

Kurze Analyse der Saar Wars Teilnehmerzahlen anhand der climbercontest.de/saarwars website

(nach Löschung einiger duplikate -> daten in den textblöcken nicht mehr aktuell)

Hintergrund: Es treten viele Duplikate auf (fällt beim Durchsehen der Liste auf) -> Daten verfälscht

Motivation: Fehlinformation, dass es über 500 Teilnehmer gibt

Benefit: Klärung wo das vermeintlich fehlende Geld der Duplikate hin ist

Annahme: Menschen, deren Benutzername in derselben Kletterhalle bereits vorhanden war, werden als Duplikat angenommen, d.h. duplikate Namen einer Kletterhalle werden herausgeschnitten, da angenommen wird, dass niemand doppelt heißt.

Dies ist aufgrund der Wahrscheinlichkeit vertretbar.

Main Points

Es gibt möglicherweise mehr als 44 Duplikate -> irgendwo ist das System fehlerhaft

Möglicherweise vermeintlich fehlendes Geld: >220€ (aktuelle daten im code print)

Am Ende des Codes gibt es eine vermutlich unvollständige, aber duplikatfreie Liste aller Teilnehmer

Abfolge der Tabellen

- 1. KBA weiblich
- 2. KBA männlich
- 3. Rocklands weiblich
- 4. Rocklands männlich
- 5. Olymp weiblich
- 6. Olymp männlich

```
In [9]: import pandas as pd
from IPython import display
from datetime import datetime

print(f'Letzter Durchlauf: {datetime.now()}')

table = pd.read_html('https://climbercontest.de/SaarWars/contest.php')

for dataframe in table:
    with pd.option_context('display.max_rows', 7):
        display.display(dataframe)

dupl_participants = 0
for df in table:
    dupl_participants += len(df)
print(f'Total participants including duplicates = {dupl_participants}')
```

letzter Durchlauf: 2022-06-21 13:42:11.310430

	0	1	2
0	1	Michaela Meisberger	51 P.
1	2	Adeline Hehetor	48 P.
2	3	Tamina Lorenz	39 P.
...	...	...	...
75	24	Sandra Rodriguez	0 P.
76	24	Elena Weiss	0 P.
77	24	Laura Petri	0 P.

78 rows x 3 columns

	0	1	2
0	1	Hendrik Junkwitsch	99 P.
1	2	Michael Becker	82 P.
2	3	Mika Scherer	79 P.
...	...	...	...
161	60	vincent renonce	0 P.
162	60	Frederic Zenner	0 P.
163	60	Georg Friedrich	0 P.

164 rows x 3 columns

	0	1	2
0	1	Carlotta Barth	85 P.
1	2	Natascha Klicker	56 P.
2	3	Jennifer Soma	46 P.
...	...	...	...
55	27	Michelle Reuter	0 P.
56	27	Jennifer Soma	0 P.
57	27	Sophie Culmann	0 P.

58 rows x 3 columns

	0	1	2
0	1	Patrick Igel	114 P.
1	2	Luca Kaul	112 P.
2	3	Finn Machwirth	103 P.
...	...	...	...
120	67	Hendrik Bock	0 P.
121	67	Felix Steinmetz	0 P.
122	67	Samuel Fuchs	0 P.

123 rows x 3 columns

	0	1	2
0	1	Julia Jene	52 P.
1	2	Tanja Trenkler	46 P.
2	3	Jana Brust	39 P.
...	...	...	...
38	25	Marzena Steuer	0 P.
39	25	Schröer Britta	0 P.
40	25	Verena Wittling	0 P.

41 rows x 3 columns

	0	1	2
0	1	Gregor Lotz	97 P.
1	2	Connor Johann	59 P.
2	3	Elias Weisser	50 P.
...	...	...	...
91	37	Jakob Mertz	0 P.
92	37	Sebastian Wendling	0 P.
93	37	Alexander Ziebert	0 P.

