

# ART OF DATA

*Predicting Final Sale Prices of Fine Art*

# THE TEAM

JEREMIAH TORIBIO



COLTON KANE



# AGENDA

- Executive Summary
- Acquisition
- Preparation
- Findings
- Models
- Conclusion

# EXECUTIVE SUMMARY

*Big Idea*

Deep learning & regression model to  
predict art price

Zhang Daqian, Verdant Mountains of Summer Spirit, 1968

# EXECUTIVE SUMMARY

## *Big Idea*

Deep learning & regression model to predict art price

## *Findings*

- Drivers: *artist, year created, canvas*
- Some artist more predictable than others

# EXECUTIVE SUMMARY

## *Big Idea*

Deep learning & regression model to predict art price

## *Goals*

Predict price of art based on previous sales to benefit artists, vendors, and collectors

## *Findings*

- Drivers: *artist, year created, canvas*
- Some artist more predictable than others

# EXECUTIVE SUMMARY

## *Big Idea*

Deep learning & regression model to predict art price

## *Goals*

Predict price of art based on previous sales to benefit artists, vendors, and collectors

Zhang Daqian, Verdant Mountains of Summer Spirit, 1968

## *Findings*

- Drivers: *artist, year created, canvas*
- Some artist more predictable than others

## *Next Steps*

- Obtain more data from different artists
- Further develop deep learning model and test other neural networks

# AQUISITION

*Artnet.com*

- Vendor since 1989
- 1000 pieces of art
- Art industry privatized

# AQUISITION

*The data we chose:*

*Proof of concept*

Most recent sales

- 10 unique artists
- Various backgrounds
- Price range from \$10,000 to \$1,000,000

# ARTISTS

---



- Zhang Daqian
- Marc Chagall
- Andy Warhol
- Leonard Tsuguharu Foujita
- Banksy
- KAWS
- Yayoi Kusama
- Rembrandt van Rijn
- Pablo Picasso
- Salvador Dali



# AQUISITION

## *Raw Data:*

- 10 PDFs, 1 for each artist with 100 pieces each
- Joined 10 for 1 PDF of all 1000 pieces
- PyPDF2 (python) to extract text from PDFs

# PREPARATION

*What it looked like before:*

Banksy Queen Vicscreenprint in colorsHeight 70.7 x Width 49.7 cm.

.Height 27.835 x Width 19.567 in.2003Edition: 50013

December 202211 months agoForum AuctionsThe British Sale

(with Dreweatts 1759) - [Lot00122]est. 16,000 u 20,000 GBP

20,000GBPest. 19,799 u 24,749 USD24,749USD

# PREPARATION

*How we changed it:*

Banksy Queen Vic screenprint in colors Height 70.7 x Width 49.7 cm.

