

# 2

Sicherstellen der Datenqualität

## Teil 2 der Abschlussprüfung

### Allgemeine Korrekturhinweise

Die Lösungs- und Bewertungshinweise zu den einzelnen Handlungsschritten sind als Korrekturhilfen zu verstehen und erheben nicht in jedem Fall Anspruch auf Vollständigkeit und Ausschließlichkeit. Neben hier beispielhaft angeführten Lösungsmöglichkeiten sind auch andere sach- und fachgerechte Lösungsalternativen bzw. Darstellungsformen mit der vorgesehenen Punktzahl zu bewerten. Der Bewertungsspielraum des Korrektors (z. B. hinsichtlich der Berücksichtigung regionaler oder branchenspezifischer Gegebenheiten) bleibt unberührt.

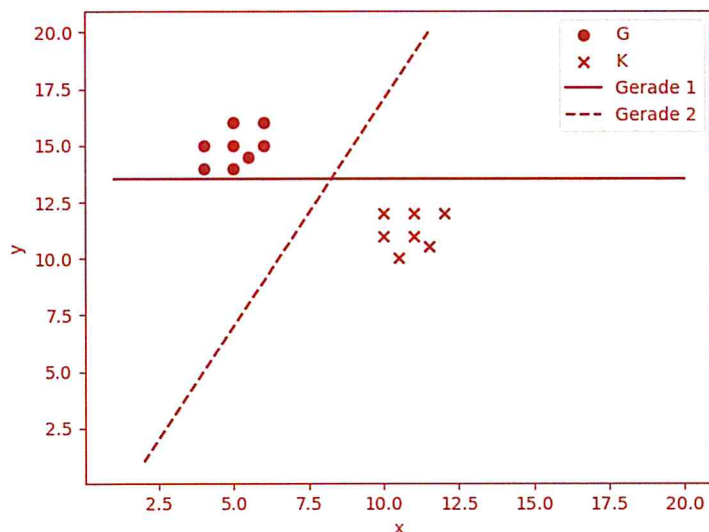
Zu beachten ist die unterschiedliche Dimension der Aufgabenstellung (nennen – erklären – beschreiben – erläutern usw.).

Für die Bewertung gilt folgender Punkte-Noten-Schlüssel:

Note 1 =	100 – 92 Punkte	Note 2 =	unter 92 – 81 Punkte
Note 3 =	unter 81 – 67 Punkte	Note 4 =	unter 67 – 50 Punkte
Note 5 =	unter 50 – 30 Punkte	Note 6 =	unter 30 – 0 Punkte

## 1. Aufgabe (27 Punkte)

a) 7 Punkte, 4 Punkte für Messdatenpunkte und je 1,5 Punkte pro Gerade



b) 2 Punkte

Steigung und y-Achsenabschnitt

c) 2 Punkte

Datenpunkte über der Geraden klassifizieren kranke Rebstöcke, Datenpunkte unter der Geraden gesunde Rebstöcke.

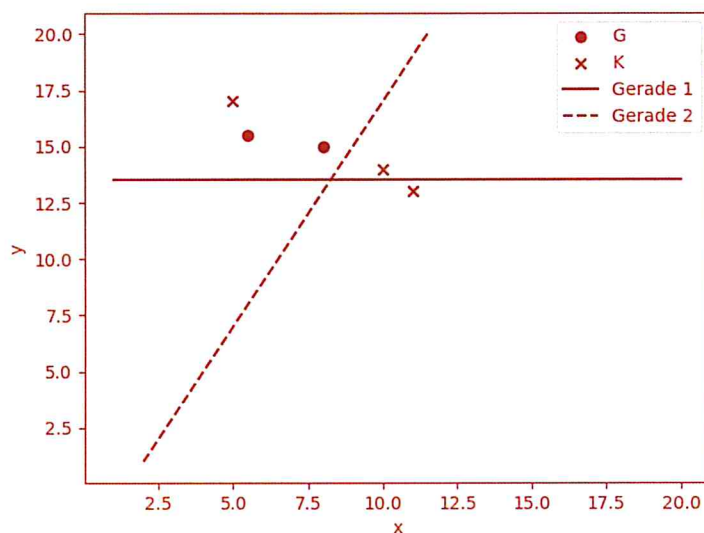
d) 4 Punkte

Gerade 2.

