

Maica Crovetto

Data Science II
Entrega final

TABLA DE CONTENIDO

1.- Contexto y Audiencia	3
2.- Hipótesis/Preguntas de interés	4
3.- Resumen de datos	4
4.- Metadata	5
5.- Análisis exploratorio	6
6.- Conclusión	11

CONTEXTO

El arte es una manifestación cultural que ha evolucionado a lo largo de la historia, y los museos desempeñan un papel clave en la preservación y difusión de este patrimonio. En la era digital, el análisis de datos ha permitido a instituciones como el Museo de Arte Moderno de Nueva York gestionar sus adquisiciones, estudiar tendencias artísticas y entender mejor el impacto de sus colecciones.

El MoMA, fundado en 1929, es una de las instituciones más influyentes en el mundo del arte moderno y contemporáneo. Su colección abarca más de 200.000 obras de arte, incluyendo pinturas, esculturas, fotografías y nuevas formas de expresión artística. El proceso de adquisición de nuevas obras es crucial para su misión de representar la diversidad del arte a nivel global. En este proyecto, se analizarán datos de sus adquisiciones, con el objetivo de identificar patrones en las obras adquiridas. A través de técnicas de análisis de datos y aprendizaje automático, se buscará responder preguntas como: ¿qué factores influyen en la selección de una obra?, ¿qué nacionalidades están más representadas?, ¿existen tendencias en los materiales o técnicas utilizadas en las adquisiciones?

AUDIENCIA

Esta investigación se enfoca en proporcionar una herramienta de análisis estratégico para los niveles gerenciales y directivos del museo. Proveerles de información relevante sobre la adquisición de obras, facilitando la toma de decisiones y la planificación estratégica a largo plazo.

HIPÓTESIS / PREGUNTAS DE INTERÉS

- ¿Cuál es la probabilidad de que una obra de un artista previamente adquirido sea seleccionada para futuras adquisiciones del MoMA?
Variables: Artista, Fecha de adquisición, Número de obras adquiridas previamente.
- ¿La obra es moderna o no? (año \geq 1980)
Variables: Departamento, Medio, Clasificación.
- ¿Cómo han evolucionado las preferencias del MoMA en cuanto a técnica o material de las obras adquiridas desde 1950?
Variables: Material, Técnica, Fecha de adquisición.
- ¿A que departamento pertenece la obra?
Variables: Departamento, Medio, Clasificación, Año de adquisición.
- ¿Qué países tienen mayor representación en las adquisiciones del MoMA y cómo ha cambiado esa tendencia a lo largo de los años?
Variables: País del artista, Año de adquisición, Técnica, Tipo de obra.
- ¿La obra es nacional (EEUU) o extranjera?
Variables: Departamento, Medio, Clasificación, Año de adquisición, Género, Artista.

