Compliance dossier — training data (AI Act, Article 10)

Automatically generated skeleton from the dataset analysis. The "to be completed by the expert" blocks require human review (provenance, compliance judgment, mitigation measures).

1. Identity & purpose

- **System / model**: _____
- **Dossier version**: _____
- **Date**: _____
- **Dataset**: `src`
- **Intended purpose**: _____

To be completed by the expert — intended purpose and what the data is meant to represent

2. Provenance & legal basis

Origin of each source, licences/contracts, and — for personal data — initial purpose and GDPR legal basis.

(!) **Potential personal data detected**: `telephone` — a GDPR legal basis is required for these columns.

To be completed by the expert — provenance per source (separate flow §16), licences, GDPR legal basis

3. Composition (filled automatically)

- **Volume**: 1000 rows, 21 columns.
- **Variables**:

Column	Type
status	VARCHAR
duration	BIGINT
credit_history	VARCHAR
purpose	VARCHAR
amount	BIGINT
savings	VARCHAR
employment_duration	VARCHAR
installment_rate	BIGINT
personal_status_sex	VARCHAR
other_debtors	VARCHAR
present_residence	BIGINT
property	VARCHAR
age	BIGINT
other_installment_plans	VARCHAR

housing	VARCHAR
number_credits	BIGINT
job	VARCHAR
people_liable	BIGINT
telephone	BOOLEAN
foreign_worker	BOOLEAN
credit_risk	BIGINT

- **Declared sensitive attributes**: `status`, `credit_history`, `purpose`, `savings`, `employment_duration`, `personal_status_sex`, `other_debtors`, `property`, `age`, `other_installment_plans`, `housing`, `job`, `foreign_worker`.
- **Detected PII columns**: `telephone`.

To be completed by the expert — geographic / contextual / behavioural scope

4. Preparation (automatic observations)

To be completed by the expert — transformation log: collection, cleaning, labelling, enrichment, aggregation

5. Quality (filled automatically)

- **Accuracy / outliers**: `duration` (1).

To be completed by the expert — representativeness and relevance to the purpose

6. Bias

Automatic analysis

status — disparity of the « credit_risk » rate: ratio 1.74 (!)

Group	Count	« credit_risk » rate
no checking account	394	88.3 %
... >= 200 DM / salary for at least 1 year	63	77.8 %
0 <= ... < 200 DM	269	61.0 %
... < 100 DM	274	50.7 %

credit_history — disparity of the « credit_risk » rate: ratio 2.21 (!)

Group	Count	« credit_risk » rate
critical account/other credits existing	293	82.9 %
delay in paying off in the past	88	68.2 %
existing credits paid back duly till now	530	68.1 %
all credits at this bank paid back duly	49	42.9 %

no credits taken/all credits paid back duly	40	37.5 %

purpose — disparity of the « credit_risk » rate: ratio 1.59 (!)

Group	Count	« credit_risk » rate
business	9	88.9 %
car (used)	103	83.5 %
domestic appliances	280	77.9 %
radio/television	181	68.0 %
repairs	12	66.7 %
others	97	64.9 %
education	22	63.6 %
car (new)	234	62.0 %
furniture/equipment	12	58.3 %
retraining	50	56.0 %

savings — disparity of the « credit_risk » rate: ratio 1.37 (!)

Group	Count	« credit_risk » rate
... >= 1000 DM	48	87.5 %
500 <= ... < 1000 DM	63	82.5 %
unknown/no savings account	183	82.5 %
100 <= ... < 500 DM	103	67.0 %
... < 100 DM	603	64.0 %

employment_duration — disparity of the « credit_risk » rate: ratio 1.31 (!)

Group	Count	« credit_risk » rate
4 <= ... < 7 years	174	77.6 %
... >= 7 years	253	74.7 %
1 <= ... < 4 years	339	69.3 %
unemployed	62	62.9 %
... < 1 year	172	59.3 %

personal_status_sex — disparity of the « credit_risk » rate: ratio 1.22 (!)

Group	Count	« credit_risk » rate
male : single	548	73.4 %

male : married/widowed	92	72.8 %
female : divorced/separated/married	310	64.8 %
male : divorced/separated	50	60.0 %

other_debtors — disparity of the « credit_risk » rate: ratio 1.44 (!)

Group	Count	« credit_risk » rate
guarantor	52	80.8 %
none	907	70.0 %
co-applicant	41	56.1 %

property — disparity of the « credit_risk » rate: ratio 1.39 (!)

Group	Count	« credit_risk » rate
real estate	282	78.7 %
building society savings agreement/ life insurance	232	69.4 %
car or other	332	69.3 %
unknown/no property	154	56.5 %

other_installment_plans — disparity of the « credit_risk » rate: ratio 1.23 (!)

Group	Count	« credit_risk » rate
none	814	72.5 %
stores	47	59.6 %
bank	139	59.0 %

housing — disparity of the « credit_risk » rate: ratio 1.25 (!)

Group	Count	« credit_risk » rate
own	713	73.9 %
rent	179	60.9 %
for free	108	59.3 %

job — disparity of the « credit_risk » rate: ratio 1.10 (!)

Group	Count	« credit_risk » rate
unskilled - resident	200	72.0 %
skilled employee/official	630	70.5 %
unemployed/unskilled - non-resident	22	68.2 %

management/self-employed/highly qualified employee/officer	148	65.5 %

foreign_worker — disparity of the « credit_risk » rate: ratio 1.29 (!)

Group	Count	« credit_risk » rate
false	37	89.2 %
true	963	69.3 %

Judgment & measures

To be completed by the expert — impact on health/safety/fundamental rights and detection/prevention/mitigation measures

7. Gaps & limitations

Detected points to examine as potential gaps:

- [low] outliers — `duration`
- [low] outliers — `amount`
- [high] pii — `telephone`
- [medium] bias — `status`
- [high] bias — `credit_history`
- [medium] bias — `purpose`

To be completed by the expert — identified and addressed gaps; out-of-scope uses

8. Governance & traceability

- **Analysis tool version**: `0.2.0`

To be completed by the expert — responsibilities, audit log, versioning, maintenance/updates