



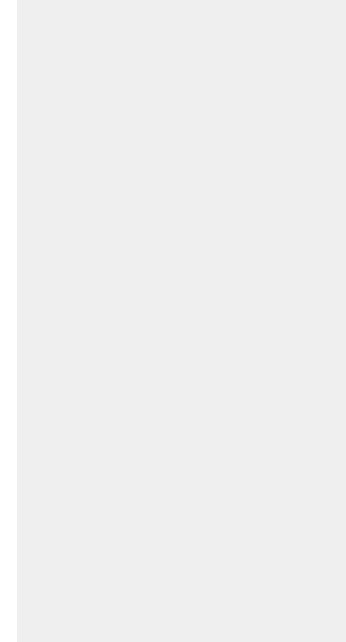
PROJET 6



DÉTECTEZ DES BILLETS FAUX

SOMMAIRE

- A. Exploration des données
- B. Analyse en composante principale
- C. Modélisation
 - A. Analyse de la modélisation
 - B. Test du programme



PRÉSENTATION DES VARIABLES



DIAGONALE

LONGUEUR

HAUTEUR

Droite

Gauche

MARGES

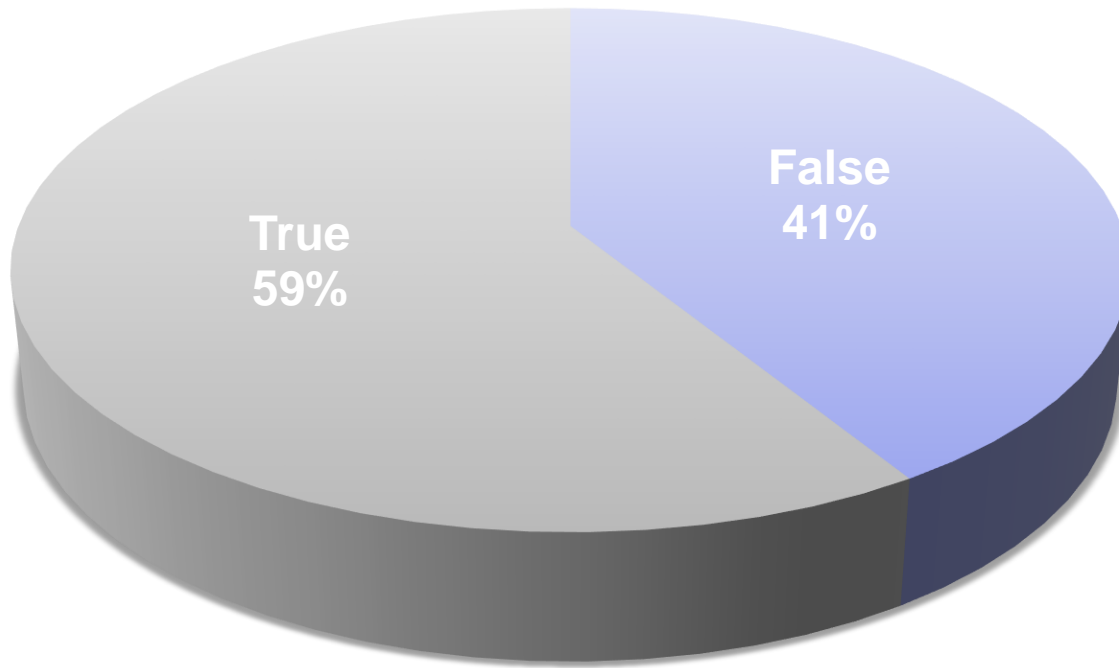
Supérieur

re

Inférieur

AUTHENTICITÉ DES BILLETS

Répartition de l'authenticité des billets



■ False ■ True

