Analysis of the Dataset: Gender and Age Distribution among One Million AI-Generated Faces

Erkhov Richard

Abstract

This report presents an analysis of a dataset named "million-faces"[1], containing one million AI-generated faces. The dataset was examined focusing on the gender and age distribution among these faces. The results reveal valuable insights into the demographics of the dataset, providing a foundation for future research and exploration in fields such as demographics analysis, facial recognition, and machine learning.

Dataset Overview

"Million-faces" is one of the largest facesets available to the public, consisting of one million faces that are all AI-generated. The images in the dataset may contain some artifacts due to the nature of AI generation. The dataset is almost fully uploaded on HuggingFace[2], a renowned platform for hosting datasets and models for the machine learning community. It serves as a significant resource for researchers and developers in AI, machine learning, and computer vision.

Usage:

The dataset is open for use in various projects and research. The community is encouraged to utilize this vast resource and consider linking back to the original project when creating something amazing, adhering to the recognition of work as a pillar of the open-source community.

Gender Distribution:

The dataset was analyzed to determine the gender distribution among the one million faces. Figure 1 illustrates the distribution of females and males.

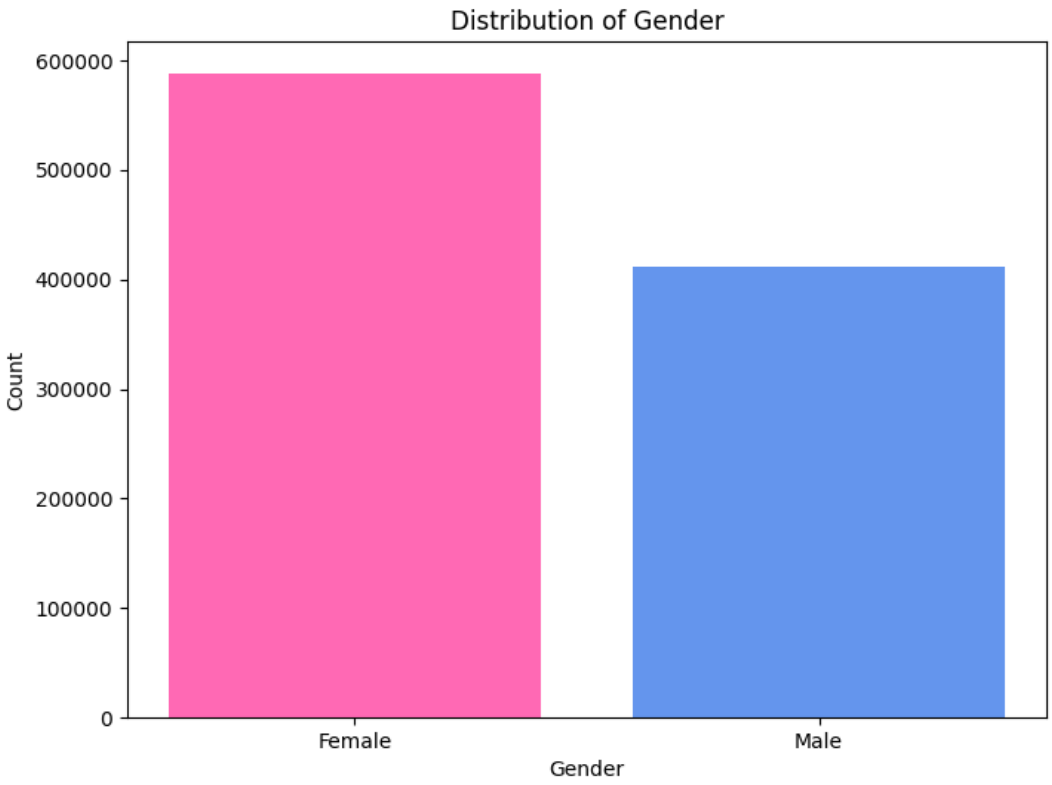

Figure 1: Distribution of Females and Males

Female: 58.81%

Male: 41.19%

Age Distribution:

To gain insights into the age demographics, the dataset's age distribution was examined. The analysis revealed the minimum and maximum ages, as well as the average and median ages. Figure 2 presents the overall age distribution.

