Data Visualisation

Correlation matrix

The correlation matrix can show if there is any correlation between the features all the features and the target. As we can see in our correlation matrix our features are not really correlated to our target. The four features the most correlated to our target are PagesValues (0.5), ProductRelated (0.2), BounceRates (-0.2) and ExitRates (-0.2)

PCA

The PCA shows us the ability to reduce the dimensionality of our dataset.   
As we can see we need to keep at least 10 axes to well represent our dataset.

Univariate Analysis

I decided to take a look at our dataset balance by looking at our taget (Purchased) and VisitorTypes.  
As we can see, the dataset is quite unbalanced. This means that we will need to downsample the data that leads to no purchase.

I tried to look at the purchase repartition across all month to see if there is any difference between ‘Purchase’ and ‘No Purchase’

Bivaiate analysis

I decided to compare the ExitRates for ‘purchase’ and ‘no purchase’. As we can see there is a difference between the two, this may explain the correlation between Revenue (Purchas/NoPurchase) and the ExitRates.

As we can see here the VisitorType is almost the same across all the different type of visitors.

Multivariate analysis

Modélisation

Processing

API Flask