A DATA-DRIVEN ANALYSIS OF POPULARITY IN AI-GENERATED ART

Fatmanur Bakar 33811

DSA210 Final Presentation

TABLE OF CONTENTS

INTRODUCTION

The rapid rise of AI-generated art has transformed creative production by lowering barriers to creation and enabling large-scale artistic experimentation. While artistic and ethical debates around AI creativity are widespread, quantitative evidence on how audiences engage with AI-generated artworks remains limited.

This project aims to **investigate whether the popularity of AI-generated artworks can be explained or predicted using metadata alone,** such as art style, generation tool, platform, region, and creation time. Understanding the limits of such predictability is important both for cultural analysis and for evaluating the applicability of data-driven methods in creative domains.

MOTIVATION

- To test whether metadata contains meaningful predictive signal for popularity in creative domains
- To evaluate the limits of machine learning when applied to cultural and artistic data

DATA SOURCES

This project uses publicly available datasets from Kaggle, focusing on AI-generated art trends and engagement metrics.

Datasets:

- **AI Generated Art Trends** (https://www.kaggle.com/datasets/waqi786/ai-generated-art-trends)
- **AI Generated Art Popularity and Market Trends** (https://www.kaggle.com/datasets/atharvasoundankar/ai-generated-art-popularity-and-market-trends)

kaggle

DATASET OVERVIEW

Used variables in dataset 1:

- Popularity_Score
- Art_Style
- Art_Genre
- Tools_Used
- Platform
- Region
- Medium
- Creation_Date

Used variables in dataset 2:

- Engagement_Score
- Views
- Likes
- Comments
- Shares

Key Difference Between Datasets

Dataset 1: Descriptive metadata + engineered temporal and rarity features

Dataset 2: Direct user interaction metrics only

DATA PREPARATION AND FEATURE ENGINEERING

Data preparation included:

- removal of identifier-like columns (e.g., artwork IDs, artist names),
- conversion of creation dates into numerical time features (year, month),
- handling missing values via imputation,
- categorical variable encoding using one-hot encoding.

Feature engineering was restricted to metadata-only transformations, including:

- temporal indicators (recency, seasonal encodings),
- simple numerical transformations derived from existing columns.

No visual or image-based features were used in order to evaluate the explanatory power of metadata alone.

EXPLORATORY DATA ANALYSIS

- **Descriptive Statistics:** Calculated summary statistics for popularity metrics such as Popularity_Score, Likes, and Engagement_Score.
- **Category Comparison:** Boxplots and bar charts were used to compare popularity across categorical variables such as art style, creator type, tool, and region.
- **Correlation Analysis:** Numerical correlations were explored through a heatmap after applying one-hot encoding to categorical variables.