Height 27.835 x Width 19.567 in. 2003 Edition: 50013

December 2022 11 months ago Forum Auctions The British Sale

(with Dreweatts 1759) - [Lot 00122] est. 16,000 u 20,000 GBP

20,000 GBP est. 19,799 u 24,749 USD 24,749 USD

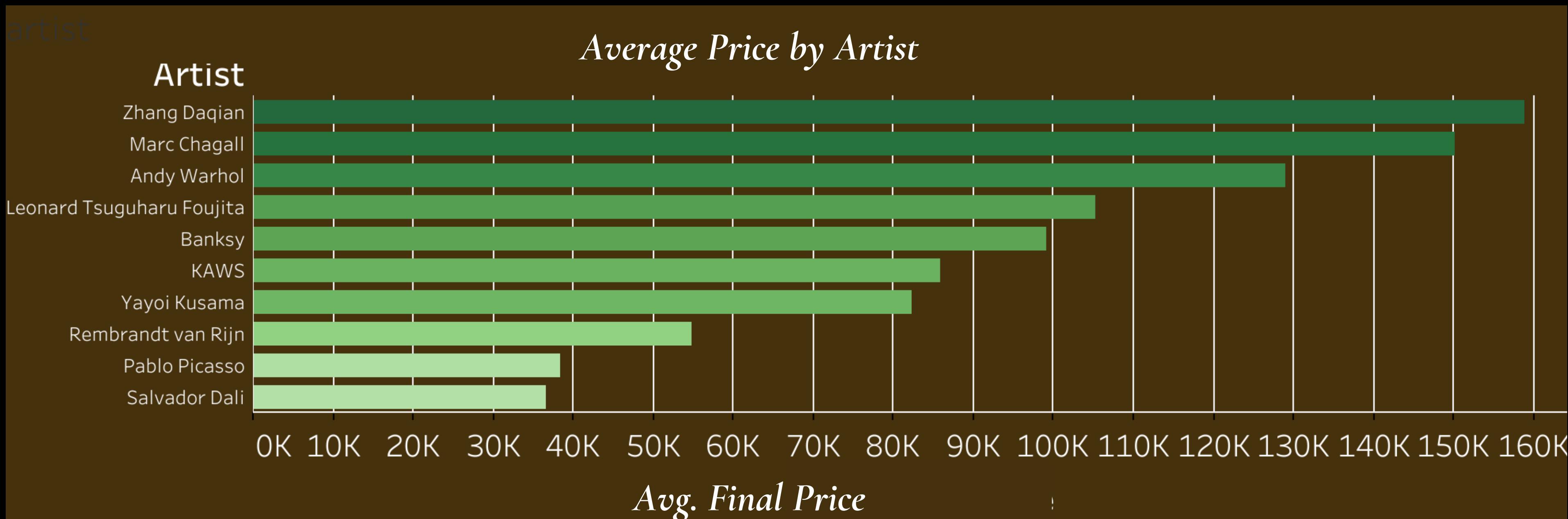
# PREPARATION

*What it looked like after:*

artist	dimension_in	year_created	date_sold	auction_house	estimate_usd_low	estimate_usd_high	hammer_price	percent_estimate	title_medium
banksy	Height 27.559 x Width 19.685 in	2003.0	December 2022	Tate Ward AuctionsUrban & Contemporary Art -	17071.0	21948.0	21338.0	0.0	Banksy Queen Victoria screenprint in colors on ...

# FINDINGS

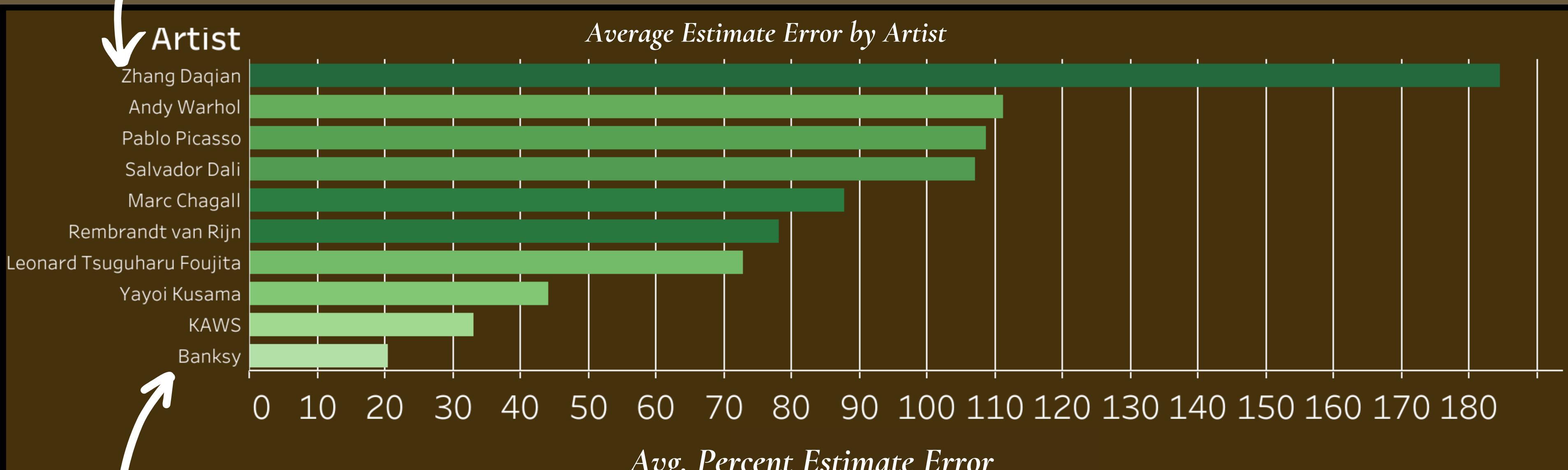
*Artist is the biggest driver of final price*



