Merkmale umcodieren

**Link zu CODAP:** <https://tinyurl.com/jim-pb50>

|  |
| --- |
| Beim **Umcodieren von Merkmalen** geht es darum, Ausprägungen eines Merkmals zu Gruppen zusammenzufassen und im Datensatz neu zu codieren, um schneller Aussagen treffen zu können.Fasst man die Ausprägungen eines Merkmals zu zwei Gruppen zusammen, so erhält man ein **binäres Merkmal**. Verschiedene Gruppierungen von Ausprägungen ermöglichen **unterschiedliche Interpretationen**! |

# Wie viele Befragte hören häufig Podcasts und wie viele Befragte tun dies selten?

Schaut man sich das Merkmal Podcast\_hören an, so stellt man fest, dass es einen großen Anteil an Personen gibt, die angegeben haben, nie oder seltener Podcasts zu hören.



Ab jetzt nehmen wir als Data Scientists Einfluss auf die Auswertung, indem wir festlegen bzw. modellieren welche Ausprägungen für uns bedeuten, dass eine Person **selten** bzw. **häufig** Podcasts hört. Hier legen wir fest, dass

* **selten** für *1\_nie*, *2\_seltener*, *3\_einmal\_im\_Monat* und *4\_einmal\_in\_14\_Tagen* steht,
* **häufig** für *5\_einmal\_pro\_Woche*, *6\_mehrmals\_pro\_Woche*, *7\_täglich* steht.

Das codieren wir nun in die Daten.

Als erstes ziehen wir das Merkmal „Podcast\_hören“ per Drag&Drop in der Tabelle ganz nach links, bis ein gelbes Feld erscheint.



Das Merkmal Podcasts\_hören liegt nun in der Tabelle auf einer höheren Ebene und es werden die sieben unterschiedlichen Ausprägungen angezeigt. Wir sehen, dass das Merkmal 7 verschiedene Ausprägungen zwischen „1\_nie“ und „7\_täglich“ hat.



Diese wollen wir nun in einem neuen Merkmal umcodieren (bzw. zusammenfassen), sodass das neue Merkmal nur noch die beiden Ausprägungen **häufig** und **selten** enthält.

Dazu erstellen wir durch Klicken auf das Plus links ein neues Merkmal.



Dieses neue Merkmal nennen wir nun passend um. Das geschieht mit einem Linksklick auf „neuesMerkmal“ und dann „umbenennen“. Als neuen Merkmalsnamen wählen wir „Neu\_Podcasts\_hören“, damit wir wissen, auf was sich das Merkmal bezieht.





Nun müssen die Ausprägungen passend umcodiert werden. Dazu sortieren wir als erstes die alten Ausprägungen, indem wir auf den Merkmalsnamen klicken und „Aufsteigend sortieren“ wählen.





Nun können die Ausprägungen des neuen Merkmals „Neu\_Podcasts\_hören“ passend zusammengefasst werden.

 1\_nie, 2\_seltener, 3\_einmal\_im\_Monat, 4\_einmal\_in\_14\_Tagen 1\_selten

5\_einmal\_pro\_Woche, 6\_mehrmals\_pro\_Woche, 7\_täglich 2\_häufig

Die entsprechenden Ausprägungen werden von Hand in die Tabelle geschrieben.



Anschließend müssen die beiden Merkmale wieder per Drag&Drop zurück in den rechten Teil der Tabelle geschoben werden.





Die Auswertung ist nun einfacher.



73% der Befragten hören selten Podcasts und 27% hören häufig Podcasts.

# Alternative Modellierung

Beim Umcodieren modelliert man die Daten. Dies kann unterschiedlich geschehen. Alternativ hätte man auch unterschieden können zwischen Befragten, die **nie** Podcasts hören im Gegensatz zu allen anderen, die angegeben überhaupt **jemals** Podcasts zu hören.

Die Modellierung und somit Umcodierung sähe dann so aus und die Auswertung dazu:

 

Die Interpretation hierzu lautet dann: 37% der Befragten hören niemals Podcasts, wohingegen 63% der Befragten angeben, überhaupt Podcasts zu hören.