ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

Intention of Online Shoppers

Friday, 15th of December 2023

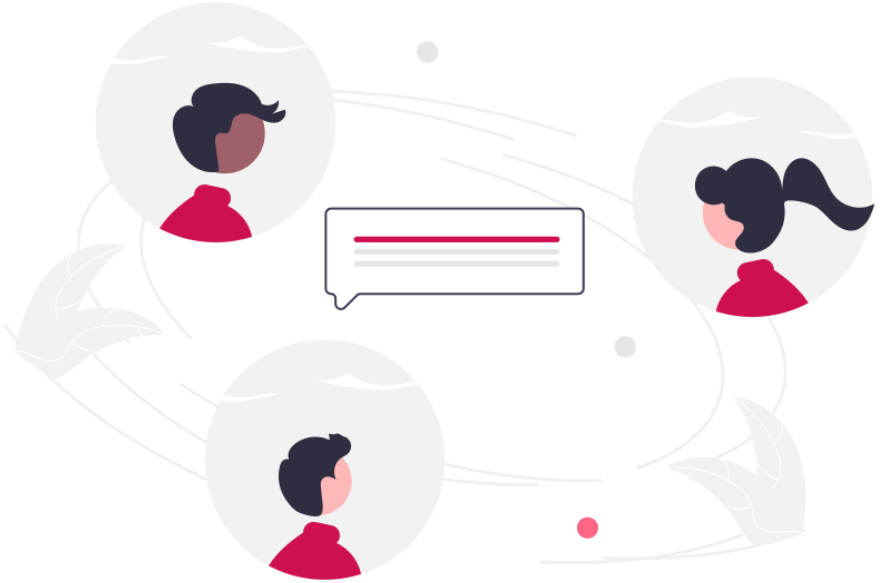
ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

OUR TEAM MEMBERS

Ahmed MAALOUL

Martin PUJOL

Sarujan DENSON



ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

The ins and the outs of the Project

Page Categories :

- Administrative, Informational, Product Related: These represent the types of pages a visitor views on an e-commerce site
- Aministrative Duration, Informational Duration, Product Related Duration: These show the time spent on each tyoe of page

Google Analytics Metrics :

- Bounce Rate: Percentage of visitors who enter a page and leave without doing anything else
- Exit Rate: Percentage of visitors for whom a page was the last in their session
- Page Value: Average value of a page visited before a transaction is completed

Special Day :

- Indicates how close a site visit is to a special day
- Values vary around specific dates (e.g., Valentine's Day), with a maximum value on the actual day

Other Features :

- Operating System, Browser, Region, Traffic Type: Information about the user's environnement
- Visitor Type: Whether the visitor is returning or new
- Weekend: Boolean indicating if the visit date in on a weekend
- Month of the Year: Indicates the month of the visit

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

DATA EXPLORATION (1)

- We have 12330 rows and 18 columns for the dataset.
- It contains 10 numerical and 8 categorical attributes.
- The 'Revenue' attribute can be used as the class label.

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

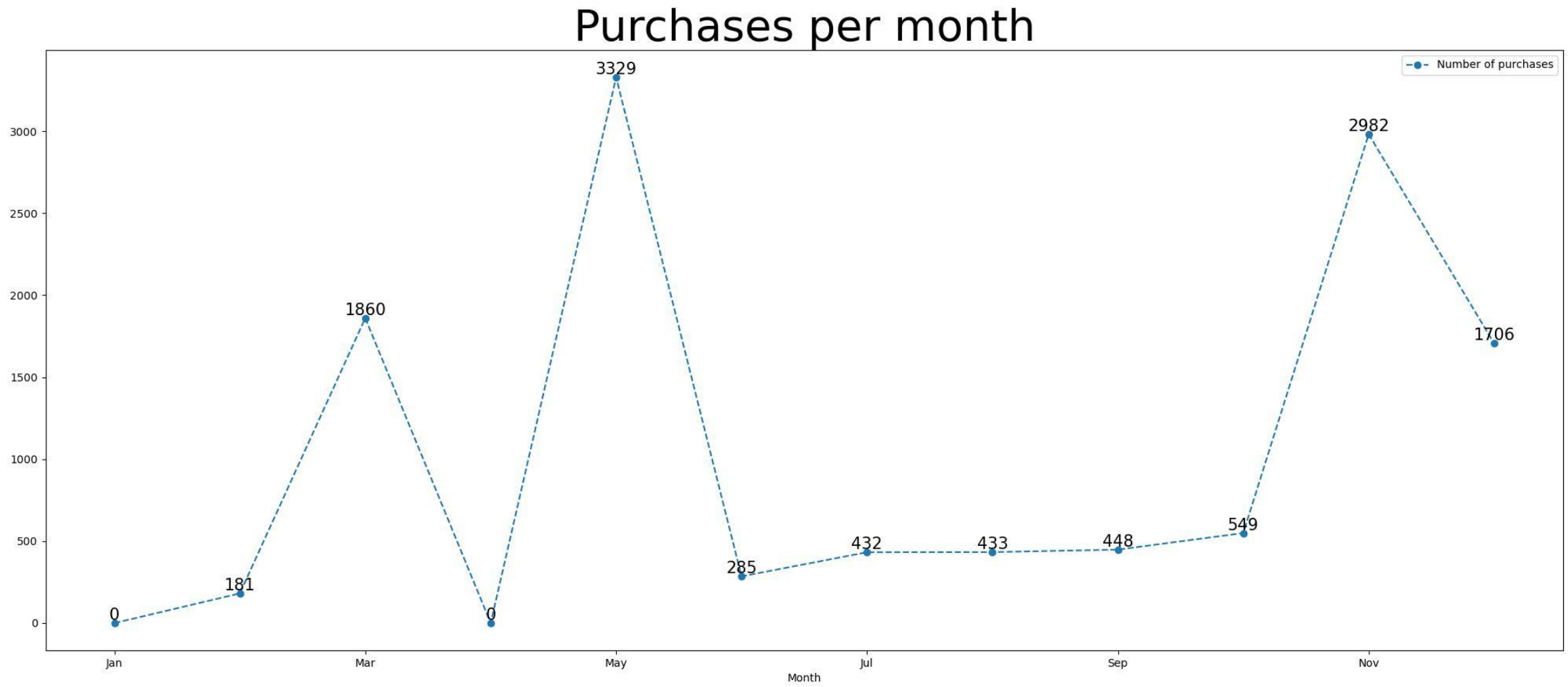
DATA EXPLORATION (2)

Is the Data Cleaned ?

