



GROUP 12

# WHAT DRIVES ANIME RATINGS?

Shanaya Mehra, Sindhuja Kumar, Songying Lyu, Gabriella Pranadi

# AGENDA

**01** OBJECTIVE

**02** DATA  
DESCRIPTION

**03** DATA  
CLEANING

**04** DATA  
DISTRIBUTION

**05** CONFOUNDING  
VARIABLES

**06** CAUSAL  
ANALYSIS

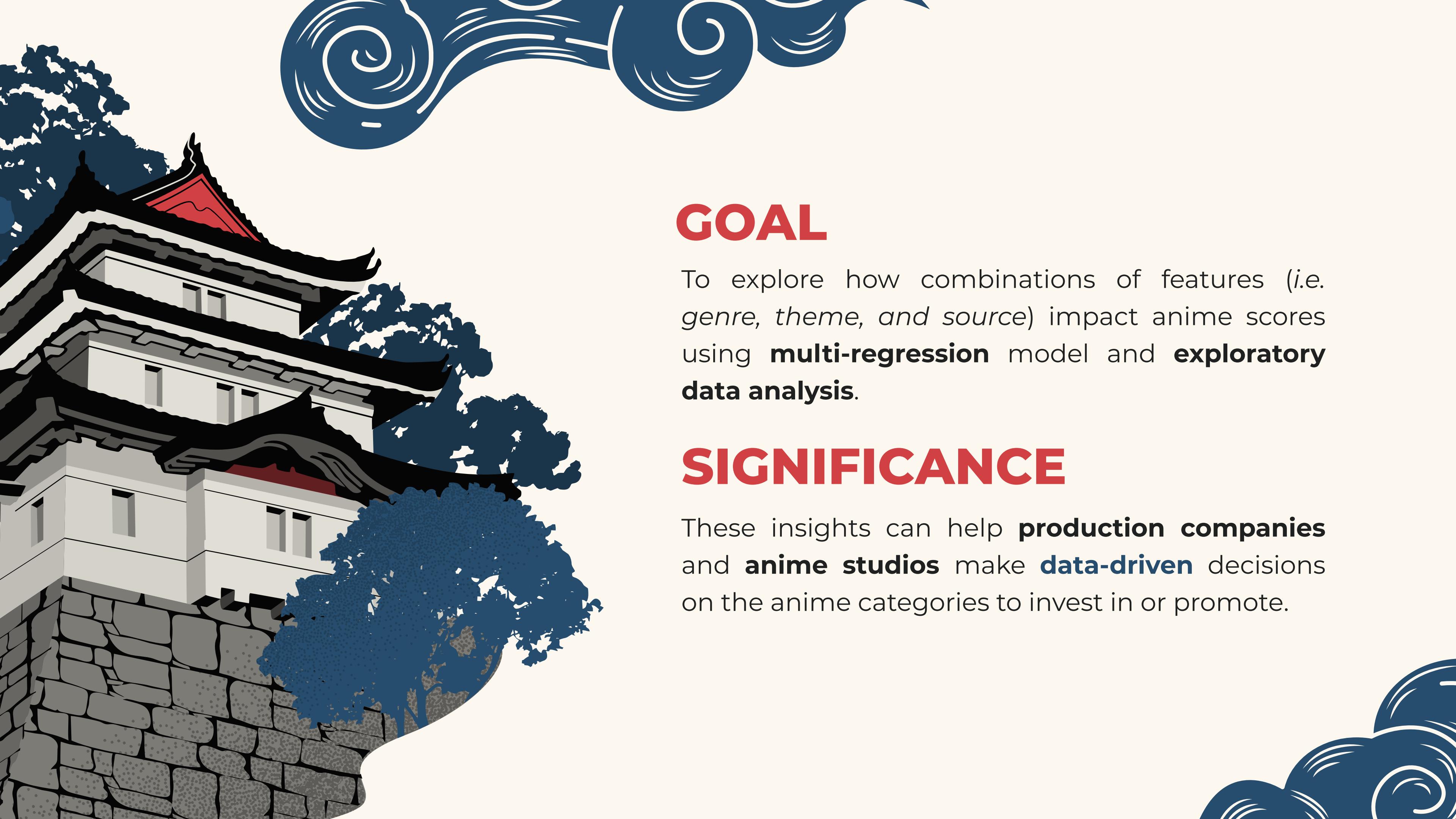
**07** REGRESSION  
MODEL

**08** CONCLUSION

01

# OBJECTIVE





## GOAL

To explore how combinations of features (i.e. genre, theme, and source) impact anime scores using **multi-regression** model and **exploratory data analysis**.