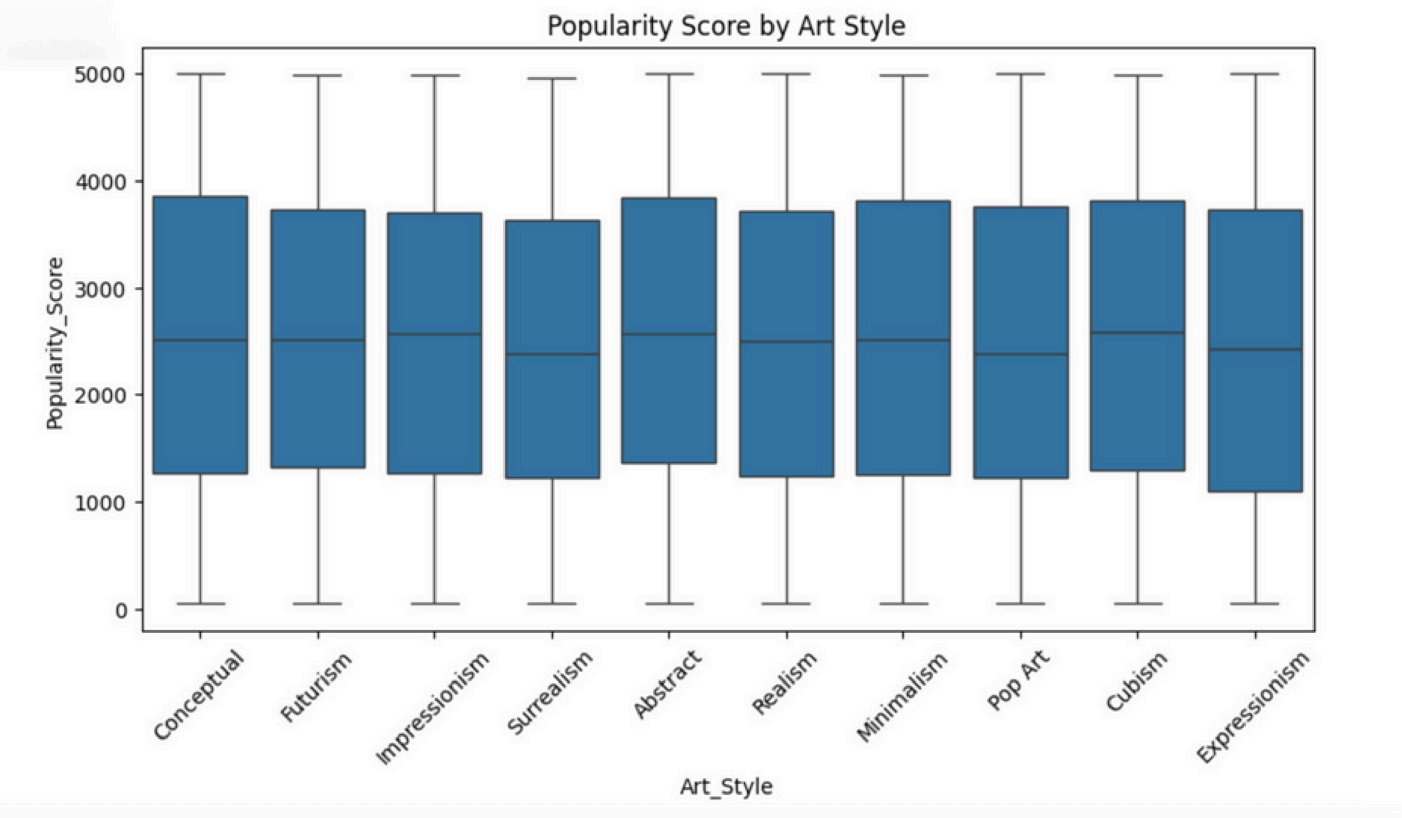
1) Descriptive Statistics:

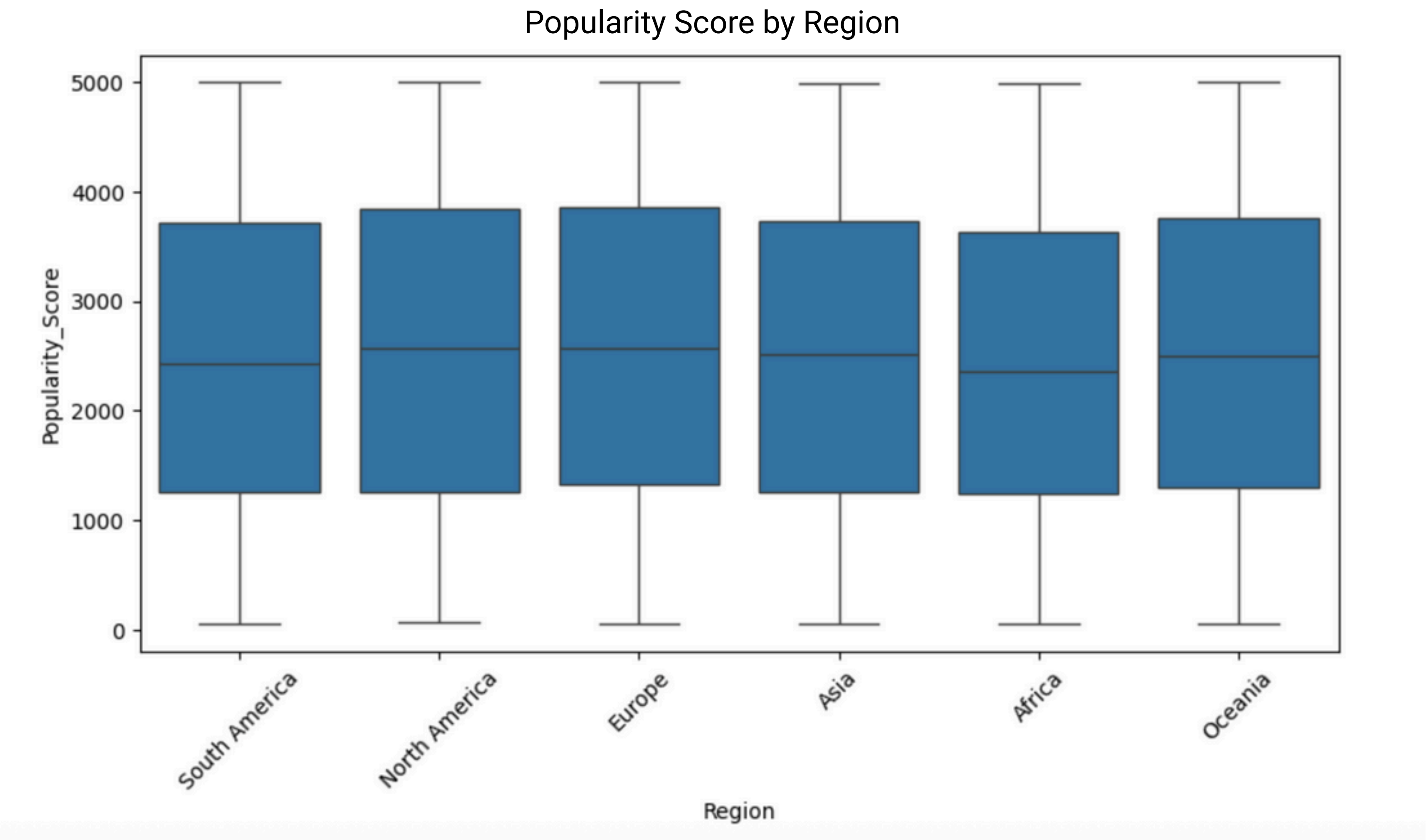
Descriptive statistics used to evaluate general statitics.

