MoMAmentum : Capturing MoMA’s Momentum

March 25th, 2025
Isha Patel
ip02891n@pace.edu
Class Name: Practical Data Science
Program Name: MS in Data Science
Seidenberg School of Computer Science and Information Systems
Pace university

Agenda

- Executive summary
- Project plan recap
- Data
- Exploratory data analysis
- Modeling methods
- Findings
- Business recommendations and technical next steps

Executive summary

Problem Statement:

- MoMA struggles to choose the right artworks for exhibitions, impacting visitor engagement. They need to identify trending mediums and artworks that attract the most visitors.

Solution Summary:

- Explore MoMA's dataset to identify trends in artwork popularity, including mediums, artists, and key features that attract visitors.
- Build models to predict which types of artwork are likely to engage more visitors in upcoming exhibitions.
- Provide insights to guide MoMA in making data-driven decisions on selecting artworks that will maximize visitor engagement.

Project plan recap

Deliverable	Due Date	Status
Data & EDA	**3/25/2025**	***COMPLETED***
Methods, Findings, and Recommendations	**4/1/2025**	***COMPLETED***
Final Presentation	**4/22/2025**	***IN PROCESS***

Data

Data

Data Details:

- **Data Source:** Museum of Modern Art Collection
- **Sample Size:** ~138,000 records
- **Time Period:** Acquisitions from 1929 to present.
  - **Included Data:** Artwork details (Title, Artist, Medium, Classification).
  - Acquisition year.
  - Dimensions (Height, Width).
  - **Excluded Data:** Missing or inconsistent acquisition dates.
  - Original data size: 138,000 rows and 21 columns
  - Data size after removing outliers and Missing Values: 112,904 rows and 15 columns

Assumptions:

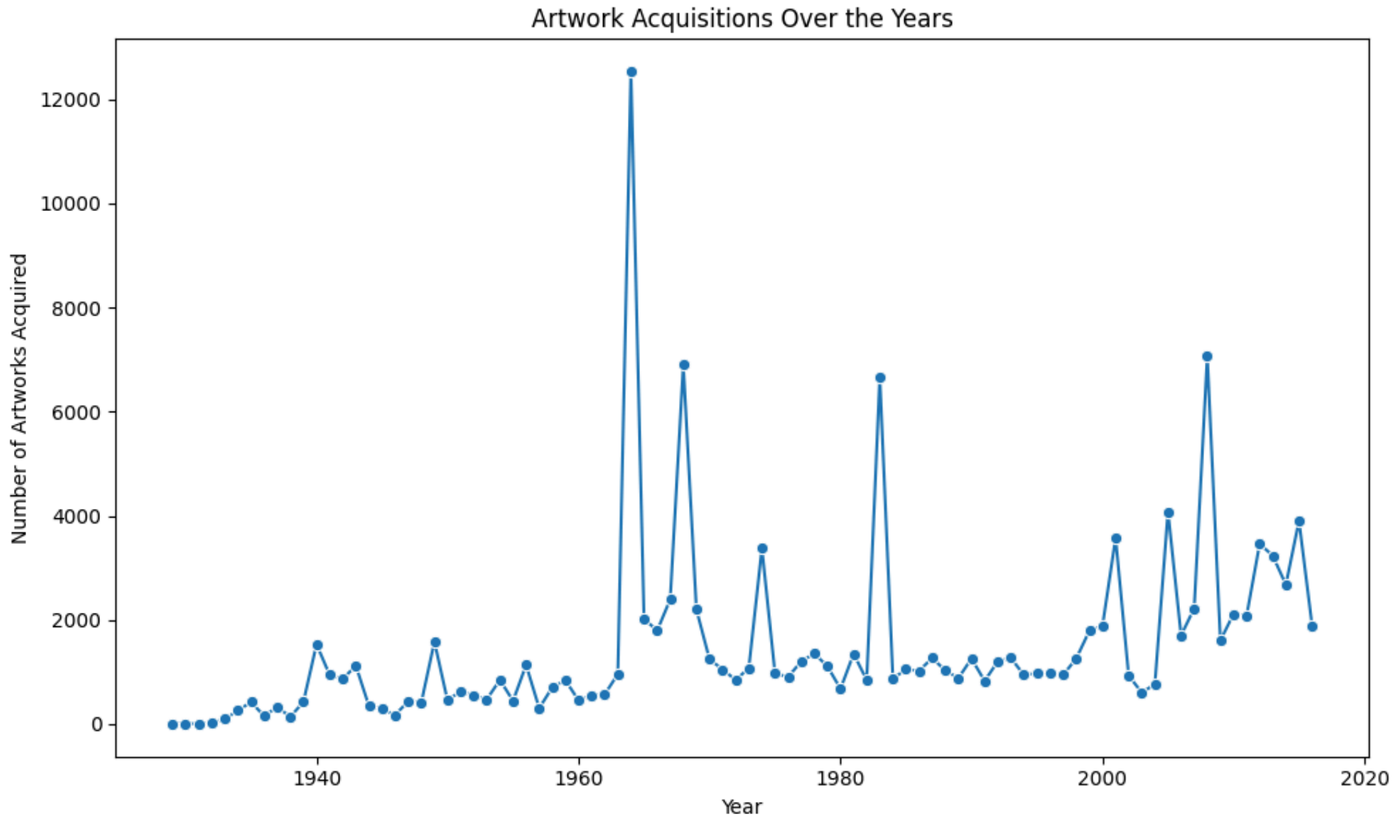
- The dataset accurately reflects MoMA's acquisitions.
- The acquisition trends observed in past data will continue in the future.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA)