# FINDINGS

---

*Zhang Daqian's art's price has been historically LESS predictable*



*Banksy's has been MORE predictable*



# DAQIAN

---

Estimated \$642 - 1,028 USD

Sold for \$63k (667% error)

Estimated \$256k - 385k USD

Sold for \$938k (144% error)



Daqian, Listening to Pine Trees in Yunshan, 1935

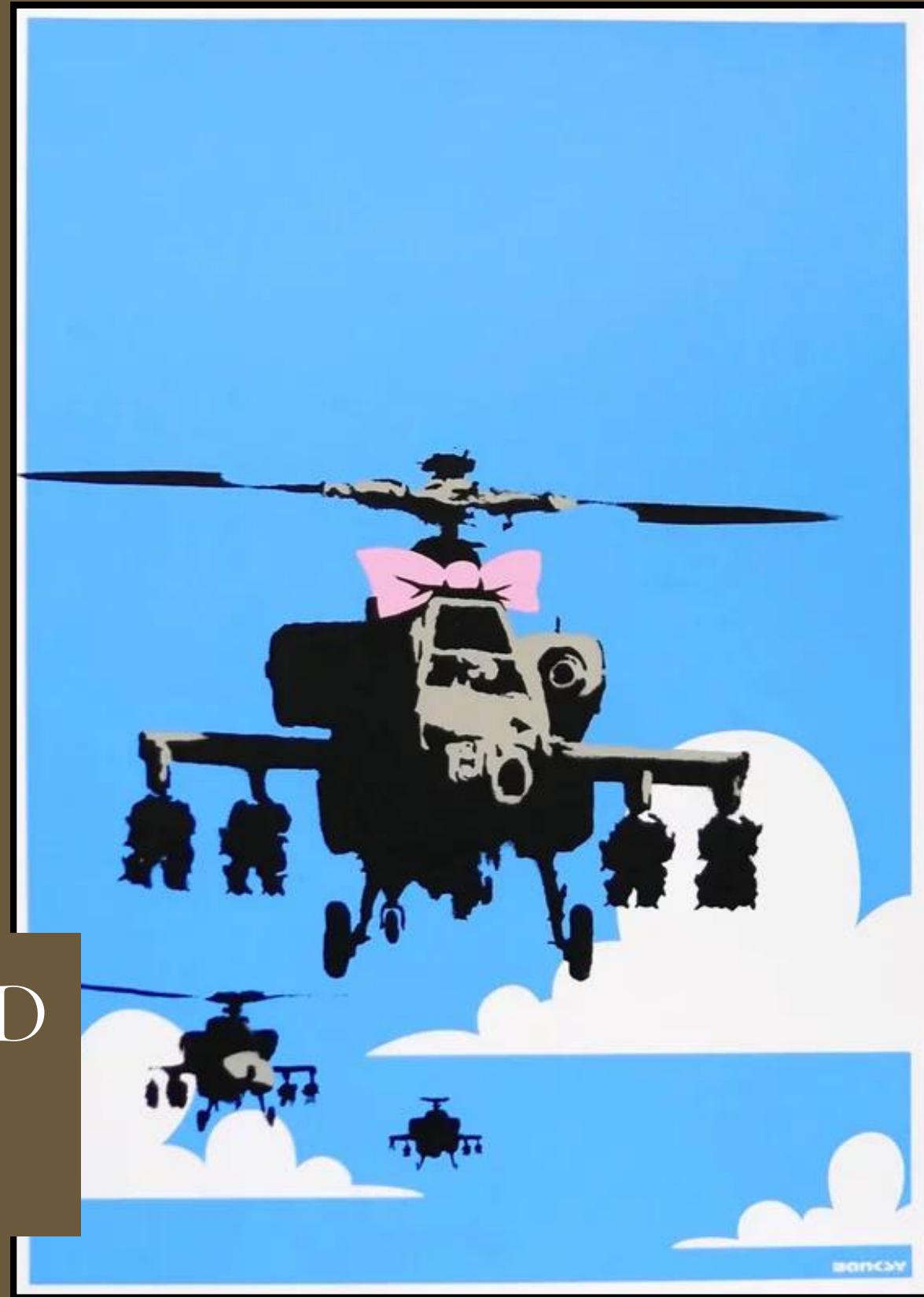
Daqian, Pavilions on Emerald Hills, 1967