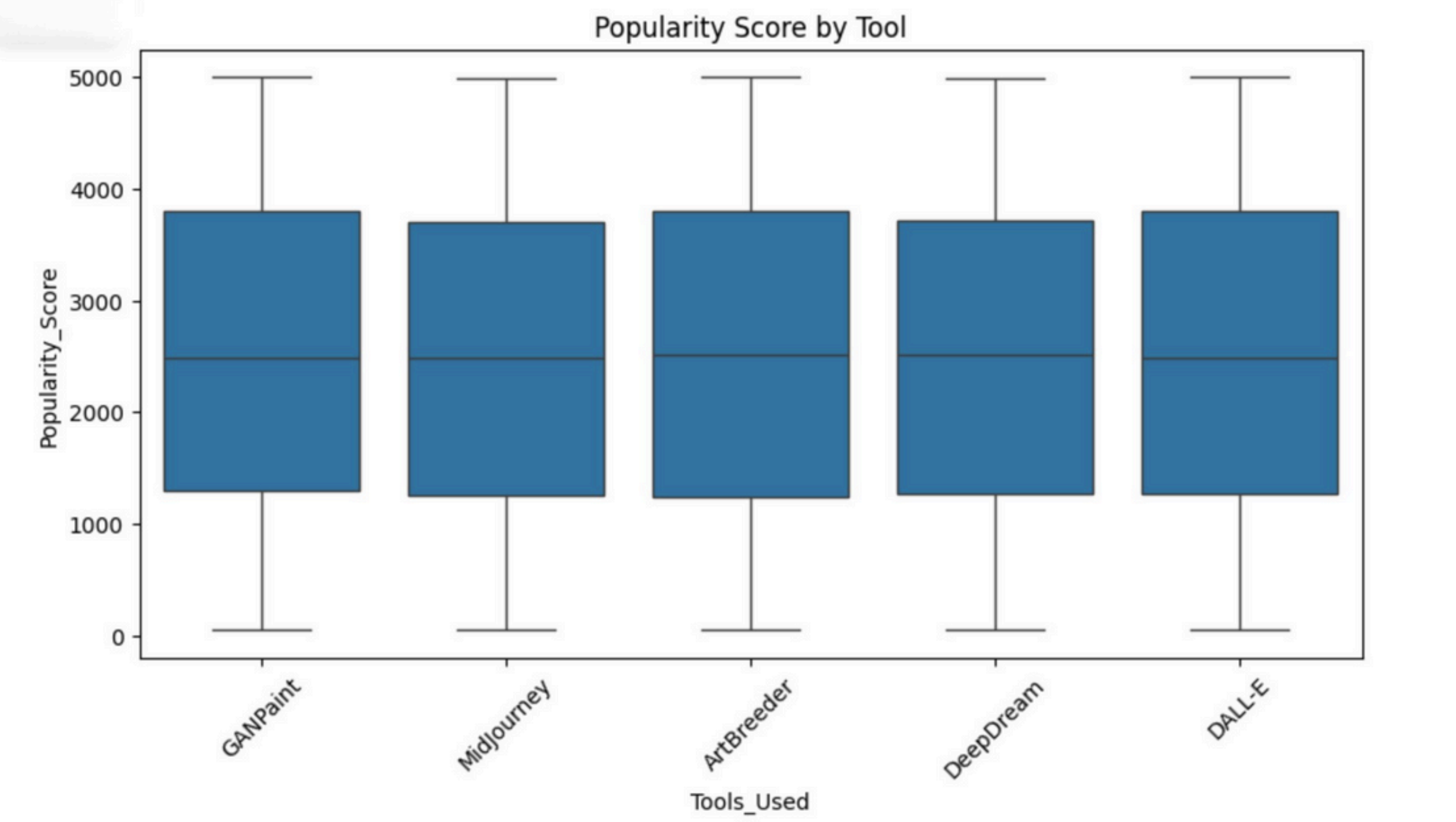
```
Most common style: Cubism
Most common tool: DeepDream
Most common region: South America
Average popularity score: 2508.1907380000002
Highest popularity score: 4999.62
```