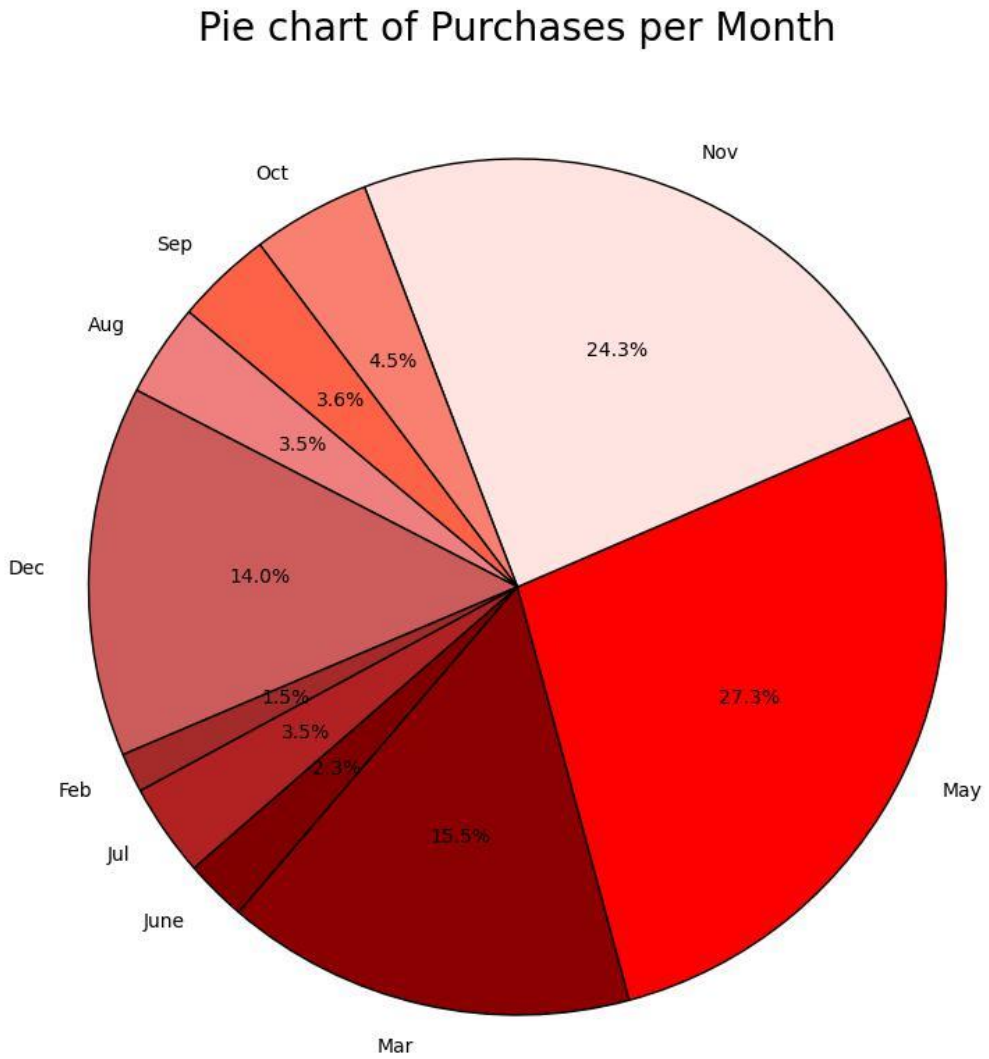
- The dataset has no missing values, all are clean.
- There are still 132 duplicated rows.
- The dataset is unbalancedand not normalized.

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

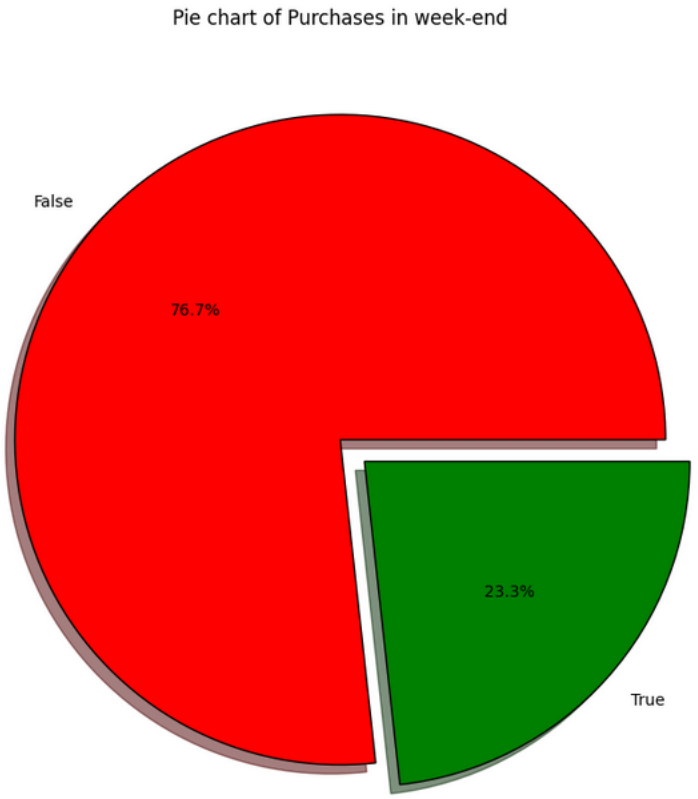
DATA VISUALIZATION (1)

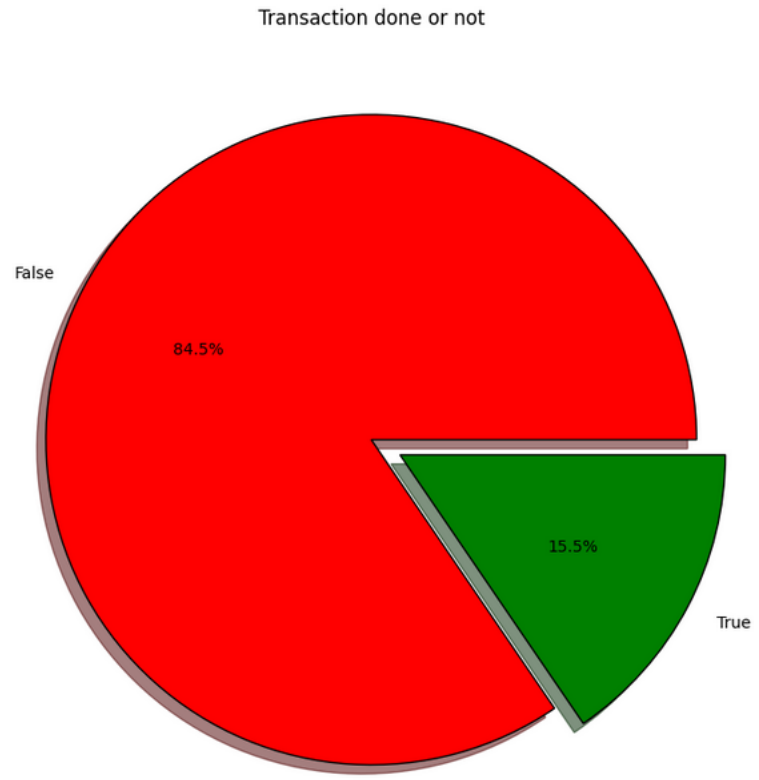


DATA VISUALIZATION (2)



