

Merkmale umcodieren

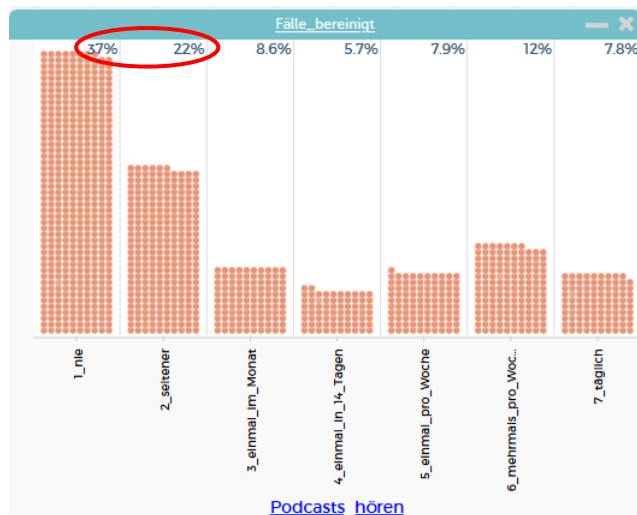
Link zu CODAP: <https://tinyurl.com/jim-pb50>

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.

JIM-PB_50_bereinigt

Fälle (1273 Fälle, 14 ablegen)

In- dex	Ge- schlecht	Klassen- stufe	Schul- form	Alter	Accounts Anzahl	Ebook le- sen	Online ng lesen	Online en lesen	Podcasts hören	Tablet nutzen	festenPC nutzen	Laptop nutzen	tragbSpi nutzen	festeSp nutz
1	weiblich	11	Cygnas...	15	8	4_einm...	1_nie	1_nie	3_einm...	7_täglich	7_täglich	2_selte...	1_nie	1_nie
2	weiblich	6	Cygnas...	11	6	1_nie	1_nie	1_nie	3_einm...	4_einm...	1_nie	7_täglich	6_mehr...	1_nie
3	weiblich	10	Cygnas...	15	3	3_einm...	5_einm...	5_einm...	6_mehr...	7_täglich	1_nie	6_mehr...	2_selte...	4_einm
4	männli...	7	Sonstig...	12	1	1_nie	1_nie	1_nie	2_selte...	7_täglich	1_nie	7_täglich	7_täglich	2_selte...
5	weiblich	9	Cygnas...	15	4	1_nie	6_mehr...	2_selte...	1_nie	6_mehr...	1_nie	7_täglich	1_nie	2_selte...
6	männli...	12	Cygnas...	17	4	1_nie	6_mehr...	6_mehr...	3_einm...	7_täglich	2_selte...	4_einm...	1_nie	5_einm
7	weiblich	12	Cygnas...	17	10	1_nie	4_einm...	1_nie	6_mehr...	6_mehr...	6_mehr...	6_mehr...	1_nie	1_nie

Merkmal fallen lassen, um eine neue Kollektion zu erzeugen

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.

JIM-PB_50_bereinigt

Podcasts_hören

In- dex	Podcasts hören
1	3_einmal_im_Mo...
2	6_mehrmals_pro...
3	1_nie
4	2_seltener
5	4_einmal_in_14_T...
6	7_täglich
7	5_einmal_pro_Wo...

Fälle (1273 Fälle, 14 ablegen)

In- dex	Ge- schlecht	Klassen- stufe	Schul- form	Alter	Accounts Anzahl	Ebook le- sen	Online ng lesen	Online en lesen	Tablet nutzen	festenPC nutzen
1	weiblich	11	Cygnas...	15	8	4_einm...	1_nie	1_nie	7_täglich	7_täglich
2	weiblich	6	Cygnas...	11	6	1_nie	1_nie	1_nie	4_einm...	1_nie
3	männli...	12	Cygnas...	17	4	1_nie	6_mehr...	6_mehr...	7_täglich	2_selte...
4	weiblich	12	Cygnas...	18	5	6_mehr...	6_mehr...	6_mehr...	7_täglich	2_selte...
5	männli...	7	Cygnas...	12	7	5_einm...	6_mehr...	2_selte...	7_täglich	1_nie
6	weiblich	9	Cygnas...	14	4	2_selte...	2_selte...	2_selte...	7_täglich	6_mehr...
7	weiblich	5	Cygnas...	11	2	1_nie	1_nie	1_nie	4_einm...	2_selte...

