Importation des bibliothéques

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import r2_score
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
```

Importation de notre database

```
db = pd.read_csv("db-scoring-2.csv")

##Visualiser les 5 premieres lignes
db.head()

   Customer_ID  Status_Checking_Acc  Duration_in_Months  \
0       100001             … < 0 USD                   6
1       100002        0 <= … < 10000                  48
2       100003  no checking account                   12
3       100004             … < 0 USD                  42
4       100005             … < 0 USD                  24

                                      Credit_History
Purposre_Credit_Taken  \
0  critical account/other credits existing(not at...
radio/television
1          existing credits paid back duly till now
radio/television
2  critical account/other credits existing(not at...
education
3          existing credits paid back duly till now
furniture/equipment
4                    delay in paying off in the past             car
(new)

   Credit_Amount                    Savings_Acc
Years_At_Present_Employment  \
0          1169  unknown/ no savings account               .. >= 7
years
1          5951                 … < 1000 USD           1 <= … < 4
years
```

```
2          2096                 … < 1000 USD            4 <= … < 7
years
3          7882                 … < 1000 USD            4 <= … < 7
years
4          4870                 … < 1000 USD            1 <= … < 4
years

   Inst_Rt_Income              Marital_Status_Gender  ...  \
0               4                        male  single  ...
1               2  female divorced/separated/married  ...
2               2                        male  single  ...
3               2                        male  single  ...
4               3                        male  single  ...

  credit history score  Credit_Amount score savings score  \
0                    0                    2             0
1                    1                    1             0
2                    0                    2             0
3                    1                    1             0
4                    0                    1             0

   Years_At_Present_Employment score Other_Debtors_Guarantors score  \
0                                  3                              0
1                                  1                              0
2                                  2                              0
3                                  2                              2
4                                  1                              0

  Current_Address_Yrs score  Job score Property score  Score
pondéré  \
0                          1          1              3            1.06

1                          0          1              3            1.03

2                          1          0              3            0.71

3                          1          1              2            1.12

4                          1          1              0            0.58


  Score pondéré binaire
0                     1
1                     1
2                     0
3                     1
4                     0

[5 rows x 34 columns]
```

```
##Visualiser les 5 dérnieres lignes
db.tail()

      Customer_ID  Status_Checking_Acc  Duration_in_Months  \
4995       104996  no checking account                  12
4996       104997             … < 0 USD                  30
4997       104998  no checking account                  12
4998       104999             … < 0 USD                  45
4999       105000       0 <= … < 10000                  45

                                         Credit_History
Purposre_Credit_Taken  \
4995          existing credits paid back duly till now
furniture/equipment
4996          existing credits paid back duly till now           car
(used)
4997          existing credits paid back duly till now
radio/television
4998          existing credits paid back duly till now
radio/television
4999  critical account/other credits existing(not at...           car
(used)

      Credit_Amount           Savings_Acc Years_At_Present_Employment
\
4995           1736          … < 1000 USD            4 <= … < 7 years

4996           3857          … < 1000 USD            1 <= … < 4 years

4997            804          … < 1000 USD                 .. >= 7 years

4998           1845          … < 1000 USD            1 <= … < 4 years

4999           4576  1000 <= … < 5000 USD                  unemployed


      Inst_Rt_Income              Marital_Status_Gender  ...  \
4995               3  female divorced/separated/married  ...
4996               4          male  divorced/separated  ...
4997               4                        male  single  ...
4998               4                        male  single  ...
4999               3                        male  single  ...

     credit history score  Credit_Amount score savings score  \
4995                    1                    2             0
4996                    1                    1             0
4997                    1                    2             0
4998                    1                    2             0
4999                    0                    1             1
```

```
      Years_At_Present_Employment score Other_Debtors_Guarantors score
\
4995                                  2                              0

4996                                  1                              0

4997                                  3                              0

4998                                  1                              0

4999                                  0                              0


     Current_Address_Yrs score  Job score Property score  Score
pondéré  \
4995                         1          0              3
0,91
4996                         1          2              2
0,93
4997                         1          1              1
0,96
4998                         1          1              0
0,91
4999                         1          1              1
0,83

     Score pondéré binaire
4995                     1
4996                     1
4997                     1
4998                     1
4999                     0

[5 rows x 34 columns]
```

```
##La somme de variables null
db.isnull().sum()
```

```
Customer_ID                           0
Status_Checking_Acc                   0
Duration_in_Months                    0
Credit_History                        0
Purposre_Credit_Taken                 0
Credit_Amount                         0
Savings_Acc                           0
Years_At_Present_Employment           0
Inst_Rt_Income                        0
Marital_Status_Gender                 0
Other_Debtors_Guarantors              0
Current_Address_Yrs                   0
```

```
Property                               0
Age                                    0
Other_Inst_Plans                       0
Housing                                0
Num_CC                                 0
Job                                    0
Dependents                             0
Telephone                              0
Foreign_Worker                         0
Default_On_Payment                     0
Customer_ID.1                          0
Status score                           0
credit history score                   0
Credit_Amount score                    0
savings score                          0
Years_At_Present_Employment score      0
Other_Debtors_Guarantors score         0
Current_Address_Yrs score              0
Job score                              0
Property score                         0
Score pondéré                          0
Score pondéré binaire                  0
dtype: int64

##Quelques informations statistiques sur notre dataset
db.describe()

         Customer_ID  Duration_in_Months  Credit_Amount
Inst_Rt_Income  \
count    5000.000000         5000.000000    5000.000000
5000.000000
mean   102500.500000           20.903000    3271.258000
2.973000
std      1443.520003           12.053989    2821.607329
1.118267
min    100001.000000            4.000000     250.000000
1.000000
25%    101250.750000           12.000000    1365.500000
2.000000
50%    102500.500000           18.000000    2319.500000
3.000000
75%    103750.250000           24.000000    3972.250000
4.000000
max    105000.000000           72.000000   18424.000000
4.000000

