

```
In [1]: import pandas as pd
import numpy as np
from matplotlib import pyplot
import seaborn as sns
import missingno as msno
```

Loading The dataset

```
In [2]: Basic_info_data = pd.read_excel("file:///C:/Users/Vasu%20Prasad/OneDrive/Documents/MCA/Internships/Unified%20Mentors/Entertainer%20Data%20Anal...")
```

```
In [3]: Basic_info_data
```

```
Out[3]:
```

| | 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 |

70 rows × 3 columns

In [8]: Basic_info_data.info()

```
<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
```

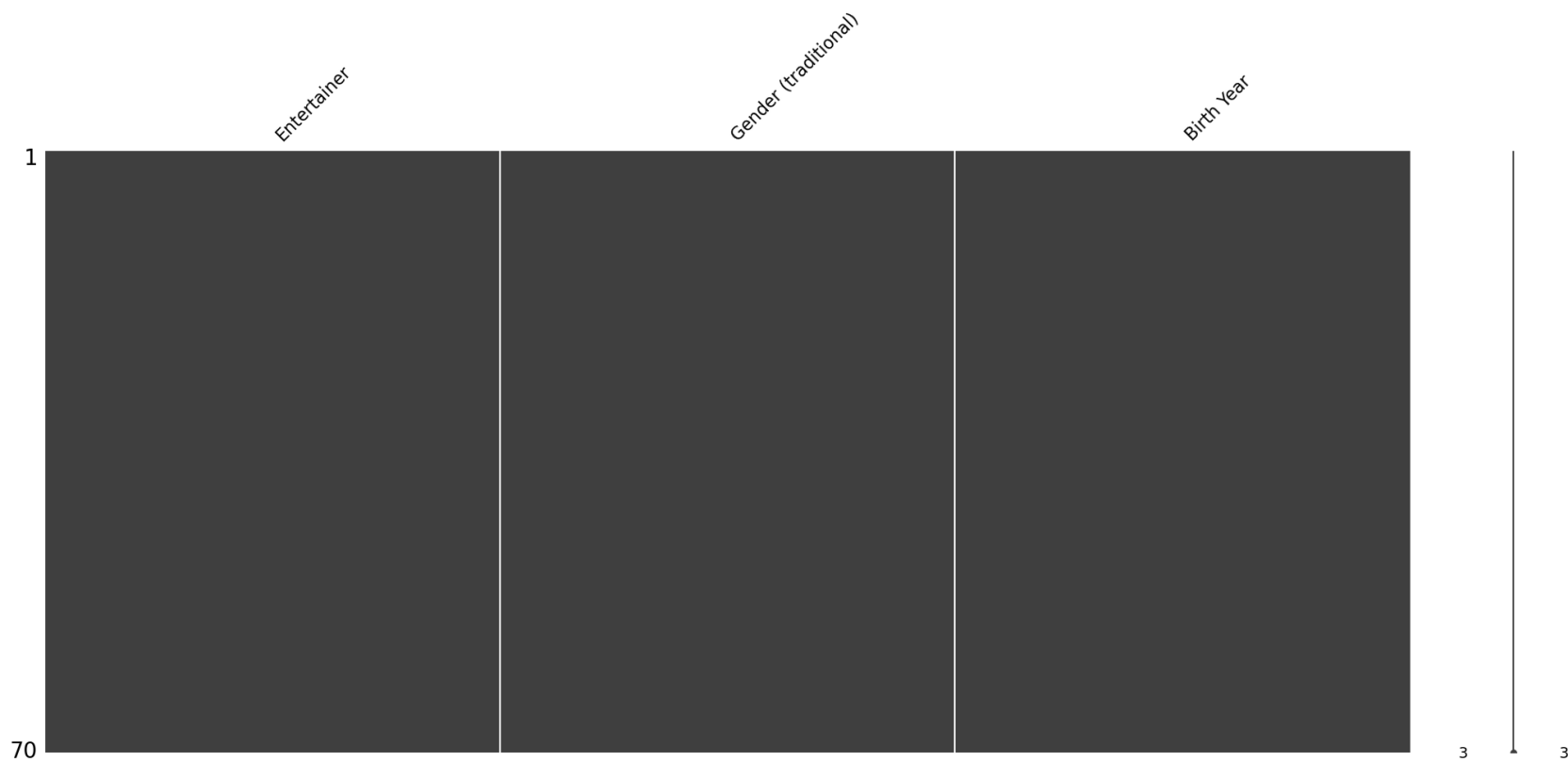
In [11]: Basic_info_data.isna().sum().reset_index().rename(columns = {0:"Count"})

Out[11]:

| | index | Count |
|---|----------------------|-------|
| 0 | Entertainer | 0 |
| 1 | Gender (traditional) | 0 |
| 2 | Birth Year | 0 |

```
In [12]: msno.matrix(Basic_info_data)
```

```
Out[12]: <Axes: >
```



```
In [4]: Breakthrough_Info = pd.read_excel("file:///C:/Users/Vasu%20Prasad/OneDrive/Documents/MCA/Internships/Unified%20Mentors/Entertainer%20Data%20Anal...
```