Nombre
de billets
faux

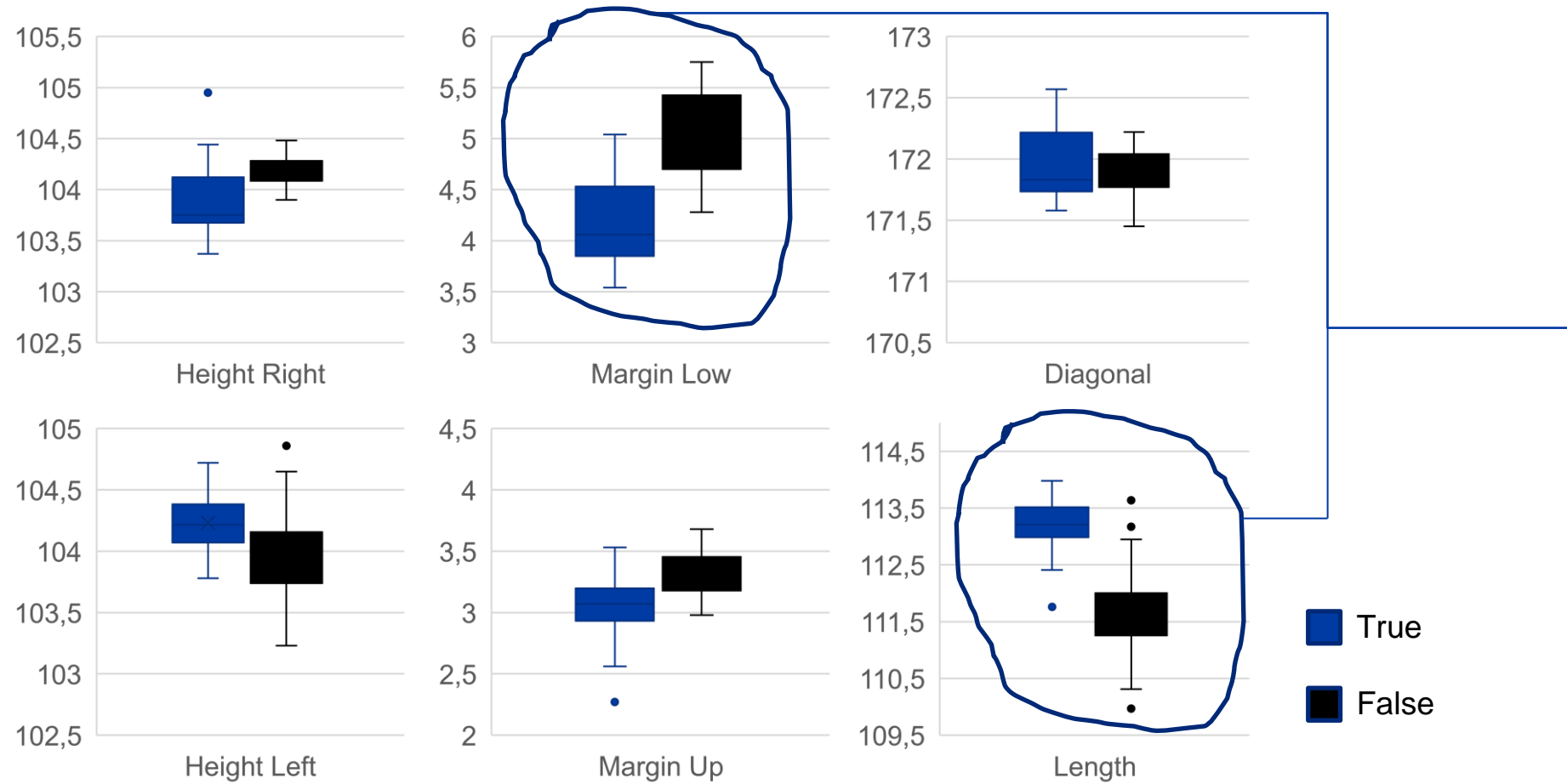
70

Nombre
de billets
vrai

100



CARACTÉRISATION DES DONNÉES



Variables les plus différentes selon l'authenticité du billets

PAIR-PLOT

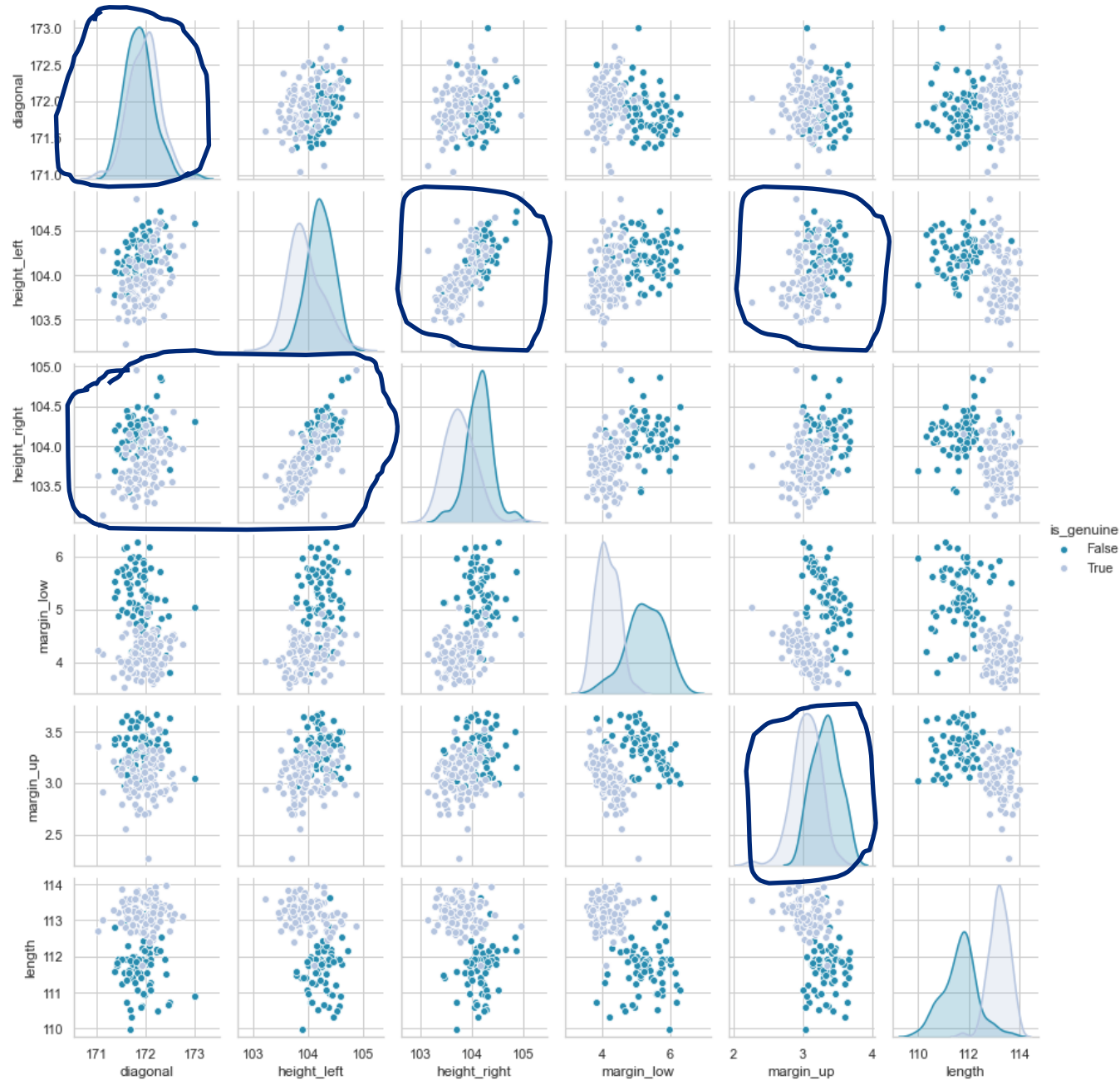
**DIAGONAL -
DIAGONAL**

**HEIGHT_RIGHT -
DIAGONAL**

**HEIGHT -
HEIGHT**

**HEIGHT_LEFT -
MARGIN
UP**

**MARGIN UP
- MARGIN
UP**



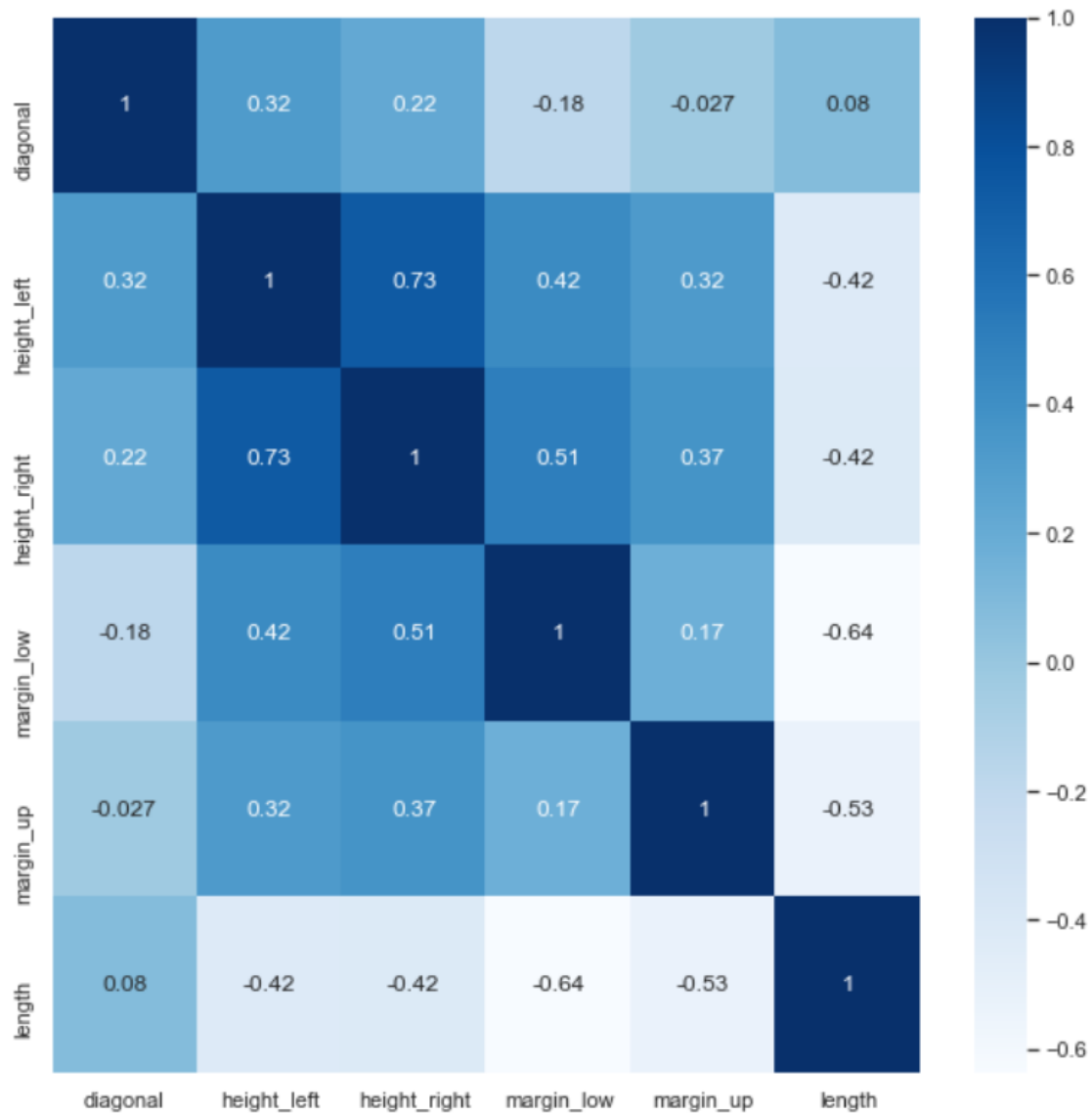


Tableau de contingence des intersections des données