94 rows x 3 columns

Total participants including duplicates = 558

```
In [10]: final_tables = []
duplicate_list = []
num = 0

for tables in table:
    indices_of_duplicates = []
    #temp = tables.drop(labels=[0, 2], axis=1) <== ersetzt durch (subset=1) Spezifikation
    temp = tables.copy()
    copy = temp.duplicated(subset=1)

    for index, dupl_value in copy.iteritems(): # iterate over pd.Series
        if dupl_value: # check if value at index is duplicate
            indices_of_duplicates.append(index)

    #print(indices_of_duplicates)

    #duplicates = tables.iloc(indices_of_duplicates)
    duplicate_list.append(temp.iloc(indices_of_duplicates)) # --> eckige klammern müssen benutzt werden bruh
    #duplicate_list = pd.DataFrame(duplicate_list)

    temp = temp.drop_duplicates(subset=1)
    final_tables.append(temp)
    num += len(indices_of_duplicates)

print(f'folgende Kandidaten können Duplikate sein ((num) Stück): ')

for dataframe in duplicate_list:
    with pd.option_context('display.max_rows', 7):
        dataframe = pd.DataFrame(dataframe)
        display.display(dataframe)
```

folgende Kandidaten können Duplikate sein (30 Stück):

	0	1	2
33	24	Sandra Rodriguez	0 P.
38	24	Madeleine Birtel	0 P.
48	24	Sandra Rodriguez	0 P.
68	24	Stefanie Dohm	0 P.
71	24	Sandra Rodriguez	0 P.
75	24	Sandra Rodriguez	0 P.

	0	1	2
70	60	Marco Ladwein	0 P.
88	60	Mike Müller	0 P.
150	60	Christian Schmidt	0 P.

	0	1	2
30	27	Tanja Biehl	0 P.
31	27	Sabrina Rauber	0 P.
36	27	Hannah Hattenbach	0 P.
48	27	Annette Baldauf	0 P.
56	27	Jennifer Soma	0 P.

	0	1	2
92	67	Moritz Huber	0 P.
109	67	Salvatore Santilippo	0 P.
121	67	Felix Steinmetz	0 P.

	0	1	2
51	37	Alexander Potapov	0 P.
62	37	Alexander Potapov	0 P.
63	37	Steven Jessel	0 P.
...	...	...	...
87	37	Jakob Mertz	0 P.
89	37	Tom Kehli	0 P.
91	37	Jakob Mertz	0 P.

13 rows x 3 columns

```
In [11]: # for dataframe in final_tables:
#     with pd.option_context('display.max_rows', None):
#         display.display(dataframe)

print(f'Total participants KBA = {len(final_tables[0])} + len(final_tables[1])} \nKBA women = {len(final_tables[0])}
print(f'Total participants Rocklands = {len(final_tables[2])} + len(final_tables[3])} \nRocklands women = {len(
print(f'Total participants Olymp = {len(final_tables[4])} + len(final_tables[5])} \nOlymp women = {len(final_tables[4])}

total_participants = 0
for df in final_tables:
    total_participants += len(df)

print(f'Total number of participants = {total_participants}')
dupl = dupl_participants - total_participants
print(f'Difference between totals = {dupl}')
print(f'ggf. fehlendes Geld = {dupl * 5}€')

-----
Total participants KBA = 233
KBA women = 72
KBA men = 161
-----
Total participants Rocklands = 173
Rocklands women = 53
Rocklands men = 120
-----
Total participants Olymp = 122
Olymp women = 41
Olymp men = 81
-----
Total number of participants = 528
Difference between totals = 30
ggf. fehlendes Geld = 150€
```

Folgender Code ist nicht mehr für die Analyse relevant

```
In [12]: # add id as index
df_complete = pd.concat(final_tables)

# generate ids
ids = list(range(0, len(df_complete)))
df_complete.insert(1,loc=0, column='id', value=ids)
df_complete = df_complete.set_index('id')

# export as json
df_complete.to_json(path_or_buf='./Teilnehmer.json')

#with pd.option_context('display.max_rows', 2):
#    display.display(df_complete)

df_json = pd.read_json('./Teilnehmer.json')
df_json = df_json.sort_index()

# export as sorted json
df_json.to_json(path_or_buf='./Teilnehmer_final.json')

df_new_json = pd.read_json('./Teilnehmer_final.json')
with pd.option_context('display.max_rows', None):
    display.display(df_json)
```

