Udemy Courses:

Pathways to Success

Benjamin Morris, Inbar Sapir, Yehuda Wexler

Abstract

What makes a successful Udemy course?

This study defines success as both profitability and learner satisfaction, as measured by high average ratings. Using a dataset of over 209,000 Udemy courses, we investigate the primary factors contributing to these two dimensions of success. Through multiple regression analysis, we examine how various course attributes, such as price, content length, number of reviews, and subscriber counts, impact profitability and ratings.

While our models identify several key predictors of success, concerns about model accuracy highlight the challenges of drawing definitive conclusions from observational data. These findings provide valuable insight for course creators aiming to optimize course design for profitability and learner satisfaction.

Research Questions

1. What are the factors of a successful Udemy course?

   a. What are the most important factors in releasing a profitable Udemy Course?

   b. What are the most important factors in releasing a highly satisfactory course?

   c. Are highly satisfactory courses more profitable?

Variable Overview

Dependent/Response Variable:

- Average Rating
- Revenue

Explanatory Variables:

- Length of Headline (Numerical)
- Category (Categorical)
- Language (Categorical)
- Duration of Course (Numerical)
- Number of Lectures (Numerical)
- Average Length of Lecture (Numerical)
- Number of Reviews (Numerical)
- Course is paid (Categorical)
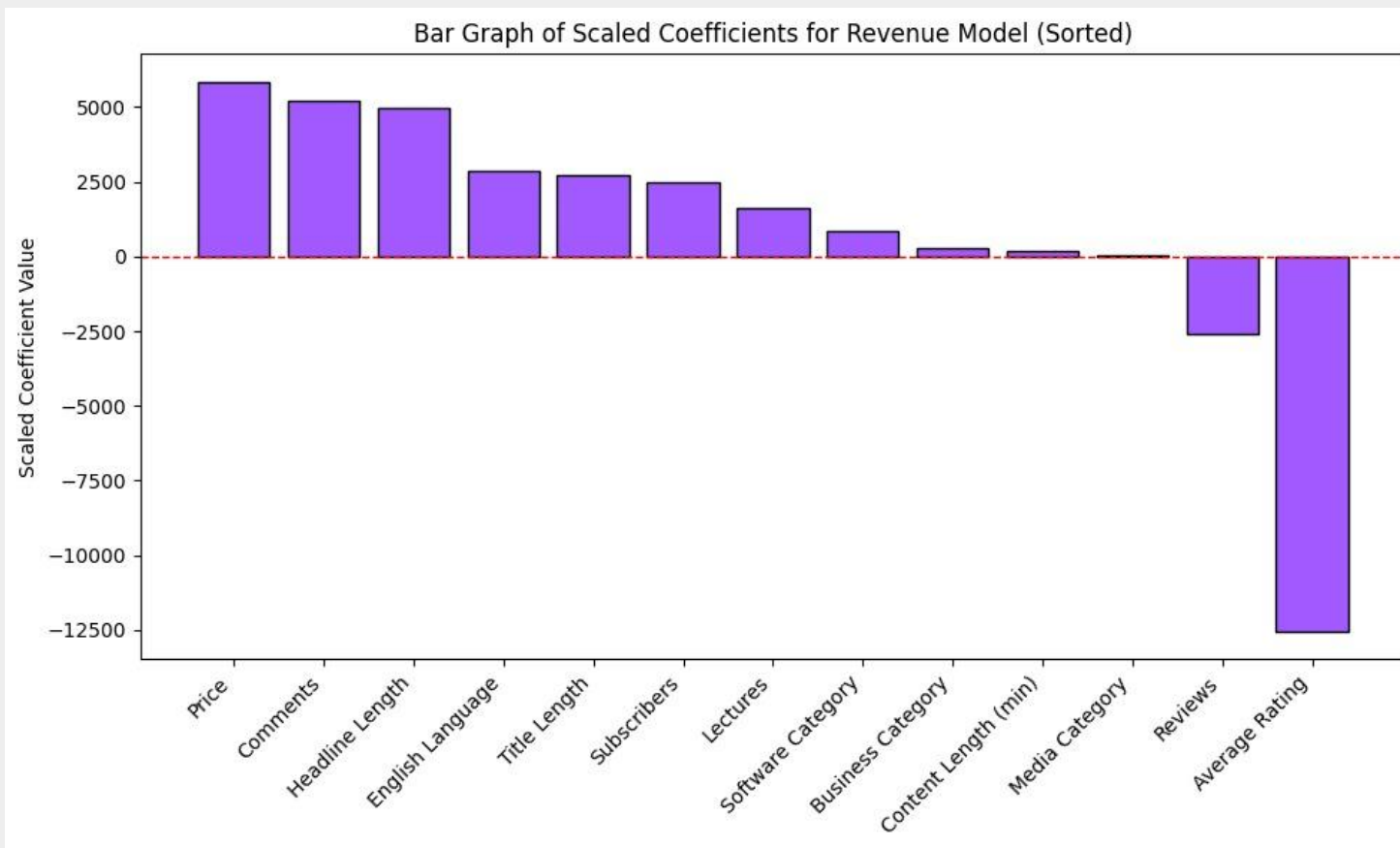- Price of Course (Numerical)

