

Détection du cancer du sein à partir d'images mammographiques



*Réalisée par : Oumaima ERRIHANI
3-ème ACI-ICSD
ESI*



01

Introduction

02

*Exploration
des données*

03

*Nettoyage
des données*

04

*Analyse exploratoire
des Données
(EDA)*

05

*Exploration
des images*

06

*Questions
Techniques +
Domaine métier*

Introduction

La détection automatique de la maladie du cancer du sein est une solution primordiale en médecine, car elle aide les spécialistes à identifier automatiquement les anomalies dans les seins, simplifiant ainsi la détection, facilitant l'identification et la localisation du cancer, et permettant une décision rapide pour initier le traitement.

*Dataset d'anomalie de
Calcification*

01

*Dataset d'anomalie de
Masse*

02

Dataset de métadonnées

03

Dataset Dicom

04

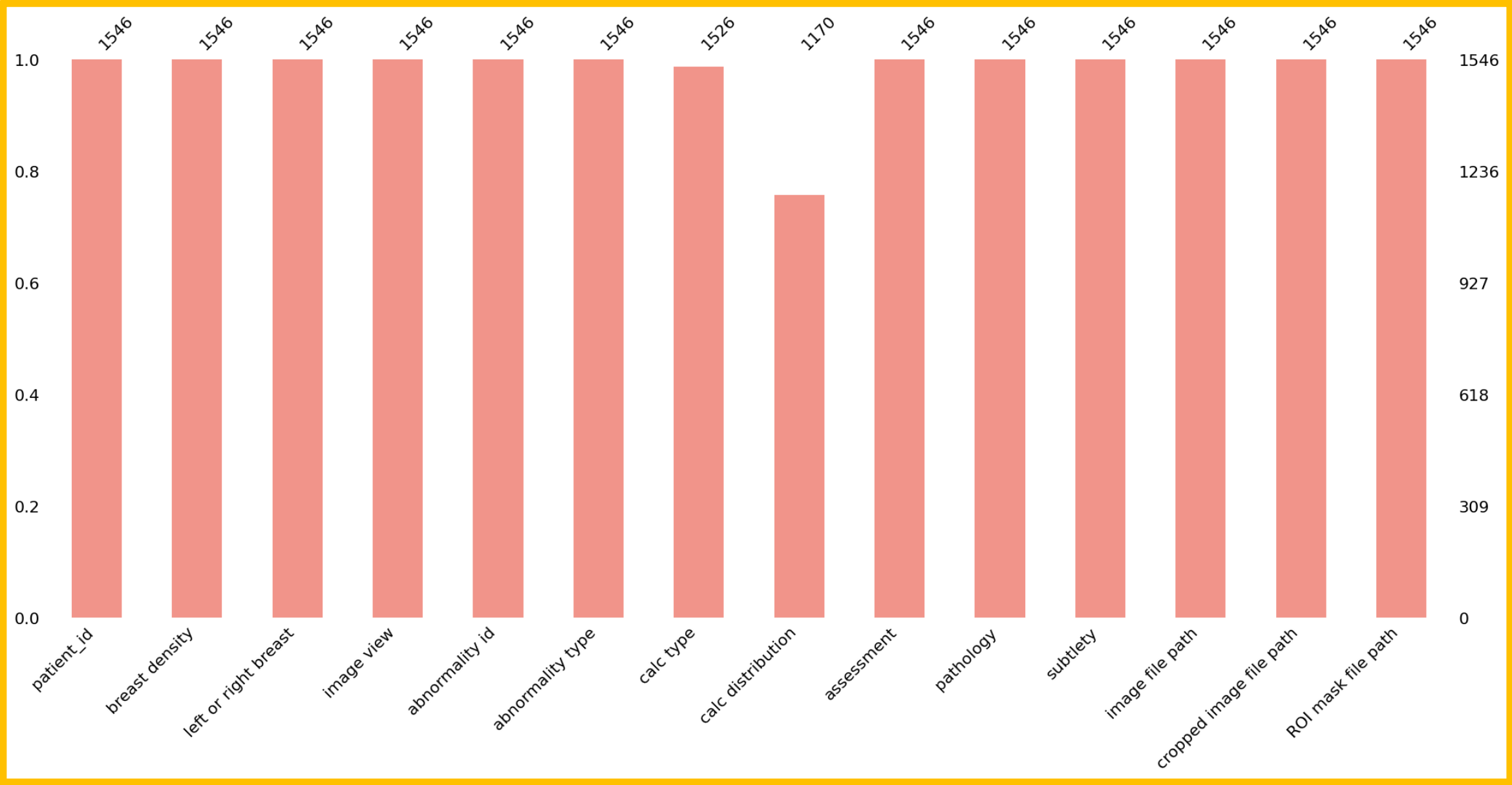
Dataset 01 : Description des Cas d'Anomalie de Calcification

	patient_id	breast density	left or right breast	image view	abnormality id	abnormality type	calc type	calc distribution	assessment	pathology	subtlety	
0	P_00005	3	RIGHT	CC	1	calcification	AMORPHOUS	CLUSTERED	3	MALIGNANT	3	Training_P_00005_RIGHT_CC/1.3.6.1.4.1.95
1	P_00005	3	RIGHT	MLO	1	calcification	AMORPHOUS	CLUSTERED	3	MALIGNANT	3	Training_P_00005_RIGHT_MLO/1.3.6.1.4.1.95
2	P_00007	4	LEFT	CC	1	calcification	PLEOMORPHIC	LINEAR	4	BENIGN	4	Training_P_00007_LEFT_CC/1.3.6.1.4.1.95
3	P_00007	4	LEFT	MLO	1	calcification	PLEOMORPHIC	LINEAR	4	BENIGN	4	Training_P_00007_LEFT_MLO/1.3.6.1.4.1.95
4	P_00008	1	LEFT	CC	1	calcification	nan	REGIONAL	2	BENIGN_WITHOUT_CALLBACK	3	Training_P_00008_LEFT_CC/1.3.6.1.4.1.95

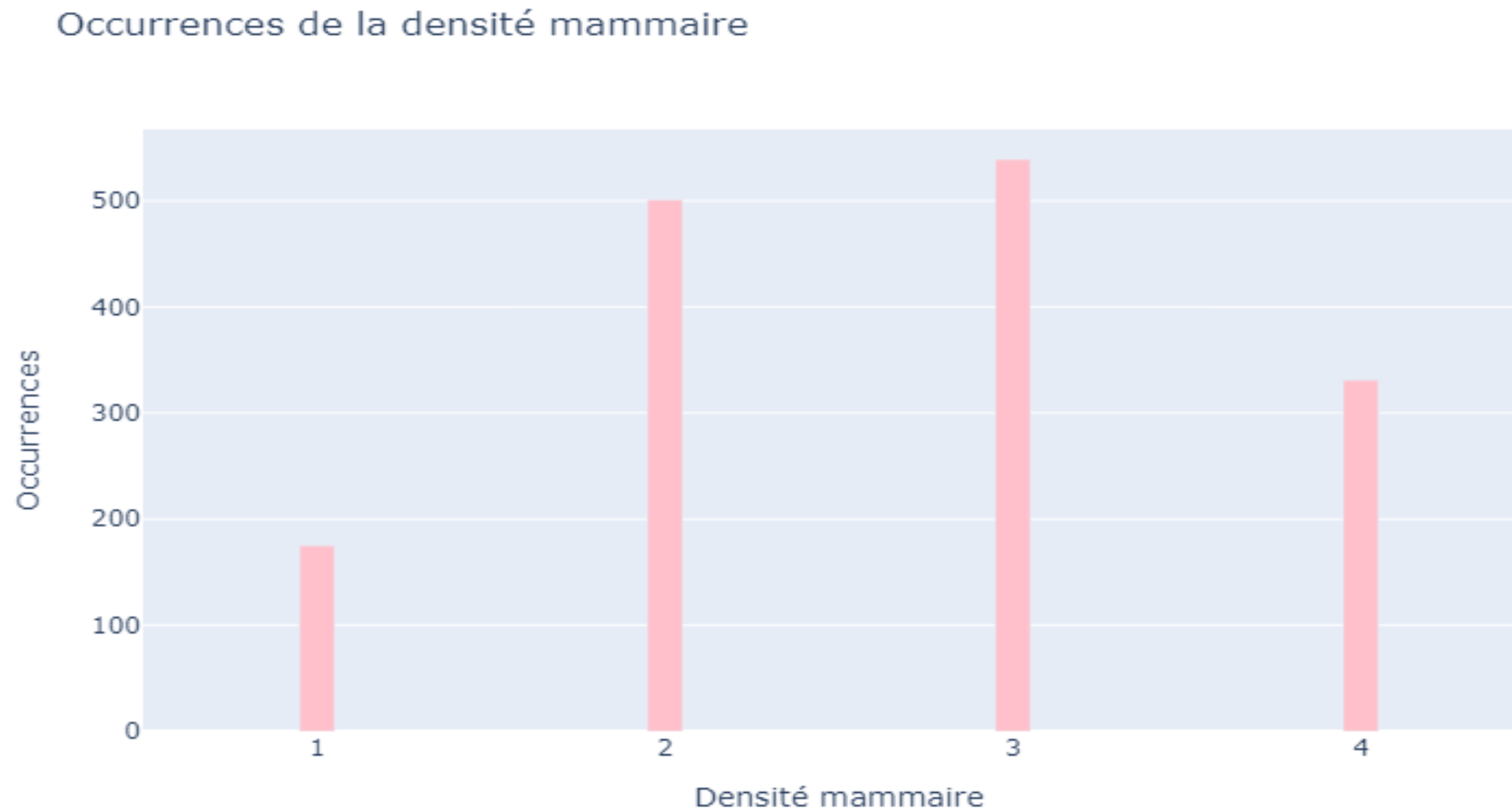
Dataset 01 : Description des Cas d'Anomalie de Calcification

***01/ Exploration des
Données***

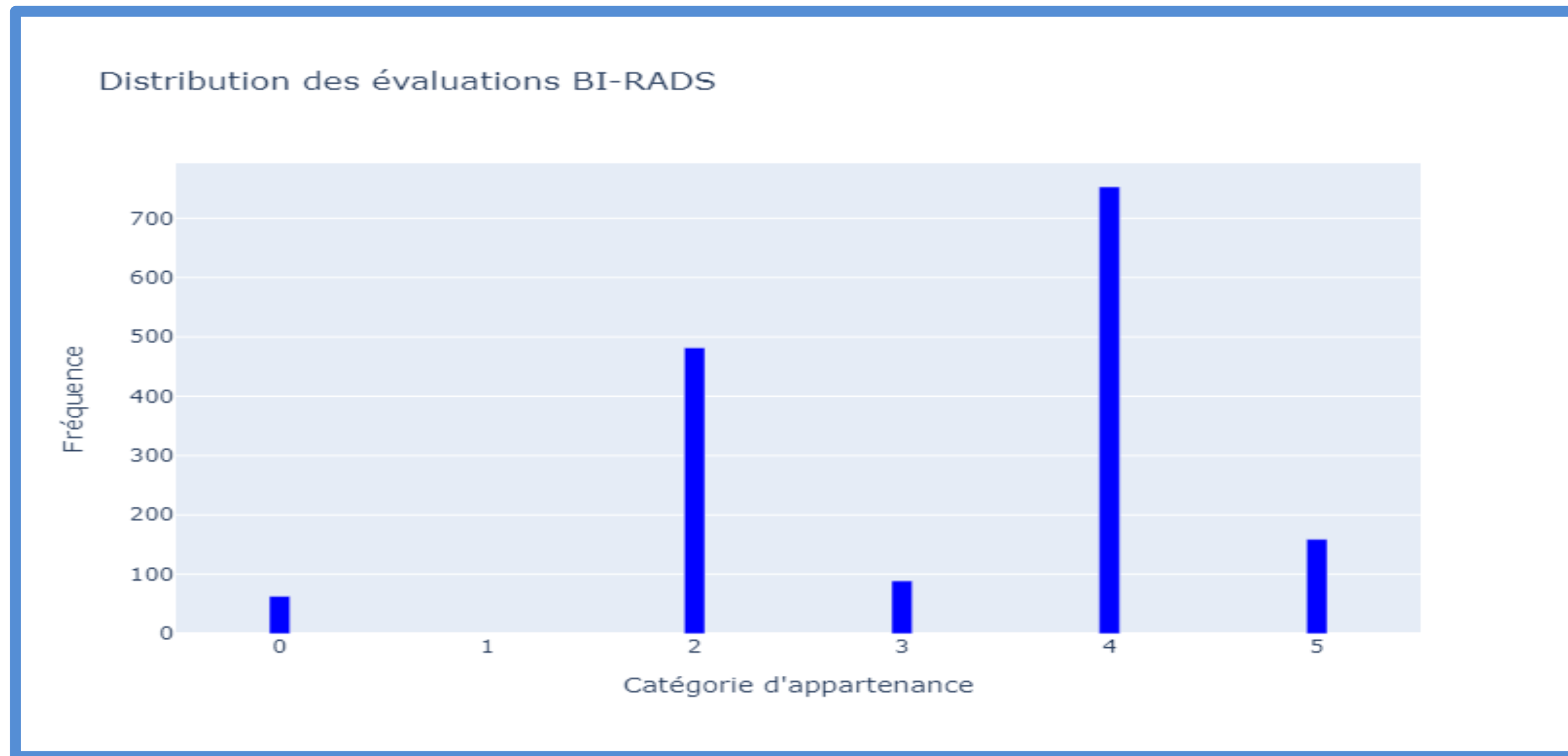
Informations de Dataset



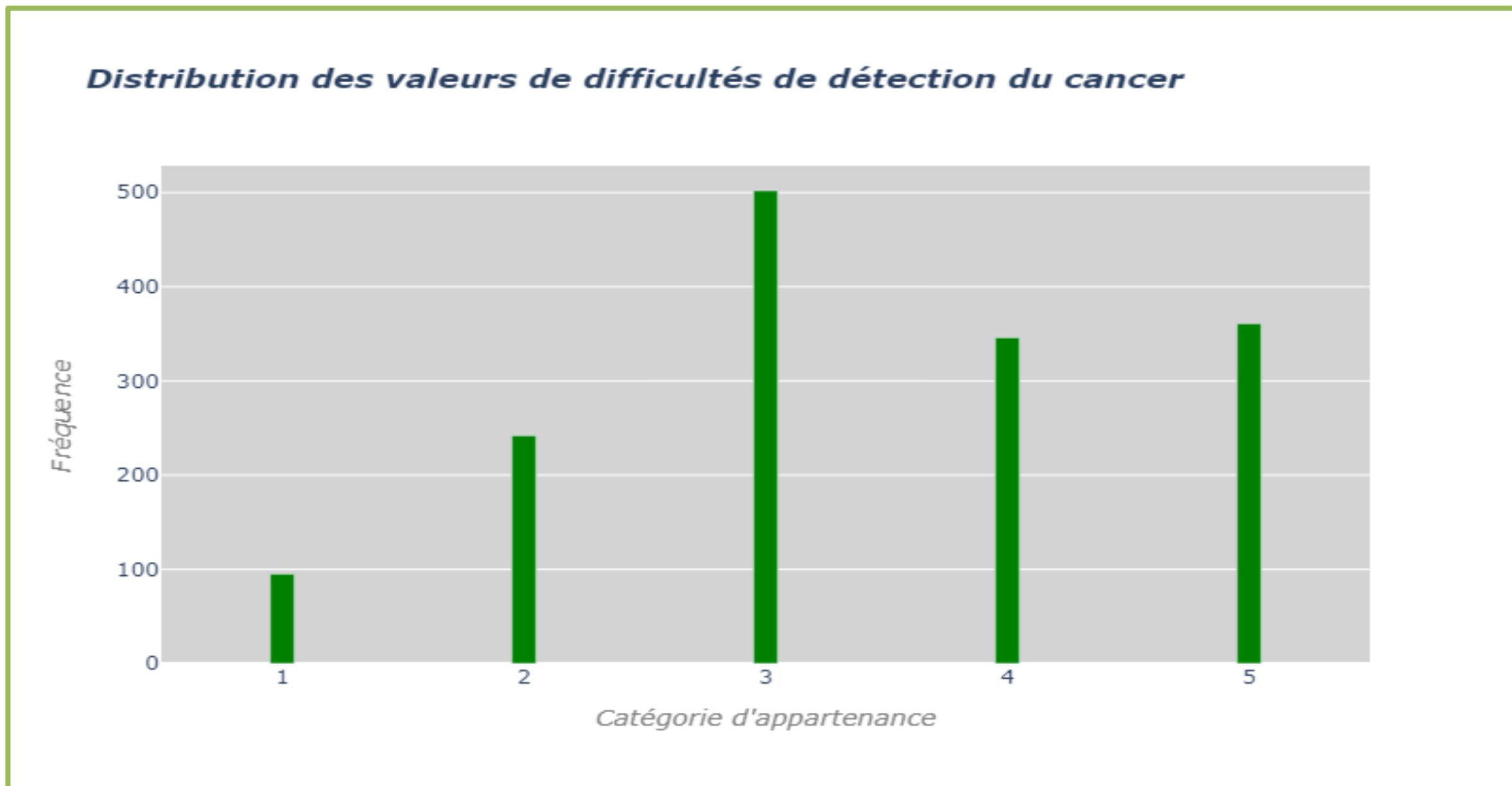
Distribution des Occurrences de la densité mammaire



Distribution des Occurrences des évaluations BI-RADS

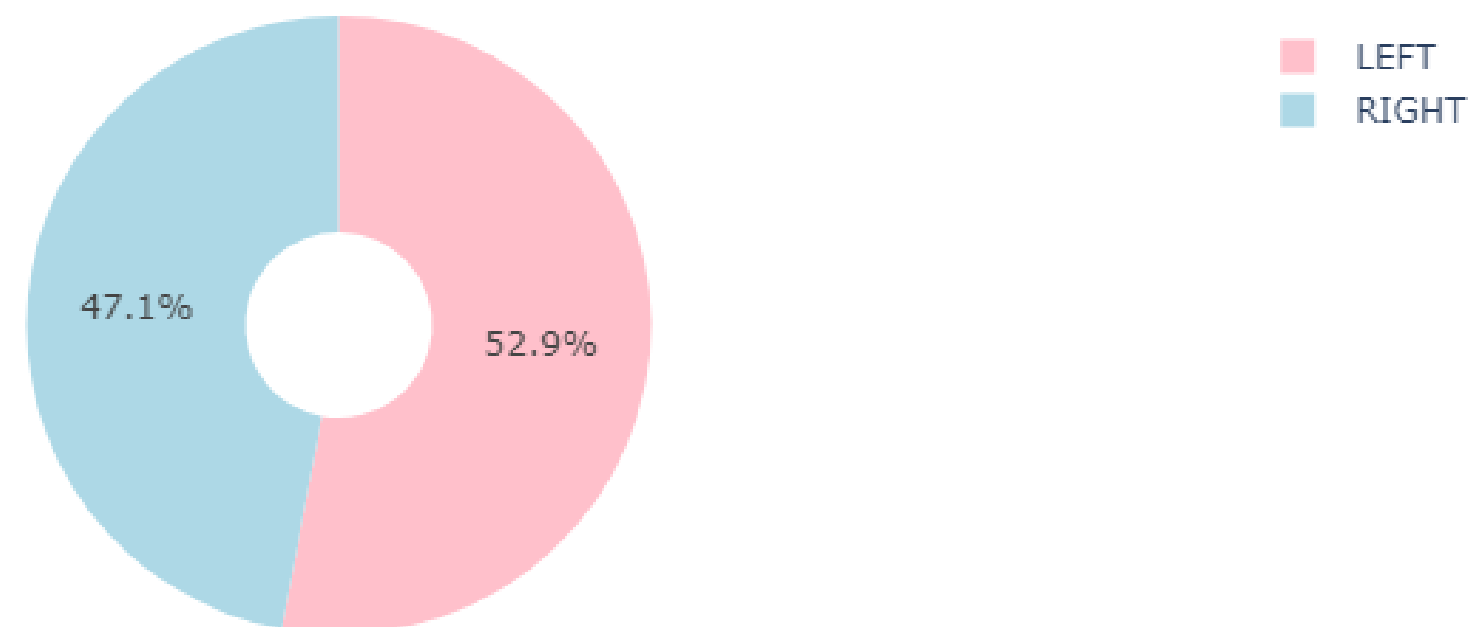


Distribution des Valeurs de difficultés du Cancer du sein



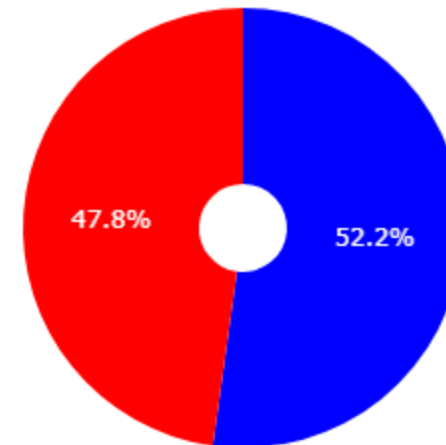
Répartition des données par côté du sein examiné

