Entertainer Data Analysis Report

1. Introduction

1.1. Project Overview

This report presents the analysis of entertainer data collected from various sources. The goal of this analysis is to merge and analyze data from different datasets to gain insights into entertainers' careers, breakthrough achievements, and last major works.

1.2. Datasets

Entertainer - Basic Info.xlsx: Contains basic information about entertainers including name, gender, and birth year.

Entertainer - Breakthrough Info.xlsx: Details on breakthrough achievements, including the year of breakthrough, breakthrough name, and year of first significant award.

Entertainer - Last Work Info.xlsx: Information on the year of last major work and year of death.

employee dataset.csv: Additional employee data, which might be related to the entertainers' roles.

2. Data Preparation

2.1. Loading Data

The data was loaded from the provided files using Pandas. The following code was used for loading the datasets:

```
import pandas as pd


# Load data

basic_info = pd.read_excel('Entertainer - Basic Info.xlsx')
```

```
breakthrough_info = pd.read_excel('Entertainer - Breakthrough Info.xlsx')

last_work_info = pd.read_excel('Entertainer - Last Work Info.xlsx')

employee_data = pd.read_csv('employee dataset.csv')
```

2.2. Data Cleaning

Column names were standardized to facilitate merging:

```
# Clean column names

basic_info.columns = basic_info.columns.str.strip().str.lower().str.replace(' ', '_')

breakthrough_info.columns =
breakthrough_info.columns.str.strip().str.lower().str.replace(' ', '_')

last_work_info.columns = last_work_info.columns.str.strip().str.lower().str.replace(' ', '_')

employee_data.columns = employee_data.columns.str.strip().str.lower().str.replace(' ',
'_')
```

2.3. Merging Datasets

The datasets were merged on the `entertainer` column:

```
# Rename columns for merging consistency

basic_info = basic_info.rename(columns={'entertainer': 'entertainer'})

breakthrough_info = breakthrough_info.rename(columns={'entertainer': 'entertainer'})

last_work_info = last_work_info.rename(columns={'entertainer': 'entertainer'})


# Merge datasets

merged_data = basic_info.merge(breakthrough_info, on='entertainer', how='left')

merged_data = merged_data.merge(last_work_info, on='entertainer', how='left')
```

3. Data Analysis

3.1. Summary Statistics

The summary statistics of the merged dataset are as follows:

Statistic	Birth year	Year of Breakthrough	Year of First Award	Year of Last Major Work	Year of Death
Count	70	70	64	70	30
Mean	1935.59	1964.23	1976.23	1998.97	1988.13
Std Dev	24.14	22.41	22.17	22.87	20.48
Min	1889	1915	1929	1933	1942
25th Percentile	1916	1949.5	1962	1980	1977
Median	1935.5	1963.5	1978	2014	1989.5
75th Percentile	1954	1983.5	1993	2016	2003.75
Max	1988	2008	2017	2016	2016

3.2. Data Visualization

Visualizations were created to analyze the distribution of various attributes:

- **Distribution of Years of Last Major Work**

import seaborn as sns

import matplotlib.pyplot as plt

```
# Convert column to numeric
merged_data['year_of_last_major_work_(arguable)'] = pd.to_numeric(merged_data['year_of_last_major_work_(arguable)'], errors='coerce')

# Plot histogram
sns.histplot(data=merged_data, x='year_of_last_major_work_(arguable)', kde=True)
plt.title('Distribution of Years of Last Major Work')
plt.xlabel('Year of Last Major Work')
plt.ylabel('Frequency')
plt.show()
```

- **Distribution of Breakthroughs**

```
sns.countplot(data=merged_data, x='year_of_breakthrough/#1_hit/award_nomination')
plt.xticks(rotation=90)
plt.title('Distribution of Breakthroughs or Hits')
plt.xlabel('Year of Breakthrough or Hit')
plt.ylabel('Count')
plt.show()
```

- **Gender Distribution**

```
sns.countplot(data=merged_data, x='gender_(traditional)')
plt.title('Gender Distribution of Entertainers')
plt.xlabel('Gender')
plt.ylabel('Count')
```

```
plt.show()
```

4. Insights and Findings

4.1. Key Insights

- **Birth Year Trends:** The average birth year of entertainers is 1935. With a wide range from 1889 to 1988, the data includes entertainers from multiple generations.
- **Breakthrough Trends:** The mean year of breakthrough is 1964, indicating a peak period of achievements around the mid-20th century.
- **Award Trends:** The year of first major award ranges from 1929 to 2017, reflecting long and diverse careers.
- **Last Major Work:** Entertainers' last major work spans from 1933 to 2016, with a mean year of 1998, showing a long period of active career.
- **Death Trends:** The average year of death is 1988, with many entertainers passing away in recent decades.

4.2. Limitations

- The data may contain missing or inconsistent entries that could affect the analysis.
- The relationship between employee data and entertainer data was not explored due to lack of direct linkage.

5. Conclusions

The analysis of the entertainer data provides valuable insights into the career trajectories of entertainers, their breakthrough moments, and the timing of their last major works. Further exploration and additional data could provide a deeper understanding of industry trends and individual career paths.