https://www.kaggle.com/code/krishna0000007/udemy-2/input?select=Course_info.csv
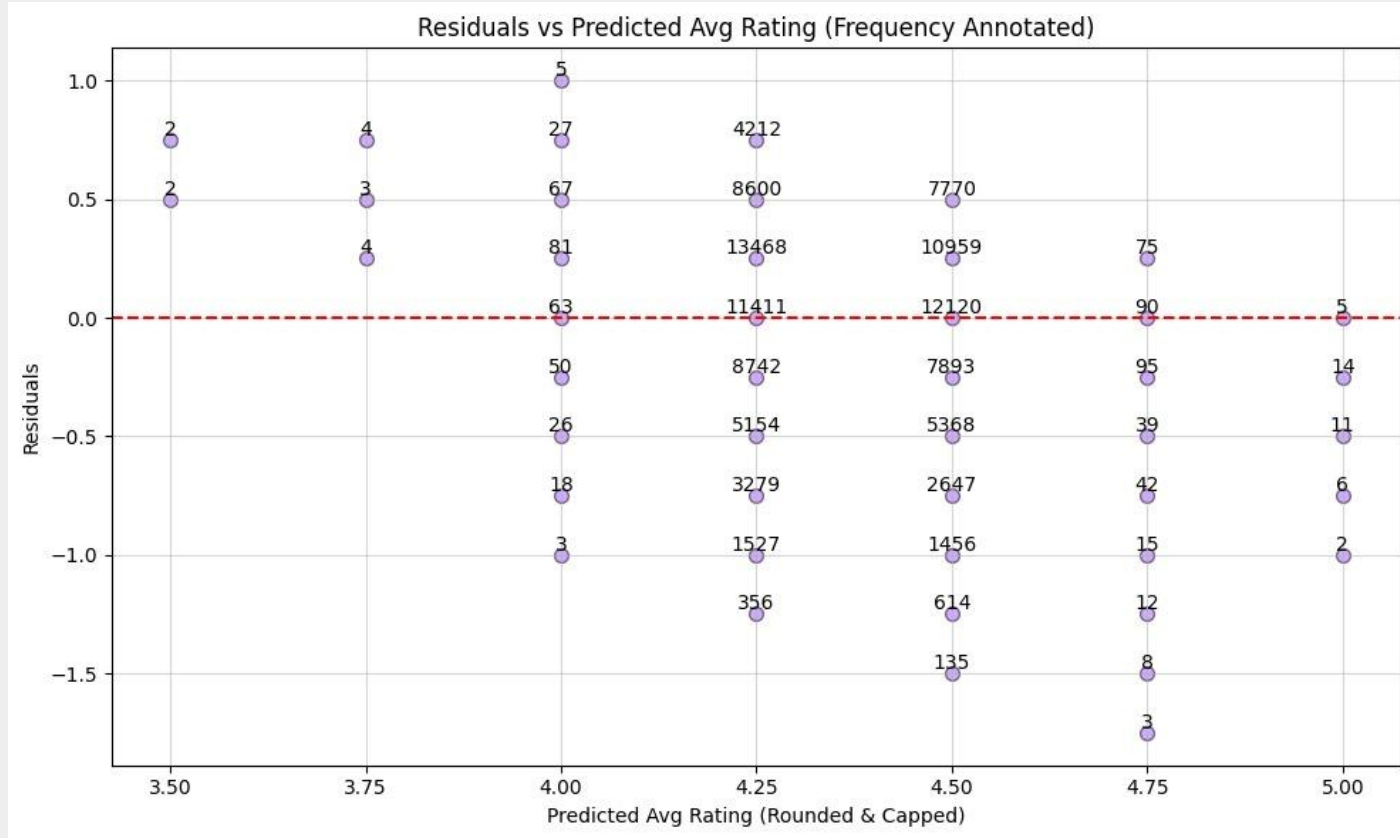
Weighted Coefficients - Avg Rating

