

Suomen Parhaat Boulderit 2018: Create Boulders Template

March 17, 2018

Load the datafile `survey_-_cleaned.csv`, which contains the form responses to the **Suomen Parhaat Boulderit 2018** survey.

```

In [1]: import pandas as pd
import numpy as np

# Load cleaned dataset
spb2018_df = pd.read_csv("data/survey_-_cleaned.csv")

# Drop duplicates (exclude the Timestamp column from comparisons)
spb2018_df = spb2018_df.drop_duplicates(subset=spb2018_df.columns.values
.tolist()[1:])
spb2018_df.head()

```

Out[1]:

	Timestamp	Suostumus	Ikä	Sukupuoli	Pituus (cm)	Vaikein *ulkona* kiipeämäsi boulder	Boulderin nimi	Olen lähettänyt (kiivennyt) kyseisen boulderin	Ku
0	12/17/2017 18:36:32	Kyllä	35 - 39	Mies	184	7A+	Muistipeli	Kyllä	erilai kolm pitkä
1	12/17/2017 18:48:44	Kyllä	30 - 34	Mies	180	7B	Voodoo	Kyllä	Tiuk hänt
2	12/17/2017 18:49:04	Kyllä	30 - 34	Mies	180	8A	One love	Kyllä	Dyna
3	12/17/2017 18:53:03	Kyllä	25 - 29	Mies	190	7A	Bitch slap and male pinch	Kyllä	kuur mieh
4	12/17/2017 19:10:14	Kyllä	25 - 29	Mies	180	8A	Maitomies	Kyllä	Herk klas

5 rows × 29 columns

Create boulders template file boulders_-_template.csv.

```
In [2]: def create_boulders_template():
        boulder_name_columns = [spb2018_df["Boulderin nimi"], spb2018_df["Boulderin nimi.1"], spb2018_df["Boulderin nimi.2"]]
        unique_boulder_names_s = pd.concat(boulder_name_columns, ignore_index=True).dropna().drop_duplicates().sort_values().reset_index(drop=True)
        unique_boulder_names_s.to_csv("data/boulders_-_template.csv", index=False)

create_boulders_template()
```