Répartition des données par côté du sein



Répartition des données en fonction des incidences des images mammographiques

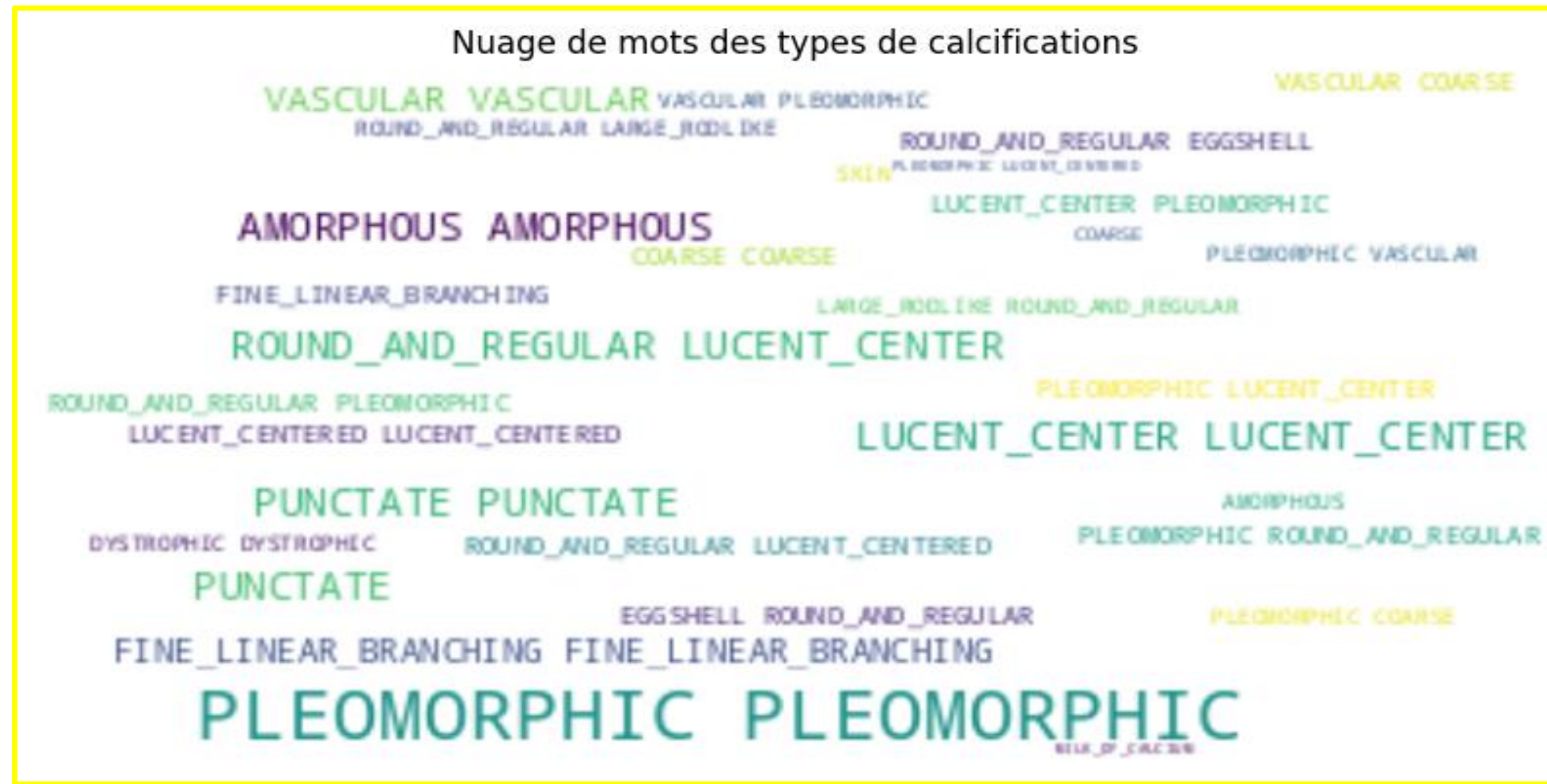
Répartition des données en fonction des incidences des images mammographiques



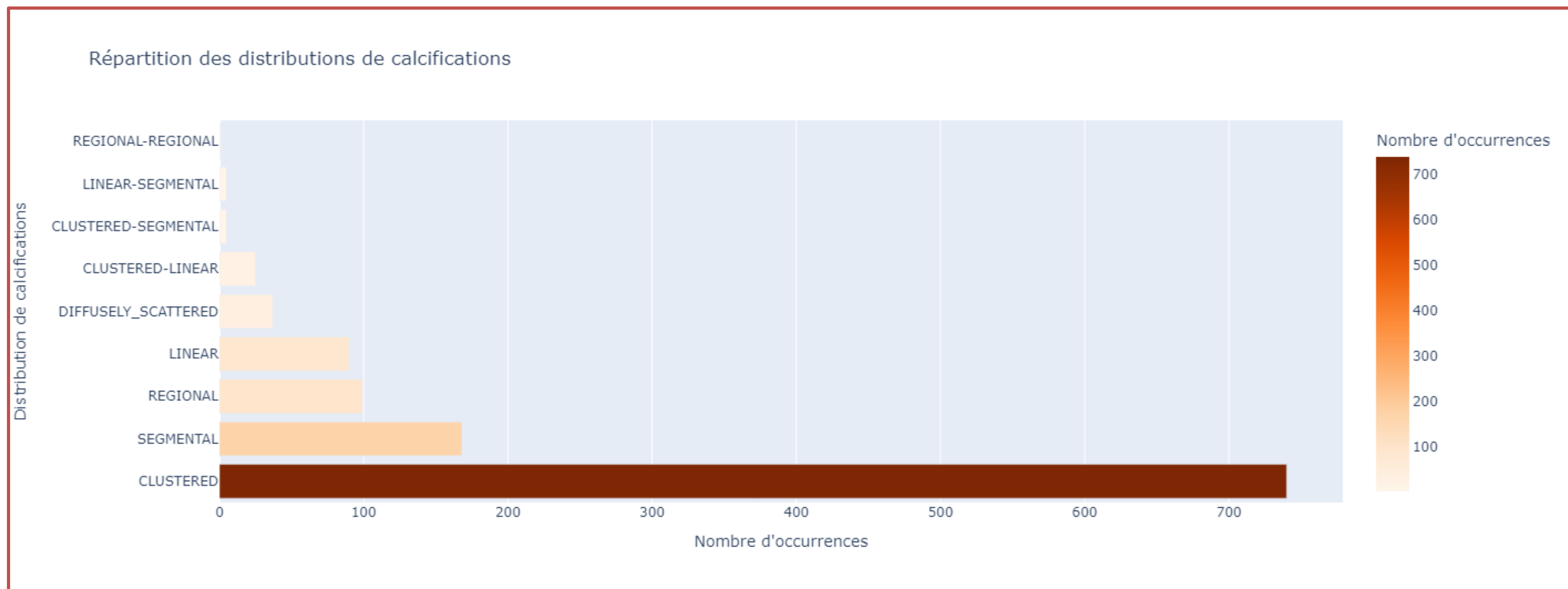
■ MLO
■ CC



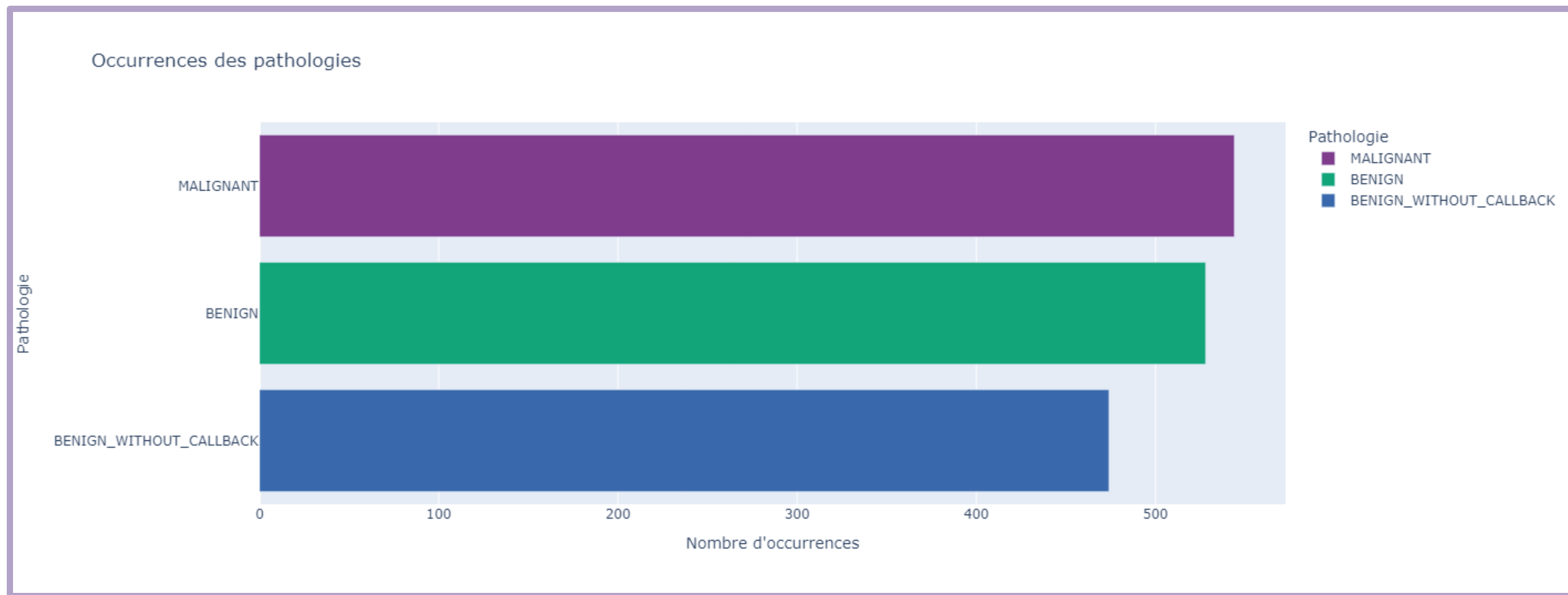
Nuage de mots de différents types de calcification



Répartition des distributions d'anomalie de calcification



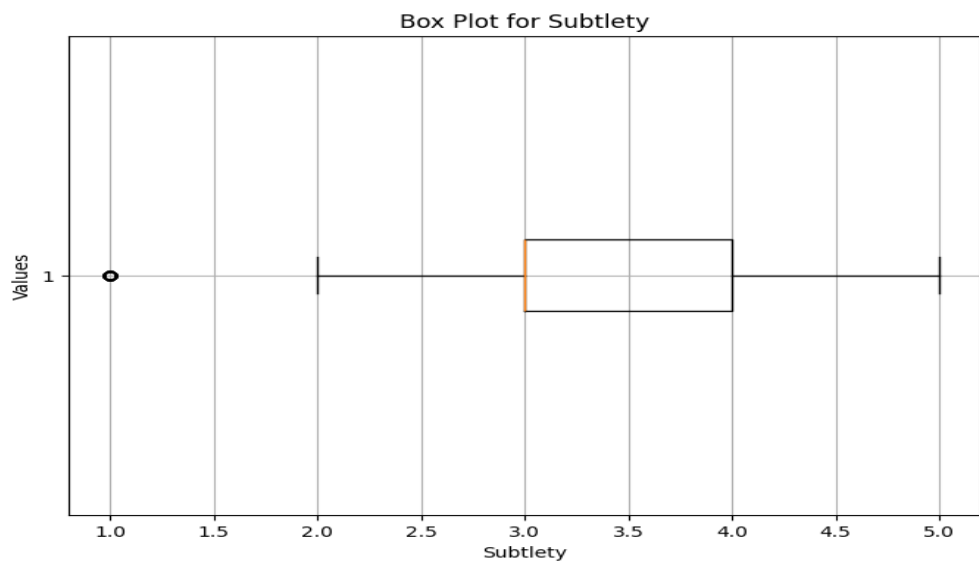
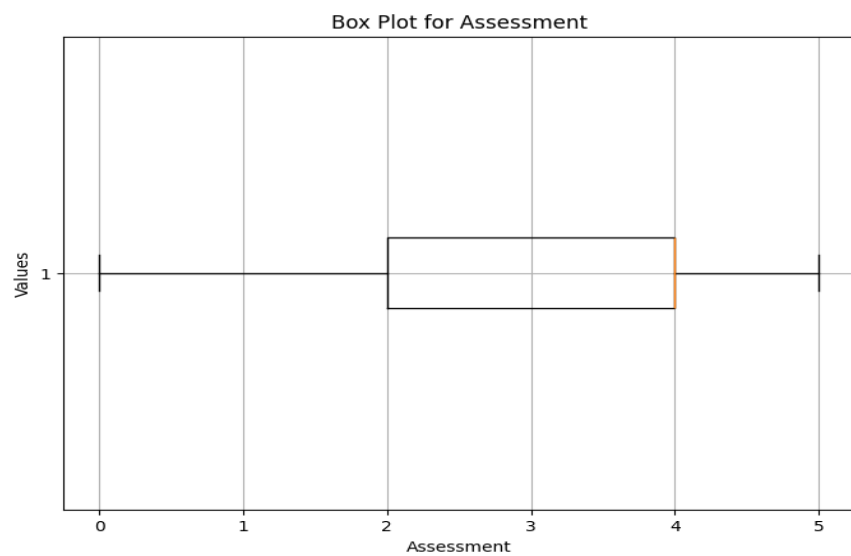
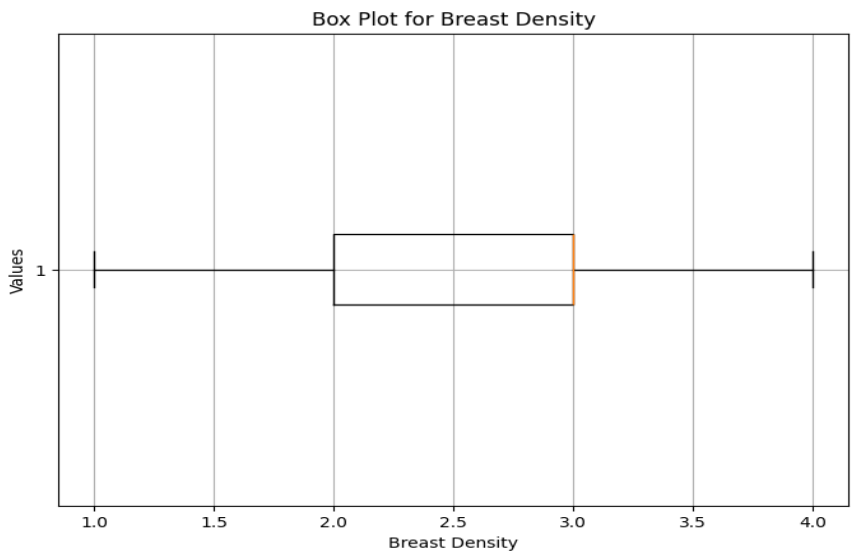
Répartition des Occurrences de Pathologies



Dataset 01 : Description des Cas d'Anomalie de Calcification

02/ Nettoyage des Données

Data cleaning



Valeurs abérantes

6-1 Traitement des valeurs manquantes dans les deux colonnes : *calc_type* et *calc_distribution*

```
#Remplacer les valeurs manquantes dans 'calc_type' par le mode de cette colonne :
calc_type_mode = calc_case_description_train_set_df['calc_type'].mode()[0]
# Et remplacer les valeurs manquantes dans 'calc_distribution' par le mode de cette colonne :
calc_distribution_mode = calc_case_description_train_set_df['calc_distribution'].mode()[0]

# Imputation des valeurs manquantes :
calc_case_description_train_set_df['calc_type'].fillna(calc_type_mode, inplace=True)
calc_case_description_train_set_df['calc_distribution'].fillna(calc_distribution_mode, inplace=True)

# Vérification que les valeurs manquantes ont été remplies
print("\n Données Après imputation :")
calc_case_description_train_set_df.isnull().sum()
```

Données Après imputation :

patient_id	0
breast density	0
left or right breast	0
image view	0
abnormality id	0
abnormality type	0
calc type	0
calc distribution	0
assessment	0
pathology	0
subtlety	0
image file path	0
cropped image file path	0
ROI mask file path	0
dtype: int64	

6-2 Renomer les colonnes de dataset

```
[ ] # Renommer les colonnes en utilisant des tirets bas pour faciliter leurs utilisations :
calc_case_description_train_set_df = calc_case_description_train_set_df.rename(columns={
    'breast density': 'breast_density',
    'left or right breast': 'left_or_right_breast',
    'image view': 'image_view',
    'abnormality id': 'abnormality_id',
    'abnormality type': 'abnormality_type',
    'calc type': 'calc_type',
    'calc distribution': 'calc_distribution',
    'image file path': 'image_file_path',
    'cropped image file path': 'cropped_image_file_path',
    'ROI mask file path': 'ROI_mask_file_path'
})

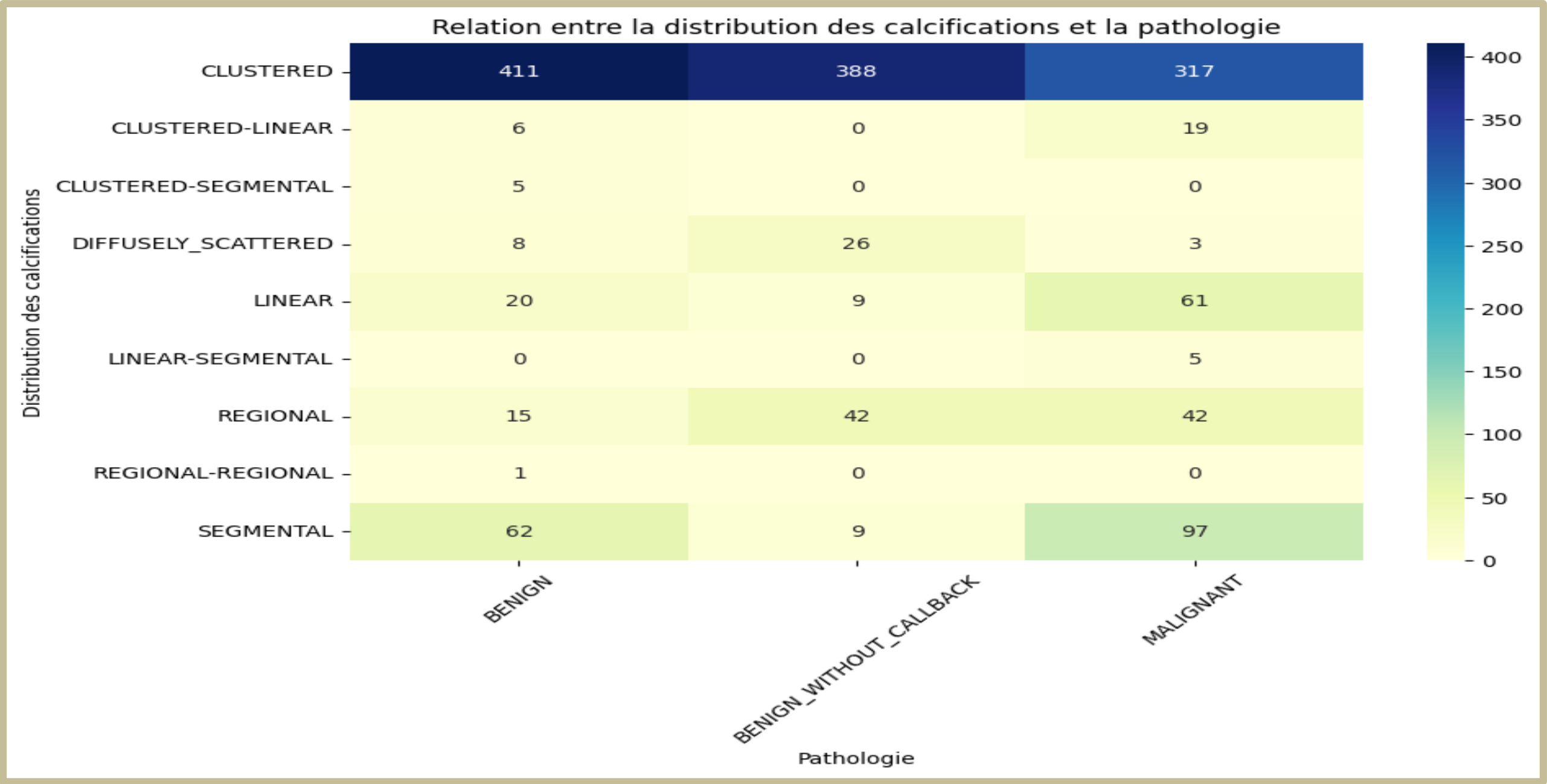
# Afficher les nouveaux noms de colonnes
print(calc_case_description_train_set_df.columns)

Index(['patient_id', 'breast_density', 'left_or_right_breast', 'image_view',
      'abnormality_id', 'abnormality_type', 'calc_type', 'calc_distribution',
      'assessment', 'pathology', 'subtlety', 'image_file_path',
      'cropped_image_file_path', 'ROI_mask_file_path'],
      dtype='object')
```

