

# Einführung in Power BI

---

## IMDB Top 250

---

### Data Source

Wir verwenden die [IMDB Top 250 User rated Movies](https://www.imdb.com/chart/top/) als unseren Datensatz. Hierzu müssen wir nur die URL kopieren: <https://www.imdb.com/chart/top/>

Wir erstellen ein neues Projekt, bei *Daten Abrufen* wählen wir *Web* aus und fügen die URL ein. Nun werden uns mehrere Tabellen angezeigt und die mit den gewünschten Daten aus, indem wir den Haken neben der Tabelle setzen. Bevor wir die Daten nun in Power BI laden wollen wir sie noch anpassen, dazu wählen wir *Daten Transformieren*.

### Daten transformieren

Der Power-Query-Editor öffnet sich. Hier können die Daten einfach angepasst werden. Zuerst entfernen wir die Spalten 5 bis 14 da sie keine relevanten Informationen enthalten. Dann können wir den Spalten aussagekräftige Namen geben: *Platzierung*, *Titel*, *Erscheinungsjahr*, *Rating*. In der Spalte *Platzierung* stehen noch eine Menge überflüssiger Informationen wie der Titel und das Jahr, das wir auch in anderen Spalten finden können, weshalb wir diese löschen möchten. Dazu wählen wir die Spalte aus → *Extrahieren* → *Text nach Trennzeichen* und geben einen Punkt ein.

Ähnliches machen wir bei der Spalte *Erscheinungsjahr*, um den "-" zu entfernen. Zuerst müssen wir die Spalte aber in das Text Format umwandeln: Rechtsklick auf *Spaltenheader* → *Typ ändern* → *Text*. Dann wieder *Extrahieren* → *Text vor Trennzeichen* und wir geben "-" ein. Dann ändern wir den Typ wieder zurück zu *Ganzer Zahl*.

Um die Daten später besser Gruppieren zu können wollen wir nun noch eine Spalte mit der Dekade hinzufügen. Dazu duplizieren wir die *Erscheinungsjahr* Spalte und geben ihr einen anderen Namen. Wir ändern den Typ auf Text und *Extrahieren* → *Erste Zeichen* → 3. Damit haben wir die ersten drei Stellen des Jahres extrahiert und müssen nur noch eine 0 am Ende einfügen: *Format* → *Suffix hinzufügen* → 0.

Nun können wir die Daten laden → *Schließen und übernehmen*.

### User Interface

Power BI ähnelt stark den MS Office Programmen: Am oberen Rand finden wir die *Ribbon* Leiste mit allen wichtigen Menüpunkten. Rechts befinden sich drei Leisten zum ausklappen mit den Namen *Felder*, *Visualisierung* und *Filter*. In *Felder* finden wir die Spalten unserer Datenquellen, unter *Visualisierung* mögliche Grafiken und Diagramme, die in der Leiste *Filter* noch genauer festgelegt werden können. Unten können wir ähnlich wie in Excel weitere Seiten anlegen. In BI nennen sich diese Berichtseiten.

Auf der linken Seite befinden sich untereinander drei Symbole für (von oben nach unten) *Bericht*, *Daten* und *Modell*. In *Bericht* können wir unsere Diagramme erstellen. Unter *Daten* können wir die Tabellen einsehen und unter *Modell* bekommen wir eine Übersicht wie die Daten (sollten wir mehrere Tabelle haben) zusammenhängen.

## Erstellen erster Grafiken

Um schnell einfach Grafiken zu produzieren wählen wir zuerst im Fenster *Felder* die gewünschten Spalten aus. Zum Beispiel können wir ein Kreisdiagramm auswählen, die Dekade in die das Feld Legende ziehen, sowie die Titel in das Werte Feld. Damit sehen wir wie viele Filme der Top 250 in welcher Dekade erschienen sind.

Wenn wir einen zeitlichen Verlauf davon sehen möchten, erstellen wir ein neues Liniendiagramm mit der Dekade auf der y-Achse und die Anzahl der Titel als Werte. Auf dem verblieben Platz auf dieser Berichtsseite können wir uns noch die Ratings Pro Film als Balken Diagramm anzeigen lassen.

Welche Grafiken sind mit dem Datensatz noch möglich?