PLUS FORTES
CORRÉLATIONS

HEIGHT - HEIGHT

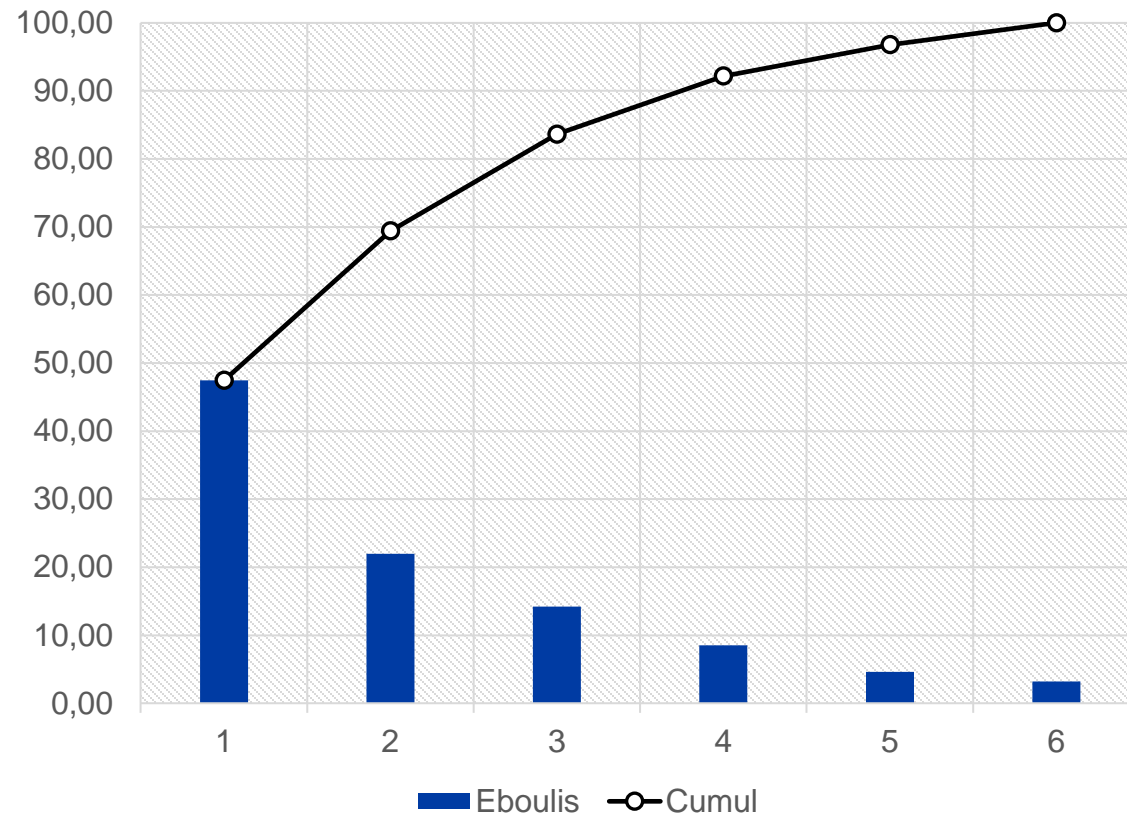
MARGIN LOW -
HEIGHT

MARGIN UP - HEIGHT



ANALYSE EN COMPOSANTE PRINCIPALE

Eboulis des valeurs propres



1^{er} Plan Factoriel

69

Pourcentage d'inertie

2^{ème} Plan Factoriel

23

Pourcentage d'inertie

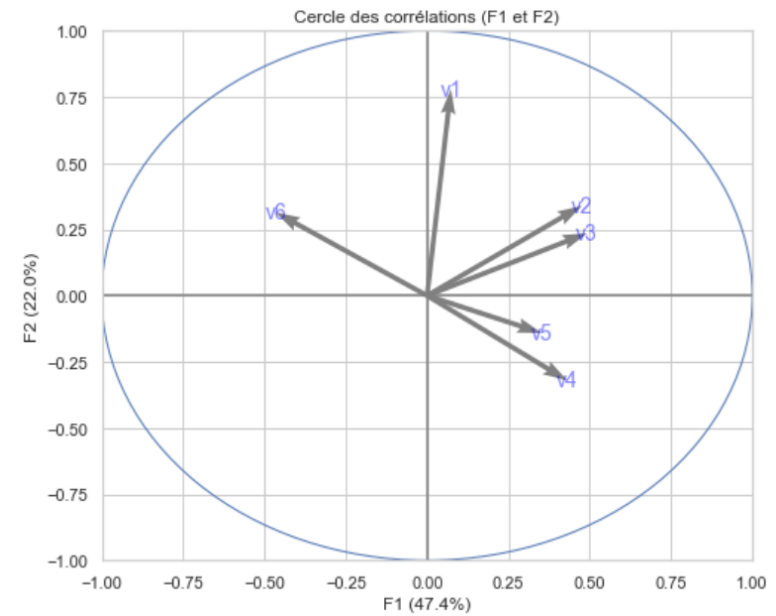


- V1** Diagonale
- V2** Hauteur gauche
- V3** Hauteur droite
- V4** Marge inférieure
- V5** Marge supérieure
- V6** Longueur