0		1	2
1	0	Michaela Meisberger	51 P.
2	1	Adeline Hebestor	48 P.
3	2	Tamia Lorenz	39 P.
4	3	Mia Siegel	36 P.
5	4	Hanna Niedermeyer	28 P.
6	5	Pauline Rietz	27 P.
7	6	Nina Stadler	24 P.
8	7	Nina Engelkamp	23 P.
9	8	Adriana Altpeter	21 P.
10	9	Marlene Fries	20 P.
11	10	Anne-Marie Schwarz	18 P.
12	11	Sandra Rodriguez	17 P.
13	12	Vanessa Rodriguez	15 P.
14	13	Theresa Müller	11 P.
15	14	Julia Wagner	11 P.
16	15	Amélie Müller Pinto	10 P.
17	16	Denise Däuber	9 P.
18	17	Madeleine Birtel	8 P.
19	18	Paula Mohr	7 P.
20	19	Louisa Bachhofen	7 P.
21	20	Amélie Clemens	4 P.
22	21	Michaela Kremer	4 P.
23	22	Verena Blank	1 P.
24	23	Isabelle Kammer	0 P.
25	24	Paola Goergen	0 P.
26	25	Marie Wolter	0 P.
27	26	Katharina Schackmar	0 P.
28	27	Eva Friedrich	0 P.
29	28	Lara Köhler	0 P.
30	29	Valerie Feis	0 P.
31	30	Stefanie Dohm	0 P.
32	31	Lena Kammer	0 P.
33	32	Malke Zimmer	0 P.
34	33	Rebecca Krieger	0 P.
35	34	Jessica Hahn	0 P.
36	35	Luise Kempf	0 P.
37	36	Franca Kierfing	0 P.
38	37	zoe moll	0 P.
39	38	Daniela Fischer	0 P.
40	39	Ida Stapf	0 P.
41	40	Nina Archut	0 P.
42	41	Cosima Hafner	0 P.
43	42	Zinzi Sarah Ngumou	0 P.
44	43	Jessica Kamp	0 P.
45	44	Lauren nautin	0 P.
46	45	Matilda Krauser	0 P.
47	46	Michaela Berger	0 P.
48	47	Verena Rothbach	0 P.
49	48	Hannah Mai	0 P.
50	49	Lexa Görg	0 P.
51	50	Leila Funke	0 P.
52	51	Stephanie Höhn	0 P.
53	52	Carlotta Hoff	0 P.
54	53	Elena Kurzenmann	0 P.
55	54	Katharina Jansen	0 P.
56	55	Svenja Pohl	0 P.
57	56	Sabrina Lang	0 P.
58	57	Sarah Brethrich	0 P.
59	58	Louise Betz	0 P.
60	59	Myriam Betz	0 P.
61	60	Leonie Wittenstein	0 P.
62	61	Fabienne Hahn	0 P.
63	62	sigrid cacspar	0 P.
64	63	Isabelle Neuhardt	0 P.
65	64	Daniela Glöbner	0 P.
66	65	Conny Stopp	0 P.
67	66	Josaine Prinz	0 P.
68	67	Stephanie Hoffmann	0 P.
69	68	Eileen Klein	0 P.
70	69	Viktoria Lima-Vaz	0 P.
71	70	Elena Weiss	0 P.
72	71	Laura Petri	0 P.
73	2	Hendrik Junkawitsch	99 P.
74	3	Michael Becker	82 P.
75	4	Mika Scherer	79 P.
76	5	Ruven Jost	78 P.
77	6	Steffen Mai	70 P.
78	7	Dennis Machwirth	70 P.
79	8	Kay Uwe Buschauer	67 P.
80	9	Jens Harder	66 P.
81	10	Michael Kasper	64 P.
82	11	Sebastian Rhiemann	63 P.
83	12	Jan Piontek	63 P.
84	13	Paul Mohr	55 P.
85	14	Richard Pfeifer	54 P.
86	15	Albert Marekka	51 P.
87	16	Jan Machwirth	50 P.
88	17	Nikolas Eckert	46 P.
89	18	Sebastian Adam	46 P.