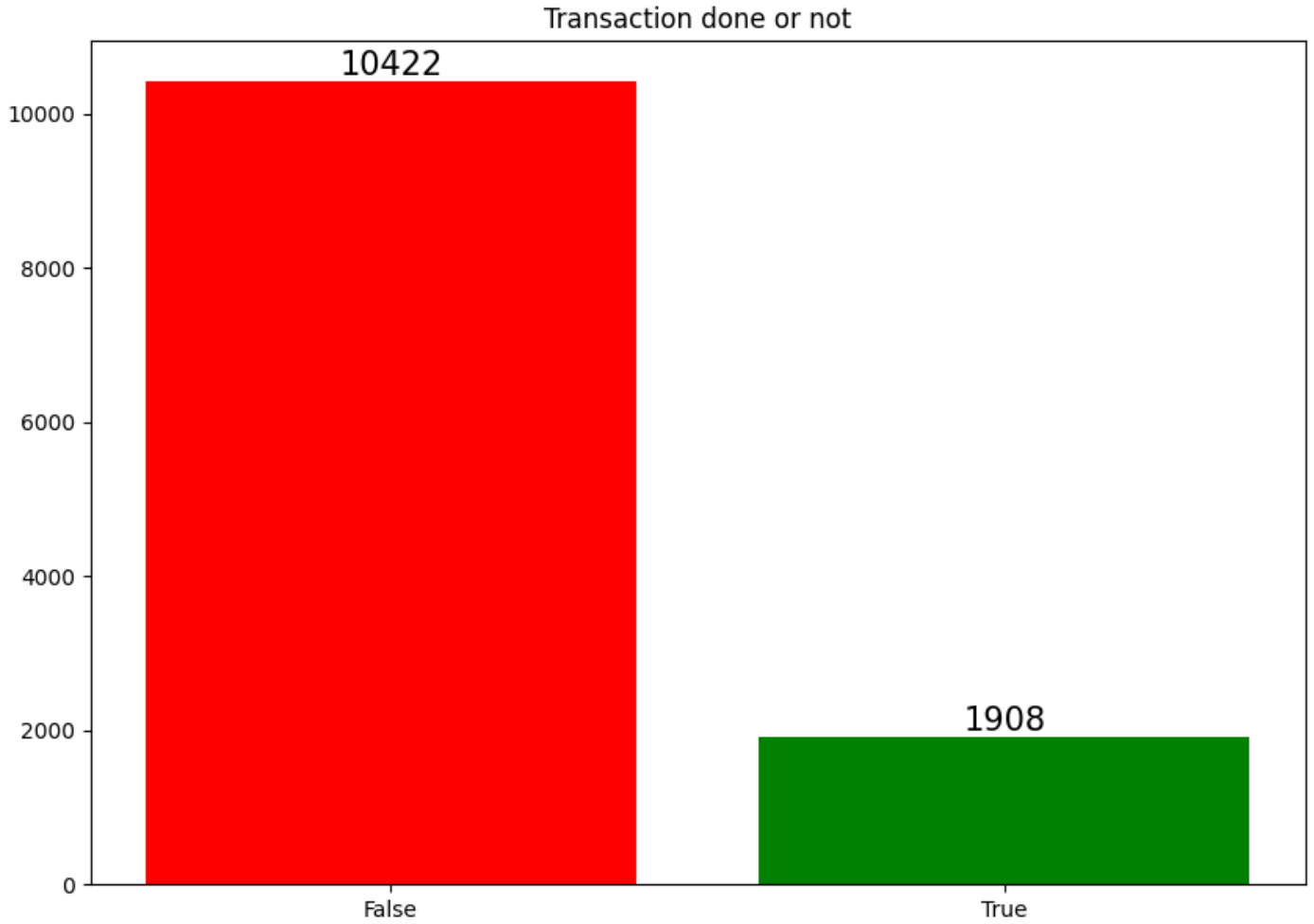
DATA VISUALIZATION (3)





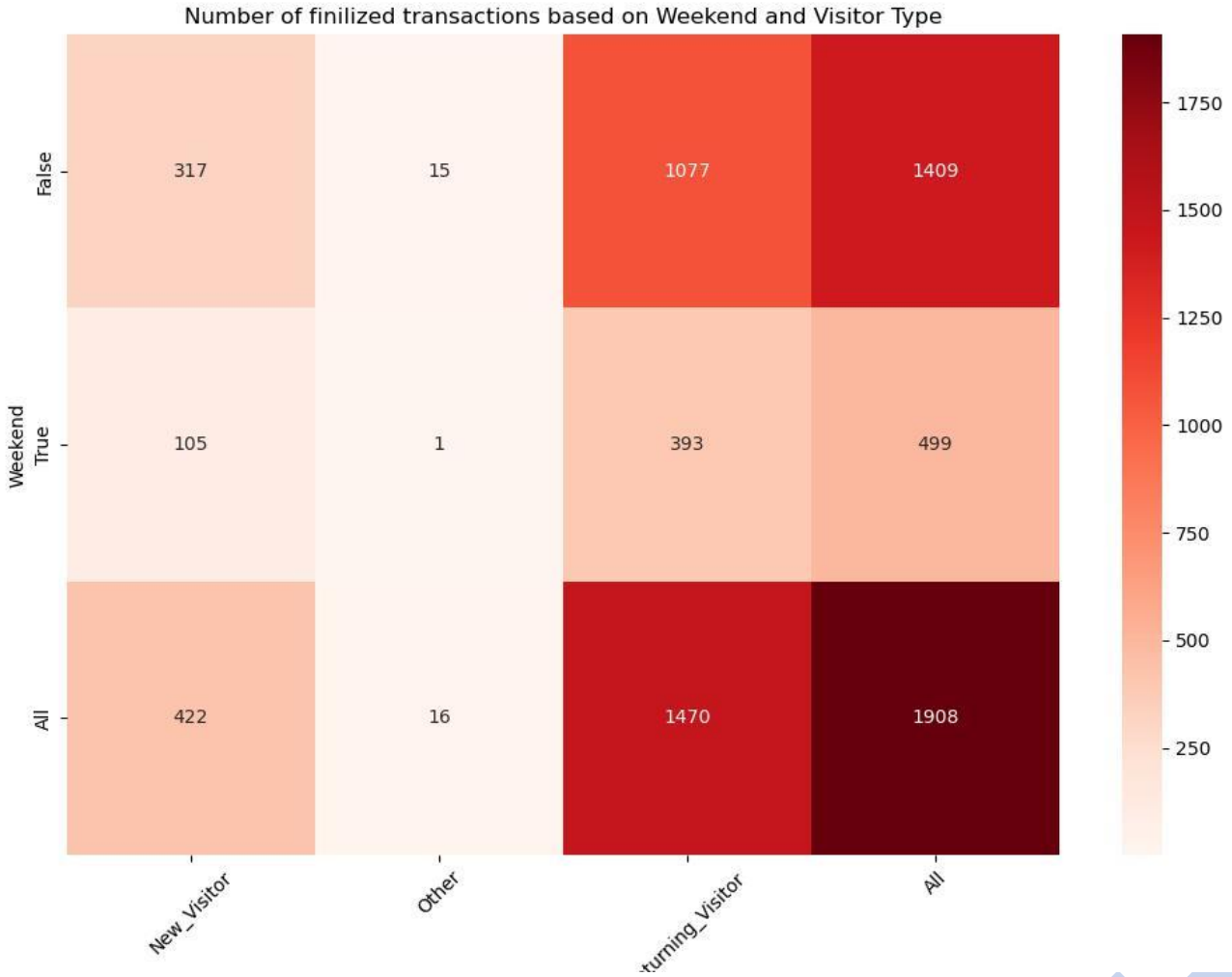
Low influence of the weekend

DATA VISUALIZATION (4)