# BANKSY

---

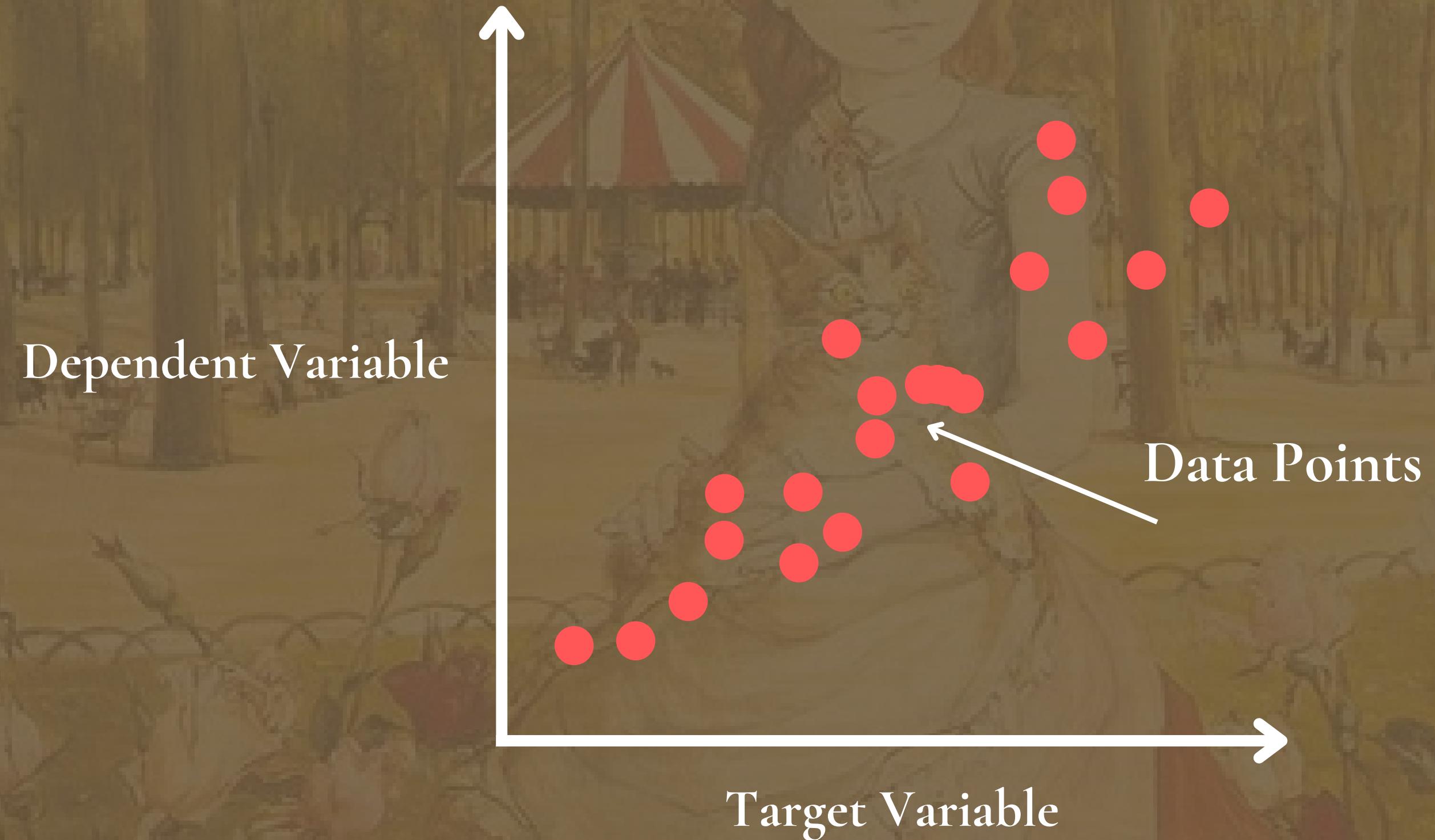


Estimated \$39k - 50k USD  
Sold for \$45k (no error)