Bar Graph of Scaled Coefficients for Avg Rating Model (Sorted)
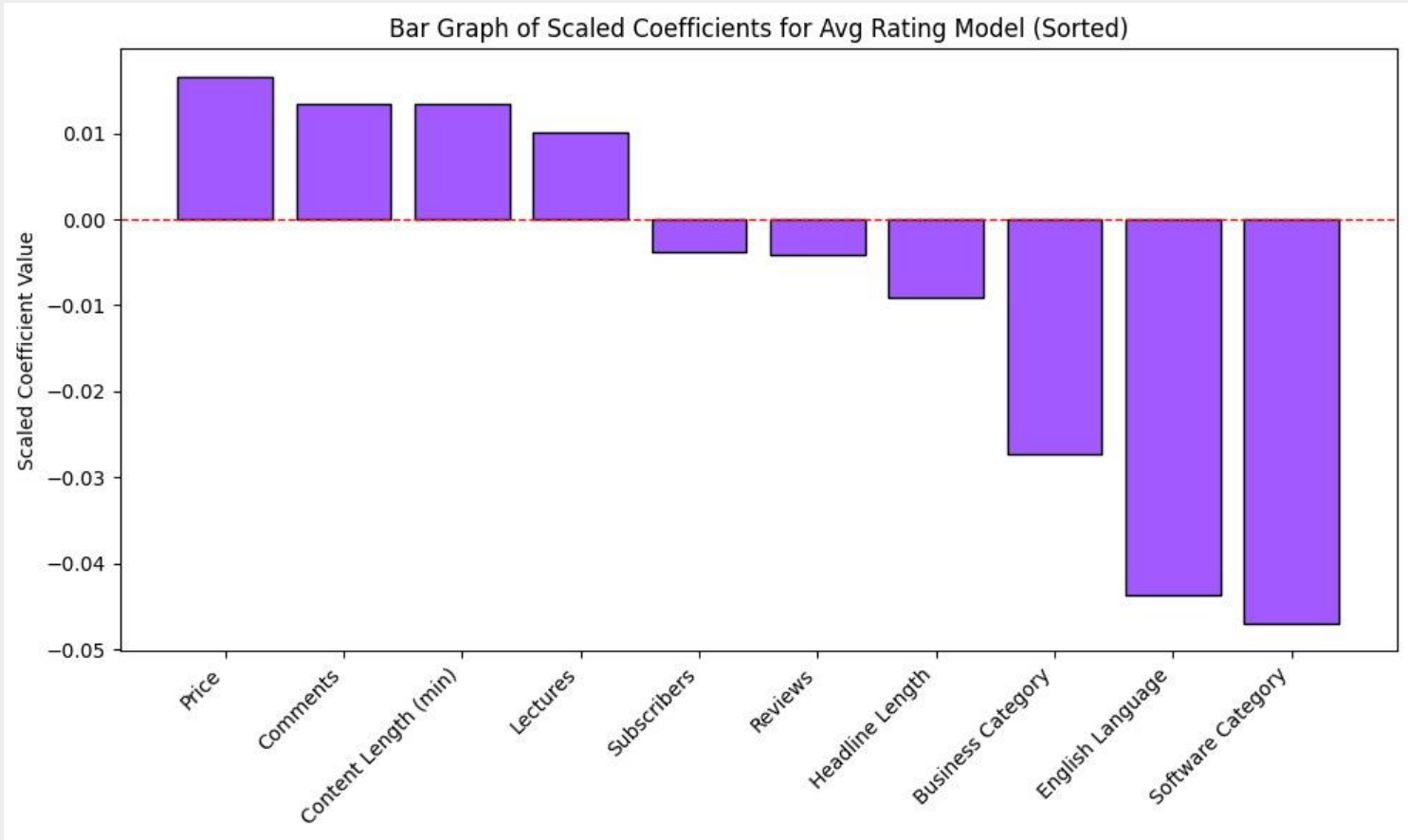
Weighted Coefficients - Revenue

Bar Graph of Scaled Coefficients for Revenue Model (Sorted)

Residuals vs Predicted - Avg Rating

Residuals vs Predicted Avg Rating (Frequency Annotated)

