3_missing_value_exercise

May 20, 2019

```python
In [ ]: import pandas as pd
        import numpy as np
```

```python
In [ ]: # Eaxmple from - https://chrisalbon.com/python/pandas_missing_data.html
        raw_data = {'first_name': ['Jason', np.nan, 'Tina', 'Jake', 'Amy'],
                'last_name': ['Miller', np.nan, 'Ali', 'Milner', 'Cooze'],
                'age': [42, np.nan, 36, 24, 73],
                'sex': ['m', np.nan, 'f', 'm', 'f'],
                'preTestScore': [4, np.nan, np.nan, 2, 3],
                'postTestScore': [25, np.nan, np.nan, 62, 70]}
        df = pd.DataFrame(raw_data, columns = ['first_name', 'last_name', \
                                        'age', 'sex', 'preTestScore', 'postTestScore'])
        df
```

```python
In [ ]: df.isnull().sum()
```

```python
In [ ]: df.isnull().sum() / len(df)
```

```python
In [ ]: df_no_missing = df.dropna()
        df_no_missing
```

```python
In [ ]: df
```

```python
In [ ]: df_cleaned = df.dropna(how='all')
        df_cleaned
```

```python
In [ ]: df['location'] = np.nan
        df
```

```python
In [ ]: df.dropna(axis=1, how='all')
```

```python
In [ ]: df
```

```python
In [ ]: df.dropna(axis=0, thresh=1)
```

```python
In [ ]: df.dropna(thresh=5)
```

```python
In [ ]: df
```

1

```
In [ ]: df.fillna(0)

In [ ]: df["preTestScore"].mean()

In [ ]: df["postTestScore"].median()

In [ ]: df["postTestScore"].mode()

In [ ]: df["preTestScore"]

In [ ]: df["preTestScore"].fillna(df["preTestScore"].mean(), inplace=True)

In [ ]: df

In [ ]: df.groupby("sex")["postTestScore"].transform("mean")

In [ ]: df["postTestScore"].fillna(
            df.groupby("sex")["postTestScore"].transform("mean"), inplace=True)
        df

In [ ]: df[df['age'].notnull() & df['sex'].notnull()]
```