## 1.1

1. Nennen Sie die drei Hauptstufen des KPI-Entwurfsprozesses.

*Aufbau eines internen KPI-Teams.*

*Bestimmung der entscheidenden Erfolgsfaktoren des Unternehmens.*

*Auswahl der richtigen zu messenden Indikatoren.*

1. Kreuzen Sie die richtigen Aussagen an.
* *Der Reingewinn ist ein möglicher wichtiger Ergebnisindikator (KRI) für ein Unternehmen.*
* Leistungsindikatoren (KPIs) können finanzieller Natur sein.
* Es ist am besten, einen KPI auf jährlicher Basis zu messen.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Um den KPI-Entwurfsprozess besser steuern zu können, sollte ein KPI-Team aus Personen gebildet werden, die derzeit nicht im Unternehmen arbeiten. Wahr/*Falsch*

## 1.2

1. Nennen Sie die beiden wichtigsten Kategorien an Leistungsindikatoren (KPIs).

*Verstehen der Kund:innen*

*Verstehen der internen Prozesse der Organisation*

1. Kreuzen Sie die richtige Aussage an.
* Die Kundenfluktuationsrate gibt an, wie viele Kund:innen das Unternehmen innerhalb eines bestimmten Zeitraums mögen.
* *Die Kosten für die Berechnung der Maschinenausfallzeiten sind nicht hoch, da die Informationen direkt aus den Maschinendaten gewonnen werden können*.
1. Vervollständigen Sie den folgenden Satz:

Die Kosten für die Erfassung der Daten zur Berechnung des CUR (*Auslastungsgrads*) sind *hoch*.

## 1.3

1. Nennen Sie mindestens drei Gründe, die für einen Wechsel zum Process-Mining sprechen.

*Perfekte und durchgehende Transparenz des Prozessmodells der Organisation*

*Eliminierung von Unterschieden in den Daten und Formaten*

*Datengesteuerte Entscheidungen*

1. Kreuzen Sie die richtigen Aussagen an.
* Process-Mining kann nicht feststellen, welche Auswirkungen die einzelnen Aktivitäten auf die verschiedenen Bereiche des Unternehmens haben.
* *Die mit Process-Mining verbundenen Techniken nutzen Ereignisprotokolle zum Finden von Engpässen.*
* *Durch Process-Mining kann ein Prozessmodell automatisch und mit geringeren Kosten erstellt werden.*
1. Kennzeichnen Sie die folgende Aussage als wahr oder falsch.

Process-Mining-Techniken ermöglichen ein besseres Verständnis des gesamten OTC-Prozesses (order-to-cash). *Wahr*/Falsch

2.1

1. Nennen Sie drei Probleme, die bei der Arbeit mit dem Bag-of-Words-Modell auftreten können.

*Dünnbesetzte Vektoren erfordern mehr Rechenressourcen.*

*Die Struktur eines Satzes geht verloren.*

*Die Bedeutung des Textes wird nicht berücksichtigt.*

*Die Auswahl der richtigen Wörter für die Aufnahme in das Vokabular kann schwierig sein.*

1. Kreuzen Sie die richtige Aussage an.
* TF bestimmt die Gesamtzahl der Vorkommen eines Wortes in einem Korpus von Dokumenten.
* *Je höher der IDF-Wert, desto höher die Bedeutung eines Wortes.*
* Die fit\_transform-Methode liefert eine Matrix, die Dokumente mit Themen verknüpft.
1. Füllen Sie die Lücken im folgenden Satz aus.

Das Paket nltk.tokenize enthält die Methode *word\_tokenize* zur Tokenisierung einer Zeichenkette und zum Entfernen von *Stoppwörtern und Satzzeichen*.

2.2

1. Führen Sie die Schritte der Latent Dirichlet Allocation auf.

*Auswahl einer Zahl K von zu entdeckenden Themen.*

*Prüfen jedes Dokuments und zufällige Zuordnung jedes Wortes zu einem der Themen.*

*Betrachten jedes Wortes in jedem Dokument.*

*Berechnen der den einzelnen Themen zugeordneten Wörter im Dokument und des Anteils der mit diesem Wort verknüpften Themenzuordnungen in allen Dokumenten.*