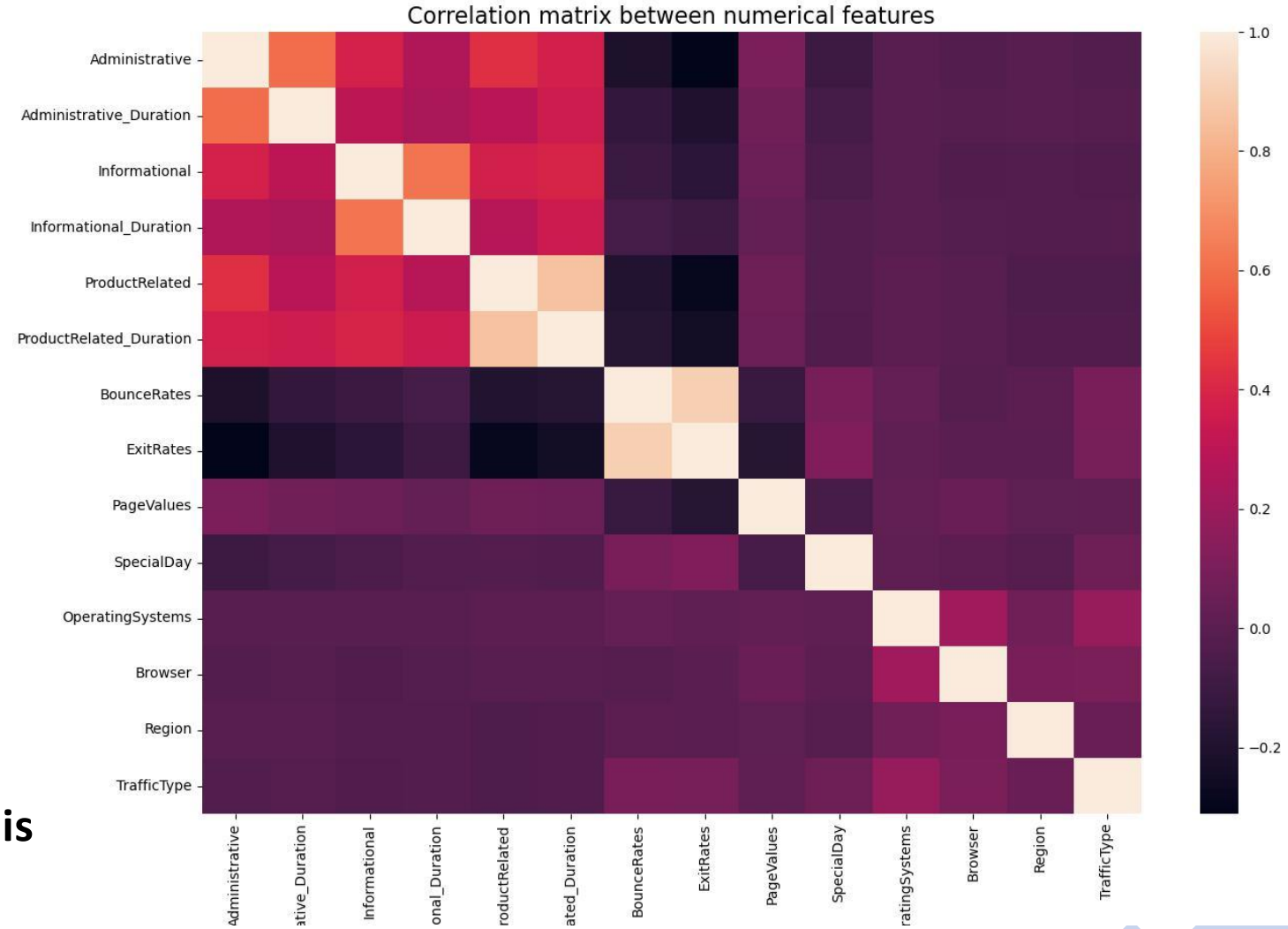


The dataset is unbalanced !

DATA VISUALIZATION (5)



DATA VISUALIZATION (6)



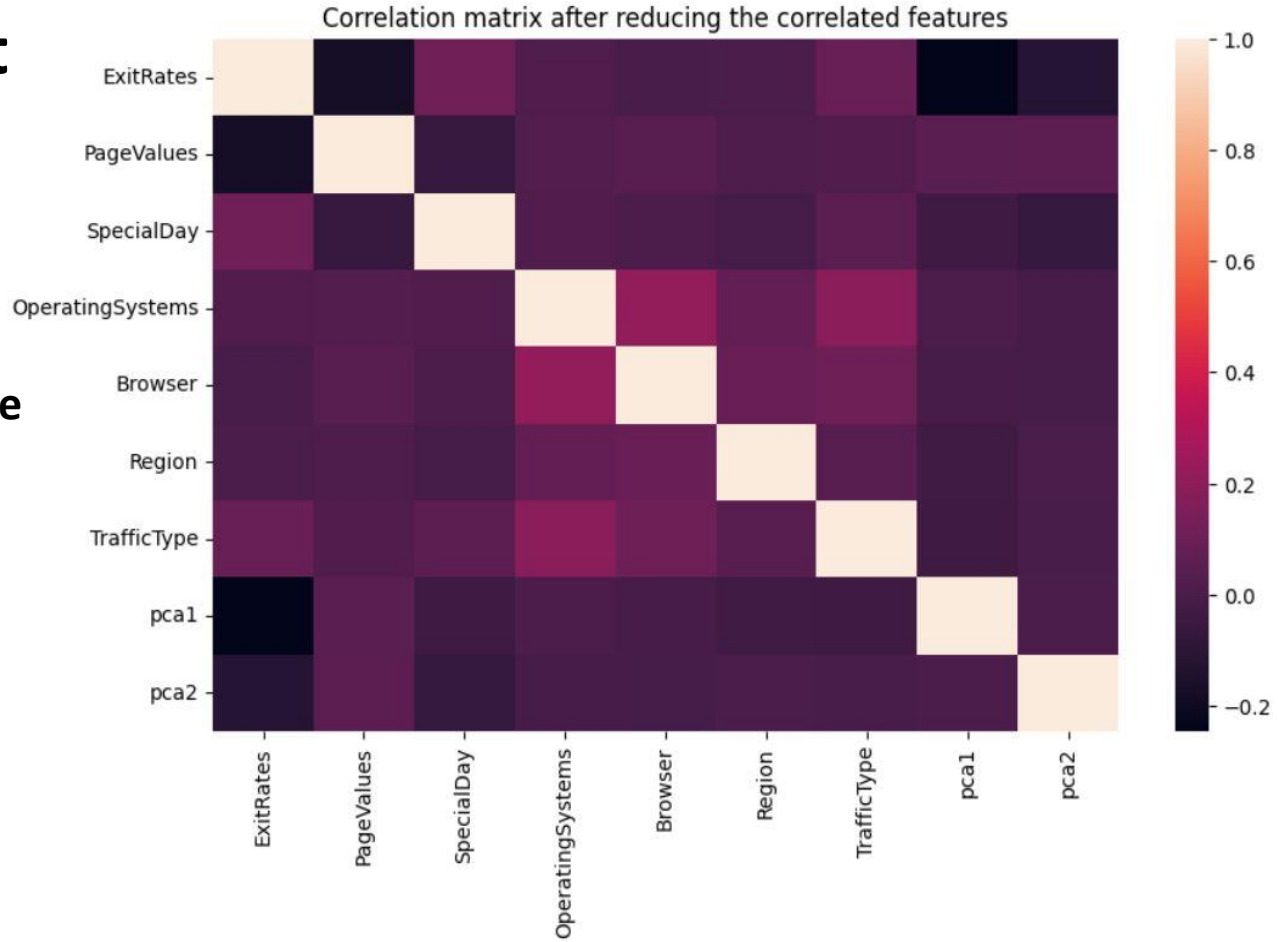
The dataset is correlated !

