

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 [1]: 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 16:36:15.459124

0 1 2
0 1 Michaela Meisberger 48 P.
1 2 Adelina Hehetror 48 P.
2 3 Tamina Lorenz 39 P.
... ..
75 24 Sandra Rodriguez O.P.
76 24 Elena Weiss O.P.
77 24 Laura Petri O.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 61 vincent renonce O.P.
162 61 Jan Backes O.P.
163 61 Nicolas Welisch O.P.

164 rows x 3 columns

0 1 2
0 1 Carotta Barth 85 P.
1 2 Natascha Kicker 56 P.
2 3 Maren Schohl 46 P.
... ..
55 28 Michelle Reuter O.P.
56 28 Jennifer Soma O.P.
57 28 Sophie Culmann O.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 O.P.
121 67 Felix Steinmetz O.P.
122 67 Samuel Fuchs O.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 O.P.
39 25 Schröder Britta O.P.
40 25 Verena Wittling O.P.

41 rows x 3 columns

0 1 2
0 1 Gregor Lotz 97 P.
1 2 Connor Johann 62 P.
2 3 Elias Weiser 59 P.
... ..
91 38 Pascal Koch O.P.
92 38 Jakob Mertz O.P.
93 38 Alexander Ziebert O.P.

94 rows x 3 columns

Total participants including duplicates = 558
```

```
In [2]: 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', None):
        dataframe = pd.DataFrame(dataframe)
        display.display(dataframe)

# generate .csv files with duplicates per gym / gender(m or f)
l = ['kba_f', 'kba_m', 'rocklands_f', 'rocklands_m', 'olymp_f', 'olymp_m']
i = 0
for df in duplicate_list:
    df.to_csv('{}.csv'.format(i))
    i += 1

folgende Kandidaten können Duplikate sein (30 Stück):

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

0 1 2
78 61 Marco Ladwein O.P.
81 61 Mike Müller O.P.
150 61 Christian Schmidt O.P.

0 1 2
27 28 Hannah Hattenbach O.P.
32 28 Sabrina Rauber O.P.
36 28 Tanja Bleh O.P.
48 28 Annette Baldauf O.P.
56 28 Jennifer Soma O.P.

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

0 1 2

0 1 2
52 38 Alexander Potapov O.P.
53 38 Alexander Potapov O.P.
64 38 Steven Jessel O.P.
66 38 Jens Zimmer O.P.
74 38 Florian Hock O.P.
75 38 Christian Schönauf O.P.
78 38 Fabian Sauerbrey O.P.
79 38 Alexander Potapov O.P.
80 38 Luca Zothor O.P.
86 38 Jens Zimmer O.P.
88 38 Jakob Mertz O.P.
90 38 Tom Kehl O.P.
92 38 Jakob Mertz O.P.
```

```
In [3]: # 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])} \nKBA men = {len(final_tables[1])}')
print(f'Total participants Rocklands = {len(final_tables[2])} + {len(final_tables[3])} \nRocklands women = {len(final_tables[2])} \nRocklands men = {len(final_tables[3])}')
print(f'Total participants Olymp = {len(final_tables[4])} + {len(final_tables[5])} \nolymp women = {len(final_tables[4])} \nolymp men = {len(final_tables[5])}')

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

print('-----')
print(f'Total number of participants = {total_participants}')
dupl = dupl_participants - total_participants
print(f'Difference between totals = {dupl}')
print(f'ypf. fehlendes Geld = {dupl * 5[6]}')

Total participants KBA = 233
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 [4]: # add id as index
df_complete = pd.concat(final_tables)