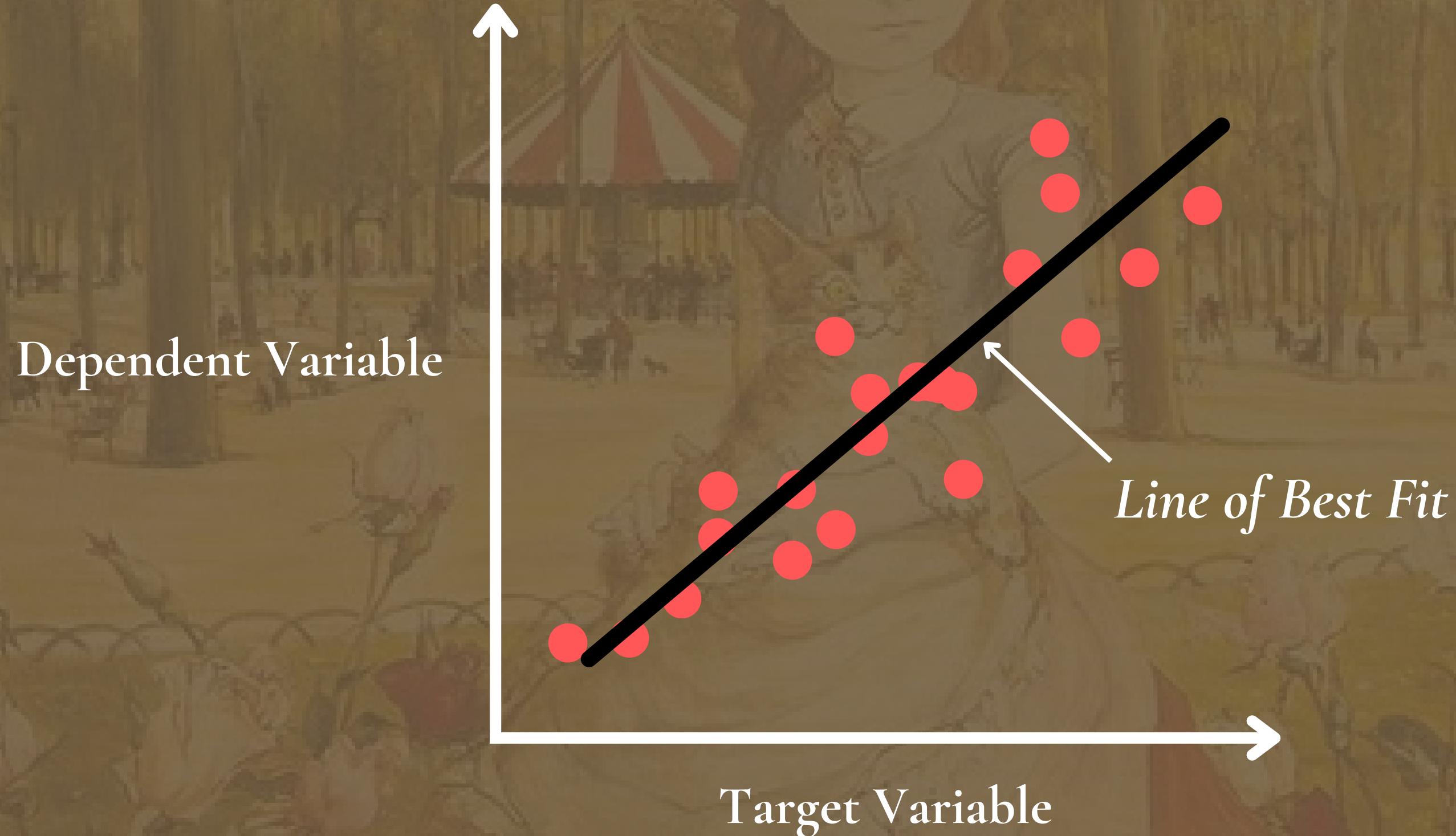


Estimated \$40k - 60k USD  
Sold for \$53k (no error)