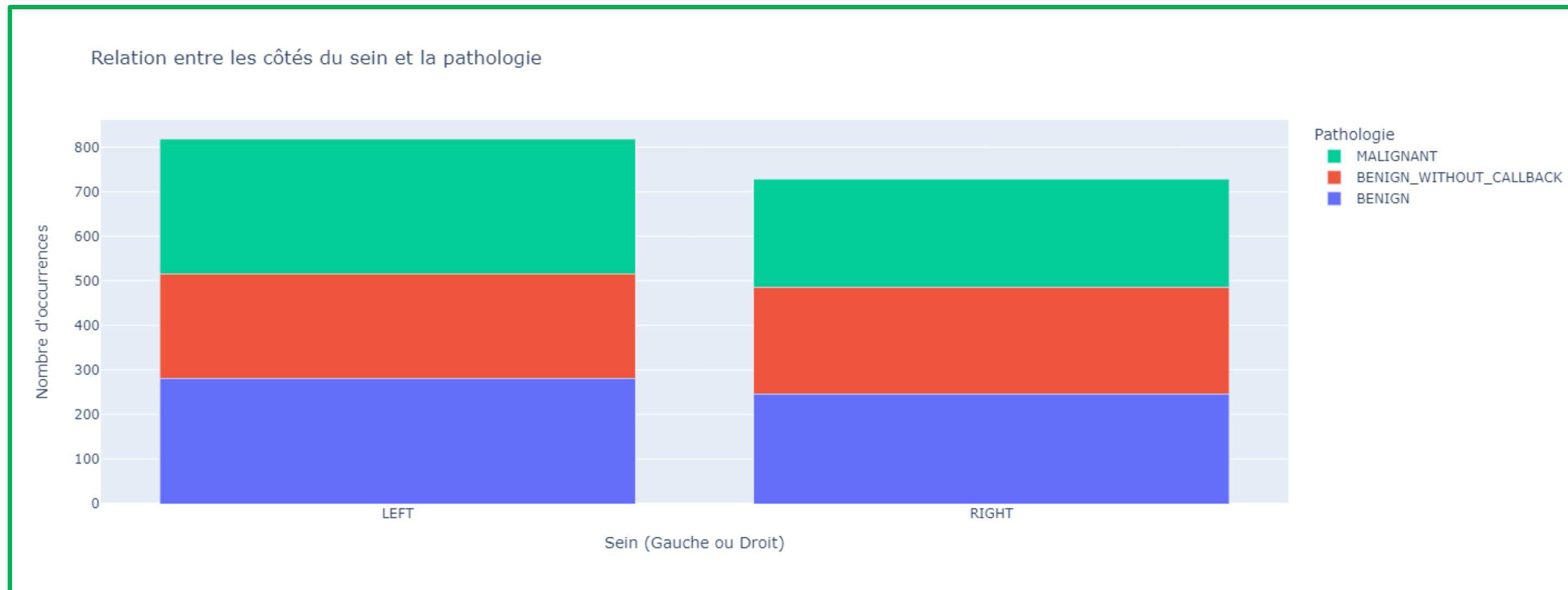
Dataset 01 : Description des Cas d'Anomalie de Calcification

03/ Analyse Exploratoire des Données

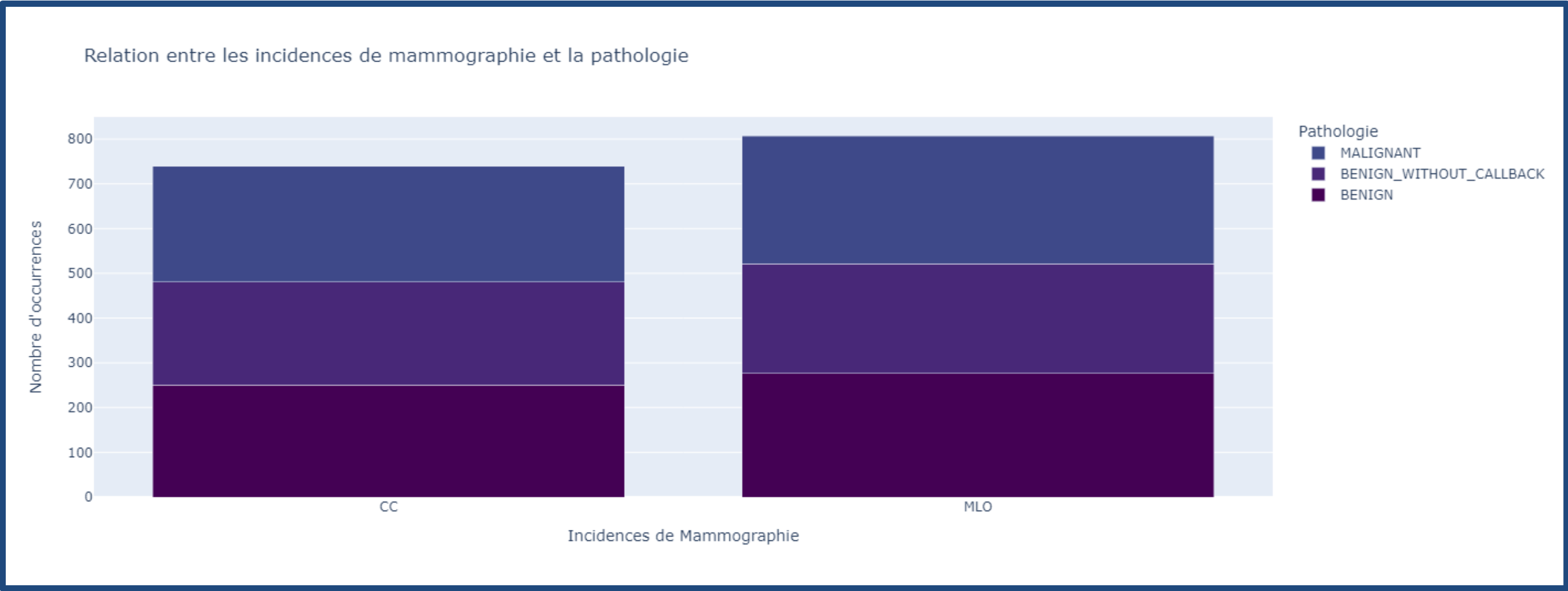
Relation Combinant la pathologie et les distributions de calcification



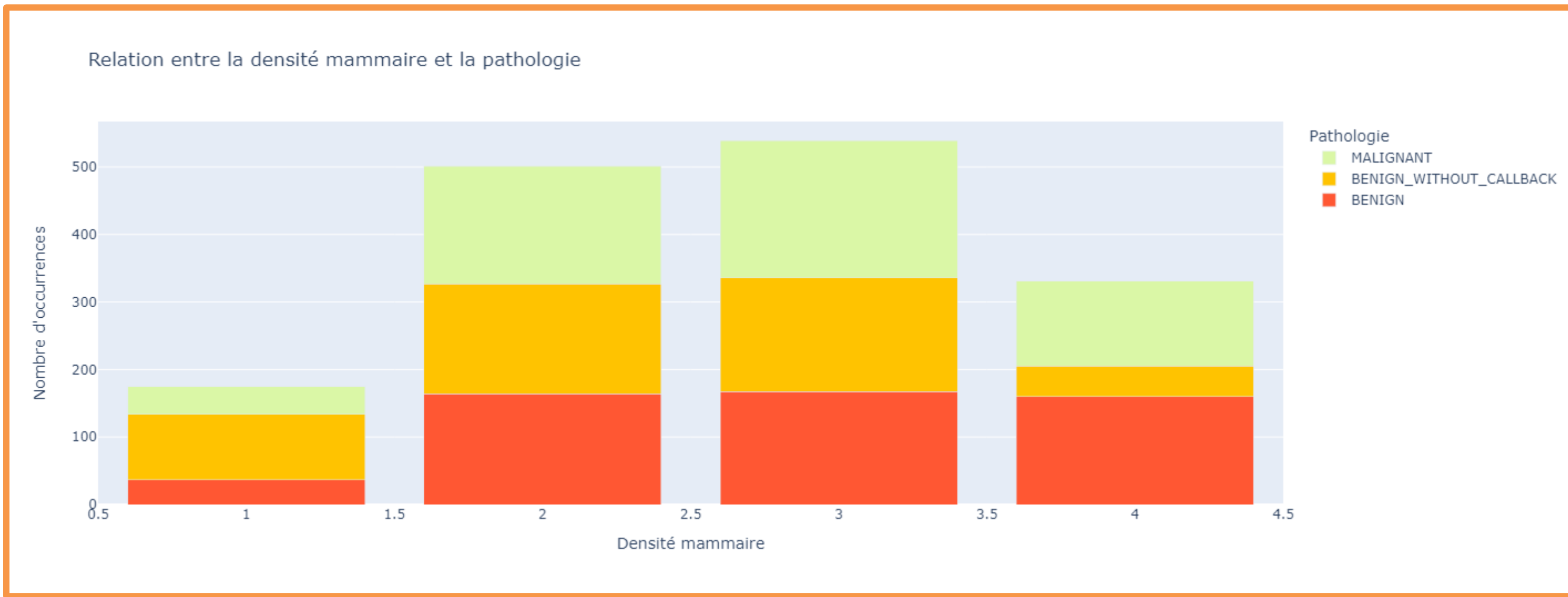
Relation entre les deux variables pathologie et Cotés des seins



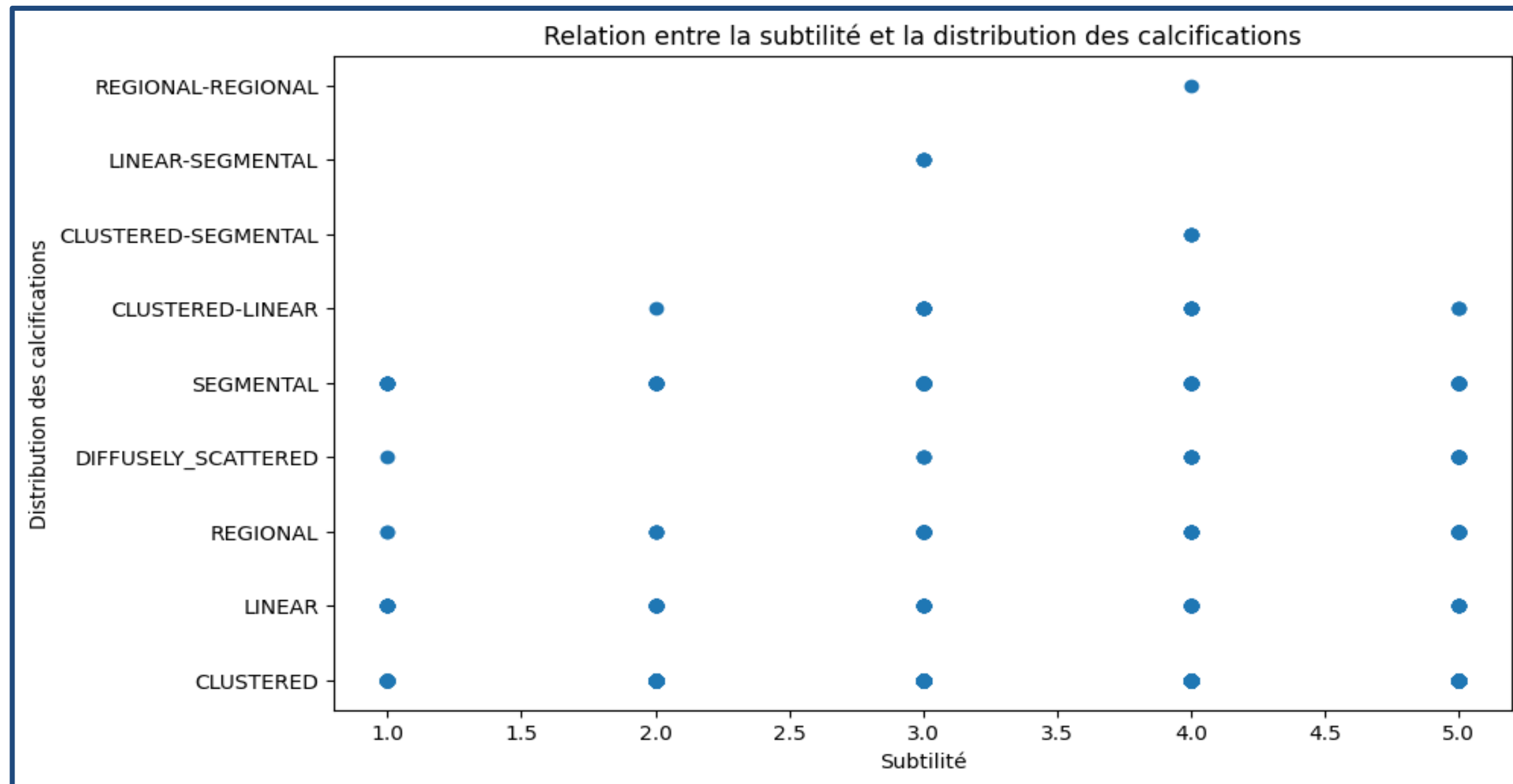
Relation entre les incidences mammographiques et la pathologie



Relation entre la densité mammaire et la pathologie



Relation entre la subtilité et la distribution des calcifications





Dataset 02 : Description des Cas d'Anomalie de Masse

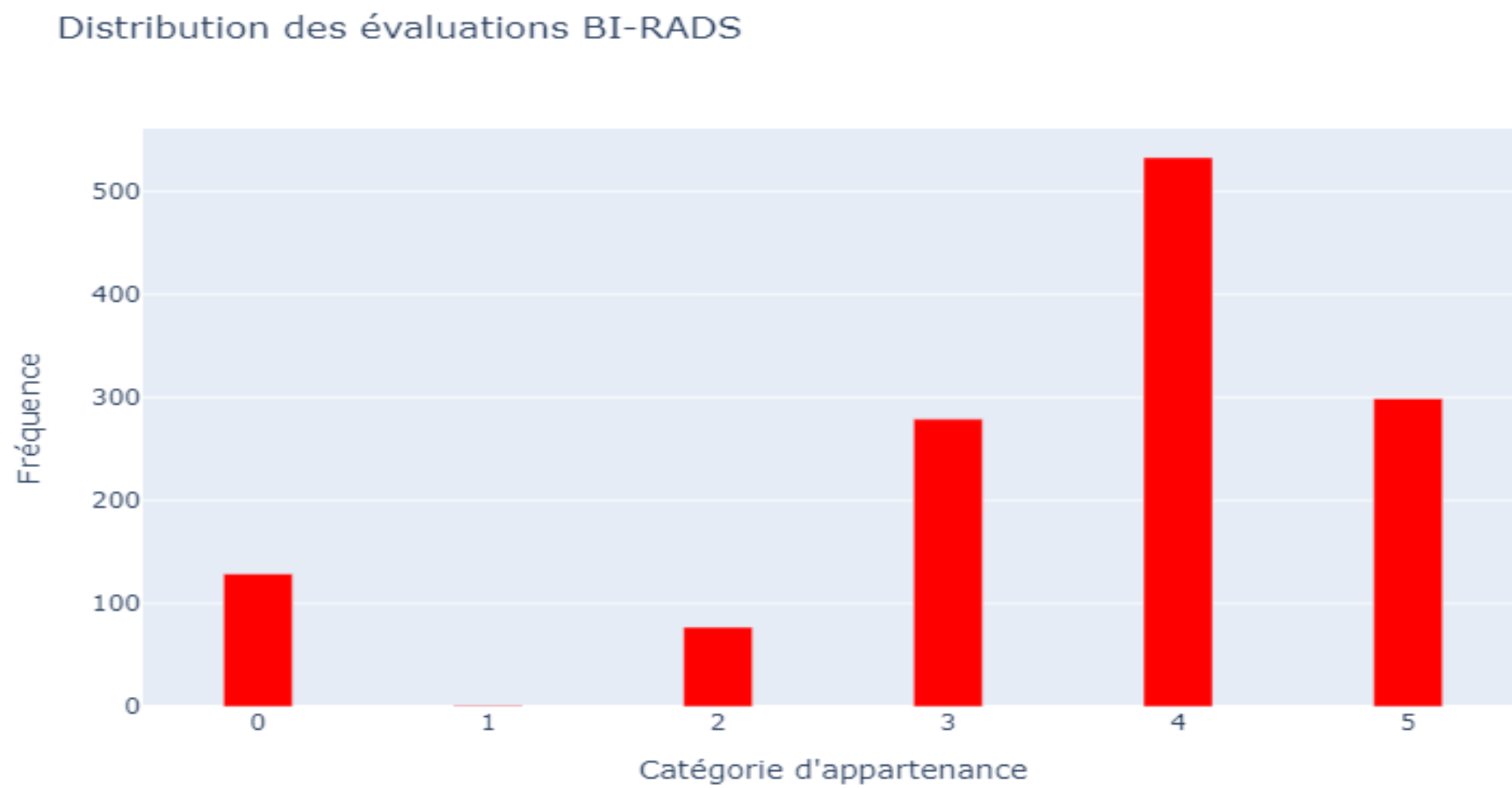


	patient_id	breast_density	left or right breast	image view	abnormality id	abnormality type	mass	shape	mass	assessment	pathology	subtlety	
				view	id	type			margin				
0	P_00001	3	LEFT	CC	1	mass	ARCHITECTURAL_DISTORTION	IRREGULAR- SPICULATED	SPICULATED	4	MALIGNANT	4	Training_P_00001_LEFT_CC/1.3.6
1	P_00001	3	LEFT	MLO	1	mass	ARCHITECTURAL_DISTORTION	IRREGULAR- SPICULATED	SPICULATED	4	MALIGNANT	4	Training_P_00001_LEFT_MLO/1.3.6
2	P_00004	3	LEFT	CC	1	mass	ARCHITECTURAL_DISTORTION	ILL_DEFINED	ILL_DEFINED	4	BENIGN	3	Training_P_00004_LEFT_CC/1.3.6
3	P_00004	3	LEFT	MLO	1	mass	ARCHITECTURAL_DISTORTION	ILL_DEFINED	ILL_DEFINED	4	BENIGN	3	Training_P_00004_LEFT_MLO/1.3.6
4	P_00004	3	RIGHT	MLO	1	mass		OVAL	CIRCUMSCRIBED	4	BENIGN	5	Training_P_00004_RIGHT_MLO/1.3.6

