

23. Pandas - Part 6

October 28, 2022

```
[ ]: import pandas as pd
```

1 How to randomly sample rows from a DataFrame?

```
[ ]: ufo_df = pd.read_csv("../data/ufo.csv")
```

```
[ ]: # sample 3 rows from the DataFrame without replacement (new in pandas 0.16.1)  
ufo_df.sample(n=3)
```

```
[ ]: # use the 'random_state' parameter for reproducibility  
ufo_df.sample(n=3, random_state=1)
```

```
[ ]: # sample 75% of the DataFrame's rows without replacement  
train = ufo_df.sample(frac=0.75, random_state=99)
```

```
[ ]: ufo_df.shape
```

```
[ ]: train.shape
```

```
[ ]: train
```

```
[ ]: train.index
```

```
[ ]: ufo_df.index
```

```
[ ]: ufo_df.index.isin(train.index)
```

```
[ ]: # store the remaining 25% of the rows in another DataFrame  
test = ufo_df.loc[~ufo_df.index.isin(train.index), :]
```

```
[ ]: test
```

```
[ ]: ufo_df
```

```
[ ]: ufo_df.sample(frac=1, random_state=99) # To Shuffle all rows , use frac=1
```

2 Display Options

```
[ ]: # read a dataset of alcohol consumption into a DataFrame
drinks_df = pd.read_csv('../data/drinks.csv')
```

```
[ ]: # only 10 rows will be displayed when printing
drinks_df
```

```
[ ]: # check the current setting for the 'max_rows' option
pd.get_option('display.max_rows')
```

```
[ ]: # overwrite the current setting so that all rows will be displayed
pd.set_option('display.max_rows', 200)
```

```
[ ]: drinks_df
```

```
[ ]: # reset the 'max_rows' option to its default
pd.reset_option('display.max_rows')
```

```
[ ]: drinks_df
```

```
[ ]: # the 'max_columns' option is similar to 'max_rows'
pd.get_option('display.max_columns')
```

```
[ ]: # read the training dataset from Kaggle's Titanic competition into a DataFrame
train_df = pd.read_csv('../data/titanic_train.csv')
```

```
[ ]: train_df.head()
```

```
[ ]: # an ellipsis is displayed in the 'Name' cell of row 1 because of the
    ↪ 'max_colwidth' option
pd.get_option('display.max_colwidth')
```

```
[ ]: # overwrite the current setting so that more characters will be displayed
pd.set_option('display.max_colwidth', 1000)
```

```
[ ]: train_df.head()
```

```
[ ]: # overwrite the 'precision' setting to display 2 digits after the decimal point
    ↪ of 'Fare'
pd.set_option('display.precision', 2)
```

```
[ ]: train_df.head()
```

3 Save Dataframe to a file

```
[ ]: # save train dataframe to a csv file
train_df.to_csv("../data/1_train_data.csv", index=False)
```

4 Pandas User Guide

The Pandas User Guide covers all of pandas by topic area. Each of the subsections introduces a topic (such as `working with missing data`), and discusses how pandas approaches the problem, with many examples throughout.

[Click here](#) for pandas user guide.