

MEDICAL COST

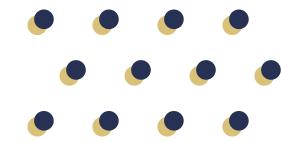


Réalisé par :
RACHIDI Widad
IMLOUL Douae

Encadré par :
M.LAMRANI



INTRODUCTION



Dans un environnement médical, comprendre les coûts médicaux est crucial pour garantir une gestion financière optimale, tant pour les patients que pour les prestataires de soins de santé. Anticiper les variations des coûts médicaux permet aux gestionnaires de mettre en place des stratégies efficaces pour réduire les dépenses imprévues et améliorer la rentabilité des services. L'importance de ce projet réside dans sa capacité à fournir une évaluation quantitative et précise des facteurs influençant les coûts médicaux, offrant ainsi aux décideurs des informations fiables pour optimiser les ressources et prendre des décisions éclairées en matière de gestion des soins de santé.





PROBLÉMATIQUE



COÛTS MÉDICAUX :: QU'EST-CE QUE C'EST ?

Les coûts médicaux englobent toutes les dépenses liées aux soins de santé, telles que les consultations, l'hospitalisation, les médicaments et les traitements. Comprendre et maîtriser ces coûts est crucial pour garantir un accès équitable aux soins tout en optimisant les ressources dans le secteur de la santé.





PROBLÉMATIQUE

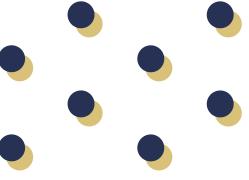
DIFFICULTÉ DE PRÉDICTION

La prédition des coûts médicaux est complexe en raison de la diversité des facteurs cliniques, économiques et sociaux qui influencent les dépenses de santé. Les coûts peuvent fluctuer en fonction de la gravité des maladies, des traitements nécessaires et des conditions spécifiques des patients.





PROBLÉMATIQUE



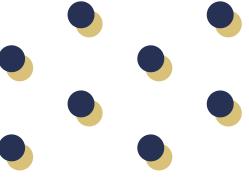
IMPORTANCE DES DONNÉES FINANCIÈRES

- **Age** : L'âge influence les coûts médicaux, avec des dépenses plus élevées généralement chez les personnes âgées.
- **Sexe** : Les besoins médicaux varient entre hommes et femmes, impactant les coûts de santé.
- **BMI (Indice de Masse Corporelle)** : Un BMI élevé est souvent lié à des problèmes de santé chroniques, entraînant des coûts plus importants.





PROBLÉMATIQUE



IMPORTANCE DES DONNÉES FINANCIÈRES

- Children : Le nombre d'enfants peut augmenter les dépenses en soins pédiatriques.
- Smoker : Les fumeurs ont des coûts médicaux plus élevés en raison des maladies liées au tabac.
- Region : Les coûts varient selon la région en fonction des politiques de santé et des services disponibles.
- Charges : Représente le coût total des soins médicaux, influencé par les facteurs ci-dessus.





MÉTHODOLOGIE

Prétraitement des Données

Construction du Modèle

- 1
- 2
- 3
- 4

Collecte des Données

Choix du Modèle de Machine Learning





MÉTHODOLOGIE



LIBRAIRIES PYTHON

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import joblib
import os
from sklearn.linear_model import LinearRegression,Lasso,Ridge,ElasticNet
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.model_selection import GridSearchCV
os.getcwd()
import warnings
warnings.filterwarnings('ignore')
```





COLLECTE DES DONNÉES



- Charges
- Age
- Sexe
- BMI (Indice de Masse Corporelle)
- Children (ENFANTS)
- Smoker (FUMEURS)
- Region





COLLECTE DES DONNÉES

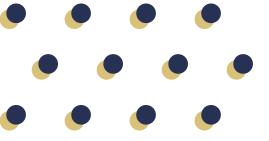
- Recherche et sélection d'une base de données contenant les variables essentielles.

```
data = pd.read_csv('insurance.csv')
```





COLLECTE DES DONNÉES



CONTENU DE LA BASE DE DONNÉES

- Total de 1338 enregistrements et 7 colonnes.
- Variable cible : Charges.
- Variables explicatives : 6 variables explicatives.

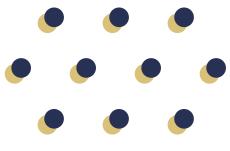
```
print("Dimensions de la BDD : ", data.shape)
```

Dimensions de la BDD : (1338, 7)





COLLECTE DES DONNÉES



CONTENU DE LA BASE DE DONNÉES

```
print("\n\n Types de variables : \n", data.dtypes)
```

```
|  
| Types de variables :  
| age           int64  
| sex          object  
| bmi         float64  
| children      int64  
| smoker        object  
| region        object  
| charges     float64  
| dtype: object
```





COLLECTE DES DONNÉES



CONTENU DE LA BASE DE DONNÉES

```
data.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520





PRÉTRAITEMENT DES DONNÉES

RÉSUMÉ DES DONNÉES

```
data.describe()
```

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010





PRÉTRAITEMENT DES DONNÉES

VÉRIFICATION DES DONNÉES MANQUANTES

```
# Vérifier s'il y a des données manquantes  
data.isna().sum()
```

```
age          0  
sex          0  
bmi          0  
children     0  
smoker       0  
region        0  
charges       0  
dtype: int64
```