## SIGNIFICANCE

These insights can help **production companies** and **anime studios** make **data-driven** decisions on the anime categories to invest in or promote.

# 02

# DATA DESCRIPTION



# DATA OVERVIEW...

**15k**

*total anime entries*

**24**

*descriptive variables*



**2**

*URLs*

**14**

*categorical*

**8**

*numerical*

# 03

# DATA

# CLEANING



# FEATURE ENGINEERING...

**15k** → **~12k**  
*total anime entries*      *anime entries*

**41**  
*descriptive  
variables*

→

**38**  
*dummy variables  
created*

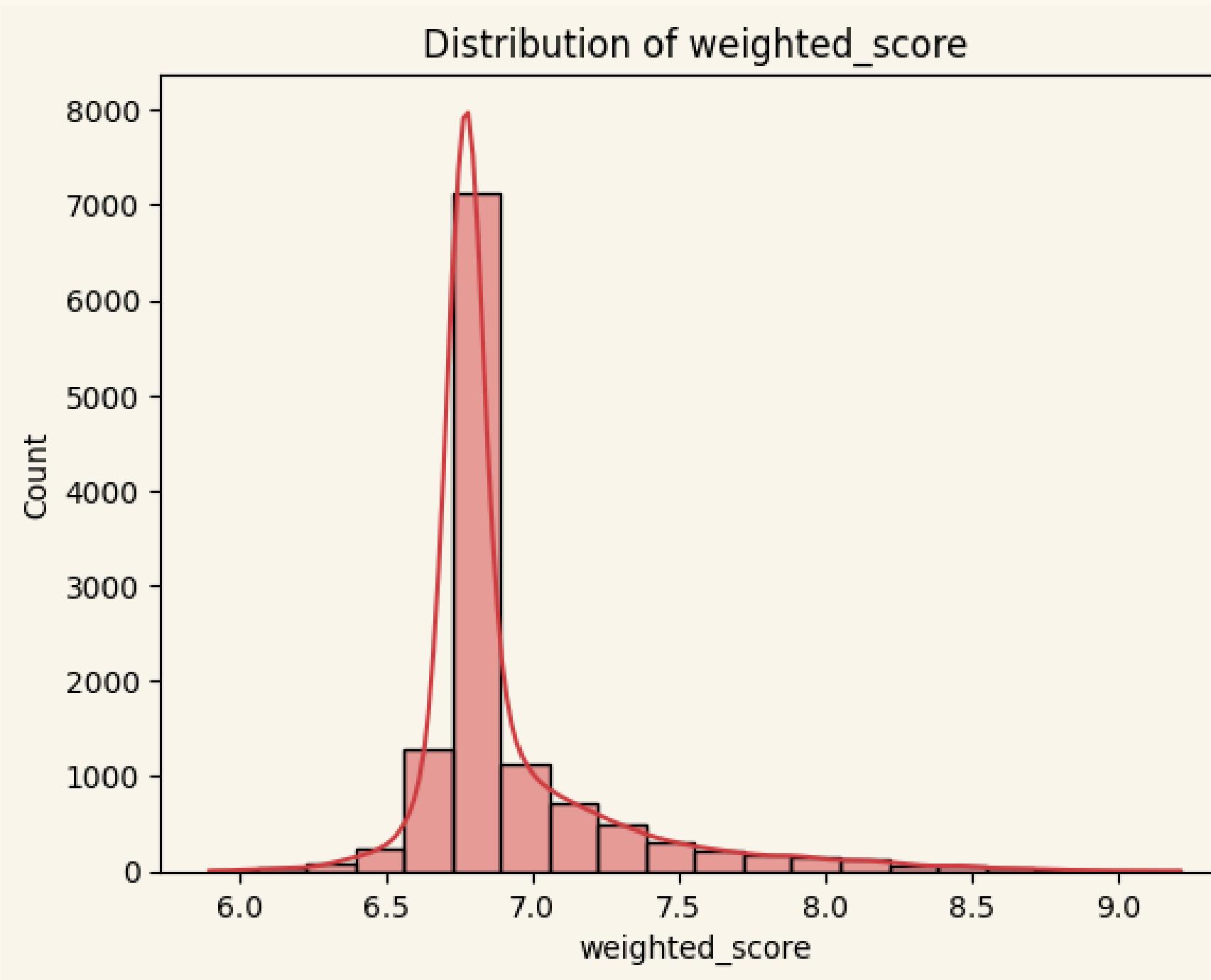
**3**  
*numerical  
variables*

# 04

# DATA DISTRIBUTION



# DEPENDENT VARIABLE...



**SHAPE**  
Slightly right-skewed

**MODALITY**  
Unimodal

**SUMMARY STATISTICS**

**Mean:** 6.91

**Median:** 6.78

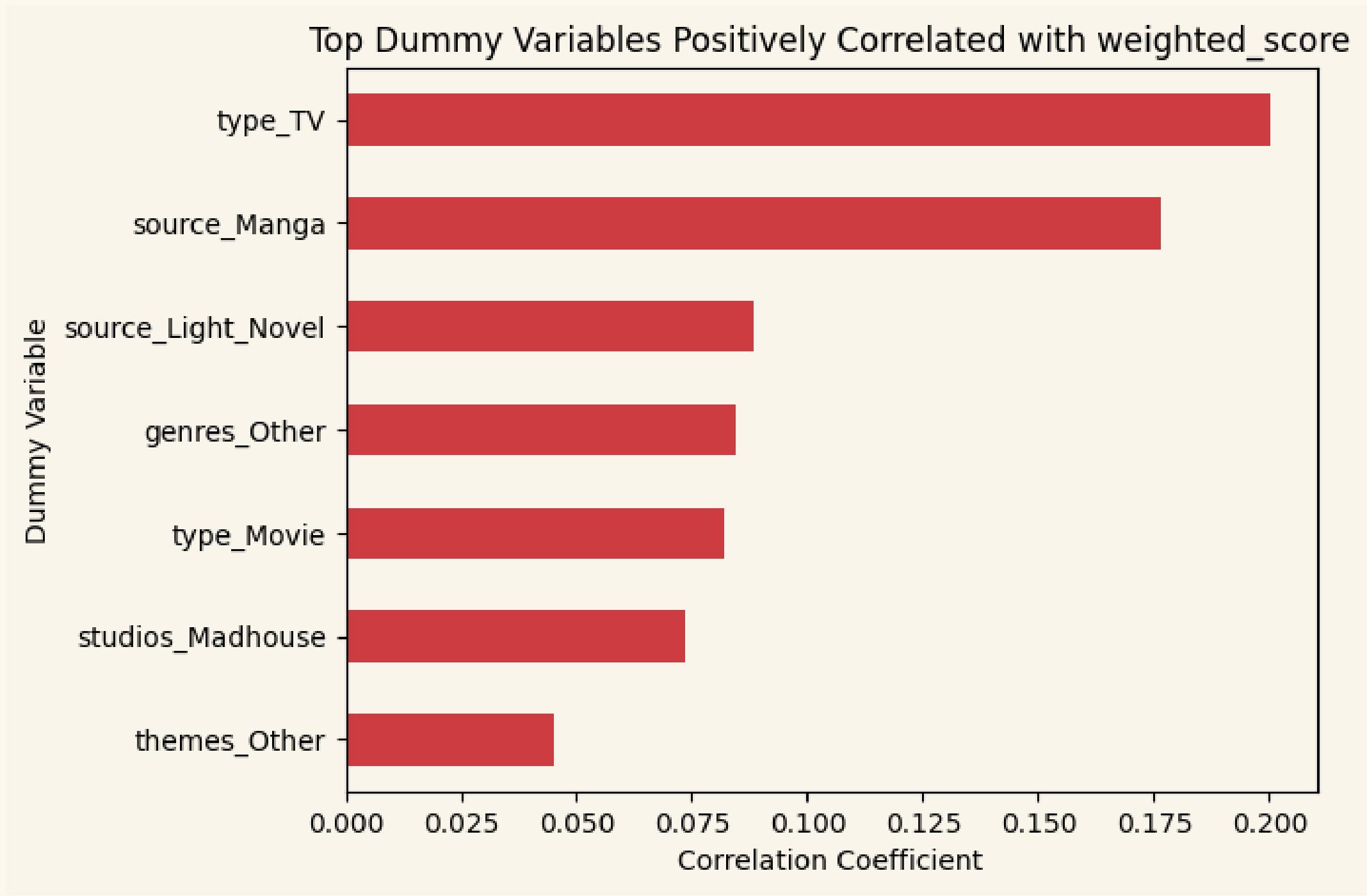
**Range:** 5.89 to 9.21

# NUMERICAL INDEPENDENT VARIABLES...

	weighted_score	episodes	duration_minutes
weighted_score	1.000000	<b>0.047046</b>	<b>0.172408</b>
episodes	0.047046	1.000000	-0.106558
duration_minutes	0.172408	-0.106558	1.000000

Both variables show **weak correlation** with weighted\_score, though duration\_minutes is **slightly more positively associated** than episodes

