In [22]:
```
import pandas as pd
from datetime import datetime
```

In [2]:
```
csv1 = pd.read_csv("C:\\Users\\soham\\Entertainer - Basic Info.csv" , names=
csv1.drop(index=[0], inplace=True)
csv1
```

Out[2]:

	Entertainer	Gender	Birth Year
1	Adele	F	1988
2	Angelina Jolie	F	1975
3	Aretha Franklin	F	1942
4	Bette Davis	F	1908
5	Betty White	F	1922
...	...	...	...
66	Tom Hanks	M	1956
67	Tony Bennett	M	1926
68	Wayne Newton	M	1942
69	Will Smith	M	1968
70	Willie Nelson	M	1933

70 rows × 3 columns

In [3]:
```
csv2 = pd.read_csv("C:\\Users\\soham\\Entertainer - Breakthrough Info.csv")
#imported the file "Entertainer - Breakthrough Info" in jupyter notebook

csv2
```

Loading [MathJax]/extensions/Safe.js

Out[3]:

	Entertainer	Year of Breakthrough/#1 Hit/Award Nomination	Breakthrough Name	Year of First Oscar/Grammy/Emmy
0	Adele	Year of Breakthrough/#1 Hit/Award Nomination	19	2009.0
1	Angelina Jolie	Year of Breakthrough/#1 Hit/Award Nomination	Girl, Interrupted	1999.0
2	Aretha Franklin	Year of Breakthrough/#1 Hit/Award Nomination	I Never Loved a Man (The Way I Love You)	1968.0
3	Bette Davis	Year of Breakthrough/#1 Hit/Award Nomination	Of Human Bondage	1935.0
4	Betty White	Year of Breakthrough/#1 Hit/Award Nomination	Life with Elilzabeth	1976.0
...	...	...	...	...
65	Tom Hanks	Year of Breakthrough/#1 Hit/Award Nomination	Splash	1993.0
66	Tony Bennett	Year of Breakthrough/#1 Hit/Award Nomination	Because of You	1963.0
67	Wayne Newton	Year of Breakthrough/#1 Hit/Award Nomination	Daddy, Don't You Walk So Fast	NaN
68	Will Smith	Year of Breakthrough/#1 Hit/Award Nomination	The Fresh Prince of Bel-Air	1988.0
69	Willie Nelson	Year of Breakthrough/#1 Hit/Award Nomination	Red Headed Stranger	1976.0

70 rows × 4 columns

In [4]:
```
csv2.rename(columns = {'Year of Breakthrough/#1 Hit/Award Nomination':'#1 Hi
csv2.rename(columns = {'Year of First Oscar/Grammy/Emmy':'Award Year'}, inpl
```

Loading [MathJax]/extensions/Safe.js

```
csv2
csv2cleaned = csv2.fillna({'#1 Hit': 0,
                           'Award Year': 0})            # HANDLING THE MISS
csv2cleaned
csv2clean = csv2cleaned.astype({"Award Year":int})  #CHANGED THE DATATYPE FR

csv2clean
csv2cleaned = csv2clean.replace(["Year of Breakthrough/#1 Hit/Award Nominati
csv2cleaned
```

Out[4]:

	Entertainer	#1 Hit	Breakthrough Name	Award Year
0	Adele	nomination year	19	2009
1	Angelina Jolie	nomination year	Girl, Interrupted	1999
2	Aretha Franklin	nomination year	I Never Loved a Man (The Way I Love You)	1968
3	Bette Davis	nomination year	Of Human Bondage	1935
4	Betty White	nomination year	Life with Elilzabeth	1976
...	...	...	...	...
65	Tom Hanks	nomination year	Splash	1993
66	Tony Bennett	nomination year	Because of You	1963
67	Wayne Newton	nomination year	Daddy, Don't You Walk So Fast	0
68	Will Smith	nomination year	The Fresh Prince of Bel-Air	1988
69	Willie Nelson	nomination year	Red Headed Stranger	1976

70 rows × 4 columns

In [5]:
```
csv3 = pd.read_csv("C:\\Users\\soham\\Entertainer - Last work Info.csv")
#imported the file "Entertainer - Breakthrough Info" in jupyter notebook

csv3
```

Loading [MathJax]/extensions/Safe.js

Out[5]:

	Entertainer	Year of Last Major Work (arguable)	Year of Death
0	Adele	2016	NaN
1	Angelina Jolie	2016	NaN
2	Aretha Franklin	2014	NaN
3	Bette Davis	1989	1989.0
4	Betty White	2016	NaN
...	...	...	...
65	Tom Hanks	2016	NaN
66	Tony Bennett	2016	NaN
67	Wayne Newton	2016	NaN
68	Will Smith	2016	NaN
69	Willie Nelson	2016	NaN

70 rows × 3 columns

In [6]:
```
csv3.rename(columns = {'Year of Last Major Work  (arguable)':'Last work'}, i
csv3


csv3c = csv3.fillna({'Year of Death': 0})              # HANDLING THE MISSINC


csv3clean = csv3c.astype({"Year of Death":int})  #CHANGED THE DATATYPE FROM
csv3clean
```

Out[6]:

	Entertainer	Last work	Year of Death
0	Adele	2016	0
1	Angelina Jolie	2016	0
2	Aretha Franklin	2014	0
3	Bette Davis	1989	1989
4	Betty White	2016	0
...	...	...	...
65	Tom Hanks	2016	0
66	Tony Bennett	2016	0
67	Wayne Newton	2016	0
68	Will Smith	2016	0
69	Willie Nelson	2016	0

70 rows × 3 columns

Loading [MathJax]/extensions/Safe.js

In [31]:
```
merged_df = csv1.merge(csv2cleaned, on='Entertainer', how='outer')\
                      .merge(csv3clean, on='Entertainer', how='outer')

merged_df['Birth Year'] = merged_df['Birth Year'].astype(int)
merged_df['Year of Death'] = merged_df['Year of Death'].astype(int)
merged_df['Award Year'] = merged_df['Award Year'].astype(int)
merged_df['last work'] = merged_df['Last work'].astype(int)
current_year = datetime.now().year

merged_df['Age'] = merged_df.apply(lambda row: current_year - row['Birth Yea
merged_df['Living Status'] = merged_df['Year of Death'].apply(lambda x: 'Liv

merged_df
```

Out[31]:

	Entertainer	Gender	Birth Year	#1 Hit	Breakthrough Name	Award Year	Last work	Yea o Death
0	Adele	F	1988	nomination year	19	2009	2016	0
1	Angelina Jolie	F	1975	nomination year	Girl, Interrupted	1999	2016	0
2	Aretha Franklin	F	1942	nomination year	I Never Loved a Man (The Way I Love You)	1968	2014	0
3	Bette Davis	F	1908	nomination year	Of Human Bondage	1935	1989	1989
4	Betty White	F	1922	nomination year	Life with Elilzabeth	1976	2016	0
...	...	...	...	...	...	...	...	...
65	Tom Hanks	M	1956	nomination year	Splash	1993	2016	0
66	Tony Bennett	M	1926	nomination year	Because of You	1963	2016	0
67	Wayne Newton	M	1942	nomination year	Daddy, Don't You Walk So Fast	0	2016	0
68	Will Smith	M	1968	nomination year	The Fresh Prince of Bel-Air	1988	2016	0
69	Willie Nelson	M	1933	nomination year	Red Headed Stranger	1976	2016	0

70 rows × 11 columns

In [33]:
```
merged_df.to_csv('Merged finalfff.csv', index=False)
```

Loading [MathJax]/extensions/Safe.js

In [ ]:

Loading [MathJax]/extensions/Safe.js