Gerade soll zu allen Datenpunkten beider Klassen einen möglichst großen Abstand haben.

e) 2 Punkte

Gerade	richtig	falsch
1	3	2
2	4	1



fa) 4 Punkte

	Messdatenpunkt (A = 3   B = 12)	Messdatenpunkt (A = 9   B = 11)
1	$B = 2 \cdot 3 - 3 = 3$	$B = 2 \cdot 9 - 3 = 15$
2	$12 - 3 = +9$	$11 - 15 = -4$

fb) 3 Punkte

Messdatenpunkt (3 / 12) → gesund

Messdatenpunkt (9 / 11) → krank

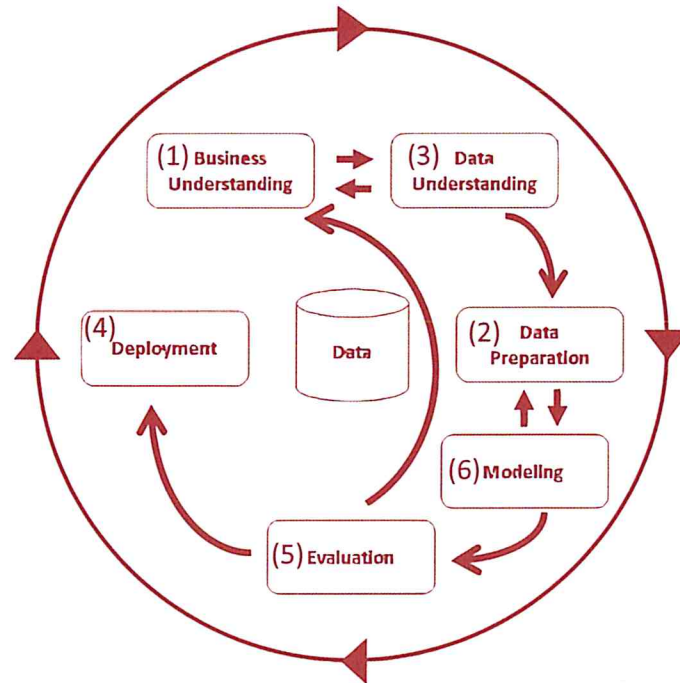
Ist die Differenz aus Messwert und Funktionswert positiv, liegt der Datenpunkt über der Trenngeraden. Ist er negativ, so liegt der Messpunkt entsprechend darunter.

g) 3 Punkte

1. Schritt	2. Schritt	3. Schritt	4. Schritt	5. Schritt	6. Schritt	7. Schritt
4	2	6	7	3	5	1

## 2. Aufgabe (25 Punkte)

a) 6 Punkte



b) 4 Punkte

```
# importieren der Panda Bibliothek
import pandas as pd

# dronen csv einlesen und in der Variable csv_df speichern
csv_df = pd.read_csv("wein/22092214.csv")

# wetter json einlesen und in der Variable wetter_df speichern
wetter_df = pd.read_json("wetter/2209.json")

# eingelesene csv_df und wetter_df verknüpfen und in data speichern
data = pd.merge(csv_df, wetter_df, on = "date", how = "inner")
```

Zeile 4: muss die letzte Zeile sein.

Zeile 6: Pfad der Datei ist „wetter/2209.json“

c) 3 Punkte

Bei einem JOIN werden Datensätze von mind. zwei Tabellen zusammengefügt (1 Punkt).

Die Tabellen müssen in einem logischen Bezug stehen/über eine Primär-/Fremdschlüsselbeziehung verfügen (1 Punkt).

In der Regel muss über ON angegeben werden, über welche Attribute die beiden Tabellen verbunden werden (1 Punkt).

Für andere richtige Aspekte können auch Punkte vergeben werden.

da) 4 Punkte

```
data[data.type == 'Piroso']
```

db) 4 Punkte

```
data[(data.status == 'gut') & (data.action == 'keine')]
```

dc) 4 Punkte

```
data[(data.status == 'Schädlinge') | (data.action == 'manuell')]
```

### 3. Aufgabe (28 Punkte)

a) 6 Punkte

```
SELECT DISTINCT W_ID, W_Name, E_Name, Wert, W_Qualität
FROM      Eigenschaft
          JOIN Messwert ON W_ID = WID
          JOIN Wein ON E_ID = EID
WHERE E_Name = "Restzuckergehalt" AND Wert < 2
ORDER BY Wert, W_Qualität DESC;
```