RESUMEN DE DATOS

Data set: **MoMA Artworks**

Cantidad de filas

132404

Cantidad de columnas

29

METADATA



Departamentos



Media and
Performance Art
2685



Design
41



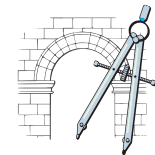
Film
3288



Painting &
Sculpture
3815



Photography
29475



Architecture
23



Drawings
11269



Prints &
Illustrated books
60604

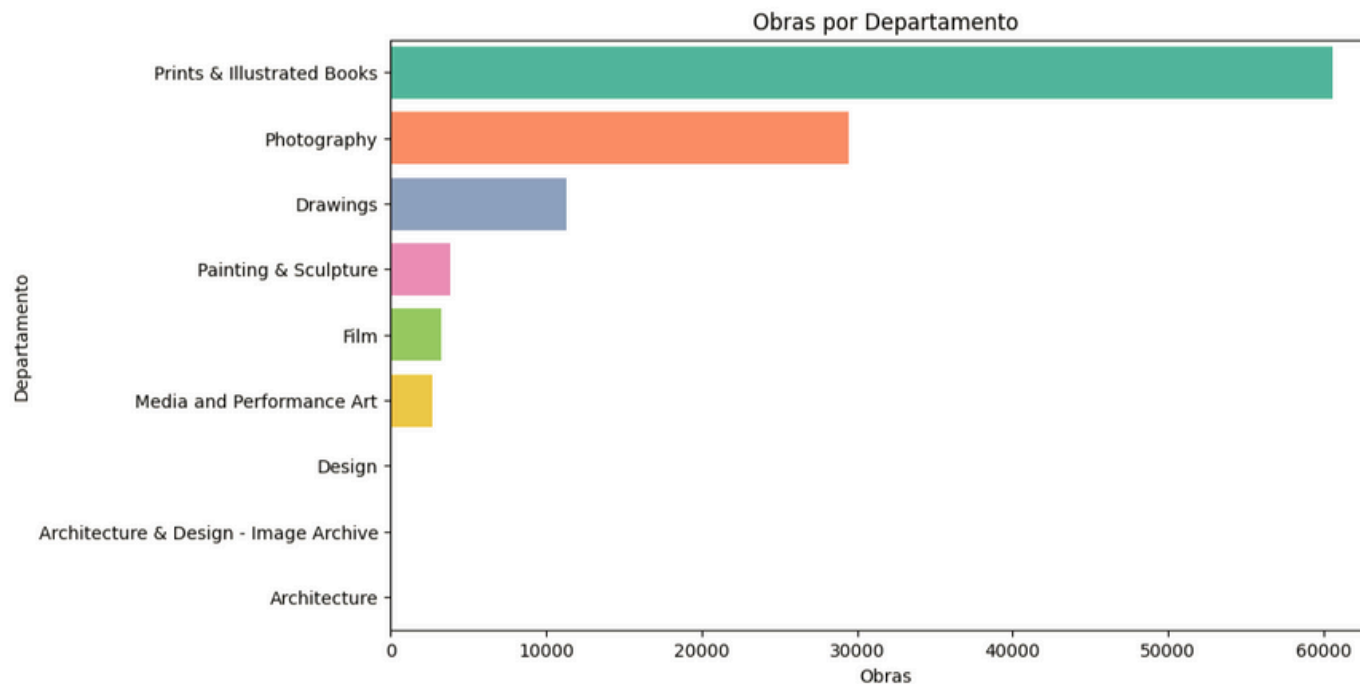
**Obras por
departamento**

ANÁLISIS EXPLORATORIO

- *Cantidad de obras actuales pertenecientes a cada departamento.*

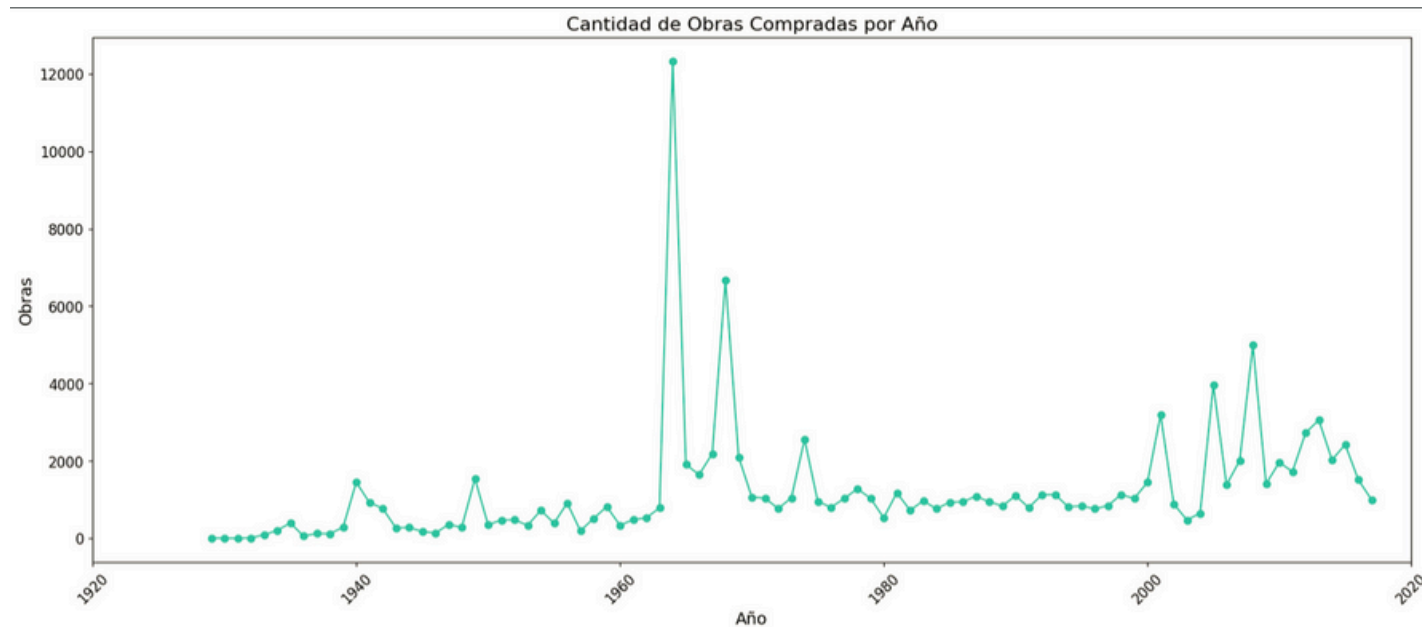
El departamento con mayor número de obras es "Prints and Illustrated books", seguido por "Photography".