# generate ids
ids = list(range(0, len(df_complete)))
df_complete.insert(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
0	1	Michaela Meisberger		51 P.
1	2	Adeline Hehretor	48 P.	
2	3	Tamia Lorenz	39 P.	
3	4	Mia Siegel	36 P.	
4	5	Hanna Niedermeyer	28 P.	
5	6	Pauline Rühle	27 P.	
6	7	Nina Stadler	24 P.	
7	8	Nina Engelgkamp	23 P.	
8	9	Anna-Marie Schwarz	22 P.	
9	10	Adriana Altpeter	21 P.	
10	11	Marlene Fries	20 P.	
11	12	Sandra Rodríguez	17 P.	
12	13	Denise Däuber	16 P.	
13	14	Vanessa Daschner	15 P.	
14	15	Theresa Müller	11 P.	
15	15	Julia Wagner	11 P.	
16	17	Amélie Müller Pinto	10 P.	
17	18	Madeleine Birtel	8 P.	
18	19	Paula Mohr	7 P.	
19	19	Louisa Bachhofen	7 P.	
20	21	Amélie Clemens	4 P.	
21	21	Michaela Kremer	4 P.	
22	23	Verena Blank	1 P.	
23	24	Isabelle Kammer	0 P.	
24	24	Paola Goergen	0 P.	
25	24	Marie Wolter	0 P.	
26	24	Katharina Schackmar	0 P.	
27	24	Eva Friedrich	0 P.	
28	24	Lara Köhler	0 P.	
29	24	Valerie Feis	0 P.	
30	24	Stefanie Dohm	0 P.	
31	24	Lena Zimmer	0 P.	
32	24	Malke Zimmer	0 P.	
33	24	Rebecca Kriier	0 P.	
34	24	Jessica Hahn	0 P.	
35	24	Luise Kempf	0 P.	
36	24	Franca Kierfing	0 P.	
37	24	zoe moll	0 P.	
38	24	Daniela Fischer	0 P.	
39	24	Ida Stapf	0 P.	
40	24	Nina Archut	0 P.	
41	24	Cosima Hafner	0 P.	
42	24	Zinzi Sarah Ngumou	0 P.	
43	24	Jessica Kamp	0 P.	
44	24	Lauren naudin	0 P.	
45	24	Matilda Krauser	0 P.	
46	24	Michaels Berger	0 P.	
47	24	Verena Roßbach	0 P.	
48	24	Hannah Mai	0 P.	
49	24	Lea Görg	0 P.	
50	24	Leila Funke	0 P.	
51	24	Stephanie Höhn	0 P.	
52	24	Carlotta Hoff	0 P.	
53	24	Elena Kurtzenmann	0 P.	
54	24	Katharina Jansen	0 P.	
55	24	Svenja Pohl	0 P.	
56	24	Sabrina Lang	0 P.	
57	24	Sarah Brettnich	0 P.	
58	24	Louise Betz	0 P.	
59	24	Myriam Betz	0 P.	
60	24	Leonie Wittenstein	0 P.	
61	24	Fabienne Hahn	0 P.	
62	24	sigrid cacspar	0 P.	
63	24	Isabelle Neuhardt	0 P.	
64	24	Daniela Glöbner	0 P.	
65	24	Conny Stopp	0 P.	
66	24	Josaline Prinz	0 P.	
67	24	Stephanie Hoffmann	0 P.	
68	24	Eileen Klein	0 P.	
69	24	Viktoria Lima-Vaz	0 P.	
70	24	Elena Weiss	0 P.	
71	24	Laura Petri	0 P.	
72	1	Hendrik Junkawitsch	99 P.	
73	2	Michael Becker	82 P.	
74	3	Mika Scherer	79 P.	
75	4	Ruven Jost	78 P.	
76	5	Kay Uwe Buschauer	77 P.	
77	6	Dennis Machwirth	70 P.	
78	6	Steffen Mai	70 P.	
79	8	Jens Harder	66 P.	
80	9	Michael Kasper	64 P.	
81	10	Sebastian Rhiemann	63 P.	
82	11	Jan Piontek	63 P.	
83	12	Paul Mohr	55 P.	
84	13	Richard Pfeiffer	54 P.	
85	14	Albert Marekka	51 P.	
86	15	Jan Machwirth	50 P.	
87	16	Nikolas Eckert	46 P.	
88	16	Sebastian Adam	46 P.	
89	18	Jakob Holzer	45 P.	
90	18	Dirk Leinenbach	45 P.	
91	20	Tobias Lesch	43 P.	
92	21	Till Martin	34 P.	
93	22	Oskar Ludwig	32 P.	
94	23	Philipp Hermes	31 P.	
95	23	Adrian Gerling	31 P.	
96	23	Julian Einspennner	31 P.	
97	26	Hans Dampf	30 P.	
98	26	Christian Schmidt	30 P.	
99	28	Naresh Schuler	29 P.	