*Zuordnen eines neuen Themas zu einem Wort.*

*Wiederholen der obigen Schritte, bis die gewünschte Zuordnung erreicht ist.*

1. Kreuzen Sie die richtigen Aussagen an.
* *LSA ist eine unüberwachte Technik in der natürlichen Sprachverarbeitung*.
* LSA geht davon aus, dass die Themen, aus denen ein Dokument besteht, jeweils eine normale Verteilung der Wörter aufweisen.
* Die beiden Parameter α und β*,* die in der Formel zur Berechnung der Signifikanz eines Wortes im LDA-Algorithmus verwendet werden, sollten mit einem Wert größer als eins berücksichtigt werden.
1. Berechnen Sie den Anteil der Wörter in einem Dokument, die T1 zugeordnet sind. Es gelte: $α = 0,2$no\_topics=2 (Anzahl der Themen), no\_words\_document=100 (Zahl der Wörter im Dokument), und die Anzahl der mit T1 verknüpften Wörter in diesem Dokument sei 20.

$$p\left(Dokument d\right)= \frac{Anzahlo\_{dt}+α}{Anzahlo\_{d }-1+no\_{t}\*α}=\frac{20+0,2}{100-1+2\*0,2}= 0,2$$

## 3.1

1. Nennen Sie 6 der wichtigsten Metriken in der Web-Analyse.

*Anzahl der Besuche*

*Auf der Website verbrachte Zeit*

*Absprungrate (Bounce Rate)*

*Ausstiegsrate (Exit Rate)*

*Konversionsrate*

*Bindung*

1. Kreuzen Sie die richtige Aussage an.
* Jedem Besuchenden, der die Website betritt, wird eine Sitzungs-ID zugeordnet.
* *Ein Cookie dient zur Identifikation einzelner Besucher:innen auf einer Website.*
* Eine gute durchschnittliche Sitzungsdauer beträgt etwa acht Minuten.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Eine niedrige Absprungrate deutet in der Regel auf ein mögliches Problem mit dem Inhalt der Website oder der Benutzeroberfläche hin. Wahr/*Falsch*

## 3.2

1. Nennen Sie einige der wichtigsten Arten des Datenverkehrs auf einer Website:

*Direkter Datenverkehr*

*Datenverkehr aus bezahlten Suchen*

*Organischer Datenverkehr*

*Sozialer Datenverkehr*

*E-Mail-Datenverkehr*

1. Markieren Sie die falschen Aussagen:
* *SEO-bezogene Daten finden Sie nur in Web-Analysewerkzeugen und nicht in externen, von Suchmaschinen angebotenen Werkzeugen.*
* Mit UTM-Parametern lassen sich nachverfolgbare URLs für die weitere Analyse in Google Analytics erstellen.
* *Die Klickrate (click rate) in einer E-Mail-Kampagne lässt sich berechnen als die Anzahl der geöffneten E-Mails geteilt durch die Gesamtzahl der versendeten E-Mails.*
1. Finden Sie den Wert der folgenden Parameter anhand der nachstehenden Beispiel-URL: Häufigkeit und Name der Kampagne.

[https://www.audio.com?utm\_source=google&utm\_medium=daily&utm\_campaign=spring\_sales](https://www.audio.com?utm_source=google&amp;amp;utm_medium=daily&amp;amp;utm_campaign=spring_sales)

*Häufigkeit: daily (täglich)*

*Name der Kampagne: spring\_sales (Frühjahrsangebote)*

## 3.3

1. Nennen Sie die wichtigsten Ansätze, die zum Aufbau von Empfehlungssystemen eingesetzt werden:

*Inhaltsbezogen*

*Kollaboratives Filtern*

*Wissensbasiert*

*Hybrid*

*Graph-basiert*

*Kontextabhängig*

1. Kreuzen Sie die richtigen Aussagen an.
* *Inhaltsbasierte Empfehlungssysteme müssen nicht auf die Daten anderer Nutzenden zugreifen.*
* *Ein hybrider Ansatz nutzt die Kombination von kollaborativem Filtern und inhaltsbasierten Ansätzen.*
* Ansätze für das kollaborative Filtern benötigen keinen Zugang zu Informationen über die früheren Aktivitäten der Nutzenden.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Das Hauptproblem bei den kollaborativen Filteransätzen ist als Kaltstart bekannt. *Wahr*/Falsch

