

A medical stethoscope with a teal tube and silver chest piece is resting on a medical chart. A blue pen is also visible in the background. The chart contains a table with a header 'Datum:' and a list of times: 13:30, 13:45, 14:00. The text 'PAVADE Aishwarya et PHIMPHRACHANH Julie' is visible in the bottom right corner of the chart.

# ***Diabetes 130-US hospitals for years 1999-2008***

PAVADE Aishwarya et PHIMPHRACHANH Julie



# ***Collecte des données***

- Date de collection : entre 1999 et 2008 (10 ans)
- Lieu de collection : 130 hôpitaux américains et réseaux de prestation intégrés
- Nombre de caractéristiques : +50 attributs
- Caractéristiques : numéro du patient, race, sexe, âge, la diagnostic, les médicaments pour le diabète, la réadmission



# Problématique

- Objectif : Réadmission à l'hôpital des patients diabétiques en se basant sur des données des patients autrefois réadmis ou non
- Interêt : Indique la qualité de l'hôpital et de son soin
- Coût hôpitaux : améliorer la qualité de vos soins en connaissant les facteurs de readmission pour réduire leurs nombre et ainsi éviter le surplus de coût liées au soins
- La réadmission est notre réponse. On le considère comme une variable qualitative/catégorique binaire. C'est donc une étude de classification.



# ***Etapes***

- Etape 1 : Data exploration and data cleaning (nb valeurs nulles, nettoyer certaines valeurs)
- Etape 2 : Features Engineering and Visualization ( lien entre notre variable de prédiction et les autres variables)
- Etape 3 : Train-Test Splitting
- Etape 4 : Encoding
- Etape 5 : Standardization and Normalization
- Etape 6 : Model creation
- Etape 7 : Hyperparameters Tuning

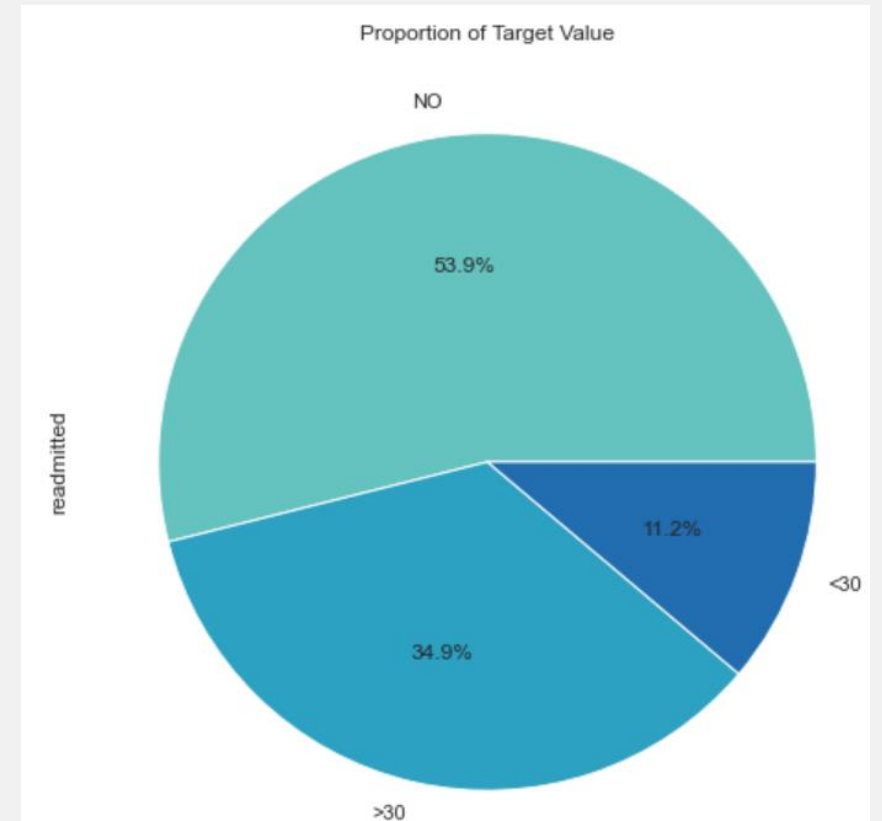
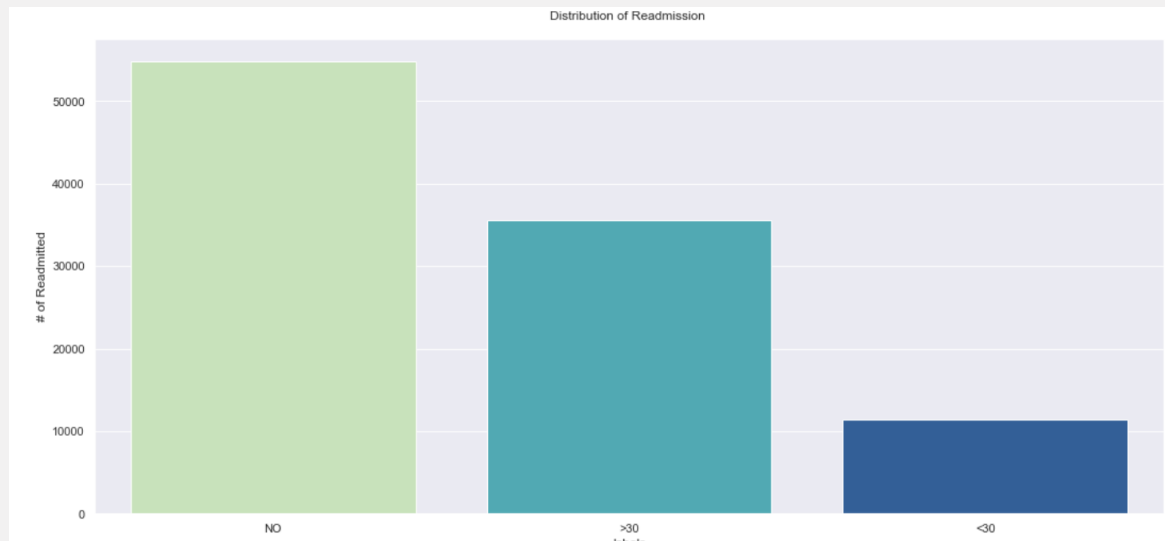