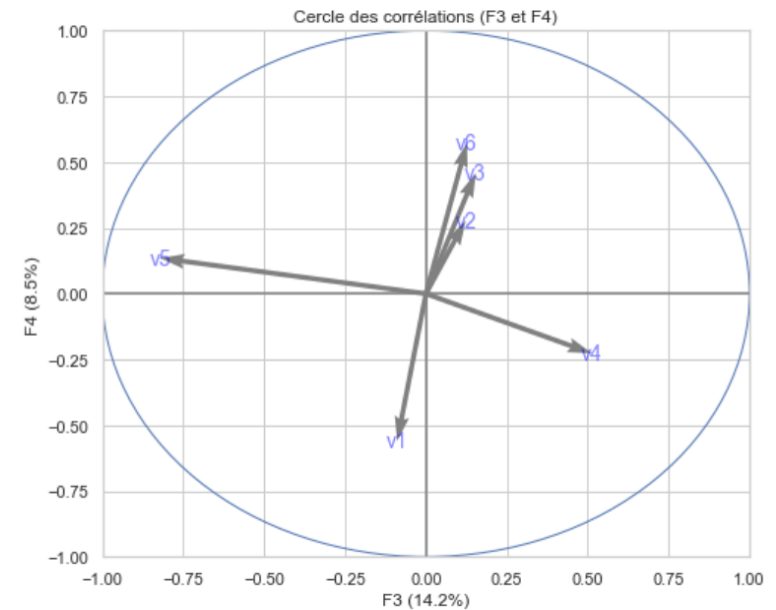
F1 Diagonales
F2 Marges

F3 Marge inférieure
F4 Longueur horizontale et verticale

1^{er} plan factoriel

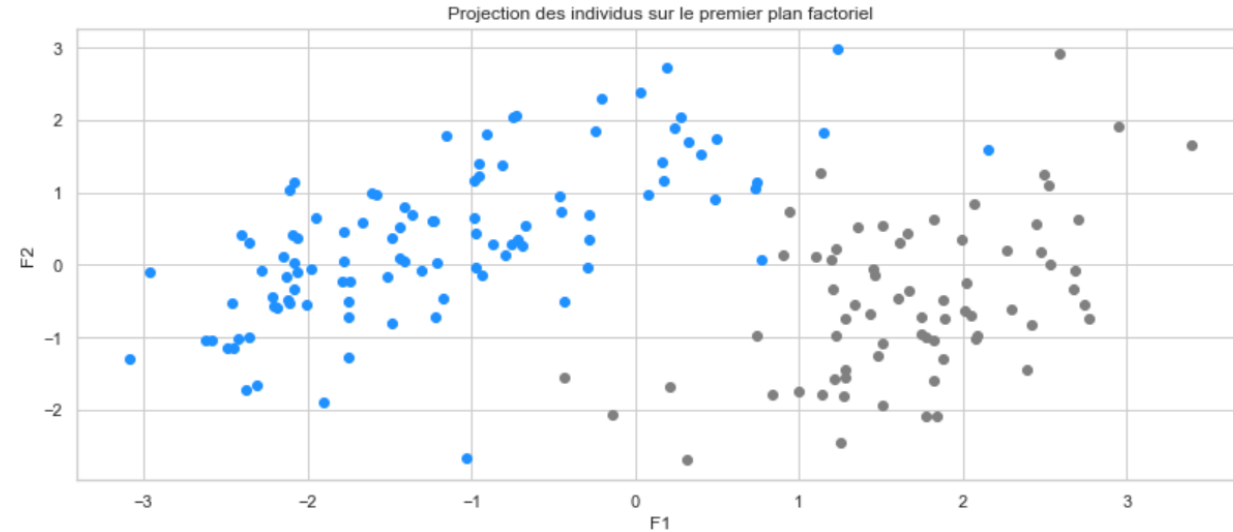


2^{ème} plan factoriel

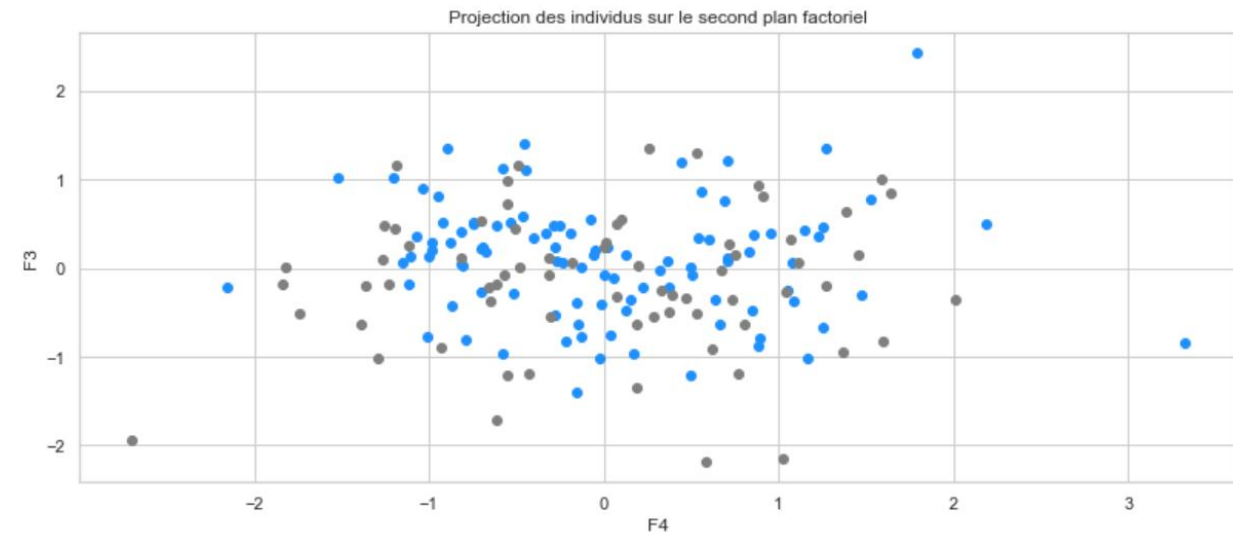


PROJECTION DES INDIVIDUS SUR LES PLANS

1^{er} Plan Factoriel