Artwork Acquisitions Over the Years

Visualization:



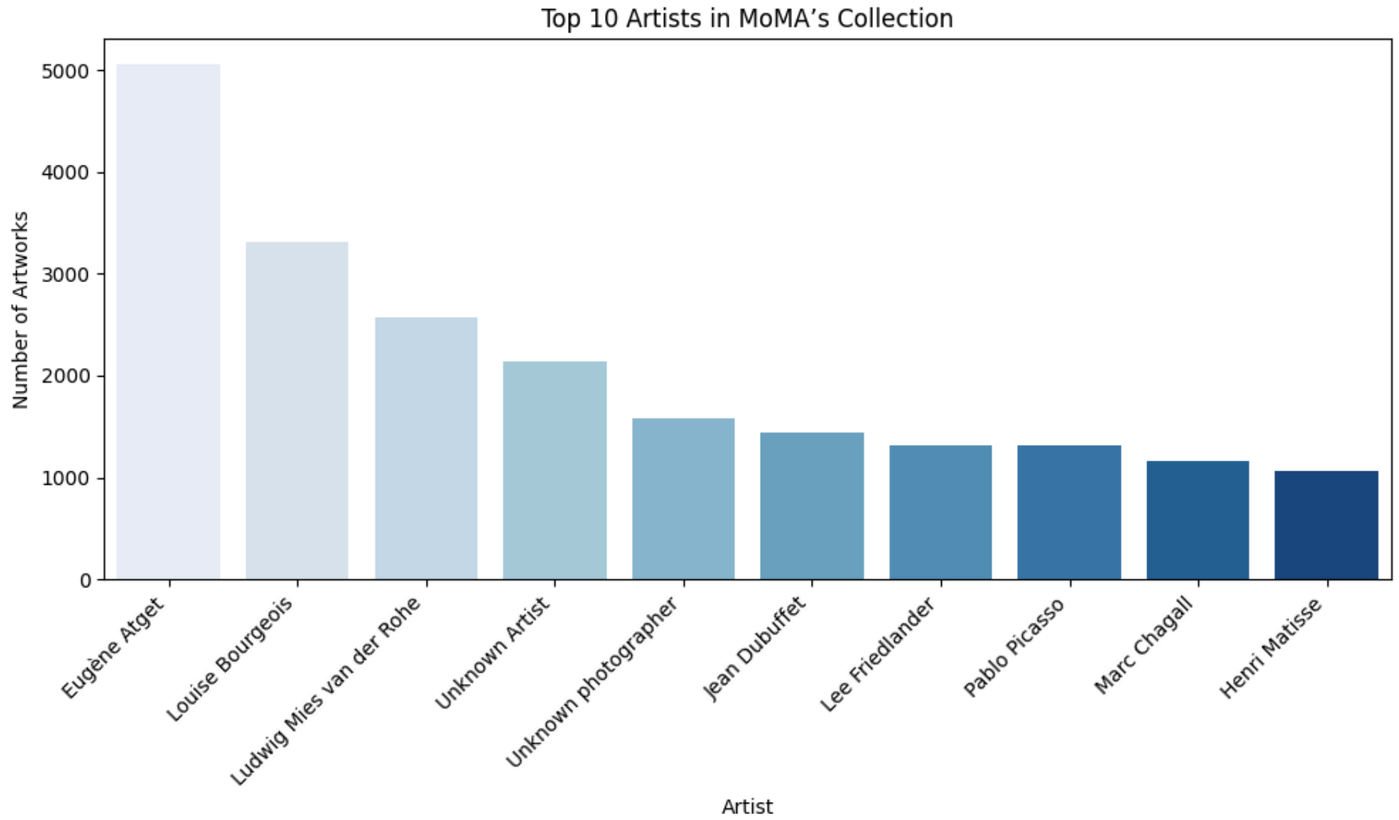
Key Takeaway-

MoMA acquired the most artworks between 1960-1970, possibly due to a cultural shift or movement that encouraged more acquisitions. After that, acquisitions dropped and stabilized around 6,000, suggesting a more consistent approach to collecting art in recent decades.

Exploratory Data Analysis (EDA)

Popular Artists in MoMA’s Collection

Visualization:



