info

September 30, 2021

[2]:
```
import pandas as pd
import numpy as np
```

[4]:
```
df = pd.read_csv("data.csv")
df
```

[4]:

	user	condition	character	meaning	success	teacher	session	\
0	0	myopic		miss	True	leitner	0	
1	0	myopic		older sister	True	leitner	0	
2	0	myopic		marriage	True	leitner	0	
3	0	myopic		miss	True	leitner	0	
4	0	myopic		older sister	True	leitner	0	
…	…	…	…	…	…	…	…	
77631	52	myopic		bright	True	myopic	6	
77632	52	myopic		duplicate	False	myopic	6	
77633	52	myopic		dance	True	myopic	6	
77634	52	myopic		escape	True	myopic	6	
77635	52	myopic		based on	True	myopic	6	

	is_eval	ts_display	ts_reply	option0	option1	\
0	False	1.601488e+09	1.601488e+09	cape	cork	
1	False	1.601488e+09	1.601488e+09	age	aggression	
2	False	1.601488e+09	1.601488e+09	comprehend	long	
3	False	1.601488e+09	1.601488e+09	bright	fat	
4	False	1.601488e+09	1.601488e+09	aim	marriage	
…	…	…	…	…	…	
77631	True	1.602136e+09	1.602136e+09	bright	knee	
77632	True	1.602136e+09	1.602136e+09	boil	cultivation	
77633	True	1.602136e+09	1.602136e+09	continue	dance	
77634	True	1.602136e+09	1.602136e+09	cultivation	escape	
77635	True	1.602136e+09	1.602136e+09	based on	cultivation	

	option2	option3	option4	option5
0	miss	stiff	strong	universal
1	conflict	measurement	miss	older sister
2	marriage	miss	older sister	reparation
3	heat	marriage	miss	older sister
4	miss	older sister	radiance	snow

```
…          …           …           …             …
77631        run         sew       steal  western style
77632        dry   duplicate        rich           sew
77633        hit     machine   set aside        winter
77634  generally        lump         run  western style
77635       five     machine         odd        salary

[77636 rows x 16 columns]
```

[6]:
```
df_demo = pd.read_csv("demographic_info.csv")
df_demo
```

[6]:
```
    user gender  age    native_lang  \
0      0      M   19        finnish
1      1      M   30           urdu
2      2      F   26        finnish
3      3      M   20        finnish
4      4      F   20        finnish
5      5      F   29        finnish
6      6      F   23        swedish
7      7      F   19        swedish
8      8      F   25        finnish
9      9      F   27        finnish
10    10      F   20        finnish
11    11      F   26        finnish
12    12      F   27        finnish
13    13      F   30        finnish
14    14      F   28        finnish
15    15      M   34        finnish
16    16      F   29        finnish
17    17      F   19        swedish
18    18      M   27        finnish
19    19      F   27        finnish
20    20      F   26        finnish
21    21      M   23        finnish
22    22      M   30        english
23    23      F   25        finnish
24    24      F   21        finnish
25    25      F   26        finnish
26    26      M   20         polish
27    27      M   25        finnish
28    28      F   21        russian
29    29      F   28        finnish
30    30      F   19        finnish
31    31      O   40        finnish
32    32      F   22        finnish
33    33      M   26        spanish
```

```
34    34      F   20        finnish
35    35      F   21        finnish
36    36      F   22        finnish
37    37      F   22        finnish
38    38      F   20        finnish
39    39      F   29        finnish
40    40      F   20        finnish
41    41      F   44        finnish
42    42      F   31        finnish
43    43      F   23        finnish
44    44      F   19        finnish
45    45      M   20        finnish
46    46      M   20        swedish
47    47      F   39        finnish
48    48      F   26        finnish
49    49      F   34         french
50    50      F   56        spanish
51    51      F   50        spanish
52    52      F   25  spanish dutch
```

```
                                        other_lang
0                           english swedish german
1                                          english
2                            english swedish german
3                           english french swedish
4                                  english swedish
5                                          english
6                                  finnish english
7                           finnish english french
8      english swedish german russian french estonian
9              english italian swedish spanish french
10                                         english
11                                  english swedish
12                    english swedish french spanish
13                          english russian swedish
14                   english swedish spanish german
15            swedish english german spanish russian
16           english swedish french spanish estonian
17                     finnish english french korean
18       english swedish german russian spanish dutch
19                          english swedish russian
20                                 english swedish
21                                  english swedish
22                                         finnish
23                          swedish english spanish
24                          english swedish spanish
25                                  english swedish
```

```
26                            english spanish finnish
27                                     english french
28                             finnish english french
29                                     english swedish
30                      english swedish french spanish
31  swedish english german russian lithuanian port…
32                              swedish english french
33                               english german french
34                              swedish english german
35              english german swedish russian italian
36                                     english swedish
37                             english swedish spanish
38                             english italian swedish
39                       english swedish italian french
40                        english swedish german french
41  english swedish french italian german spanish …
42                             english swedish spanish
43                              english swedish korean
44                              english swedish german
45                                             english
46                             finnish english spanish
47                       english swedish german korean
48                             english swedish spanish
49  finnish english spanish russian swedish german…
50                                             english
51                                             english
52                                             english
```