       Current_Address_Yrs          Age       Num_CC   Dependents  \
count          5000.000000  5000.000000  5000.000000  5000.000000
mean              2.845000    35.546000     1.407000     1.155000
std               1.103276    11.370917     0.577423     0.361941
```

```
min               1.000000    19.000000     1.000000     1.000000
25%               2.000000    27.000000     1.000000     1.000000
50%               3.000000    33.000000     1.000000     1.000000
75%               4.000000    42.000000     2.000000     1.000000
max               4.000000    75.000000     4.000000     2.000000

       Customer_ID.1  Status score  credit history score  
Credit_Amount score  \
count    5000.000000   5000.000000           5000.000000   
5000.000000   
mean   102500.500000      1.001000              0.659000   
1.580000   
std      1443.520003      0.956651              0.552069   
0.568915   
min    100001.000000      0.000000              0.000000   
0.000000   
25%    101250.750000      0.000000              0.000000   
1.000000   
50%    102500.500000      1.000000              1.000000   
2.000000   
75%    103750.250000      2.000000              1.000000   
2.000000   
max    105000.000000      3.000000              2.000000   
2.000000   

       savings score  Years_At_Present_Employment score  \
count    5000.000000                        5000.000000   
mean        0.373000                           1.446000   
std         0.804985                           1.105137   
min         0.000000                           0.000000   
25%         0.000000                           1.000000   
50%         0.000000                           1.000000   
75%         0.000000                           3.000000   
max         3.000000                           3.000000   

       Other_Debtors_Guarantors score  Current_Address_Yrs score  
Job score  \
count                     5000.000000                5000.000000   
5000.000000   
mean                         0.145000                   0.562000   
0.926000   
std                          0.477515                   0.496191   
0.603819   
min                          0.000000                   0.000000   
0.000000   
25%                          0.000000                   0.000000   
1.000000   
50%                          0.000000                   1.000000   
1.000000   
75%                          0.000000                   1.000000   
```

```
1.000000
max                              2.000000                 1.000000
2.000000

       Property score  Score pondéré  Score pondéré binaire
count     5000.000000    5000.000000            5000.000000
mean         1.642000       0.914050               0.513000
std          1.049789       0.276339               0.499881
min          0.000000       0.230000               0.000000
25%          1.000000       0.710000               0.000000
50%          2.000000       0.910000               1.000000
75%          3.000000       1.110000               1.000000
max          3.000000       1.760000               1.000000
```

```
##pour avoir le nombre de lignes et colonnes
db.shape
```

```
(5000, 34)
```

```
##Pour avoir les types de nos colonnes
db.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 34 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Customer_ID                        5000 non-null   int64
 1   Status_Checking_Acc                5000 non-null   object
 2   Duration_in_Months                 5000 non-null   int64
 3   Credit_History                     5000 non-null   object
 4   Purposre_Credit_Taken              5000 non-null   object
 5   Credit_Amount                      5000 non-null   int64
 6   Savings_Acc                        5000 non-null   object
 7   Years_At_Present_Employment        5000 non-null   object
 8   Inst_Rt_Income                     5000 non-null   int64
 9   Marital_Status_Gender              5000 non-null   object
 10  Other_Debtors_Guarantors           5000 non-null   object
 11  Current_Address_Yrs                5000 non-null   int64
 12  Property                           5000 non-null   object
 13  Age                                5000 non-null   int64
 14  Other_Inst_Plans                   5000 non-null   object
 15  Housing                            5000 non-null   object
 16  Num_CC                             5000 non-null   int64
 17  Job                                5000 non-null   object
 18  Dependents                         5000 non-null   int64
 19  Telephone                          5000 non-null   object
 20  Foreign_Worker                     5000 non-null   object
 21  Default_On_Payment                 5000 non-null   object
 22  Customer_ID.1                      5000 non-null   int64
```

```
 23  Status score                          5000 non-null   int64
 24  credit history score                  5000 non-null   int64
 25  Credit_Amount score                   5000 non-null   int64
 26  savings score                         5000 non-null   int64
 27  Years_At_Present_Employment score  5000 non-null   int64
 28  Other_Debtors_Guarantors score        5000 non-null   int64
 29  Current_Address_Yrs score             5000 non-null   int64
 30  Job score                             5000 non-null   int64
 31  Property score                        5000 non-null   int64
 32  Score pondéré                         5000 non-null   float64
 33  Score pondéré binaire                 5000 non-null   int64
dtypes: float64(1), int64(19), object(14)
memory usage: 1.3+ MB
```

```
##pour avoir les noms de colonnes
db.columns
```

```
Index(['Customer_ID', 'Status_Checking_Acc', 'Duration_in_Months',
       'Credit_History', 'Purposre_Credit_Taken', 'Credit_Amount',
       'Savings_Acc', 'Years_At_Present_Employment', 'Inst_Rt_Income',
       'Marital_Status_Gender', 'Other_Debtors_Guarantors',
       'Current_Address_Yrs', 'Property', 'Age', 'Other_Inst_Plans',
'Housing',
       'Num_CC', 'Job', 'Dependents', 'Telephone', 'Foreign_Worker',
       'Default_On_Payment', 'Customer_ID.1', 'Status score',
       'credit history score', 'Credit_Amount score', 'savings score',
       'Years_At_Present_Employment score', 'Other_Debtors_Guarantors
score',
       'Current_Address_Yrs score', 'Job score', 'Property score',
       'Score pondéré', 'Score pondéré binaire'],
      dtype='object')
```

