Code

Python 3 (ipyk

[ ]:
```
#datacleaning
import pandas as pd
import numpy as np
```

[ ]:
```
df=pd.read_excel(r"C:\Users\dell\Downloads\Entertainer Data\Entertainer Data Analysis\Entertainer - Basic Info.xlsx")
df1=pd.read_excel(r"C:\Users\dell\Downloads\Entertainer Data\Entertainer Data Analysis\Entertainer - Breakthrough Info.xlsx")
df2=pd.read_excel(r"C:\Users\dell\Downloads\Entertainer Data\Entertainer Data Analysis\Entertainer - Last work Info.xlsx")#importing 3 files
```

*[92]:
```
df.info()#first fil
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70 entries, 0 to 69
Data columns (total 3 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Entertainer           70 non-null     object
 1   Gender (traditional)  70 non-null     object
 2   Birth Year            70 non-null     int64 
dtypes: int64(1), object(2)
memory usage: 1.8+ KB
```

[93]:
```
df.describe()
```

[93]:

	Birth Year
count	70.000000
mean	1935.585714
std	24.135783

min	1889.000000
25%	1916.000000
50%	1935.500000
75%	1954.000000
max	1988.000000

[94]:
```
df.head()
```

[94]:

	Entertainer	Gender (traditional)	Birth Year
0	Adele	F	1988
1	Angelina Jolie	F	1975
2	Aretha Franklin	F	1942
3	Bette Davis	F	1908
4	Betty White	F	1922

*[95]:
```
df1.info()#second file info
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70 entries, 0 to 69
Data columns (total 4 columns):
 #   Column                  Non-Null Count  Dtype
---  ------
```

Code

Python 3 (ipykernel)

```
 0   Entertainer                                 70 non-null     object
 1   Year of Breakthrough/#1 Hit/Award Nomination  70 non-null   int64
 2   Breakthrough Name                           70 non-null     object
 3   Year of First Oscar/Grammy/Emmy             64 non-null     float64
dtypes: float64(1), int64(1), object(2)
memory usage: 2.3+ KB
```

[111]:
```
df2.info()#third file info
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70 entries, 0 to 69
Data columns (total 2 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Entertainer                     70 non-null     object
 1   Year of Last Major Work (arguable)  70 non-null  int64
dtypes: int64(1), object(1)
memory usage: 1.2+ KB
```

[100]:
```
df2.head()#third file
```

[100]:

	Entertainer	Year of Last Major Work (arguable)
0	Adele	2016
1	Angelina Jolie	2016
2	Aretha Franklin	2014
3	Bette Davis	1989
4	Betty White	2016

6	Bob Hope	1938	The Big Broadcast of 1938	1940.0
7	Carol Burnett	1959	The Garry Moore Show	1962.0
8	Carole Lombard	1934	Twentieth Century	NaN
9	Carrie Fisher	1977	Star Wars	NaN
10	Cary Grant	1933	She Done Him Wrong, I'm No Angel	1970.0
11	Charlie Chaplin	1915	The Tramp	1929.0
12	Clara Bow	1926	Mantrap	NaN
13	Clark Gable	1934	It Happened One Night	1934.0
14	David Letterman	1982	Late Night with David Letterman	1981.0
15	Debbie Reynolds	1952	Singin' in the Rain	NaN
16	Denzel Washington	1989	Glory	1989.0
17	Dick Van Dyke	1961	Bye Bye Birdie, The Dick Van Dyke Show	1964.0
18	Donald Sutherland	1967	The Dirty Dozen	1995.0
19	Dustin Hoffman	1967	The Graduate	1980.0
20	Ed Sullivan	1948	Toast of the Town	1956.0
21	Eddie Murphy	1980	Saturday Night Live	2001.0
22	Elton John	1972	Honky Chateau	1987.0

23	Elvis Presley	1956	Heartbreak Hotel	1959.0
24	Frank Sinatra	1940	I'll Never Smile Again	1946.0
25	Gene Hackman	1967	Bonnie and Clyde	1971.0
26	George Michael	1984	Wake Me Up Before You Go-Go	1988.0
27	Gregory Peck	1944	The Keys of the Kingdom	1962.0
28	Greta Garbo	1930	Anna Christie	1954.0
29	Humphrey Bogart	1936	The Petrified Forest	1951.0

*[102]:
```
df1['Year of First Oscar/Grammy/Emmy'].fillna(df1['Year of First Oscar/Grammy/Emmy'].median(), inplace=True)#adding median to the null values in second file
```

[103]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70 entries, 0 to 69
Data columns (total 4 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   Entertainer                               70 non-null     object
 1   Year of Breakthrough/#1 Hit/Award Nomination  70 non-null     int64
 2   Breakthrough Name                         70 non-null     object
 3   Year of First Oscar/Grammy/Emmy           70 non-null     float64
dtypes: float64(1), int64(1), object(2)
memory usage: 2.3+ KB
```

Entertainer data analytics

Code

Python 3 (ipykernel)

```
memory usage: 2.3+ KB
```

```
[104]: Entertainer_data=pd.merge(df,df1, on='Entertainer', how='outer')#merging first file and second file
```

```
[105]: Entertainer_data
```

[105]:

	Entertainer	Gender (traditional)	Birth Year	Year of Breakthrough/#1 Hit/Award Nomination	Breakthrough Name	Year of First Oscar/Grammy/Emmy
0	Adele	F	1988	2008	19	2009.0
1	Angelina Jolie	F	1975	1999	Girl, Interrupted	1999.0
2	Aretha Franklin	F	1942	1967	I Never Loved a Man (The Way I Love You)	1968.0
3	Bette Davis	F	1908	1934	Of Human Bondage	1935.0
4	Betty White	F	1922	1952	Life with Elilzabeth	1976.0
...	...	...	...	...	...	...
65	Tom Hanks	M	1956	1984	Splash	1993.0
66	Tony Bennett	M	1926	1951	Because of You	1963.0
67	Wayne Newton	M	1942	1972	Daddy, Don't You Walk So Fast	1978.0
68	Will Smith	M	1968	1990	The Fresh Prince of Bel-Air	1988.0
69	Willie Nelson	M	1933	1975	Red Headed Stranger	1976.0

70 rows × 6 columns

Entertainer data analytics

Code

Python 3 (ipykernel)

[108]:
```
Entertainer_data1.describe()
```

```
...
```

*[109]:
```
Entertainer_data1.info()#info of all three merged files
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 70 entries, 0 to 69
Data columns (total 7 columns):
 #   Column                                      Non-Null Count  Dtype  
---  ------                                      --------------  -----  
 0   Entertainer                                 70 non-null     object 
 1   Gender (traditional)                        70 non-null     object 
 2   Birth Year                                  70 non-null     int64  
 3   Year of Breakthrough/#1 Hit/Award Nomination  70 non-null   int64  
 4   Breakthrough Name                           70 non-null     object 
 5   Year of First Oscar/Grammy/Emmy             70 non-null     float64
 6   Year of Last Major Work (arguable)          70 non-null     int64  
dtypes: float64(1), int64(3), object(3)
memory usage: 4.4+ KB
```

*[110]:
```
Entertainer_data1.to_csv('Entertainer_data1.csv' , index=False)#coverting updated data to csv file
```

[ ]:

[ ]: