



# Analyzing the Relationship Between an Artist's Background and the Popularity of Their Works in MoMA

**Lina Li**

Shenyang Conservatory of Music, Shenyang 110818, Liaoning, China

DOI: 10.32629/asc.v4i1.1311

---

**Abstract:** This study delves into the intricate relationship between an artist's background (including nationality and gender) and the popularity of their artworks in the Museum of Modern Art (MoMA) in New York. Leveraging statistical methods, including Chi-squared tests and ANOVA, significant correlations between an artist's nationality, gender, and the popularity of their artworks were identified. Time series analysis further underscored evolving trends in MoMA's acquisition patterns over the years. The research also utilized a Random Forest classification model to predict artwork popularity, achieving an accuracy rate of 0.75. The findings shed light on the dynamics of art acquisitions and offer implications for diversifying and enhancing museum collections in the future.

**Keywords:** background, popularity, the Museum of Modern Art (MoMA), statistical method

---

## 1. Introduction

Since its establishment in 1929, the Museum of Modern Art (MoMA) has firmly secured its leadership in the global art sphere[1]. Located in the heart of New York City, MoMA annually attracts a myriad of tourists, artists, and scholars for visits and research[2]. Its comprehensive and profound collection spans from Impressionism to contemporary art, encompassing diverse art forms from sculpture and painting to photography[3]. Furthermore, through its insightful exhibitions and educational programs, MoMA further cements its authoritative position in the domains of modern art and design. Behind every piece of art lies a story intricately tied to the artist's background. Factors like nationality, gender, educational background, and experiences can significantly influence an artist's creative style, content, and audience reception[4]. Taking the French Impressionism and Japanese Ukiyo-e as examples, though these art forms emerged in the same era, they exhibit stark contrasts in style and content due to cultural and personal differences[4]. Thus, a deep exploration into the relationship between an artist's background and the popularity of their works in renowned institutions like MoMA holds significant academic and practical value[4]. This not only aids in understanding the multifaceted factors behind art creation and appreciation but also offers invaluable insights for art institutions, curators, educators, and researchers[5]. Moreover, for art investors and collectors, understanding this relationship can provide robust guidance for their decisions in the art market[6].

## 2. Research Background

Since its inception in 1929, the Museum of Modern Art (MoMA) has progressively established its avant-garde position in the global art community[1]. The museum was born amidst a desire to challenge traditional art museum paradigms and pursue innovative and experimental art[7]. Throughout its rich history, MoMA has been dedicated to propelling groundbreaking exhibitions, educational programs, and academic research, leaving an indelible mark on the evolution of modern and contemporary art. From its early acquisitions of Impressionist pieces to its extensive coverage of modern multimedia art, MoMA has emerged as a global epicenter for art, design, and film[8]. Delving deeper into the roots of artistic creation, the background of an artist emerges as an indispensable perspective[9]. This backdrop pertains to the environment in which an artist grew up, such as nationality and region, which profoundly influence their artistic endeavors[10]. Additionally, an artist's gender, identity, educational journey, and even the socio-political milieu they inhabit can shape their unique creative style and viewpoints[11]. As for the popularity of artistic works, it can be assessed from multiple dimensions. Direct audience feedback, exhibition and sales data, critiques from art critics, and even a work's position and impact in art history all offer multifaceted viewpoints to evaluate its reception[12].

## 3. Methodology

### 3.1 Data Source and Collection

The data utilized in this study was sourced from the Museum of Modern Art (MoMA)[1]. Founded in 1929, MoMA

acquired its inaugural collection that year. Today, the Museum's collection has evolved, encompassing nearly 200,000 works of art from across the globe, spanning the past 150 years. These collections cater to an ever-expanding range of visual expression, covering painting, sculpture, prints, sketches, photography, architecture, design, film, as well as media and performance arts.

The MoMA dataset meticulously captures detailed information about the museum's artworks and artists[13]. Specifically, this dataset comprises 130,262 entries, representing all the works that have been included in MoMA's collection and cataloged in its database. Each artwork's metadata, including title, artist, creation date, medium, dimensions, and the museum's acquisition date, is encompassed. It's noteworthy that some entries contain incomplete information and are marked as "not approved by the curator". Additionally, the artist dataset incorporates 15,091 entries, representing every artist with works in MoMA's collection that have been cataloged in the database. Essential metadata for each artist, such as name, nationality, gender, birth year, and year of passing, are included.

