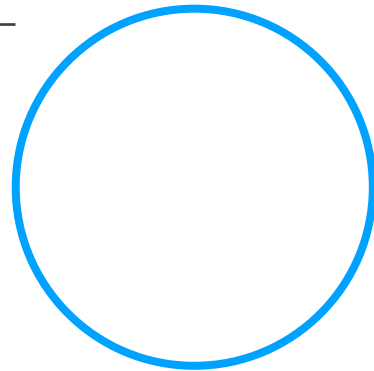




Détection de faux billets

MAXIMEBCH – DATA ANALYST



INTRODUCTION

- Mission pour le compte de l'Office central pour la répression du faux monnayage (Ministère de l'intérieur)
- Objectif : Conception d'un algorithme permettant de détecter les faux billets



INTRODUCTION

- Le **faux-monnayage** est une activité bien connue du crime organisé, mais a également été utilisé par des pays pour affaiblir l'économie de pays rivaux (Napoléon 1^{er}, UK pendant la guerre d'indépendance américaine, régime nazi...).
- En France, les faux-monnayeurs sont condamnés à mort jusqu'en 1932. Actuellement, la peine est de trente ans de réclusion criminelle et 450 000 EUR d'amende.
- Mesures préventives : inclusion de détails très fins difficiles à reproduire pour reconnaître facilement la fausse monnaie (hologrammes, micro impressions, encres d'impression optique...).
- Selon la BCE, on compte 17 contrefaçons par million de vrai billets (sur 25 milliards en circulation).



Fausse monnaie (Empire romain)



ANALYSE DES DONNÉES

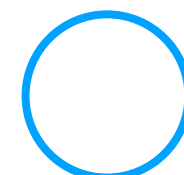
FICHER SOURCE



1 fichier .csv fourni par la police judiciaire :

- Contient les dimensions de 70 faux billets et 100 vrais billets
- Les dimensions sont exprimées par les variables : *diagonal*, *height_left*, *height_right*, *margin_low*, *margin_up* et *length*.

	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length
0	True	171.81	104.86	104.95	4.52	2.89	112.83
1	True	171.67	103.74	103.70	4.01	2.87	113.29
2	True	171.83	103.76	103.76	4.40	2.88	113.84
3	True	171.80	103.78	103.65	3.73	3.12	113.63
4	True	172.05	103.70	103.75	5.04	2.27	113.55
...
165	False	172.11	104.23	104.45	5.24	3.58	111.78
166	False	173.01	104.59	104.31	5.04	3.05	110.91
167	False	172.47	104.27	104.10	4.88	3.33	110.68
168	False	171.82	103.97	103.88	4.73	3.55	111.87
169	False	171.96	104.00	103.95	5.63	3.26	110.96