```
##Diviser notre data
##db = datascore + data
##datascore pour entrainer notre model et data pour creer des
visualisations

datascore =db.drop(['Customer_ID', 'Status_Checking_Acc',
'Duration_in_Months',
       'Credit_History', 'Purposre_Credit_Taken', 'Credit_Amount',
       'Savings_Acc', 'Years_At_Present_Employment', 'Inst_Rt_Income',
       'Marital_Status_Gender', 'Other_Debtors_Guarantors',
       'Current_Address_Yrs', 'Property', 'Age', 'Other_Inst_Plans',
'Housing',
       'Num_CC', 'Job', 'Dependents', 'Telephone', 'Foreign_Worker',
       'Default_On_Payment', 'Customer_ID.1'],axis=1)
data = db.drop(['Customer_ID','Customer_ID.1', 'Status score',
       'credit history score', 'Credit_Amount score', 'savings score',
       'Years_At_Present_Employment score', 'Other_Debtors_Guarantors
score',
```

```
       'Current_Address_Yrs score', 'Job score', 'Property score',
       'Score pondéré', 'Score pondéré binaire'],axis=1)

datascore.head()

   Status score  credit history score  Credit_Amount score  savings
score  \
0             1                     0                    2
0
1             2                     1                    1
0
2             0                     0                    2
0
3             1                     1                    1
0
4             1                     0                    1
0

   Years_At_Present_Employment score  Other_Debtors_Guarantors
score  \
0                                  3                         0

1                                  1                         0

2                                  2                         0

3                                  2                         2

4                                  1                         0


   Current_Address_Yrs score  Job score  Property score Score pondéré
\
0                          1          1               3          1,06

1                          0          1               3          1,03

2                          1          0               3          0,71

3                          1          1               2          1,12

4                          1          1               0          0,58


   Score pondéré binaire
0                      1
1                      1
2                      0
3                      1
4                      0
```

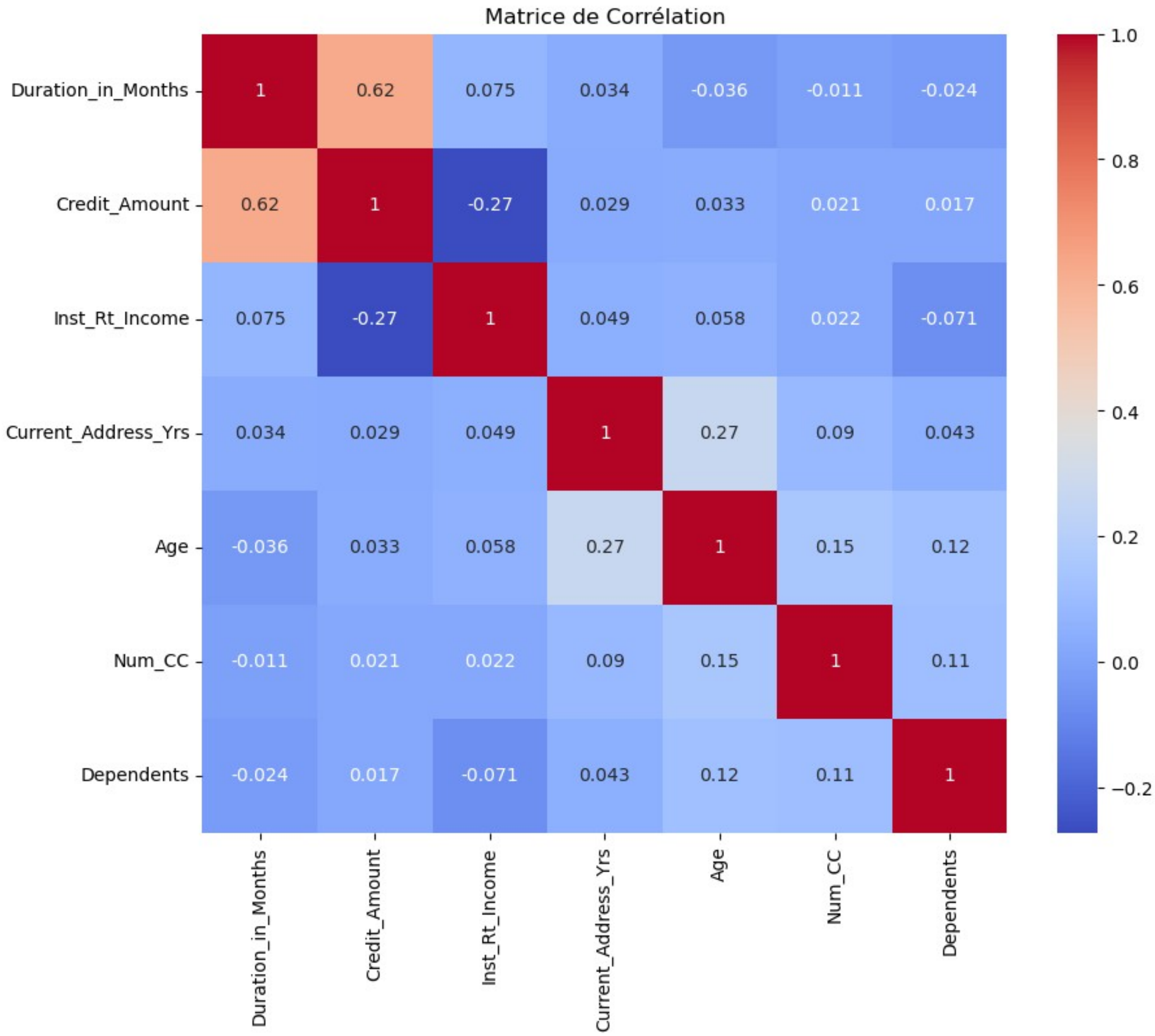
```
data.head()
```

```
   Status_Checking_Acc  Duration_in_Months  \
0             … < 0 USD                   6
1       0 <= … < 10000                   48
2  no checking account                   12
3             … < 0 USD                  42
4             … < 0 USD                  24