Dataset 02 : Description des Cas d'Anomalie de Masse

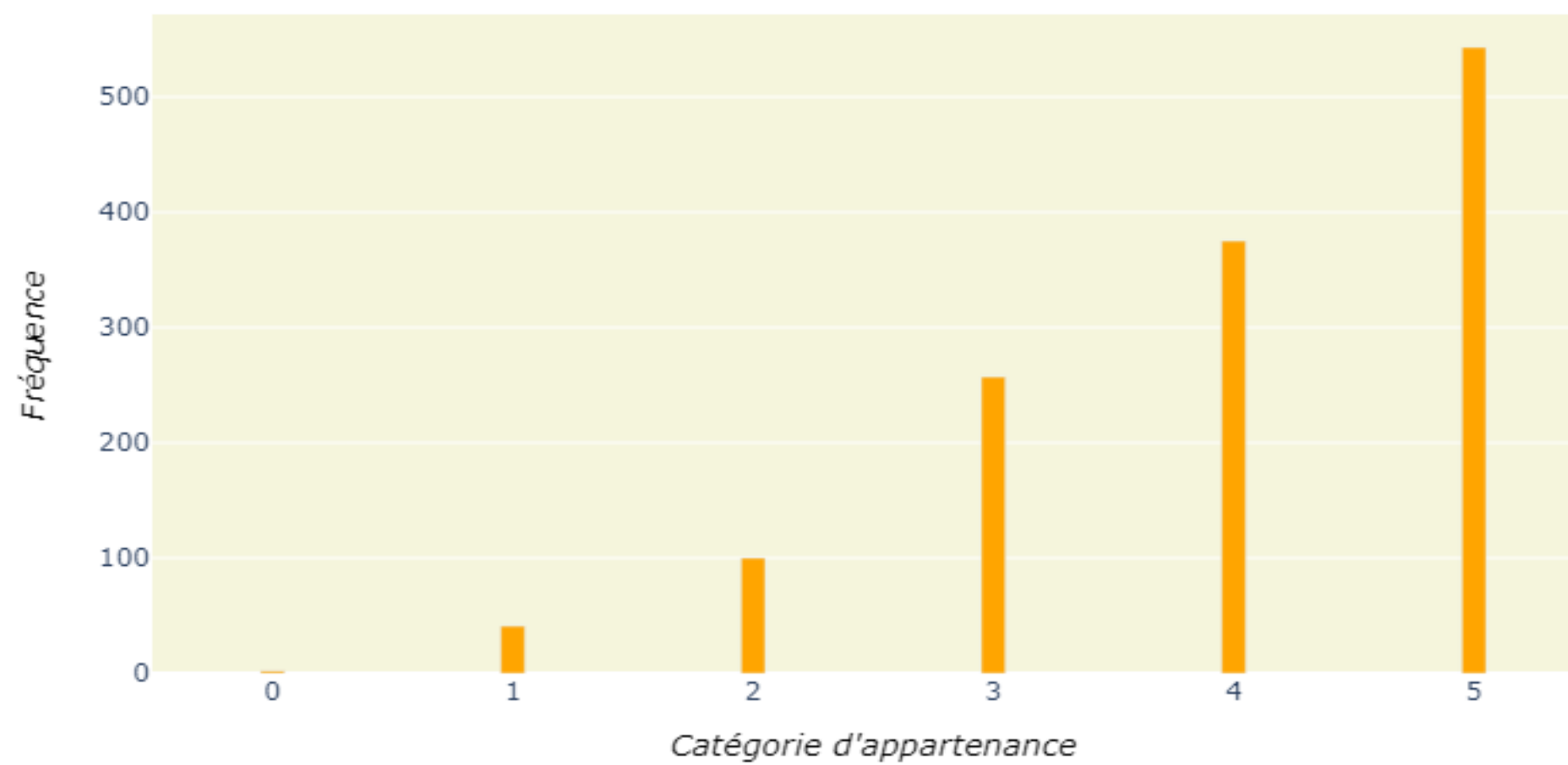
01/ Exploration des données

Distribution des évaluations BI-RADS



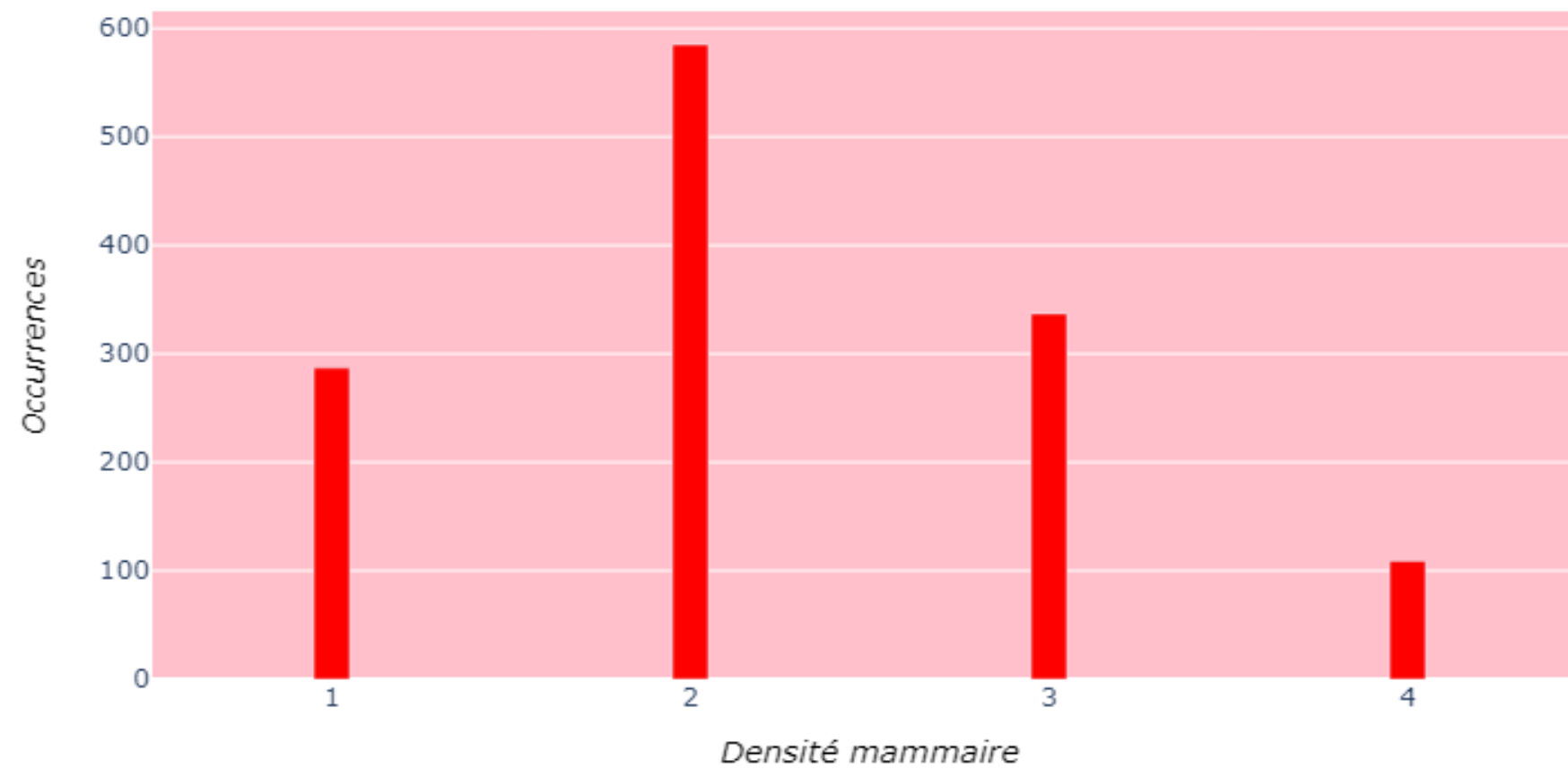
Distribution des Subtilités de la masse détectée

Distribution des Valeurs de difficultés de détection du cancer

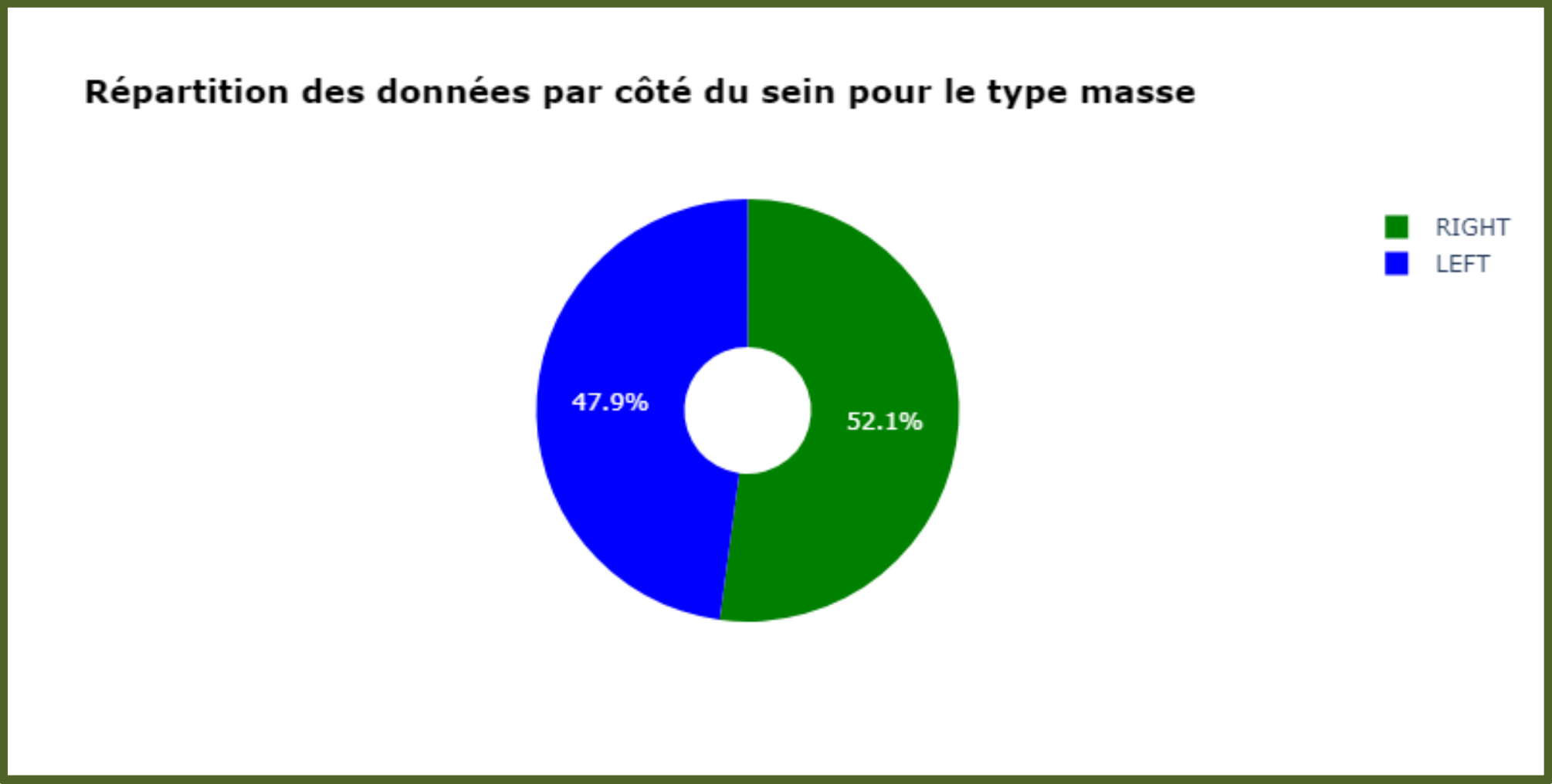


Distribution de la densité mammaire

Occurrences de la densité mammaire

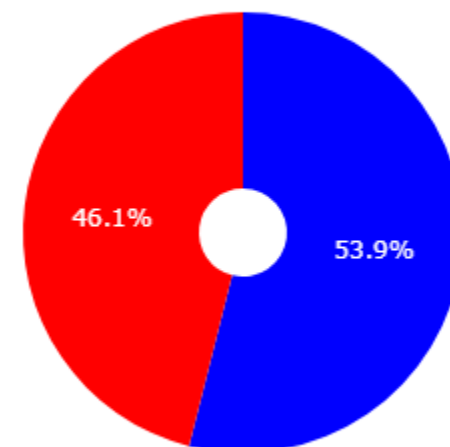


Distribution des données par côté du sein



Distribution des données par les images des incidences mammographiques

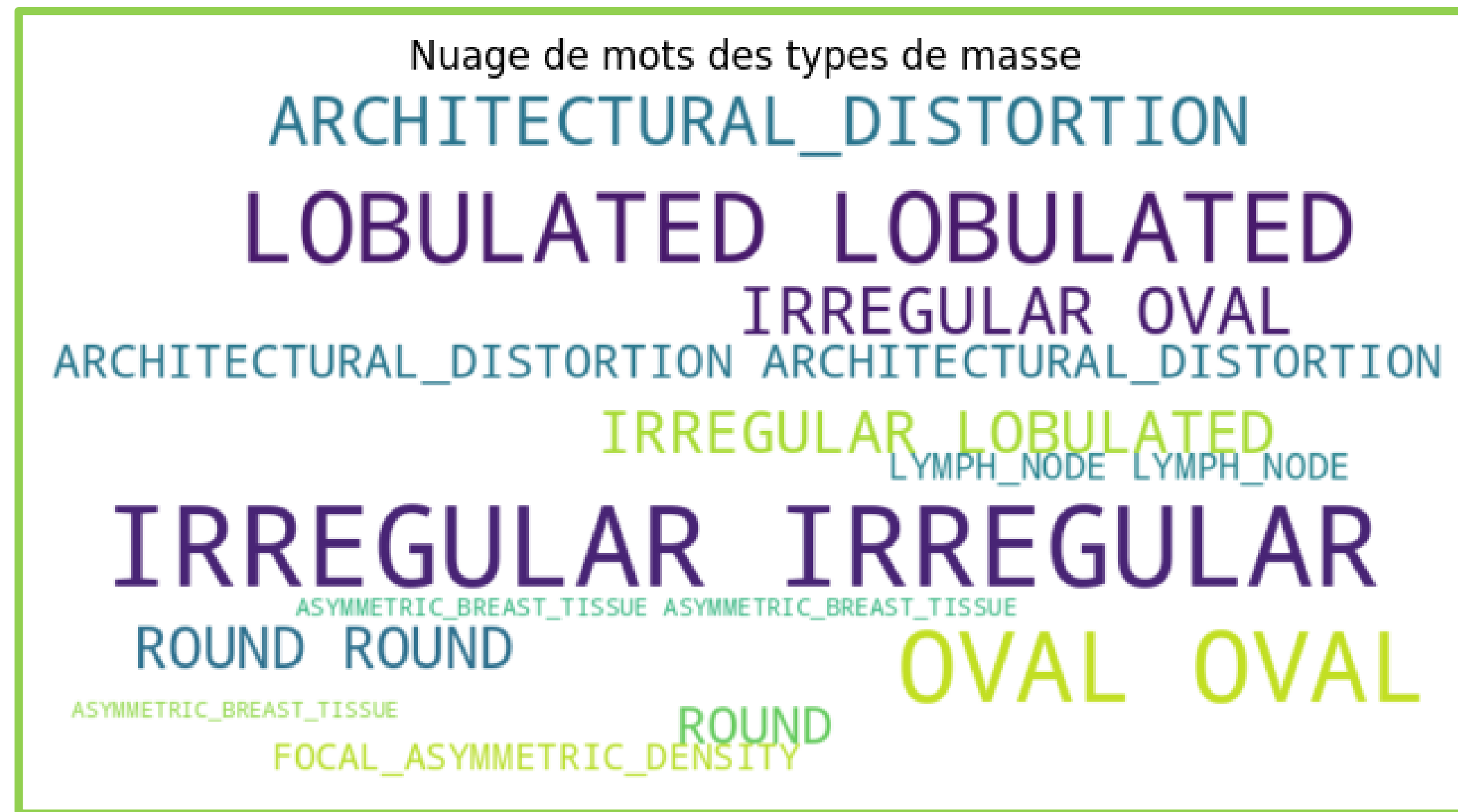
Répartition des données par les incidences des images mammographiques



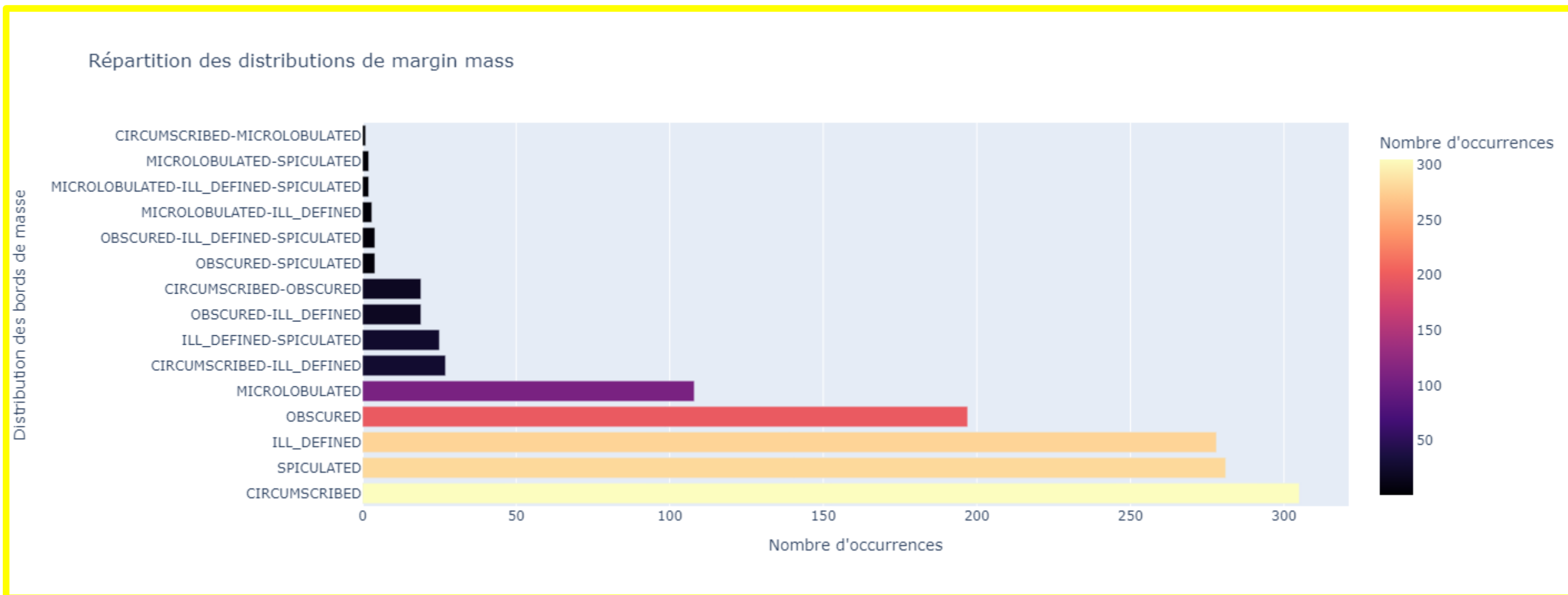
■ MLO
■ CC



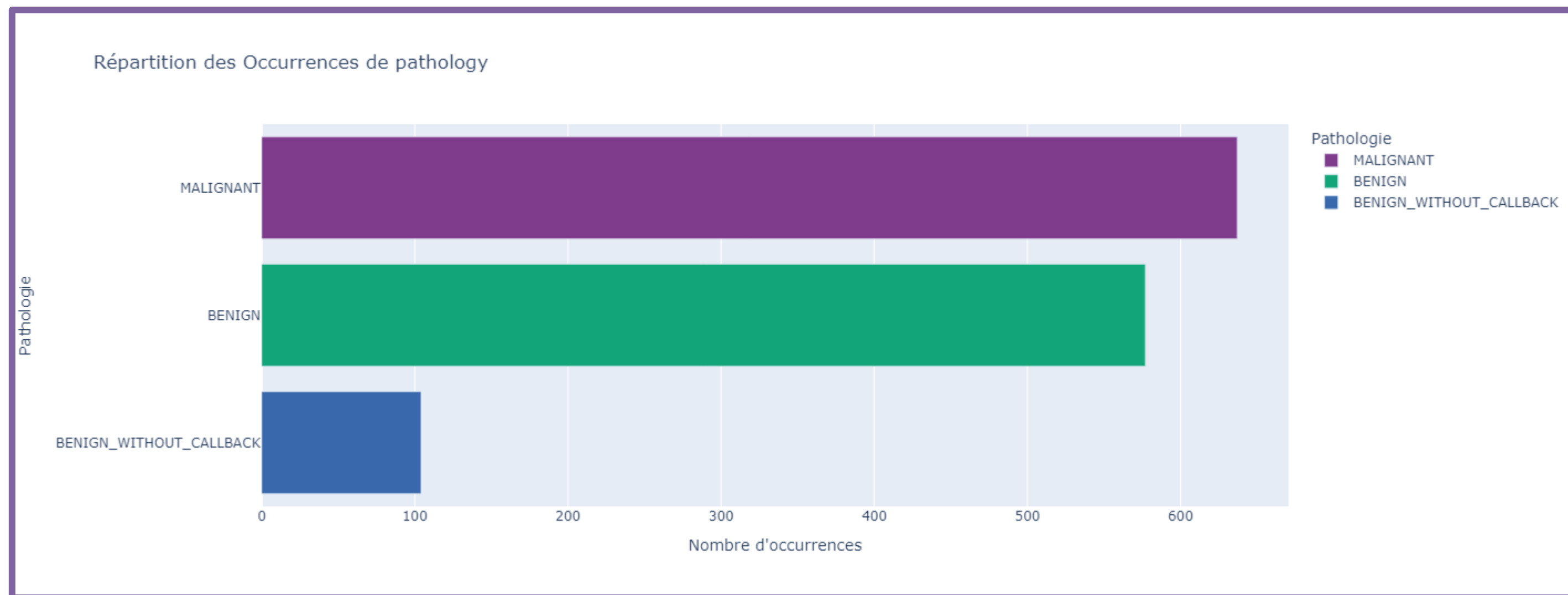
Distribution des types d'anomalie de masse