# ***Visualisation***

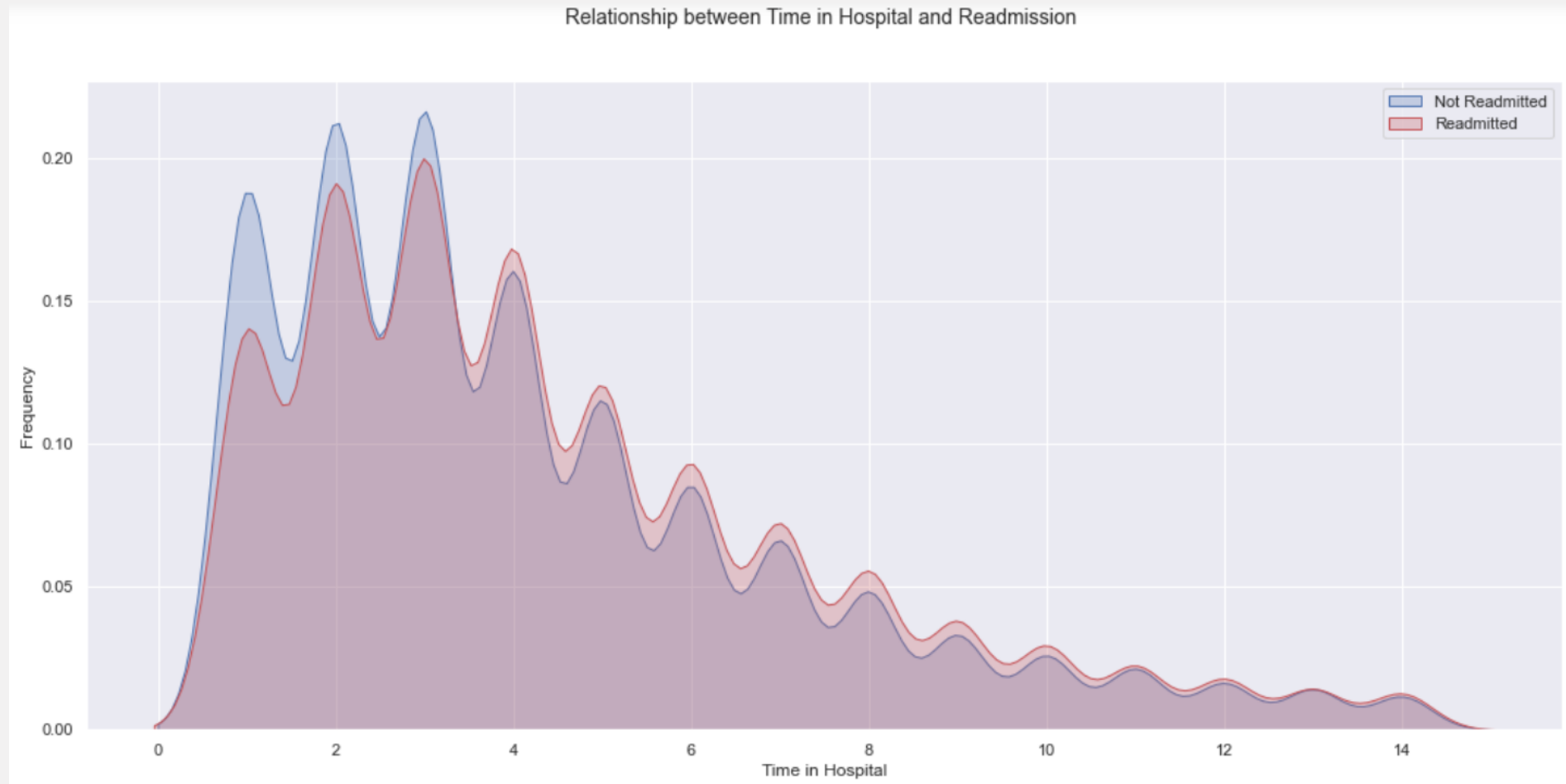
Lien entre notre variable de prédiction et nos autres variables