                                      Credit_History
Purposre_Credit_Taken  \
0  critical account/other credits existing(not at...
radio/television
1           existing credits paid back duly till now
radio/television
2  critical account/other credits existing(not at...
education
3           existing credits paid back duly till now
furniture/equipment
4                    delay in paying off in the past             car
(new)

   Credit_Amount                  Savings_Acc
Years_At_Present_Employment  \
0           1169  unknown/ no savings account               .. >= 7
years
1           5951                 … < 1000 USD           1 <= … < 4
years
2           2096                 … < 1000 USD           4 <= … < 7
years
3           7882                 … < 1000 USD           4 <= … < 7
years
4           4870                 … < 1000 USD           1 <= … < 4
years

   Inst_Rt_Income             Marital_Status_Gender
Other_Debtors_Guarantors  \
0               4                       male  single
none
1               2  female divorced/separated/married
none
2               2                       male  single
none
3               2                       male  single
guarantor
4               3                       male  single
none

   ...                                              Property Age  \
0  ...                                           real estate  67
```

```
1  ...                                        real estate  22
2  ...                                        real estate  49
3  ...  building society savings agreement/life insurance  45
4  ...                              unknown / no property  53

   Other_Inst_Plans   Housing Num_CC                          Job
Dependents  \
0              none       own      2  skilled employee / official
1
1              none       own      1  skilled employee / official
1
2              none       own      1         unskilled - resident
2
3              none  for free      1  skilled employee / official
2
4              none  for free      2  skilled employee / official
2

                                   Telephone Foreign_Worker
Default_On_Payment
0  yes, registered under the customer’s name            yes
Defaulted
1                                       none            yes
Defaulted
2                                       none            yes
Defaulted
3                                       none            yes
Defaulted
4                                       none            yes         No
Default

[5 rows x 21 columns]
```

Etude de corrélation entre les variables

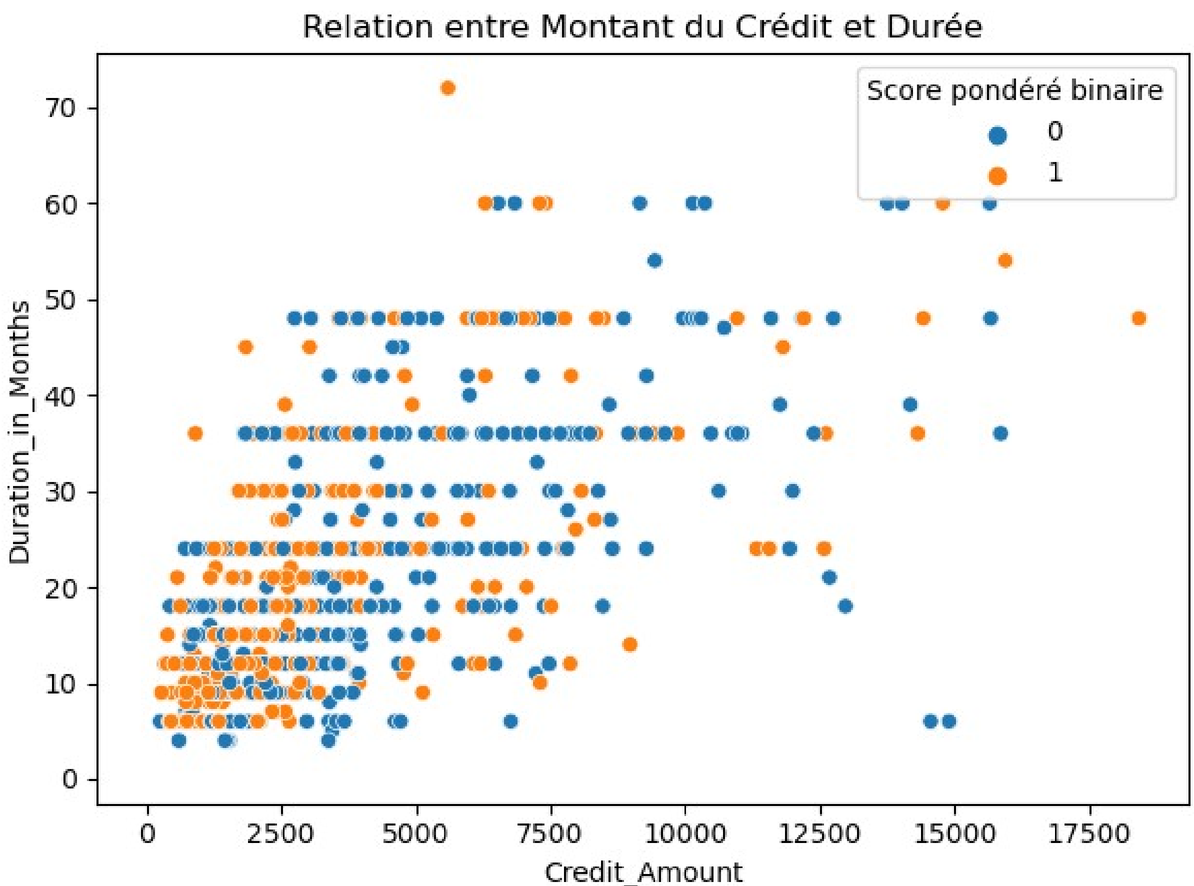
```
# Calculer la matrice de corrélation
correlation_matrix = data.corr()