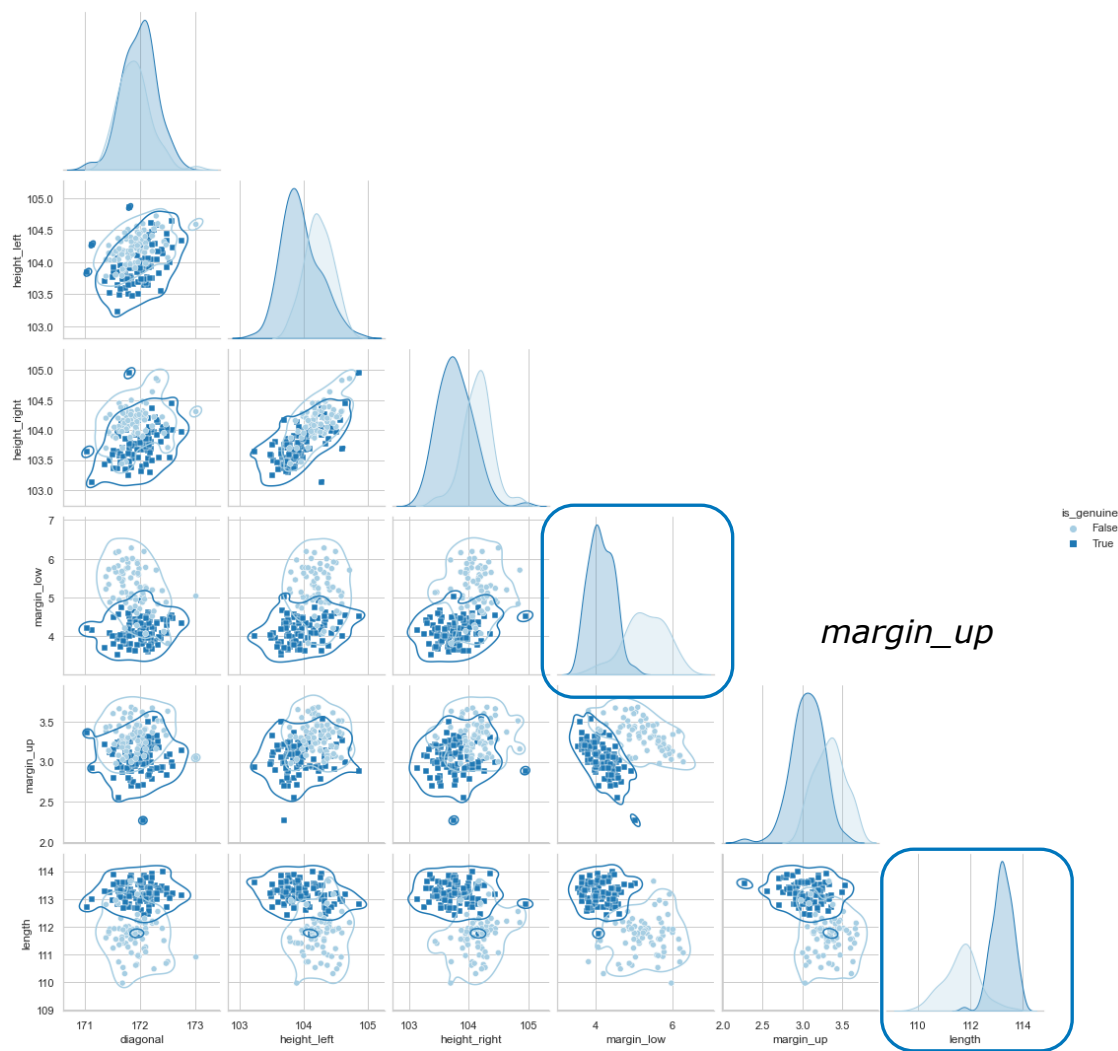


ANALYSE DES DONNÉES

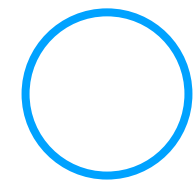
DISTRIBUTION VRAIS/FAUX BILLETS PAR VARIABLE

 **Vrai**

 **Faux**



- Les vrais billets semblent avoir de fortes valeurs dans les variables *margin_up* et *length*