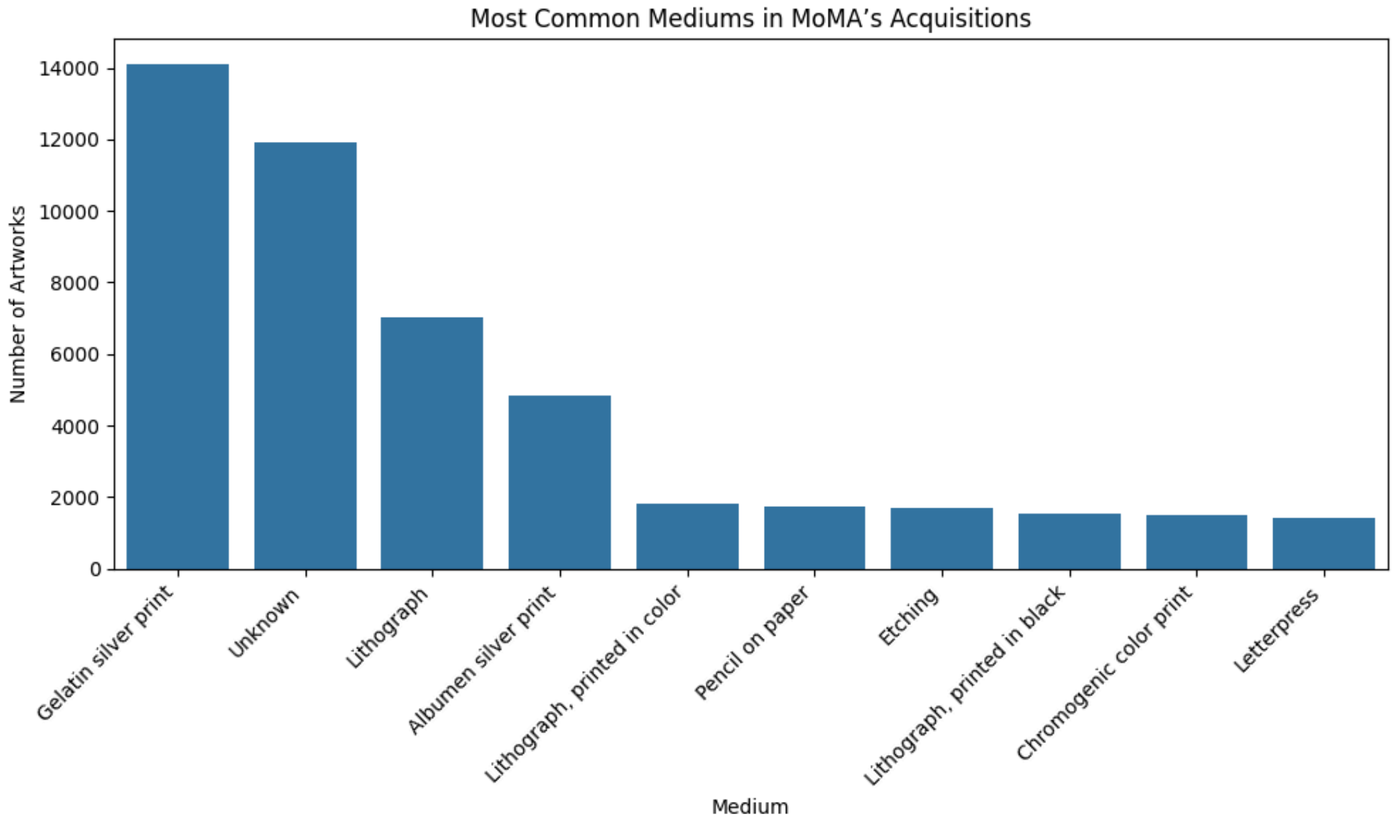
Key Takeaway-

Artists like Eugène Atget and Louise Bourgeois have more artworks in MoMA’s collection because they were active for longer periods. Bourgeois, especially in the 20th and 21st centuries, benefited from a time when modern and contemporary art was more popular. Recently, MoMA seems to be collecting from a wider range of artists, focusing on diversity instead of just a few big names.

Exploratory Data Analysis (EDA)

Distribution of Artworks by Medium