Esto podría reflejar un enfoque curatorial específico hacia estas dos formas de expresión artística, o una mayor disponibilidad de este tipo de obras en el mercado. El dominio de "Prints and Illustrated books" sugiere que las técnicas de impresión han sido altamente valoradas y adquiridas a lo largo del tiempo.



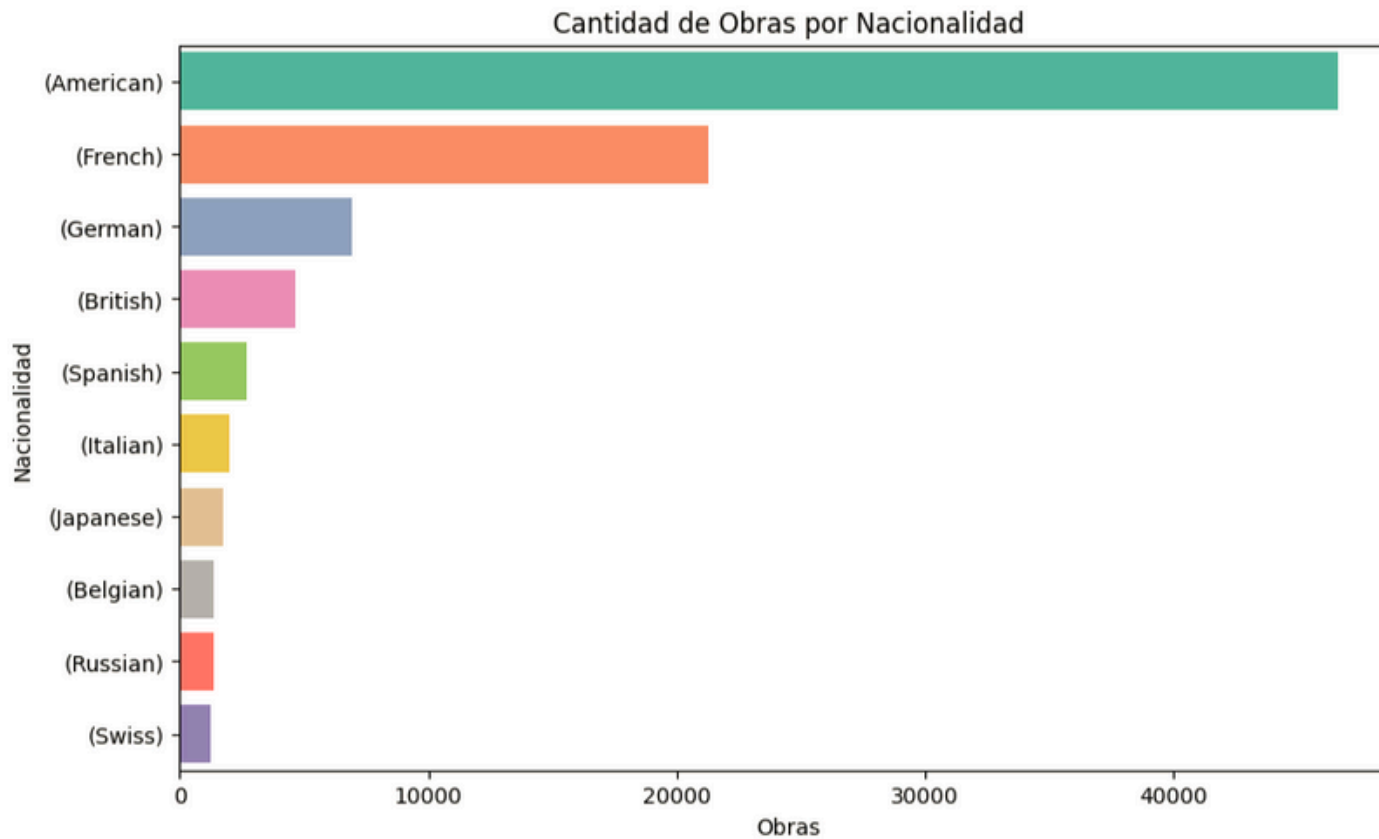
● *Cantidad de obras adquiridas por año*