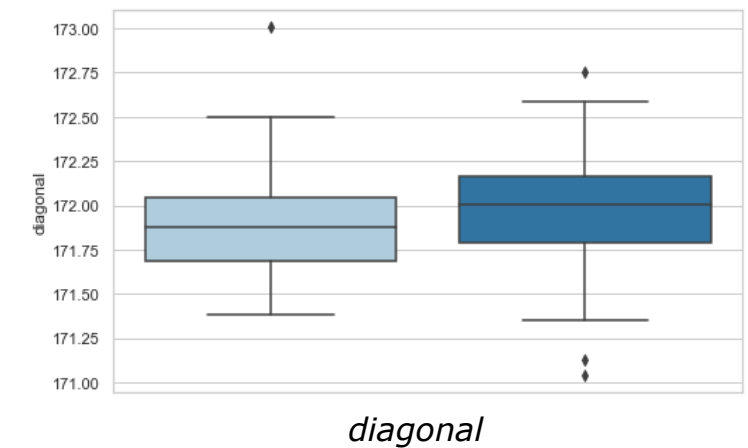
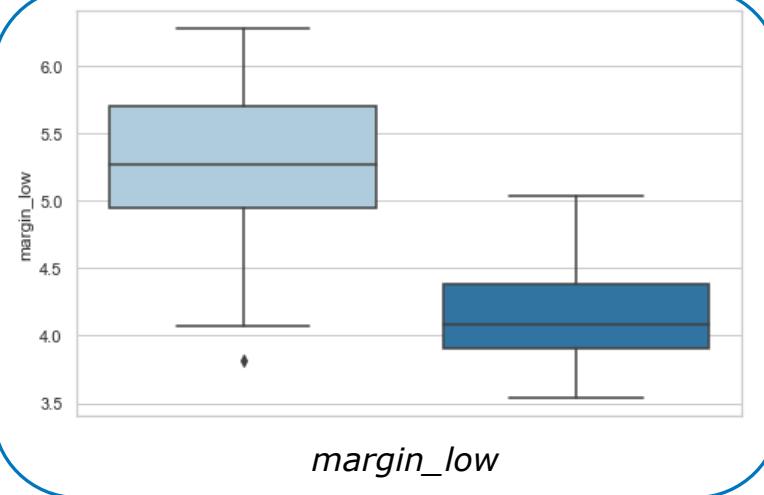
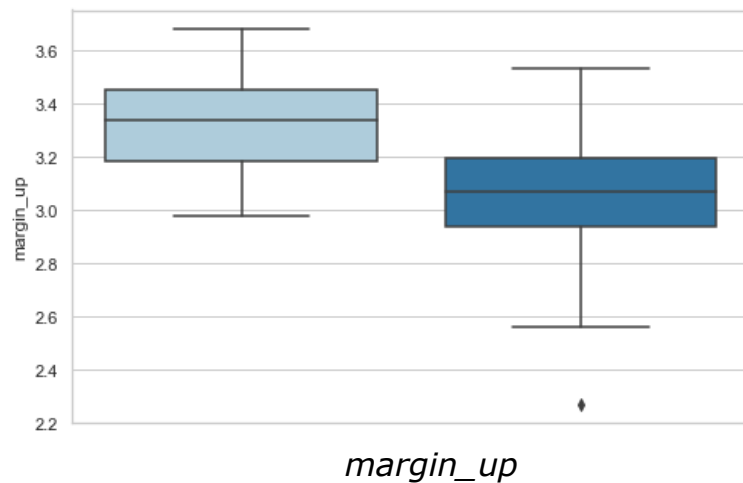
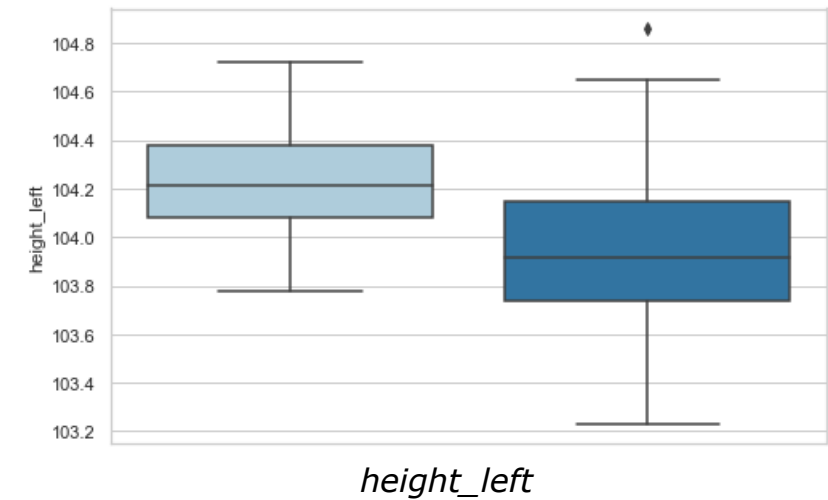
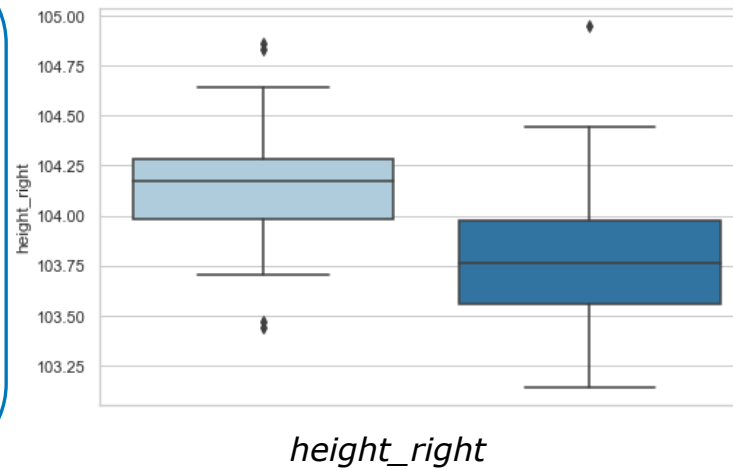
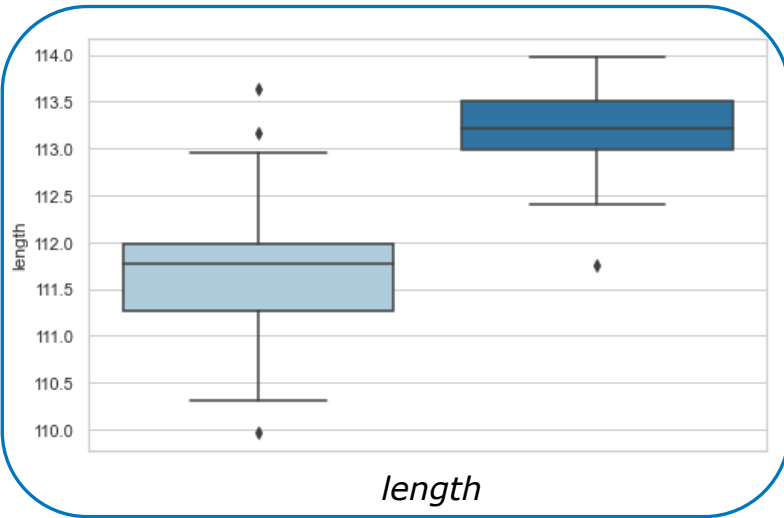


