

Dossier de conformité — données d'entraînement (AI Act, article 10)

Squelette généré automatiquement à partir de l'analyse du jeu de données. Les blocs « à compléter par l'expert » nécessitent une revue humaine (provenance, jugement de conformité, mesures d'atténuation).

1. Identité & finalité

- **Système / modèle** : _____
- **Version du dossier** : _____
- **Date d'analyse** : _____
- **Jeu de données** : src
- **Finalité prévue** : _____

À compléter par l'expert — finalité prévue et ce que les données sont censées représenter

2. Provenance & base légale

Origine de chaque source, licences/contrats, et — pour les données personnelles — finalité initiale et base légale RGPD.

(!) **Données personnelles potentielles détectées** : name, first, last, dob — une base légale RGPD est requise pour ces colonnes.

À compléter par l'expert — provenance par source (flux séparé §16), licences, base légale RGPD

3. Composition (rempli automatiquement)

- **Volume** : 7214 lignes, 53 colonnes.
- **Période des données** (compas_screening_date) : 2013-01-01 -> 2014-12-31.
- **Variables** :

Colonne	Type
id	BIGINT
name	VARCHAR
first	VARCHAR
last	VARCHAR
compas_screening_date	DATE
sex	VARCHAR
dob	DATE
age	BIGINT
age_cat	VARCHAR
race	VARCHAR
juv_fel_count	BIGINT
decile_score	BIGINT
juv_misd_count	BIGINT

juv_other_count	BIGINT
priors_count	BIGINT
days_b_screening_arrest	BIGINT
c_jail_in	TIMESTAMP
c_jail_out	TIMESTAMP
c_case_number	VARCHAR
c_offense_date	DATE
c_arrest_date	DATE
c_days_from_compas	BIGINT
c_charge_degree	VARCHAR
c_charge_desc	VARCHAR
is_recid	BIGINT
r_case_number	VARCHAR
r_charge_degree	VARCHAR
r_days_from_arrest	BIGINT
r_offense_date	DATE
r_charge_desc	VARCHAR
r_jail_in	DATE
r_jail_out	DATE
violent_recid	VARCHAR
is_violent_recid	BIGINT
vr_case_number	VARCHAR
vr_charge_degree	VARCHAR
vr_offense_date	DATE
vr_charge_desc	VARCHAR
type_of_assessment	VARCHAR
decile_score_1	BIGINT
score_text	VARCHAR
screening_date	DATE
v_type_of_assessment	VARCHAR
v_decile_score	BIGINT
v_score_text	VARCHAR
v_screening_date	DATE

in_custody	DATE
out_custody	DATE
priors_count_1	BIGINT
start	BIGINT
end	BIGINT
event	BIGINT
two_year_recid	BIGINT

- **Attributs sensibles déclarés** : race, sex, age_cat.
- **Colonnes PII détectées** : name, first, last, dob.

À compléter par l'expert — périmètre géographique / contextuel / comportemental

4. Préparation (observations automatiques)

- **Valeurs manquantes** : days_b_screening_arrest (307) -> stratégie de traitement à documenter.

À compléter par l'expert — journal des transformations : collecte, nettoyage, étiquetage, enrichissement, agrégation

5. Qualité (rempli automatiquement)

- **Complétude** : 307 valeur(s) manquante(s) (days_b_screening_arrest).
- **Exactitude / aberrations** : age (1).

À compléter par l'expert — représentativité et pertinence vis-à-vis de la finalité

6. Biais

Analyse automatique

Caractéristique	Mesure	Cible	Verdict
Complétude	81.4 %	"e 95.0 %	à traiter
Unicité	100.0 %	"e 99.0 %	conforme
Validité	96.7 %	"e 98.0 %	à traiter

race — « score_text » : **disparate impact** (ratio) 4.83 · **parité statistique** (écart) 26.4 % (!)

_(!) Disparité **non robuste** : plus petit groupe n=18 — écart non significatif ou porté par un petit effectif.
 À ne pas interpréter comme un biais avéré._

Groupe	Effectif	Taux de « score_text »
Native American	18	33.3 %
African-American	3696	27.7 %
Caucasian	2454	11.2 %

Hispanic	637	10.5 %
Asian	32	9.4 %
Other	377	6.9 %

sex — « score_text » : **disparate impact** (ratio) 1.53 · **parité statistique** (écart) 7.2 % (!)

Groupe	Effectif	Taux de « score_text »
Male	5819	20.8 %
Female	1395	13.6 %

age_cat — « score_text » : **disparate impact** (ratio) 3.65 · **parité statistique** (écart) 21.5 % (!)

Groupe	Effectif	Taux de « score_text »
Less than 25	1529	29.6 %
25 - 45	4109	20.0 %
Greater than 45	1576	8.1 %

Jugement & mesures

À compléter par l'expert — impact sur santé/sécurité/droits fondamentaux et mesures de détection/prévention/atténuation

7. Lacunes & limites

Points détectés à examiner comme lacunes potentielles :

- [low] missing — days_b_screening_arrest
- [low] missing — c_jail_in
- [low] missing — c_jail_out
- [low] missing — c_case_number
- [low] missing — c_offense_date
- [high] missing — c_arrest_date
- [low] missing — c_days_from_compas
- [low] missing — c_charge_desc
- [high] missing — r_case_number
- [high] missing — r_charge_degree
- [high] missing — r_days_from_arrest
- [high] missing — r_offense_date
- [high] missing — r_charge_desc
- [high] missing — r_jail_in
- [high] missing — r_jail_out
- [high] missing — violent_recid
- [high] missing — vr_case_number
- [high] missing — vr_charge_degree
- [high] missing — vr_offense_date
- [high] missing — vr_charge_desc
- [low] missing — in_custody
- [low] missing — out_custody
- [low] outliers — age
- [low] outliers — priors_count
- [low] outliers — days_b_screening_arrest
- [low] outliers — c_days_from_compas
- [low] outliers — r_days_from_arrest
- [low] outliers — priors_count_1

- [low] outliers — start
- [high] pii — name
- [high] pii — first
- [high] pii — last
- [high] pii — dob
- [low] bias — race
- [medium] bias — sex
- [high] bias — age_cat

À compléter par l'expert — manques identifiés et traités ; usages hors périmètre

8. Gouvernance & traçabilité

- **Version de l'outil d'analyse** : 0 . 4 . 1
- **Empreinte du jeu (schéma + volume)** : fee3a183fbb13d6f

À compléter par l'expert — responsabilités, journal d'audit, versioning, maintenance/mise à jour