90	19	Jakob Holzer	45 P.
91	20	Dirk Leinenbach	45 P.
92	21	Tobias Lesch	43 P.
93	22	Till Martin	34 P.
94	23	Oskar Ludwig	32 P.
95	24	Adrian Gerling	31 P.
96	25	Philipp Hermes	31 P.
97	26	Julian Einspennner	31 P.
98	27	Hans Dampf	30 P.
99	28	Christian Schmidt	30 P.
100	29	Nareesh Schuler	29 P.
101	30	Bastian Meisberger	28 P.
102	31	Kevin Klesen	24 P.
103	32	Lukas Wahl	24 P.
104	33	The old man Boulderling	23 P.
105	34	Marco Ronconi	21 P.
106	35	Stefan Stornjek	21 P.
107	36	Kevin Küsters	20 P.
108	37	Fabian Revilla	17 P.
109	38	Andreas Leibfried	17 P.
110	39	Joel Schanz	17 P.
111	40	Frank Steinacker	16 P.
112	41	Fynn Simmet	14 P.
113	42	Daniel Gusenburger	14 P.
114	43	Christian Saar	13 P.
115	44	Alexander Dahmen	11 P.
116	45	Andreas Bechthold	11 P.
117	46	David Losert	11 P.
118	47	Lukas Tafeshi	11 P.
119	48	Fabian Loth	11 P.
120	49	Daniel Dietz	10 P.
121	50	Martin Jacobs	10 P.
122	51	Edin Suljic	10 P.
123	52	Till Hedder	10 P.
124	53	Tim Janßen	9 P.
125	54	Oliver Hoyer	9 P.
126	55	Luca Nicodemo	8 P.
127	56	Kevin Leiden	6 P.
128	57	Aljosa Lellig	5 P.
129	58	Marco Ladwein	4 P.
130	59	Oliver Däuber	3 P.
131	60	Jan Federmann	3 P.
132	61	Mike Müller	0 P.
133	62	Cyrrill Wagner	0 P.
134	63	Felix Rozek	0 P.
135	64	Kevin Michel	0 P.
136	65	Savitskiy Olexandr	0 P.
137	66	Demian Dietrich	0 P.
138	67	Claus Kammer	0 P.
139	68	Niklas Braun	0 P.
140	69	Paul Burgert	0 P.
141	70	Savitskiy Timofiy	0 P.
142	71	Bastian Klein	0 P.
143	72	Tobias Elbert	0 P.
144	73	Luca Strauss	0 P.
145	74	Jannik Feld	0 P.
146	75	Tobias Müller	0 P.
147	76	Stephan Pauly	0 P.
148	77	Jakob Bucksch	0 P.
149	78	Johannes Richters	0 P.
150	79	Marvin Stenger	0 P.
151	80	Christopher Jung	0 P.
152	81	Tobias Kalmes	0 P.
153	82	Sebastian Hahn	0 P.
154	83	Max Jakob	0 P.
155	84	Marc Austin	0 P.
156	85	Thomas Lichius	0 P.
157	86	yannick schaefer	0 P.
158	87	Bernhard Speicher	0 P.
159	88	Till Schmelczysch	0 P.
160	89	Frederic Landwehr	0 P.
161	90	Domenik Lauer	0 P.
162	91	Jakob Kneip	0 P.
163	92	Lukas Jedzejewski	0 P.
164	93	Nicolas Landwehr	0 P.
165	94	Gerhard Traut	0 P.
166	95	Marco Dohm	0 P.
167	96	Niklas Brünken	0 P.
168	97	Lars Proske	0 P.
169	98	Julian Becker	0 P.
170	99	Niklas Müller	0 P.
171	00	André Rodriguez	0 P.
172	01	Mark Scherer	0 P.
173	02	Nicolas Welsch	0 P.
174	03	Chris Braeunig	0 P.