ANALYSE DES DONNÉES

DISTRIBUTION DES VARIABLES PAR AUTHENTICITÉ

 Vrai

 Faux



ANALYSE DES DONNÉES

CORRÉLATIONS À LA VARIABLE « IS_GENUINE »

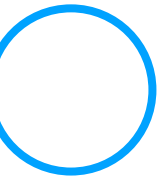
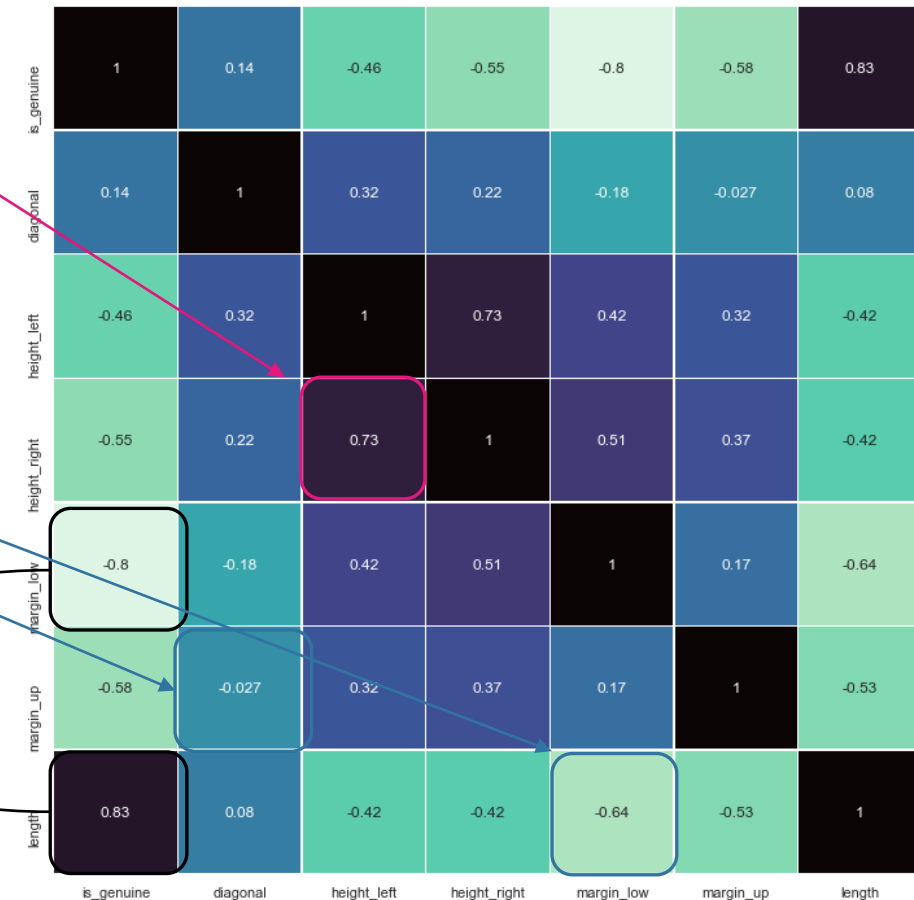