1,5 Punkte

1,5 Punkte

2 Punkte

1 Punkt

b) 7 Punkte

```
SELECT G_Region,
       MIN(W_Qualität) AS "niedrigste Qualität",
       MAX(W_Qualität) AS "höchste Qualität",
       AVG(W_Qualität) AS "Durchschnittliche Qualität"
FROM   Wein JOIN Weingut ON G_ID = W_GID
GROUP BY G_Region;
```

0,5 Punkte

1 Punkt

1 Punkt

1 Punkt

0,5 Punkte

2 Punkte

c) 7 Punkte

```
SELECT G_Name, COUNT(*) AS "Anzahl produzierte Weine"
FROM   Wein JOIN Weingut ON G_ID = W_GID
WHERE W_Qualität >=7
GROUP BY G_ID
HAVING COUNT(*) >2
ORDER BY COUNT(*) DESC;
```

2 Punkte

0,5 Punkte

0,5 Punkte

1 Punkt

2 Punkte

1 Punkt

d) 5 Punkte

```
SELECT W_Name,
FROM   Wein
WHERE W_Qualität = (SELECT MAX(W_Qualität) FROM Wein);
```

0,5 Punkte

0,5 Punkte

4 Punkte

e) 3 Punkte

```
DELETE FROM   Weingut
WHERE G_Name = "Richards Rebe";
```

2 Punkte

1 Punkt



#### 4. Aufgabe (20 Punkte)

a) 4 Punkte, je Nachteil 2 Punkte

Zwei von folgenden:

- Wenn das Internet nicht funktioniert (Downtime), kann man nicht mit den Daten arbeiten.
- Datensicherheit und Datenschutz beinhalten ein viel größeres Risiko, da auf alle Daten über das Internet zugegriffen werden wird.
- Mögliche Abhängigkeit vom Cloud-Anbieter bzw. des dauerhaften Angebots der konkreten genutzten Cloud-Dienstleistung
- I. d. R. kein direkter (persönlicher) Kontakt zum Dienstleister der Cloud-Dienstleistung, sondern nur über Web-Oberflächen und technische Schnittstellen (aufgrund on-demand self service)
- Das Potenzial für Cyber-Angriffe ist größer als bei unternehmensintern gespeicherten Daten.

Andere sinnvolle Lösungen sind auch zulässig.

b) 16 Punkte

2 Punkte: je Einordnung 0,5

6 Punkte: je Begründung für sehr hohen Schutzbedarf „sehr hoch“

Risiko des Eintretens → 1,5 Punkte

Auswirkung des Eintretens → 1,5 Punkte

8 Punkte: je weitere Begründung → 2 Punkte

Weil die Auswirkung bestimmend für den Schutzbedarf ist und nicht die Wahrscheinlichkeit des Eintretens

Gefahr	Schutzbedarf		Begründung
Datenbank in der Cloud	Verfügbarkeit	Sehr hoch	1. Es besteht ein hohes Risiko, dass Daten durch Cyberangriffe zerstört werden können. 1,5 P 2. Ohne Zugriff auf die gesammelten Daten kann keine Analyse der aktuellen Situation durchgeführt werden. 1,5 P
	Integrität	Sehr hoch 0,5 P	1. Es besteht ein hohes Risiko, dass Daten durch Cyberangriffe zerstört werden können. Dadurch kann die Aussagekraft der Daten beeinträchtigt werden. 1,5 P 2. Sind die Daten unvollständig bzw. nicht konsistent, können die abgeleiteten Entscheidungen falsch sein. Das kann zu immensem wirtschaftlichem Schaden führen. 1,5 P
	Vertraulichkeit	Normal	Es werden keine besonders sensiblen personenbezogenen Daten (nur Benutzerzugangsdaten) in der Cloud gespeichert, daher ist dieses Risiko als voraussichtlich niedrig einzustufen. 2 P
KI-Analysetools in der Cloud	Verfügbarkeit	Hoch 0,5 P	Das Winzergeschäft ist nicht zeitlich sensibel. Kurze Ausfallzeiten können gut verkraftet werden. Sollten die Tools mittel- oder langfristig nicht zur Verfügung stehen, dann fehlen wirtschaftlich signifikante Informationen, die zu hohen Verlusten führen können. 2 P
	Integrität	Normal 0,5 P	Sollte lediglich die Integrität der Tools beeinträchtigt sein, kann die Software neu installiert werden. 2 P
	Vertraulichkeit	Normal 0,5 P	Es werden keine personenbezogenen Daten in der Cloud gespeichert. Daher ist dieses Risiko niedrig. 2 P

Andere sinnvolle Lösungen sind auch zulässig.