### **3.2 Data Preprocessing**

To ensure the accuracy and integrity of the analysis, a series of data preprocessing steps were undertaken[14]. Firstly, all entries with missing critical information were removed from the dataset. Subsequently, for numerical data, an interquartile range (IQR) method was adopted to identify and handle outliers. Specifically, values below  $Q1 - 1.5 * IQR$  or above  $Q1 + 1.5 * IQR$  were defined as outliers and excluded from the dataset [15].

### **3.3 Data Visualization and Statistical Analysis**

To gain a deeper understanding of the data distribution and characteristics, data visualization and statistical analyses were conducted. Initially, bar graphs were employed to visualize the representation of artists from different countries/regions and gender in MoMA. Furthermore, to explore the relationship between an artist's nationality and gender with the popularity of their works, a chi-square test was performed. The chi-square test is a statistical method employed to test the independence between two categorical variables.

### **3.4 Time Series Analysis**

To uncover the trends in art collection quantities over time, the data was grouped based on the creation date of artworks, and the number of artworks for each period was tallied. A line chart was utilized to display this trend, facilitating a better comprehension of art collection tendencies across different historical epochs. After grouping by year, the number of artworks acquired each year was calculated. Additionally, data grouped by the year of artwork acquisition and artist's gender, and by the year of artwork acquisition and artist's nationality, were analyzed to chart the annual count of artworks from the top ten countries/regions. Line graphs elucidated these statistical trends.

### **3.5 Machine Learning Model**

To further excavate latent insights from the data and forecast trends, a Random Forest algorithm was employed as the machine learning model[16]. Initially, the data was randomly partitioned into an 80% training set and a 20% testing set. The training set was then employed to train the Random Forest model. The trained model was subsequently validated on the testing set to evaluate its predictive performance. The objective was to predict the popularity of artworks in MoMA based on the background features of the artists. For this purpose, gender, nationality, and the artist's birth year (converted to age) were considered as features, while the "Acquisition Period" field (containing the categories 'Early', 'Mid', 'Recent') was used as the target variable. During the data preprocessing stage, categorical features were encoded, and numerical features were standardized.

## **4. Results**

The collection of the Museum of Modern Art (MoMA) in New York reveals distinctive patterns in the distribution of artists and their works based on nationality/region. Visualization indicates that the top ten countries/regions represented in MoMA's collection likely include art powerhouses like the United States, France, and Germany (As shown in Figure 1).

