**Struktur: 6-Wochen-Projekt mit Scrum (3 Sprints à 2 Wochen)**

* **Sprint 1** (Woche 1–2): Ziel + Datensuche, Datendownload, Sortierung, Bereinigung
* **Sprint 2** (Woche 3–4): Erweiterte Analyse
* **Sprint 3** (Woche 5–6): Feinschliff, Doku, Testing, Präsentation

**Zeitplan Woche für Woche**

**🧭 Woche 1: — Vision, Planung**

**23.06-29.06**

Meeting 23.06.2025, 08:30

* Ziel definieren
* Projektidee formulieren: Buchmarktanalyse
* Andere Projekte anschauen
* *Rollen klären (Product Owner, Scrum Master, Entwickler)*
* Erste Aufgaben:
  + GitHub Projekt erstellen
  + Rohdatenimport
  + Ausgewählte Test-Datenstruktur analysieren
  + Erstes Thema: Bestselleranalyse/50 berühmte Büche🡪 Streemlit einfach vesualisieren
* **Sprint Planning 1**

Datensuche und Download:

🡪Siehe Dateien**: links\_mit\_info**

🡪 Ordner: **Rohdaten**

**Woche 2 — Datenbeschaffung & Basis-Dashboard**

**30.06-06.07**

Meeting 30.06.2025, 09:00

* Verschiedene Datenquelle analysieren
* Möglichkeiten Daten-scraping ausprobieren
* Größere Datei, über 1000 Bücher als csv-Datei downloaden
* Streamlit: Basis-Dashboard erstellen
* Datenbeschaffung & -integration:
* Echte Bestseller-Daten (1070 Datensätze) importiert
* Synthetisch erstellte Nicht-Bestseller-Daten(500 Datensätze) hinzugefügt. Aber fürs End Tool darauf verzichtet
* Datensätze in ein gemeinsames DataFrame zusammengeführt
* Datensatz mit Bestseller und Labels, ob verfilmt oder nicht🡪 als Basis für ML Analyse

**Warum synthetische Daten?**

Echte Nicht-Bestseller-Daten sind schwer verfügbar und oft unvollständig. Um das Datenungleichgewicht auszugleichen und das Modell besser zu trainieren, haben wir synthetische Nicht-Bestseller-Daten erstellt. So verbessern wir die Datenbasis und die Vorhersagequalität.

* **Sprint Review & Retrospektive 1**

**Woche 3 — Tiefgehende Datenvorbereitung und erste ML Analyse**

**07.07-13.07**

Meeting 07.07.2025, 09:00

Meeting: 11.07.2025, 11:00

* CSV Dateien erklären
* 3 Themenbereichen festlegen
* **Sprint Planning 2**

**Woche 4 — Komplexere Visualisierungen & Optimierung**

**14.07-20.07**

Meeting 14.07.2025, 11:00

Meeting 18.07.2025, 09:30

* Fortgeschrittene Darstellungen: Heatmaps/*paiplot/Boxplot/skatterplot*
* Kategorische Daten🡪 Nummerische umwandeln, fertig für ML Modele
* ML Modele testen und festlegen
* Streamlit Basis für 3 Bereichen festlegen
* **Sprint Review & Retrospektive 2**

**Woche 5 — Testung, Fehlerbehandlung, Dokumentation**

**21.07-27.07**

Meetings täglich um 9:00

* Dokumentation schreiben (README, Installationsanleitung, Architektur)
* Datei für die neue Test-Bücher vorbereiten, die für Vorhersage benutz werden
* Streamlit gestalten
* **Sprint Planung 3**

**Woche 6 — Präsentationsvorbereitung & Projektabschluss**

**28.07-01.08(Präsentationtag)**

Meetings täglich um 9:00

* Präsentation oder Demo bauen (direkt im Streamlit-Dashboard)
* Review mit Fokus auf Zielerreichung
* Finales Deployment
* **Sprint Review & Retrospektive 3**

**Scrum-Elemente regelmäßig einsetzen**

| **Scrum-Element** | **Wann/Wo** |
| --- | --- |
| Product Backlog | Direkt in Woche 1 starten |
| Sprint Planning | Immer zu Beginn von Woche 1, 3 und 5 |
| Daily Scrums | Jeden Arbeitstag, ca. 15 Minuten |
| Sprint Review | Immer Ende von Woche 2, 4 und 6 |
| Sprint Retrospektive | Direkt nach dem Review |