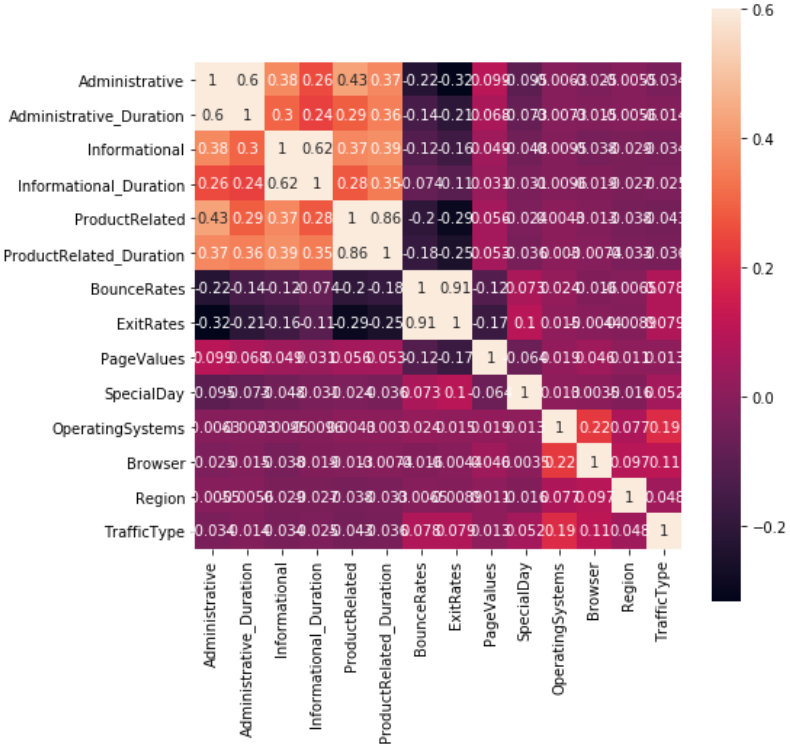

Browser with Revenue=True
2
64.10%
19.13%
6.81%
4.51%
1.68%
5
10
4
1


Region with Revenue=True
1
40.41%
18.29%
9.85%
9.17%
6.24%
5.87%
4.51%
2.94%
2.73%
3
2
4
7
6
9
8
5


Traffic type with Revenue=True
2
44.39%
13.73%
9.43%
8.65%
4.98%
2.94%
2.72%
2.62%
2.46%
1
3
4
8
10
5
6
20
11
13


Visitor type with Revenue=True
Returning_Visitor
77.04%
0.84%
Other
22.12%
New_Visitor


Weekend versus Revenue
1.0
0.8
0.6
0.4
0.2
0.0
Revenue
False
True
False
True
Weekend



3. **Define a reasonable problem (classification, regression, clustering...) and predict the results.**

   Use linear regression to predict whether the online shoppers will purchase the product or not.

4. **Explain how you improved your results step-by-step.**

   (1) First, preprocess the dataset. Manage categorial variables, such as ‘Month’, ‘VisitorType’, ‘Weekend’ and ‘Revenue’.

   (2) Second, divide the dataset into training data and testing data.

   (3) Third, define predictor/independent and target/dependent variables.

   (4) Next, fit the model, predict class and probability.

   (5) Finally, evaluate the accuracy rate.