

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

## Abstract

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

Motivation: Glaube, 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 ausgelassen, d.h. duplikate Namen einer Kletterhalle werden herausgeschnitten, da angenommen wird, dass niemand doppelt heißt.

Da dies aufgrund der Wahrscheinlichkeit vertretbar.

## Main Points

Es gibt 44 Duplikate -> irgendwo ist das System fehlerhaft

Möglicherweise vermeintlich fehlendes Geld: 220€

```
In [4]: import pandas as pd
        from IPython import display

        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}')
```

Total participants including duplicates = 532

```
In [5]: final_tables = []
        for tables in table:
            temp = tables.drop(labels=[0, 2], axis=1)
            temp = temp.drop_duplicates()
            final_tables.append(temp)
```

```
In [6]: # 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('-----')
        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('-----')
        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}')
        print(f'Difference between totals = {dupl_participants - total_participants}')
```

Total participants KBA = 220

KBA women = 68

KBA men = 152

-----

Total participants Rocklands = 159

Rocklands women = 51

Rocklands men = 108

-----

Total participants Olymp = 109

Olymp women = 38

Olymp men = 71

-----

Total number of participants = 488

Difference between totals = 44