

✓ Importing Libraries

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

✓ Reading Datasets

```
entertainer_data = pd.read_excel('/content/Entertainer - Basic Info.xlsx')
breakthrough_data = pd.read_excel('/content/Entertainer - Breakthrough Info.xlsx')
last_major_work_data = pd.read_excel('/content/Entertainer - Last work Info.xlsx')
```

```
# Convert data to dataframes
entertainer_df = pd.DataFrame(entertainer_data)
breakthrough_df = pd.DataFrame(breakthrough_data)
last_major_work_df = pd.DataFrame(last_major_work_data)
```

```
#Transform
merged_df = pd.merge(entertainer_df, breakthrough_df, on='Entertainer')
merged_df = pd.merge(merged_df, last_major_work_df, on='Entertainer')
```

```
# Display the transformed data
print(merged_df)
```

	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	
..

65	Tom Hanks	M	1956
66	Tony Bennett	M	1926
67	Wayne Newton	M	1942
68	Will Smith	M	1968
69	Willie Nelson	M	1933

	Year of Breakthrough/#1 Hit/Award Nomination \
0	2008
1	1999
2	1967
3	1934
4	1952
..	...
65	1984
66	1951
67	1972
68	1990
69	1975

	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 Elilizabeth	1976.0
..
65	Splash	1993.0
66	Because of You	1963.0
67	Daddy, Don't You Walk So Fast	NaN
68	The Fresh Prince of Bel-Air	1988.0
69	Red Headed Stranger	1976.0

	Year of Last Major Work (arguable)	Year of Death
0	2016	NaN
1	2016	NaN
2	2014	NaN
3	1989	1989.0
4	2016	NaN
..
65	2016	NaN
66	2016	NaN
67	2016	NaN

68	2016	NaN
69	2016	NaN

[70 rows x 8 columns]

✓ Getting information about datasets

```
df=merged_df
```

```
#shape of dataset
```

```
df.shape
```

```
(70, 8)
```

```
#Finding unique columns
```

```
df.columns
```

```
Index(['Entertainer', 'Gender (traditional)', 'Birth Year',
       'Year of Breakthrough/#1 Hit/Award Nomination', 'Breakthrough Name',
       'Year of First Oscar/Grammy/Emmy', 'Year of Last Major Work (arguable)',
       'Year of Death'],
      dtype='object')
```

```
#Size of dataset
```

```
df.size
```

```
560
```

```
#information about dataset
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 70 entries, 0 to 69
Data columns (total 8 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 64 non-null    float64
6  Year of Last Major Work (arguable) 70 non-null    int64
7  Year of Death         30 non-null    float64
dtypes: float64(2), int64(3), object(3)
memory usage: 4.9+ KB

```

```

#Descriptive statistics
df.describe()

```

	Birth Year	Year of Breakthrough/#1 Hit/Award Nomination	Year of First Oscar/Grammy/Emmy	Year of Last Major Work (arguable)	Year of Death
count	70.000000	70.000000	64.000000	70.000000	30.000000
mean	1935.585714	1964.228571	1976.234375	1998.971429	1988.133333
std	24.135783	22.411935	22.170152	22.874561	20.483355
min	1889.000000	1915.000000	1929.000000	1933.000000	1942.000000
25%	1916.000000	1949.500000	1962.000000	1980.000000	1977.000000
50%	1935.500000	1963.500000	1978.000000	2014.000000	1989.500000
75%	1954.000000	1983.500000	1993.000000	2016.000000	2003.750000
max	1988.000000	2008.000000	2017.000000	2016.000000	2016.000000

```

#Dimensions
df.ndim

```

```
2
```

```

#Finding number NULL values in dataset

```

```
df.isnull().sum()
```

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