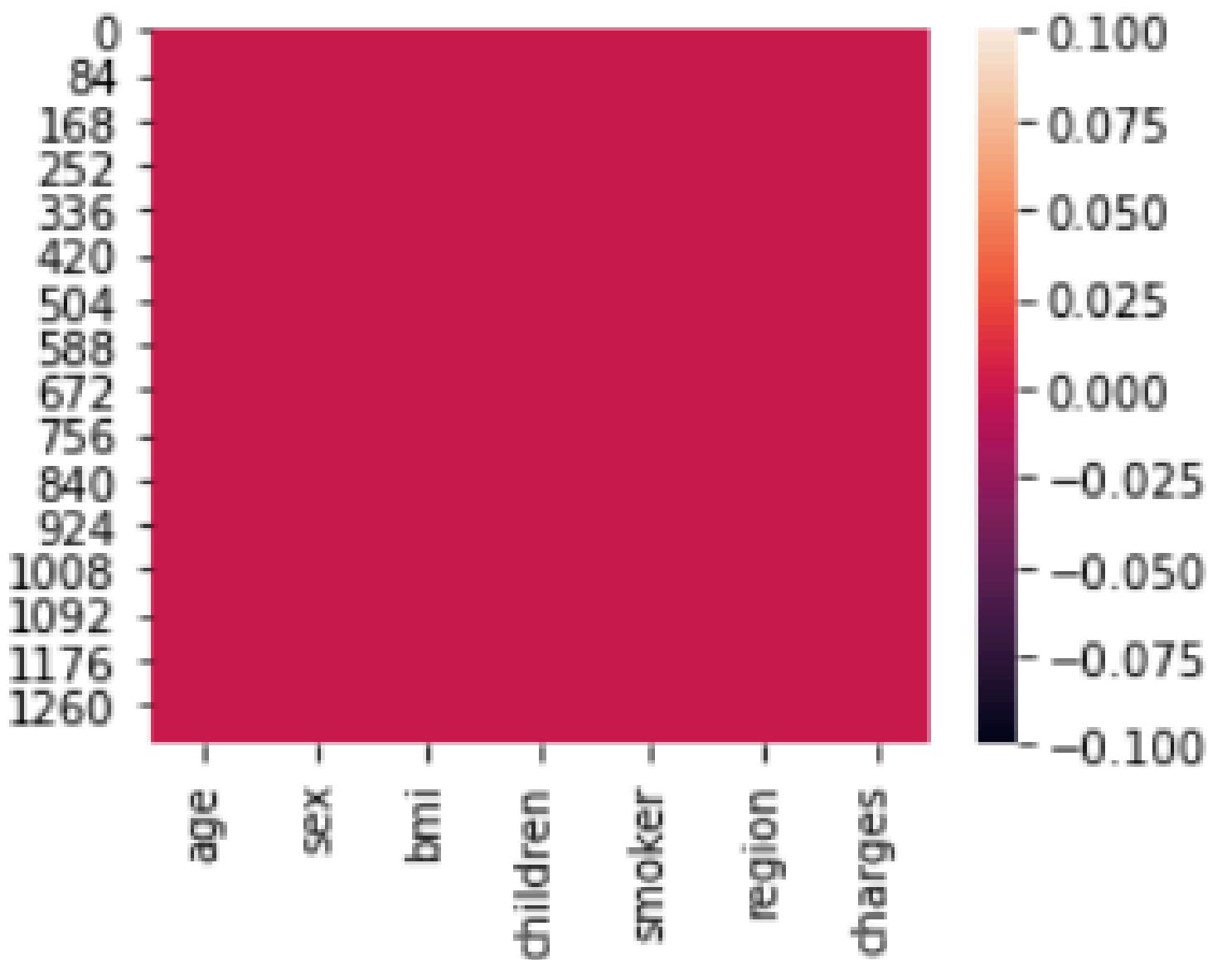


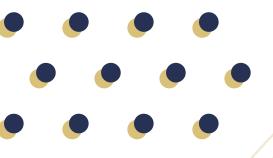


PRÉTRAITEMENT DES DONNÉES

VÉRIFICATION DES DONNÉES MANQUANTES

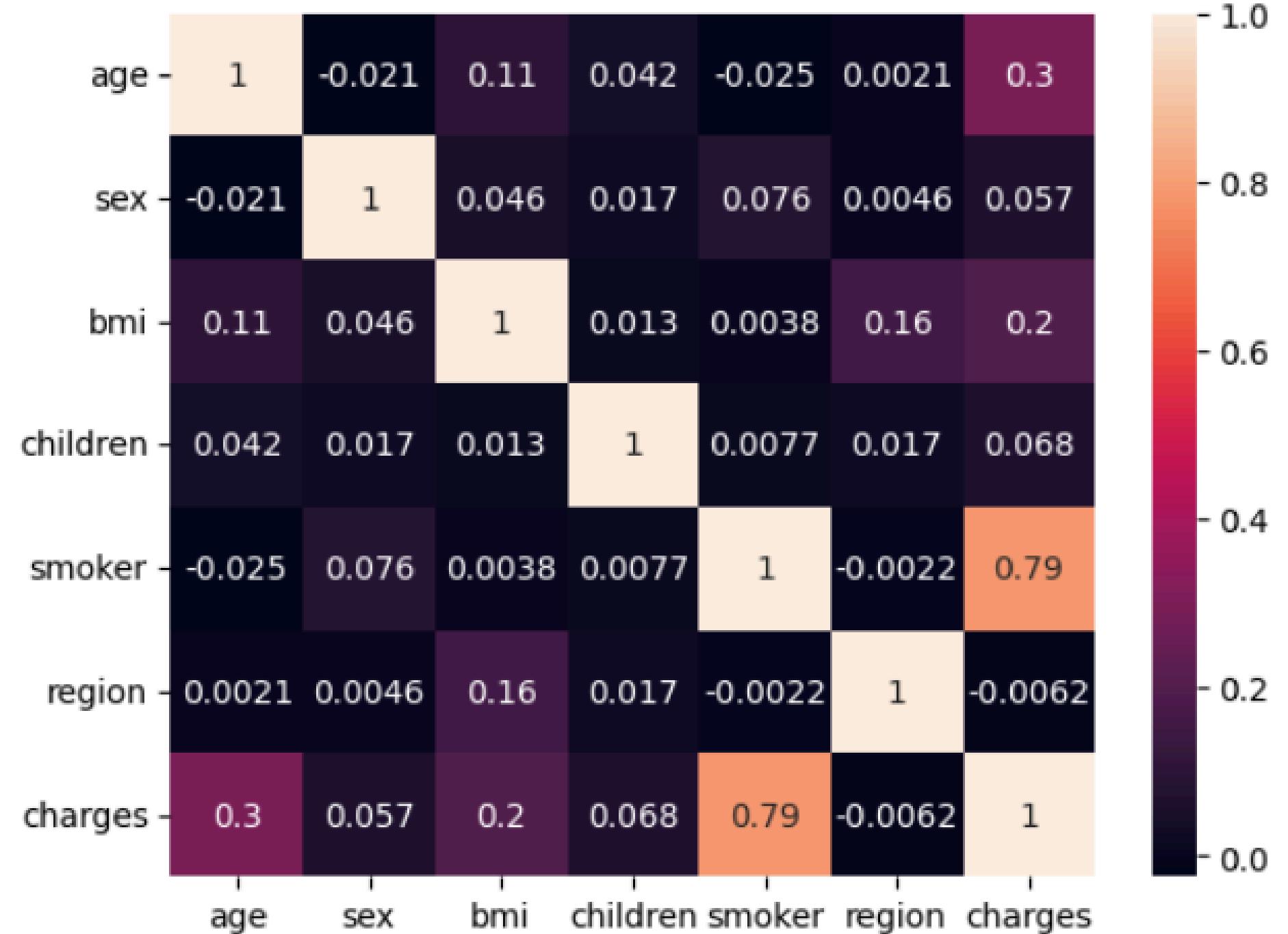
<AxesSubplot:>





PRÉTRAITEMENT DES DONNÉES

MATRICE DE CORRÉLATION





PRÉTRAITEMENT DES DONNÉES

DISTRIBUTION DES VALEURS

```
cat_features = data.select_dtypes(include = 'object').columns  
  
num_features = data.select_dtypes(include = ['int', 'float']).columns
```

```
cat_features
```

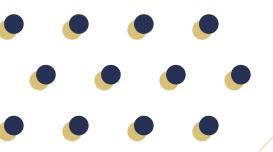
```
Index([], dtype='object')
```

```
num_features
```

```
Index(['sex', 'bmi', 'smoker', 'region', 'charges'], dtype='object')
```

```
for i in cat_features:  
  
    print(data[i].value_counts())  
  
    print("\n"+"-*50")
```