# CATEGORICAL INDEPENDENT VARIABLES...



type\_TV, source\_Manga, and source\_Light\_Novel, have the **highest positive correlations** with weighted\_score

# 05

# CONFOUNDING VARIABLES



# IDENTIFYING CONFOUNDING VARIABLES...

31

*confounding  
variables*



## PROCESS

Ran regression analysis to test whether the variable is correlated with both independent and dependent variables

## ADJUSTMENT

Added into the regression model to keep them controlled.  
Eliminate omitted variable bias to model the true association between predictors and outcome variable.



# 06

# CAUSAL ANALYSIS



# COMPARING MODELS...



	Univariate Regression Model Between Weighted Score and Anime TV Type	Multivariate Regression Model Controlling for <u>All</u> Confounding Variables	Multivariate Regression Model Controlling for <u>A</u> <u>Few</u> Confounding Variables
Beta Coefficient of the Anime TV Type	0.1473	0.1901	<b>0.1486</b>
Standard Error	0.007	0.016	<b>0.008</b>
Sample Size (Number of Observations)	12232	12232	<b>12232</b>
Number of Confounding Variables	0	31	<b>13</b>
Adjusted R-Squared	0.040	0.149	<b>0.139</b>



# 07

# REGRESSION

# MODEL



# MULTIVARIATE REGRESSION...

**Weighted  
Score**

$$= 5.8 + 1.5(\text{TV}) + 1.1(\text{Manga}) + 0.1(\text{Shounen}) + 1.0(\text{Light Novel}) + \\ 0.1(\text{Madhouse}) + 0.9(\text{Game}) - 0.04(\text{OVA}) + 1.0(\text{Source_Other}) + 0.02(\text{ONA}) \\ + 0.9(\text{Original}) + 1.0(\text{Visual Novel}) + 0.01(\text{genres_Other}) + 0.003(\text{duration})$$

**41**

descriptive  
variables

**38**

dummy variables

**3**

numerical

Adjusted R-Squared: 0.139

# 08

# CONCLUSION



# KEY TAKEAWAYS...



**Model Performance:** The model explained approximately **14%** of the variance in anime ratings, suggesting moderate explanatory power, but other unaccounted factors may contribute to ratings. While the analysis reveals correlations, it cannot establish causality!

**Significant Factors:** **TV-type**, **Manga-based**, and **Light Novel-based** anime consistently perform better in ratings. Anime targeted at the **Shounen** demographic and produced by **Madhouse** also significantly influence ratings.

**Confounding Variables:** The model successfully controlled for confounders like certain demographics and studio, which **eliminated omitted variable bias**.





**THANK  
YOU**