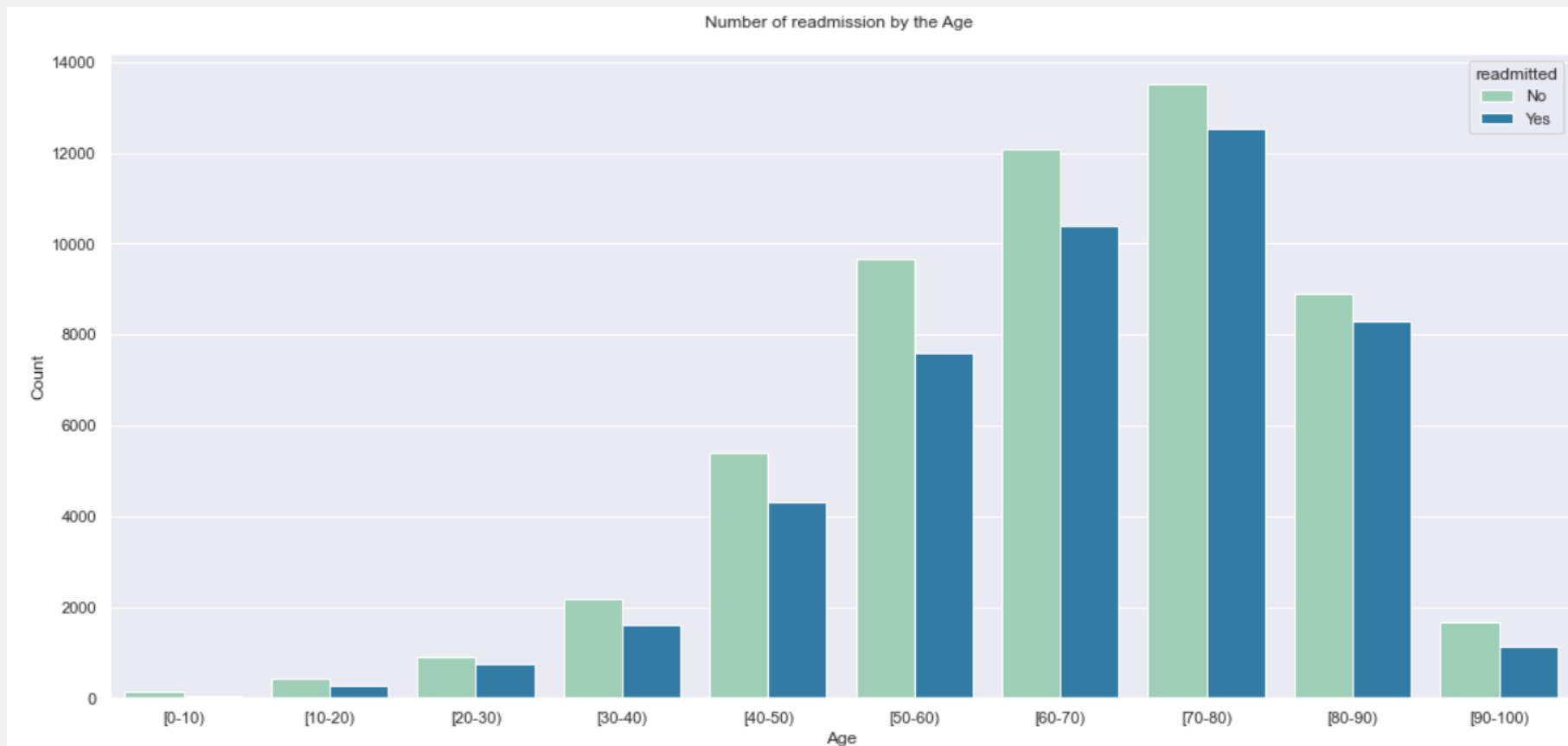
# Combien de patients sont réadmis ?