PRÉTRAITEMENT DES DONNÉES

RÉSULTAT

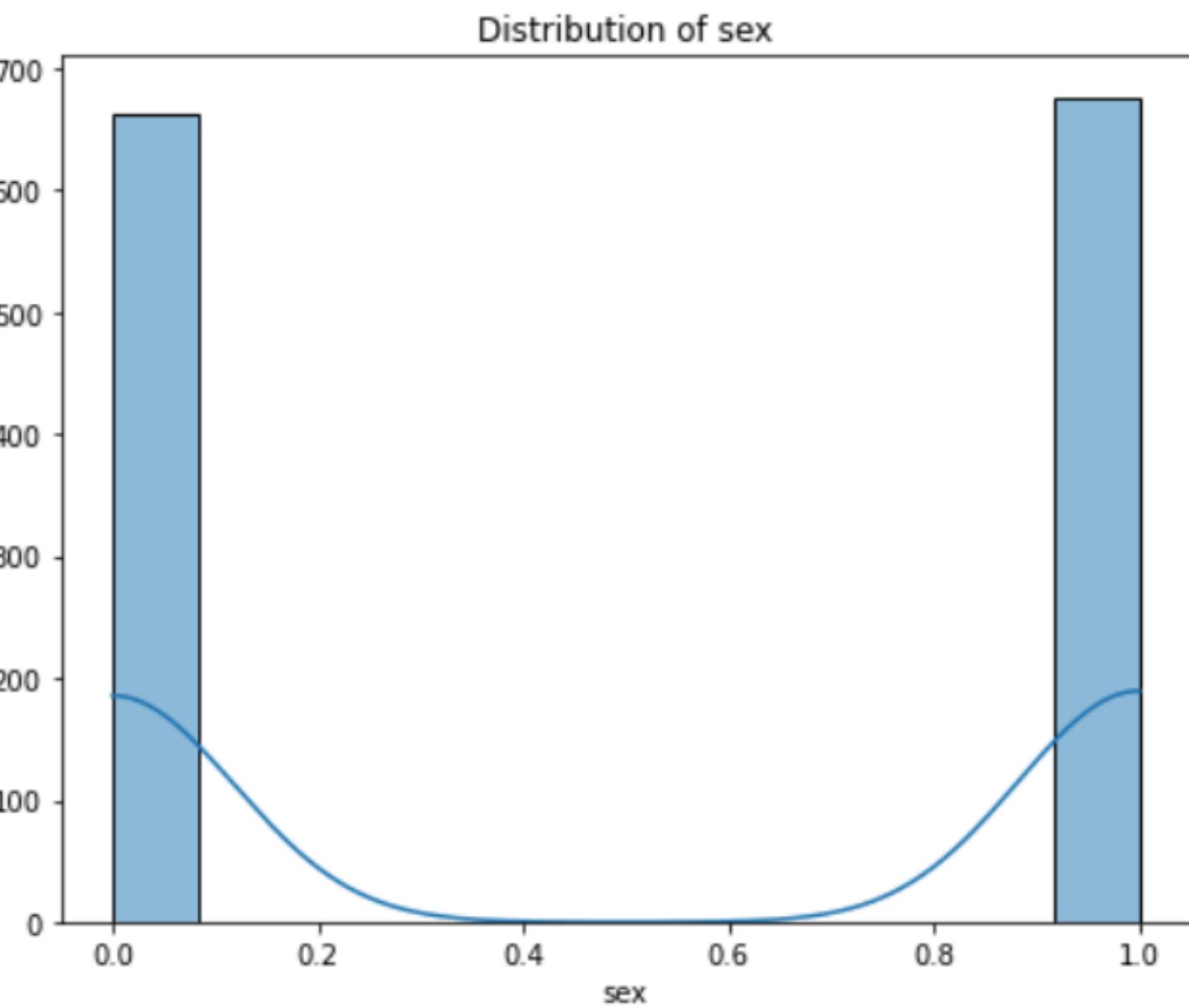
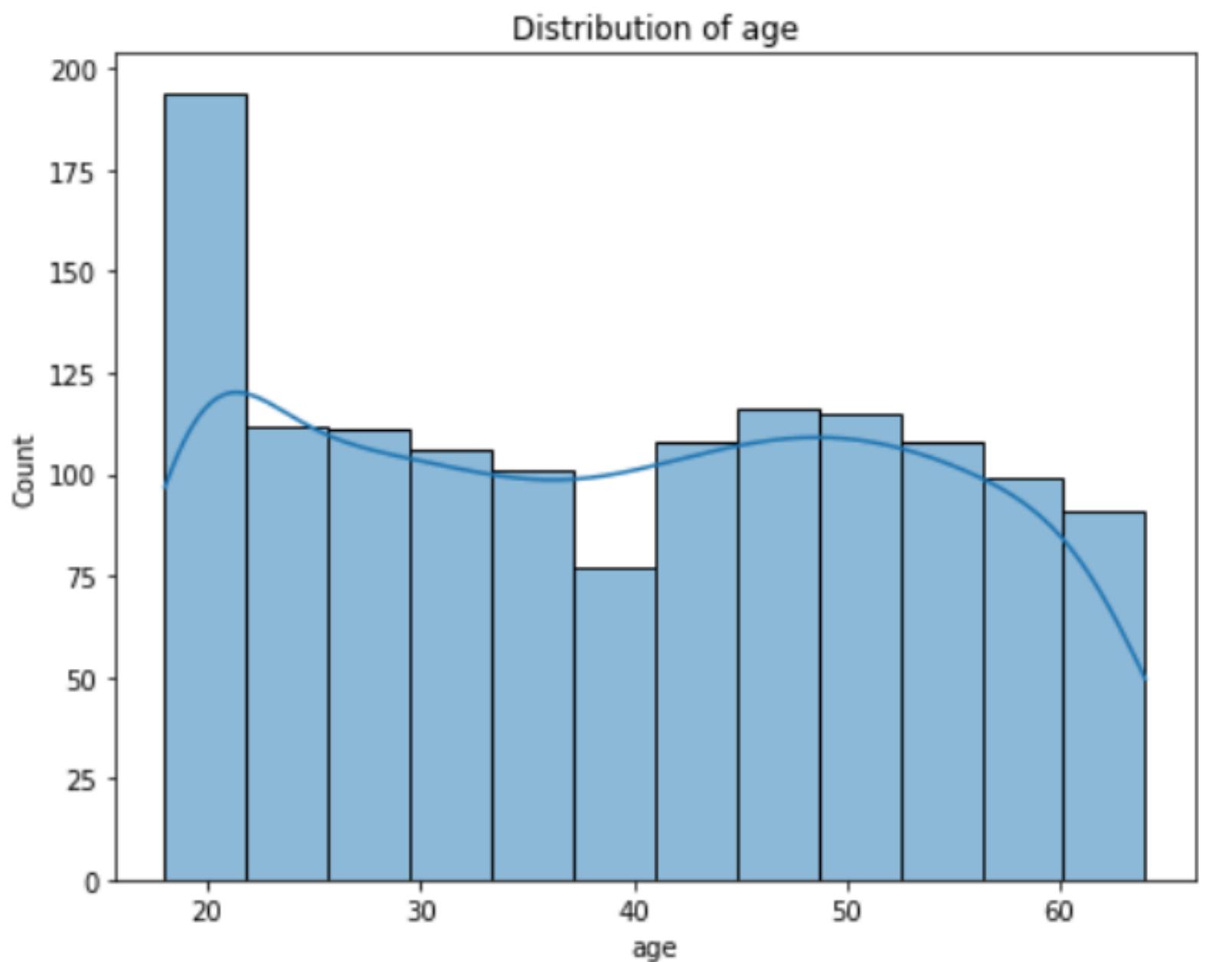
	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520