Répartition des distributions de margin masse



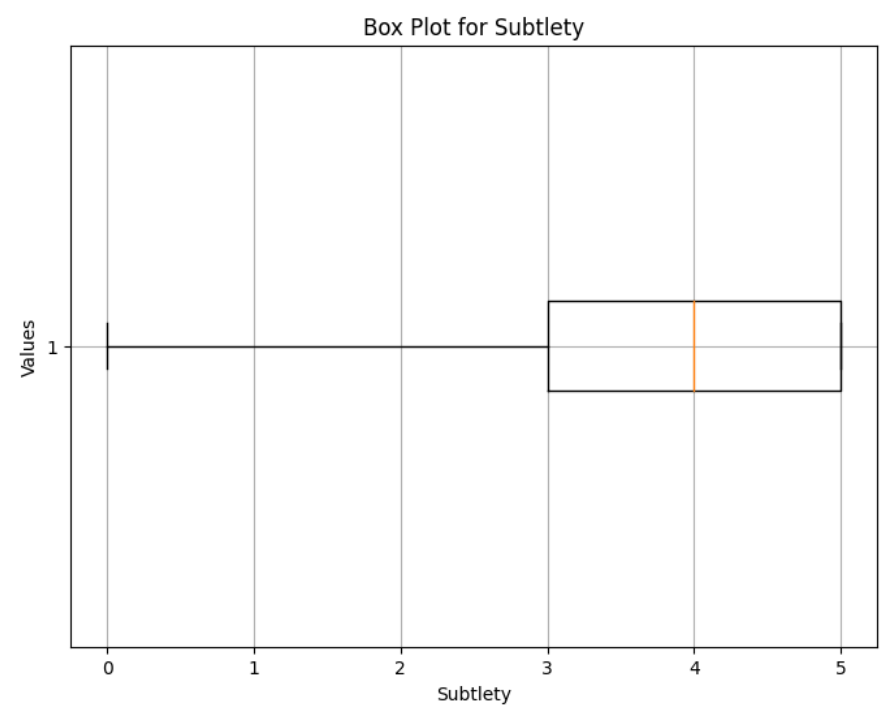
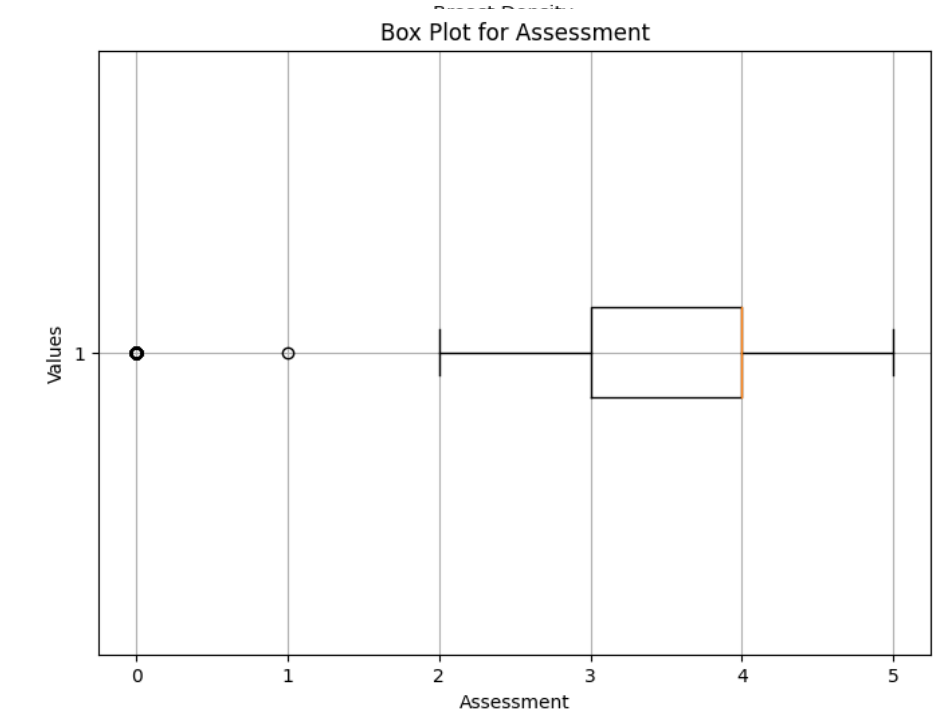
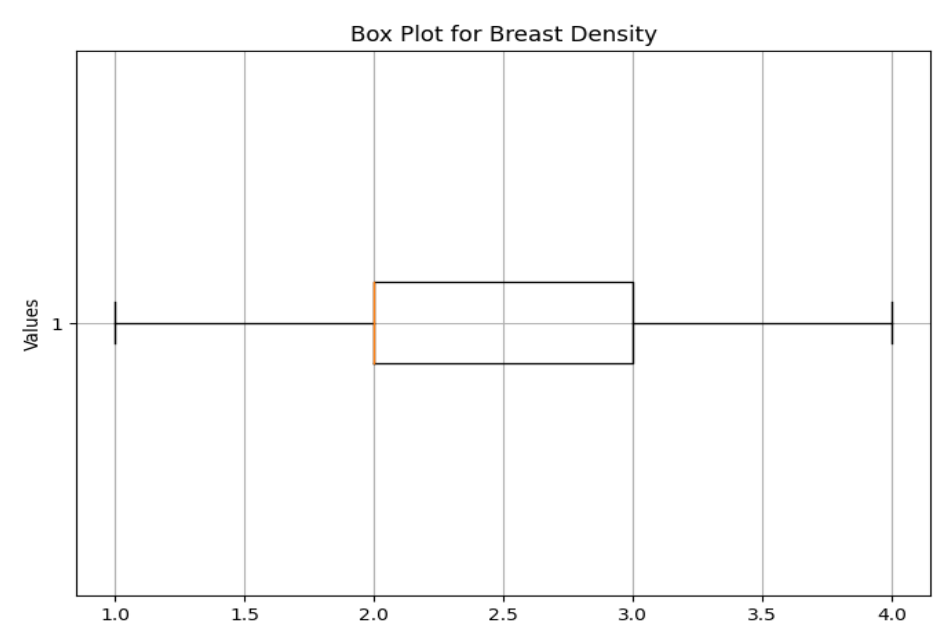
Répartition des occurrences de classes de la pathologie



Dataset 02 : Description des Cas d'Anomalie de Masse

02/ Nettoyage des Données

Data cleaning



Valeurs abérantes

4-1 Traitement des valeurs manquantes dans les deux colonnes : mass shape et mass margins

```
[ ] # Calculer le mode pour les colonnes 'mass shape' et 'mass margins' :
mode_mass_shape = mass_case_description_train_set_df['mass shape'].mode()[0]
mode_mass_margins = mass_case_description_train_set_df['mass margins'].mode()[0]

# Remplacer les valeurs manquantes par le mode dans les colonnes 'mass shape' et 'mass margins' :
mass_case_description_train_set_df['mass shape'].fillna(mode_mass_shape, inplace=True)
mass_case_description_train_set_df['mass margins'].fillna(mode_mass_margins, inplace=True)

# Vérifier si les valeurs manquantes ont été remplacées :
print("Nombre de valeurs manquantes dans 'mass shape' après remplissage :", mass_case_description_train_set_df['mass shape'].isnull().sum())
print("Nombre de valeurs manquantes dans 'mass margins' après remplissage :", mass_case_description_train_set_df['mass margins'].isnull().sum())
```

Nombre de valeurs manquantes dans 'mass shape' après remplissage : 0
Nombre de valeurs manquantes dans 'mass margins' après remplissage : 0

4-2 Renommer les colonnes de dataset

```
# Renommer les colonnes en utilisant des tirets bas pour faciliter leurs utilisations :
mass_case_description_train_set_df = mass_case_description_train_set_df.rename(columns={

    'left or right breast': 'left_or_right_breast',
    'image view': 'image_view',
    'abnormality id': 'abnormality_id',
    'abnormality type': 'abnormality_type',
    'mass shape': 'mass_shape',
    'mass margins': 'mass_margins',
    'image file path': 'image_file_path',
    'cropped image file path': 'cropped_image_file_path',
    'ROI mask file path': 'ROI_mask_file_path'
})

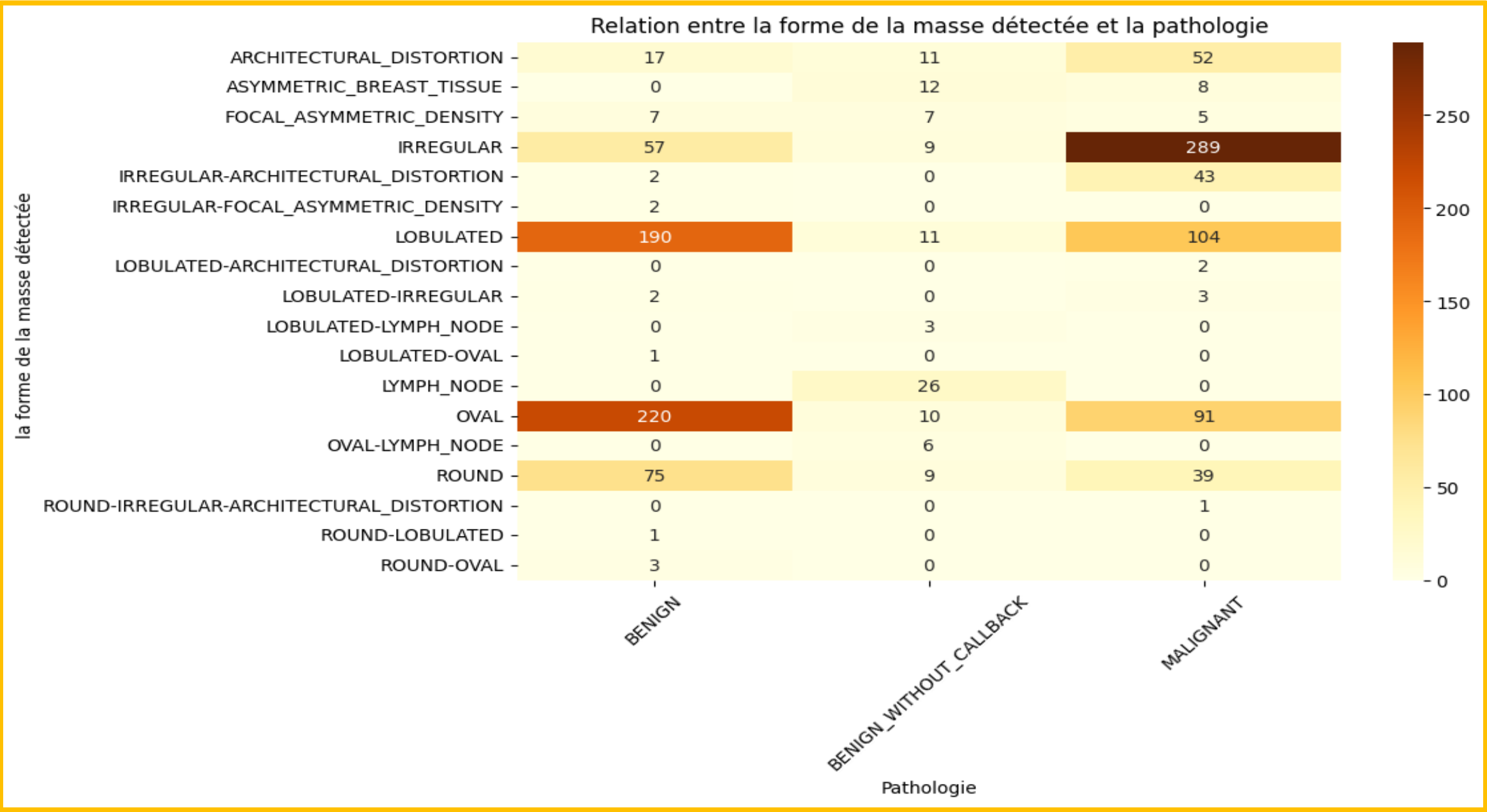
# Afficher les nouveaux noms de colonnes
print(mass_case_description_train_set_df.columns)
```

Index(['patient_id', 'breast_density', 'left_or_right_breast', 'image_view',
'abnormality_id', 'abnormality_type', 'mass_shape', 'mass_margins',
'assessment', 'pathology', 'subtlety', 'image_file_path',
'cropped_image_file_path', 'ROI_mask_file_path'],
dtype='object')

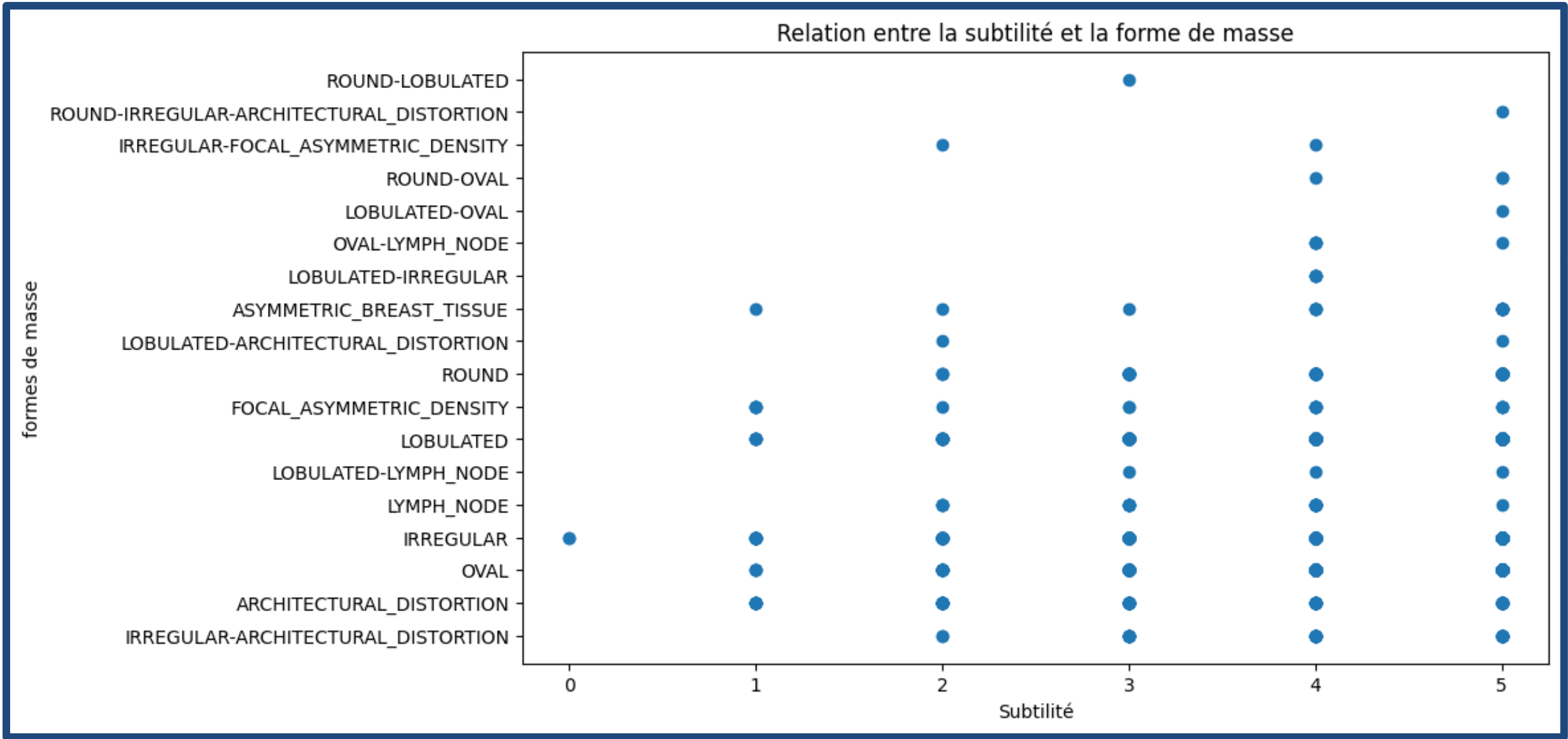
Dataset 02 : Description des Cas d'Anomalie de Masse

03/ Analyse Exploratoire des Données

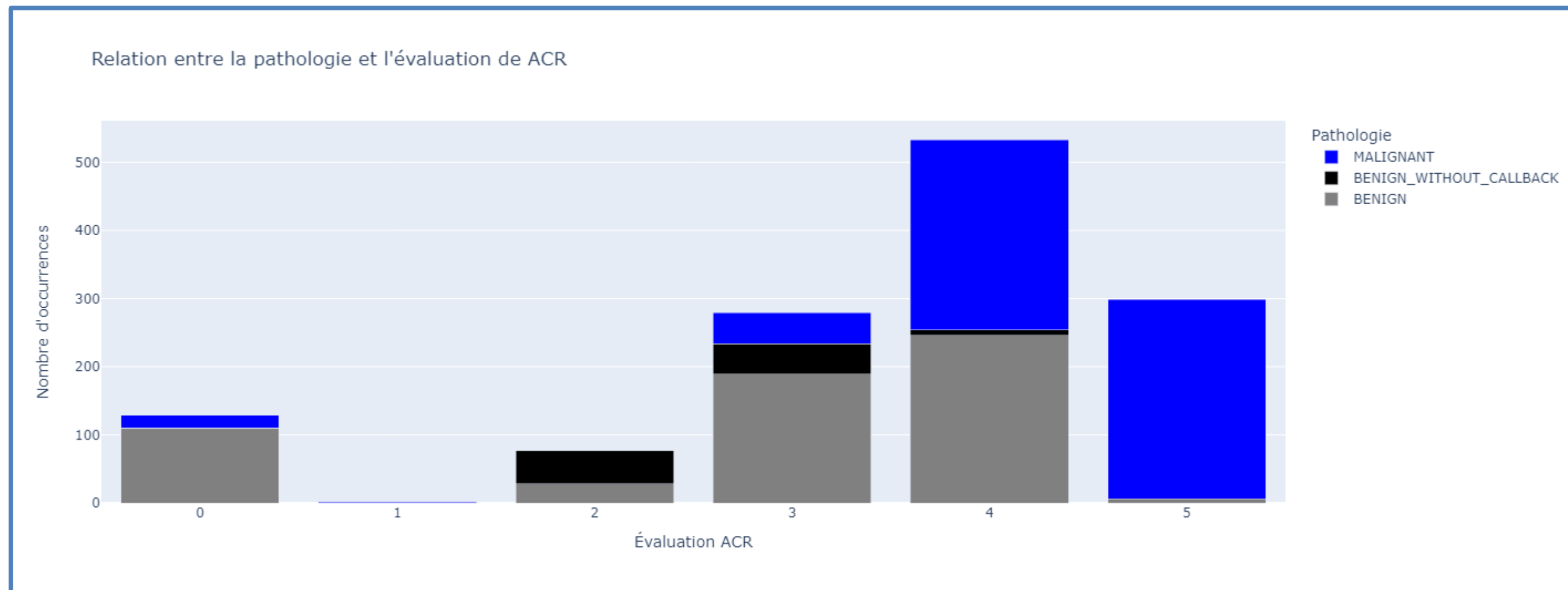
Visualisation bivariée entre la forme de la masse détectée et la pathologie



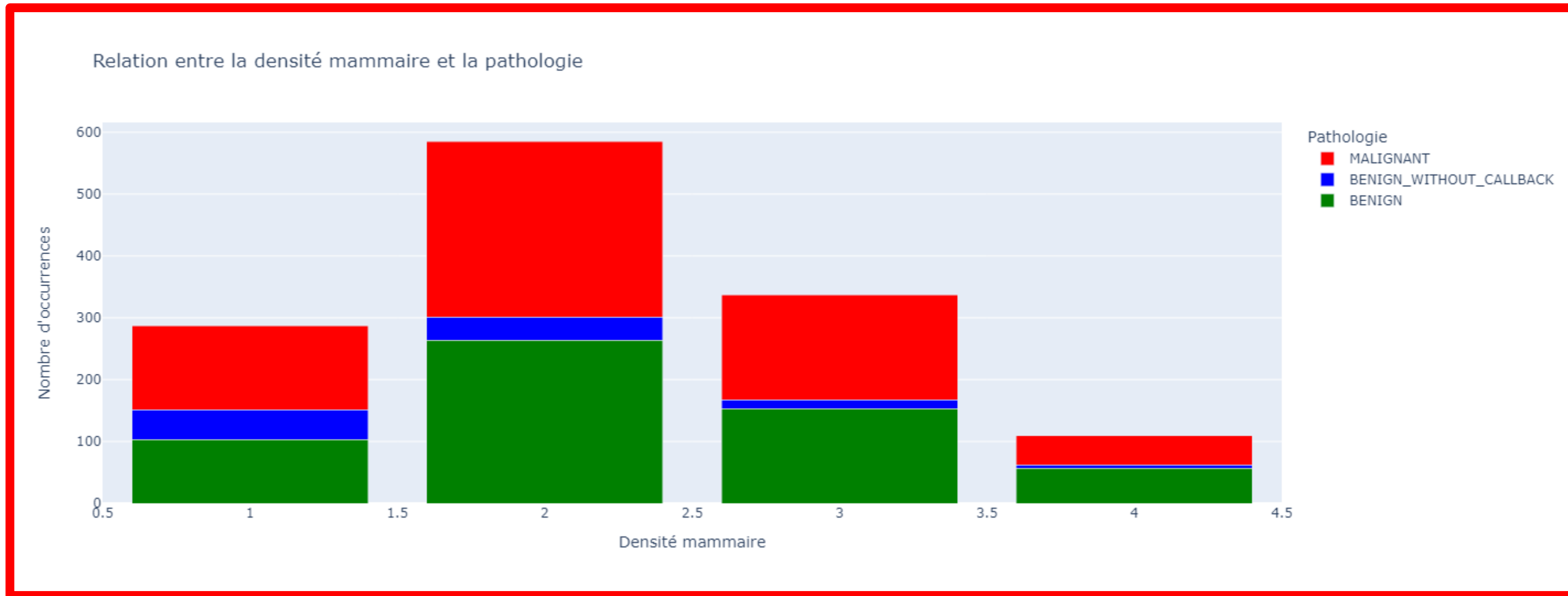
Visualisation bivariée entre la subtilité et la forme de masse