In [5]: Breakthrough_Info

Out[5]:

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

70 rows × 4 columns

```
In [13]: Breakthrough_Info.describe()
```

```
Out[13]:
```

| | Year of Breakthrough/#1 Hit/Award Nomination | Year of First Oscar/Grammy/Emmy |
|-------|--|---------------------------------|
| count | 70.000000 | 64.000000 |
| mean | 1964.228571 | 1976.234375 |
| std | 22.411935 | 22.170152 |
| min | 1915.000000 | 1929.000000 |
| 25% | 1949.500000 | 1962.000000 |
| 50% | 1963.500000 | 1978.000000 |
| 75% | 1983.500000 | 1993.000000 |
| max | 2008.000000 | 2017.000000 |

```
In [14]: Breakthrough_Info.isna().sum().reset_index().rename(columns = {0: "Count"})
```

```
Out[14]:
```

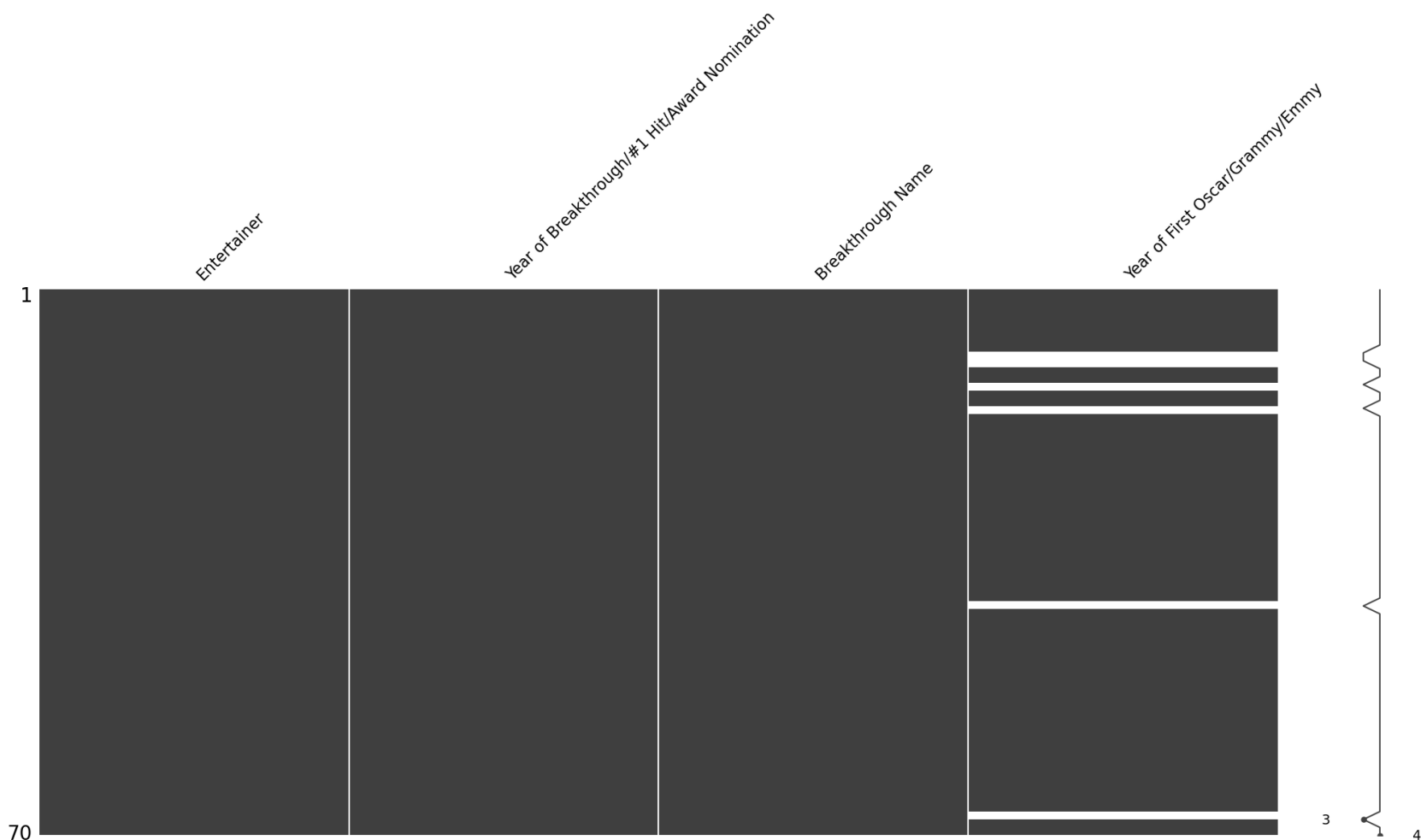
| | index | Count |
|---|--|-------|
| 0 | Entertainer | 0 |
| 1 | Year of Breakthrough/#1 Hit/Award Nomination | 0 |
| 2 | Breakthrough Name | 0 |
| 3 | Year of First Oscar/Grammy/Emmy | 6 |

```
In [15]: Breakthrough_Info.duplicated().sum()
```

```
Out[15]: 0
```

```
In [17]: msno.matrix(Breakthrough_Info)
```

```
Out[17]: <Axes: >
```



```
In [6]: Last_work_Info = pd.read_excel("file:///C:/Users/Vasu%20Prasad/OneDrive/Documents/MCA/Internships/Unified%20M...
```

```
In [7]: Last_work_Info
```

```
Out[7]:
```