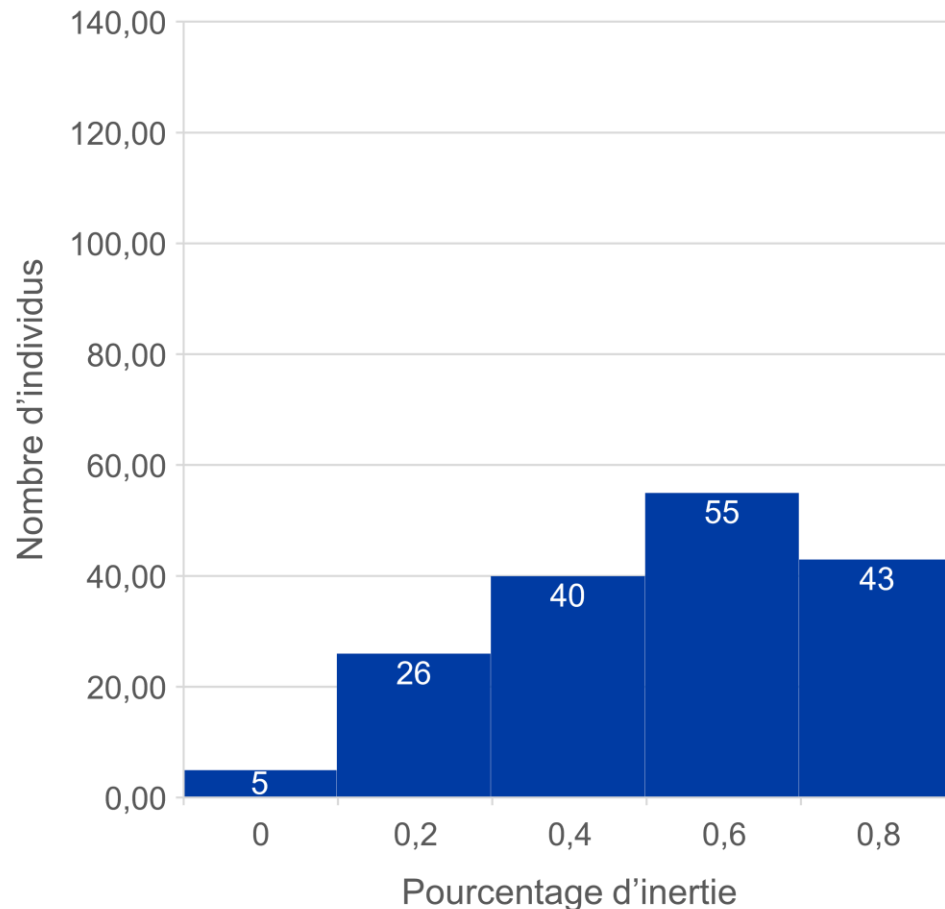


2^{ème} Plan Factoriel



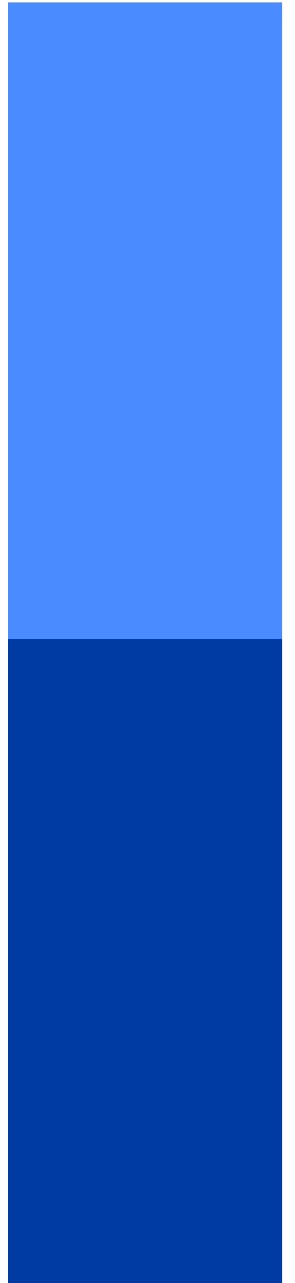
QUALITÉ DE REPRÉSENTATION DES INDIVIDUS SUR LE 1^{er} PLAN FACTORIEL AVEC UN CUMUL DES AXES

1^{er} Plan



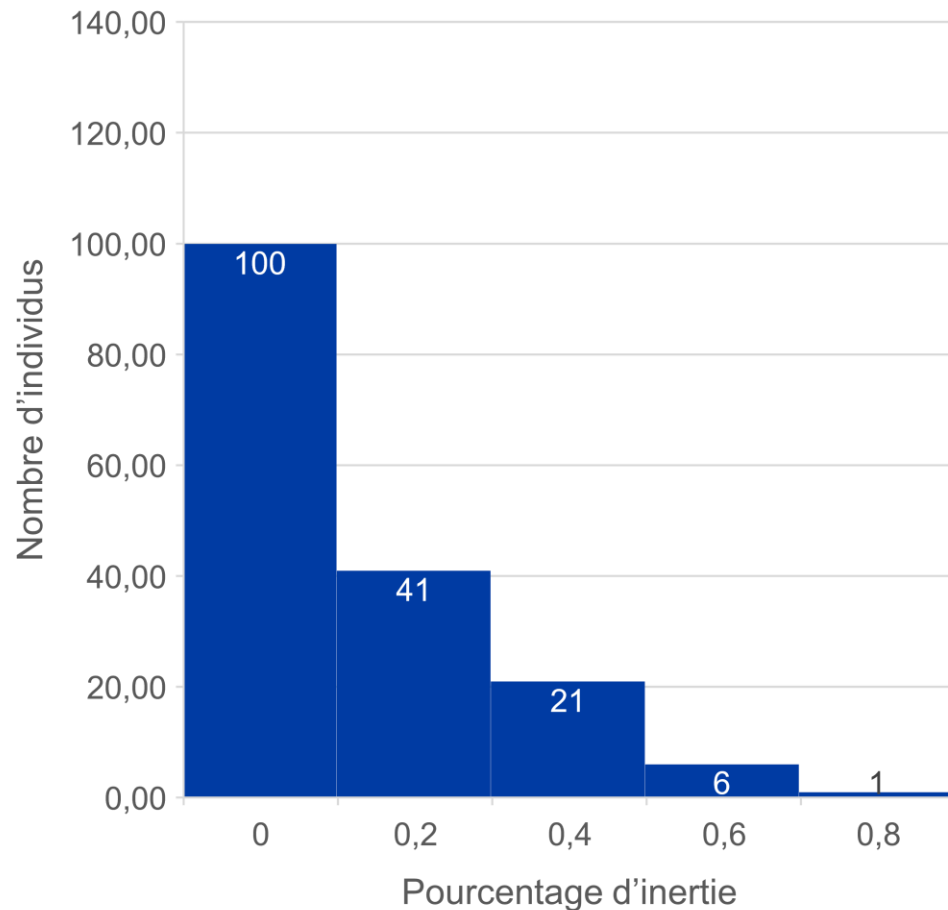
Pourcentage
d'individus avec moins
de 40 % d'inertie