[ ]:
```python
# Copy actual item ID in a new column
df["item_id"] = pd.factorize(df.character)[0]
# Create new ids starting from zero
for i, i_id in enumerate(df.item_id.unique()):
    df.loc[df.item_id == i_id, 'item'] = i

# Total number of user
n_u = len(df.user.unique())

# Number of observations per user
n_o_by_u = np.zeros(shape=n_u, dtype=int)
for u, (user, user_df) in enumerate(df.groupby("user")):
    # Do not count first presentation
    n_o_by_u[u] = len(user_df) - len(user_df.item.unique())

# Total number of observation
n_obs = n_o_by_u.sum()

# Replies (1: success, 0: error)
```

```
y = np.zeros(shape=n_obs, dtype=int)
# Time elapsed since the last presentation of the same item (in seconds)
x = np.zeros(shape=n_obs, dtype=float)
# Number of repetition (number of presentation - 1)
r = np.zeros(shape=n_obs, dtype=int)
# Item ID
w = np.zeros(shape=n_obs, dtype=int)
# User ID
u = np.zeros(shape=n_obs, dtype=int)

# Fill the containers `y`, `x`, `r`, `w`, `u`
idx = 0
for i_u, (user, user_df) in enumerate(df.groupby("user")):

    # Extract data from user `u`
    user_df = user_df.sort_values(by="ts_reply")
    seen = user_df.item.unique()
    w_u = user_df.item.values
    ts_u = user_df.ts_reply.values
    y_u = user_df.success.values

    # Initialize counts of repetition for each words at -1
    counts = {word: -1 for word in seen}
    # Initialize time of last presentation at None
    last_pres = {word: None for word in seen}

    # Number of observations for user `u` including first presentations
    n_obs_u_incl_first = len(user_df)

    # Number of repetitions for user `u`
    r_u = np.zeros(n_obs_u_incl_first)
    # Time elapsed since last repetition for user `u`
    x_u = np.zeros(n_obs_u_incl_first)

    # Loop over each entry for user `u`:
    for i in range(n_obs_u_incl_first):

        # Get info for iteration `i`
        word = w_u[i]
        ts = ts_u[i]
        r_u[i] = counts[word]

        # Compute time elasped since last presentation
        if last_pres[word] is not None:
            x_u[i] = ts - last_pres[word]

        # Update count of repetition
```

```
        counts[word] += 1
        # Update last presentation
        last_pres[word] = ts

    # Keep only observations that are not the first presentation of an item
    to_keep = r_u >= 0
    y_u = y_u[to_keep]
    r_u = r_u[to_keep]
    w_u = w_u[to_keep]
    x_u = x_u[to_keep]

    # Number of observations for user `u` excluding first presentations
    n_obs_u = len(y_u)

    # Fill containers
    y[idx:idx + n_obs_u] = y_u
    x[idx:idx + n_obs_u] = x_u
    w[idx:idx + n_obs_u] = w_u
    r[idx:idx + n_obs_u] = r_u
    u[idx:idx + n_obs_u] = i_u

    # Update index
    idx += n_obs_u

n_w = len(np.unique(w))
n_o_max = n_o_by_u.max()
n_o_min = n_o_by_u.min()
print("number of user", n_u)
print("number of items", n_w)
print("total number of observations (excluding first presentation)", n_obs)
print("minimum number of observation for a single user", n_o_min)
print("maximum number of observation for a single user", n_o_max)

pd.DataFrame({
    'u': u,  # User ID
    'w': w,  # Item ID
    'x': x,  # Time elapsed since the last presentation of the same item (in␣
 ↪seconds)
    'r': r,  # Number of repetition (number of presentation - 1)
    'y': y   # Replies (0: error, 1: success)
})
```