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

DATA PRE-PROCESSING (1)

How to handle correlated / Redundant variables ?

- Reduce the input space dimension by using PCA ! Reduce 6 variables to 2 !
- We observe that we have 125 duplicates rows in the dataset, we deleted them .



ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

DATA PRE-PROCESSING (1)

How to Balance the data set ?

We don’t want our model to just return false to have 90 % of accuracy !

Downsampling

Less data to train on

```
Revenue
0    1908
1    1908
Name: count, dtype: int64
```

Upsampling

Might lead to overfitting due to redundancy on reduced dataset

```
Revenue
0    10297
1    10297
Name: count, dtype: int64
```

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (1)

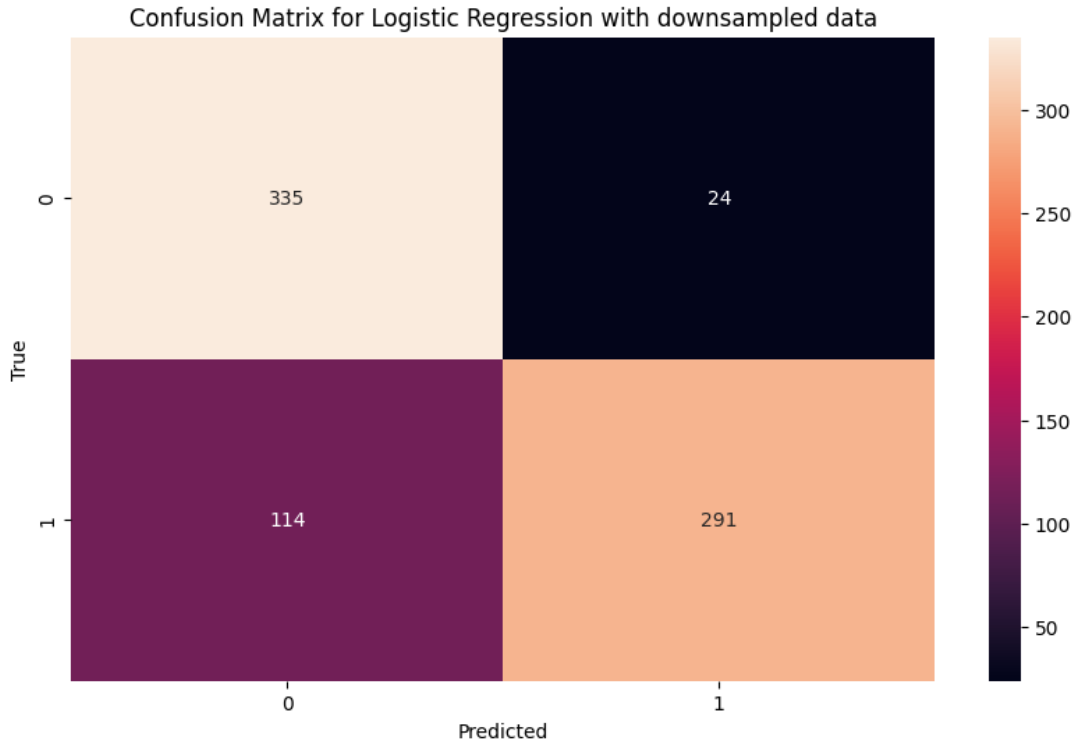
For understanding the performance of our classification models (Logistic regression, Neural Network, Random Forest), we can use confusion matrix since it shows the number of correct and incorrect predictions for each class

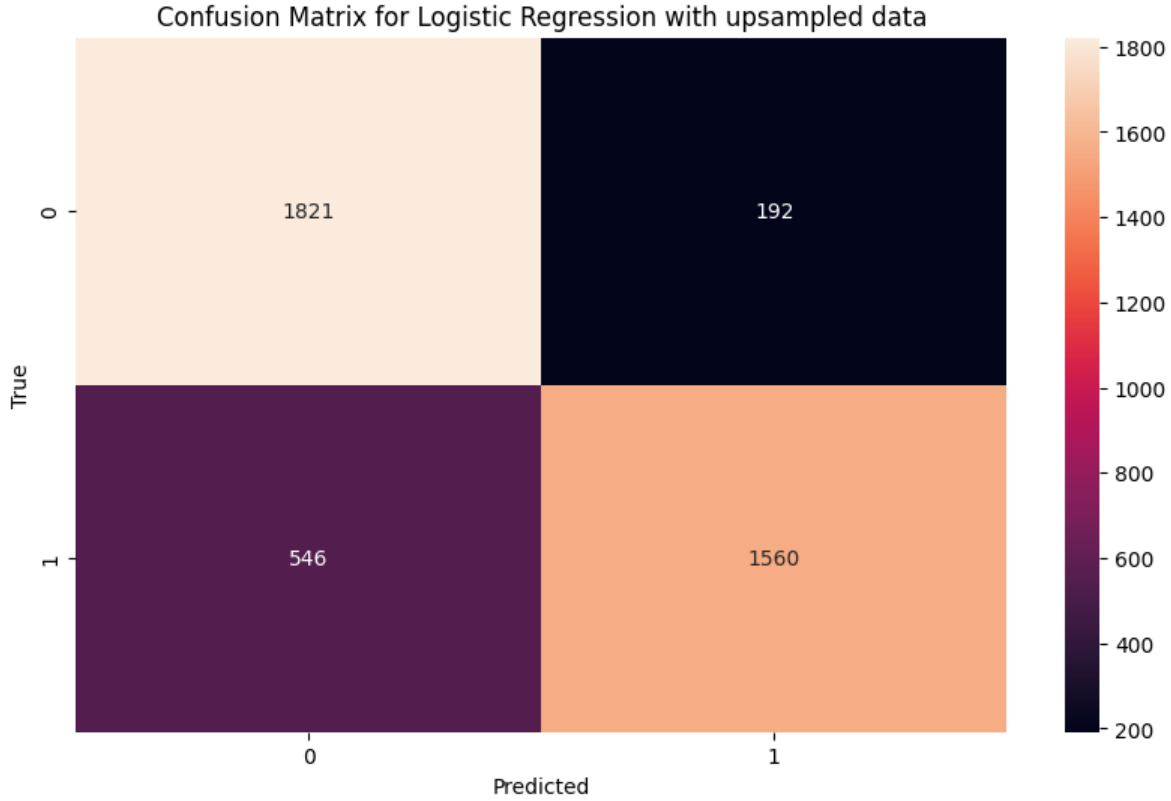
ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (2)

Logistic Regression

```
param_grid = {
    'C': [0.001, 0.01, 0.1, 1, 10, 100],
}
```