18
%



QUALITÉ DE REPRÉSENTATION DES INDIVIDUS SUR LE 2^{ème} PLAN FACTORIEL AVEC UN CUMUL DES AXES

2^{ème} Plan

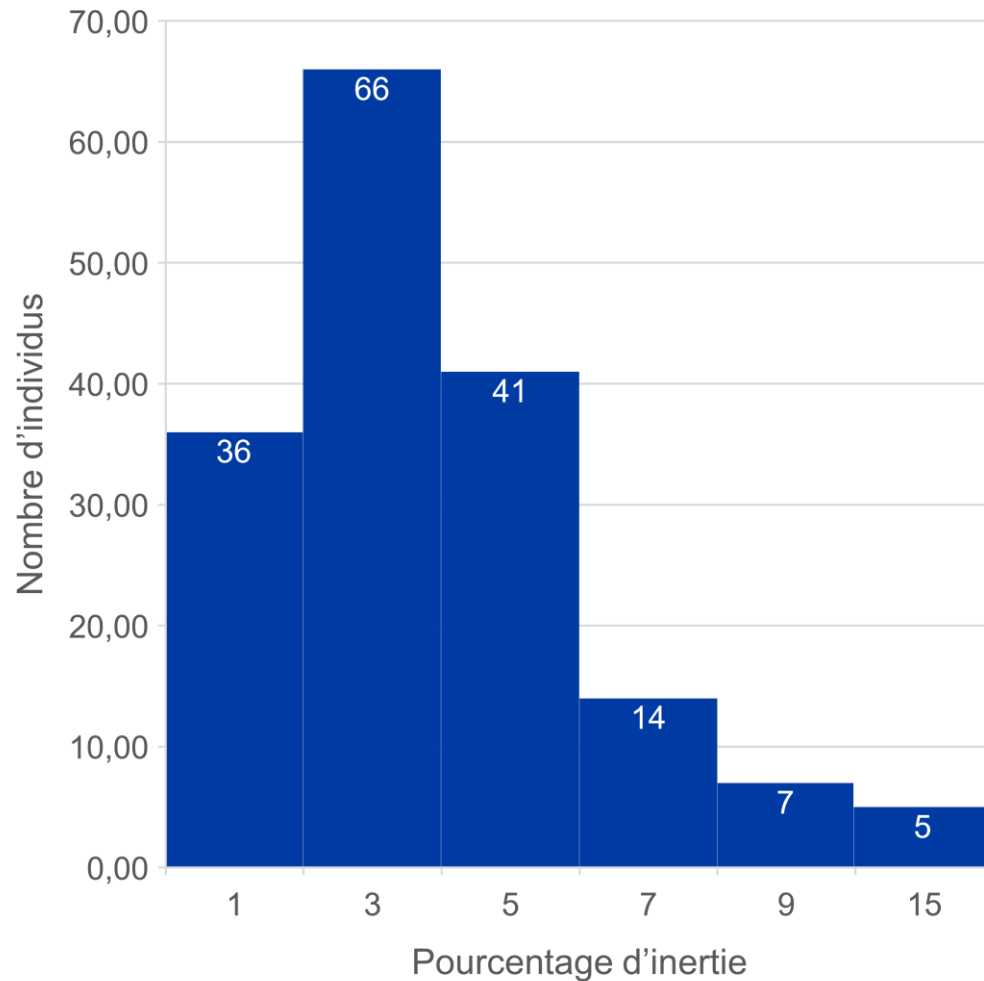


Pourcentage
d'individus avec moins
de 40 % d'inertie

83

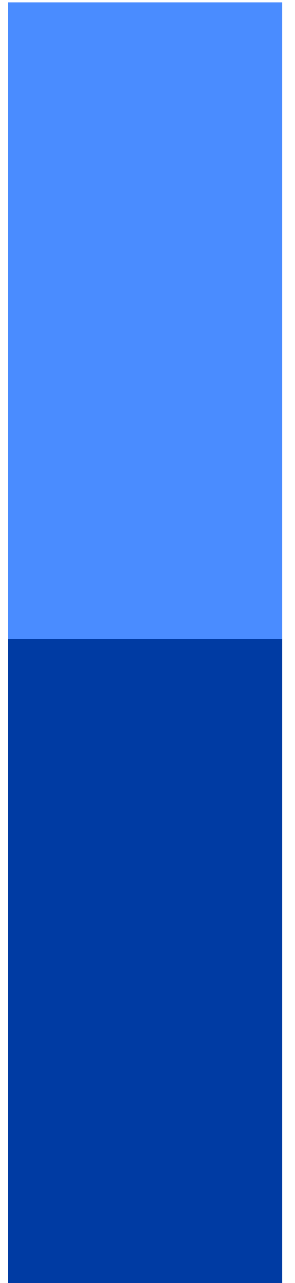
84

CONTRIBUTION DES INDIVIDUS AUX AXES



Pourcentage d'individus
contribuant à moins de 4 %
d'inertie

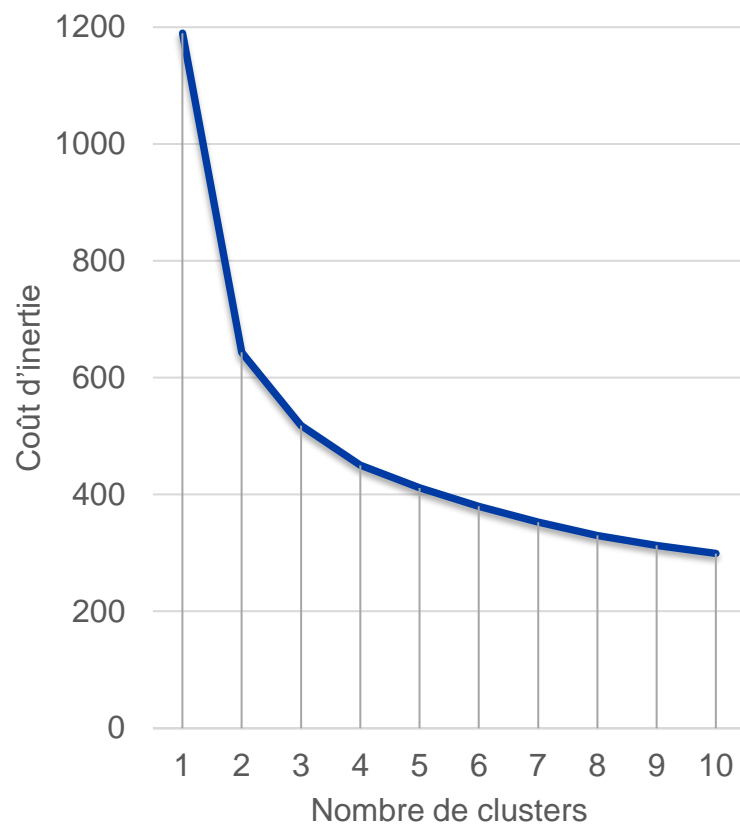
60 %