PRÉTRAITEMENT DES DONNÉES

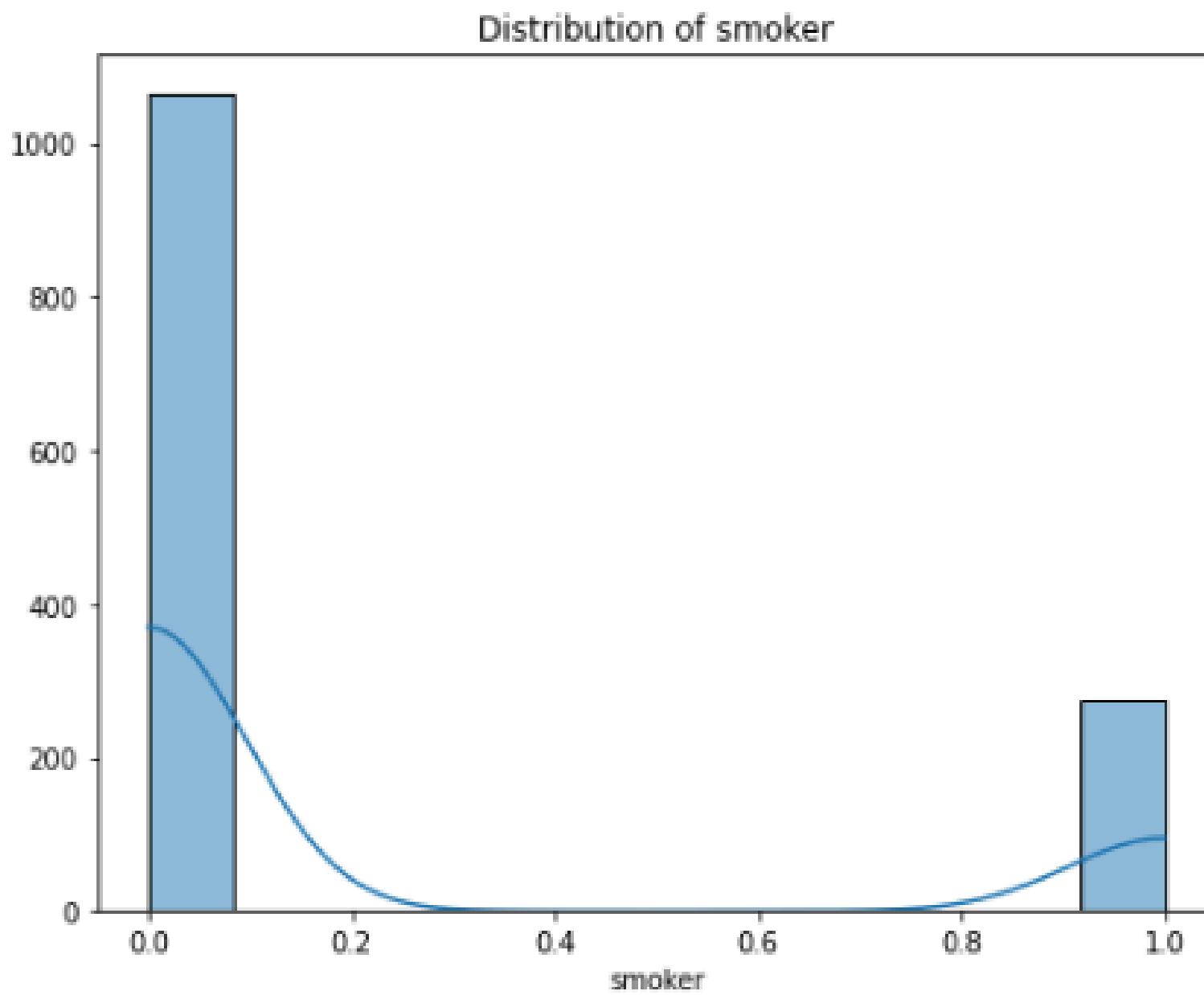
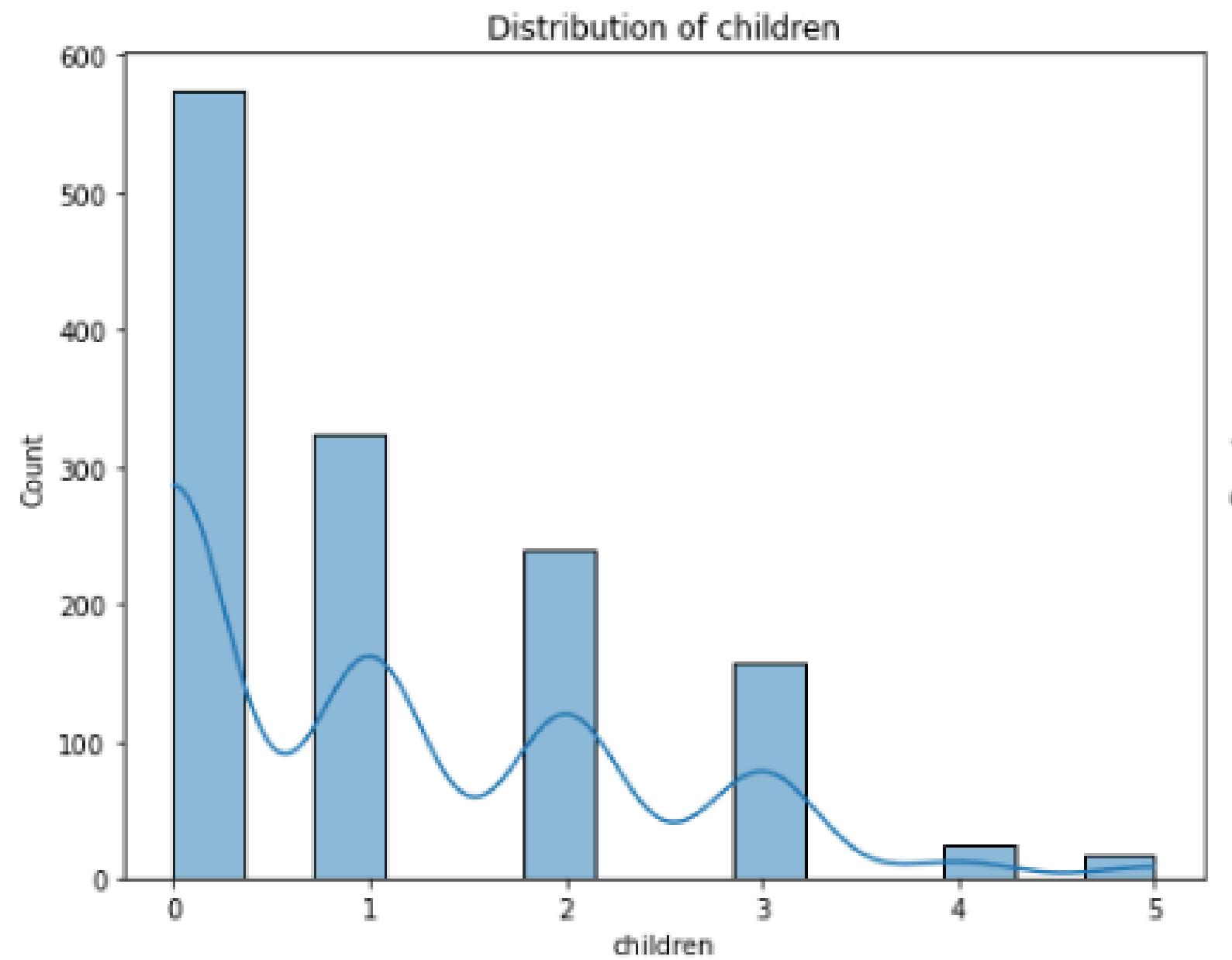
DISTRIBUTION : AGE - SEXE