# ***Lien entre la durée du séjour à l'hôpital et la réadmission du patient***

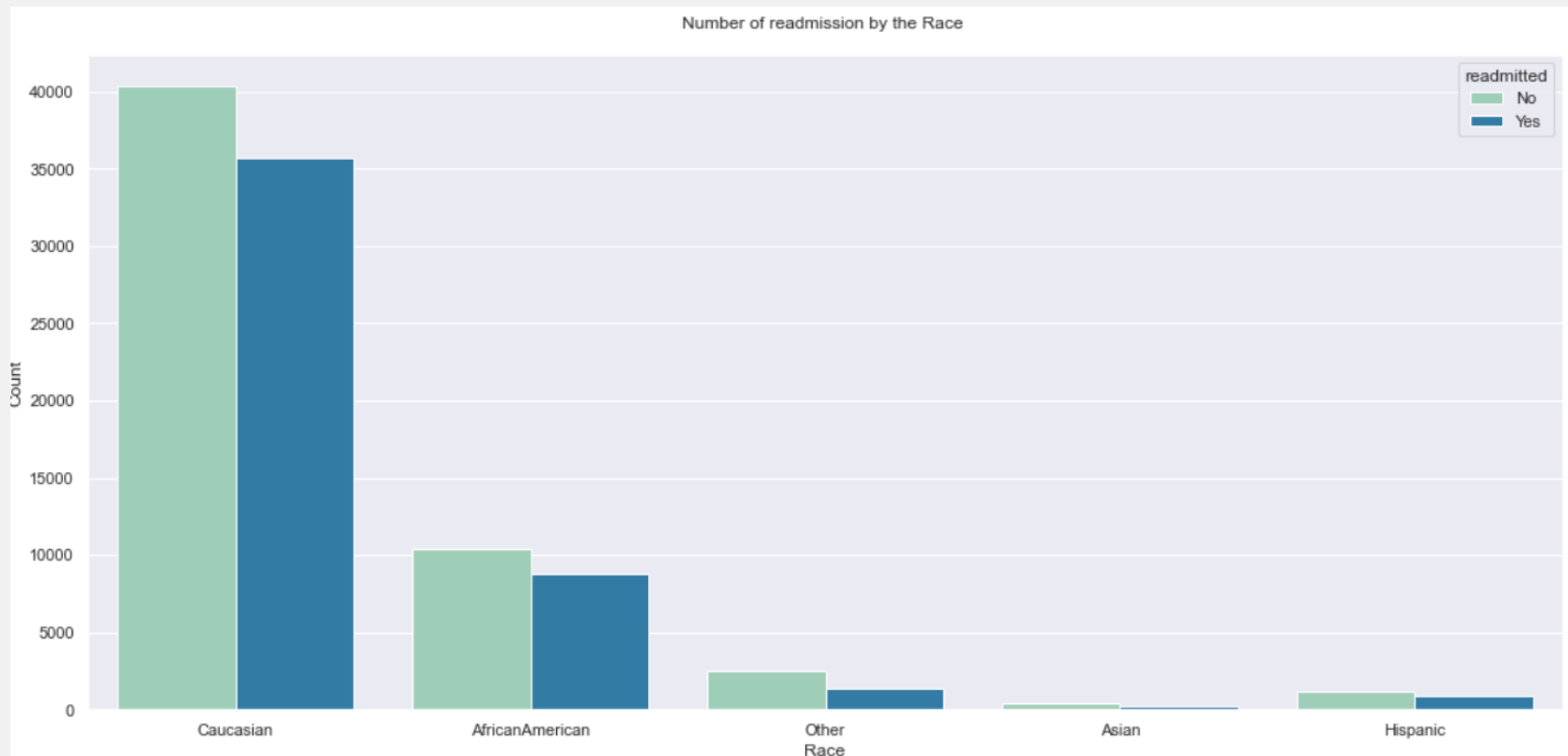


# ***Lien entre l'âge du patient et sa réadmission