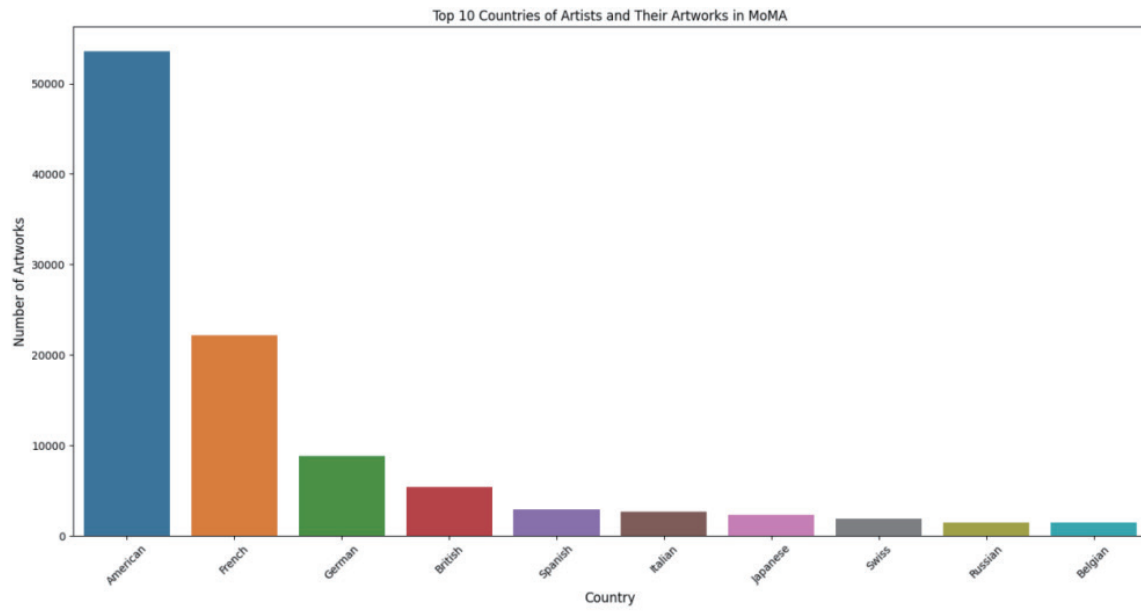


Figure 1. The distribution of artists and their works based on nationality/region.

Regarding gender distribution (Figure 2), works by male artists have historically dominated the collection. However, in recent years, as emphasis on gender diversity has grown, the number of works by female artists has been on the rise.