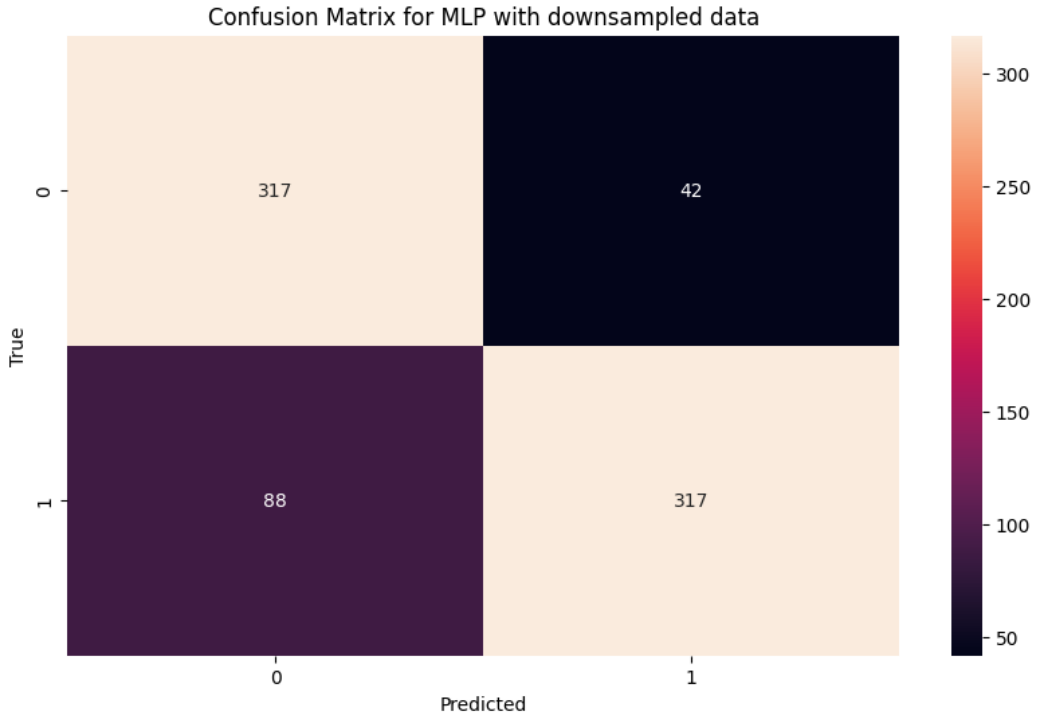


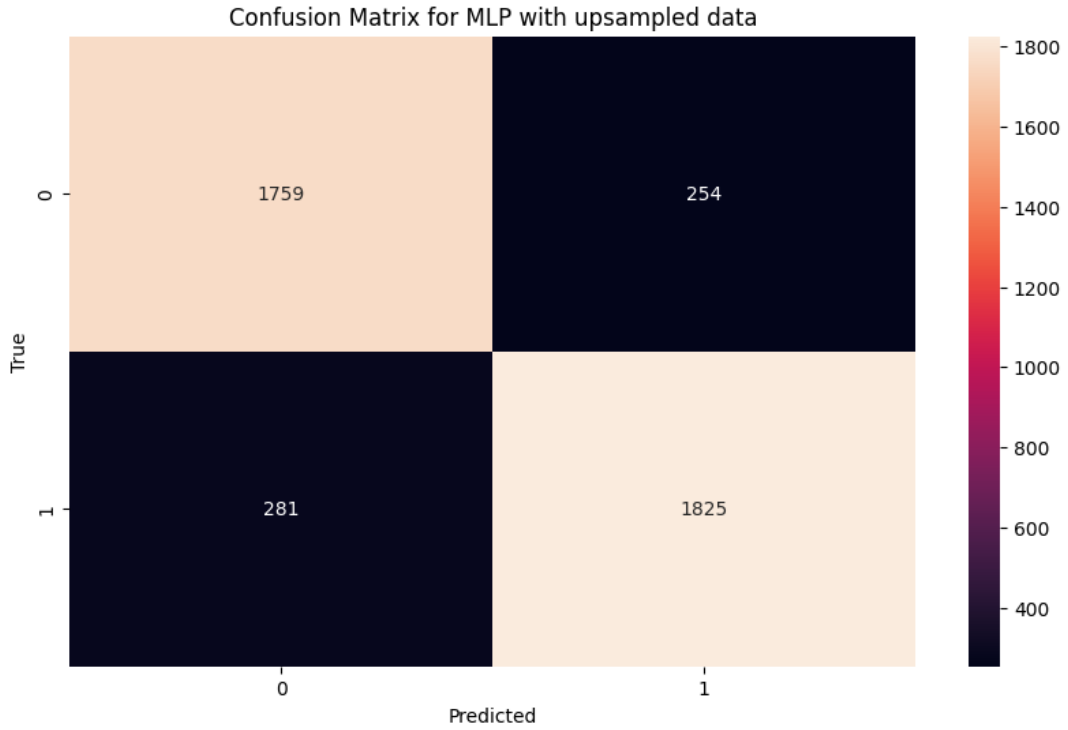
ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (3)

Multi-Layer Perceptron

```
param_grid_mlp = {
    'hidden_layer_sizes': [(20,20), (10,10), (20,), (10,)],
    'activation': ['tanh', 'relu'],
}
```

