Import Libraries and Load Data

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the data
basic_info_df = pd.read_excel('Entertainer - Basic Info.xlsx')
breakthrough_info_df = pd.read_excel('Entertainer - Breakthrough
Info.xlsx')
last_work_info_df = pd.read_excel('Entertainer - Last work Info.xlsx')
```

Data Overview

```
# Display the first few rows of each dataframe
print(basic_info_df.head())
print(breakthrough_info_df.head())
print(last_work_info_df.head())

# Summary statistics
print(basic_info_df.describe())
print(breakthrough_info_df.describe())
print(last_work_info_df.describe())
```

```
       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
       Entertainer  Year of Breakthrough/#1 Hit/Award Nomination  \
0            Adele                                          2008   
1   Angelina Jolie                                          1999   
2  Aretha Franklin                                          1967   
3      Bette Davis                                          1934   
4      Betty White                                          1952   

                          Breakthrough Name  Year of First
Oscar/Grammy/Emmy
0                                        19
2009.0
1                          Girl, Interrupted
1999.0
2  I Never Loved a Man (The Way I Love You)
1968.0
3                            Of Human Bondage
1935.0
4                      Life with Elilzabeth
```

```
1976.0
       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
        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
       Year of Breakthrough/#1 Hit/Award Nomination  \
count                                     70.000000
mean                                    1964.228571
std                                       22.411935
min                                     1915.000000
25%                                     1949.500000
50%                                     1963.500000
75%                                     1983.500000
max                                     2008.000000

       Year of First Oscar/Grammy/Emmy
count                        64.000000
mean                       1976.234375
std                          22.170152
min                        1929.000000
25%                        1962.000000
50%                        1978.000000
75%                        1993.000000
max                        2017.000000
       Year of Last Major Work (arguable)  Year of Death
count                           70.000000      30.000000
mean                          1998.971429    1988.133333
std                             22.874561      20.483355
min                           1933.000000    1942.000000
25%                           1980.000000    1977.000000
50%                           2014.000000    1989.500000
75%                           2016.000000    2003.750000
max                           2016.000000    2016.000000
```

Data Cleaning

```
# Check for missing values
print(basic_info_df.isnull().sum())
print(breakthrough_info_df.isnull().sum())
print(last_work_info_df.isnull().sum())

# Fill or drop missing values as necessary
basic_info_df = basic_info_df.dropna()
breakthrough_info_df = breakthrough_info_df.dropna()
last_work_info_df = last_work_info_df.dropna()
```