Visualization:



Key Takeaway-

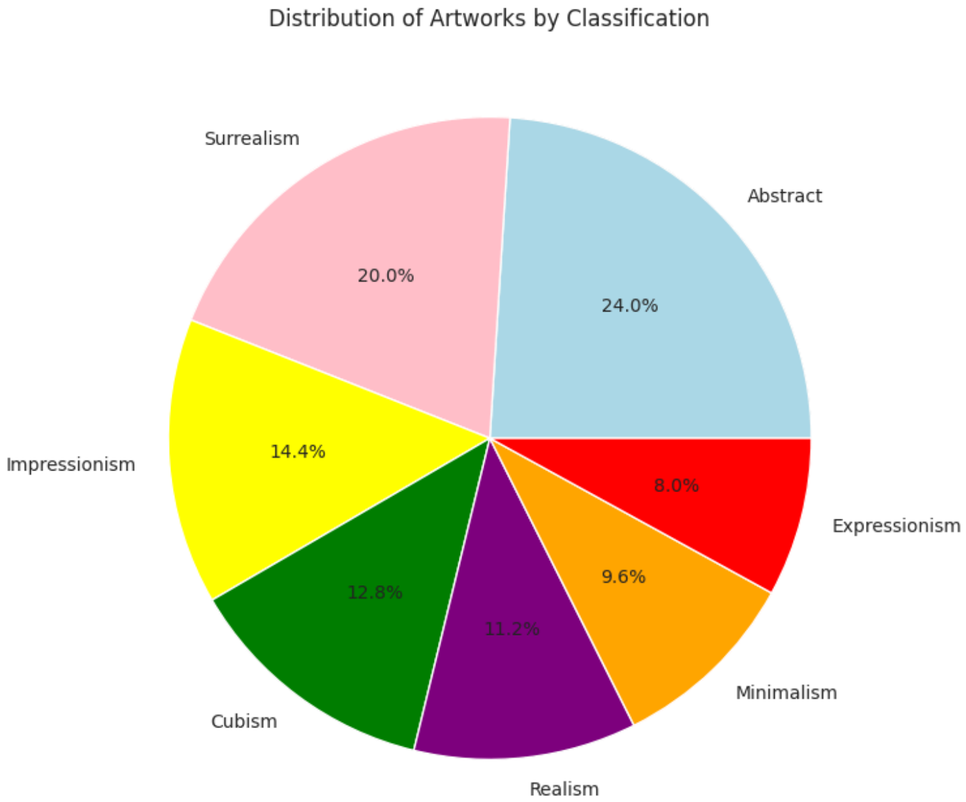
If certain mediums are growing in popularity, like Gelatin Silver Print, are becoming more popular, MoMA could consider featuring these mediums more in future exhibitions to capture visitor interest.

There's a large portion labeled "Unknown" with around 12,000 artworks, which might make it harder to fully understand the distribution of mediums.

Exploratory Data Analysis (EDA)

Distribution of Artworks by Classification

Visualization:



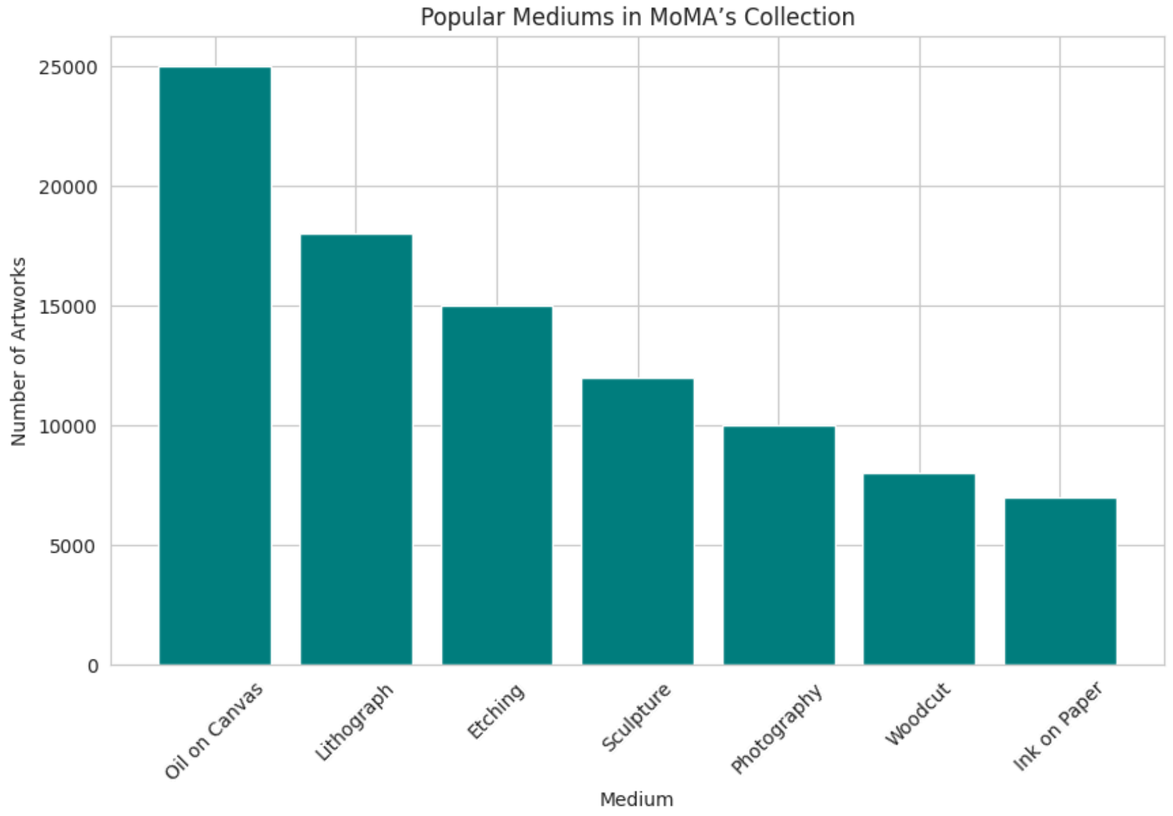
Key Takeaway-

MoMA's collection is heavily focused on Abstract art, which makes up 24% of the collection, but also maintains a diverse representation of various art movements, such as Surrealism and Impressionism, showcasing a broad mix of artistic styles.

Exploratory Data Analysis (EDA)

Popular Mediums in MoMA’s Collection

Visualization:



Key Takeaway-

The prominence of "Oil on Canvas" highlights MoMA's strong focus on traditional painting techniques, while other mediums like "Ink on Paper" and "Woodcut" remain less emphasized, potentially indicating areas for future curatorial exploration.

Modeling Methods

Model Overview

Model Chosen: Logistic Regression

- **Why Logistic Regression?**

Logistic Regression is a simple and effective model for predicting outcomes, such as whether an artwork is likely to be exhibited or not. Logistic Regression is well-suited for classified data because it estimates the probability of a certain class.

- **Why this model?**

We chose Logistic Regression because:

  - It's straightforward to understand and communicate.
  - It efficiently handles the binary classification problem (exhibited or not).
  - We can interpret the coefficients (importance of each feature), providing insights into which factors matter most for exhibition likelihood.

Features

(What Drives Predictions)

Key Features:

- **Acquisition Year:** The year when the artwork was acquired.
  Hypothesis: Newer acquisitions might have a higher chance of being exhibited.
- **Medium:** The material or medium used in the artwork (e.g., oil, sculpture).
  Hypothesis: Certain mediums could be more aligned with current exhibition trends.
- **Height & Width:** The physical size of the artwork.
  Hypothesis: Larger artworks might have higher chances of being exhibited in large-scale exhibitions.
- **Department:** The specific category or department the artwork belongs to, such as Painting & Sculpture, Architecture, etc.
  Hypothesis: Some departments may have a higher exhibition likelihood than others.