- Forte corrélation positive entre *height_right* et *height_left*

- Fortes corrélations négatives entre :

- *Margin_low* et *length*
- *Diagonal* et *margin_up*

- Comme vu précédemment, les vrais billets ont plutôt une forte *length* (corrélation positive) et une faible *margin_low*

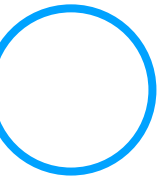
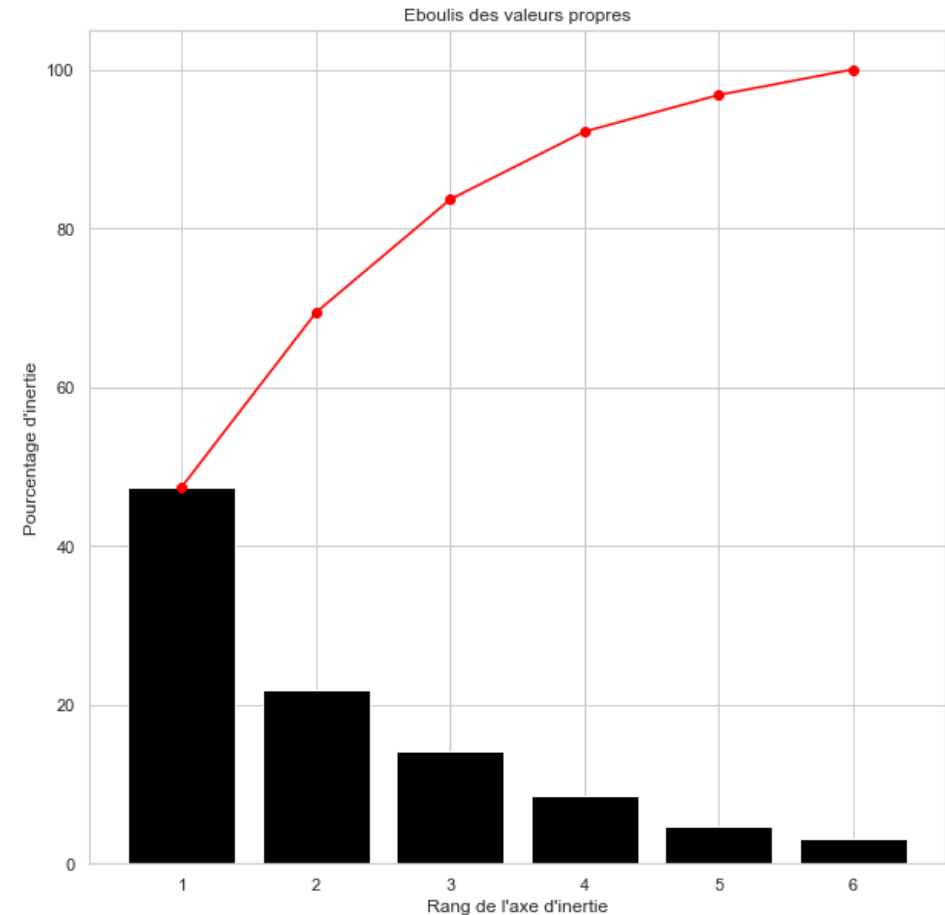