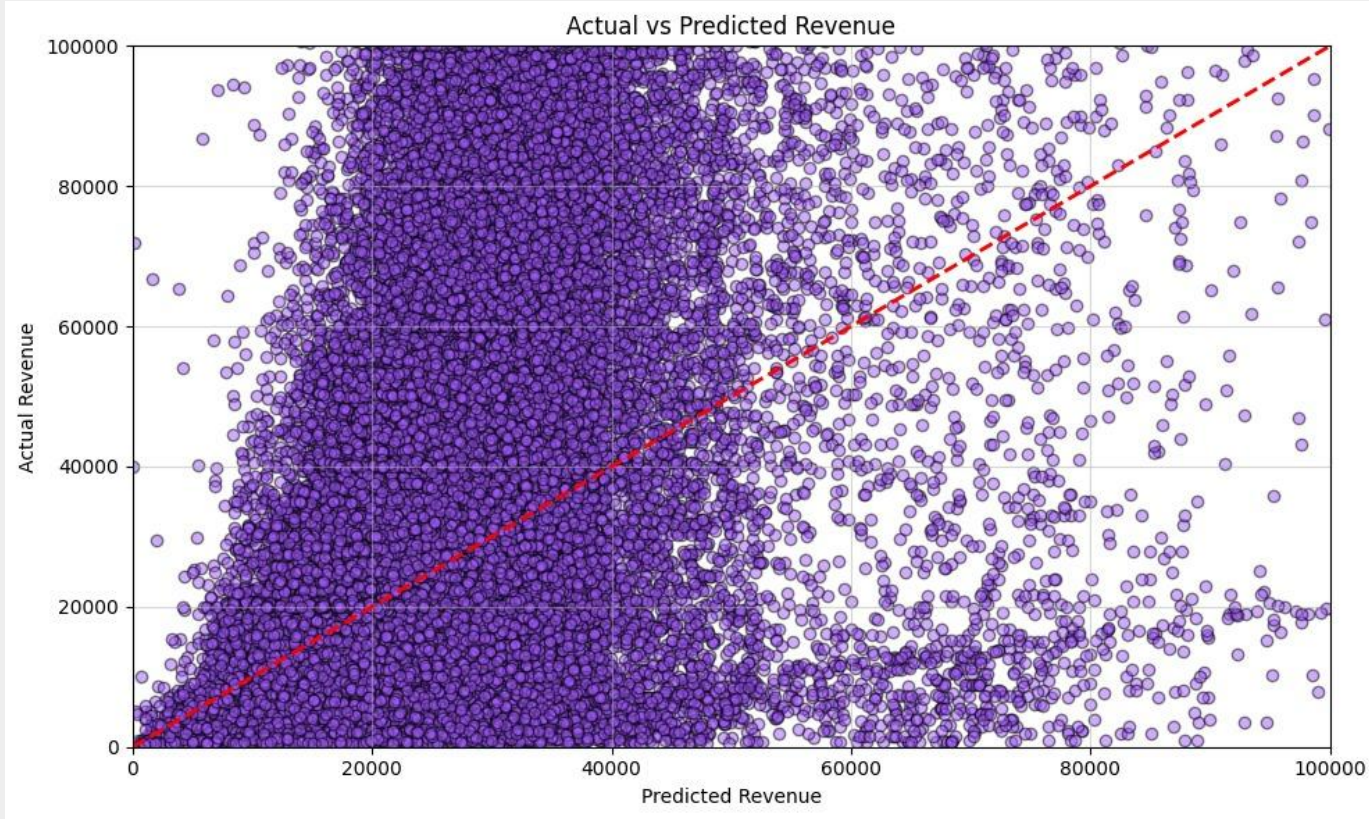# Residuals vs Predicted - Avg Revenue

Actual vs Predicted Avg Revenue