| | 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

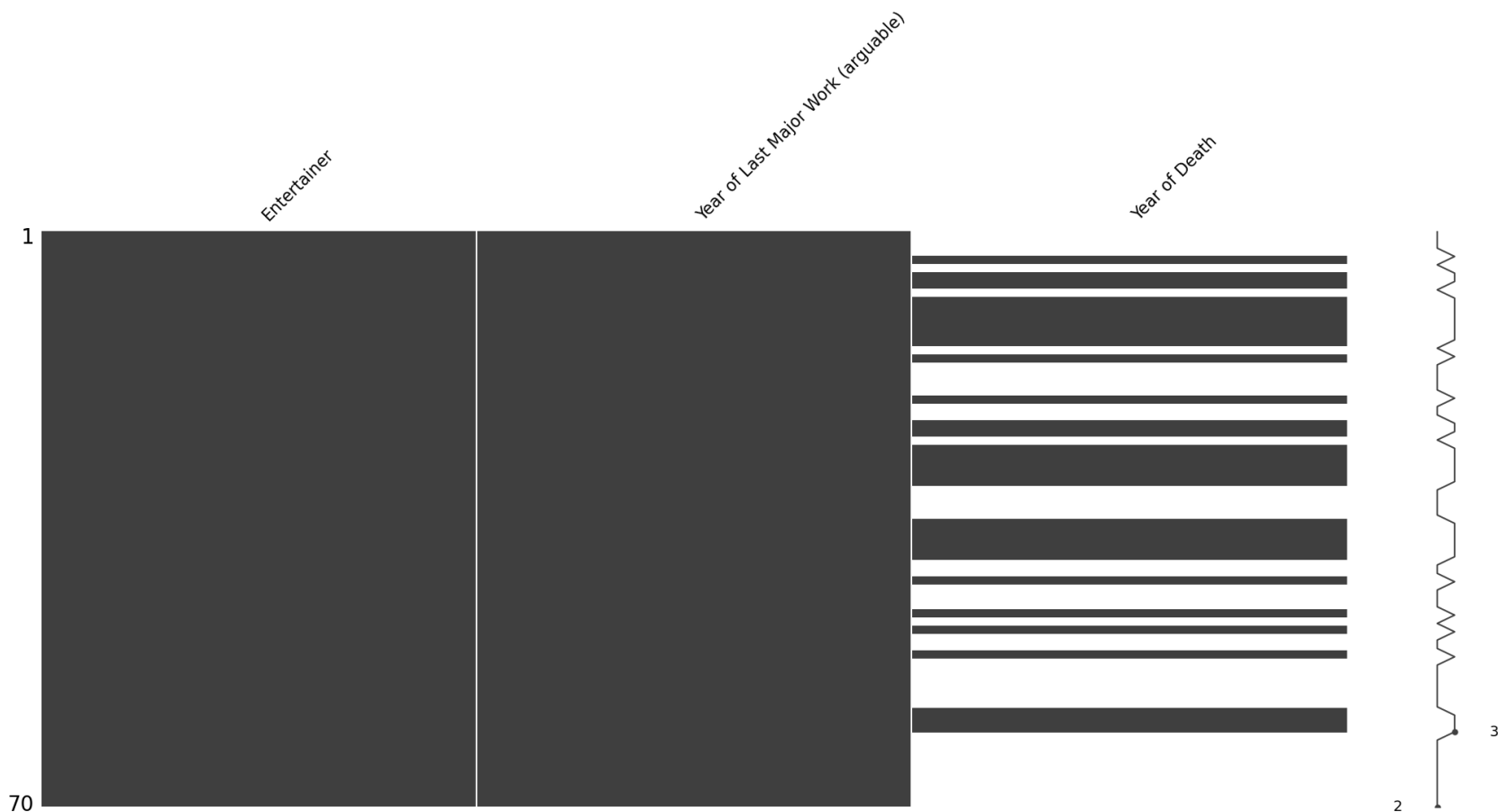
```
In [18]: .isna().sum().reset_index().rename(columns = {0:"Count"})
```

```
Out[18]:
```

| | index | Count |
|---|------------------------------------|-------|
| 0 | Entertainer | 0 |
| 1 | Year of Last Major Work (arguable) | 0 |
| 2 | Year of Death | 40 |

```
In [20]: msno.matrix>Last_work_Info)
```

```
Out[20]: <Axes: >
```



```
In [22]: Last_work_Info.duplicated().sum()
```

```
Out[22]: 0
```

```
In [ ]:
```