Übung zu Metriken der Datenqualität

Markieren Sie in der folgenden Tabelle mit Kundendaten eines Versicherungs-unternehmens diejenigen Zellen mangelnder Datenqualität in den Dimensionen Fehlerfreiheit, Einheitlichkeit und Vollständigkeit. Benennen Sie je Spalte die betroffenen Qualitätsdimensionen. Berechnen Sie je Qualitätsdimension Fehlerprozente in Bezug auf jeweils eine Spalte und 10 Beispieldatensätze.

Hinweise: Weisen Sie einen Fehler nur einer Dimension zu. Treffen Sie je Dimension sinnvolle Vorgaben. Bezugsdatum für das Alter ist Geburtsdatum 31.12.2022.

Erreichbare Punktzahl: 14

**1 2 3 4 5 6**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| KundenNr | Alter in Jahren | Geburtsdatum | Geschlecht | Adresse | Einkommen jährlich |
| 43 | 73 | * **V** | w | Europaring 46, 51109 Köln | 96000 |
| 44 | 14 J. **E** | 02.01.1990 **F** | weiblich **E** | Willy-Brandt-Platz 3, 50679 Köln | 34000 |
| 48 | 72 | 12.07.1950 | m | Europaring 61, 51109 Koeln **E** | 36750 |
| 48 **F** | 22 | 18.02.2000 | ja **F** | Venloer Straße 419-421, 50825 Köln | 67321,21 **E** |
| 49 | 242 | 18.02.1780 **F** | w | Venloer Straße 123, 50825 Köln | **V** |
| 50 | 23 | ? **V** | m | Hebammenpraxis Kugelrund, Köln **F** | **V** |
| 51 | 87 | **V** | d | Wiener Platz 2a, 51065 Köln | 10000 bis 25000 **E** |
| 52 | 45 | 1977-02-12 **E** | m | Wiener Platz 1, Köln **F** | **V** |
| 96 | 1 | 30.08.2021 | m | **V** | 0 |
| 147 | 23 | 12.08.99 **E** | m | **V** | 3,5 Mio **E** |

Notieren Sie hier Ihre Antwort:

Im Folgenden wird auf die laufende Nummer der Spalten, wie oben markiert, Bezug genommen. Fehler sind in der Tabelle oben durch Buchstaben markiert. Kunden jünger 19 und älter 120 Jahren Jahren werden als nicht geschäftsfähig angesehen. Da Alter als berechnetes Attribut interpretiert wird, werden Fehler beim Geburtsdatum markiert.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DQ Dimension | 1 | 2 | 3 | 4 | 5 | 6 |
| Fehlerfreiheit | 10% | 0% | 20% | 10% | 20% | 0% |
| Einheitlichkeit | 0% | 10% | 20% | 10% | 10% | 30% |
| Vollständigkeit | 0% | 0% | 30% | 0% | 20% | 30% |