# Visualiser la matrice de corrélation avec une heatmap de seaborn
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.title("Matrice de Corrélation")
plt.show()
```



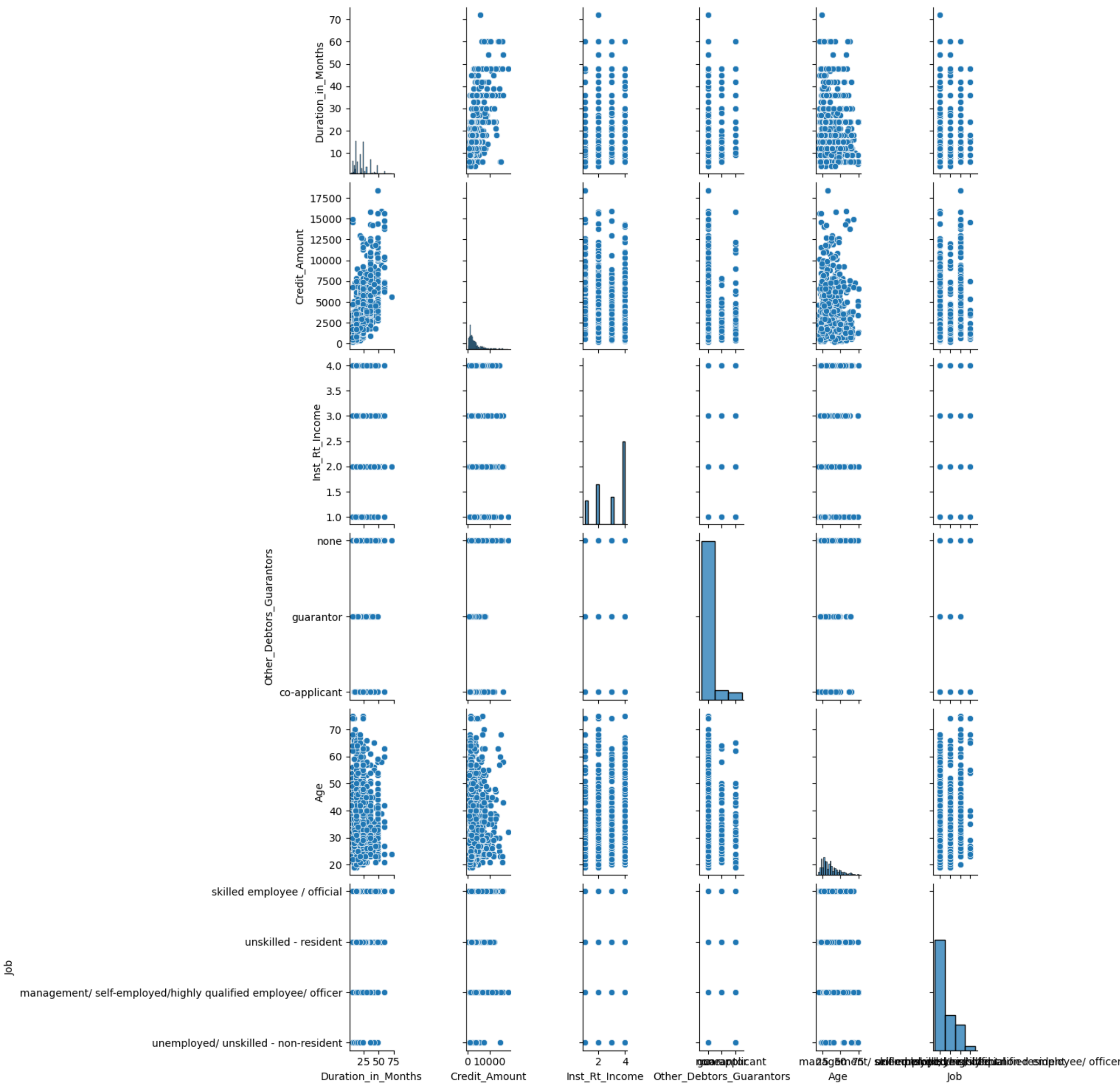
Créer des plots pour comprendre mieux notre dataset

```
# Créer des scatter plots
plt.figure(figsize=(12, 6))

<Figure size 1200x600 with 0 Axes>

<Figure size 1200x600 with 0 Axes>

# Créer un scatter plot pour etudier la relation entre le montant de
crédit et la durée
plt.plot(1, 2, 1)
sns.scatterplot(data=db, x="Credit_Amount", y="Duration_in_Months",
```

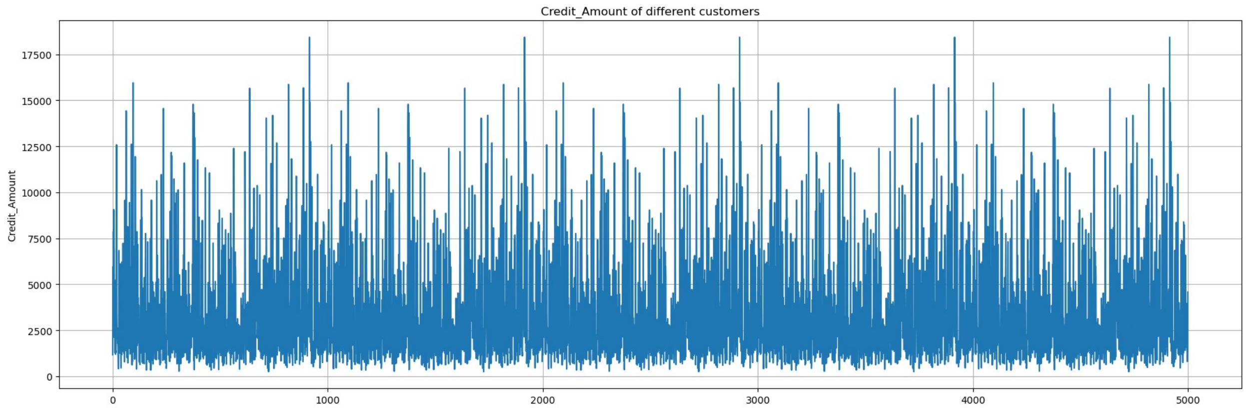
```
hue="Score pondéré binaire")
plt.title("Relation entre Montant du Crédit et Durée")
plt.tight_layout()
plt.show()
```




