

Online-Zertifikatslehrgang

# Data Analyst IHK

Die neue Generation digitaler  
IHK-Weiterbildungen

# INHALT

ANFORDERUNG FÜR  
ABSCHLUSS-PROJEKTE

DATENSÄTZE FÜR  
ABSCHLUSS-PROJEKTE



**Data  
Analyst**<sub>(IHK)</sub>

# Anforderungen für die Abschlussprojekte

# Inhalt der Projektarbeiten (pro Gruppe)

## Projektbeschreibung

### 1. Einleitung und Ist-Analyse:

- Beschreibung der Ausgangssituation und des Hintergrundes der Daten (Unternehmen, Branche, Prozesse, etc.).
- Erläuterung der Zusammensetzung der Daten
- Bewertung der Datenqualität.

### 2. Backlog und Projektziele:

- Sammlung von User Storys und Potenzialen.
- Bewertung und Priorisierung nach Aufwand und Ertrag (Low Hanging Fruits)
- Auswahl der erreichbaren Projektziele für den ersten Sprint.

### 3. Projektausarbeitung:

- Formulierung von Hypothesen, die die Machbarkeit der Funktionen der User Storys und Potenziale untermauern.
- Zuordnung der Hypothesen zu User Storys und Potenzialen.

### 4. Projektmetriken:

- Wie lassen sich Fortschritt und Erfolg des Projektes und die Qualität der Analyse bzw. des Dashboards im Einsatz messen?
- Definitions of done für die Mehrwerte der ausgewählten Ziele.

### 5. Planung des ersten Sprints:

- Formulierung der einzelnen Aufgabenpakete mit ihrem berechneten oder geschätzten Aufwand.

### 6. Technische Beschreibung des ersten Sprints anhand des Machine Learning Canvas.

### 7. Ergebnisse der EDA und der Hypothesenüberprüfung.

### 8. Optional: Bedienungsanleitungen und Ergänzungen zur Dokumentation (Dashboard bzw. Workflow).

→ Abgabe: Dokument „Projektbeschreibung“ (PDF)

# Inhalt der Projektarbeiten (pro Gruppe)

## Datenanalyse

- Strukturierter Workflow (Datenimport, Datenbereinigung, Datentransformation, Datenanalyse, Datenexport)
- EDA zu den importierten Daten
- Datenanalyse mit Bewertung, Optimierungs- und Stabilitätstest.
- Einfache (grafische) Darstellung der Ergebnisse

→ Abgabe: Workflow mit Dokumentation und Input-Daten als Archiv (.knar)

## Datenvisualisierung

- Bericht oder Dashboard mit Visualisierung der Daten
- Interaktive Elemente zur flexiblen Erkundung der Informationen
- Darstellung des Zusammenhangs der Informationen (Welche Geschichte soll erzählt werden?)
- Kurze Dokumentation der Datenbearbeitung

→ Abgabe: Interaktiver Bericht (.pbix) mit kurzer Funktionsbeschreibung und Input-Daten

# Projektgruppen

Gruppe	Name, Teilnehmer				Datensatz
Gruppe 1					
Gruppe 2					
Gruppe 3					
Gruppe 4					
Gruppe 5					



The background of the slide features a blue-toned image of a hand reaching out to touch a wireframe globe. The globe is composed of a network of white lines and dots, symbolizing global connectivity and digital technology. The hand is positioned in the lower-left foreground, with fingers slightly curled as if about to make contact with the globe.

# **Datensätze für die Abschlussprojekte**

# Vorbereitete Datensätze



# P1: Bank Marketing

Die Daten stammen aus verschiedenen Direktmarketingkampagnen (Telefonanrufe) eines portugiesischen Bankinstituts. Dabei wurden dem Kunden ein Anlageprodukt empfohlen.

Typ: Klassifizierung

Das Klassifizierungsziel: Tätigt der Kunde eine Anlage (Variable y)?

20 Attribute in den Bereichen:

- Kundendaten
- Kontaktinformationen
- Verschiedene Kampagneninformationen
- Sozialer und ökonomische Informationen zum Kunden

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

# P2: Fahrerumfrage zu Gutscheinen

Die Daten stammen aus einer Kundenbefragung von Autofahrern, denen verschiedene Gutscheine angeboten wurden (Restaurant, Cafe,... )

Typ: Klassifikation

Klassifikationsziel: Wird ein empfohlener Gutschein akzeptiert?

23 Attribute in den Bereichen:

- Fahrtbedingung
- Gutscheininformationen
- Fahrerinformationen

<http://archive.ics.uci.edu/ml/datasets/in-vehicle+coupon+recommendation#>

Wang, Tong, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. 'A bayesian framework for learning rule sets for interpretable classification.' The Journal of Machine Learning Research 18, no. 1 (2017): 2357-2393.

# P3: Einkaufsabschlüsse im Online-Shop

Die Daten stammen aus dem aufgezeichnetem Kundenverhalten einer Online-Shop-Seite. Dabei wurden bestimmte Aspekte zum Kunden aber auch zur Einkaufssituation erfasst.

Typ: Klassifikation

Klassifikationsziel: Wurde der Kauf abgeschlossen?

18 Attribute in den Bereichen:

- Kundenverhalten
- Einkaufssituation
- Produktinformation

<http://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset#>

Sakar, C.O., Polat, S.O., Katircioglu, M. et al. Neural Comput & Applic (2018). [Web Link]

# P4: Kreditverhalten von Bankkunden

Die Daten stammen aus dem Kundengeschäft einer süddeutschen Bank aus den Jahren 1973-75. Sie beschreiben das Kreditverhalten anhand verschiedener Angaben zum Kunden und des Kredites.

Typ: Klassifikation

Klassifikationsziel: Wurde die Bedingungen des Kredites eingehalten?

21 Attribute in den Bereichen:

- Kundeninformation
- Kreditinformation

<http://archive.ics.uci.edu/ml/datasets/South+German+Credit+%28UPDATE%29#>

Grömping, U. (2019). South German Credit Data: Correcting a Widely Used Data Set. Report 4/2019, Reports in Mathematics, Physics and Chemistry, Department II, Beuth University of Applied Sciences Berlin.

# P5: Weinqualität

Die Daten stammen aus der Bewertung von portugiesischen Weinen. Dabei wurde die Qualität anhand verschiedener Inhaltsstoffe bewertet.

Typ: Regression bzw. Klassifikation

Die Qualitätsbeurteilungen liegen als Werte im Bereich zwischen 0,3 und 0,9. je größer der Wert, desto besser wurde die Qualität des Weines eingeschätzt. Eine Regression der Daten ist nicht ganz einfach. Alternativ können die Weine in Kategorien eingeteilt werden (z.B. gut und schlecht) und dann als Klassifikation vorhergesagt werden.

12 Attribute in den Bereichen:

- Inhaltsstoffe und Messwerte

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.

Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.