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.

JIM-PB_50_bereinigt

Podcasts_hörens (7 Fälle, 14 ablegen)

In- dex	Podcasts hören
1	3_einmal_im_Monat
2	6_mehrmals_pro_Woche
3	1_nie
4	2_seltener
5	4_einmal_in_14_Tagen
6	7_täglich
7	5_einmal_pro_Woche

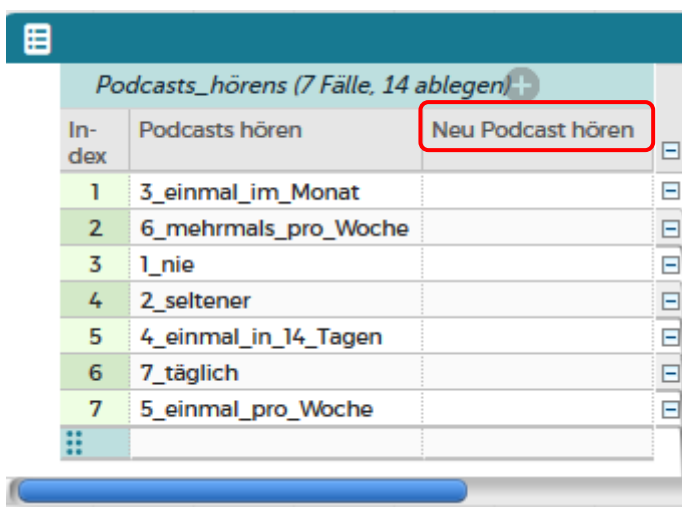
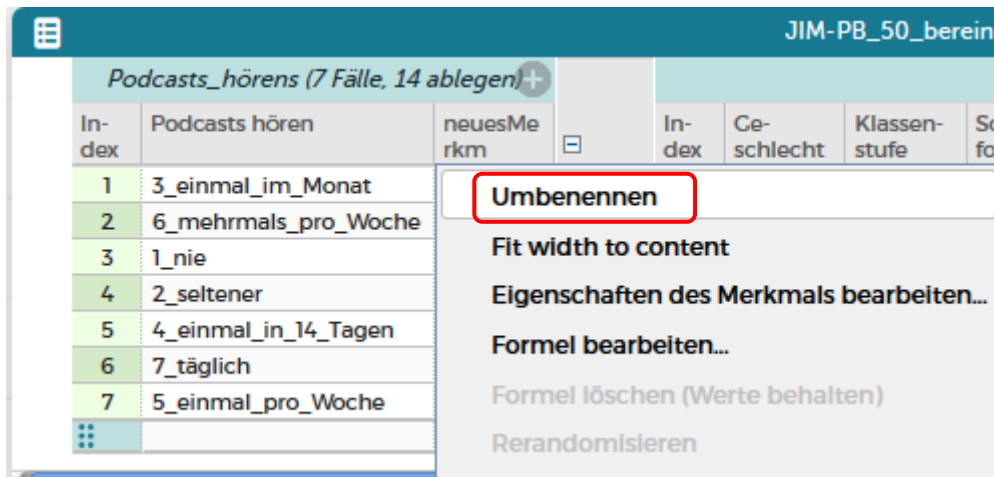
Der Tabelle ein neues Merkmal hinzufügen