Feature Grouping:

These features are categorized into:

- Artwork Characteristics (e.g., Acquisition Year, Medium)
- Physical Traits (e.g., Height, Width)
- Exhibition Potential (e.g., Department)

Outcome Variable

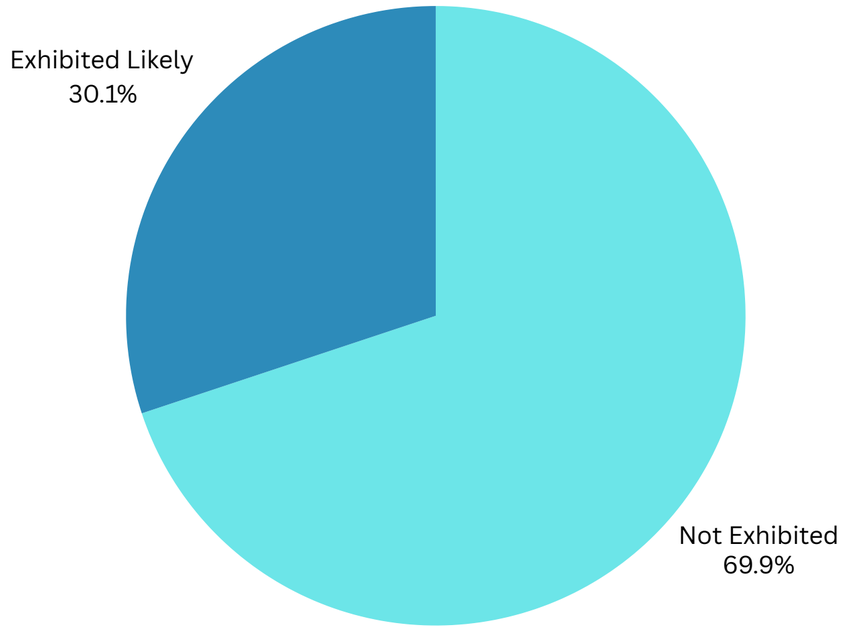
(What we're trying to predict)

The outcome variable is whether an artwork is likely to be exhibited in the future. Predicting this helps MoMA prioritize which pieces to showcase based on acquisition year, medium, and other factors.

- **Exhibited Likely:**
  - 0 = The artwork is **NOT** likely to be exhibited.
  - 1 = The artwork **IS** likely to be exhibited.

Output Breakdown:

- 0 (Not Exhibited): **69.9%** of the artworks in our dataset are predicted to not be exhibited.
- 1 (Exhibited Likely): **30.1%** of the artworks are predicted to be exhibited.



Findings

Key Findings & Insights

Model Chosen: Logistic Regression

Model Performance: Accuracy = 98.9%

The key factors that influence whether an artwork is exhibited are:

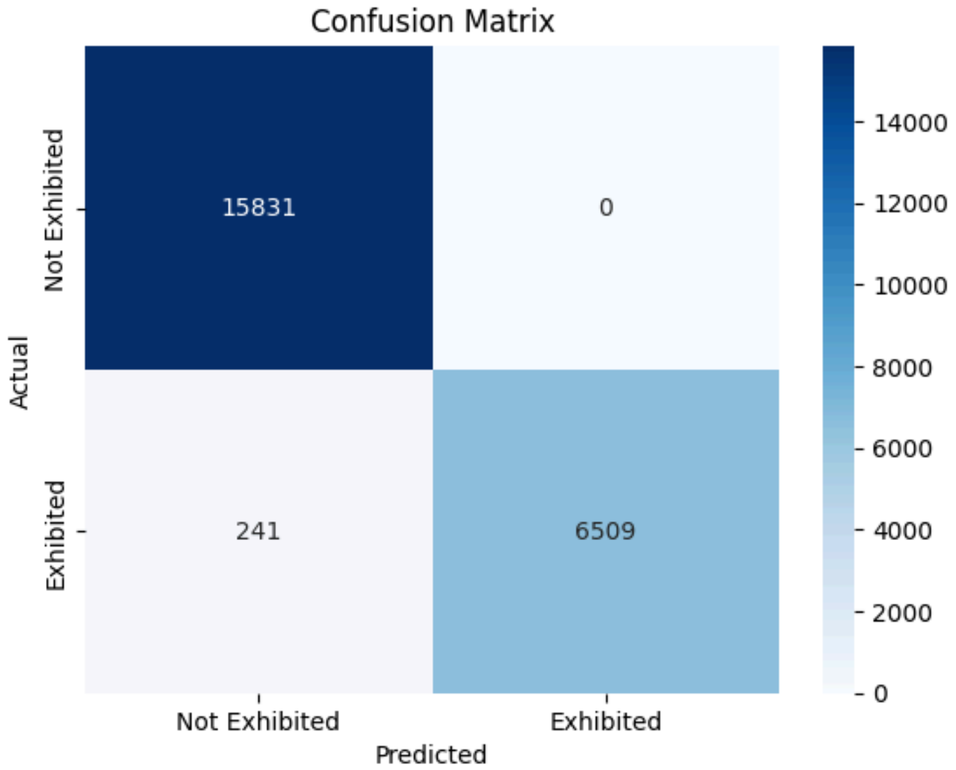
- **Acquisition Year:** More recent acquisitions have a higher likelihood.
- **Medium:** certain materials, may be more popular in exhibitions
- **Department:** Painting & Sculpture has the highest exhibition rate.