ANALYSE DES DONNÉES

ANALYSE EN COMPOSANTES PRINCIPALES

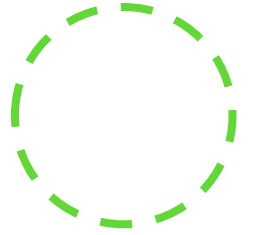


- Critère du coude :
 - Le premier axe retient 47% de l'inertie totale
 - Le deuxième axe retient 22% de l'inertie totale, pour un taux d'inertie expliquée de 69%

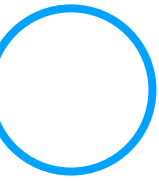
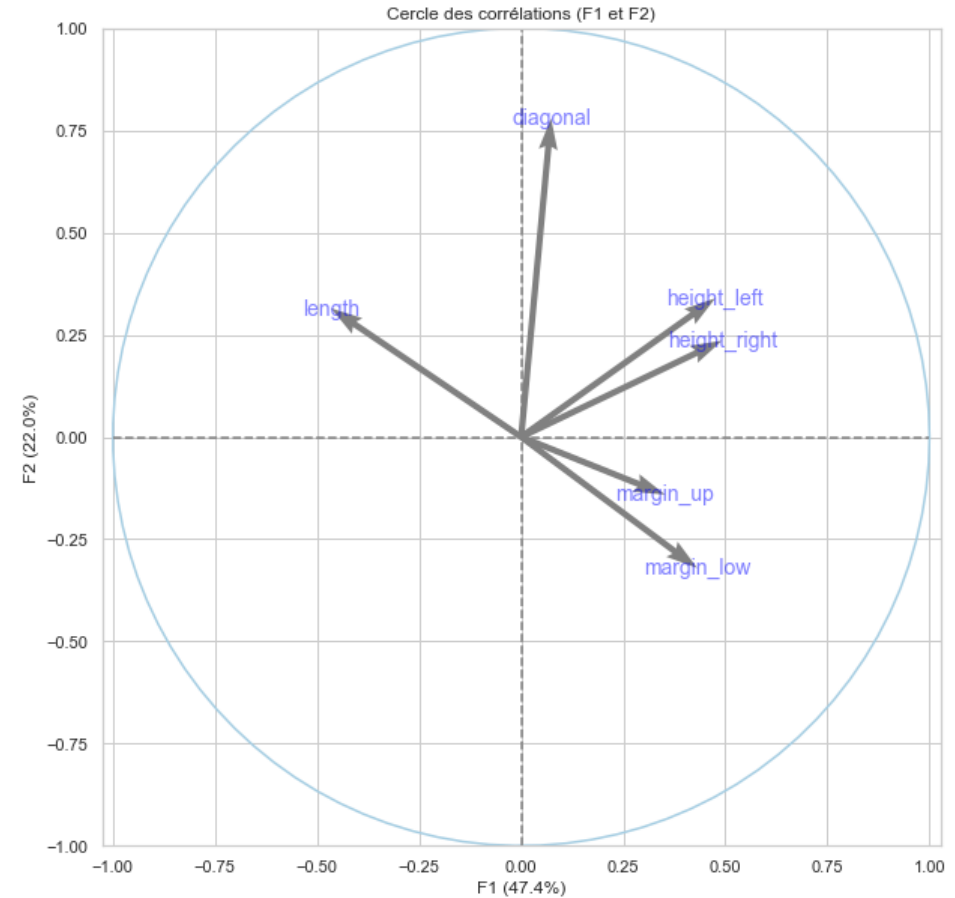