## 4.1

1. Nennen Sie die 5 wichtigsten Schritte, die zur Analyse sozialer Medien (Social Media Analytics) gehören:

*Identifikation*

*Datensammlung*

*Datenbereinigung und ‑vorverarbeitung*

*Analyse der Daten*

*Präsentation und Interpretation der Ergebnisse:*

1. Kreuzen Sie die richtige Aussage an.
* *Geschäftsanalysen sind Analysen der vertraulichen Daten eines Unternehmens.*
* Der größte Teil der in sozialen Medien generierten Daten ist nicht öffentlich zugänglich.
* Die in den Verwaltungssystemen eines Unternehmens gespeicherten Daten sind nicht gut strukturiert.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Die auf Social-Media-Plattformen generierten Daten liegen meist in einem unstrukturierten Format vor. *Wahr*/Falsch

## 4.2

1. Nennen Sie die drei Haupttypen von Twitter-Streaming-APIs.

*Öffentliche Datenströme: Datenströme (Streams), die öffentliche Tweets enthalten.*

*Benutzer-Datenströme: Datenströme zu den Tweets eines bestimmten Nutzenden.*

*Site-Datenströme: Datenströme, die sich auf mehr als einen Nutzenden beziehen.*

1. Kreuzen Sie die richtige Aussage an.
* Das Pareto-Prinzip besagt, dass 20 % der Begriffsvorkommen aus 80 % der Einzelbegriffe in einem Dokumentenkorpus stammen.
* *Die REST-API ist von Vorteil, wenn wir auf bereits veröffentlichte Tweets zugreifen möchten.*
* Das Attribut in\_reply\_to\_status\_id enthält die ID des Nutzenden, der den Tweet ursprünglich erstellt hat.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Facebook und Twitter können sowohl als soziale Graphen als auch als Interessengraphen betrachtet werden. *Wahr*/Falsch

5.1

1. Nennen Sie die vier wichtigsten Schritte, die bei der Durchführung eines A/B-Tests befolgt werden sollten.

*Den Zweck der Website verstehen*

*Daten sammeln und mögliche Engpässe identifizieren*

*Eine Hypothese festlegen*

*Tests starten*

1. Kreuzen Sie die richtigen Aussagen an.
* *A/B-Tests sind eine der kostengünstigsten Experimentiermethoden.*
* *A/B-Tests bieten eine gute Möglichkeit, eventuelle Schwachstellen in der Marketingstrategie einer Institution zu finden.*
* Ein fehlgeschlagener Test liefert den Testpersonen keine wichtigen Informationen.
1. Kennzeichnen Sie den folgenden Satz als wahr oder falsch.

Die Ergebnisse von A/B-Tests können diejenige Komponente einer Seite aufzeigen, die eine höhere Konversionsrate bietet. Wahr/*Falsch*

5.2

1. Nennen Sie die beiden wichtigsten multivariaten Testmethoden.

*Vollständige Versuchspläne (full factorial testing)*

*Teilfaktorpläne (fractional factorial testing)*

1. Kreuzen Sie die richtigen Aussagen an.
* Für multivariate Tests ist es nicht erforderlich, dass eine Seite hohen Datenverkehr aufweist.
* *Multivariate Tests sind komplexer als A/B-Tests.*
* *Teilfaktorpläne liefern schnellere Ergebnisse als vollständige Versuchspläne.*
* Multivariate Tests sind schneller als A/B-Tests.
1. Füllen Sie die folgende Lücke aus.

Die Startseite einer Website besteht aus drei wichtigen Elementen. Wenn wir beschließen, zwei Versionen dieser Elemente zu testen, ergäben sich in einem multivariaten Test *acht* mögliche Kombinationen.

## 5.3

1. Definieren Sie Erkundung (Exploration) und Ausbeutung (Exploitation) im Zusammenhang mit Multi-armed-Bandit-Algorithmen.

*Erkundung (Exploration) bedeutet, ständig zu experimentieren, um die beste Version einer Seite auf einer Website zu finden.*

*Ausbeutung (Exploitation) bedeutet, die Version zu wählen, die den Gewinn eines Unternehmens maximiert.*

1. Kreuzen Sie die richtige Aussage an.
* *Ein Multi-armed-Bandit-Algorithmus teilt den Datenverkehr dynamisch zu.*
* A/B-Tests sind die beste Option, wenn eine Kennzahl in kurzer Zeit optimiert werden muss.
* Ein guter Multi-armed-Bandit-Algorithmus konvergiert nie.
1. Geben Sie an, ob die folgende Aussage wahr oder falsch ist.

Bandit-Algorithmen reduzieren langsam die Erkundungszeit und konzentrieren die Ressourcen auf eine Version der Website.

*Wahr*/Falsch