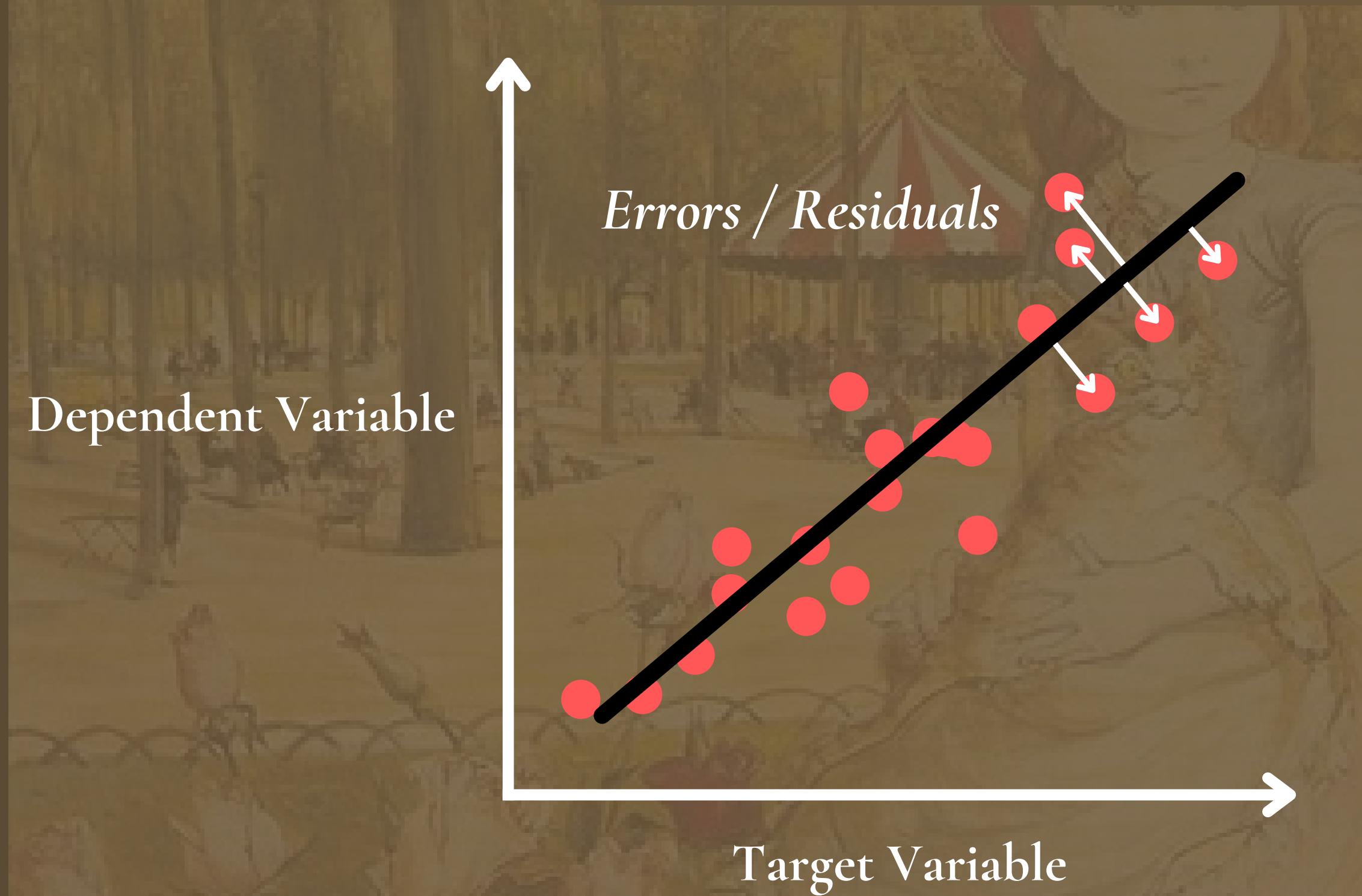
# REGRESSION



# REGRESSION



# REGRESSION



Root Mean Square Error  
(RMSE):

- Shorter distance to line is better

# MODELING

20,000

15,000

10,000

5,000

0

Model 1

Model 2

LassoLars

Regression models in (RMSE):

- *Model 1: 15700*
- *Model 2: 15200*

• *Lasso-Lars (Least Angle Regression):*

15125

# DEEP LEARNING MODEL

Deep Learning:

- *Computer Vision*: Images to arrays of numbers
- *TF-AutoKeras*
- *ResNet50*
