|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gruppe | Studierende 1 | | Studierende 2 | |
|  | Kürzel | Name, Vorname | Kürzel | Name, Vorname |
|  |  |  |  |
| Studierende 3 | | Studierende 4 | |
| Kürzel | Name, Vorname | Kürzel | Name, Vorname |
|  |  |  |  |

### Thema

Prozessmodelle und Datenbewirtschaftung

### Lernziel

Sie können die Prinzipien des Software Engineering in klassischen und in agilen Projekten nennen und in Beziehung setzen.

Sie können die Datenbewirtschaftung am Beispiel des Vorgehensmodells CRISP DM erklären und anwenden.

### Hinweise

* Bitte tragen Sie die fehlenden Daten im Kopfbereich ein.
* Bitte bearbeiten Sie Ihre Lösungen in diesem Dokument.
* Bitte verwenden Sie folgende Dateinamen-Syntax: E1\_2021ss\_Gruppe  
  Ändern Sie Gruppe entsprechend.
* Bitte verwenden Sie für den Upload das Dateiformat PDF. Sollten Sie mehrere Dateien abgeben, dann packen Sie alles in ein ZIP-Archiv.
* Die Abgabe erfolgt fristgerecht durch den Gruppenverantwortlichen per Upload in ILIAS.
* Die Bewertung berücksichtigt teilweise korrekte Lösungen. Deshalb werden die Punkte summiert und letztlich durch 100 geteilt, um auf die endgültige Punktzahl zu kommen,  
  die Sie als Startwert für die Klausur anrechnen können.

# Aufgabenstellung

[Aufgabe 1: Prozessmodelle (10 Punkte) 2](#_Toc73431061)

[Aufgabe 2: Datenbewirtschaftung (15 Punkte) 5](#_Toc73431062)

## Aufgabe 1: Prozessmodelle (10 Punkte)

### Thema

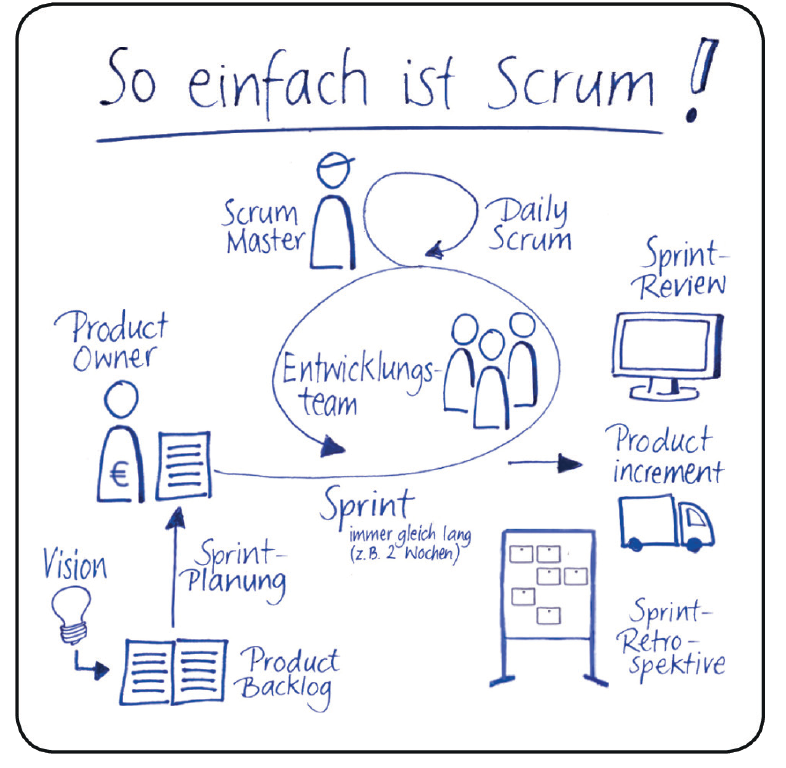
Scrum als agiles und iteratives Prozessmodell im Kontext eines Data Science Projekts.