```
# Drawing pairplot
sns.pairplot(db,
             vars=[ 'Duration_in_Months', 'Credit_Amount',
'Inst_Rt_Income','Other_Debtors_Guarantors',
         'Age', 'Job'],
             y_vars=['Score pondéré binaire'])
plt.show()
```

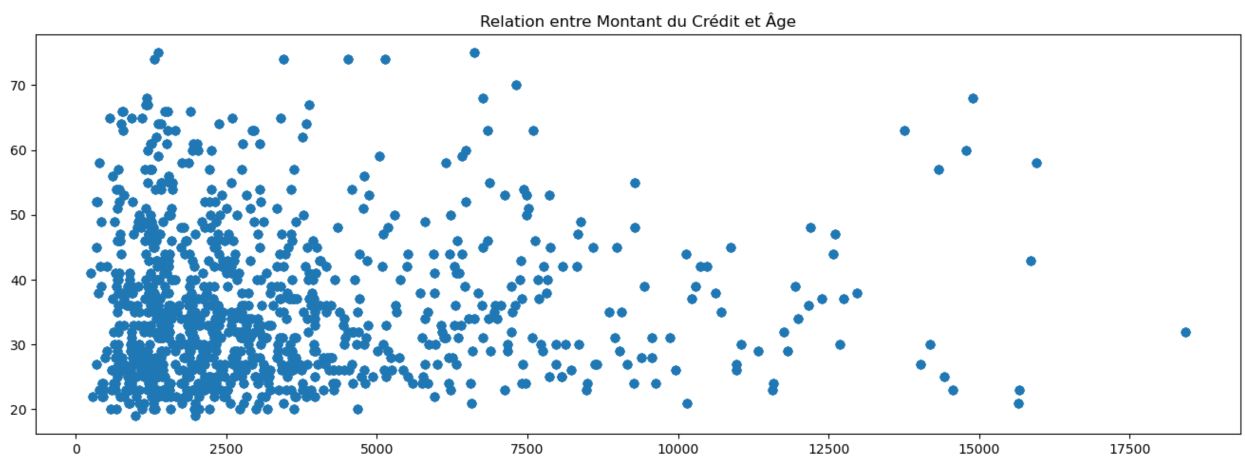



```python
# Visualisation des montants de credits
plt.figure(figsize=(18, 6))
data['Credit_Amount'].plot()
plt.title("Credit_Amount of different customers")
plt.ylabel("Credit_Amount")
plt.grid(True)
plt.tight_layout()
plt.show()
```



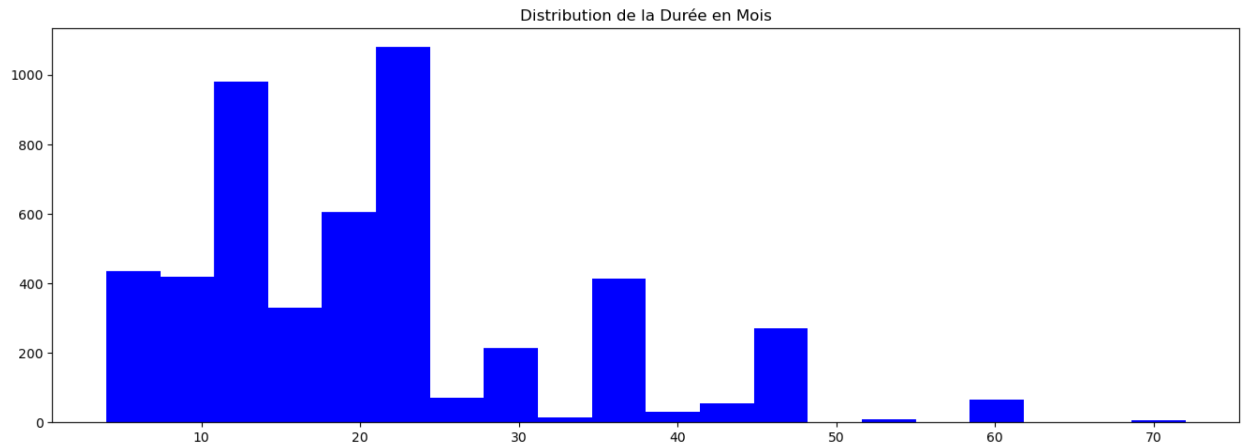

```
### Créer de visualisation pour voir la Relation entre Montant du Crédit et Âge
fig = plt.figure(figsize=(36, 12))
spec = gridspec.GridSpec(2, 2, figure=fig)

ax1 = fig.add_subplot(spec[0, 0])
ax1.scatter(db["Credit_Amount"],db["Age"])
ax1.set_title("Relation entre Montant du Crédit et Âge")
```