La gráfica muestra un notable incremento en las adquisiciones de obras durante la década de 1960, y otro pico significativo en los años 2000. Esto podría deberse a factores históricos o económicos, como el auge del arte moderno y contemporáneo en los 60, o una expansión de la política de adquisiciones del museo en los 2000. Los períodos de mayor adquisición podrían estar asociados con movimientos artísticos importantes o cambios en las prioridades curatoriales.



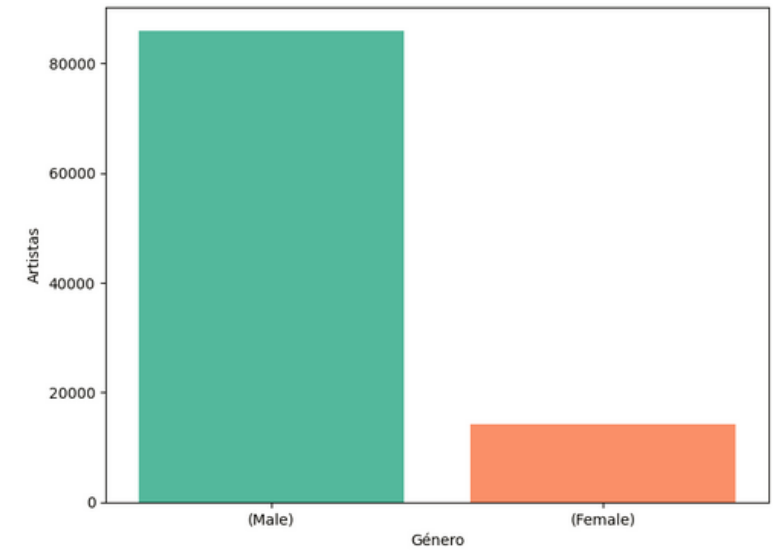
● *Nacionalidades con mayor representación*

Las nacionalidades que poseen mayor cantidad de obras en el museo son la americana, francesa y alemana. Esto sugiere que el MoMA ha puesto un énfasis particular en adquirir obras de artistas de estos países, lo que puede estar vinculado a la influencia de estas naciones en la historia del arte moderno y contemporáneo.