### Quellen

Vorlesung DSCB230\_MMV\_1\_Grundlagen.pdf

Referat Kilo-Scrum.pdf

Wolf\_2015\_AgileSoftwareentwicklung.pdf



|  |  |  |
| --- | --- | --- |
| Wer oder Was | Information | Details |
| Product Owner (PO) | Matthias Mruzek-Vering |  |
| Scrum Master (SM) | Raphaele Licciardo  *Hält während der Sprints externe Störungen vom Entwicklungsteam fern und löst auftretende Probleme, die nicht die Entwicklung betreffen, zeitnah.* |  |
| Entwickler | Siehe Seite 1 | [*GitHub Teams*](https://github.com/orgs/hka-mmv/teams/dscb230-students) |
| Projekt Kickoff | Montag, 31.05.2021, 14:00 Uhr |  |
| Sprint Meeting 1 | Montag, 07.06.2021, 9:50 Uhr |  |
| Sprint Meeting 2 | Montag, 14.06.2021, 9:50 Uhr |  |
| Projekt Ende | Montag, 28.06.2021, 18 Uhr, Upload in ILIAS |  |
| Product Backlog | Gesamte Anforderungen an die Datenauswertung werden im Product Backlog festgehalten, in das der PO jederzeit Ideen für neue  Anforderungen eintragen kann. | *[GitHub Project](https://github.com/orgs/hka-mmv/projects) via*  *[GitHub Issues](https://github.com/hka-mmv/dscb230-exercise/issues) mit Label „requirement”* |
| Sprint Meeting | Besprechung der Anforderungen, die tatsächlich im nächsten Sprint umgesetzt werden sollen. Festgelegt durch PO. Die Entwickler bestimmen, wie viele dieser Anforderungen im Sprint möglich sind.  Wann: siehe Termine. | *Sprint Backlog* |
| Sprint Backlog | Die Anforderungen für den nächsten Sprint. Änderungen am Sprint Backlog während des Sprints sind nicht erlaubt. | *Sprint Backlog* |
| Sprint | Entwickler arbeiten autonom alle Anforderungen, die für den Sprint vorgesehen sind. |  |
| Daily Scrum | Meeting der Entwickler mit SM für max. 15 min.  Inhalt: Reflektieren über die Arbeit des letzten Tages und planen den kommenden Arbeitstag. Leitfragen:   1. Was habe ich seit dem letzten Daily Scrum erledigt? 2. Was hat mich dabei behindert? 3. Was werde ich bis zum nächsten Daily Scrum tun?   Wann: Täglich zur gleichen Zeit | *In unserem Fall findet dies an den Tagen Freitag im Tutorium via Zoom Meeting in Breakout-Rooms statt.* |
| Sprint Review | Entwickler präsentieren dem PO die Zwischenergebnisse, um Feedback zu bekommen. Teile des Feedbacks werden zu neuen Einträgen im Product Backlog.  Wann: Am Ende jedes Sprints. | *In unserem Fall findet dies im Sprint Meeting statt.*  *Präsentiert werden die Inhalte im GitHub Gruppenverzeichnis.* |
| Sprint Retrospektive | Scrum Team reflektiert den Entwicklungsprozess, um diesen für die Zukunft zu verbessern  Artefakte: GitHub Gruppenverzeichnis, Dateiname  Wann: Am Ende jedes Sprints. | *In unserem Fall findet dies im Sprint Meeting statt.*  *Präsentiert werden die Inhalte im GitHub Gruppenverzeichnis.* |

### Tätigkeiten

1. Bitte lesen Sie sich in das Thema ein und machen sich damit vertraut.
2. Bitte richten Sie die Kommunikationswege mit den Projektbeteiligten ein, z. B. E-Mail-Adressen, Chat/Messanger-Gruppe, Zoom Link etc.
3. Bitte organisieren Sie Ihre Entwicklungswerkzeuge insbesondere den reibungslosen Upload in Ihr GitHub Repo.
4. Bitte folgenden Sie den Hinweisen am Anfang dieses Dokuments.
5. Bitte erstellen Sie benötigte Dokumente in Ihrem GitHub Repo im entsprechendem Gruppenverzeichnis in e1.
   1. Gruppenverzeichnis erstellen, falls noch nicht gemacht
   2. Erstellen Sie eine Readme Datei, in der Sie Ihre Teammitglieder auflisten.