à l'hôpital***

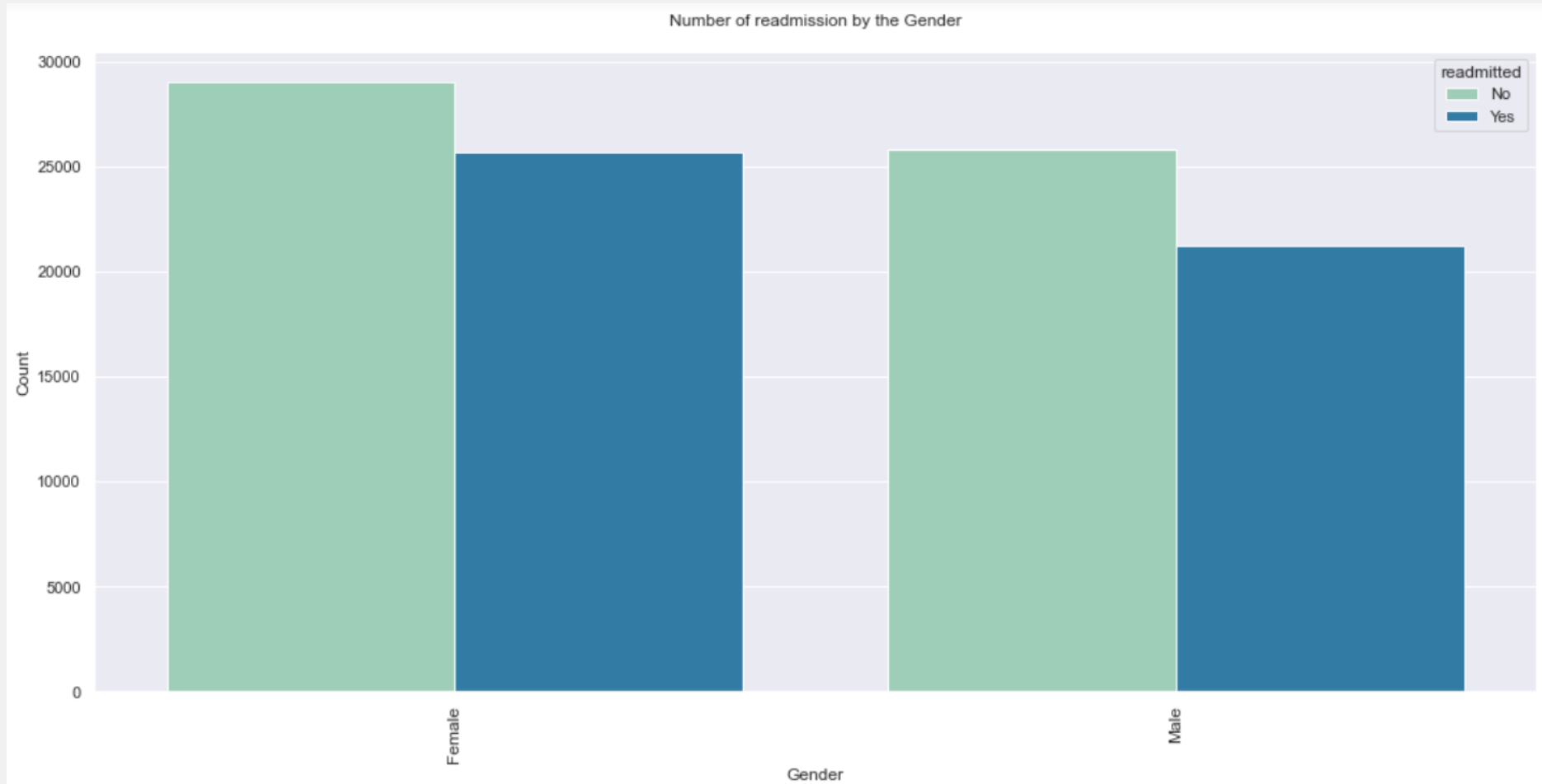




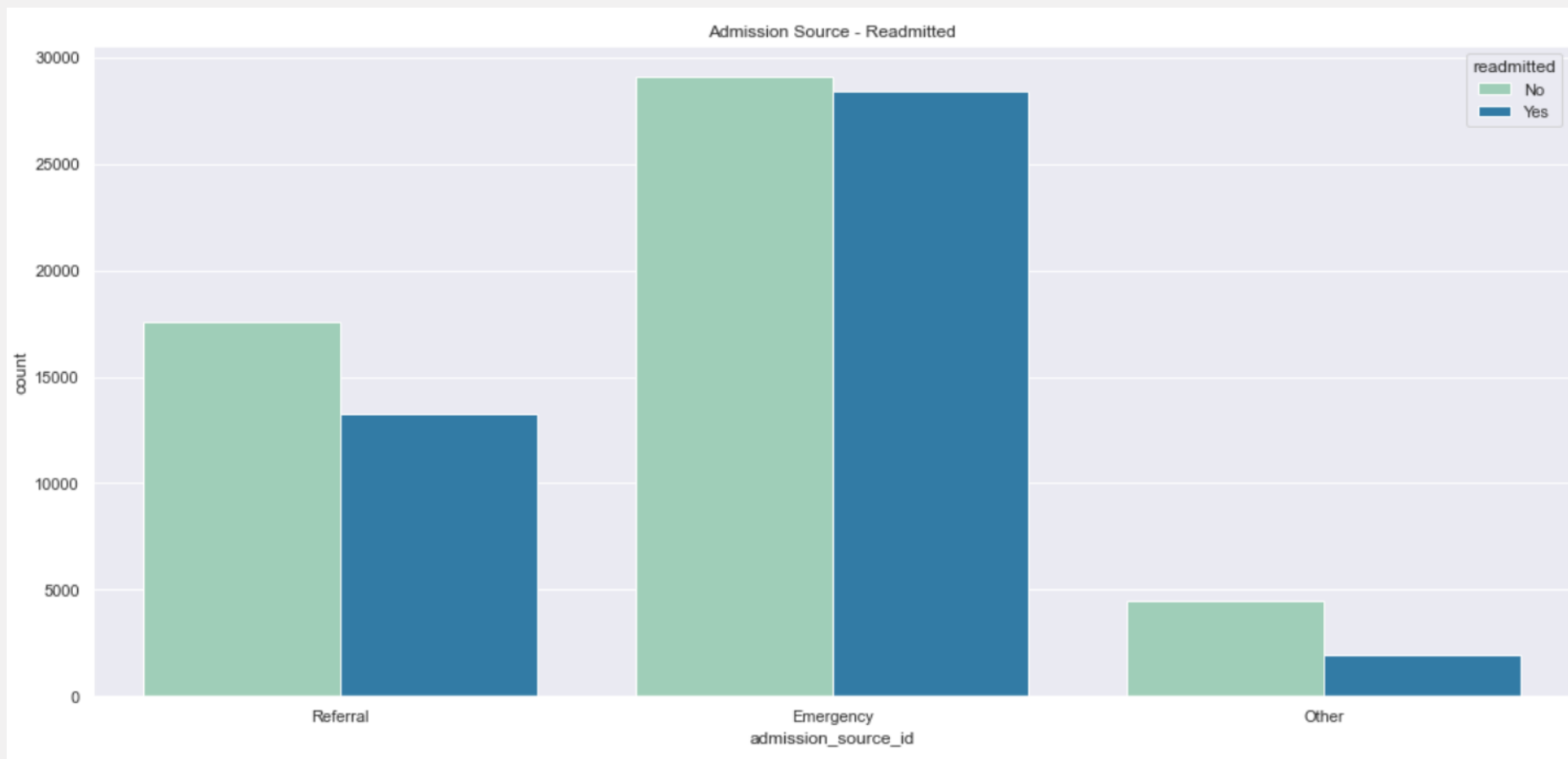
# ***Lien entre la race du patient et sa réadmission à l'hôpital***