neuesMe

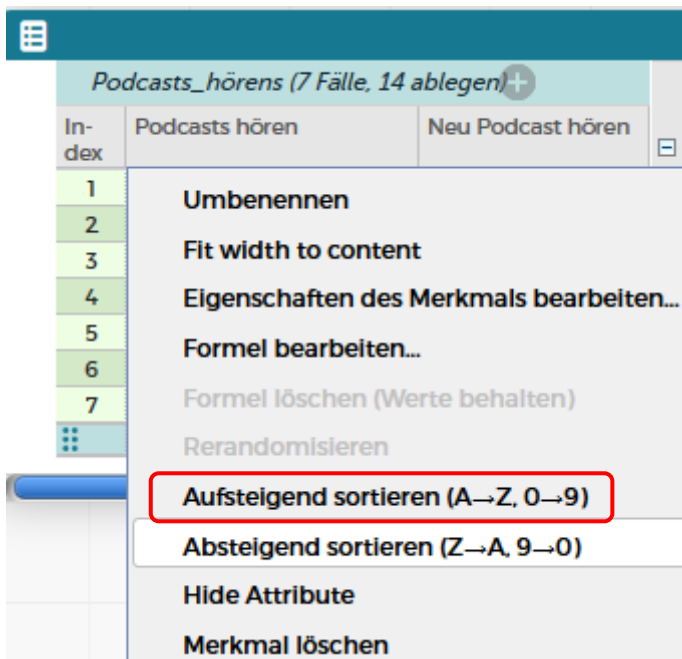
Fälle (1273 Fälle, 14 ablegen)

In- dex	Ge- schlecht	Klassen- stufe	Schul- form	Alter	Accounts Anzahl	Ebook le- sen
1	weiblich	11	Cygnas...	15	8	4_einm...
2	weiblich	6	Cygnas...	11	6	1_nie
3	männli...	12	Cygnas...	17	4	1_nie
4	weiblich	12	Cygnas...	18	5	6_mehr...
5	männli...	7	Cygnas...	12	7	5_einm...
6	weiblich	9	Cygnas...	14	4	2_selte...
7	weiblich	5	Cygnas...	11	2	1_nie

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.



Podcasts_hörens (7 Fälle, 14 ablegen)		
In- dex	Podcasts hören	Neu Podcast hören
1	1_nie	
2	2_seltener	
3	3_einmal_im_Monat	
4	4_einmal_in_14_Tagen	
5	5_einmal_pro_Woche	
6	6_mehrmals_pro_Woche	
7	7_täglich	

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.

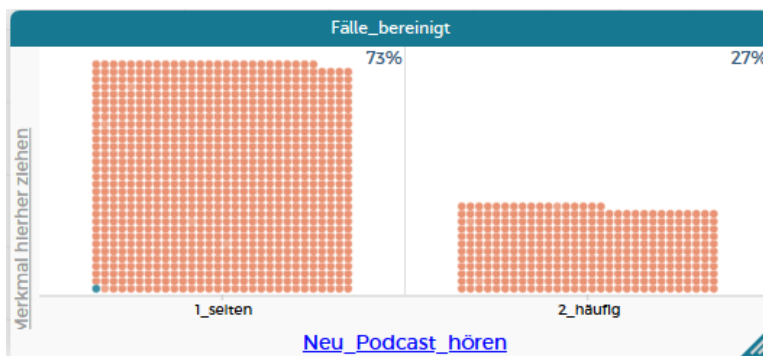
Podcasts_hörens (7 Fälle, 14 ablegen)		
In- dex	Podcasts hören	Neu Podcast hören
1	1_nie	1_selten
2	2_seltener	1_selten
3	3_einmal_im_Monat	1_selten
4	4_einmal_in_14_Tagen	1_selten
5	5_einmal_pro_Woche	2_häufig
6	6_mehrmals_pro_Woche	2_häufig
7	7_täglich	2_häufig