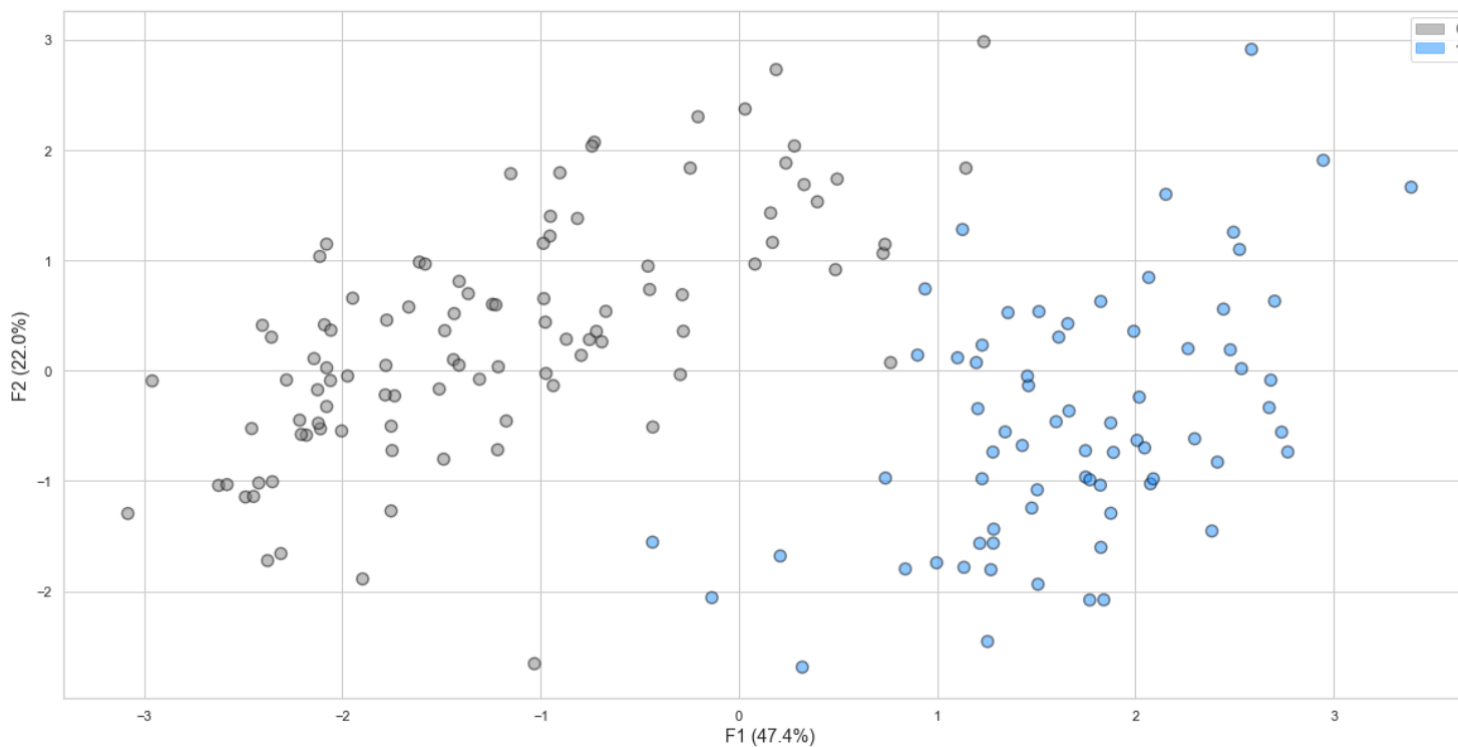


K-MEANS

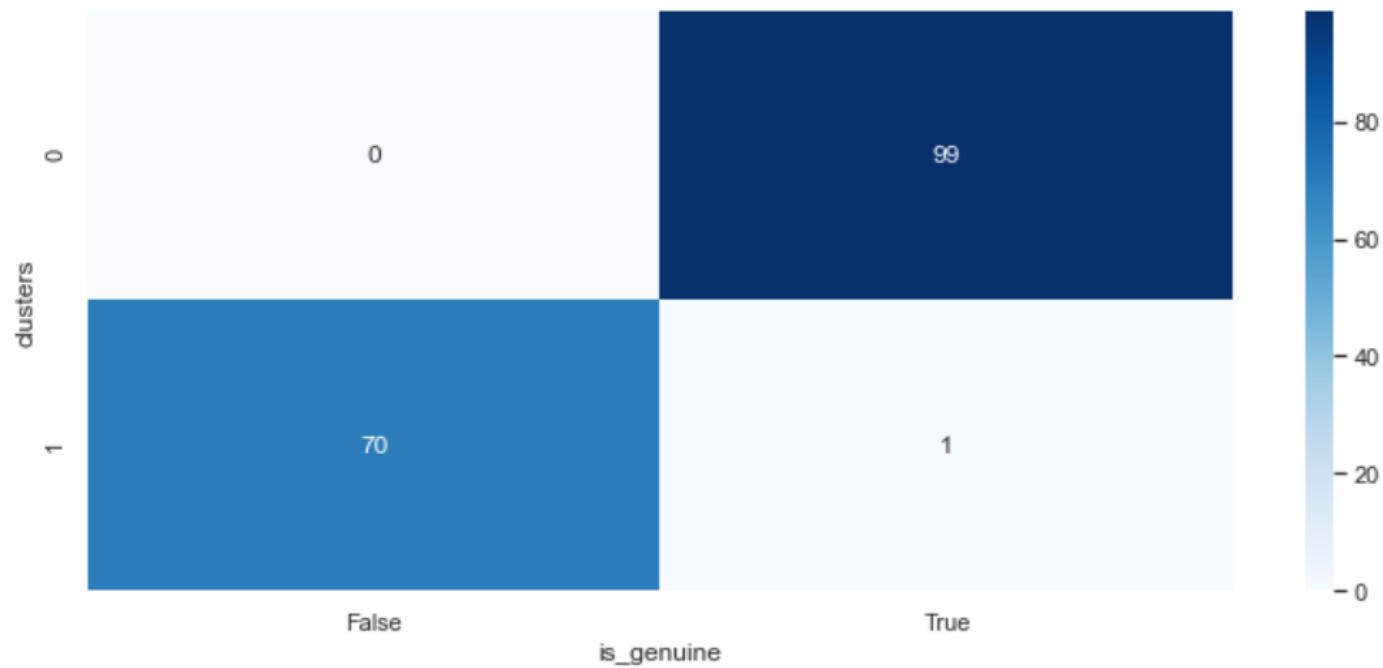
Méthode du coude




Projection en 2 clusters des 170 individus sur le 1er plan factoriel



HEATMAP



**Une seule
erreur a
été
comptabilisé**



MODÈLE PRÉDICTIF

RÉGRESSION LOGISTIQUE

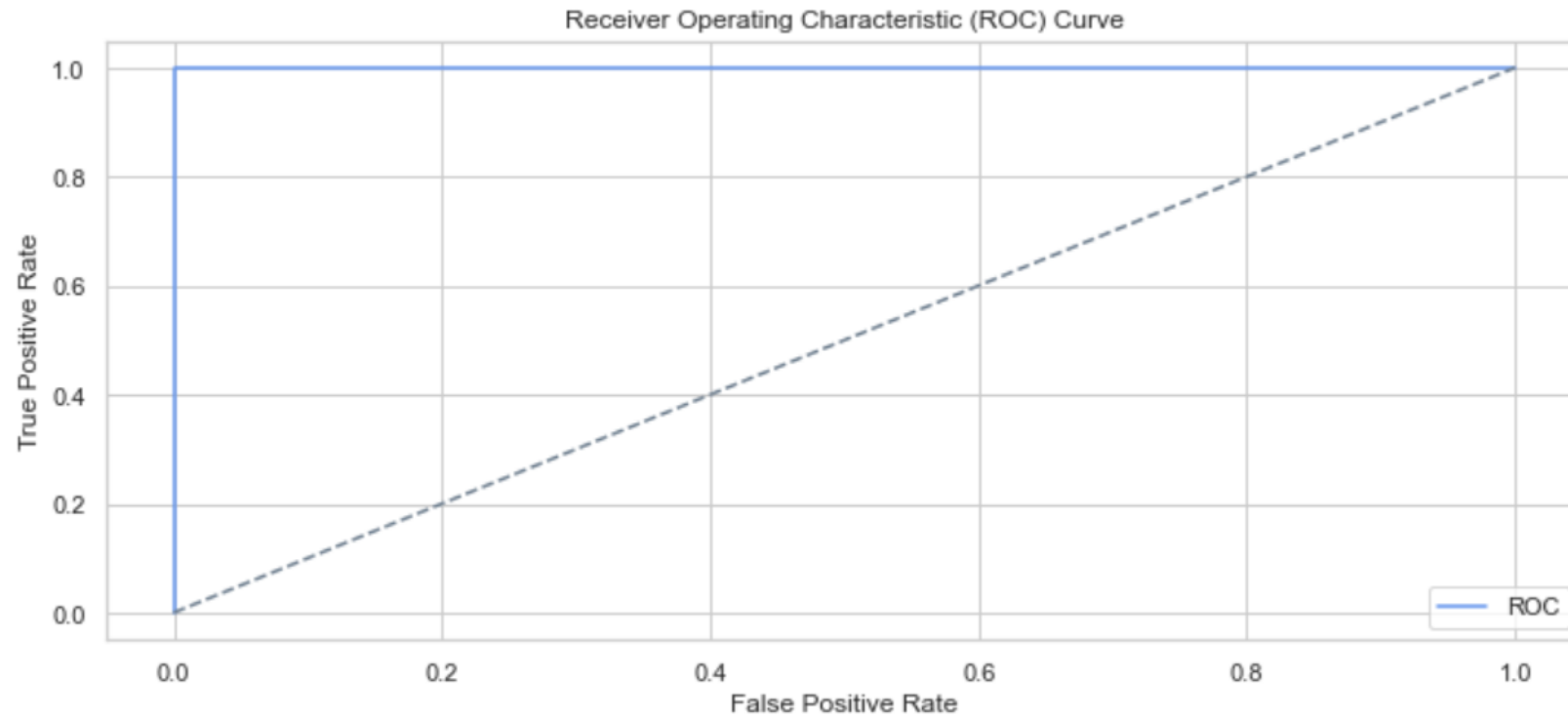
Modèle

- Modèle de régression binomiales
- Prédit la probabilité que la variable soit égale à une certaine valeur