Actual vs Predicted Avg Rating (Frequency Annotated)
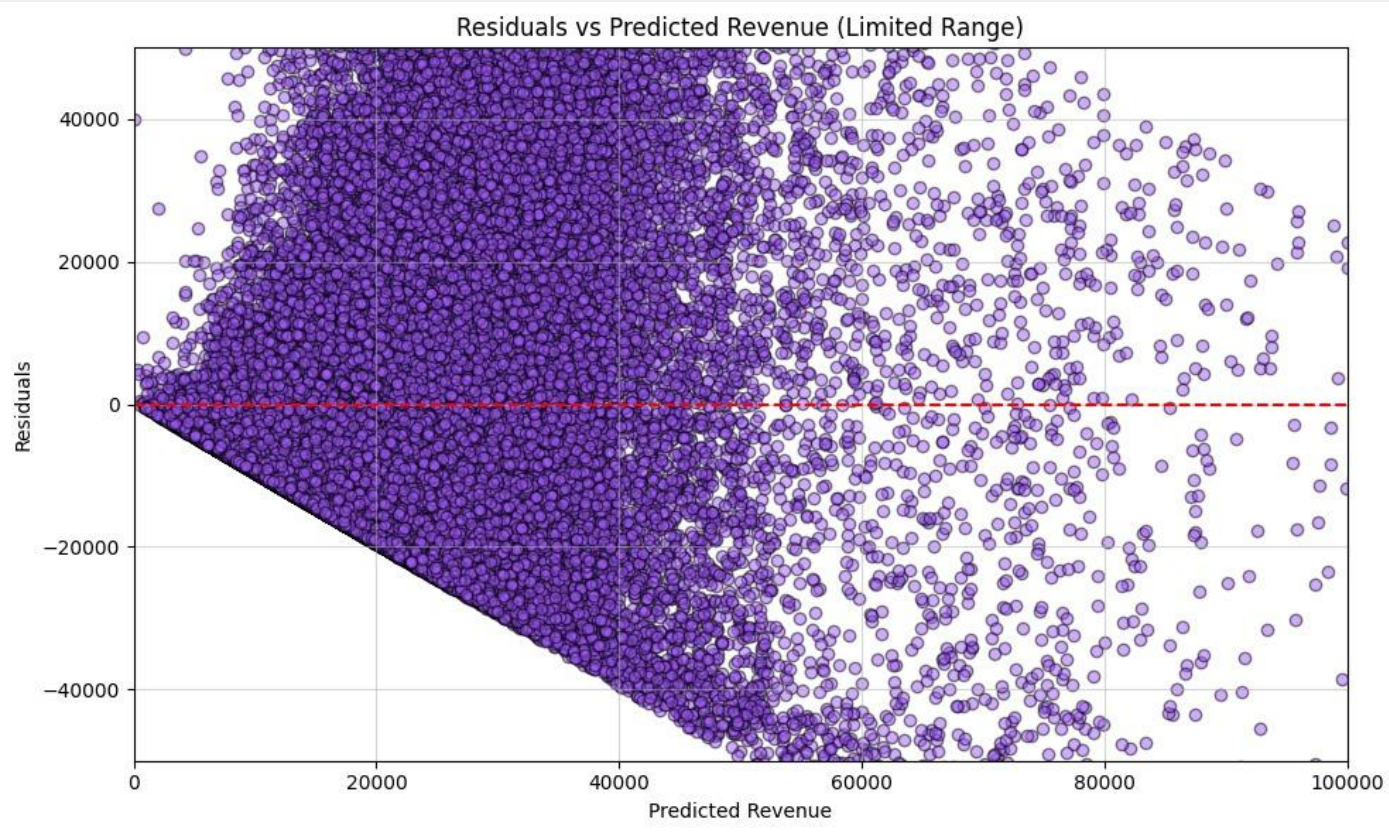