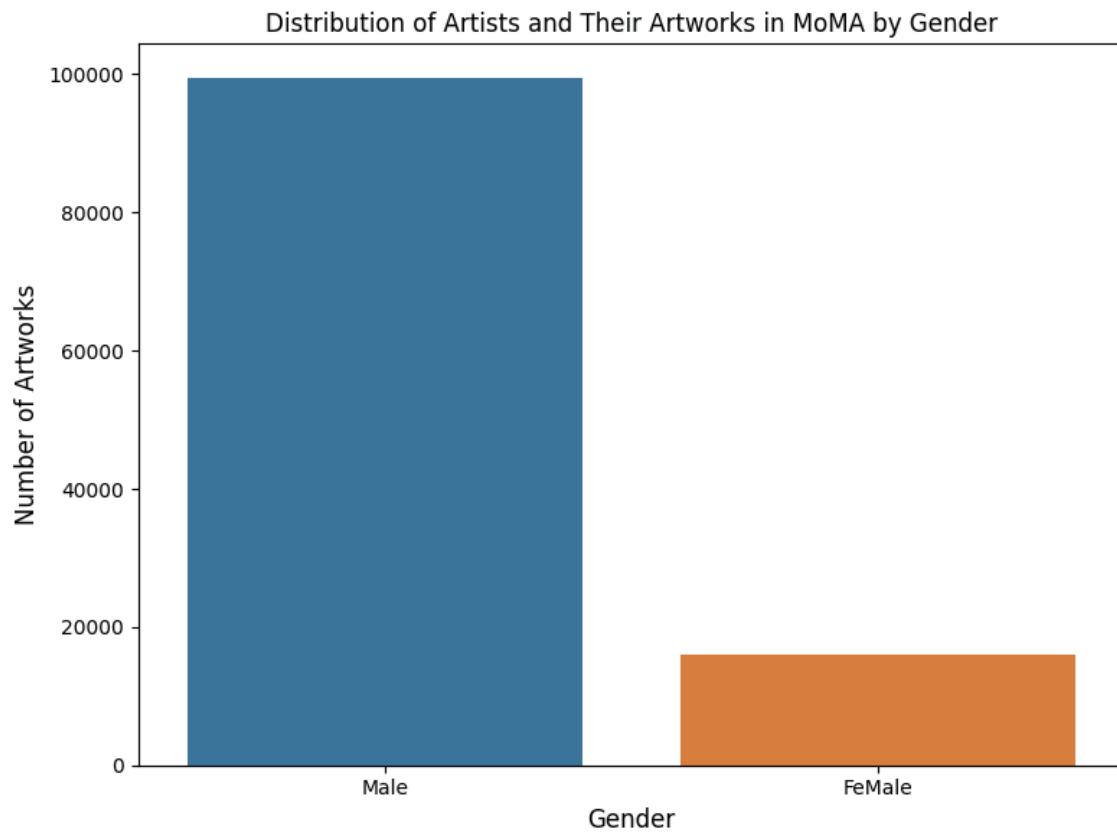
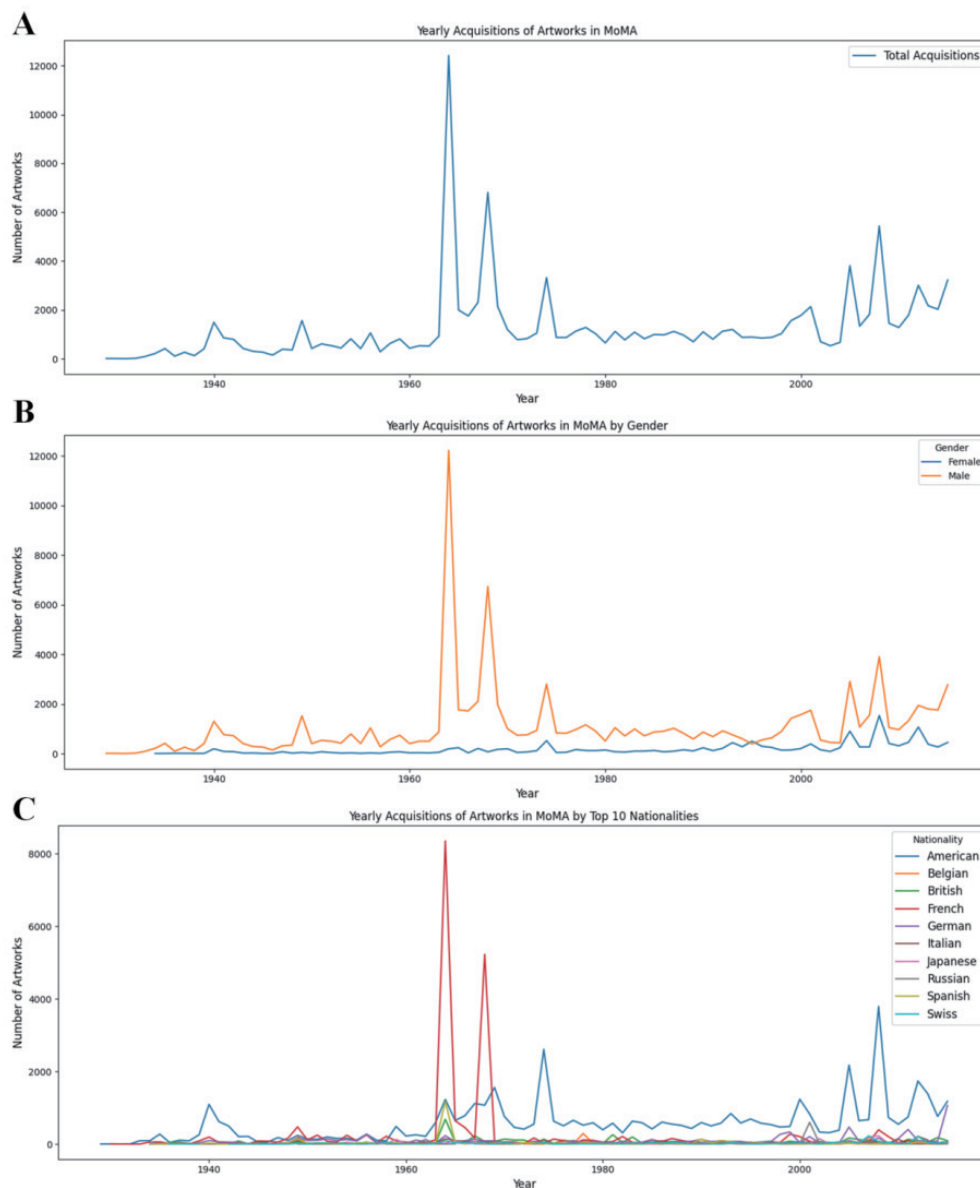


Figure 2. The gender distribution of artists.