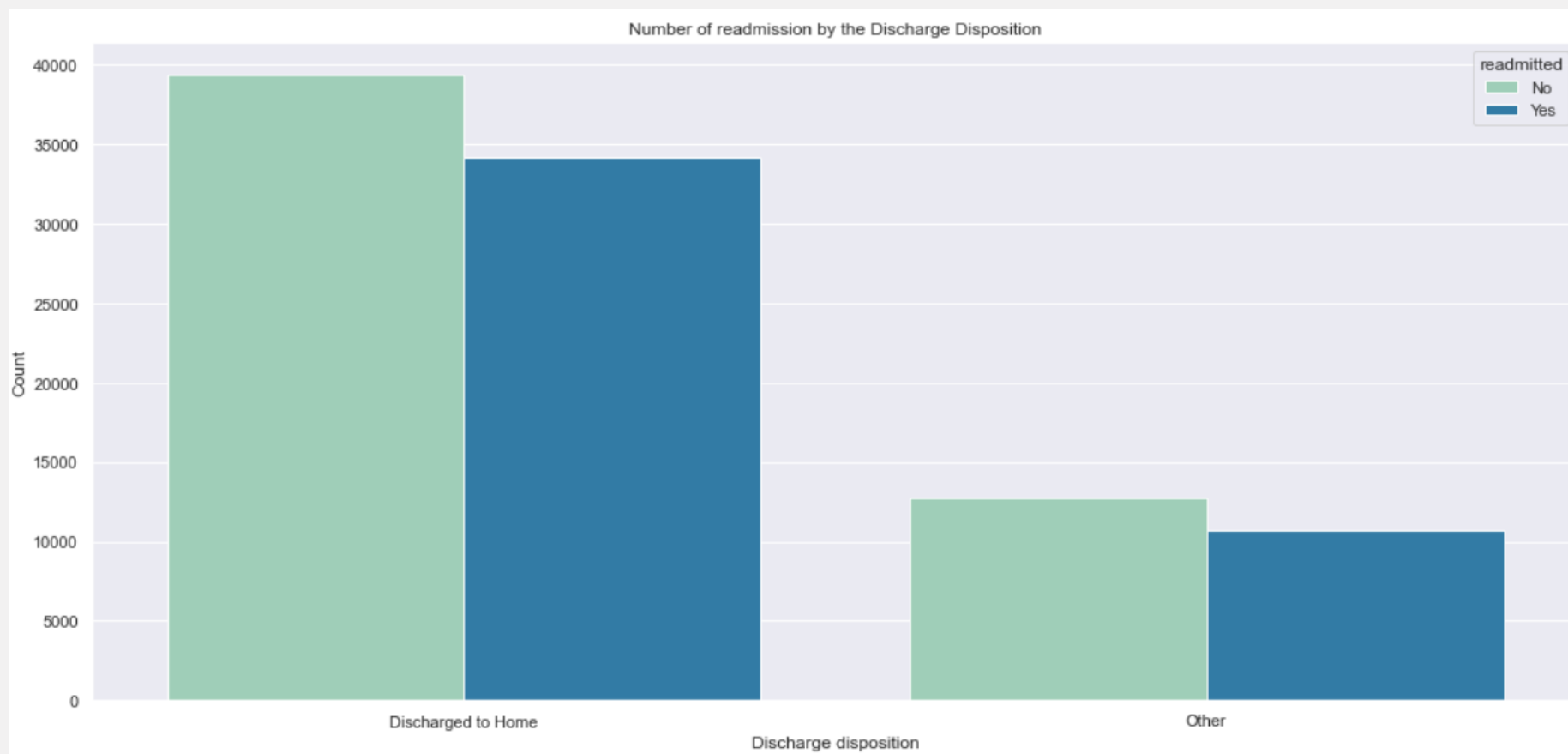
# ***Lien entre le genre du patient et sa réadmission à l'hôpital***



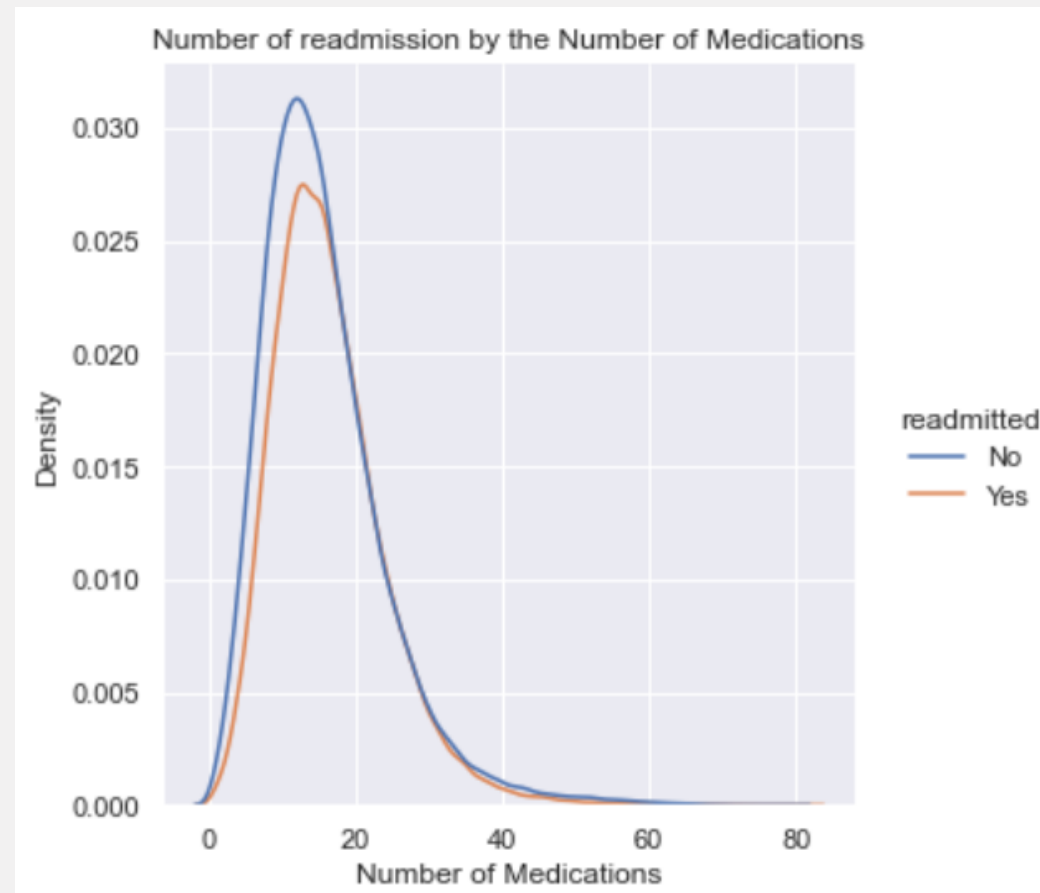
# ***Lien entre l'ID de la source d'admission et la réadmission***