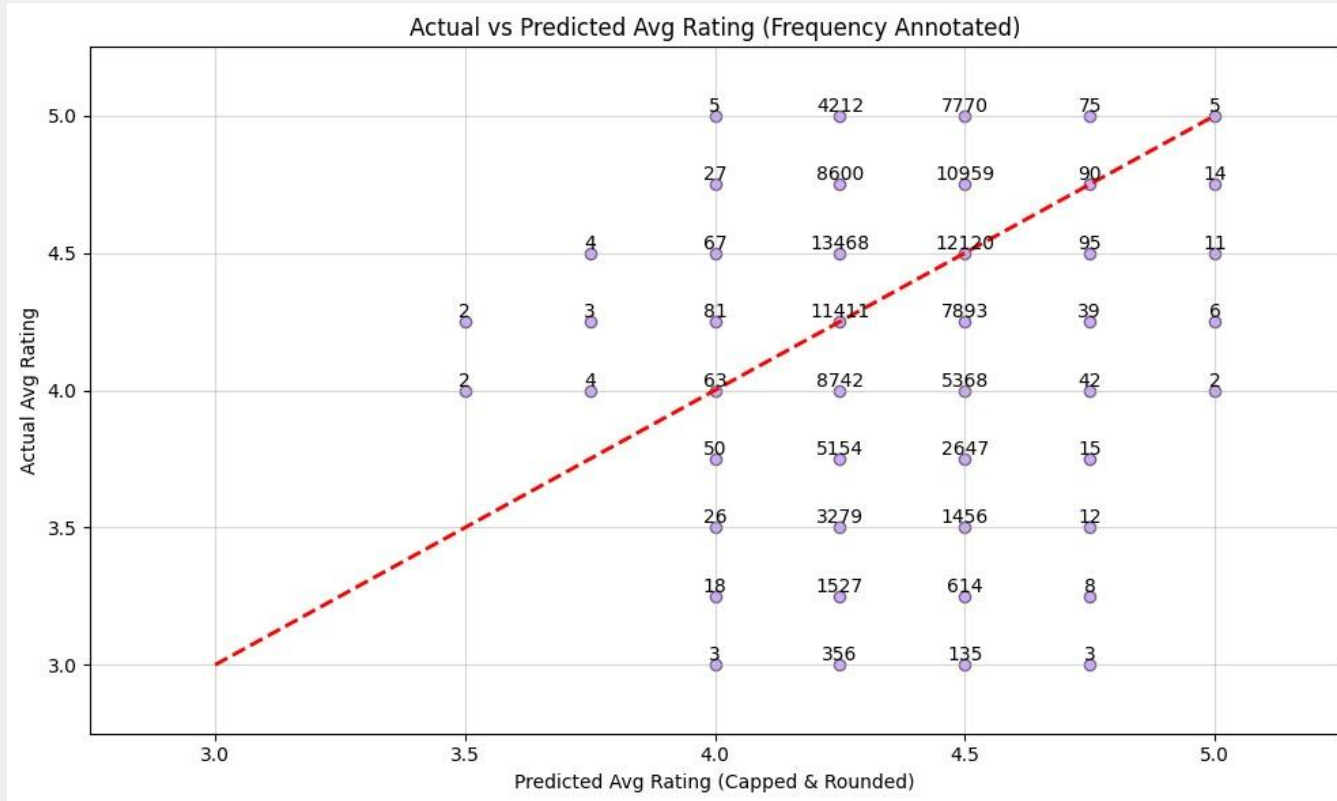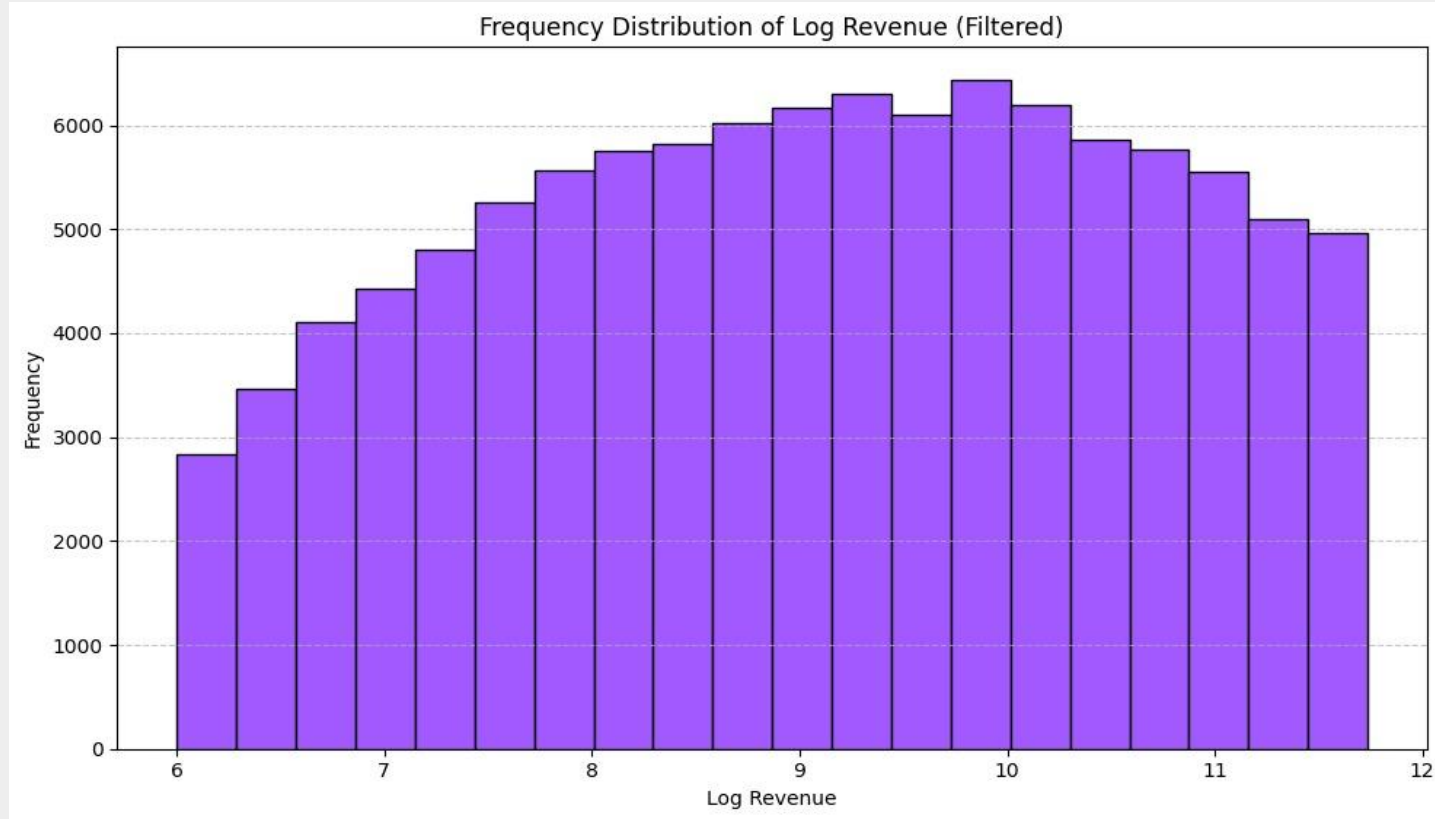
Actual vs Predicted Revenue