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

Podcasts_hörens (7 Fälle, 14 ablegen)			Fälle (1273 Fälle, 14 ablegen)										
In- dex	Podcasts hören	Neu Podcast hören	In- dex	Ge- schlecht	Klassen- stufe	Schul- form	Alter	Accounts Anzahl	Ebook le- sen	Online ng lesen	Online en les	Tablet nutzen	fest putz
1	1_nie	1_selten	1	männli...	7	Sonstig...	12	1	1_nie	1_nie	1_nie	2_selte...	7_täg
2	2_seltener	1_selten	2	weiblich	9	Cygnas...	15	4	1_nie	6_mehr...	2_selte...	6_mehr...	1_nie
3	3_einmal_im_Monat	1_selten	3	männli...	9	Cygnas...	14	3	1_nie	1_nie	2_selte...	4_einm...	7_täg
4	4_einmal_in_14_Tagen	1_selten	4	männli...	11	Cygnas...	16	5	1_nie	2_selte...	1_nie	7_täglic	1_nie
5	5_einmal_pro_Woche	2_häufig	5	männli...	5	Cesamt...	12	0	1_nie	1_nie	1_nie	6_mehr...	1_nie
6	6_mehrmals_pro_Woche	2_häufig	6	männli...	8	Sonstig...	15	22	1_nie	1_nie	1_nie	2_selte...	1_nie
7	7_täglich	2_häufig	7	weiblich	6	Cesamt...	11	0	1_nie	1_nie	1_nie	5_einm...	6_me

JIM-PB_50_bereinigt											
Fälle (1273 Fälle, 14 ablegen)											
In- dex	Ce- schlecht	Klassen- stufe	Schul- form	Alter	Accounts Anzahl	Ebook le- sen	Online ...ng lesen	Online ...en lesen	Podcasts hören	Neu Podcast hören	Tablet nutzen
1	männli...	7	Sonstig...	12	1	1_nie	1_nie	1_nie	1_nie	1_selten	2_selte...
2	weiblich	9	Cymnas...	15	4	1_nie	6_mehr...	2_selte...	1_nie	1_selten	6_mehr...
3	männli...	9	Cymnas...	14	3	1_nie	1_nie	2_selte...	1_nie	1_selten	4_einm...
4	männli...	11	Cymnas...	16	5	1_nie	2_selte...	1_nie	1_nie	1_selten	7_täglich
5	männli...	5	Cesamt...	12	0	1_nie	1_nie	1_nie	1_nie	1_selten	6_mehr...
6	männli...	8	Sonstig...	15	22	1_nie	1_nie	1_nie	1_nie	1_selten	2_selte...
7	weiblich	6	Cesamt...	11	0	1_nie	1_nie	1_nie	1_nie	1_selten	5_einm...
8	weiblich	12	Cymnas...	17	7	1_nie	1_nie	1_nie	1_nie	1_selten	1_nie

Die Auswertung ist nun einfacher.

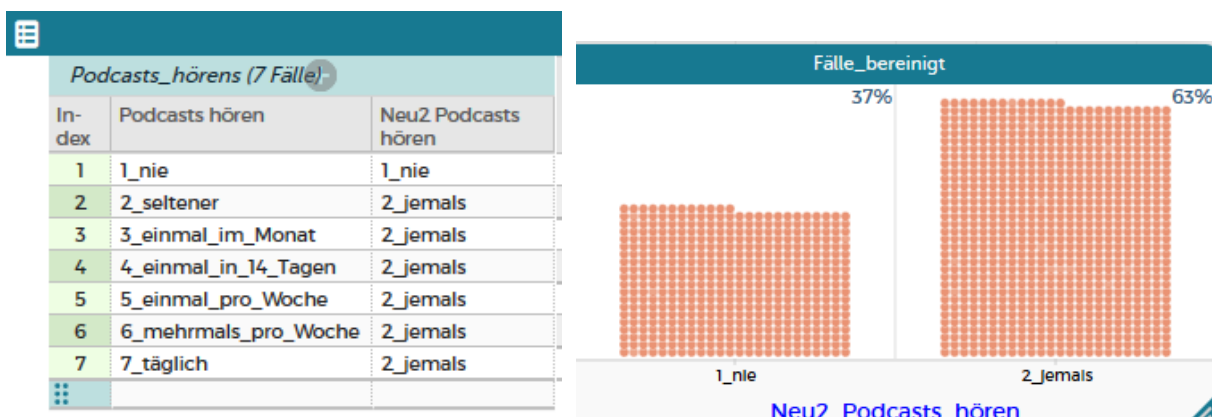


73% der Befragten hören selten Podcasts und 27% hören häufig Podcasts.

Alternative Modellierung

Beim Umcodieren modelliert man die Daten. Dies kann unterschiedlich geschehen. Alternativ hätte man auch unterscheiden 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.