2) Category Comparison:

2.1. Boxplots

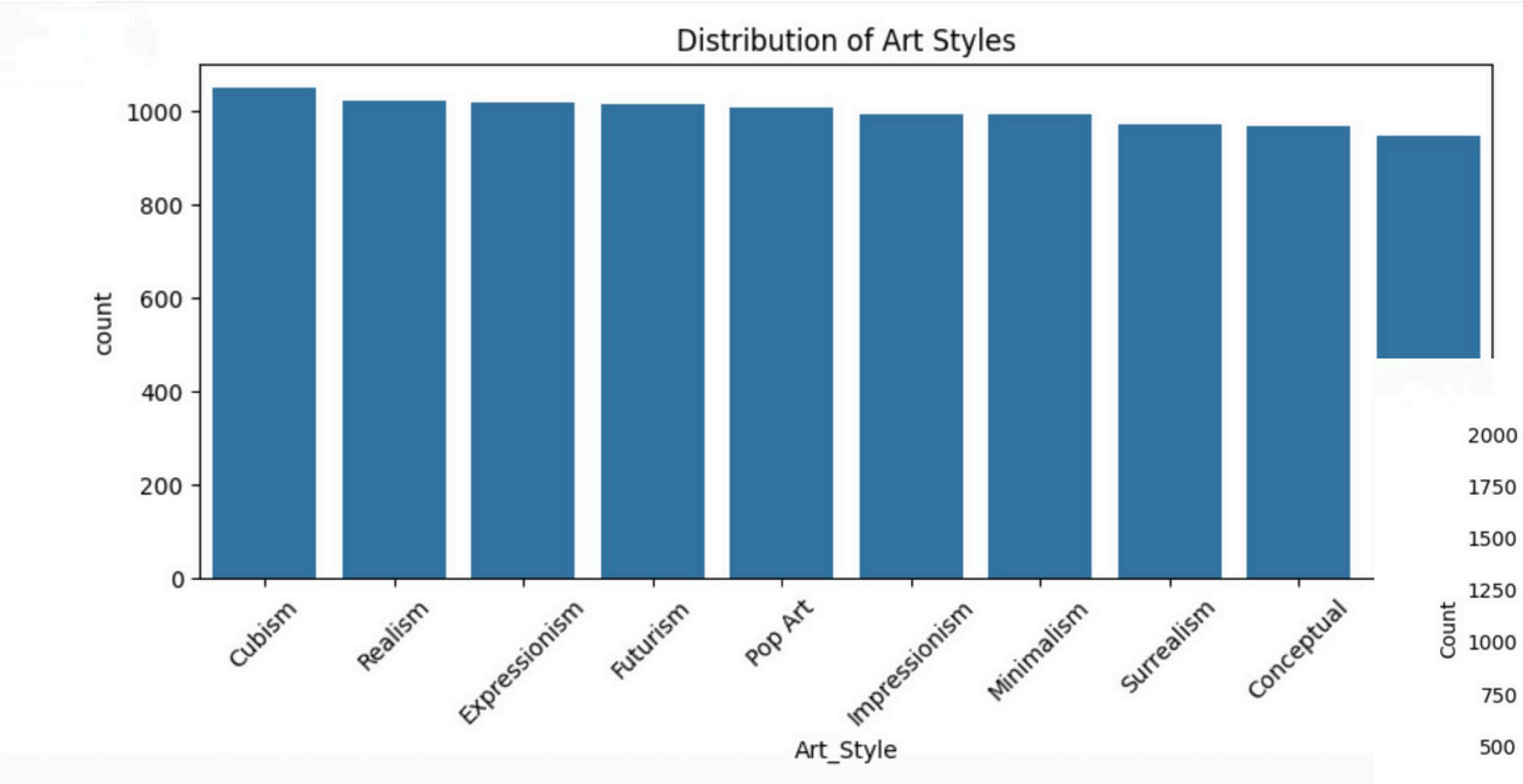


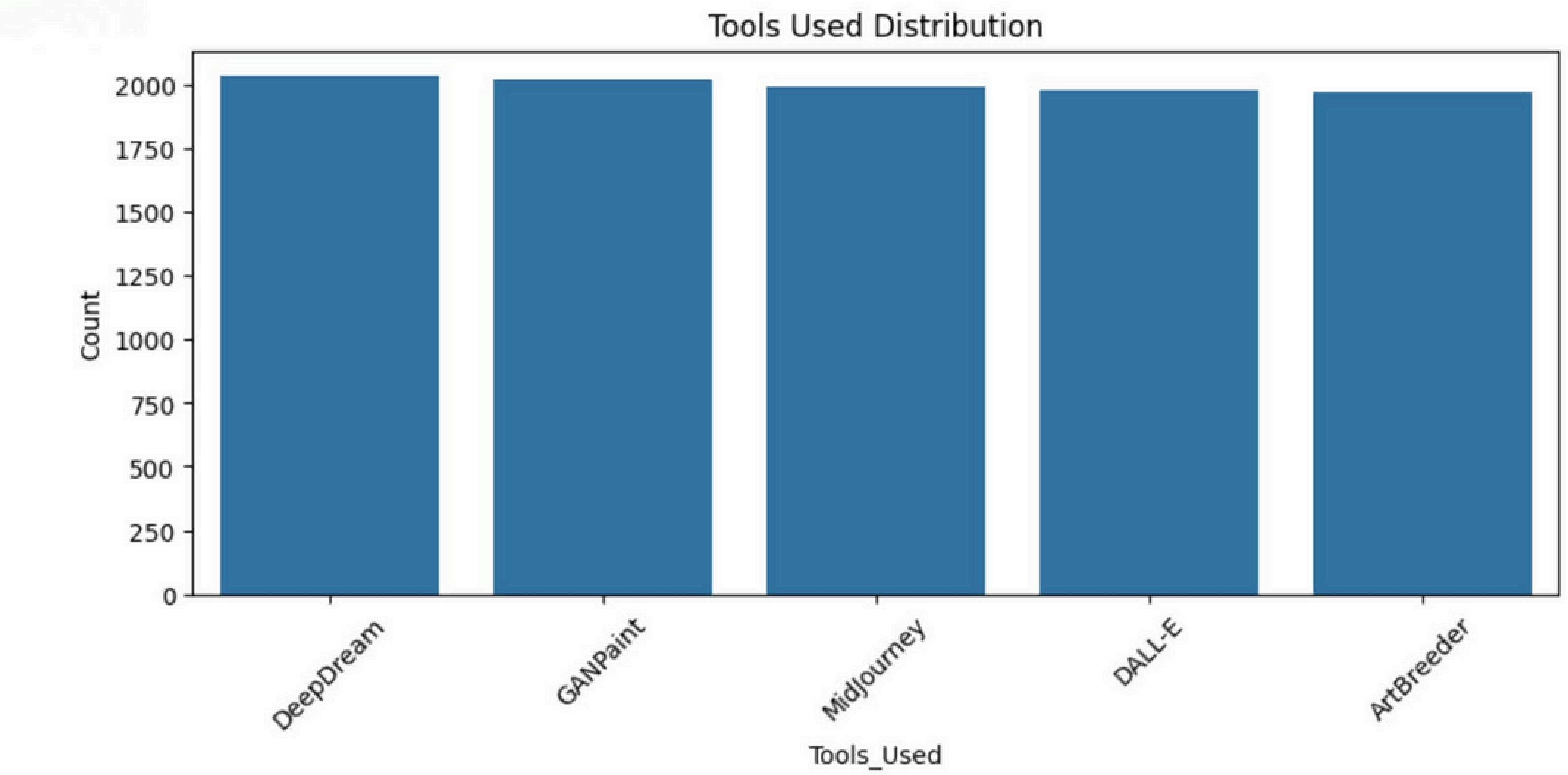




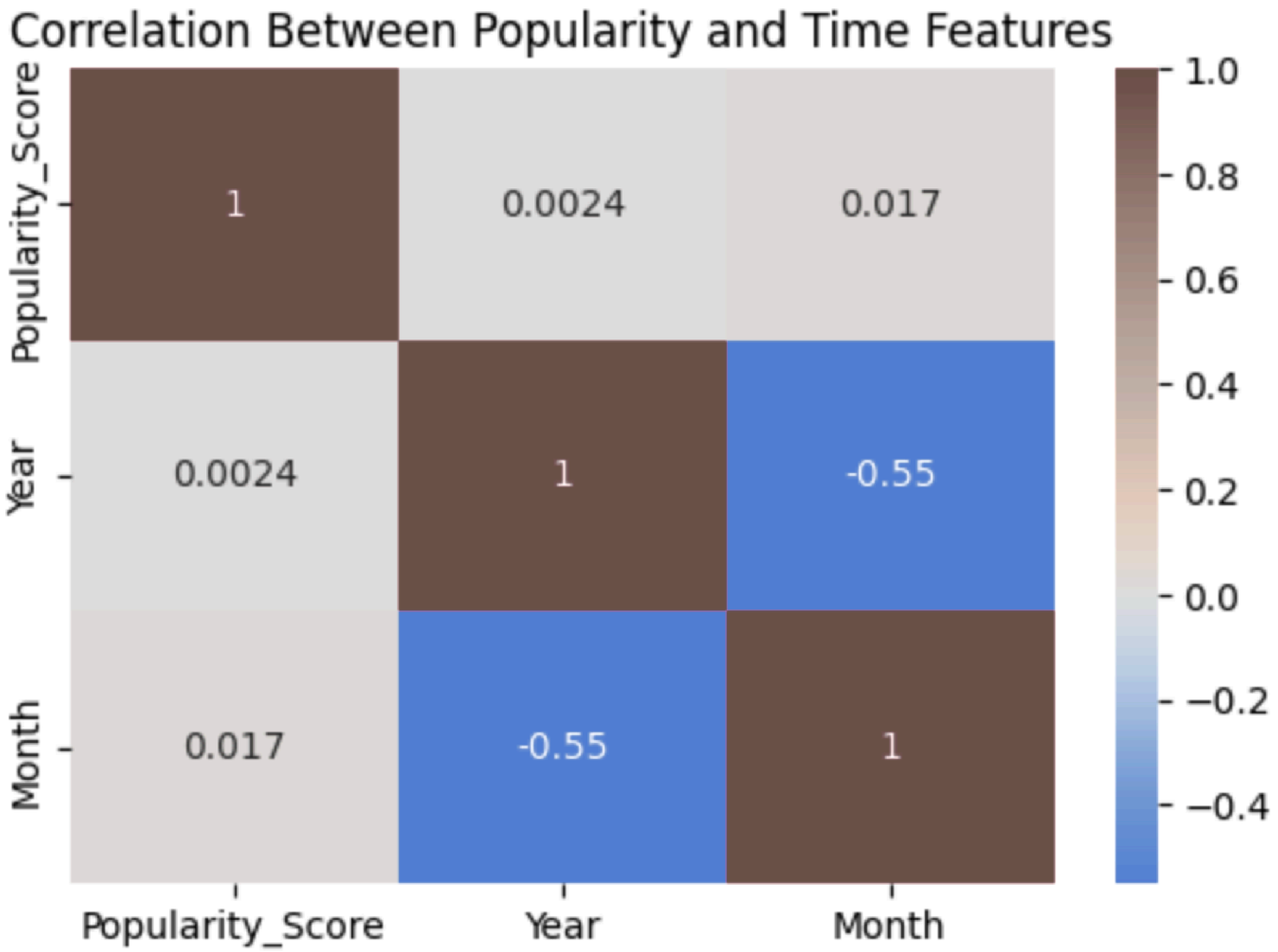
2) Category Comparison:

2.2. Bar charts





2) Correlation Analysis:



The correlation analysis indicates that popularity has almost no linear relationship with either year or month.
This suggests that AI-generated artwork popularity does not follow a clear temporal or seasonal trend.
The negative correlation between year and month reflects the data distribution rather than an effect on popularity.

HYPOTHESIS TESTING

The following hypotheses were tested using statistical methods and exploratory analysis:

- **H1:** Popularity has significantly increased after 2022.

- **H2:** Certain tools correlate with higher average engagement.

- **H3:** Human-edited (Hybrid) artworks perform better than fully AI-generated ones.

- **H4:** Art style significantly affects popularity.

H1: Popularity has significantly increased after 2022.

Result: A **Welch's t-test** was conducted to compare popularity scores before and after 2022. Although the mean popularity slightly increased in recent years, the difference was not statistically significant ($p \geq 0.05$). Therefore, we fail to reject the null hypothesis. This indicates that the **dataset does not provide sufficient evidence that AI-generated artwork popularity has significantly increased after 2022.**

H2: Certain tools correlate with higher average engagement.

