

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.

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

3. Composition (rempli automatiquement)

- **Volume** : 32561 lignes, 15 colonnes.
- **Variables** :

Colonne	Type
age	BIGINT
workclass	VARCHAR
fnlwgt	BIGINT
education	VARCHAR
education_num	BIGINT
marital_status	VARCHAR
occupation	VARCHAR
relationship	VARCHAR
race	VARCHAR
sex	VARCHAR
capital_gain	BIGINT
capital_loss	BIGINT
hours_per_week	BIGINT
native_country	VARCHAR
income	VARCHAR

• **Attributs sensibles déclarés** : age, workclass, education, marital_status, occupation, relationship, race, sex.

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

4. Préparation (observations automatiques)

- **Doublons** : 24 ligne(s) en double -> étape de dédoublonnage à tracer.

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

5. Qualité (rempli automatiquement)

- **Exactitude / aberrations** : fnlwgt (152).
- **Unicité** : 24 doublon(s).

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

6. Biais

Analyse automatique

Caractéristique	Mesure	Cible	Verdict
Complétude	100.0 %	"e 95.0 %	conforme
Unicité	99.9 %	"e 99.0 %	conforme
Validité	97.7 %	"e 98.0 %	à traiter

workclass — « income » : **disparate impact** (ratio) 5.36 · **parité statistique** (écart) 55.7 % (!)

Groupe	Effectif	Taux de « income »
Self-emp-inc	1116	55.7 %
Federal-gov	960	38.6 %
Local-gov	2093	29.5 %
Self-emp-not-inc	2541	28.5 %
State-gov	1298	27.2 %
Private	22696	21.9 %
?	1836	10.4 %
Without-pay	14	0.0 %
Never-worked	7	0.0 %

education — « income » : **disparate impact** (ratio) 20.75 · **parité statistique** (écart) 74.1 % (!)

Groupe	Effectif	Taux de « income »
Doctorate	413	74.1 %
Prof-school	576	73.4 %

Masters	1723	55.7 %
Bachelors	5355	41.5 %
Assoc-voc	1382	26.1 %
Assoc-acdm	1067	24.8 %
Some-college	7291	19.0 %
HS-grad	10501	16.0 %
12th	433	7.6 %
10th	933	6.6 %
7th-8th	646	6.2 %
9th	514	5.3 %
11th	1175	5.1 %
5th-6th	333	4.8 %
1st-4th	168	3.6 %
Preschool	51	0.0 %

marital_status — « income » : **disparate impact** (ratio) 9.72 · **parité statistique** (écart) 40.1 % (!)

Groupe	Effectif	Taux de « income »
Married-civ-spouse	14976	44.7 %
Married-AF-spouse	23	43.5 %
Divorced	4443	10.4 %
Widowed	993	8.6 %
Married-spouse-absent	418	8.1 %
Separated	1025	6.4 %
Never-married	10683	4.6 %

occupation — « income » : **disparate impact** (ratio) 72.12 · **parité statistique** (écart) 47.7 % (!)

Groupe	Effectif	Taux de « income »
Exec-managerial	4066	48.4 %
Prof-specialty	4140	44.9 %
Protective-serv	649	32.5 %
Tech-support	928	30.5 %
Sales	3650	26.9 %
Craft-repair	4099	22.7 %
Transport-moving	1597	20.0 %

Adm-clerical	3770	13.4 %
Machine-op-inspct	2002	12.5 %
Farming-fishing	994	11.6 %
Armed-Forces	9	11.1 %
?	1843	10.4 %
Handlers-cleaners	1370	6.3 %
Other-service	3295	4.2 %
Priv-house-serv	149	0.7 %

relationship — « income » : **disparate impact** (ratio) 35.94 · **parité statistique** (écart) 46.2 % (!)

Groupe	Effectif	Taux de « income »
Wife	1568	47.5 %
Husband	13193	44.9 %
Not-in-family	8305	10.3 %
Unmarried	3446	6.3 %
Other-relative	981	3.8 %
Own-child	5068	1.3 %

race — « income » : **disparate impact** (ratio) 2.88 · **parité statistique** (écart) 17.3 % (!)

Groupe	Effectif	Taux de « income »
Asian-Pac-Islander	1039	26.6 %
White	27816	25.6 %
Black	3124	12.4 %
Amer-Indian-Eskimo	311	11.6 %
Other	271	9.2 %

sex — « income » : **disparate impact** (ratio) 2.79 · **parité statistique** (écart) 19.6 % (!)

Groupe	Effectif	Taux de « income »
Male	21790	30.6 %
Female	10771	10.9 %

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 :

- [medium] duplicates
- [low] outliers — fnlwgt
- [low] outliers — hours_per_week
- [high] bias — workclass
- [high] bias — education
- [high] bias — marital_status
- [high] bias — occupation
- [high] bias — relationship
- [high] bias — race
- [high] bias — sex

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)** : c09f6f88fcbcff58

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