In [ ]:
```
"""
Notebook to demonstrate how Pandas simplified analysis.
- Reading csv with converters
- Rename columns
- Boolean Indexing
- Aggregation
- Frequency counts
- Slicing
- Unique
"""
```

In [4]:
```
import pandas as pd
import numpy as np
```

In [5]:
```
songs = pd.read_csv('data/rock.csv')
songs.head()
```

Out[5]:

	Song Clean	ARTIST CLEAN	Release Year	COMBINED	First?	Year?	PlayCount	F*G
0	Caught Up in You	.38 Special	1982	Caught Up in You by .38 Special	1	1	82	82
1	Fantasy Girl	.38 Special	NaN	Fantasy Girl by .38 Special	1	0	3	0
2	Hold On Loosely	.38 Special	1981	Hold On Loosely by .38 Special	1	1	85	85
3	Rockin' Into the Night	.38 Special	1980	Rockin' Into the Night by .38 Special	1	1	18	18
4	Art For Arts Sake	10cc	1975	Art For Arts Sake by 10cc	1	1	1	1

In [6]:
```
# Clean data as you read it in using converters
songs = pd.read_csv(
    'data/rock.csv',
    usecols=['Song Clean', 'ARTIST CLEAN', 'Release Year', 'PlayCount'],
    converters={'Release Year': lambda x: int(x) if x.isdigit() and int(x) >
)

# Rename to columns that are easier to work with
songs = songs.rename(columns={
        'Song Clean': 'title',
        'ARTIST CLEAN': 'artist',
        'Release Year': 'year',
        'PlayCount': 'count'})
songs.head()
```

Out[6]:

	title	artist	year	count
0	Caught Up in You	.38 Special	1982	82
1	Fantasy Girl	.38 Special	NaN	3
2	Hold On Loosely	.38 Special	1981	85
3	Rockin' Into the Night	.38 Special	1980	18
4	Art For Arts Sake	10cc	1975	1

In [4]:
```
# Get number of songs released in year 1981
released_in_1981 = songs['year'] == 1981
print 'There were {} songs released in 1981'.format(
    # Boolean Indexing
    len(songs[released_in_1981])
)
```

```
There were 61 songs released in 1981
```

In [4]:
```
# Get number of songs released before 1984
before_1984 = songs['year'] < 1984
print 'There were {} songs released before 1984'.format(
    len(songs[before_1984])
)
```

```
There were 1218 songs released before 1984
```

In [5]:
```
# Earliest release year
print 'The first rock song was written in: {:.0f}'.format(
    # Series Aggregation
    songs['year'].min()
)
```

```
The first rock song was written in: 1955
```

In [12]:
```
# Top 20 songs by play count
top_20 = songs.sort('count', ascending=False)[:20]
top_20[['title', 'count']]
```

Out[12]:

	title	count
49	Dream On	142
868	All Along the Watchtower	141
65	Sweet Emotion	141
38	You Shook Me All Night Long	138
267	More Than a Feeling	134
968	Carry On Wayward Son	134
269	Peace of Mind	132
774	Crazy On You	125
2220	Legs	121
2224	Sharp Dressed Man	120
989	Rock And Roll All Nite	119
1354	Bohemian Rhapsody	119
1675	Renegade	116
784	Magic Man	115
956	Wheel in the Sky	114
1106	Blinded by the Light	114
1621	The Joker	114
1523	Tom Sawyer	114
1866	Just What I Needed	113
772	Barracuda	113

In [13]:
```
# Top 10 prolific artists
songs['artist'].value_counts()[:10]
```

Out[13]:
```
The Beatles                      100
Led Zeppelin                      69
Rolling Stones                    55
Van Halen                         44
Pink Floyd                        39
Aerosmith                         31
The Who                           31
Tom Petty & The Heartbreakers     29
AC/DC                             29
Bob Seger                         24
dtype: int64
```

In [17]:
```
# Number of different artists
uniques = songs['artist'].unique()
print 'There are {} different artists in this data set'.format(
    len(uniques)
)
```

```
There are 475 different artists in this data set
```

In [7]:
```
# Get all the songs with the word 'rock' in the title
mask = songs['title'].apply(lambda x: 'rock' in x.lower())
print songs[mask].head()
print '\n'
print 'In total, there are {} songs with the word "rock" in the title'.format
```

```
                      title       artist  year  count
3    Rockin' Into the Night  .38 Special  1980     18
13   CAN'T STOP ROCK'N'ROLL        AC/DC   NaN      5
15  For Those About To Rock        AC/DC  1981     46
17           Hard As A Rock        AC/DC  1995      1
23        Let There Be Rock        AC/DC  1977      3


In total, there are 60 songs with the word "rock" in the title
```

In [ ]:
```

```