A **one-way ANOVA test** was used to examine whether different AI tools (MidJourney, GANPaint, ArtBreeder, DeepDream) produce significantly different popularity outcomes. The results show no statistically significant difference among tools ($p \geq 0.05$). Thus, the null hypothesis cannot be rejected. This suggests that in this dataset, **tool choice alone does not strongly determine engagement or popularity of AI-generated artworks.**

H3: Human-edited (Hybrid) artworks perform better than fully AI-generated ones.

An **ANOVA test** was conducted to examine whether the creator type (Hybrid, AI Model, Individual) has a significant effect on Likes. The test returned a non-significant result ($p = 0.69$), meaning that we fail to reject the null hypothesis. This indicates that the popularity of AI-generated artworks in this dataset cannot be explained by creator type alone, and engagement levels are likely influenced by other factors. **Although exploratory plots show minor differences in mean likes, these differences are not statistically significant.**

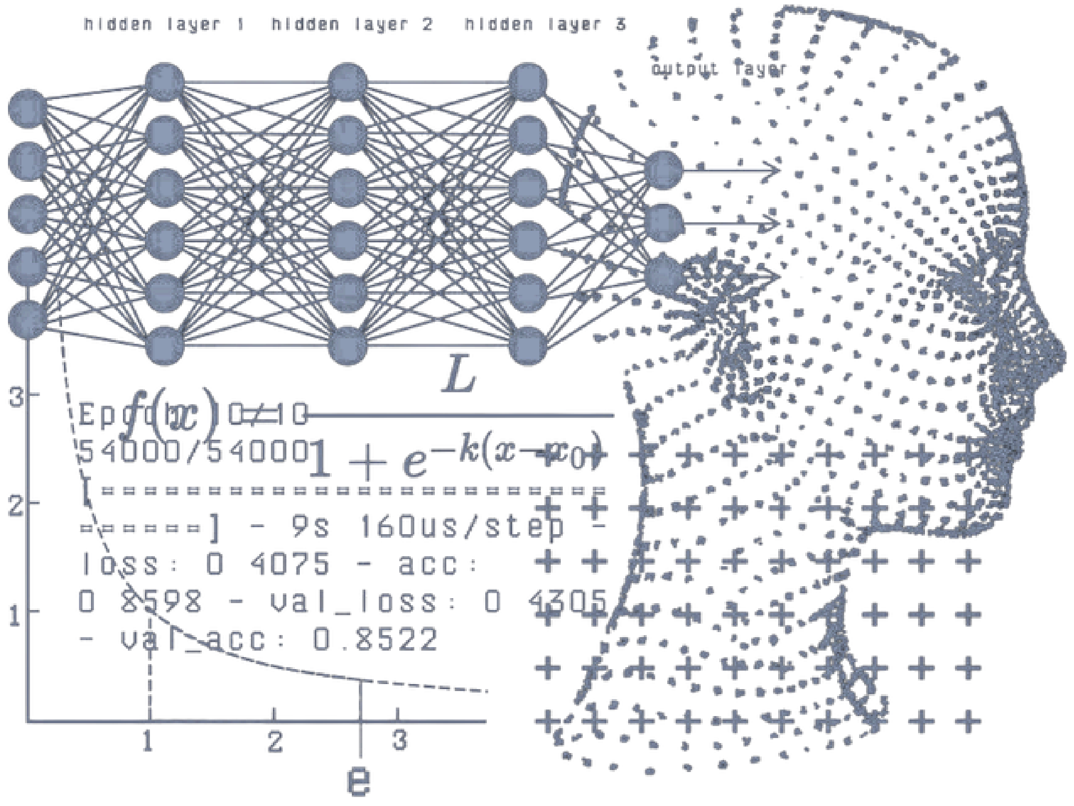
H4: Art style significantly affects popularity.

An **ANOVA test** was conducted to examine, and the results based on the available dataset show that, there is no statistically significant evidence that artworks of different styles receive different popularity scores. Although minor variations in mean popularity exist across styles in the exploratory plots, **these differences are not strong enough to be statistically confirmed.**

EDA & HYPOTHESIS TESTING CONCLUSION

While statistical significance was not reached, exploratory plots still show differences in mean popularity across tools and styles. These patterns may require a larger dataset or additional variables to detect statistically confirmed effects.

