Online Shoppers Purchasing Intention Dataset

The data is collected from an online shopping platform. Each instance is a session of a user on the platform. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

The dataset has the data of 12,330 sessions and 84.5% were negative class samples that did not end up with a profit (nothing bought).

The dataset consists of 10 numerical and 8 categorical features. The revenue is used as the class label.

Feature label	**Description**
Administrative	Number of the times visited Administrative pages
Administrative Duration	Total time spent on Administrative pages
Informational	Number of the times visited Informational pages
Informational Duration	Total time spent on Informational pages
Product-Related	Number of the times visited Product-Related pages
Product-Related Duration	Total time spent on Product-Related pages
Bounce Rate	Percentage of visitors who enter the site from that page and then leave
Exit Rate	Feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.
Page Value	The average value for a web page that a user visited before completing an e-commerce transaction.
Special Day	The closeness of the site visiting time to a specific special day (e.g. Mother’s Day, Valentine's Day). The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date.

	For example, for Valentine's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.
Month	Self-explanatory
operating system	Self-explanatory
browser	Self-explanatory
region	Self-explanatory
traffic type	Self-explanatory
visitor type	returning or new visitor
weekend	A boolean value indicating whether the date of the visit is weekend

Instructions:

- You should formulate a model to predict the revenue class using the features given.
- You have the freedom to use any method or techniques to analyze the data, train models and evaluate the results. But you should only use standard python libraries + scipy + pandas tools. No third-party libraries are allowed.
- You must deliver your project in the form of a jupyter notebook
- Jupyter notebooks are all about telling a story using the data. So make your notebook that way, present everything nicely and have a good flow.
- We will not mark you on the final accuracy you get. We will mark you on how well you explain the decisions you have made in every stage of your project, reasoning behind your decisions and proper data representation, evaluation and analyzing methods. All of these must be shown within the notebook itself.
- Make sure you attend labs and complete assignments, that will help you a lot.