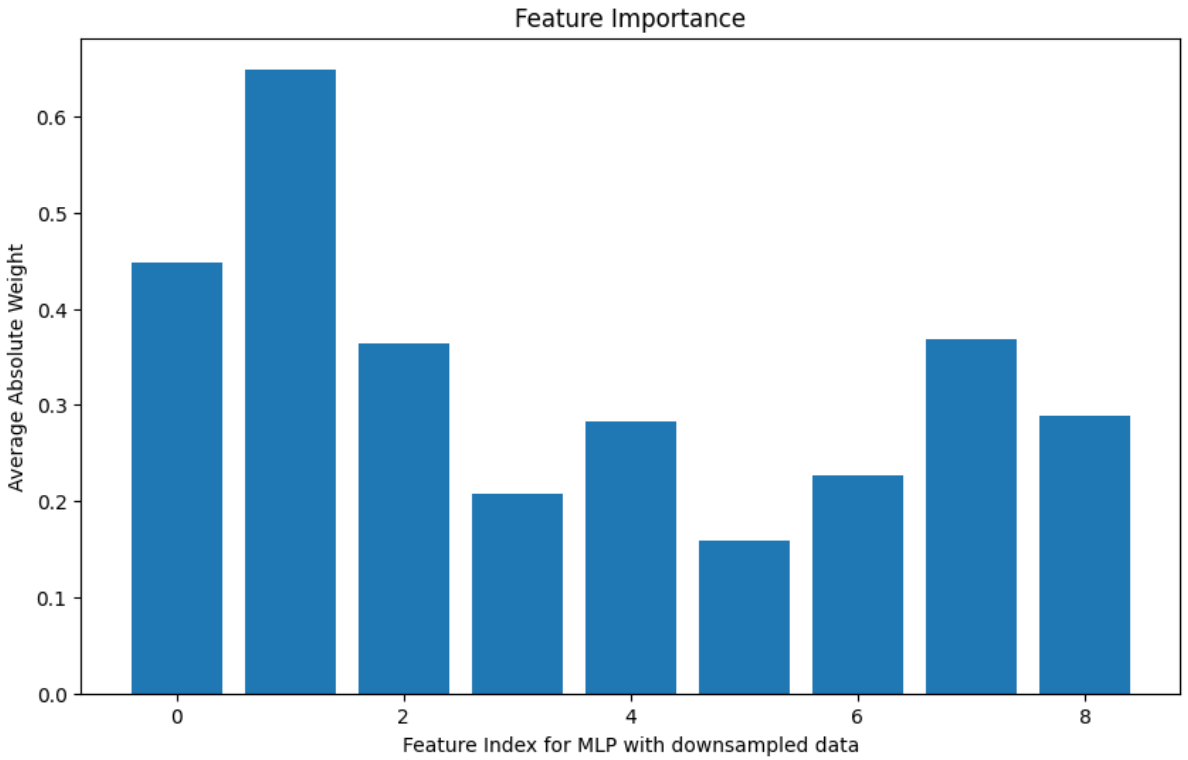


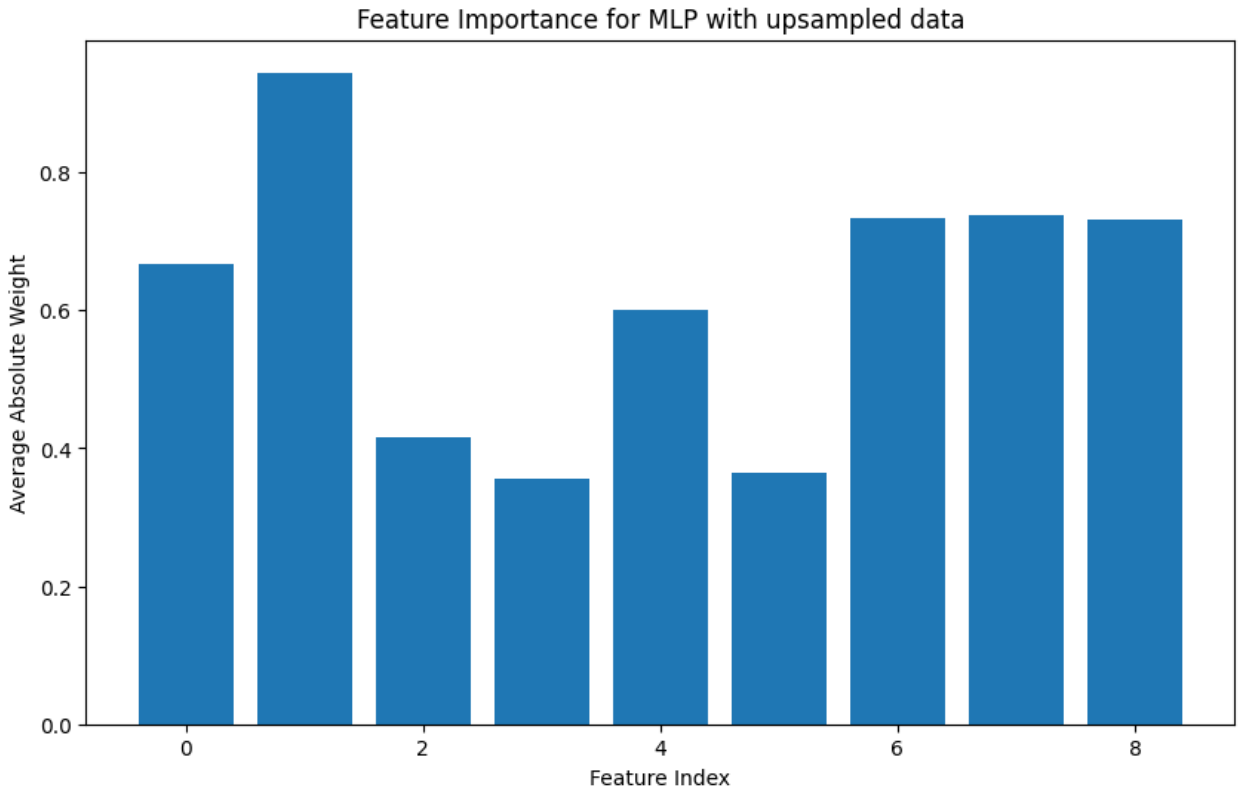


ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (4)

Logistic Regression



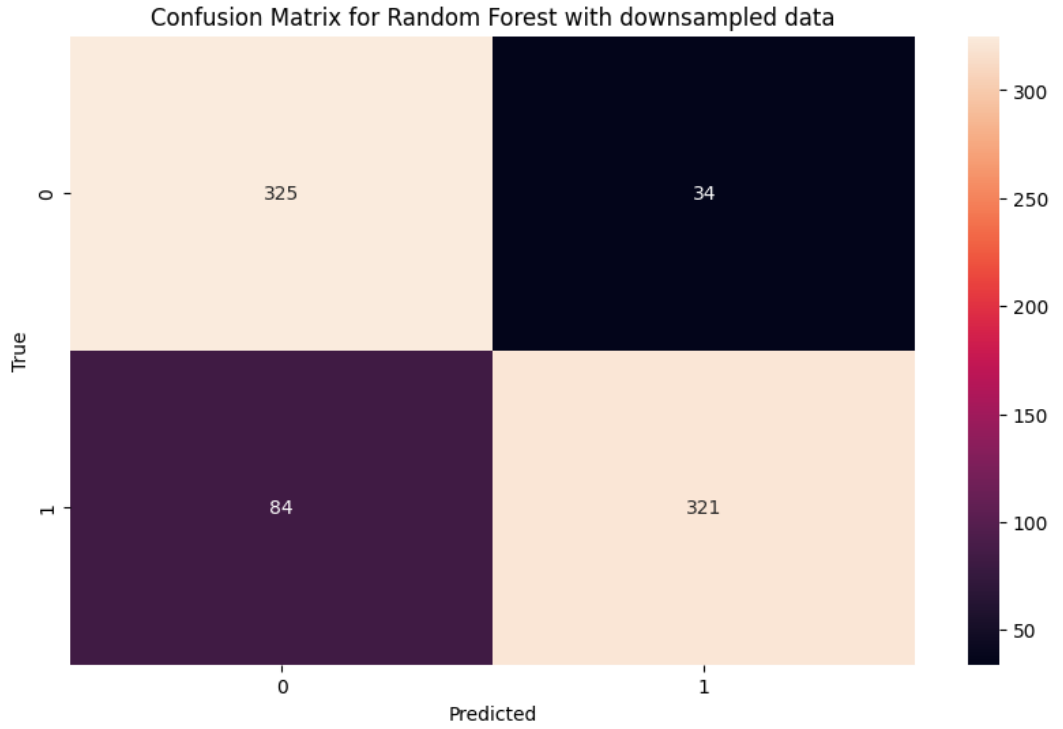


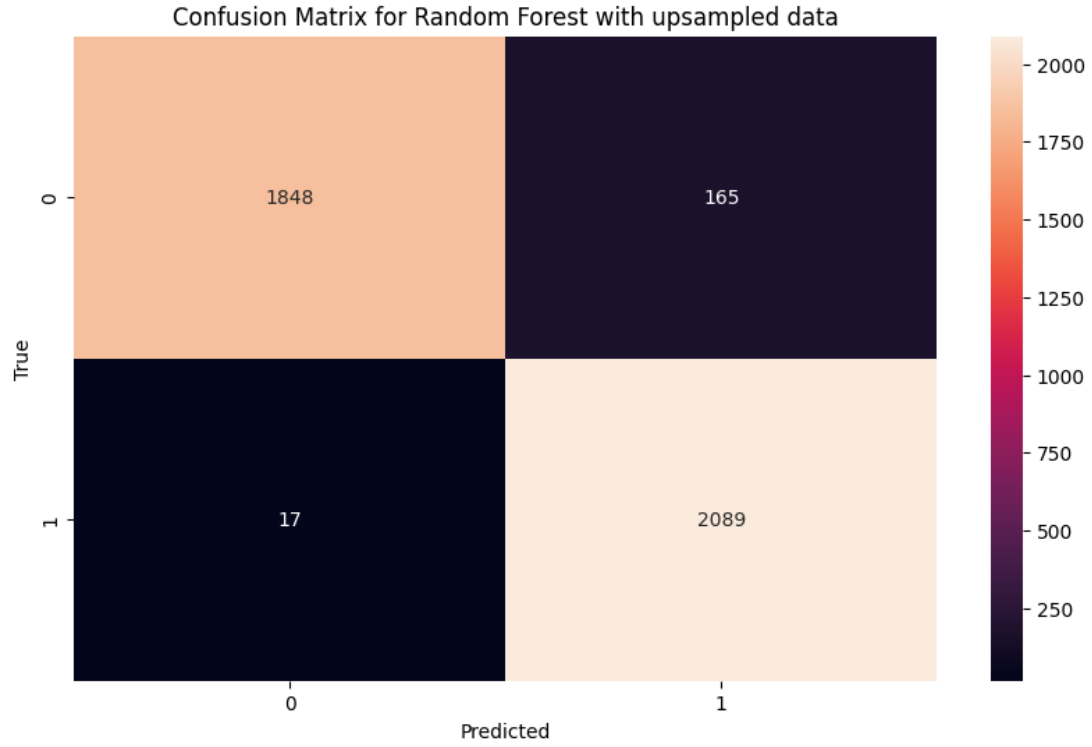
ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (5)