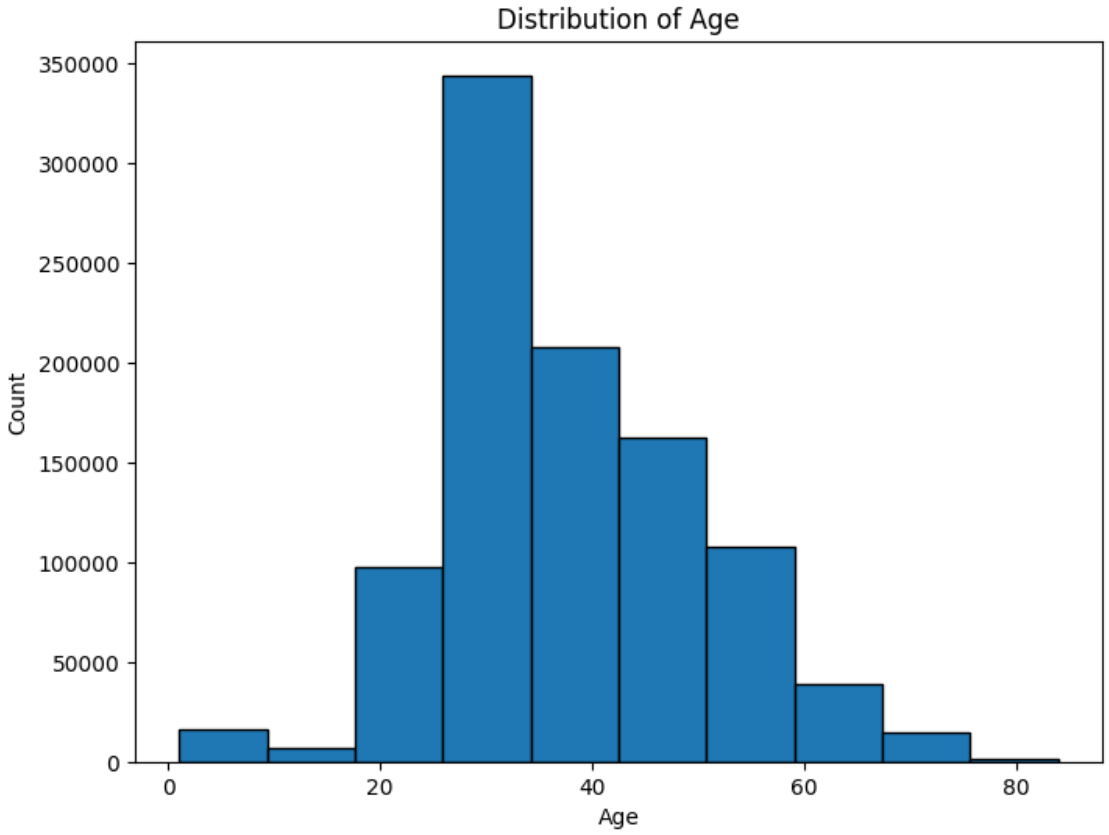

Figure 2: Overall Age Distribution

Minimum Age: 1

Maximum Age: 84

Average Age: 37.75

Median Age: 36.0

Age Distribution by Gender:

Further analysis was conducted to explore the age distribution by gender. Figure 3 showcases the distribution of males and females across different age groups.