100	29	Bastian Meisberger	28 P.	
101	30	Lukas Wahl	24 P.	
102	30	Kevin Klesen	24 P.	
103	32	The old man Boulderling	23 P.	
104	33	Frederic Zenner	22 P.	
105	34	Marco Ronconi	21 P.	
106	34	Stefan Stonjek	21 P.	
107	36	Kevin Küsters	20 P.	
108	37	Alexander Dahmen	19 P.	
109	38	Andreas Leibfried	17 P.	
110	38	Fabian Revilla	17 P.	
111	38	Joel Schanz	17 P.	
112	41	Frank Steinacker	16 P.	
113	42	Oliver Däuber	15 P.	
114	43	Fynn Simmet	14 P.	
115	43	Daniel Gusenburger	14 P.	
116	45	Christian Saar	13 P.	
117	46	David Losert	11 P.	
118	46	Andreas Bechthold	11 P.	
119	46	Fabian Loth	11 P.	
120	46	Lukas Tafreshi	11 P.	
121	50	Martin Jacobs	10 P.	
122	50	Daniel Dietz	10 P.	
123	50	Edin Suljic	10 P.	
124	50	Till Heidler	10 P.	
125	54	Tim Janßen	9 P.	
126	54	Oliver Bürtel	9 P.	
127	56	Luca Nicodemio	8 P.	
128	57	Kevin Leinen	6 P.	
129	58	Aljosha Lellin	5 P.	
130	59	Marco Ladwein	4 P.	
131	60	Jan Federmann	3 P.	
132	61	Tobias Elbert	0 P.	
133	61	Mike Müller	0 P.	
134	61	Felix Rozek	0 P.	
135	61	Cyрил Wagner	0 P.	
136	61	Bastian Klein	0 P.	
137	61	Niklas Braun	0 P.	
138	61	Paul Bungert	0 P.	
139	61	Savitsky Timofiy	0 P.	
140	61	Luca Strauss	0 P.	
141	61	Kevin Michel	0 P.	
142	61	Savitsky Olexander	0 P.	
143	61	Jannik Feld	0 P.	
144	61	Claus Kammer	0 P.	
145	61	Denlian Dietrich	0 P.	
146	61	Tobias Müller	0 P.	
147	61	Till Schmeltzysch	0 P.	
148	61	Jakob Bucksch	0 P.	
149	61	Johannes Richters	0 P.	
150	61	Marvin Stenger	0 P.	
151	61	Christopher Jung	0 P.	
152	61	Tobias Kalmes	0 P.	
153	61	Sebastian Hahn	0 P.	
154	61	Max Jakob	0 P.	
155	61	Marc Austin	0 P.	
156	61	Thomas Lilius	0 P.	
157	61	yannick schaefer	0 P.	
158	61	Bernhard Speicher	0 P.	
159	61	Frederic Landwehr	0 P.	
160	61	Stephan Pauly	0 P.	
161	61	Domenik Lauer	0 P.	
162	61	Jakob Kneip	0 P.	
163	61	Lukas Jedrzejewski	0 P.	
164	61	Nicolas Landwehr	0 P.	
165	61	Gerhard Traut	0 P.	
166	61	Marco Dohm	0 P.	
167	61	Niklas Brünken	0 P.	
168	61	Lars Proske	0 P.	
169	61	Julian Becker	0 P.	
170	61	Niklas Müller	0 P.	
171	61	André Rodriguez	0 P.	
172	61	Mark Scherer	0 P.	
173	61	Chris Braeunig	0 P.	
174	61	Marius Müller	0 P.	
175	61	Marco Lamberti	0 P.	
176	61	daniel kempfer	0 P.	
177	61	Benedikt Lübert	0 P.	
178	61	Jonas Maier	0 P.	
179	61	Lukas Ruoss	0 P.	
180	61	Peter Krauser	0 P.	
181	61	Fynn Krauser	0 P.	
182	61	Michael Roth	0 P.	
183	61	Sascha Gress	0 P.	
184	61	David Groß	0 P.	
185	61	Felix Maurer	0 P.	
186	61	Fabio Cupelli	0 P.	
187	61	Tarnik Weinem	0 P.	
188	61	Tobias Pfeil	0 P.	
189	61	Johannes Frey	0 P.	
190	61	Jan-Philipp Klotz	0 P.	
191	61	Yoshua Haas	0 P.	
192	61	Julius Auer	0 P.	
193	61	Sebastian Dahmen	0 P.	
194	61	Nils Siegel	0 P.	
195	61	Aaron Schwarz	0 P.	
196	61	Felix Francois	0 P.	
197	61	Linus Fuchs	0 P.	
198	61	Christoph Pauly	0 P.	
199	61	Michael Höhn	0 P.	
200	61	Levi Höhn	0 P.	
201	61	Niklas Flentje	0 P.	
202	61	Lars König	0 P.	