Random Forest

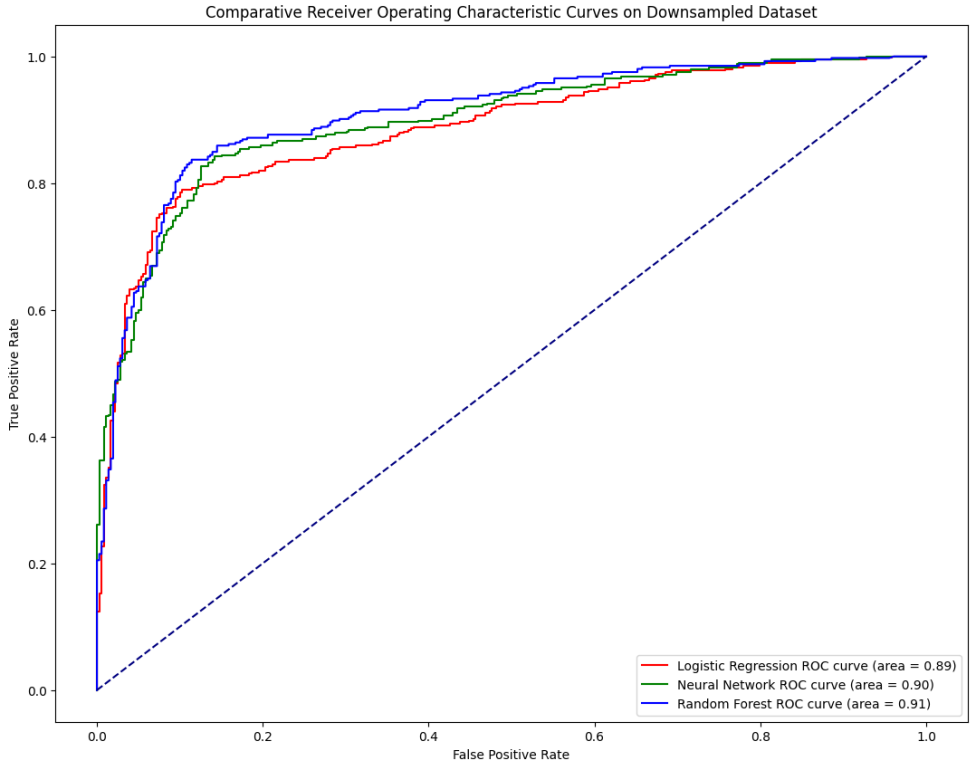
```
param_grid_rfc = {
    'n_estimators': [30, 50, 100, 200],
    'max_depth': [2, 5, 10, 20],
    'min_samples_split': [2, 10],
}
```

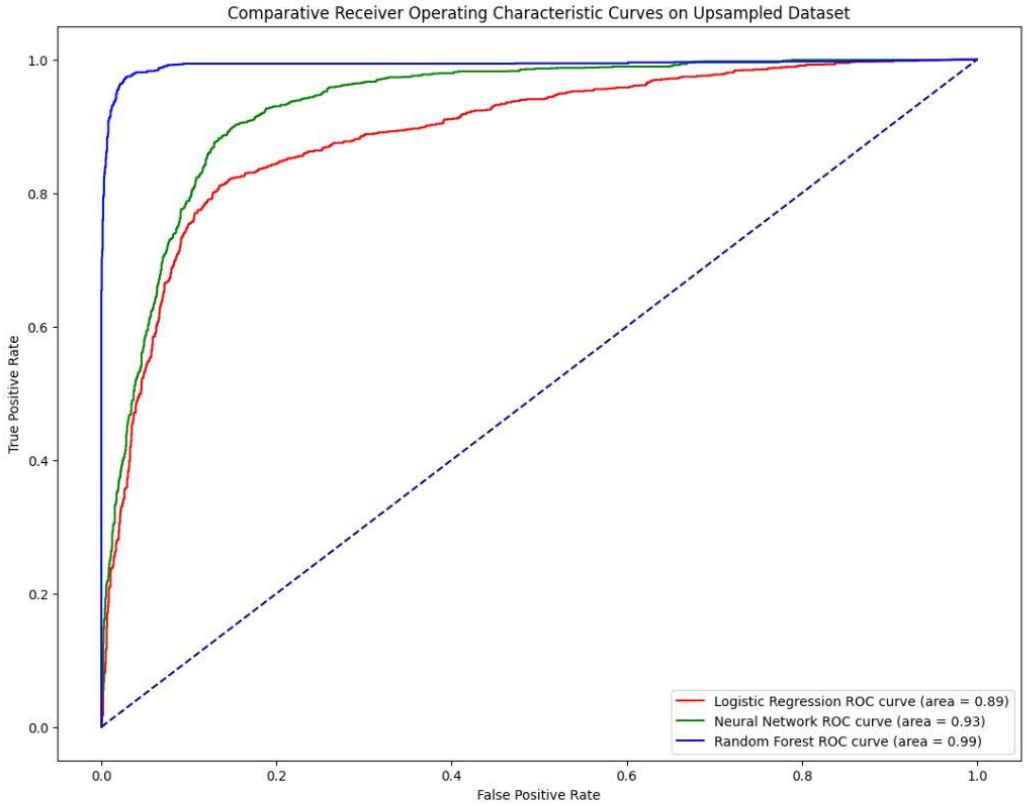




ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (6)





ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

MODELISATION (7)

Selected MODEL

```
Meilleurs hyperparamètres: {'max_depth': 20, 'min_samples_split': 2, 'n_estimators': 100}
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.92      0.95      2013
           1       0.93      0.99      0.96      2106

    accuracy                           0.96      4119
   macro avg       0.96      0.95      0.96      4119
weighted avg       0.96      0.96      0.96      4119
```

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

API (1)

Enjoy predicting the shopper intenti

Administrative: | Administrative Duration:

Informational: | Product Related:

Informational duration:

Product Related Duration: | Bounce Rates:

Exit Rates: | Page Values:

Special Day: | Operating Systems:

Browser: | Region:

Traffic Type: | Month: January

Vistor Type: Returning Visitor

Weekend: ☐

Some magic

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

API (2)

Result when the client doesn’t buy

Ta

Try the magic again

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

API (3)

Result when the client buys