```
Entertainer             0
Gender (traditional)    0
Birth Year              0
dtype: int64
Entertainer                                     0
Year of Breakthrough/#1 Hit/Award Nomination    0
Breakthrough Name                               0
Year of First Oscar/Grammy/Emmy                 6
dtype: int64
Entertainer                            0
Year of Last Major Work (arguable)     0
Year of Death                         40
dtype: int64
```

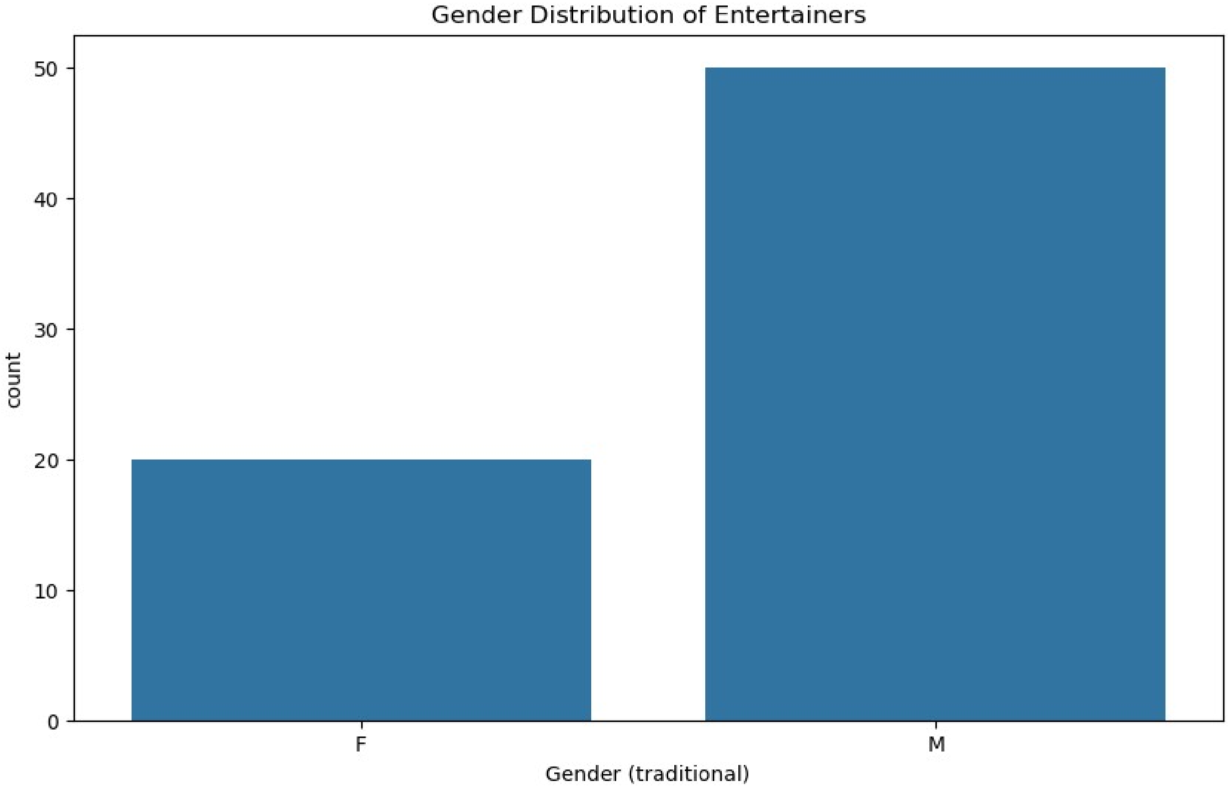
Analysis

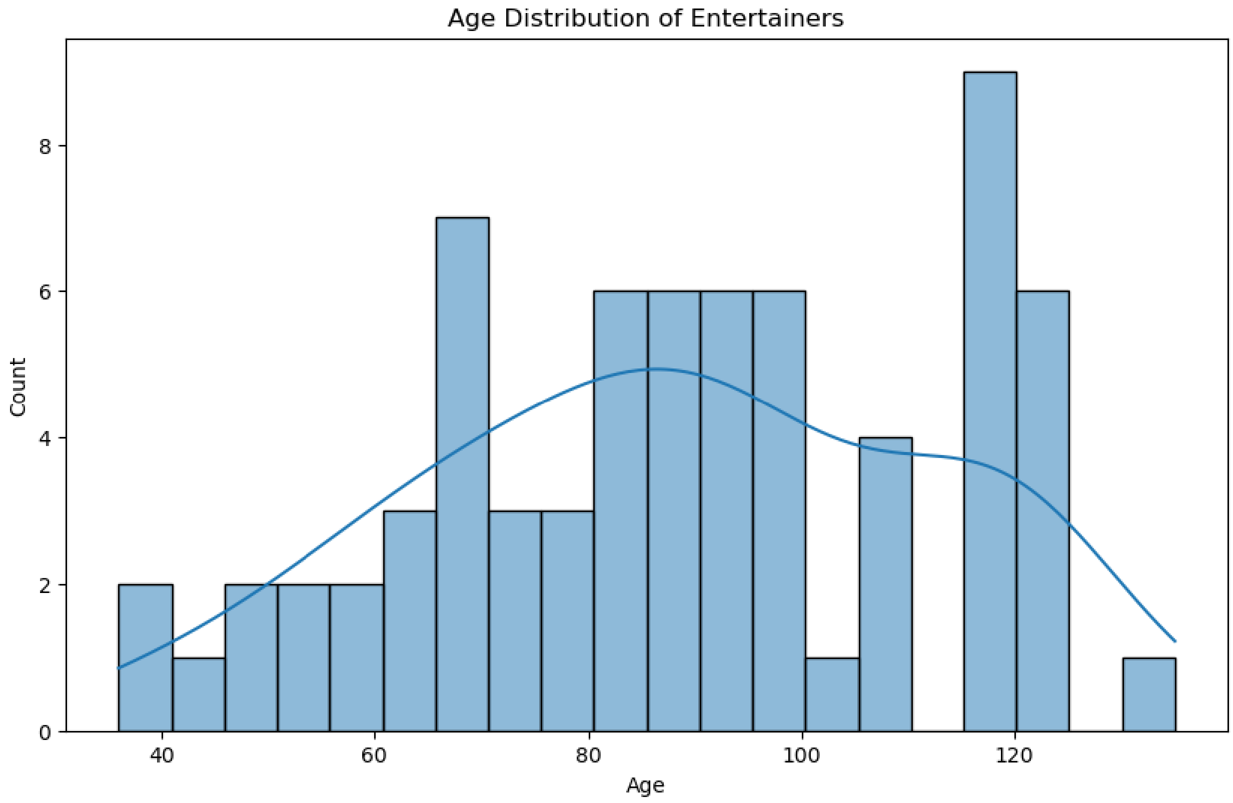
Demographics Analysis