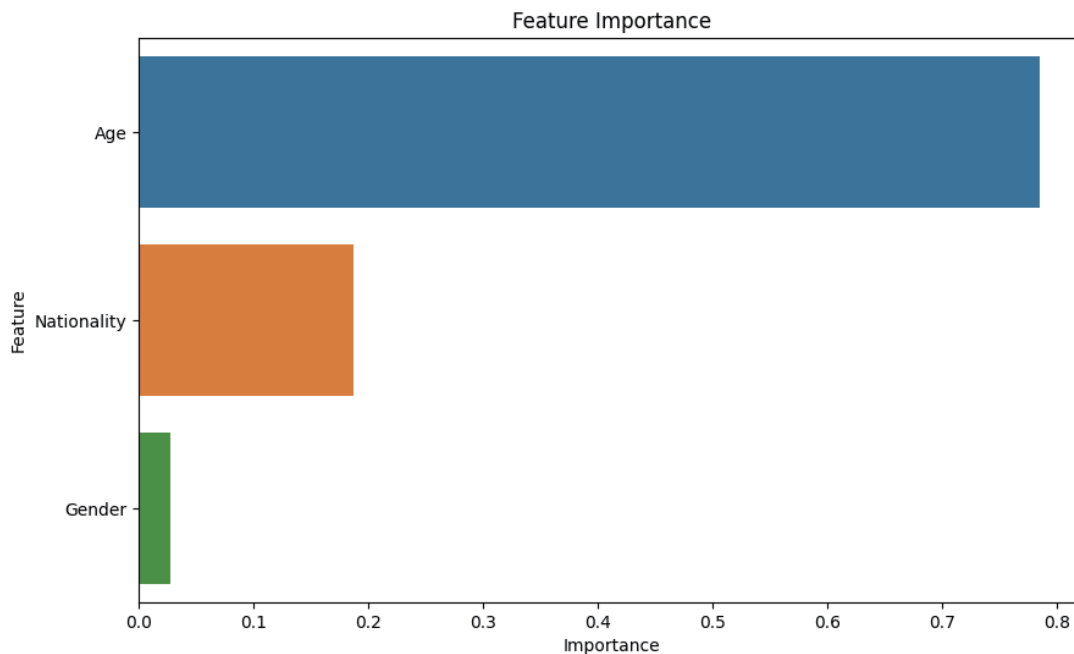


**Figure 3. The yearly distributions. A. Total acquisitions. B. Gender. C. Nationalities**

Delving deeper to explore the connection between an artist’s background and the popularity of their work within MoMA, several statistical analyses were conducted. The “popularity” of the artworks was inferred based on their acquisition date, hypothesizing that more recently acquired artworks are more popular. Accordingly, artworks were categorized into “Early”, “Mid”, and “Recent” acquisition periods. The chi-squared test revealed a significant relationship between an artist’s nationality and the popularity of their work, with a chi-squared statistic of 15376.89 and a p-value of 0.0. Similarly, ANOVA results indicated a significant relationship between the gender of the artist and the popularity of their work, with an F-statistic of 4572.16 and a p-value of 0.0.

In the time series analysis, the aim was to discern the influence of an artist’s background on the popularity of their work across different historical periods. Overall, MoMA’s annual acquisition rate of artworks has been increasing, especially post-2000. While acquisitions of works by male artists have consistently been higher, there has been a noticeable uptrend in acquisitions of female artists’ works in recent years. Moreover, in terms of nationality, acquisitions of works by American artists have consistently been the highest, while other countries, such as France and Germany, have maintained a stable acquisition rate.

Subsequent research aimed at predicting the popularity of artworks in MoMA based on artists’ background features using the Random Forest classification model.



**Figure 4. The feature importance for prediction model.**

This model achieved an overall accuracy of 0.75, indicating that 75% of its predictions were accurate. The performance across categories was relatively balanced, with metrics like precision, recall, and F1-score ranging between 0.70 and 0.78. The feature importance can be found in Figure 4. Among them, Age and gender is the most and less important parameters.

## 5. Conclusion

The art collection of the Museum of Modern Art (MoMA) in New York underscores a significant relationship between an artist's background, particularly nationality and gender, and the popularity of their works within MoMA. The findings suggest a pronounced representation of American artists in MoMA's collection, with a notable dominance of male artists' works over their female counterparts. This might reflect MoMA's historical acquisition trends or hint at the importance of artists with specific backgrounds in the narrative of modern art. Over time, especially post-2000, there has been an uptick in MoMA's annual art acquisitions. While works by male artists have consistently outnumbered their female counterparts, there's an emerging trend of increasing acquisitions of female artists' works, possibly reflecting a growing emphasis on gender diversity. Additionally, while American artists' works have always been predominantly acquired, other countries, such as France and Germany, have sustained a steady acquisition volume.