```
Text(0.5, 1.0, 'Relation entre Montant du Crédit et Âge')
```



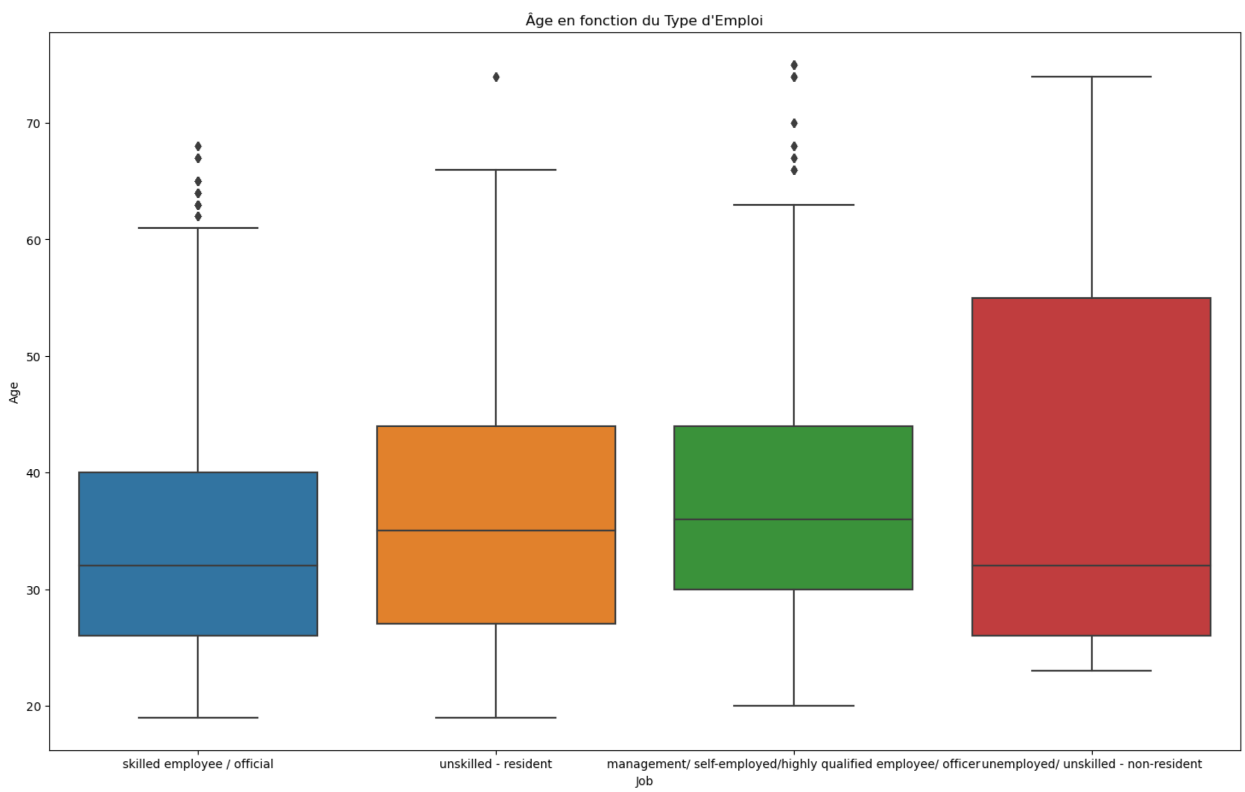

```
## Distribution de la Durée en Mois
fig = plt.figure(figsize=(36, 12))
spec = gridspec.GridSpec(2, 2, figure=fig)

ax2 = fig.add_subplot(spec[0, 1])
ax2.hist(db["Duration_in_Months"], bins=20, color="blue")
ax2.set_title("Distribution de la Durée en Mois")
```

```
Text(0.5, 1.0, 'Distribution de la Durée en Mois')
```

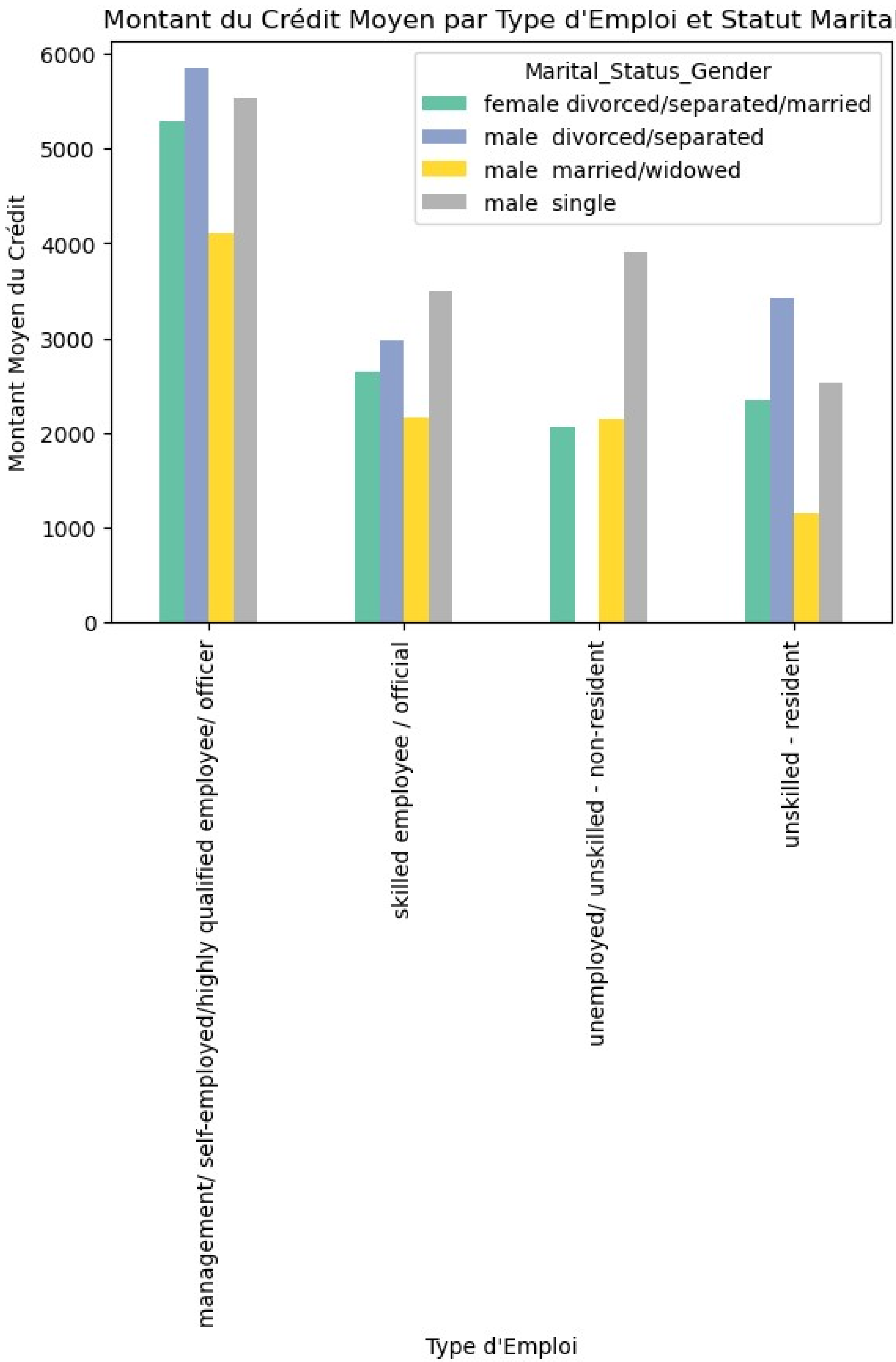



```python
##Âge en fonction du Type d'Emploi
fig = plt.figure(figsize=(18, 24))
spec = gridspec.GridSpec(2, 2, figure=fig)

ax3 = fig.add_subplot(spec[1, :])
sns.boxplot(data=db, x="Job", y="Age")
ax3.set_title("Âge en fonction du Type d'Emploi")
```

```
Text(0.5, 1.0, "Âge en fonction du Type d'Emploi")
```



```
##Montant du Crédit Moyen par Type d'Emploi et Statut Marital
pivot_table = db.pivot_table(index="Job",
columns="Marital_Status_Gender", values="Credit_Amount",
aggfunc="mean")
pivot_table.plot(kind="bar", cmap="Set2")
plt.title("Montant du Crédit Moyen par Type d'Emploi et Statut
Marital")
plt.ylabel("Montant Moyen du Crédit")
plt.xlabel("Type d'Emploi")
plt.show()
```



Montant du Crédit Moyen par Type d'Emploi et Statut Marital
Montant Moyen du Crédit
6000
5000
4000
3000
2000
1000
0
Marital_Status_Gender
female divorced/separated/married
male divorced/separated
male married/widowed
male single
management/ self-employed/highly qualified employee/ officer
skilled employee / official
unemployed/ unskilled - non-resident
unskilled - resident
Type d'Emploi

Application de modèle régression logestique avec 80% de données entrainées et 20% à tester

```
X = datascore.drop(["Score pondéré binaire"], axis=1)  # Supprimer la colonnes Y
y = datascore["Score pondéré binaire"]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LogisticRegression()

model.fit(X_train, y_train)

y_pred = model.predict(X_test)
```

Mesuring of Accuracy and R-squared (R2)

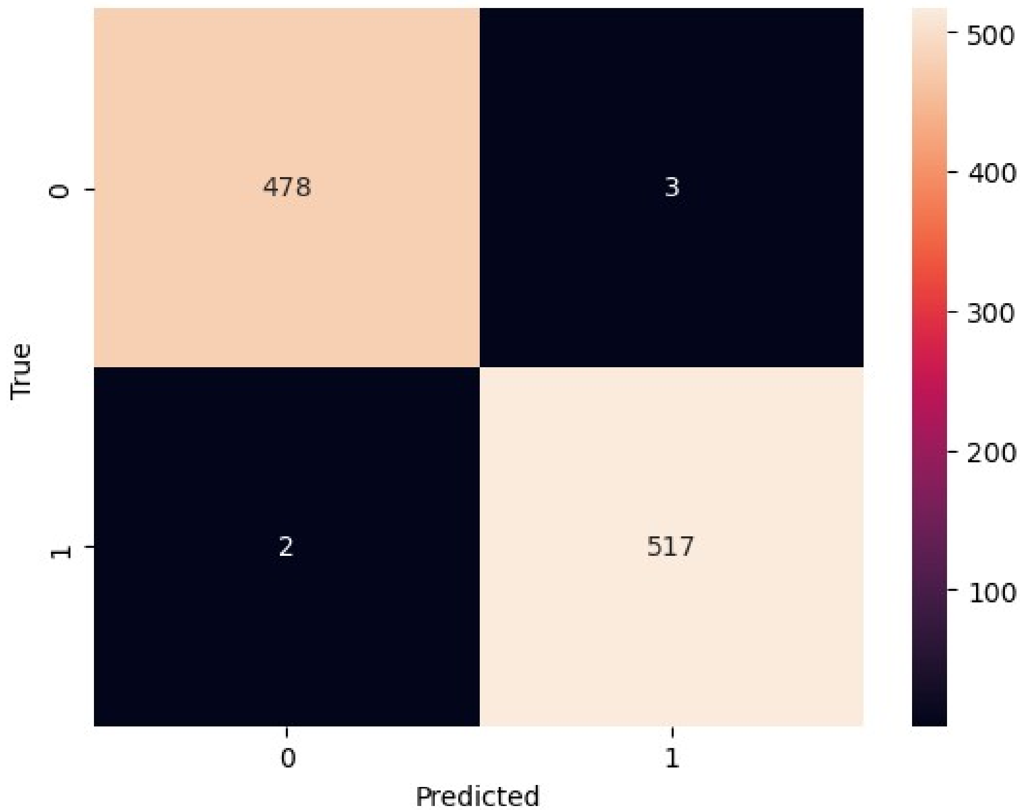
```
accuracy = accuracy_score(y_test, y_pred)

print("Précision du modèle:", accuracy)
```

```
Précision du modèle: 0.995
```

```
r2 = r2_score(y_test, y_pred)
print("R-squared value:", r2)
```

```
R-squared value: 0.9799710782369742
```

```
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt="d")
plt.xlabel('Predicted')
plt.ylabel('True')
plt.show()
```



478
3
2
517
True
Predicted
0
1
0
1
500
400
300
200
100