Frequency Distribution Revenue

Frequency Distribution of Revenue (Filtered)
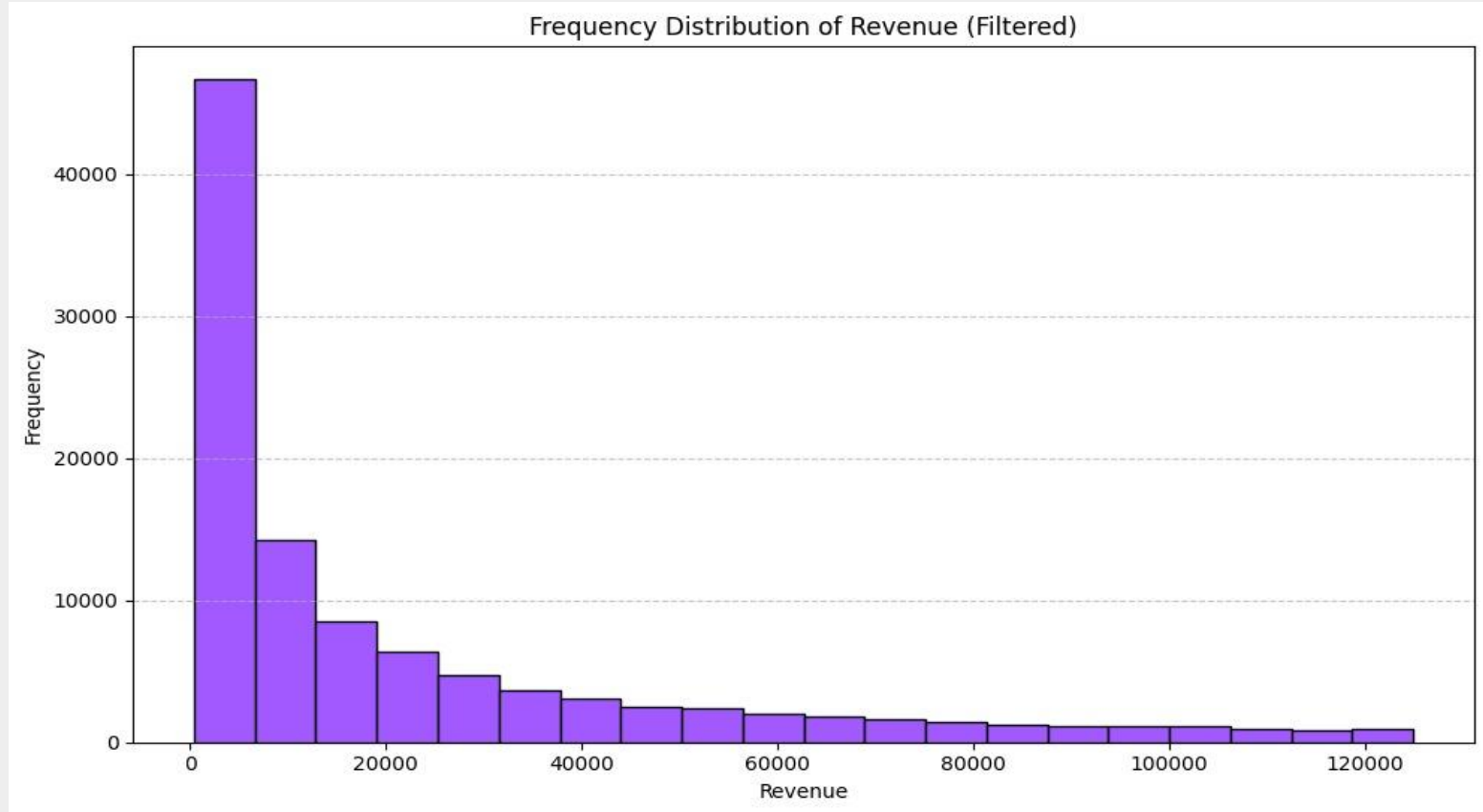
Frequency Distribution Log Revenue