In our analysis, we observed that "Age" is the most significant feature in determining art acquisitions, accounting for 78.3% of importance. This suggests that the age of the artist might have a substantial influence on the likelihood of their artwork being acquired. It could be that as artists age, their works are more likely to be considered historically valuable or unique, thus making them more attractive for art institutions to acquire.

Furthermore, the "Nationality" feature holds an importance of 18.9%, which is significantly higher than the "Gender" feature at 2.7%. This implies that an artist's nationality has a more profound impact on the acquisition potential of their work compared to their gender. This may reflect the influence and acceptance of art from certain countries or regions in the international art market.

Based on these observations, several recommendations are presented for MoMA and other art institutions. To enrich and enhance the diversity of their collections, institutions should amplify and promote artists from backgrounds that are underrepresented in their collections, such as female artists or artists from non-dominant countries. This diversity can not only augment the institution's art collection but also present the public with a broader spectrum of artistic perspectives. Moreover, given technological advancements and data availability, art institutions should adopt data-driven approaches to optimize their art acquisition strategies, ensuring comprehensive and representative collections. The prediction using the Random Forest classification model, with an accuracy of 0.75, offers robust decision-making support for art acquisitions.

However, this study has its limitations. Using MoMA's acquisition date as a proxy for the popularity of artworks might not fully encapsulate the true preferences of the general public. Due to the absence of other pivotal data, such as the number of times an artwork is viewed or public comments, the study's conclusions might not capture the complete picture of an artwork's actual popularity. For future research, it is recommended to explore more data sources and consider other potential influencing factors, like the artist's educational background or the style/theme of the work. Additionally, qualitative research, like audience interviews, could provide more in-depth and nuanced insights into this topic.

## References

---

- [1] ASSIL C. MoMA and American Art: A Re-Evaluation of the Museum of Modern Art and its Presumed Eurocentric Attitude [J]. 2023.
- [2] RADBOURNE J. Arts management: A practical guide [M]. Taylor & Francis, 2023.
- [3] CURTIN A. Kathleen O'Connor of Paris [M]. Fremantle Press, 2018.
- [4] WERSKEY G, WILSON N J J O A S. The Artists of the Atlas: Their Role in Creating Settler-Colonial Australia's Visual Culture [J]. 2023, 47(2): 322-43.
- [5] GAGLIARDI D M, CHIARELLA S G, MARENGHI F, et al. Redefining the role of Artificial Intelligence in artistic production [J]. 2023.
- [6] FINDLAY M. The Value of Art (New, expanded edition): Money. Power. Beauty [M]. Prestel Verlag, 2022.
- [7] DUARTE A, PÉREZ-IBÁÑEZ M. Introducing the Art Market and the Global South [M]. The Art Market and the Global South. Brill. 2023: 1-15.
- [8] ZHANG L. National Discourse National Discourse, Local Voice Local Voice: Contemporary Chinese Art of the 1980s [M]. The Spirit of Individualism: Shanghai Avant-Garde Art in the 1980s. Springer. 2023: 35-59.
- [9] VUONG Q-H, BUI Q-K, LA V-P, et al. Cultural evolution in Vietnam's early 20th century: A Bayesian networks analysis of Hanoi Franco-Chinese house designs [J]. 2019, 1(1): 100001.
- [10] QUEMIN A J S, ANTROPOLOGIA. The impact of nationality on the contemporary art market [J]. 2015, 5: 825-56.
- [11] CHILDS S, KROOK M L J P. Gender and politics: The state of the art [J]. 2006, 26(1): 18-28.
- [12] FREYTES FREY A, CROSS C J A R. Overcoming poor youth stigmatization and invisibility through art: A participatory action research experience in Greater Buenos Aires [J]. 2011, 9(1): 65-82.
- [13] SRINIVASAN R, DENTON E, FAMULARO J, et al. Artsheets for Art Datasets; proceedings of the Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), F, 2021 [C].
- [14] ZHANG L, ZHANG L, DU B J I G, et al. Deep learning for remote sensing data: A technical tutorial on the state of the art [J]. 2016, 4(2): 22-40.
- [15] Interquartile Range [M]. Wiley StatsRef: Statistics Reference Online, 2014.
- [16] LIANG J J W C, COMPUTING M. Problems and Solutions of Art Professional Service Rural Revitalization Strategy Based on Random Forest Algorithm [J]. 2022, 2022(1): 1-11.