# Numerische Merkmale – Statistische Übersicht & Plausibilität (nur 148.025 gelabelte Transaktionen, davon 4.656 Fraud)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Merkmal TA**  **(Fraud)** | **Minimum**  **Maximum** | **Durchschnitt**  **Median** | **Fehlende Werte** | **Imputation** | **Auffälligkeiten**? | **Analyse-Vorschlag** | **Ergebnis** |
| total\_amount | 0.00 (0,00)  1.932 (1.722) | 98,51 (99,96)  62,55 (62,87) | 0 |  | Ja  Minimum = 0 ? | Transaktion überprüfen | OK |
| transaction\_duration | 3,66s (5,59s)  1.245s (796s) | 77,81s (83,74s)  53,21s (58,63s) | 0 |  | Ja  Maximum = 1.245s |  | OK |
| customer\_feedback | 1,0 (2,0)  10,0 (10,0) | 9,33 (9,8)  10 (10) | 137.231 | ???? | meisten Werte fehlen  Fast nur 10-Bewertung | Imputation prüfen |  |
| damage | 0,00 (0,03)  110,32 (110,32 | 0,24 (7,55)  0,00 (5,24) | 0 |  | Ja  0,03 Schaden? | Minimum und Maximum ansehen | OK |
| camera\_certainty | camera\_certainty in [0, 1] | 0 ≤ certainty ≤ 1 | nein |  | Nein |  | OK |
| n\_lines | 0 (0)  219 (135) | 10,6 (11,4)  7,0 (7,0) | 0 |  | Ja Minimum = 0 ? | Transaktion überprüfen | OK |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Merkmal**  **Produkte** | **Minimum**  **Maximum** | **Durchschnitt**  **Median** | **Fehlende Werte** | **Imputation** | **Auffälligkeiten**? | **Analyse-Vorschlag** | **Ergebnis** |
| Price | 0,19 €  589,49 € | 6,71 €  3,39 € | 0 |  | nein |  |  |
| popularity | 0,00  0,193 | 0,09  0,11 | 0 |  | nein |  |  |
| Weight | 0,1 kg  7,85 kg | 0,99 kg  0,75 kg | 2.505 | Nein? | Betrifft Artikel verkauft nach Gewicht 🡪 Differenz |  |  |

# 2. Kategorische Merkmale – Übersicht & Auffälligkeiten (nur gelabelte Transaktionen)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Merkmal | Anzahl Ausprägungen | Häufigste Werte | Auffälligkeiten? | Analyse-Vorschlag | Ergebnis |
| label | Normal  FRAUD | NORMAL, FRAUD | Nein |  | OK |
| payment\_medium | Credit Card  Cash | Credit Card | Ja | FRAUD-Fälle 🡪 mehr Cash | Relevantes Merkmal |
| Land | 7 | NRW: 7 | Nein |  | OK |
| category | 14 | Fruit\_Veg: 2.155  Bakery : 131 | Nein |  | OK |
| urbanization | City  Towns  Rurals | 8  7  3 | Nein |  | OK |
| age\_restricted | 2 | True, False | Nein |  | OK |
| sold\_by\_weight | False:  Treu: | 5.965  2.155 | Ja  2.505 |  | OK |

# 3. Weitere Plausibilitätschecks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prüfung | Logik | Auffälligkeiten? | Empfohlene Maßnahmen | Ergebnis |
| Liegt jeder timestamp (line) zwischen transaction\_start und transaction\_end? | timestamp ∈ [start, end] | nein |  | OK |
| Summe sales\_price = total\_amount? | groupby(transaction\_id).sum(sales\_price) ≈ total\_amount | Abweichungen möglich | Stornos oder Discounts prüfen | Abweichungen wegen Stornos |
| sales\_price ≈ Menge × price | sales\_price ≈ pieces\_or\_weight × base\_price | Teilweise große Differenz | Rabatte, Stornos oder Messfehler prüfen | Abweichung entspricht häufig damage |
| damage > 0 nur bei label = FRAUD | damage > 0 → label = FRAUD | nein |  | OK |