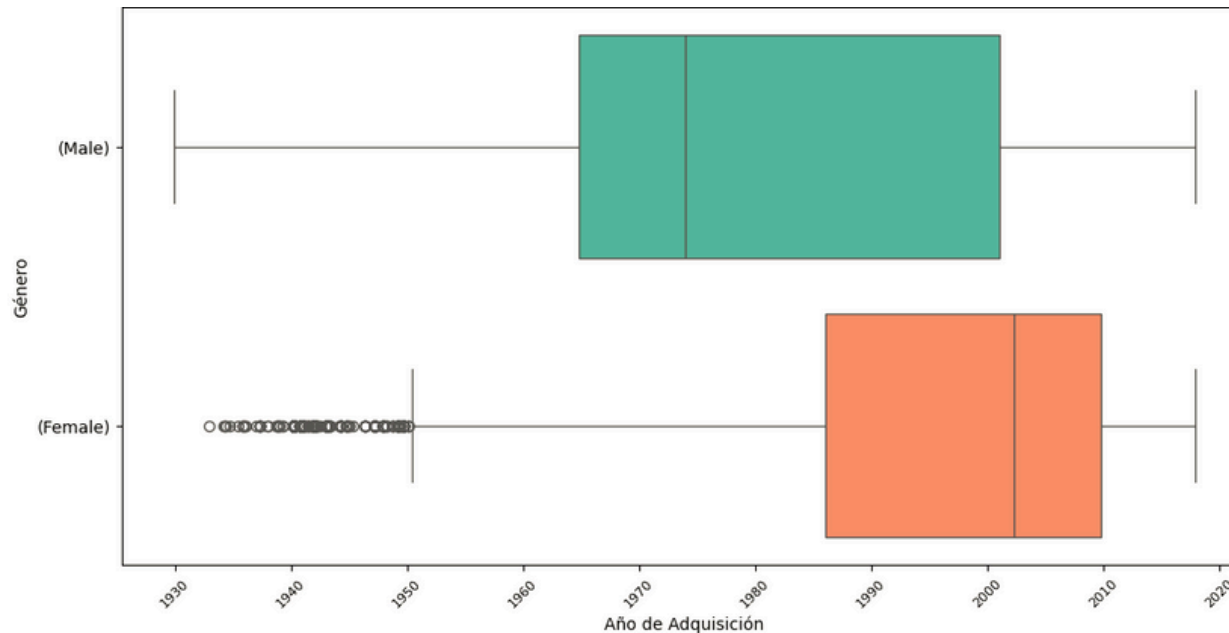
● *Cantidad de artistas por género*

El gráfico muestra una clara diferencia en el número de artistas masculinos frente a los femeninos, con una mayor representación de hombres. Esta disparidad refleja un sesgo histórico en la representación de artistas en museos, y podría ser el resultado de barreras de género en el mundo del arte a lo largo de la historia. Este dato pone en evidencia la importancia de revisar las políticas de adquisiciones para asegurar una mayor inclusión de artistas femeninas en el futuro.



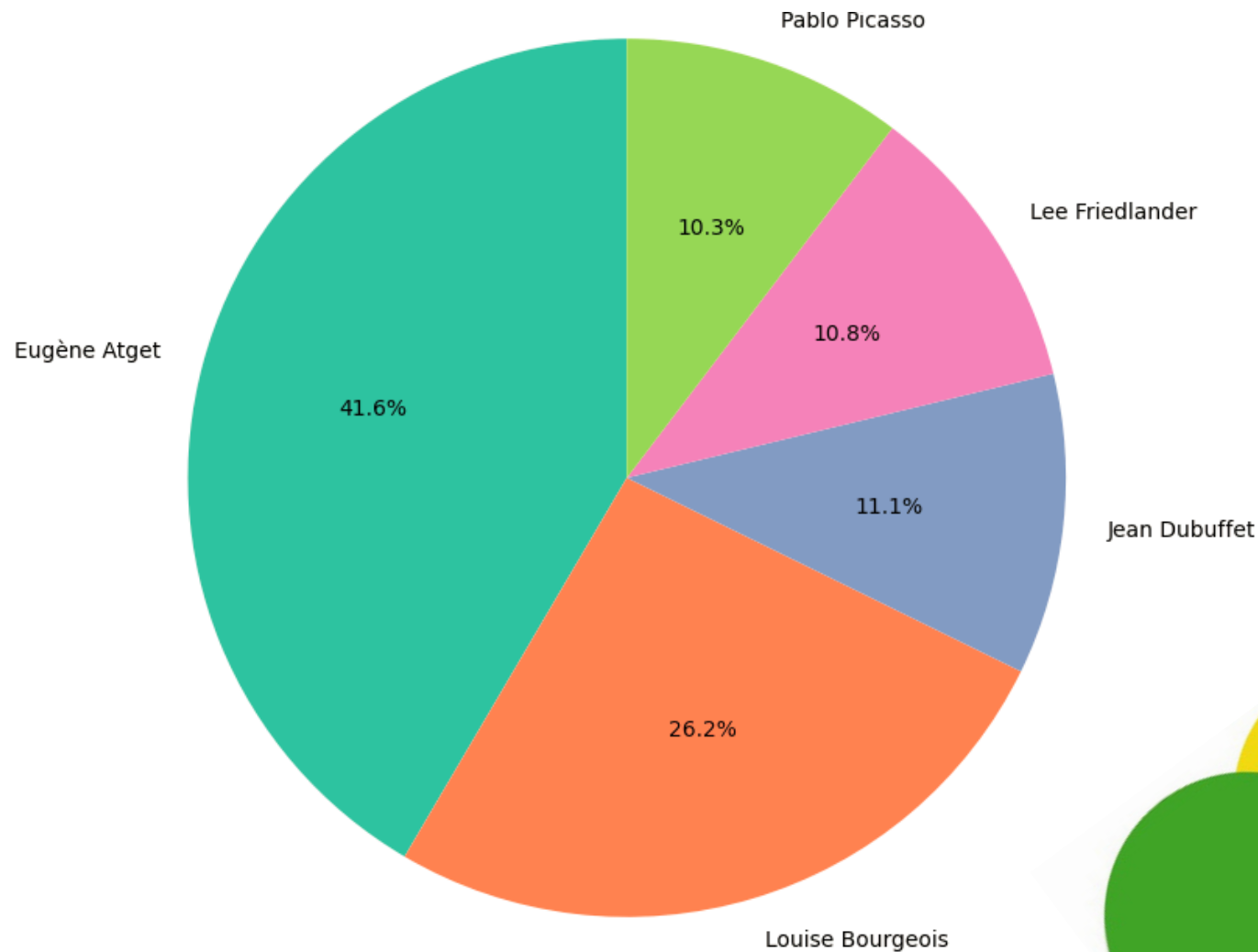
● *Distribución de adquisiciones por año y género*