Surprising Finding:

- Physical Dimensions didn't matter as much!

Key Takeaway:

- **Our model is reliable in identifying which artworks are most likely to be exhibited, helping MoMA make data-driven decisions.**



Example

Scenario:

Imagine it's 2013, and MoMA is curating an exhibition. Oil paintings are trending, and we want to feature the top 5 newly acquired artworks from 2012.

	Artwork ID	Title	Artist ID	Name	Date	Medium	Dimensions	Acquisition Date	Catalogue	Department	Classification	Object Number	Height (cm)	Width (cm)	Exhibited Likely	Acquisition Year
110018	156383	Uphe	28538	Tomma Abts	2011	polymer	(48.1 x 38.1	2/15/12	Y	Sculpture	Painting	524.2012	48.101346	38.100076	1	2012
110094	156760	Untitled	4026	Joan Mitchell	1957	Oil on paper	19 1/2 x 17 1/2" (49.5 x 44.5 cm)	2/14/12	Y	Drawings	Drawing	499.2012	49.530099	44.450089	1	2012
110117	156800	Untitled	41332	Bernardo Ortiz Campo	2002	Oil on paper	8 1/2 x 12 7/8" (21.6 x 32.7 cm)	2/14/12	Y	Drawings	Drawing	507.2012	21.590043	32.702565	1	2012
110118	156801	Untitled	41332	Bernardo Ortiz Campo	1995	Typewriting and linseed oil on photocopy	13 x 8 1/2" (33 x 21.6 cm)	2/14/12	Y	Drawings	Drawing	508.2012	33.020066	21.590043	1	2012
110124	156807	Untitled	41332	Bernardo Ortiz Campo	2010	Oil on paper	8 1/8 x 6 1/8" (20.6 x 15.6 cm)	2/14/12	Y	Drawings	Drawing	512.2012	20.637541	15.557531	1	2012

Key Takeaway:

- MoMA's 2012 acquisitions confirm that oil paintings remain a priority, with all selected works exhibited. However, the dominance of a single artist raises the need for more diversity. Featuring these in the 2013 exhibition aligns with visitor engagement trends, but exploring other mediums could enhance variety.

Business Recommendations & Technical Next Steps

Business Recommendations

1. Finding 1: Recent acquisitions dominate exhibitions

- **Insight:** The model confirms that newly acquired artworks are highly exhibited.
- **Actionable Step:** MoMA should prioritize acquiring and promoting emerging artists' works to maintain engagement with contemporary trends.

2. Finding 2: Oil paintings continue to attract attention

- **Insight:** Oil paintings remain a strong presence in exhibitions, but trends may shift.
- **Actionable Step:** MoMA should experiment with different mediums and track audience interest to diversify future collections.

Technical Next Steps

- **Expand Dataset –** Collect data on artworks from different periods and mediums to provide a broader perspective.
- **Predict Future Trends –** Develop a forecasting model that identifies upcoming artistic trends based on acquisition patterns and audience engagement.
- **Evaluate Model for Production –** Assess whether the current model's predictions are reliable enough for long-term use or if adjustments are needed.

Key Takeaway:

- Our model helps predict which artworks are more likely to be exhibited based on certain factors. With a solid accuracy of 98.9%, it's a useful tool for MoMA to make smarter decisions about future exhibitions. While the model is already performing well, there's still room to improve it by testing more advanced techniques and gathering more data. This could help MoMA stay ahead of trends and give visitors an even better experience.

We will get back with the final Touch Soon...

Thank You