MACHINE LEARNING APPLICATIONS



1) Regression Task: Predicting Popularity Score

At first, I formulate the problem as a regression task, where the goal is to predict the numerical popularity score of an artwork from metadata features, and to test whether a linear relationship exists between metadata and popularity

From Regression to Classification

Predicting the exact popularity score is hard and noisy, so I then reframe the task as a binary classification problem, and asked **"Can metadata distinguish high-performing vs low-performing artworks"** and proceed to:

2) Classification Models: High vs. Low Popularity

3) Additional Machine Learning Experiment

1) REGRESSION TASK: PREDICTING POPULARITY SCORE

Objective: Predict the numerical popularity score from metadata features.

Models used:

- **Ridge Regression (linear baseline):** The regression model yields a negative R^2 value, indicating that it performs worse than simply predicting the mean popularity score for all artworks. This result suggests that: Linear relationships between metadata and popularity are weak or nonexistent, and popularity is likely influenced by complex, non-linear, or unobserved factors.
- **Random Forest Regressor (non-linear ensemble):** Random Forest did not improve performance because the R^2 score remains negative, and prediction error stays high. Since an ensemble model still fails, the issue is likely insufficient information in metadata, not model choice.

Evaluation metrics:

- **Root Mean Squared Error (RMSE), R^2 score**

2) CLASSIFICATION MODELS: HIGH VS. LOW POPULARITY

Objective: Reformulate the task as binary classification to distinguish between high- and low-popularity artworks (based on median popularity).

Models used:

- **Logistic Regression**
- **Random Forest Classifier**

Both Logistic Regression and Random Forest achieved ROC-AUC values close to 0.50, which is near random guessing. This indicates that even after simplifying the target into two classes, the available metadata does not provide enough signal to separate high vs low popularity.

Evaluation metrics:

- **Accuracy, ROC-AUC, Confusion Matrix, Precision, Recall, F1-score**

MACHINE LEARNING CONCLUSION FOR DATASET 1

Across regression and classification experiments, machine learning models consistently **fail to predict** or distinguish artwork popularity using metadata alone.

Unlike the previous dataset, **dataset 2 contains direct user interaction metrics such as views, likes, comments, and shares, and allows testing whether direct engagement metrics** contain predictive signal beyond descriptive metadata Therefore, I then evaluated whether these features can predict the overall engagement score.

3) ADDITIONAL MACHINE LEARNING EXPERIMENT

Models used:

3.1. Regression

- **Ridge Regression:** Since the engagement score is already a normalized or aggregated function of interaction metrics (such as views, likes, comments, and shares), attempting to predict it using these same variables resembles reverse engineering and limits the learnable signal.
- **Dummy Regressor:** To assess whether regression models learn meaningful patterns, a Dummy Regressor is used as a baseline that predicts the mean engagement score. Model performance is compared against this baseline. Regression models do not outperform a simple dummy baseline, indicating that the available features do not provide meaningful predictive information for engagement score estimation.

3.2. Classification

- **Logistic Regression**
- **Random Forest Classifier**

Both Logistic Regression and Random Forest achieve accuracy and ROC-AUC values close to random guessing, suggesting that high and low engagement levels cannot be reliably distinguished using the current feature set.

FINDINGS AND INSIGHTS

Across all stages—EDA, hypothesis testing, regression, and classification—the **results consistently point to the same conclusion:**

AI-generated artwork popularity cannot be reliably explained or predicted using metadata alone.

Key insights:

- Neither linear nor non-linear models capture meaningful predictive signals.
- Popularity appears to be driven by latent factors not present in the dataset, such as:
  - visual and aesthetic qualities of the artwork,
  - platform recommendation algorithms,
  - viral diffusion and social exposure,
  - external reputation effects.

Machine learning played a crucial role in **validating the limits of predictability**, rather than merely optimizing performance.

LIMITATIONS

- The analysis relies exclusively on metadata and does not incorporate image content.
- Popularity measures may be affected by platform-specific dynamics and external exposure mechanisms.
- Causal relationships cannot be inferred from the available data.

FUTURE WORK

Future extensions could:

- Incorporate image-based features using convolutional neural networks (CNNs),
- Combine visual and metadata features in multimodal models,
- Include platform-level exposure metrics (e.g., impressions, follower counts),
- Apply causal inference methods to better understand drivers of popularity.

CONCLUSION

This project demonstrates that null or negative predictive results can be meaningful outcomes in data science. By systematically applying statistical analysis and machine learning techniques, we show that metadata alone is insufficient to explain or predict popularity in AI-generated art.

Rather than indicating failure, these results highlight the complexity of cultural and creative domains and emphasize the importance of data scope, feature selection, and problem formulation when applying machine learning to real-world phenomena.

THANK YOU
FOR YOUR ATTENTION