```
# Gender distribution
gender_counts = basic_info_df['Gender (traditional)'].value_counts()
print(gender_counts)

# Age distribution
current_year = 2024
basic_info_df['Age'] = current_year - basic_info_df['Birth Year']
age_distribution = basic_info_df['Age'].describe()
age_distribution.head(50)

# Plotting gender distribution
plt.figure(figsize=(10, 6))
sns.countplot(data=basic_info_df, x='Gender (traditional)')
plt.title('Gender Distribution of Entertainers')
plt.show()

# Plotting age distribution
plt.figure(figsize=(10, 6))
sns.histplot(basic_info_df['Age'], bins=20, kde=True)
plt.title('Age Distribution of Entertainers')
plt.show()
```

```
Gender (traditional)
M    50
```

```
F     20
Name: count, dtype: int64
```

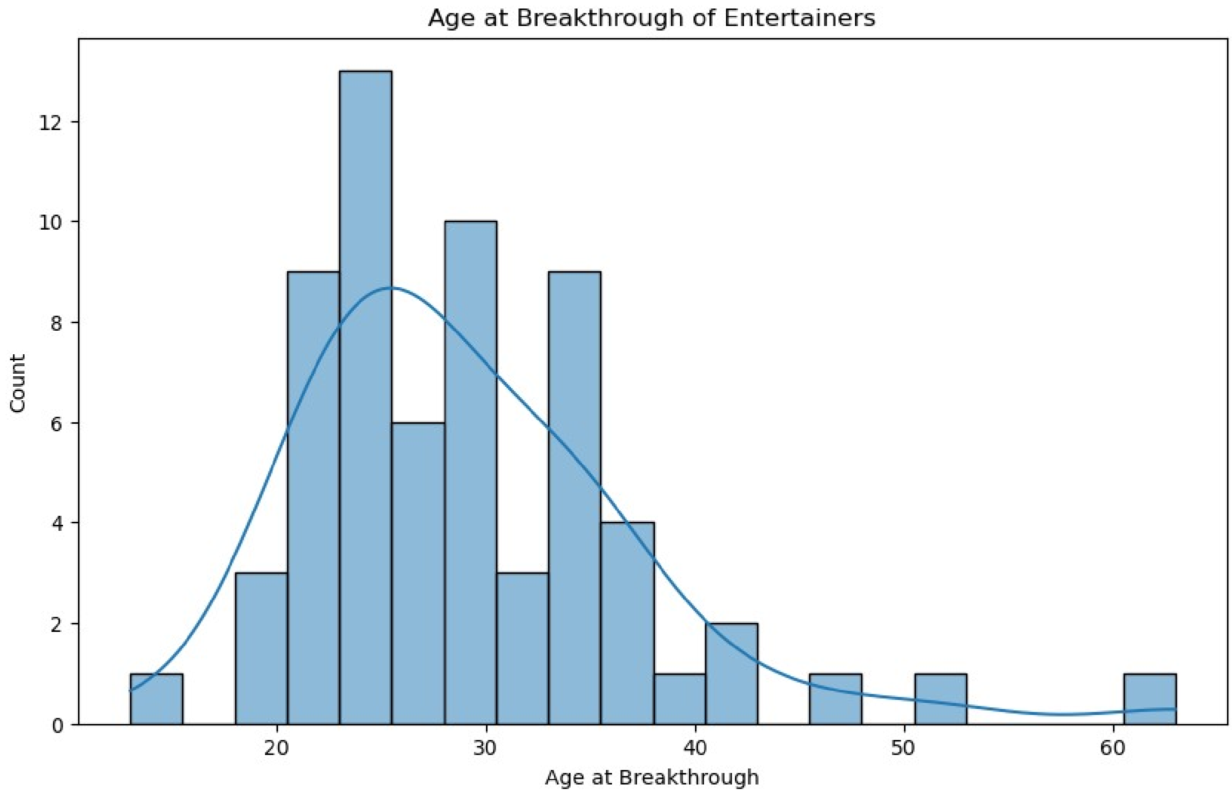




Career Milestones

```
# Assuming breakthrough_info_df has columns 'Entertainer' and 'Breakthrough Year'
# Calculate the age at breakthrough
breakthrough_info_df = breakthrough_info_df.merge(basic_info_df[['Entertainer', 'Birth Year']], on='Entertainer')
breakthrough_info_df['Age at Breakthrough'] = breakthrough_info_df['Year of Breakthrough/#1 Hit/Award Nomination'] - breakthrough_info_df['Birth Year']

# Summary statistics for age at breakthrough
age_breakthrough_stats = breakthrough_info_df['Age at Breakthrough'].describe()
print(age_breakthrough_stats)