# ***Lien entre le type de discharge et la réadmission du patient***



# ***Lien entre le nombre de médicament et la réadmission du patient***



# **Models**

- Logistic regression
- Linear SVC
- Linear Discriminant Analysis
- Decision Tree Classifier
- Random Forest Classifier
- Gradient Boosting Classifier
- Cat Boost Classifier

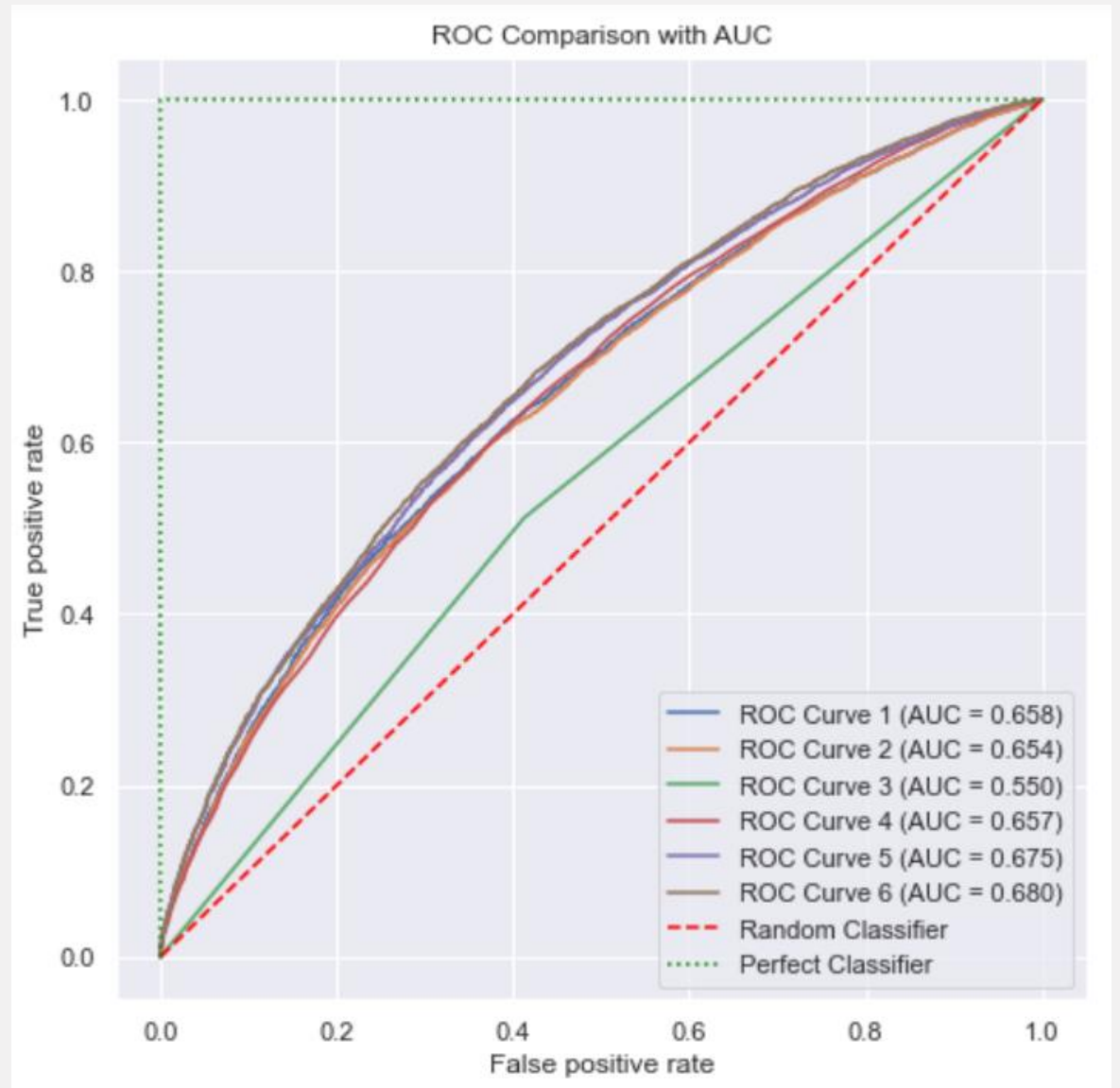


# Quels modèles choisir

- Situation 1 : Le modèle prédit que le patient sera réadmis (Predicted Positive) mais en fin de compte il n'est pas réadmis (Actual Negative) => False Positive
- Situation 2 : Le modèle prédit que le patient ne sera pas réadmis (Predicted Negative) mais finalement il est réadmis (Actual Positive) => False Negative
- La situation 1 n'est pas très grave mais la situation 2 est à éviter car cela signifiera que notre modèle n'est pas correct.
- On doit se concentrer sur le recall de nos modèles car il y a un coût plus important associé à la situation 2 (False Negative).