ANALYSE DES DONNÉES

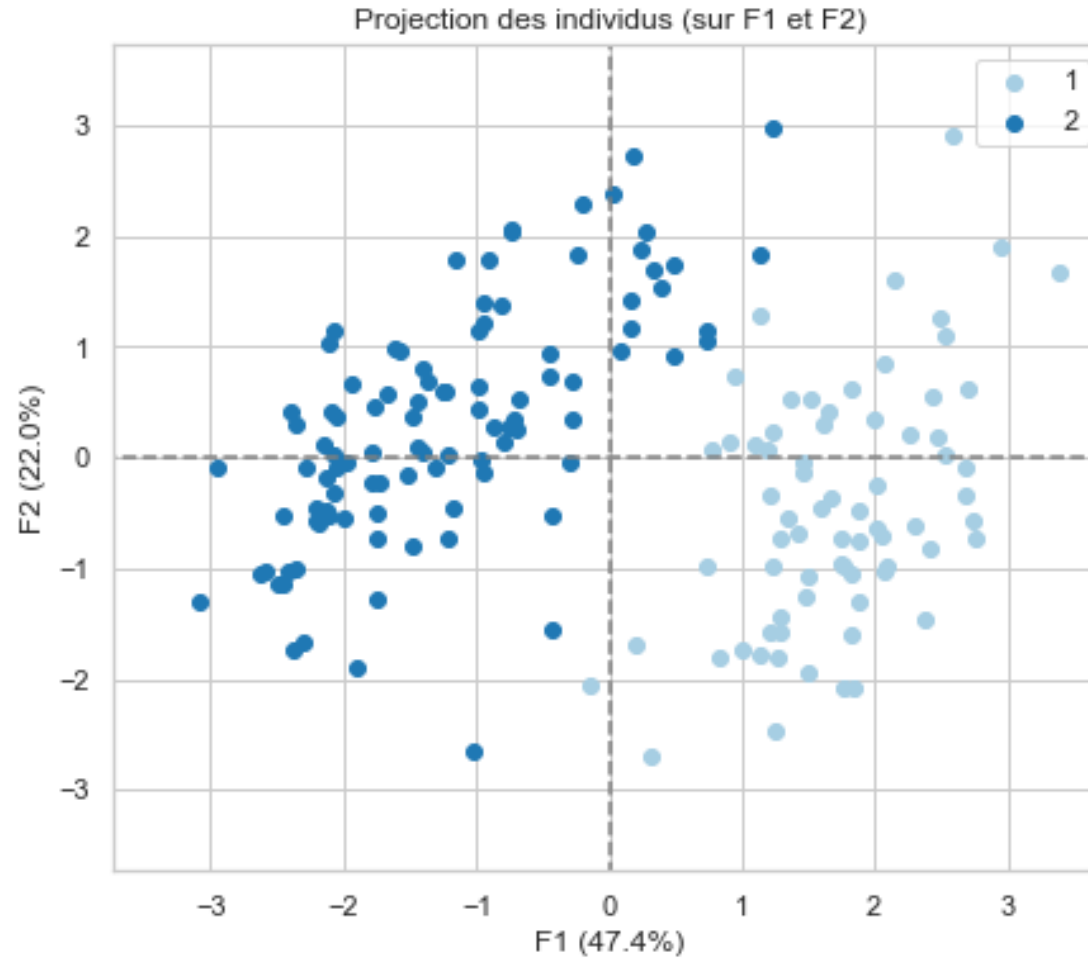
ANALYSE EN COMPOSANTES PRINCIPALES



- Cercle des corrélations :
 - F1 correspond aux informations de marge et hauteur
 - F2 correspond aux informations de longueur



ANALYSE DES DONNÉES CLASSIFICATION



MODÉLISATION DES DONNÉES

RÉGRESSION LOGISTIQUE

- Partition aléatoire du jeu de données :
 - 80% pour créer le modèle
 - 20 % pour tester le modèle
- Régression logistique et évaluation de la précision du modèle

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)
```

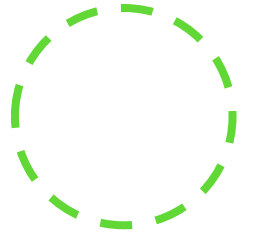
	precision	recall	f1-score	support
False	1.00	1.00	1.00	16
True	1.00	1.00	1.00	18
accuracy			1.00	34
macro avg	1.00	1.00	1.00	34
weighted avg	1.00	1.00	1.00	34

	Predicted Negative	Predicted Positive
Actual Negative	16	0
Actual Positive	0	18



MODÉLISATION DES DONNÉES

TEST DU NOUVEL ÉCHANTILLON



- Application du modèle sur les nouvelles données, calcul des probabilités d'affectation et détermination de la prédiction

	id	probalité_true	probalité_false	prédiction
0	A_1	0.054613	0.945387	False
1	A_2	0.013634	0.986366	False
2	A_3	0.032018	0.967982	False
3	A_4	0.860021	0.139979	True
4	A_5	0.995387	0.004613	True