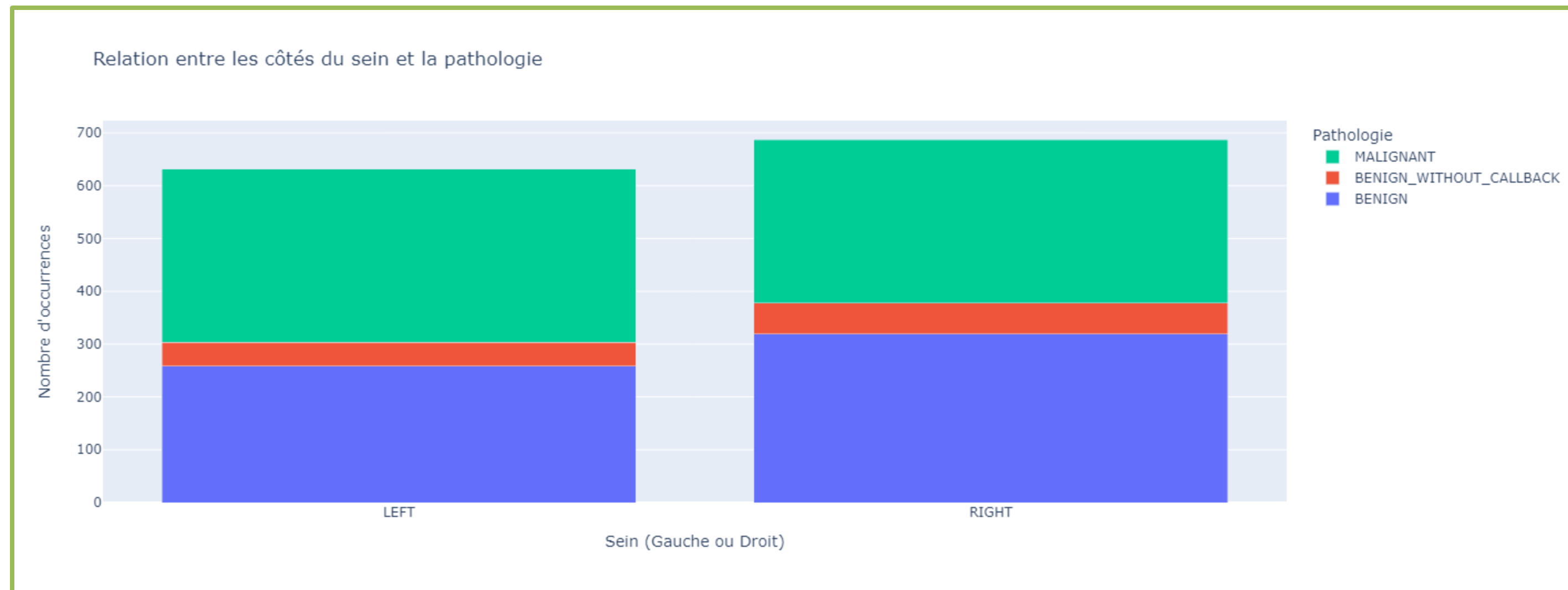
Relation entre la pathologie et l'évaluation de ACR



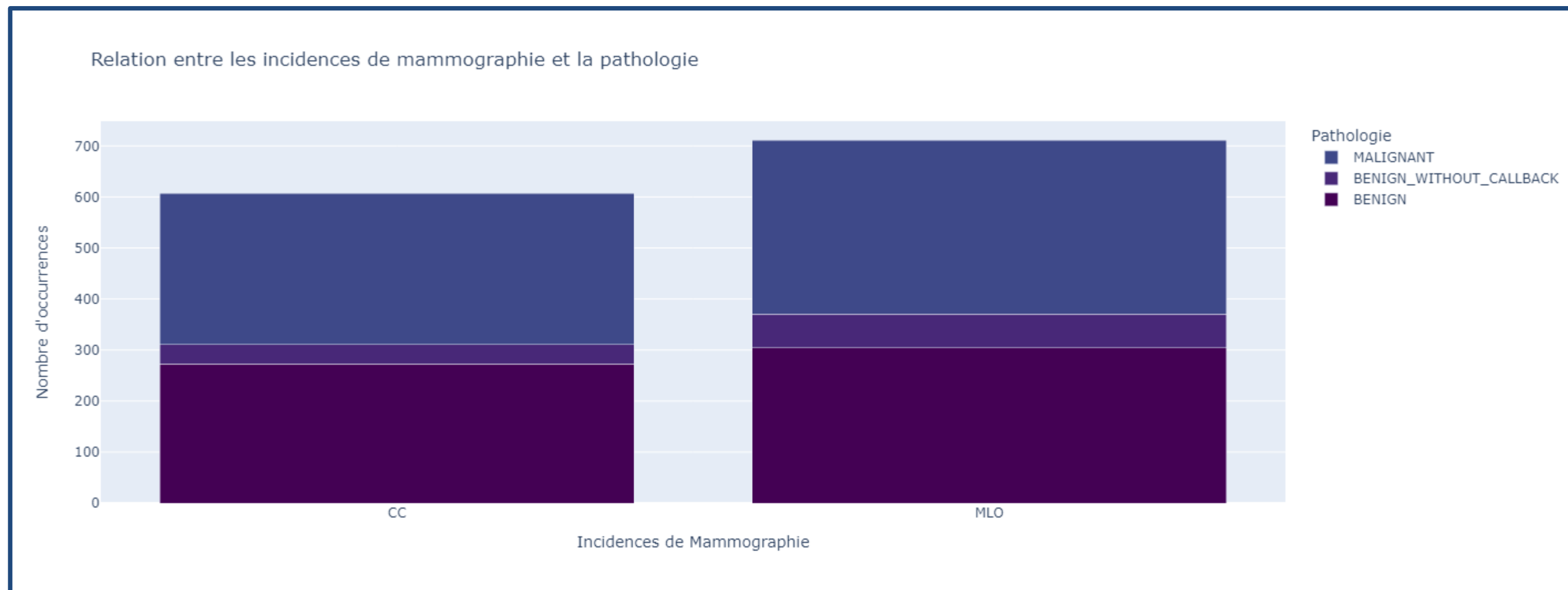
Relation entre la pathologie et la densité mammaire



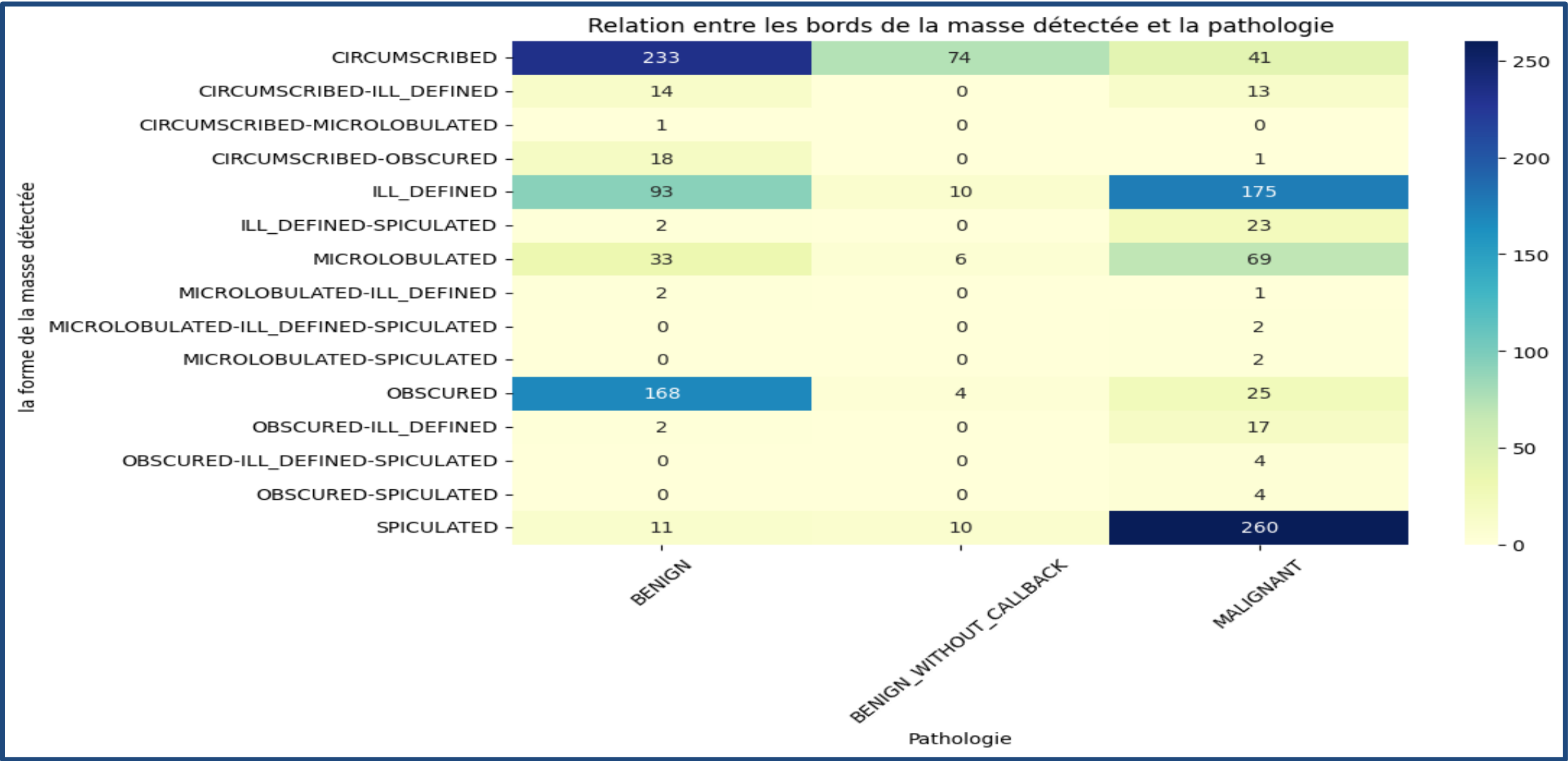
Relation entre les côtés du sein et la pathologie



Relation entre les incidences de mammographies et la pathologie



Relation entre les bords de la masse détectée et la pathologie





Dataset 03 : Dataset des données associées aux séries de mammographies



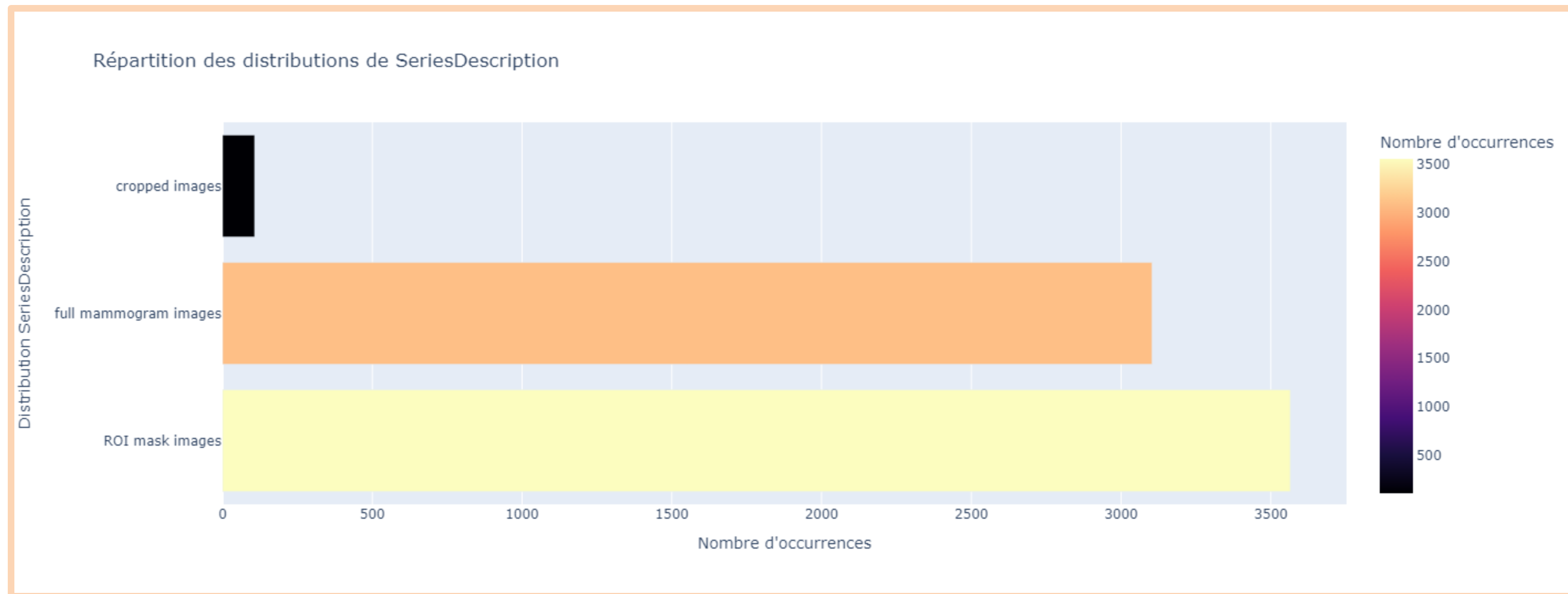
	SeriesInstanceUID	StudyInstanceUID	Modality	SeriesDescription	BodyPartExamined	SeriesNumber
0	1.3.6.1.4.1.9590.100.1.2.117041576511324414842508325652101471266	1.3.6.1.4.1.9590.100.1.2.229361142710768138411679379233064924540	MG	ROI mask images	BREAST	
1	1.3.6.1.4.1.9590.100.1.2.43873839610761788013224723323225482381	1.3.6.1.4.1.9590.100.1.2.195593486612988388325770883972107282733	MG	ROI mask images	BREAST	
2	1.3.6.1.4.1.9590.100.1.2.76741674113167646338262765132488965294	1.3.6.1.4.1.9590.100.1.2.257901172612530623323924356380431605062	MG	ROI mask images	BREAST	
3	1.3.6.1.4.1.9590.100.1.2.296931352612305599800425693833437698626	1.3.6.1.4.1.9590.100.1.2.109468616710242115222536802734027827120	MG	ROI mask images	BREAST	
4	1.3.6.1.4.1.9590.100.1.2.43665767012035310007732414810147712942	1.3.6.1.4.1.9590.100.1.2.380627129513562450304304820723973964760	MG	ROI mask images	BREAST	

Informations des séries de mammographie durant l'examen médical

Colonne	Signification
SeriesInstanceUID	Identifiant unique pour une série d'images dans un examen d'imagerie médicale.
StudyInstanceUID	Identifiant unique pour un examen d'imagerie médicale (étude).
Modality	Le type de modalité d'imagerie utilisée (dans notre cas MG).
SeriesDescription	Une description textuelle de la série d'images.
BodyPartExamined	La partie du corps examinée dans la série d'images.
SeriesNumber	Un nombre attribué à la série pour la distinguer d'autres séries dans la même étude.
Collection	Le nom de la collection ou de l'ensemble de données auquel appartient la série d'images.
Visibility	Peut représenter un indicateur de la visibilité de l'image (plus le chiffre est bas, mieux c'est).
ImageCount	Le nombre d'images contenues dans la série.



Distribution de Séries d'images MGs





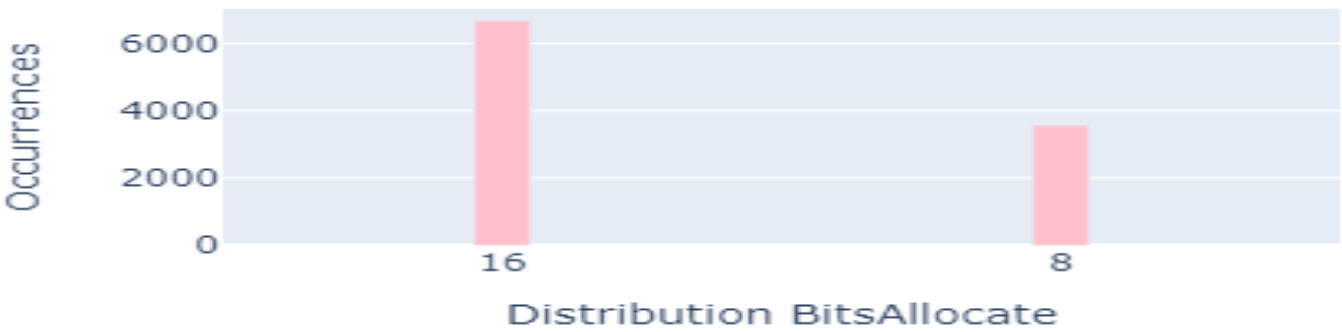
Dataset 04 : Dataset Dicom((Digital Imaging and Communications in Medicine)



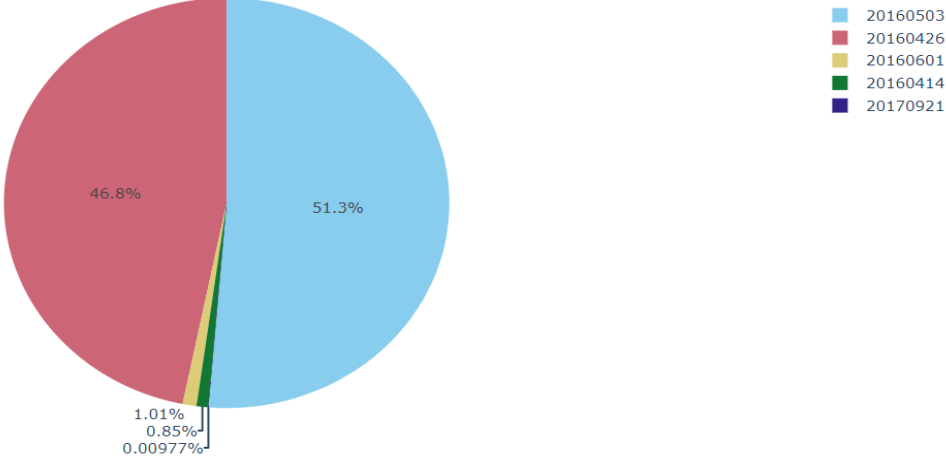
	SeriesInstanceUID	StudyInstanceUID	Modality	SeriesDescription	BodyPartExamined	SeriesNumber
0	1.3.6.1.4.1.9590.100.1.2.117041576511324414842508325652101471266	1.3.6.1.4.1.9590.100.1.2.229361142710768138411679379233064924540	MG	ROI mask images	BREAST	
1	1.3.6.1.4.1.9590.100.1.2.43873839610761788013224723323225482381	1.3.6.1.4.1.9590.100.1.2.195593486612988388325770883972107282733	MG	ROI mask images	BREAST	
2	1.3.6.1.4.1.9590.100.1.2.76741674113167646338262765132488965294	1.3.6.1.4.1.9590.100.1.2.257901172612530623323924356380431605062	MG	ROI mask images	BREAST	
3	1.3.6.1.4.1.9590.100.1.2.296931352612305599800425693833437698626	1.3.6.1.4.1.9590.100.1.2.109468616710242115222536802734027827120	MG	ROI mask images	BREAST	
4	1.3.6.1.4.1.9590.100.1.2.43665767012035310007732414810147712942	1.3.6.1.4.1.9590.100.1.2.380627129513562450304304820723973964760	MG	ROI mask images	BREAST	

Norme d'imagerie médicale numérique pour les stocker, les visualizer et les échanger

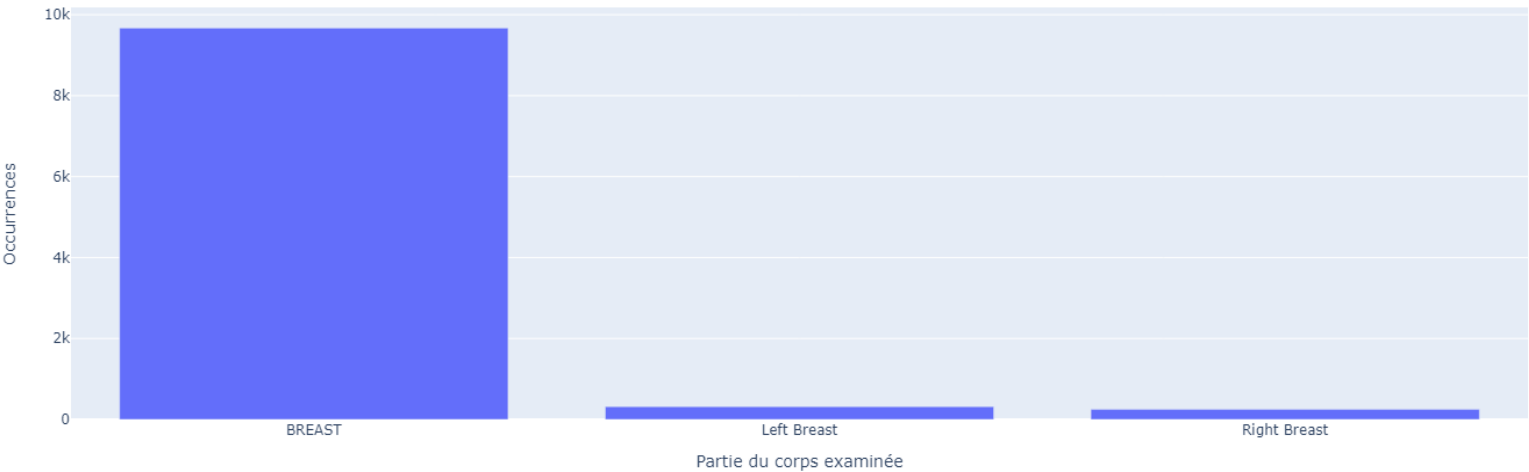
Répartition des distributions de BitsAllocated