PRÉTRAITEMENT DES DONNÉES

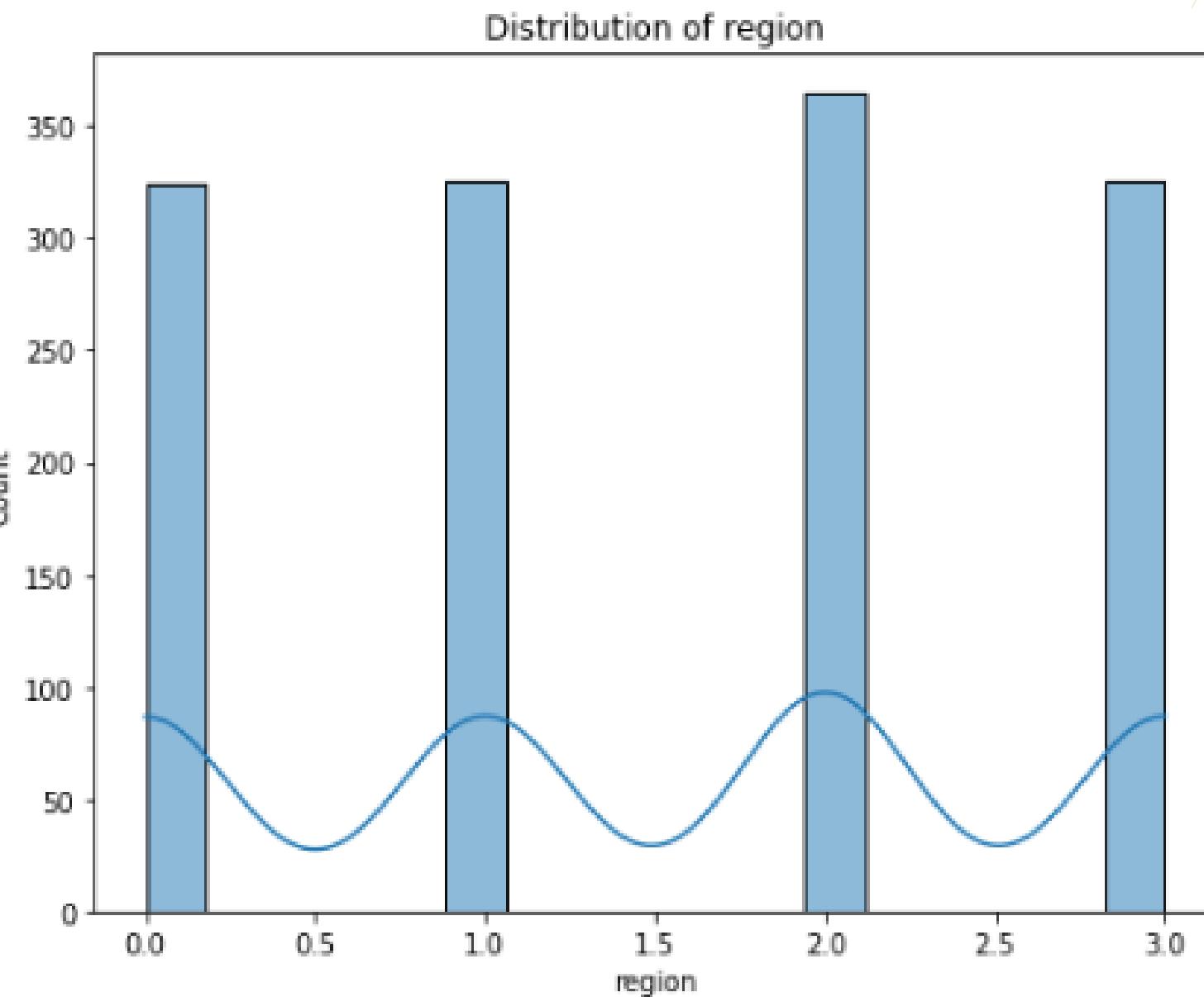
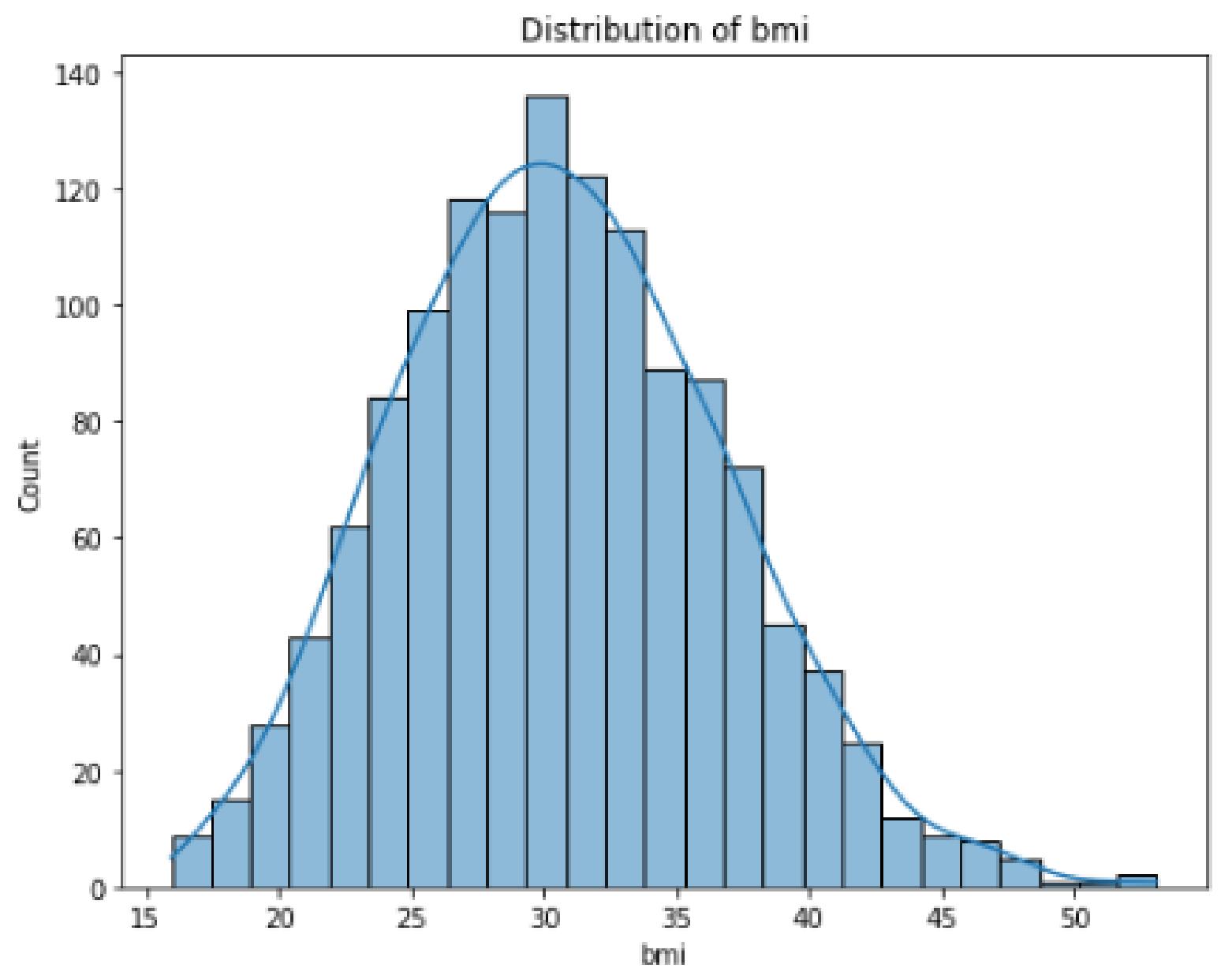
DISTRIBUTION : ENFANTS - FUMEURS