# **ROC / AUC**





# ***Modèles choisis***

- Random Forest Classifier

- + Train :

- Accuracy : 0.999841281292836
    - Recall : 0.9997804816859006

- + Test :

- Accuracy : 0.6165522017660299
    - Recall : 0.5013888888888889

- Cat Boost Classifier

- + Train :

- Accuracy : 0.6759685448380348
    - Recall : 0.5488585047666834

- + Test :

- Accuracy : 0.6347896346742079
    - Recall : 0.5023989898989899



# Hyperparameter Tuning

- Random Forest Classifier

- + Train :

- Accuracy : 0.6415121564100714
    - Recall : 0.49037255393878576

- + Test :

- Accuracy : 0.6302879898424424
    - Recall : 0.4726010101010101

- Cat Boost Classifier

- + Train :

- Accuracy : 0.7280138518144434
    - Recall : 0.6172541394882087

- + Test :

- Accuracy : 0.6314999711433024
    - Recall : 0.51010101010101



# API FLASK

```
1 # Save your model
2 import joblib
3 joblib.dump(cbc_opt_model, 'model.pkl')
4 print("Model dumped!")
5
6 # Load the model that you just saved
7 cbc_opt_model = joblib.load('model.pkl')
8
9 # Saving the data columns from training
10 model_train_columns = list(X_train.columns)
11 joblib.dump(model_train_columns, 'model_columns.pkl')
12 print("Models columns dumped!")
```

Model dumped!

Models columns dumped!

Dans notre fichier où on a  
créé notre modèle de  
Machine Learning



# API FLASK

## Notre fichier API.ipynb

```
Model loaded
Model columns loaded
* Serving Flask app "__main__" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: off
```

```
* Running on http://127.0.0.1:12345/ (Press CTRL+C to quit)
127.0.0.1 - - [06/Dec/2022 00:18:01] "GET /predict HTTP/1.1" 405 -
127.0.0.1 - - [06/Dec/2022 00:18:01] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [06/Dec/2022 00:20:58] "POST /predict HTTP/1.1" 200 -
```

None

```
127.0.0.1 - - [06/Dec/2022 00:41:51] "POST /predict HTTP/1.1" 200 -
```

```
[{'Intercept': 1, 'race': 'Caucasian', 'gender': 'Female', 'age': 1, 'admission_type_id': 0, 'discharge_disposition': 'Discharged to Home', 'admission_source_id': 0, 'time_in_hospital': 3, 'num_lab_procedures': 59, 'num_procedures': 0, 'num_medications': 18, 'number_outpatient': 0, 'number_emergency': 0, 'number_inpatient': 0, 'number_diagnoses': 9, 'max_glu_serum': 'None', 'A1Cresult': 'None', 'metformin': 'No', 'repaglinide': 'No', 'glimepiride': 'No', 'glipizide': 'No', 'glyburide': 'No', 'pioglitazone': 'No', 'rosiglitazone': 'No', 'citoglipton': 'No', 'insulin': 'Yes', 'change': 'Ch', 'diabetesMed': 'Yes'}]
```



# POSTMAN

POST

localhost:22222/predict

Send

Params

Authorization

Headers (8)

Body

Pre-request Script

Tests

Settings

Cookies

none

form-data

x-www-form-urlencoded

raw

binary

GraphQL

JSON

Beautify

```
1  [  
2    {  
3      "Intercept":1,"race":"Caucasian","gender":"Male","age":1,"admission_type_id":0,  
4      "discharge_disposition":"Discharged to Home","admission_source_id":0,"time_in_hospital":3,  
      "num_lab_procedures":10,"num_procedures":0,"num_medications":0,"number_outpatient":0,  
      "number_emergency":0,"number_inpatient":0,"number_diagnoses":9,"max_glu_serum":"None",  
      "A1Cresult":"None","metformin":"Yes","repaglinide":"No","glimepiride":"No","glipizide":"No",  
      "glyburide":"No","pioglitazone":"No","rosiglitazone":"No","citoglipton":"No","insulin":"Yes",  
      "change":"Ch","diabetesMed":"Yes"  
    }  
  ]
```

Body

Cookies

Headers (4)

Test Results

200 OK

83 ms

166 B

Save Response

Pretty

Raw

Preview

Visualize

JSON

```
1  {  
2    "prediction": "[0]"  
3  }
```



# POSTMAN

POST

localhost:22222/predict

Send

Params

Authorization

Headers (8)

Body

Pre-request Script

Tests

Settings

Cookies

none

form-data

x-www-form-urlencoded

raw

binary

GraphQL

JSON

Beautify

```
1  [  
2  ...{"Intercept":1,"race":"Caucasian","gender":"Male","age":1,"admission_type_id":0,  
      "discharge_disposition":"Discharged to Home","admission_source_id":0,"time_in_hospital":3,  
      "num_lab_procedures":34,"num_procedures":0,"num_medications":2,"number_outpatient":0,  
      "number_emergency":0,"number_inpatient":0,"number_diagnoses":9,"max_glu_serum":"None",  
      "A1Cresult":"None","metformin":"Yes","repaglinide":"No","glimepiride":"No","glipizide":"No",  
      "glyburide":"No","pioglitazone":"No","rosiglitazone":"No","citoglipton":"No","insulin":"Yes",  
      "change":"Ch","diabetesMed":"Yes"}  
3  ]  
4
```

Body

Cookies

Headers (4)

Test Results

200 OK 98 ms 166 B

Save Response

Pretty

Raw

Preview

Visualize

JSON

```
1  {  
2  "prediction": "[1]"  
3  }
```



# ***Des questions ?***

Merci pour votre attention !