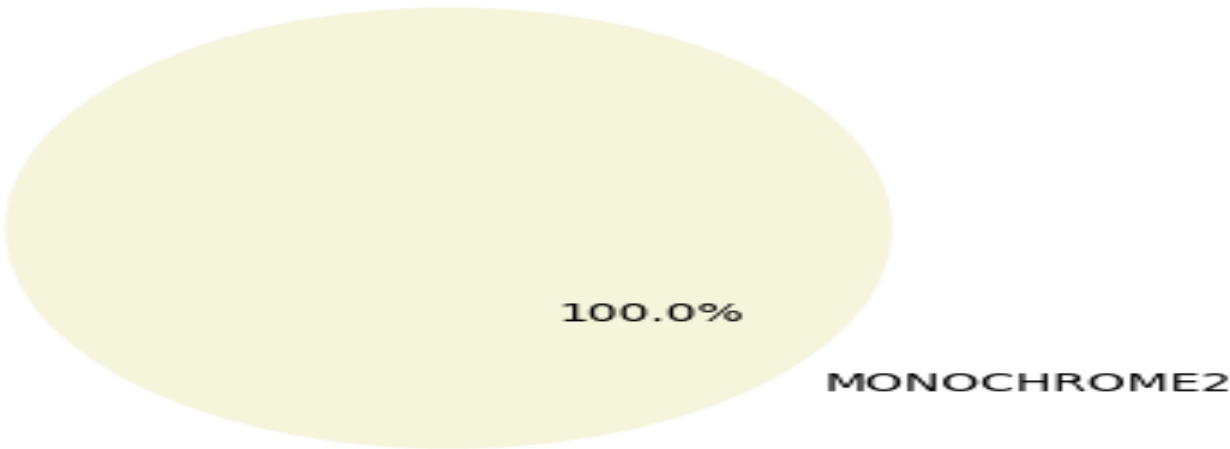
Occurrences de la colonne ContentDate



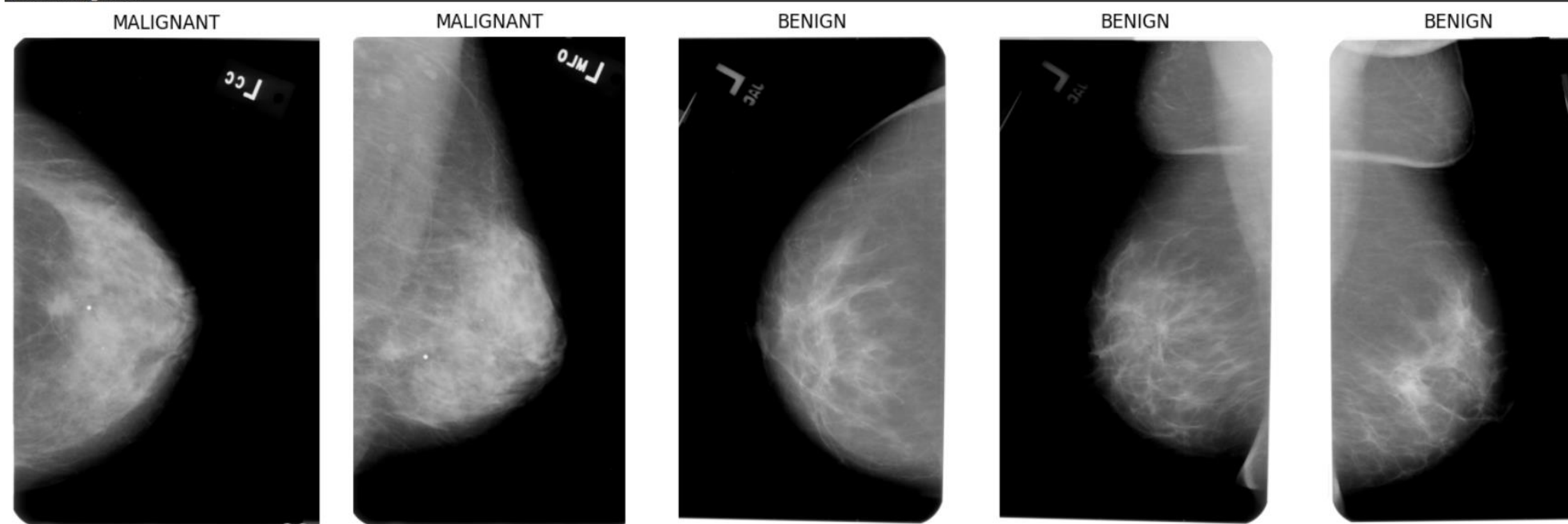
Occurrences de la colonne BodyPartExamined



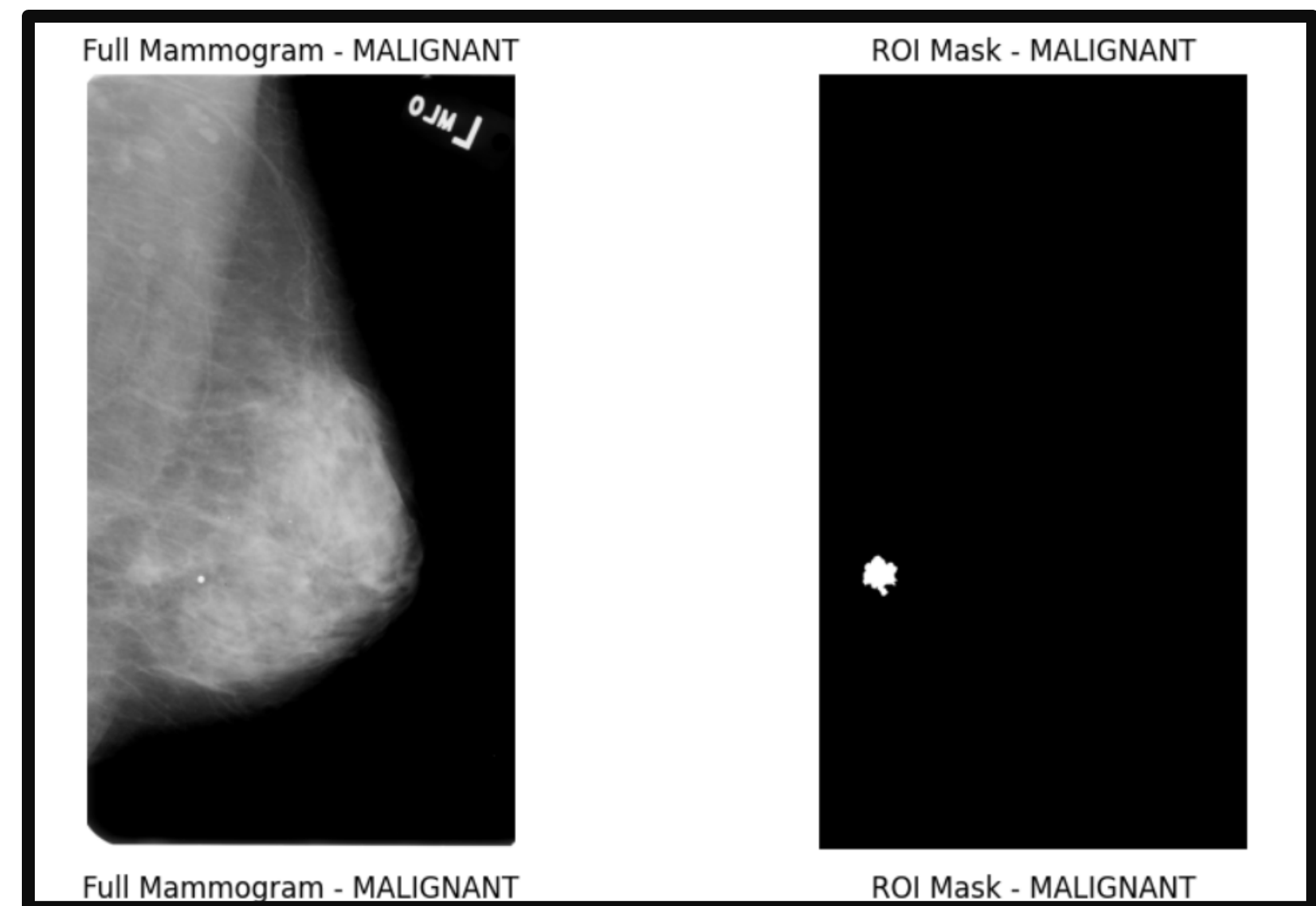
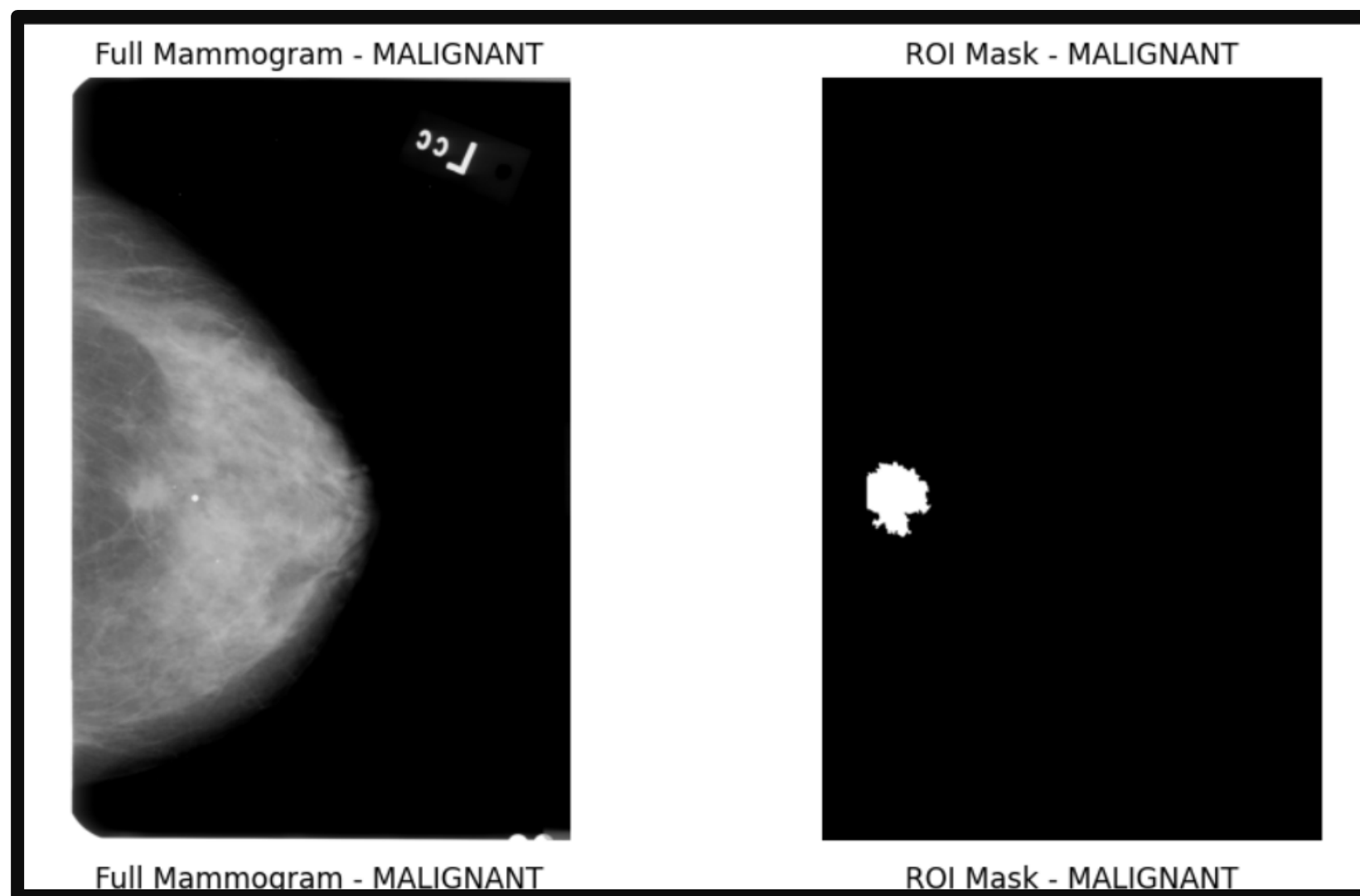
Occurrences de la colonne PhotometricInterpretation



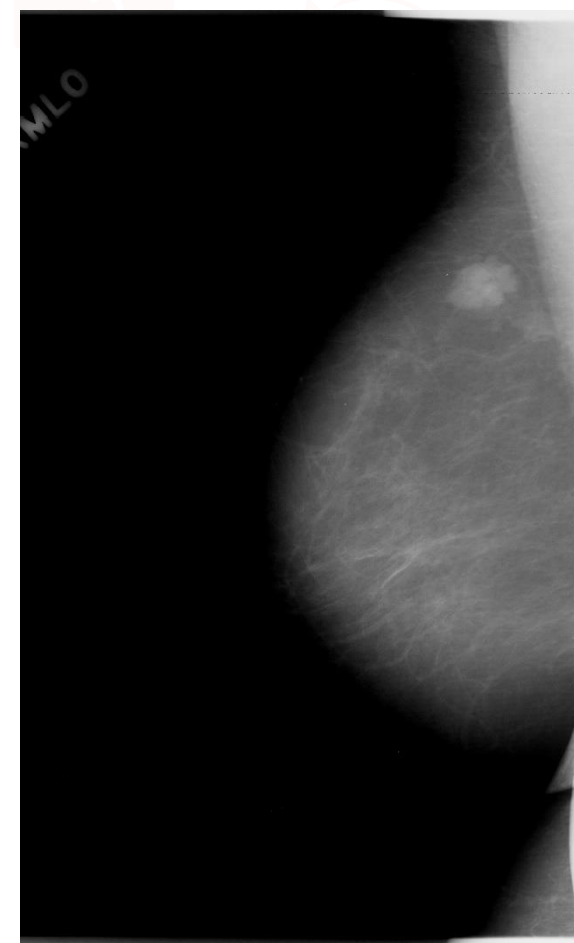
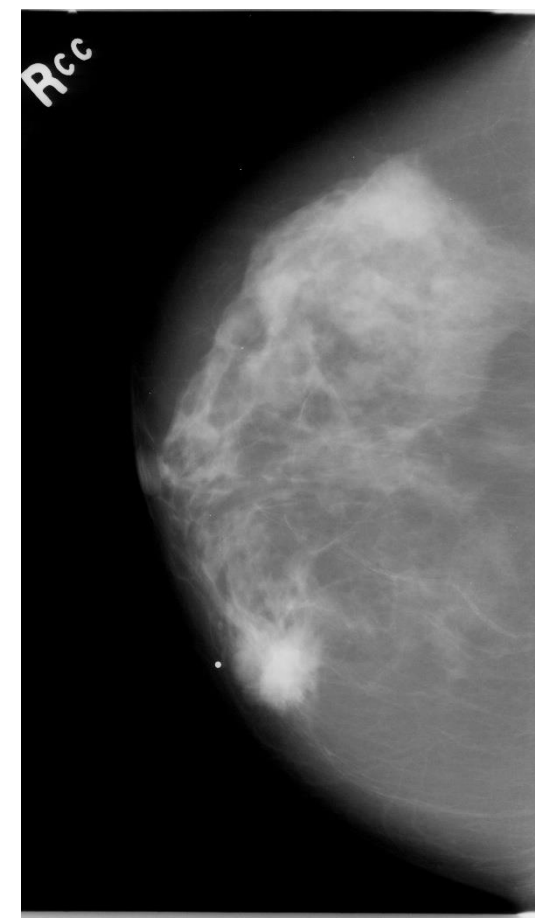
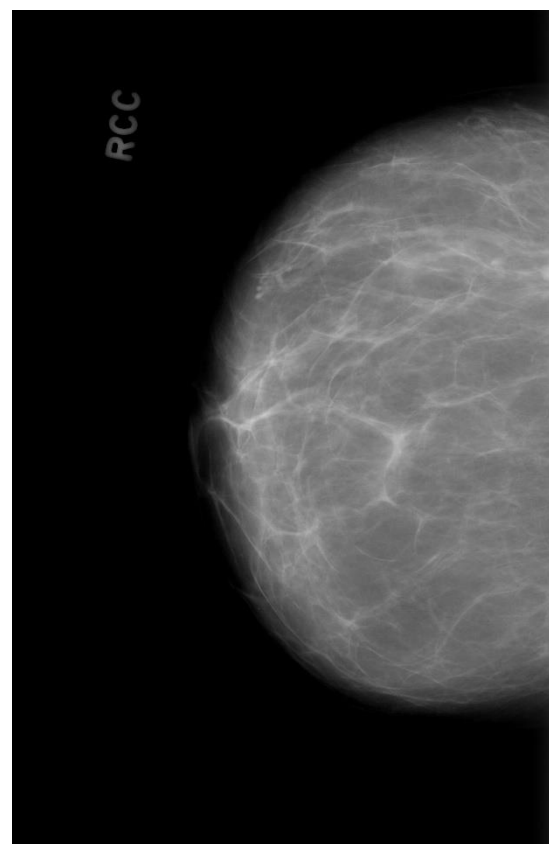
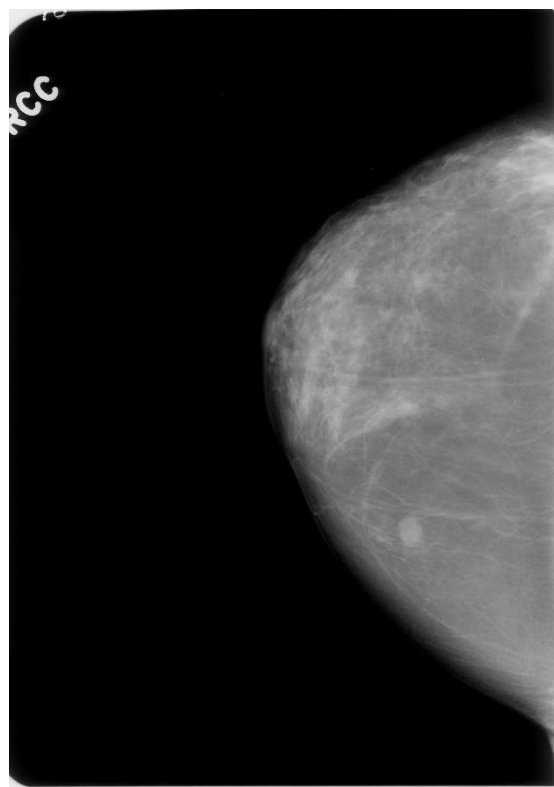
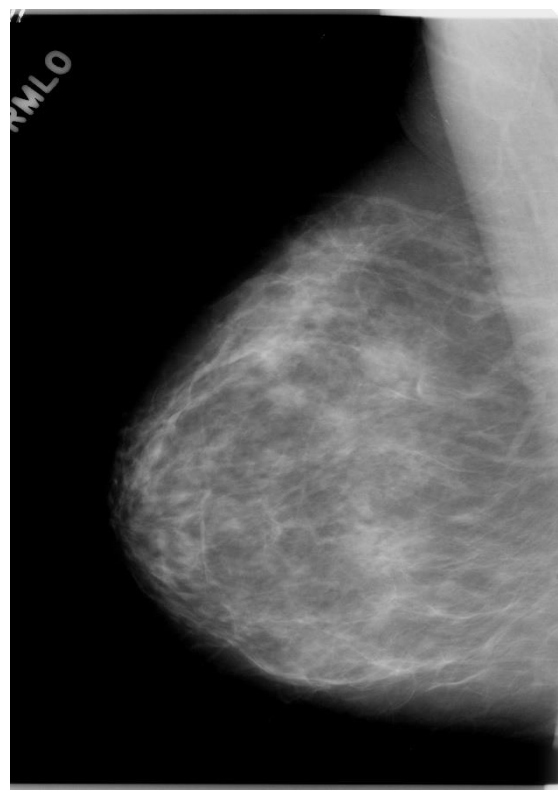
Dataset d'imagerie mammographiques