**Wichtig!** Leere Verzeichnisse verschwinden in GitHub, deshalb entweder eine Dummy-Datei anlegen oder die leere Datei .gitkeep.

### Legende zur Bezeichnung der Backlog Items bzw. Issues

|  |  |  |
| --- | --- | --- |
| Thema | Abkürzung | Beispiel |
| Business Unterstanding | BU | 1-BU\_GRUPPE\_E1\_Issue-Description |
| Data Unterstanding | DU | 2-DU\_GRUPPE\_E1\_Issue-Description |
| Date Preparation | DP | 3-DP\_GRUPPE\_E1\_Issue-Description |
| Modeling | MD | 4-MD\_GRUPPE\_E1\_Issue-Description |
| Evaluation | EV | 5-EV\_GRUPPE\_E1\_Issue-Description |
| Deployment | DT | 6-DT\_GRUPPE\_E1\_Issue-Description |

### Hinweis

Die Aufgabenstellung aus diesem Dokument wird noch in GitHub hochgeladen.  
<https://github.com/hka-mmv/dscb230-exercise>

## Aufgabe 2: Datenbewirtschaftung (15 Punkte)

### Thema

Datenanalyse am Beispiel Airbnb New York unter Verwendung der Methode CRISP-DM

### Quellen

Vorlesung DSCB230\_MMV\_1\_Grundlagen.pdf

<https://www.datascience-pm.com/crisp-dm-2/>

<https://datadrivencompany.de/crisp-dm-das-modell-einfach-erklaert-und-visualisiert/>

[https://www.kaggle.com/faressayah/analysis-of-airbnb-data-new-york-citylization](https://www.kaggle.com/faressayah/analysis-of-airbnb-data-new-york-city#2.-Data-Visualization)

|  |  |  |
| --- | --- | --- |
| Thema | Inhalt | Detail |
| Geschäftsverständnis  BUSINESS UNDERSTANDING  *Anforderungen und Rahmenbedingungen von PO an die Datenanalyse* | * Hosts, Art der Unterkunft, Stadtteile * Geografische Analyse in Form einer HeatMap * Günstigster Preis auf Landkarte farblich anzeigen * Weitere Analyseziele z.B. Verknüpfung mit anderem Datensatz wie Kriminalstatistik (JOIN via Neighborhood) | [*GitHub Issues*](https://github.com/hka-mmv/dscb230-exercise/issues) *mit Label „requirement”* |
| Datenverständnis  DATA UNDERSTANDING  *Skizzierung und Sichtung der benötigten Daten*  *Auswahl von Werkzeugen und Anforderungen an Formate* | **TODO**   * Datenquellen * Metadaten * Datentypen * Speicherort | **TODO** |
| Datenvorbereitung  DATA PREPARATION  *Konstruktion einer grundlegenden Population* | * Metadaten der Datenquelle erfassen * Auswahl relevanter Daten * Daten organisieren und ggf. in Data Sets aufteilen * Daten Bereinigen | Scrum Team untersucht die Datenquelle und macht PO Vorschläge der relevanten Daten mit Begründung.  Data Sets je nach Analysekontext |
| Modellierung  MODELING  *Konzeption eines Modells, das die Anforderungen maximal*  *erfüllt* | *Normalerweise wird diese Method angewandt.*  *Für unsere Übung ist es allerdings nicht passen, deshalb lassen wir dies weg.* |  |
| Analyse und Interpretation  EVALUATION & DEPLOYMENT  *Aufbereitung und Präsentation der Ergebnisse* | **TODO**  Aufbereitung der Data Preparation zu jedem Sprint Meeting.  Die Abschlusspräsentation wird in der Projektarbeit bearbeitet. | **TODO**  Übung 3: Optionale Unit Tests in GitHub Verzeichnis e3  Deployment der Abschlusspräsentation in GitHub Verzeichnis p |