203	61	Sebastian Kirsch	0 P.	
204	61	Sanny Schmitt	0 P.	
205	61	Ralf Müller	0 P.	
206	61	Roman Glasman	0 P.	
207	61	Fabian Hans	0 P.	
208	61	Silas Klein	0 P.	
209	61	Christian Egler	0 P.	
210	61	Luis Kreuzer	0 P.	
211	61	Jonas Schikorski	0 P.	
212	61	Tim Engel	0 P.	
213	61	Thorsten Müller	0 P.	
214	61	Maximilian Klose	0 P.	
215	61	Rolf Wilmes	0 P.	
216	61	Georg Spaniol	0 P.	
217	61	Georg Friedrich	0 P.	
218	61	Mark Reden	0 P.	
219	61	Sean Lipicar	0 P.	
220	61	Andreas Weilert	0 P.	
221	61	stefan rodenmayer	0 P.	
222	61	Uwe Moik	0 P.	
223	61	Matthias Hock	0 P.	
224	61	Matthias Stelis	0 P.	
225	61	Noel Hahn	0 P.	
226	61	Reimar KBA	0 P.	
227	61	Lukas Morbe	0 P.	
228	61	Ruven Zimmermann	0 P.	
229	61	Johannes Mertzt	0 P.	
230	61	vincent renonce	0 P.	
231	61	Jan Backes	0 P.	
232	61	Nicolas Weisch	0 P.	
233	1	Carlotta Barth	85 P.	
234	2	Natascha Klicker	56 P.	
235	3	Maren Schöhl	46 P.	
236	3	Jennifer Soma	46 P.	
237	5	Yasmine Briefs	43 P.	
238	6	Vanessa Schilling	40 P.	
239	7	Tanja Bielei	32 P.	
240	8	Leandra Schiller	31 P.	
241	9	Ida Reisbeck	30 P.	
242	10	Jenny Engel	28 P.	
243	11	Hannah Hattenbach	27 P.	
244	12	Lena Himbert	24 P.	
245	13	Sabrina Bock	23 P.	
246	13	Charmaine Utzig	23 P.	
247	15	Christel Btown	18 P.	
248	16	Emilia Müller	17 P.	
249	17	Melissa Hilgert	13 P.	
250	18	Annette Baldauf	11 P.	
251	18	Marie Rauber	11 P.	
252	18	Lea Oster	11 P.	
253	21	Joanna Paulus	10 P.	
254	21	Alisha Klicker	10 P.	
255	23	Jennifer Schu	8 P.	
256	23	Sarina Turra	8 P.	
257	25	Yvonne Thome	6 P.	
258	25	Climbing Is Not Good For You	6 P.	
259	27	Josefina Klein	4 P.	
260	28	Alice Oberacker	0 P.	
261	28	Sabrina Rauber	0 P.	
262	28	Eva Schmitt	0 P.	
263	28	Carolin Leidner	0 P.	
264	28	Marie Lehnert	0 P.	
265	28	Elli Lübmann	0 P.	
266	28	Lea Lübmann	0 P.	
267	28	Diana Hocke	0 P.	
268	28	Celine Brachmann	0 P.	
269	28	Paulina Zöhler	0 P.	
270	28	Anke Mautes	0 P.	
271	28	Sophia Backes	0 P.	
272	28	Jolina Backes	0 P.	
273	28	Tabae Frenger	0 P.	
274	28	Leoni Jacob-Holderbaum	0 P.	
275	28	Elena Scheer	0 P.	
276	28	Louisa Hähnel	0 P.	
277	28	Katharina Wolf	0 P.	
278	28	Kristina Pink	0 P.	
279	28	Rebecca Prass	0 P.	
280	28	Anita Backes	0 P.	
281	28	Phoebe Becker	0 P.	
282	28	Jana Fromberg	0 P.	
283	28	Marupl Weisgerber	0 P.	
284	28	Michelle Reuter	0 P.	
285	28	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	8	Alexander Blick	76 P.	
295	10	Max Rausch-Dupont	74 P.	
296	11	Lutz Sänger	61 P.	
297	12	Jörg Schallmo	60 P.	
298	13	Thomas Kraft	58 P.	
299	14	Scott Neu	57 P.	
300	15	Fynn Klicker	56 P.	
301	15	Florian Kamp	56 P.	
302	17	Leon Horras	52 P.	
303	18	Christian Wegmann	51 P.	
304	19	Christian Klaassen	50 P.	
305	19	Hättoewierter assi i' hoste	50 P.	
306	21	Leon Kornfeld	48 P.	
307	22	Manuel Recktenwald	47 P.	
308	23	Til Klein	43 P.	
309	23	Hannes Köther	43 P.	
310	25	Felix Steinmetz		