Mammographies avec leurs masques



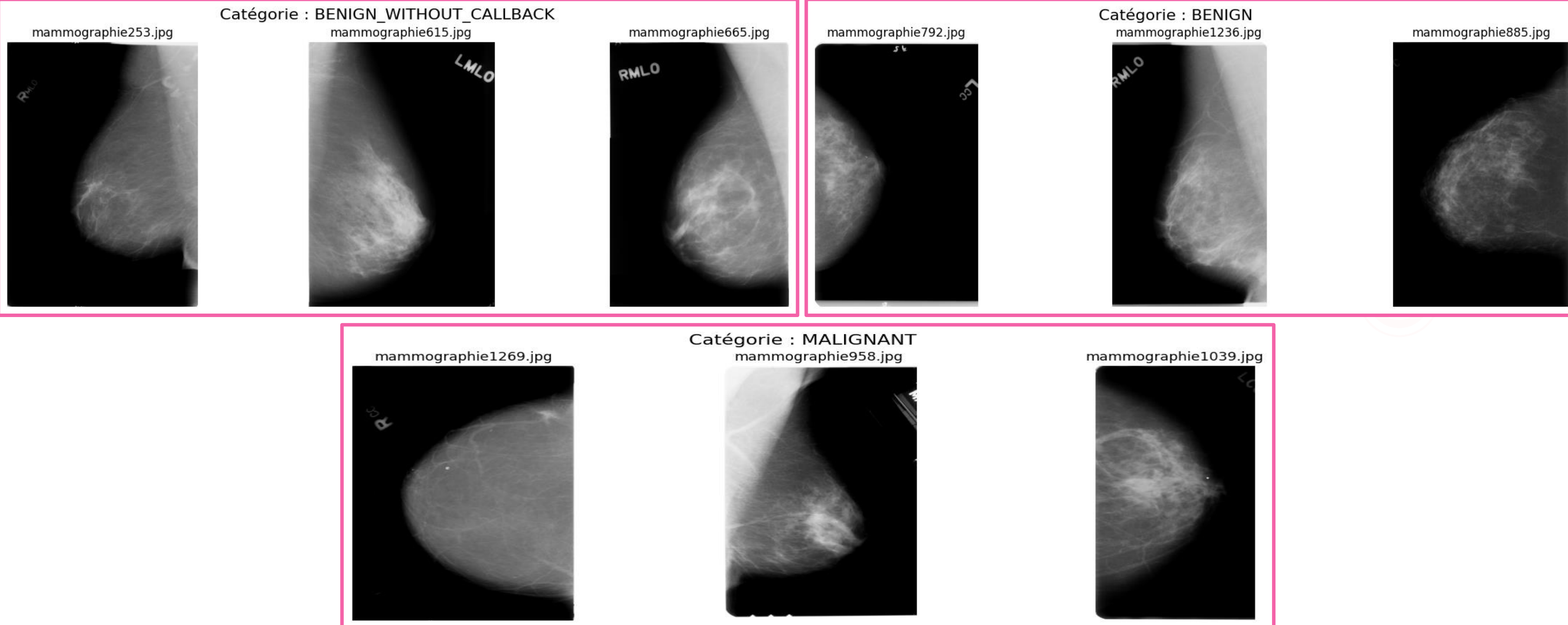
Distribution de base de données d'images mammographiques



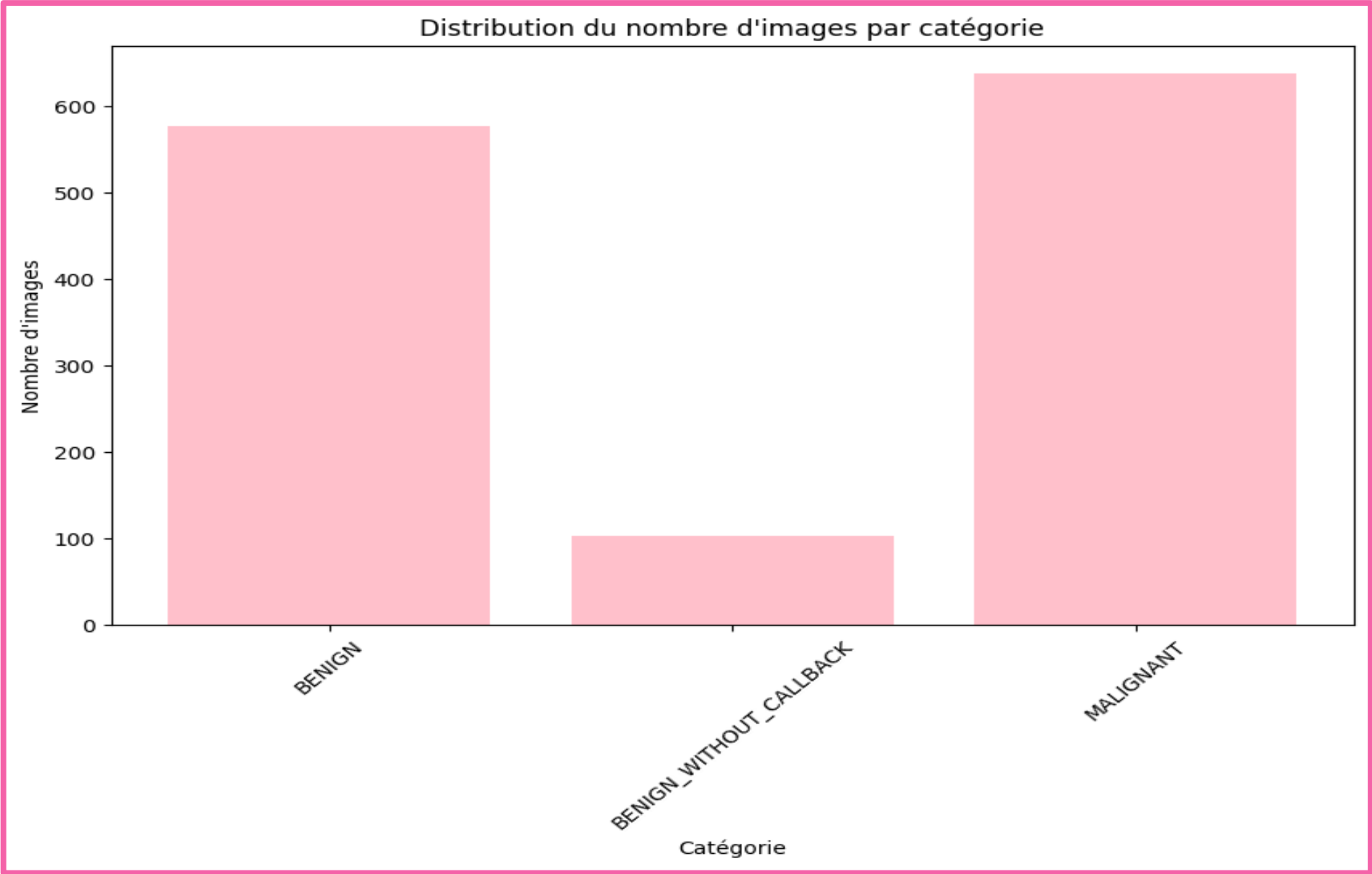
Dataset d'imagerie mammographiques

01/ Exploration des images

Affichage d'un échantillon de dataset

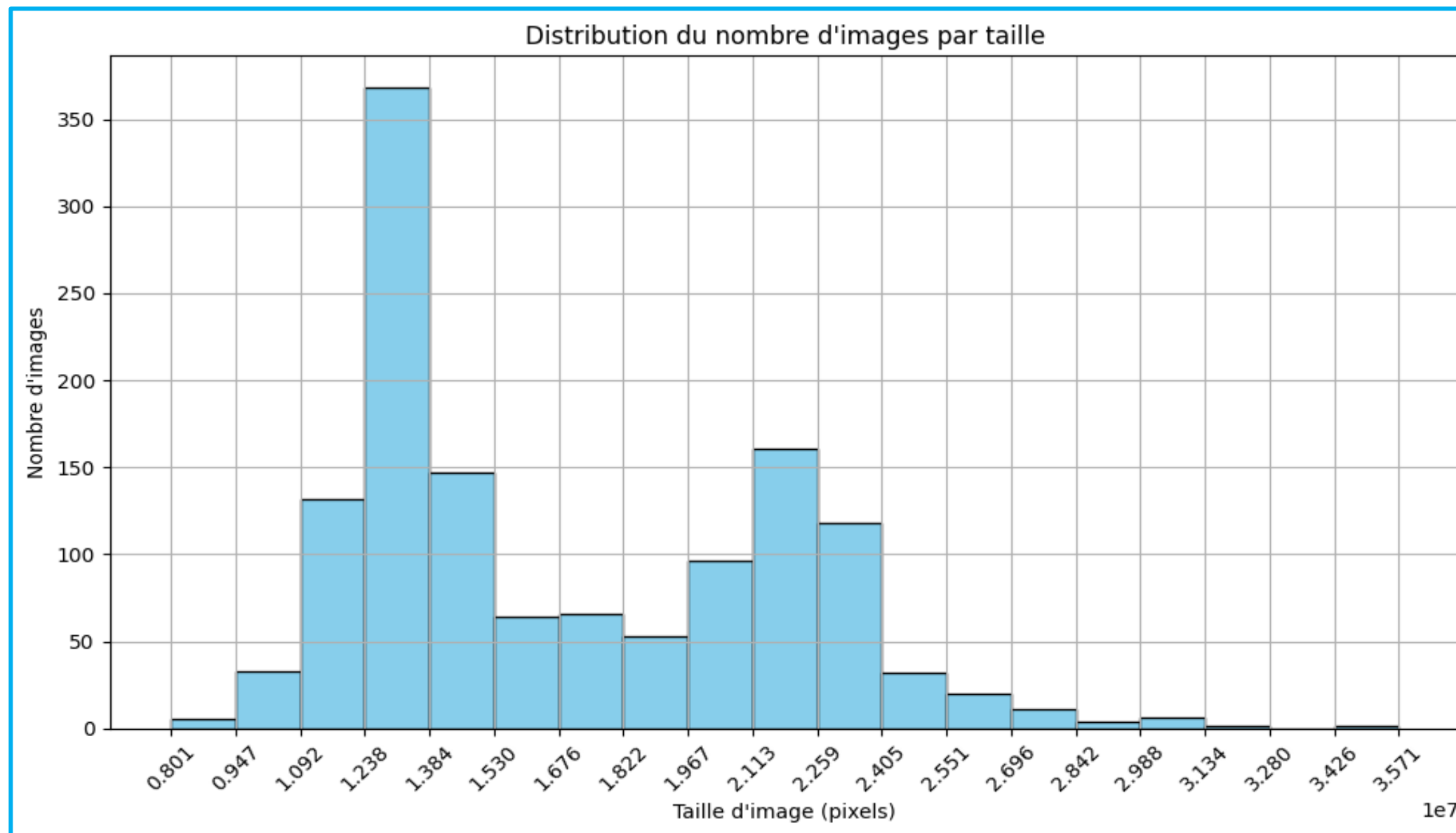


Distribution du nombre d'images par catégorie



Classe	Nombre d'images
BENIGN	577
BENIGN_WITHOUT_CALLBACK	104
MALIGNANT	637

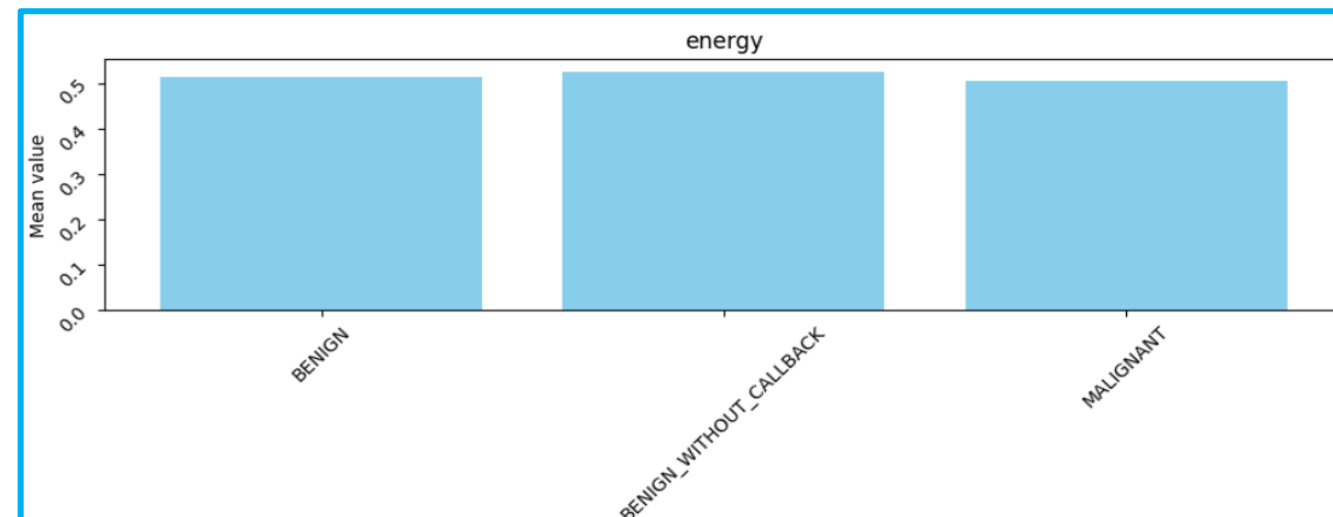
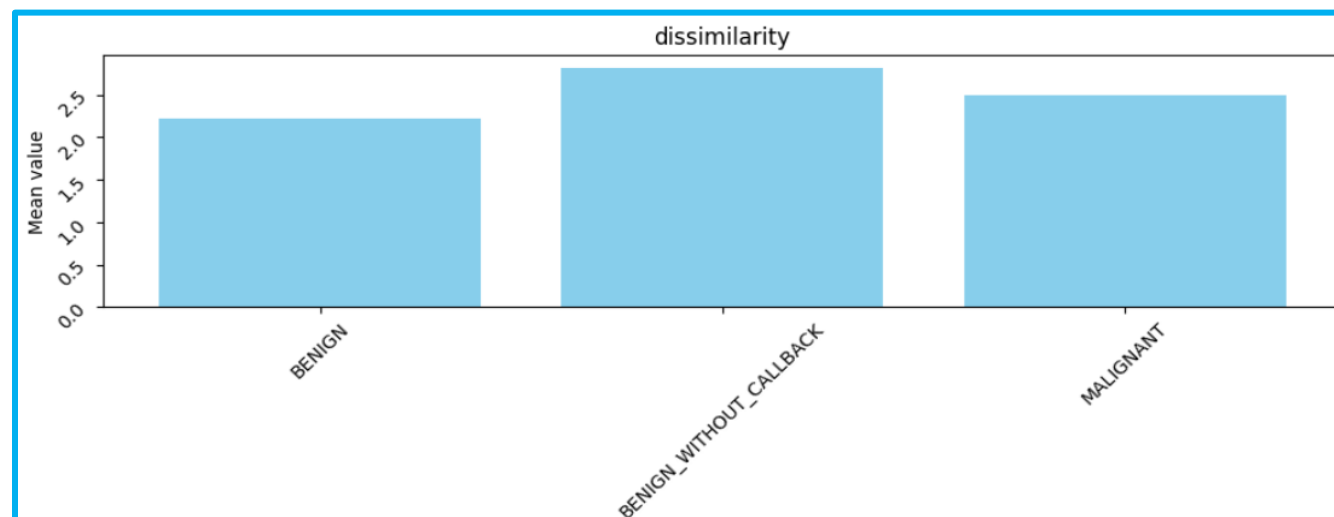
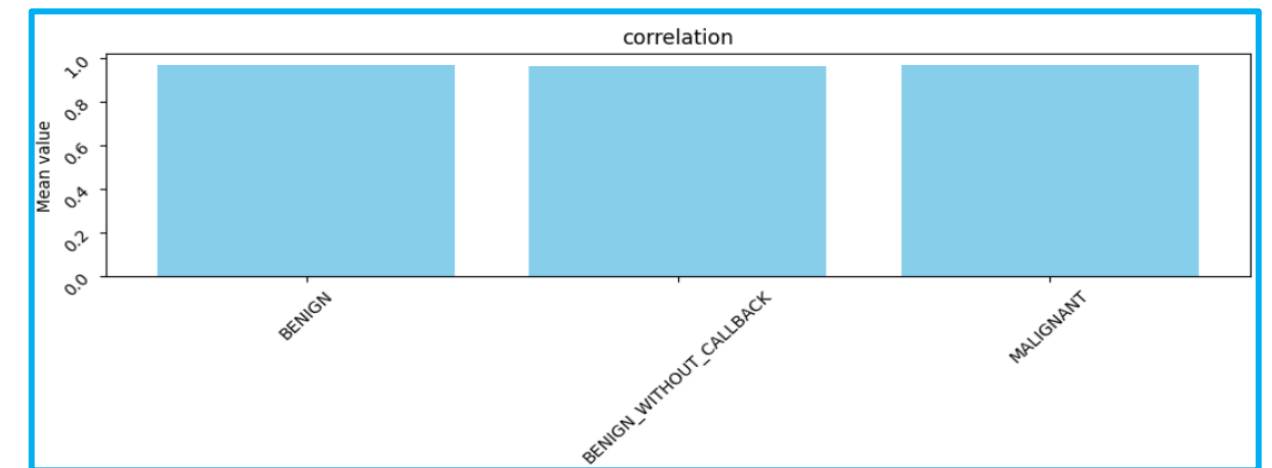
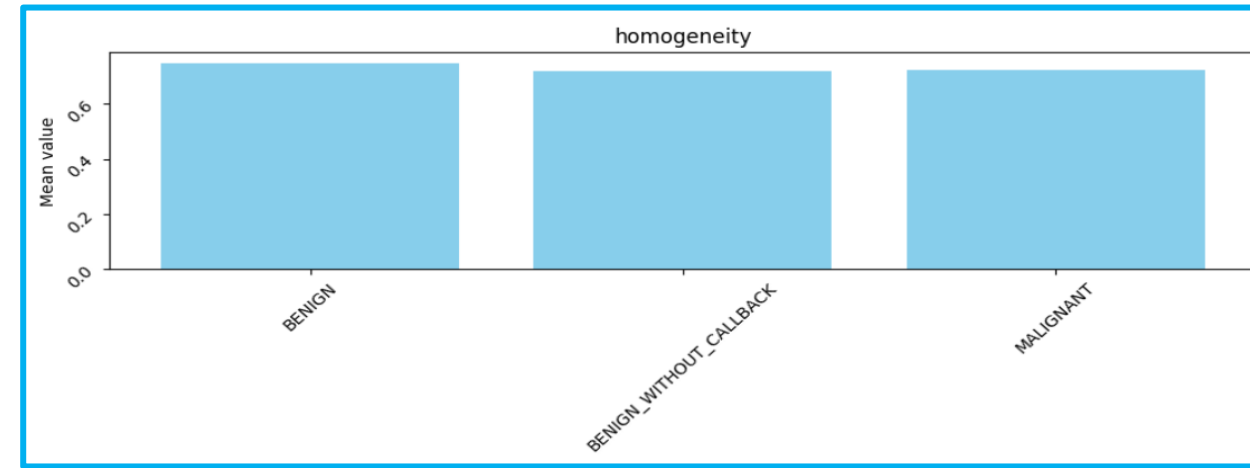
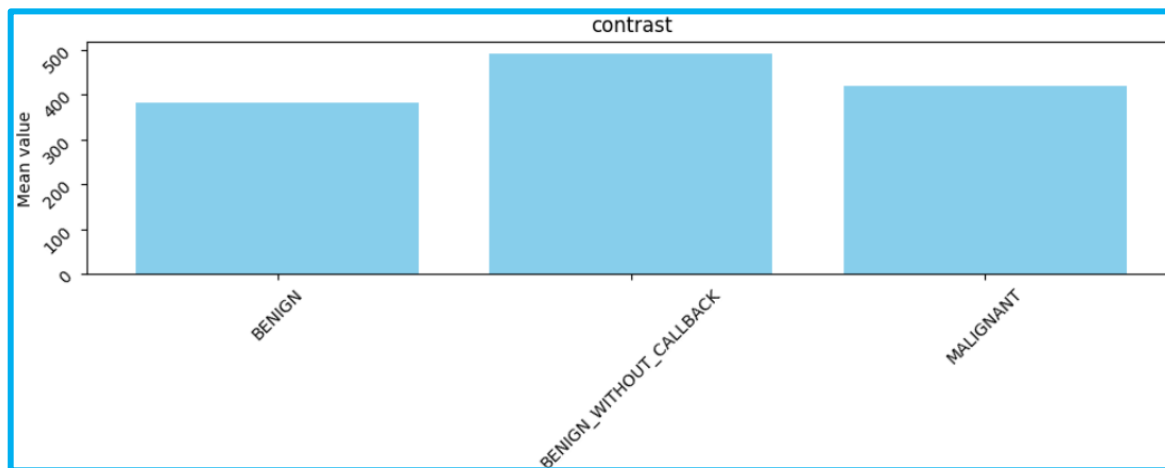
Distribution du nombre d'images par taille



Distribution des résolutions d'images



Analyse des caractéristiques de texture des images



Technique : la matrice de co-occurrence de niveaux de gris (GLCM)

Merci de Votre Attention

Questions ?