# Plotting age at breakthrough
plt.figure(figsize=(10, 6))
sns.histplot(breakthrough_info_df['Age at Breakthrough'], bins=20, kde=True)
plt.title('Age at Breakthrough of Entertainers')
plt.show()

count    64.000000
mean     29.109375
```

```
std         8.279094
min        13.000000
25%        24.000000
50%        27.500000
75%        33.000000
max        63.000000
Name: Age at Breakthrough, dtype: float64
```

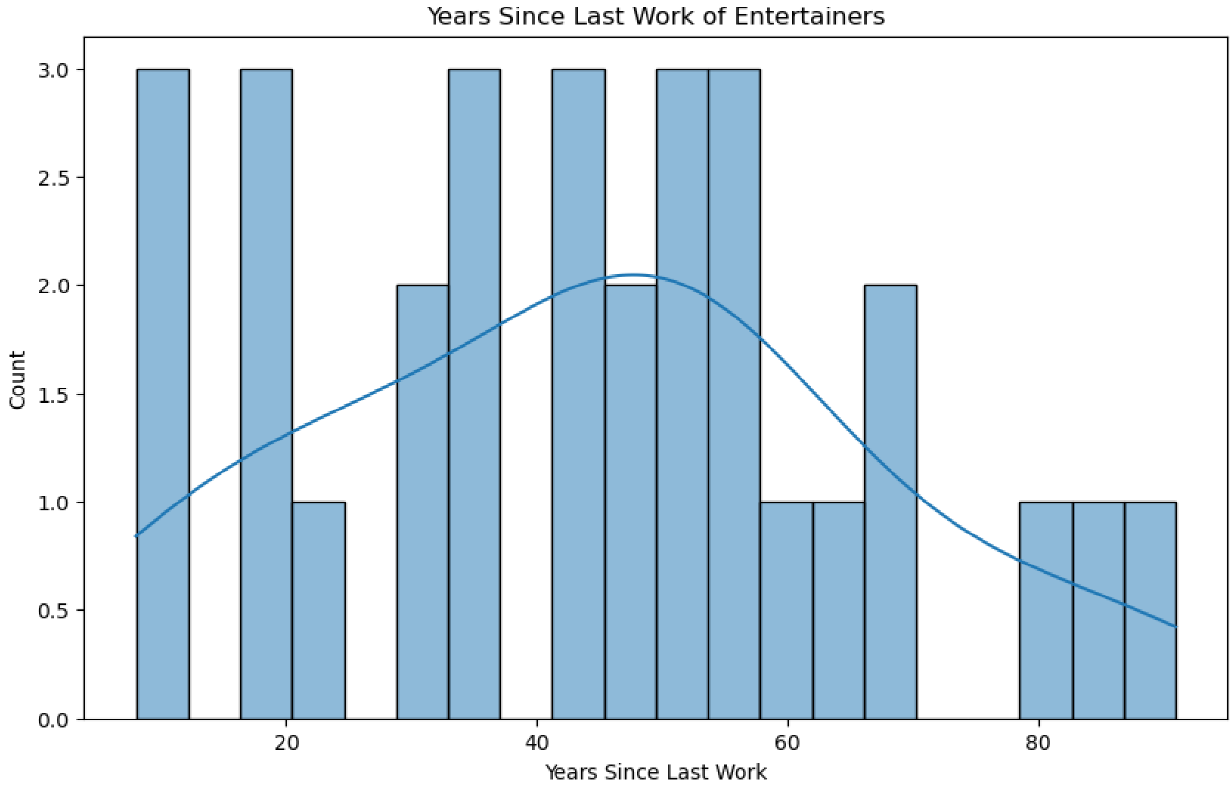


Recent Works

```
# Assuming last_work_info_df has columns 'Entertainer' and 'Last Work
Year'
# Calculate the years since last work
last_work_info_df =
last_work_info_df.merge(basic_info_df[['Entertainer', 'Birth Year']],
on='Entertainer')
last_work_info_df['Years Since Last Work'] = current_year -
last_work_info_df['Year of Last Major Work (arguable)']

# Summary statistics for years since last work
years_last_work_stats = last_work_info_df['Years Since Last
Work'].describe()
print(years_last_work_stats)

# Plotting years since last work
```

```
plt.figure(figsize=(10, 6))
sns.histplot(last_work_info_df['Years Since Last Work'], bins=20,
kde=True)
plt.title('Years Since Last Work of Entertainers')
plt.show()
```

```
count    30.000000
mean     44.400000
std      22.074482
min       8.000000
25%      30.500000
50%      45.500000
75%      56.250000
max      91.000000
Name: Years Since Last Work, dtype: float64
```


Years Since Last Work of Entertainers
Count
Years Since Last Work