175	04	Markus Müller	0 P.
176	05	Marco Lamberti	0 P.
177	06	daniel kempfer	0 P.
178	07	Benedikt Luberto	0 P.
179	08	Jonas Maier	0 P.
180	09	Lukas Ruoss	0 P.
181	10	Peter Krauser	0 P.
182	11	Michael Roth	0 P.
183	12	Sascha Gress	0 P.
184	13	David Groß	0 P.
185	14	Felix Maurer	0 P.
186	15	Fabio Cupelli	0 P.
187	16	Tarnik Weinem	0 P.
188	17	Tobias Pfeil	0 P.
189	18	Johannes Frey	0 P.
190	19	Jan-Philipp Klotz	0 P.
191	20	Yoshua Haas	0 P.
192	21	Julius Auer	0 P.
193	22	Sebastian Dahmen	0 P.
194	23	Nils Siegel	0 P.
195	24	Aaron Schwarz	0 P.
196	25	Felix Francois	0 P.
197	26	Linus Fuchs	0 P.
198	27	Christoph Pauly	0 P.
199	28	Michael Höhn	0 P.
200	29	Levi Höhn	0 P.
201	30	Niklas Flentje	0 P.
202	31	Lars König	0 P.
203	32	Sebastian Kirsch	0 P.
204	33	Sanny Schmitt	0 P.
205	34	Mark Reden	0 P.
206	35	Ralf Müller	0 P.
207	36	Roman Glasman	0 P.
208	37	Fabian Hans	0 P.
209	38	Silas Klein	0 P.
210	39	Christian Egler	0 P.
211	40	Luis Kreutzer	0 P.
212	41	Jonas Schikorski	0 P.
213	42	Tim Engel	0 P.
214	43	Thorsten Müller	0 P.
215	44	Maximilian Klose	0 P.
216	45	Rolf Wilmes	0 P.
217	46	Georg Spaniol	0 P.
218	47	Sean Lipicar	0 P.
219	48	Jan Backes	0 P.
220	49	Andreas Weibert	0 P.
221	50	stefan rodenmayer	0 P.
222	51	Uwe Moik	0 P.
223	52	Matthias Hock	0 P.
224	53	Matthias Stelis	0 P.
225	54	Noel Hahn	0 P.
226	55	Reimar KBA	0 P.
227	56	Lukas Morbe	0 P.
228	57	Ruven Zimmermann	0 P.
229	58	Johannes Mertzt	0 P.
230	59	vincent renonce	0 P.
231	60	Frederic Zenner	0 P.
232	61	Georg Friedrich	0 P.
233	1	Carlotta Barth	85 P.
234	2	Natascha Klicker	56 P.
235	3	Jennifer Soma	46 P.
236	4	Yasmine Briefs	43 P.
237	5	Vanessa Schilling	40 P.
238	6	Tanja Bielei	32 P.
239	7	Leandra Schiller	31 P.
240	8	Ida Reisbeck	30 P.
241	9	Jenny Engel	28 P.
242	10	Hannah Hattenbach	27 P.
243	11	Lena Himbert	24 P.
244	12	Charmaine Utzig	23 P.
245	13	Sabrina Bock	23 P.
246	14	Christel Btown	18 P.
247	15	Emilia Müller	17 P.
248	16	Melissa Hilgert	13 P.
249	17	Lea Oster	11 P.
250	18	Marie Rauber	11 P.
251	19	Annette Baldauf	11 P.
252	20	Joanna Paulus	10 P.
253	21	Alisha Klicker	10 P.
254	22	Sarina Turra	8 P.
255	23	Jennifer Schu	8 P.
256	24	Climbing Is Not Good For You	6 P.
257	25	Yvonne Thome	6 P.
258	26	Josefine Klein	4 P.
259	27	Carolin Leldner	0 P.
260	28	Alice Oberacker	0 P.
261	29	Sabrina Rauber	0 P.
262	30	Eva Lehmit	0 P.
263	31	Marie Lehnert	0 P.
264	32	Pauline Schöner	0 P.
265	33	Celine Brachmann	0 P.
266	34	Lea Löffmann	0 P.
267	35	Diana Hocke	0 P.
268	36	Elli Löffmann	0 P.
269	37	Maren Schöhl	0 P.
270	38	Anke Mautes	0 P.
271	39	Sophia Backes	0 P.
272	40	Jolina Backes	0 P.
273	41	Tabae Frenger	0 P.
274	42	Leoni Jacob-Holderbaum	0 P.
275	43	Elena Scheer	0 P.
276	44	Louisa Hähnel	0 P.
277	45	Katharina Wolf	0 P.
278	46	Kristina Pink	0 P.
279	47	Rebecca Praass	0 P.
280	48	Anita Backes	0 P.
281	49	Phoebe Becker	0 P.
282	50	Jana Fromberg	0 P.
283	51	Marupl Weisgerber	0 P.
284	52	Michelle Reuter	0 P.
285	53	Sophie Cullmann	0 P.
286	1	Patrick Igel	114 P.
287	2	Luca Kaul	112 P.
288	3	Finn Machwirth	103 P.
289	4	Pinky The Kid	97 P.
290	5	David Springborn	94 P.
291	6	Bernd Möller	85 P.
292	7	Jens Maltter	80 P.
293	8	Thomas Schweil	76 P.
294	9	Alexander Blick	76 P.
295	10	Lutz Sängler	61 P.
296	11	Thomas Kraß	58 P.
297	12	Scott Neu	57 P.
298	13	Fynn Kicker	56 P.
299	14	Max Rausch-Dupont	54 P.
300	15	Leon Horras	52 P.
301	16	Christian Wegmann	51 P.
302	17	Ittboowierter assi i hoste	50 P.
303	18	Christian Klaassen	50 P.
304	19	Florian Kamp	48 P.
305	20	Leon Kornfeld	48 P.
306	21	Manuel Recktenwald	47 P.
307	22	Jörg Schallmo	44 P.
308	23	Tili Klein	43 P.
309	24	Hannes Köthner	43 P.
310	25	Felix Steinmetz	41 P.
311	26	Nils Hörnlein	41 P.
312	27	Marcel Lachnit	40 P.
313	28	Bastian Jäger	39 P.
314	29	Florian Maurer	39 P.
315	30	Timo Brückner	37 P.
316	31	Alessandro Diederich	37 P.
317	32	Alexander Baldauf	36 P.
318	33	Eric Spaniol	36 P.
319	34	Nils Lambert	33 P.
320	35	Henry Junk	32 P.
321	36	Johannes Kraß	32 P.
322	37	Javier Matias Echeverria	32 P.
323	38	Maxime Schwenk	32 P.
324	39	Marius Feldt	32 P.
325	40	Felix Smarsly	31 P.
326	41	Chris Brüske	28 P.
327	42	Tobias Böhlen	28 P.
328	43	Luca Lorang	24 P.
329	44	Lukas Buschauer	24 P.
330	45	Benjamin Jochum	24 P.
331	46	Moritz Schnur	22 P.
332	47	Pascal Burkholz	21 P.
333	48	Felix Klicker	20 P.
334	49	Nico Fries	19 P.
335	50	Philipp Neis	19 P.
336	51	Frank Turra	19 P.
337	52	Felix Cullmann	18 P.
338	53	Torsten Fuss	14 P.
339	54	Climb Master Flash Vögelchen	14 P.
340	55	Christan Thawes	14 P.
341	56	Tom Thewes	12 P.
342	57	Sascha Soma	12 P.
343	58	Tim Staß	12 P.
344	59	Pascal Hehn	12 P.
3			