Evaluation du modèle

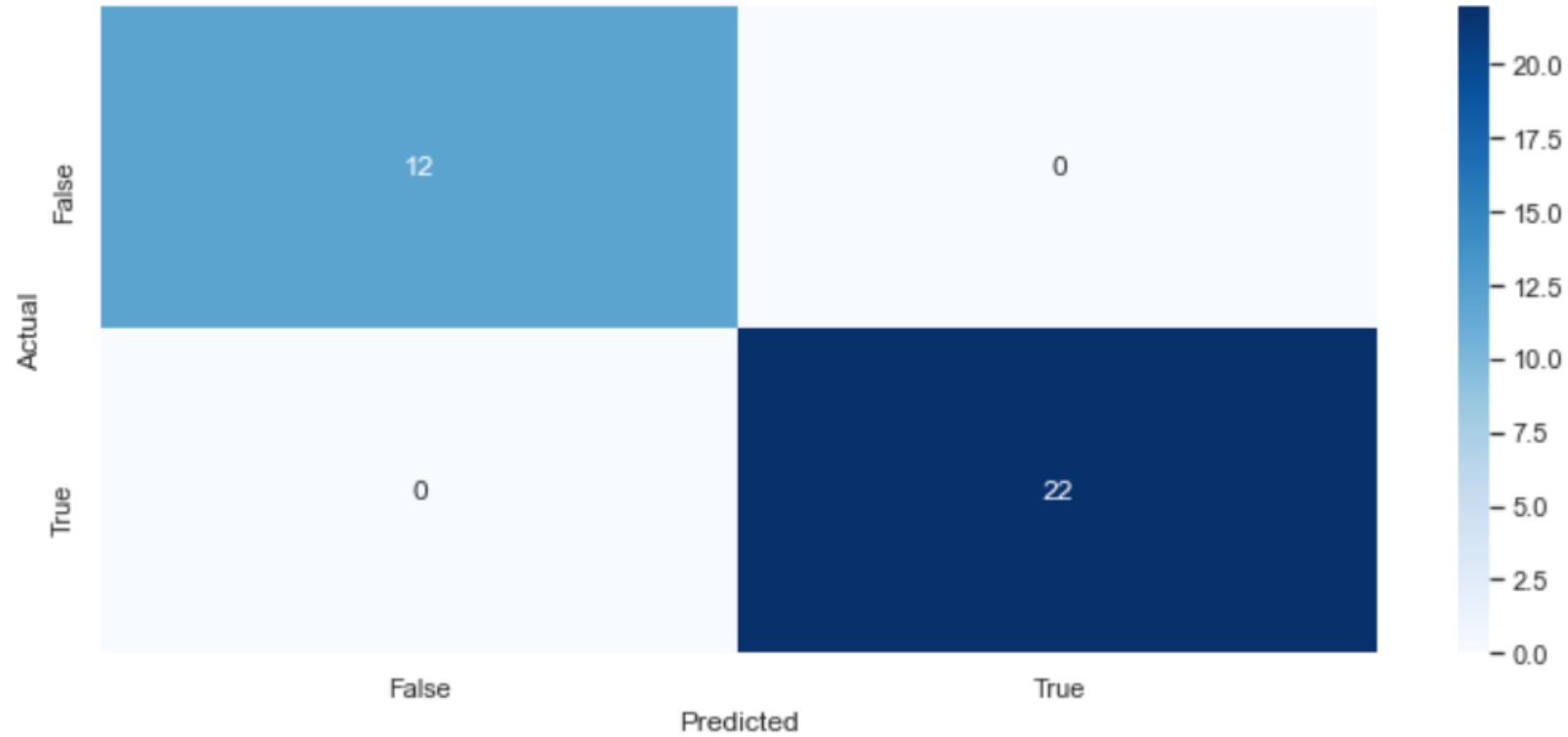
- Score du modèle
- Courbe de ROC
- Matrice de confusion
- Métriques
 - Précision
 - Recall
 - Accuracy
 - F1 Score

Courbe ROC



**La courbe
est
représentati
on parfaite**

MATRICE DE CONFUSION



Aucune erreur

MÉTRIQUES

PRÉCISION

1.00

RECALL

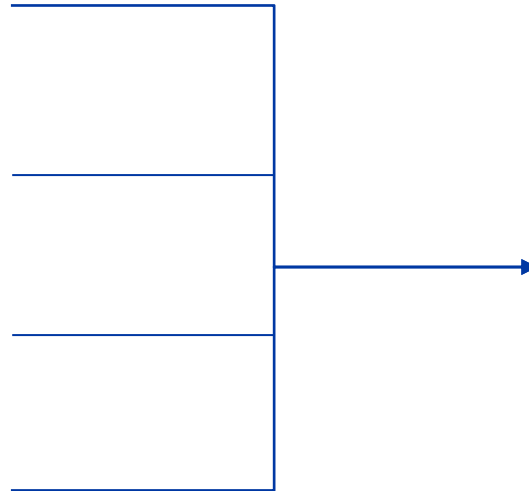
1.00

ACCURACY

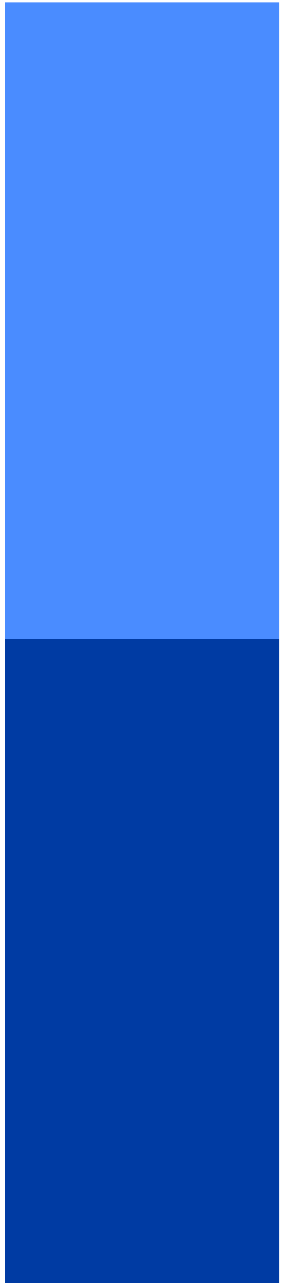
1.00

F1 SCORE

1.00



Le modèle ne
contient
aucune erreur



PRÉDICTION DES VALEURS D'EXEMPLES

	diagonal	height_left	height_right	margin_low	margin_up	length	id	predict	pourcentage
0	171.76	104.01	103.54	5.21	3.30	111.42	A_1	False	0.07
1	171.87	104.17	104.13	6.00	3.31	112.09	A_2	False	0.01
2	172.00	104.58	104.29	4.99	3.39	111.57	A_3	False	0.05
3	172.49	104.55	104.34	4.44	3.03	113.20	A_4	True	0.87
4	171.65	103.63	103.56	3.77	3.16	113.33	A_5	True	1.00

Pourcentage de billets
faux
dans le fichier exemple

60
%

Merci pour votre attention