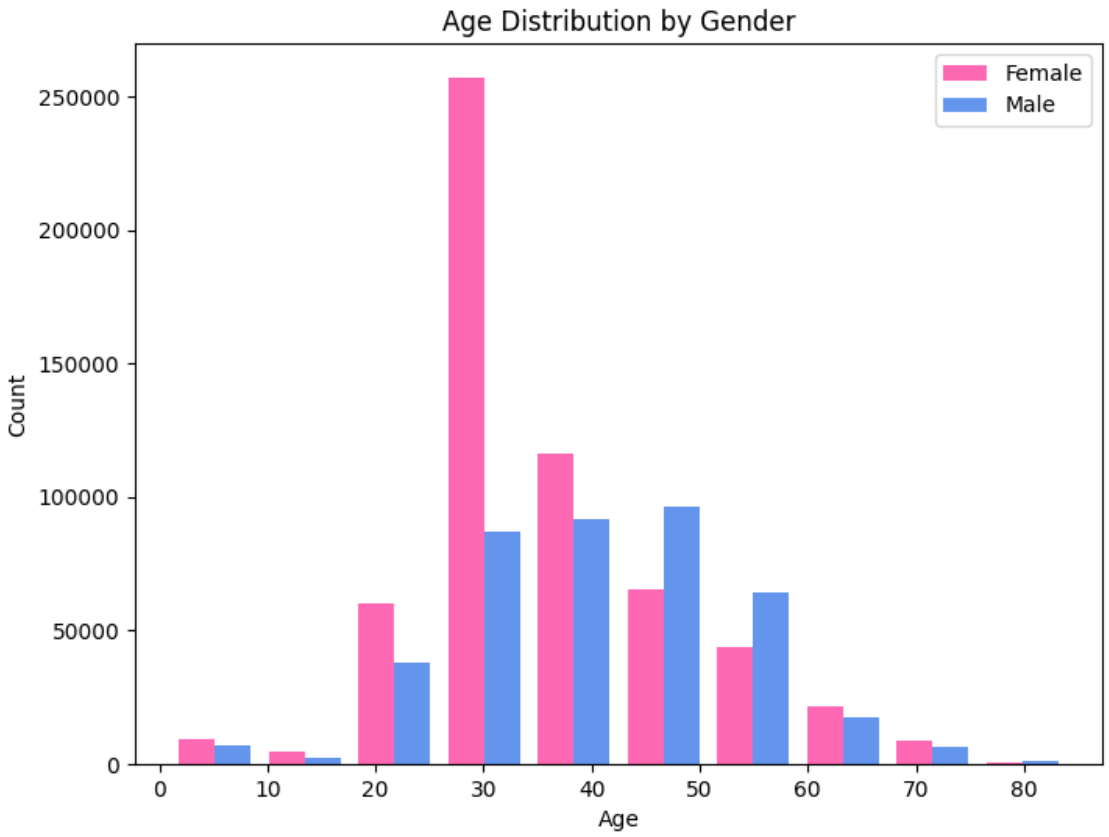

Figure 3: Distribution of Male/Female by Age

Female:

- Minimum Age: 1
- Maximum Age: 83
- Average Age: 35.78
- Median Age: 33.0

Male:

- Minimum Age: 1
- Maximum Age: 84
- Average Age: 40.56
- Median Age: 41.0

Future Work

While the initial analysis of the dataset offers valuable insights into gender and age distribution, there is a potential for further research. The creator of the dataset is looking to collect more information about the dataset. Furthermore, there are plans for a fast face enhancement/upscaling project, indicating the dataset's potential for contributing to advancements in the field of AI and machine learning.

Conclusion

The analysis of the one million faces dataset provided valuable insights into the gender and age distribution. The dataset consisted of 58.81% females and 41.19% males, highlighting a slight imbalance between the genders. The age range spanned from 1 to 84, with an average age of 37.75 and a median age of 36.

References

[1] Richard Erkhov, “Million Faces” dataset github page, https://github.com/RichardErkhov/million-faces

[2] Richard Erkhov, “Million Faces” dataset published on Huggingface, https://huggingface.co/datasets/RichardErkhov/OneMillionFaces