# 0109_pickle_data_frame

January 30, 2019

## 1 Saving a Panda DataFrame with 'pickle'

Sometimes a DataFrame may have content in it that will not save well in text (e.g. csv) format. For example a DataFrame may contain lists, and these will be saved as a text string in a text format.

Python has a library (pickle) for saving Python objects intact so that they may saved and loaded without having to generate them again.

Pandas has built in 'pckling' capability which makes it very easy to save and load intact dataframes.

Let's first generate a dataframe that contains lists.

```
In [1]: import pandas as pd

        my_df = pd.DataFrame()

        names = ['Bob', 'Sam', 'Jo', 'Bill']

        favourite_sports = [['Tennis', 'Motorsports'],
                            ['Football', 'Rugby', 'Hockey'],
                            ['Table tennis', 'Swimming', 'Athletics'],
                            ['Eating cheese']]

        my_df['name'] = names
        my_df['favourite_sport'] = favourite_sports

        print(my_df)
```

```
   name                       favourite_sport
0   Bob                [Tennis, Motorsports]
1   Sam            [Football, Rugby, Hockey]
2    Jo  [Table tennis, Swimming, Athletics]
3  Bill                       [Eating cheese]
```

### 1.1 Save and load DataFrame using Pandas built in pickle methods (recommended)

```
In [2]: # Save DataFrame to pickle object
        my_df.to_pickle('test_df.p')
```

```
        # Load DataFrame with pickle object
        test_df_load_1 = pd.read_pickle('test_df.p')

        print (test_df_load_1)
```

```
   name                        favourite_sport
0   Bob                  [Tennis, Motorsports]
1   Sam              [Football, Rugby, Hockey]
2    Jo  [Table tennis, Swimming, Athletics]
3  Bill                        [Eating cheese]
```

### 1.2 Save and load DataFrame using standard Python pickle library

With DataFrames you will probably always want to use the df.ti_pickle and pd.read_pickle methods for ease. But below is an example of using the Python pickle library - this method can be used with other types of complex Python objects (such as trained machine learning models) as well.

```
In [8]: import pickle

        # Save using pickle
        # (the b in rb denotes binary mode which is required for more complex obecjtes)
        filename = 'pickled_df.p'
        with open(filename, 'wb') as filehandler:
            pickle.dump(my_df, filehandler)

        # Load using pickle
        filename = 'pickled_df.p'
        with open(filename, 'rb') as filehandler:
            reloaded_df = pickle.load(filehandler)

        print (reloaded_df)
```

```
   name                        favourite_sport
0   Bob                  [Tennis, Motorsports]
1   Sam              [Football, Rugby, Hockey]
2    Jo  [Table tennis, Swimming, Athletics]
3  Bill                        [Eating cheese]
```