PRÉTRAITEMENT DES DONNÉES

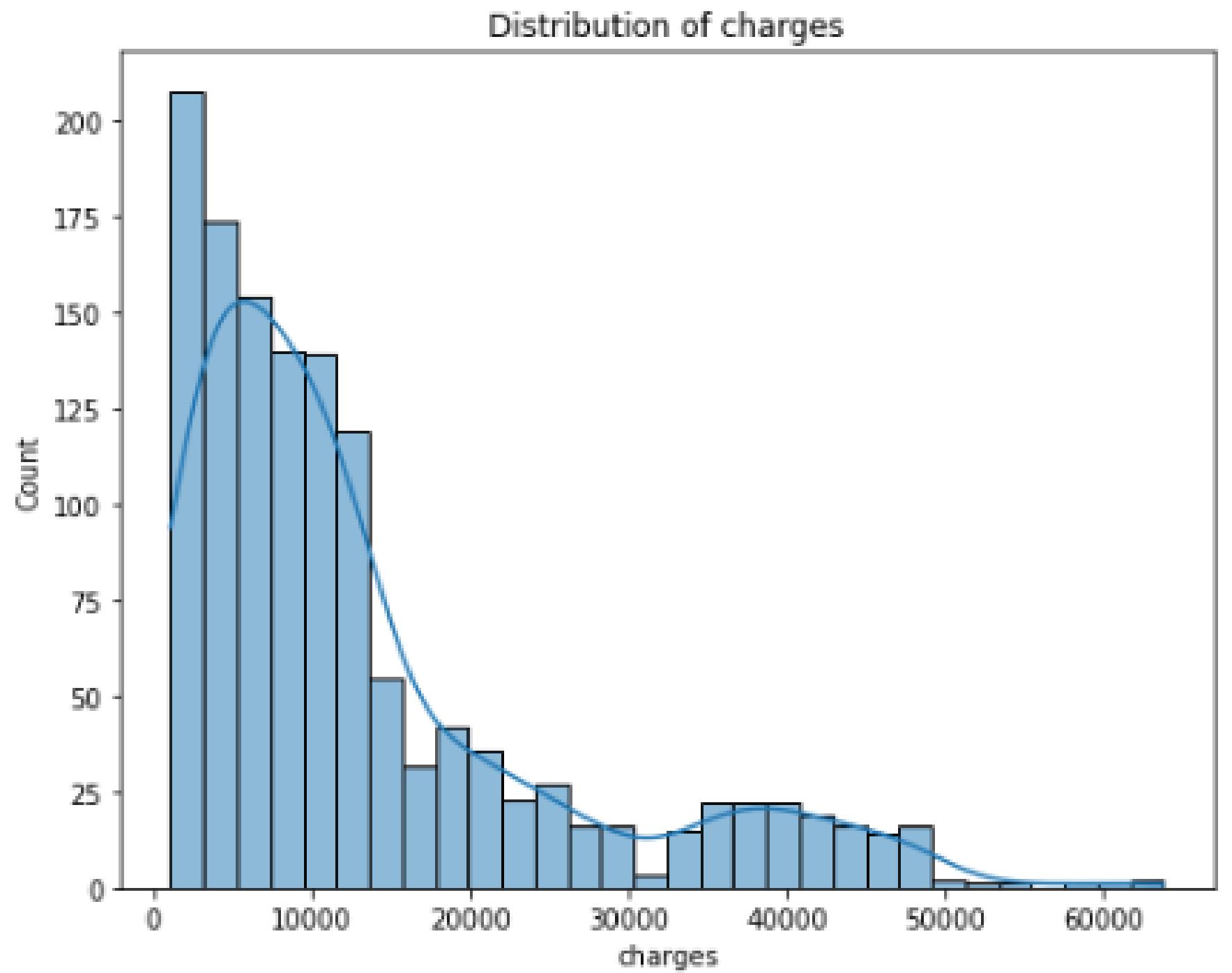
DISTRIBUTION : BMI - REGION





PRÉTRAITEMENT DES DONNÉES

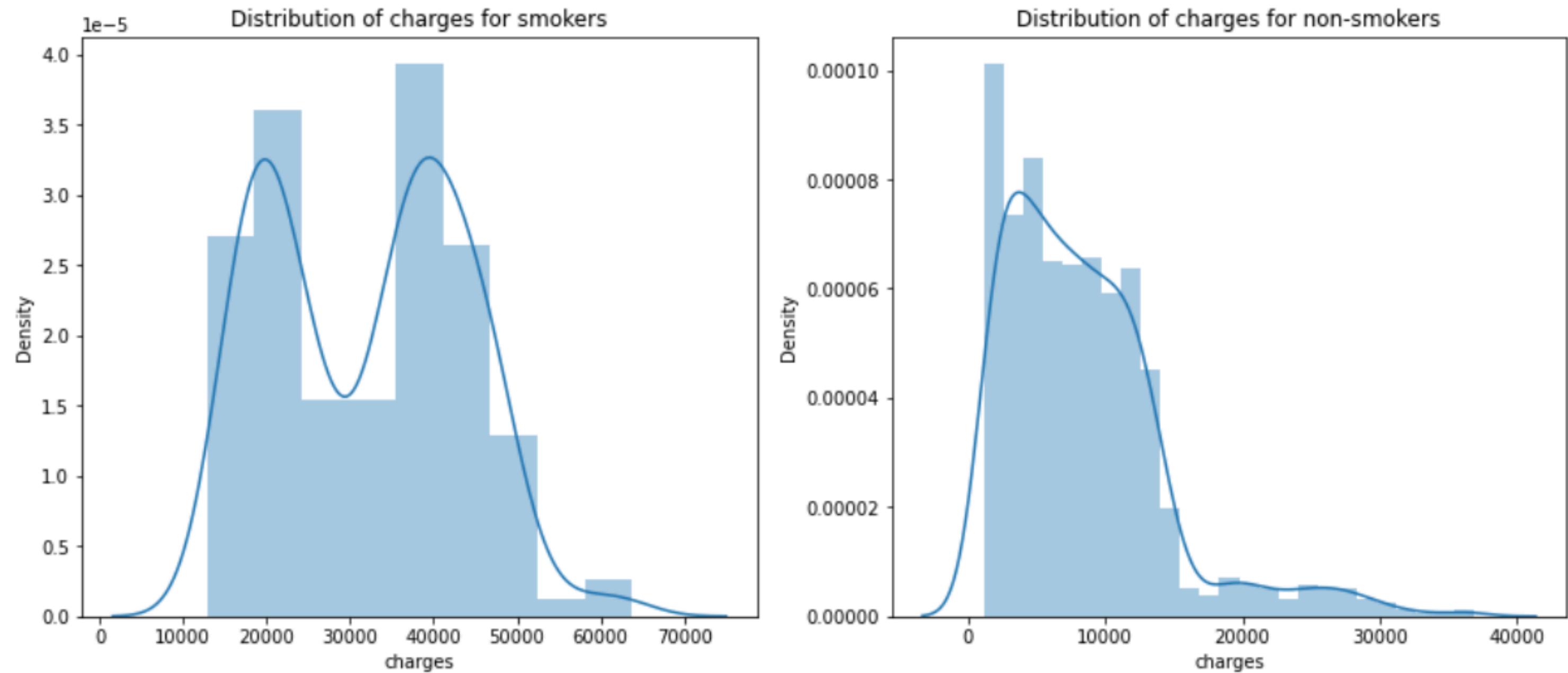
DISTRIBUTION : CHARGES





PRÉTRAITEMENT DES DONNÉES

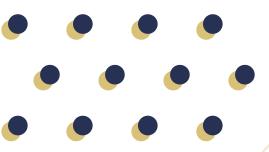
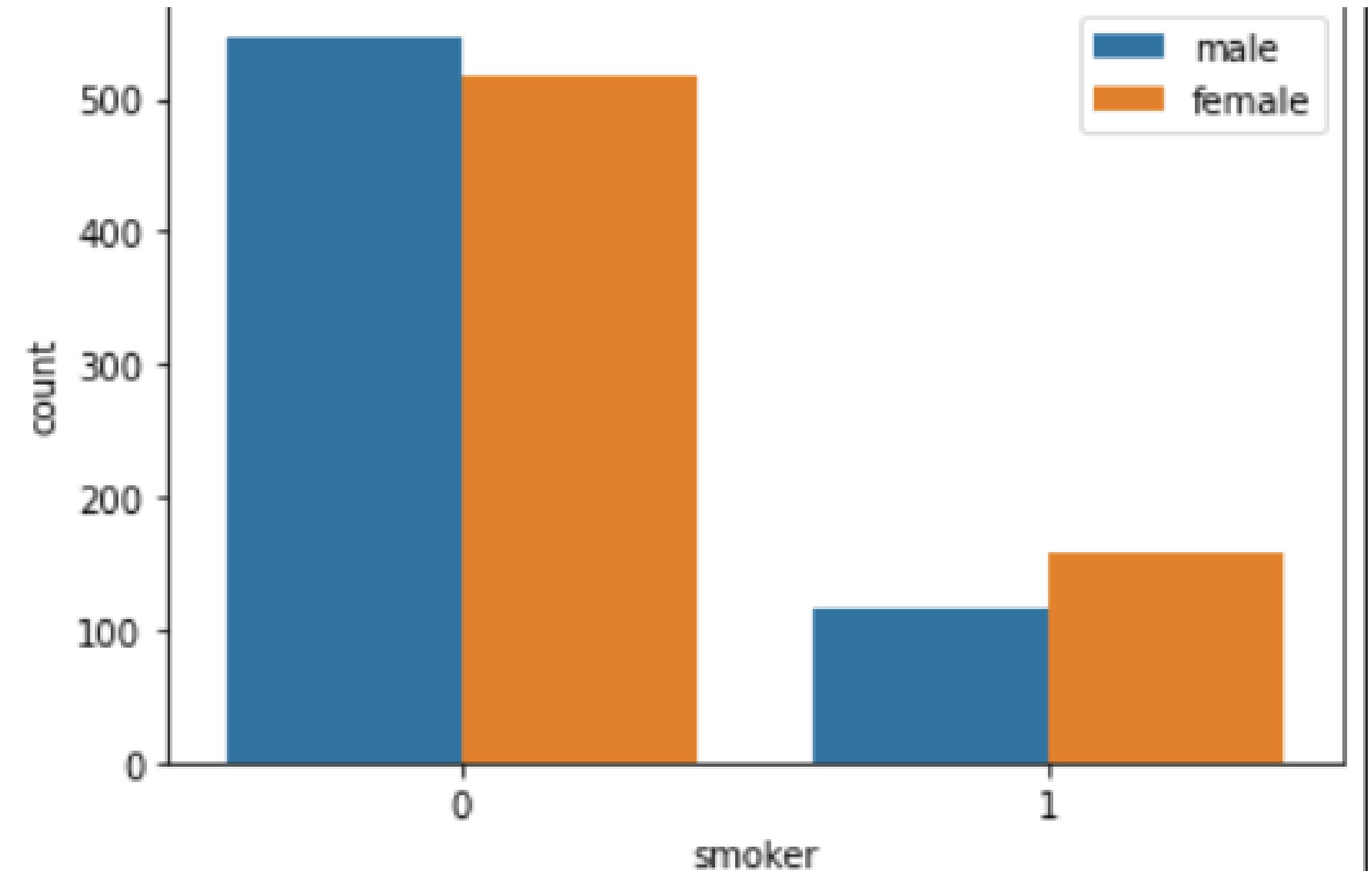
COMPARAISON : FUMEUR - NON FUMEUR





PRÉTRAITEMENT DES DONNÉES

RÉPARTITION FUMEURS (NON-FUMEUR) SELON LE SEXE





CHOIX DU MODÈLE DE MACHINE LEARNING

ENTRAINEMENT DU MODÈLE

Linear Regression R² score: 0.7833

Linear Regression R² score: 0.7833

Ridge Regression R² score: 0.7831

Ridge Regression R² score: 0.7831

LASSO Regression R² score: 0.7833

LASSO Regression R² score: 0.7833

Elastic Net R² score: 0.4193

Elastic Net R² score: 0.4193

Random Forest R² score: 0.8643

Random Forest R² score: 0.8643





CHOIX DU MODÈLE DE MACHINE LEARNING

RANDOM FOREST

Le Random Forest est un algorithme qui crée plusieurs arbres de décision sur des sous-ensembles aléatoires des données. Lors de la prédiction, il agrège les résultats des arbres pour obtenir une estimation plus précise, réduisant ainsi le risque de surapprentissage. Ce modèle est robuste, efficace et convient bien pour les problèmes complexes avec de nombreuses variables.