- *Root Mean Square Error*: 170k

# BEST MODEL

Regression:

- *Lasso-Lars (Least Angle Regression)*
- *Root Mean Square Error:*  
\$15,125
- $R^2$ : .35

# CONCLUSION

## TAKEAWAYS:

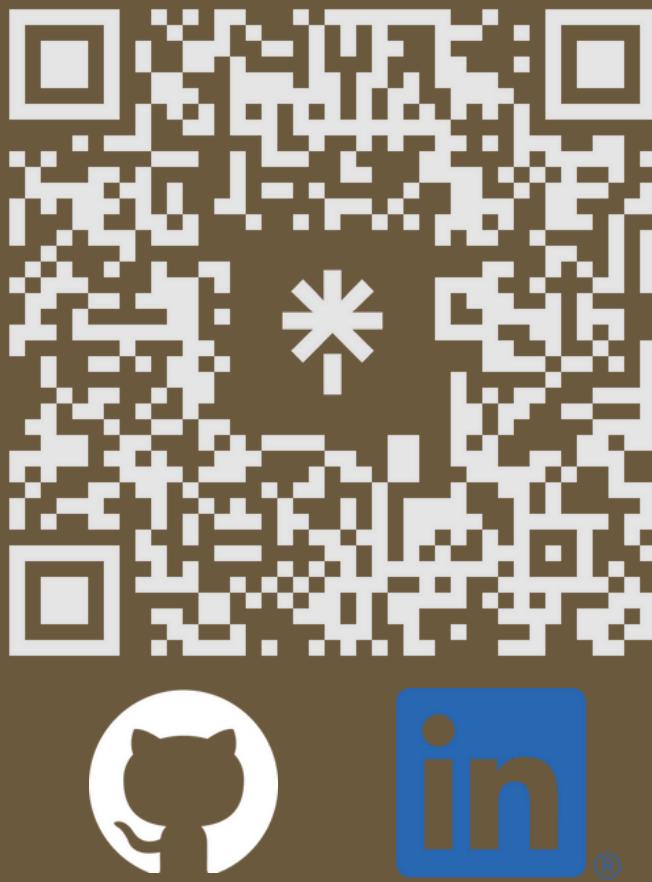
- Artist past 2000s in dataset more predictable
- Deep Learning not always the best option
- LassoLars off about \$15,000

## NEXT STEPS:

- Test DL model further
- Obtain more data points from more artists

# CONNECT WITH US

JEREMIAH TORIBIO



COLTON KANE

