

merge-multiple-csv-files-with-python

December 11, 2022

```
[1]: import os
import glob
import pandas as pd
data_folder = './nhanes_output_data/outputfilestogether'
os.chdir(data_folder)
```

```
[4]: fileTypesToMerge = 'Albumin' # Dem Albumin or Dietary
```

```
[5]: extension = 'csv';
all_filenames = [i for i in glob.glob('*' + fileTypesToMerge + '.*{ }'.
    ↪format(extension))]
all_filenames
```

```
[5]: ['2011-2012-3_Albumin & Creatinine - Urine-ALB_CR_G.XPT.csv',
'2013-2014-3_Albumin & Creatinine - Urine_ALB_CR_H.XPT.csv',
'2015-2016-3_Albumin & Creatinine - Urine-ALB_CR_I.XPT.csv',
'2017-2018-3_Albumin & Creatinine - Urine-ALB_CR_J.XPT.csv']
```

```
[6]: row_total_count = 0
for f in all_filenames:
    df_s = pd.read_csv(f)
    row_total_count += df_s.shape[0]

row_total_count
```

```
[6]: 32660
```

```
[7]: for f in all_filenames:
    df = pd.read_csv(f)
    print(f);
    print(df.head(1));
```

```
2011-2012-3_Albumin & Creatinine - Urine-ALB_CR_G.XPT.csv
                                SEQN                                URXUMA \
0  SEQN - Respondent sequence number  URXUMA - Albumin, urine (ug/mL)

                                URXUMS                                URXUCR \
0  URXUMS - Albumin, urine (mg/L)  URXUCR - Creatinine, urine (mg/dL)
```

```

                                URXCRS \
0  URXCRS - Creatinine, urine (umol/L)

                                URDACT  Unnamed: 6
0  URDACT - Albumin creatinine ratio (mg/g)      NaN
2013-2014-3_Albumin & Creatinine - Urine_ALB_CR_H.XPT.csv
                                SEQN                                URXUMA \
0  SEQN - Respondent sequence number  URXUMA - Albumin, urine (ug/mL)

                                URXUMS                                URXUCR \
0  URXUMS - Albumin, urine (mg/L)  URXUCR - Creatinine, urine (mg/dL)

                                URXCRS \
0  URXCRS - Creatinine, urine (umol/L)

                                URDACT  Unnamed: 6
0  URDACT - Albumin creatinine ratio (mg/g)      NaN
2015-2016-3_Albumin & Creatinine - Urine-ALB_CR_I.XPT.csv
                                SEQN                                URXUMA \
0  SEQN - Respondent sequence number  URXUMA - Albumin, urine (ug/mL)

                                URDUMALC                                URXUMS \
0  URDUMALC - Albumin, urine comment code  URXUMS - Albumin, urine (mg/L)

                                URXUCR \
0  URXUCR - Creatinine, urine (mg/dL)

                                URDUCRLC \
0  URDUCRLC - Creatinine, urine comment code

                                URXCRS \
0  URXCRS - Creatinine, urine (umol/L)

                                URDACT  Unnamed: 8
0  URDACT - Albumin creatinine ratio (mg/g)      NaN
2017-2018-3_Albumin & Creatinine - Urine-ALB_CR_J.XPT.csv
                                SEQN                                URXUMA \
0  SEQN - Respondent sequence number  URXUMA - Albumin, urine (ug/mL)

                                URXUMS                                URDUMALC \
0  URXUMS - Albumin, urine (mg/L)  URDUMALC - Albumin, urine comment code

                                URXUCR                                URXCRS \
0  URXUCR - Creatinine, urine (mg/dL)  URXCRS - Creatinine, urine (umol/L)

                                URDUCRLC \
0  URDUCRLC - Creatinine, urine comment code

```

```

                                URDACT  Unnamed: 8
0  URDACT - Albumin creatinine ratio (mg/g)          NaN

```

```
[14]: os.listdir('./')
```

```
[14]: ['2011-2012-0_Demographic_Variables_Sample_Weights_DEMO_G.XPT.csv',
      '2011-2012-1_Dietary Interview - Individual Foods, First Day-DR1IFF_G.XPT.csv',
      '2011-2012-2_Dietary Interview - Individual Foods, Second Day-DR2IFF_G.XPT.csv',
      '2011-2012-3_Albumin & Creatinine - Urine-ALB_CR_G.XPT.csv',
      '2013-2014-0_Demographic Variables and Sample Weights-DEMO_H.XPT.csv',
      '2013-2014-1_Dietary Interview - Individual Foods, First Day-DR1IFF_H.XPT.csv',
      '2013-2014-2_Dietary Interview - Individual Foods, Second Day-DR2IFF_H.XPT.csv',
      '2013-2014-3_Albumin & Creatinine - Urine-ALB_CR_H.XPT.csv',
      '2015-2016-0_Demographic Variables and Sample Weights-DEMO_I.XPT.csv',
      '2015-2016-1_Dietary Interview - Individual Foods, First Day-DR1IFF_I.XPT.csv',
      '2015-2016-2_Dietary Interview - Individual Foods, First Day-DR2IFF_I.XPT.csv',
      '2015-2016-3_Albumin & Creatinine - Urine-ALB_CR_I.XPT.csv',
      '2017-2018-0_demographics-DEMO_J.XPT.csv',
      '2017-2018-1_Dietary Interview - Individual Foods, First Day-DR1IFF_J.XPT.csv',
      '2017-2018-2_Dietary Interview - Individual Foods, Second Day-DR2IFF_J.XPT.csv',
      '2017-2018-3_Albumin & Creatinine - Urine-ALB_CR_J.XPT.csv',
      'all-merged.csv']
```

```
[16]: #combine all files in the list
combined_csv = pd.concat([pd.read_csv(f) for f in all_filenames])

#export to csv
combined_csv.to_csv( "../merged_files/" + fileTypeToMerge + "-all-merged.csv",
                    index=False, encoding='utf-8-sig')
```

```
[17]: df = pd.read_csv('../all-merged.csv')
df.tail()

df.head()
```

```
[17]:
```

	SEQN	URXUMA \
0	SEQN - Respondent sequence number	URXUMA - Albumin, urine (ug/mL)
1	62161.0	5.2
2	62163.0	21.4
3	62164.0	14.2
4	62165.0	25.5

	URXUMS	URXUCR \
--	--------	----------

	URXUMS - Albumin, urine (mg/L)	URXUCR - Creatinine, urine (mg/dL)
0		
1	5.2	134.0
2	21.4	182.0
3	14.2	166.0
4	25.5	41.0

	URXCRS \
0	URXCRS - Creatinine, urine (umol/L)
1	11845.6
2	16088.8
3	14674.4
4	3624.4

	URDACT	Unnamed: 6	URDUMALC	URDUCRLC \
0	URDACT - Albumin creatinine ratio (mg/g)	NaN	NaN	NaN
1	3.88	NaN	NaN	NaN
2	11.76	NaN	NaN	NaN
3	8.55	NaN	NaN	NaN
4	62.2	NaN	NaN	NaN

	Unnamed: 8
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

```
[18]: df.shape
```

```
[18]: (32660, 10)
```

```
[ ]:
```