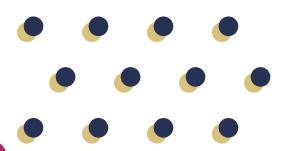
CHOIX DU MODÈLE DE MACHINE LEARNING

RANDOM FOREST

```
# Fit the scaler on the training data and transform both the training and testing
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

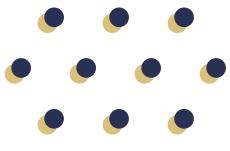
# Find the model with the highest R^2 score
best_model = max(scores, key=scores.get)
print(f"\nBest model: {best_model} with R^2 score: {scores[best_model]:.4f}")
```

Best model: Random Forest with R^2 score: 0.8643





CONSTRUCTION DU MODÈLE



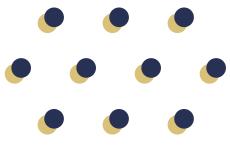
LANGUAGES UTILISÉS

- HTML (HyperText Markup Language) : Le langage de balisage principal, utilisé pour structurer le contenu de l'interface. Il nous a permis de concevoir la disposition des éléments tels que les champs de saisie, les boutons, les formulaires et autres composants visibles par l'utilisateur.
- CSS (Cascading Style Sheets) : Utilisé pour styliser et personnaliser l'apparence visuelle de l'interface. Grâce au CSS, nous avons pu améliorer l'esthétique du site en définissant des couleurs, des polices, des marges, des alignements et en rendant l'interface agréable et accessible.





CONSTRUCTION DU MODÈLE



LANGUAGES UTILISÉS

- **JavaScript** : Ce langage de programmation a été employé pour ajouter de l'interactivité à notre interface. Il permet de rendre le site dynamique, notamment en réagissant aux actions de l'utilisateur, en validant les formulaires, et en rendant les interactions fluides et intuitives.
- **Flask (framework Python)** : Pour le back-end, nous avons utilisé Flask, un micro-framework Python, afin de gérer la logique côté serveur. Flask nous a permis de connecter l'interface utilisateur avec le modèle de prédiction des coûts médicaux, en assurant la communication entre le front-end (HTML, CSS, JavaScript) et les fonctionnalités analytiques de notre application.





CONSTRUCTION DU MODÈLE

Medical Cost Prediction.

Home Predict Team

Stabilizing Through Early Risk Detection

Our AI tool helps healthcare professionals predict medical costs, enabling proactive budget planning and patient care."

[Start Prediction](#)



What is Medical Cost Prediction??

Understand how medical cost prediction can help optimize healthcare budgets, improve financial planning, and support patient care decisions.

[Learn More](#)



Tips for Managing Medical Expenses

Get practical tips on planning for medical costs, choosing the right insurance, and reducing out-of-pocket expenses.

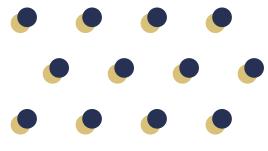
[Learn more](#)

© 2024 Medical Cost Prediction. All rights reserved to IMLOUI Douae and RACHIDI Widad.





CONSTRUCTION DU MODÈLE



About Health insurance and medical cost prediction

Medical Cost Prediction uses data and machine learning to forecast healthcare expenses, including treatments and medications, by analyzing patient demographics, medical history, and lifestyle factors. It helps optimize budgeting and improve care decisions.

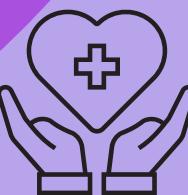
How Medical Cost Prediction Helps?

Optimizes healthcare budgets. Improves financial planning. Helps set accurate premiums. Predicts future healthcare needs. Enhances resource allocation. Supports better care decisions. Reduces unexpected medical costs.

And that is the main goal of our project !

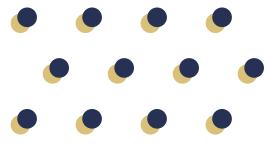


For more information about How to Keep Medical Expenses Under Control, please check [this link](#).





CONSTRUCTION DU MODÈLE



Medical Cost Prediction.

Home Predict Team

Tips to better deal with medical costs

Whether you're dealing with unexpected medical emergencies or navigating routine healthcare needs, taking proactive steps to manage your medical expenses can significantly alleviate financial strain.

Let's face it—medical expenses can be a major source of stress and anxiety for many people. Whether you're dealing with chronic conditions, unexpected illnesses, or routine healthcare needs, the cost of medical care can quickly add up. But don't panic! There are strategies you can use to navigate the world of medical expenses with confidence and ease. By taking a proactive approach to managing your healthcare costs, you can minimize financial stress and focus on what matters most—your health and well-being.

And that is the main goal of our project !





CONSTRUCTION DU MODÈLE



PAGE DE PRÉDICTION

Enter All Variables

Age
30

Sex
Male

BMI (Body Mass Index)
25

Number of Children
0

Smoking Status
Smoker

Region
North-East

Predict



CONSTRUCTION DU MODÈLE

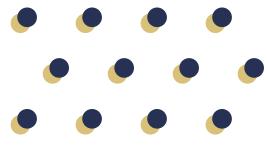
- Prediction: High Charges
- Predicted Value: 13659.5439

Close





CONSTRUCTION DU MODÈLE



Medical Cost Prediction.

Home Predict Team

Project Leaders



IMLOUL

Douae

Data Scientist



RACHIDI

Widad

Data Scientist

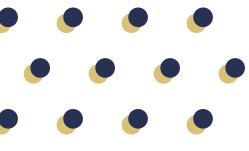


© 2024 Medical Cost Prediction. All rights reserved.





CONSTRUCTION DU MODÈLE



DOCKERISATION

La dockerisation permet d'encapsuler notre application et ses dépendances dans un conteneur, assurant ainsi portabilité et uniformité de l'environnement. Grâce à Docker, le déploiement devient plus simple et rapide, garantissant que l'application fonctionne de manière cohérente, de l'environnement de développement à la production.



CONSTRUCTION DU MODÈLE

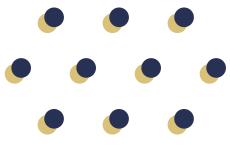
The screenshot shows the Docker Desktop interface with the following details:

- Containers** tab selected in the sidebar.
- Container CPU usage:** 0.30% / 400% (4 CPUs available).
- Container memory usage:** 244.3MB / 7.53GB.
- Show charts** button.
- Search bar:** Search for images, containers, volume... Ctrl+K.
- Filter:** Only show running containers.
- Table:** Displays one container entry.
- Container Details:**
 - Name:** elated_ran
 - Image:** medical-cost:3099c311ea <none>
 - Status:** Running
 - Port(s):** 8080:8080 ↗
 - CPU (%):** 0.3%
 - Last started:** 2 hours ago
- Actions:** A vertical menu with icons for Stop, Start, and Delete.
- Message:** Showing 1 item.
- Walkthroughs:** Multi-container applications and Containerize your application.
- Bottom status bar:** Engine running, RAM 1.58 GB, CPU 12.99%, Disk -- GB avail. of -- GB, docker init, Terminal, New version available.



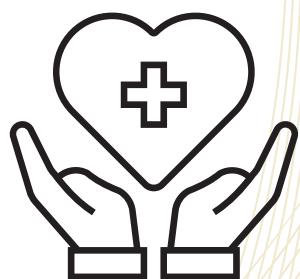


CONSTRUCTION DU MODÈLE



GOOGLE CLOUD

Le déploiement de notre application sur Google Cloud offre une infrastructure flexible et performante pour héberger et gérer les services. En utilisant Google Cloud Platform (GCP), nous bénéficions de l'évolutivité, de la sécurité et de la haute disponibilité pour notre application.





**THANK
YOU**