La mayor parte de las adquisiciones de obras de artistas masculinos se concentran entre 1950 y 1990, mientras que las obras de artistas femeninas comienzan a ser adquiridas en mayor cantidad a partir de finales de la década de 1980. Los outliers observados en las adquisiciones de mujeres antes de 1950 indican que, aunque hubo compras de obras realizadas por artistas femeninas, estas fueron esporádicas y no reflejan una tendencia de inclusión generalizada en esos años.



● *Artistas con mayor representación*

Eugene Atget es el artista con mayor número de obras adquiridas por el MoMA. Este hecho subraya la importancia de su obra dentro de la colección del museo, especialmente en el campo de la fotografía. Atget, conocido por sus registros documentales de París, ha sido un referente clave para la fotografía moderna, lo que explica su prominencia en la colección.



CONCLUSIÓN

Este proyecto buscó analizar las adquisiciones del MoMA con técnicas de Machine Learning para identificar patrones y tendencias relacionadas con los artistas, nacionalidades, técnicas y evolución temporal de las obras.

Hallazgos clave

- Se observaron cambios en la representación por nacionalidad a lo largo del tiempo, con un crecimiento en obras extranjeras en décadas recientes.
- Las técnicas y materiales más frecuentes han variado significativamente desde 1950, mostrando una preferencia creciente por ciertos medios contemporáneos.
- A través del modelado, se comprobó que variables como popularidad del artista, antigüedad de la adquisición y clasificación del objeto tienen impacto en si una obra es moderna o no.

Rendimiento de los modelos


- Se aplicaron modelos de regresión y clasificación. Los modelos con variables sintéticas mejoraron el rendimiento general.
- El mejor modelo fue un RandomForestClassifier para predecir si una obra fue adquirida después de 1980, con una accuracy sólida y buena generalización.
- También se entrenó un modelo utilizando variables sintéticas creadas a partir de los datos originales, como la antigüedad de adquisición (YearsSinceAcquired), la popularidad del artista (ArtistWorkCount), y una clasificación del tamaño estimado de la obra (SizeCategory). Este modelo logró una performance sólida al predecir si una obra era moderna (posterior a 1980), demostrando que el enriquecimiento de variables puede mejorar significativamente la capacidad predictiva del modelo.
- Se observó que los modelos simples tendían a underfitting, mientras que otros con mayor complejidad o sin tuning presentaban overfitting. El uso de GridSearchCV ayudó a mitigar este problema.

Reflexión técnica: bias-variance tradeoff

Se identificó la importancia de encontrar un balance entre modelos complejos que ajustan bien al set de entrenamiento, y la capacidad de generalizar. En algunos casos, la diferencia entre train y test fue alta, lo cual evidenció overfitting, mientras que otros modelos con bajo rendimiento general mostraron underfitting. Los modelos optimizados mediante GridSearchCV ofrecieron un mejor balance entre bias y varianza.

Aprendizaje y próximos pasos

- Incluir más variables como el rol del curador o evento de adquisición.
- Explorar modelos basados en texto, por ejemplo utilizando embeddings para el campo Title o Medium.
- Evaluar la interpretabilidad de los modelos con herramientas como SHAP o feature_importance_.

 El uso de Machine Learning permitió no solo modelar tendencias del MoMA, sino también entender mejor los factores que influyen en sus decisiones curatoriales.