Frequency Distribution of Log Revenue (Filtered)

Frequency Distribution Rating

Distribution of Average Rating
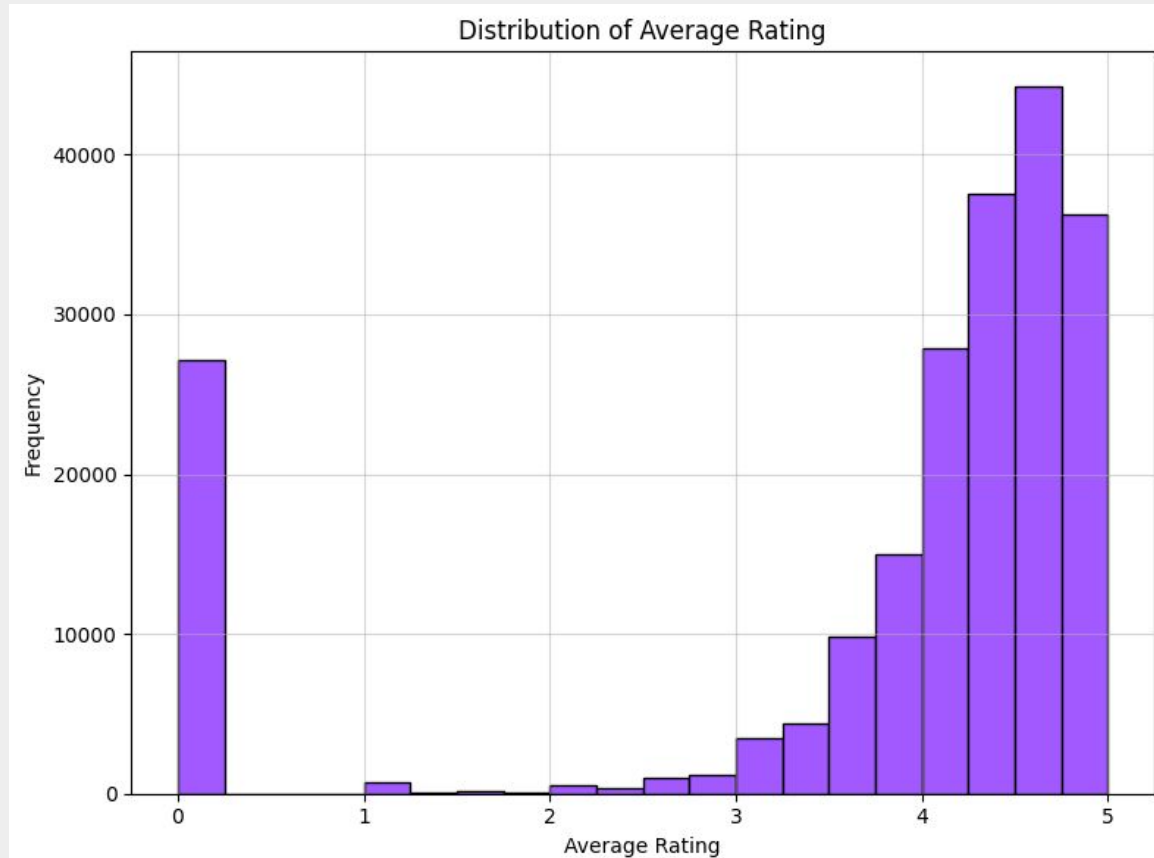
Frequency Distribution Rating (After removing < 3)
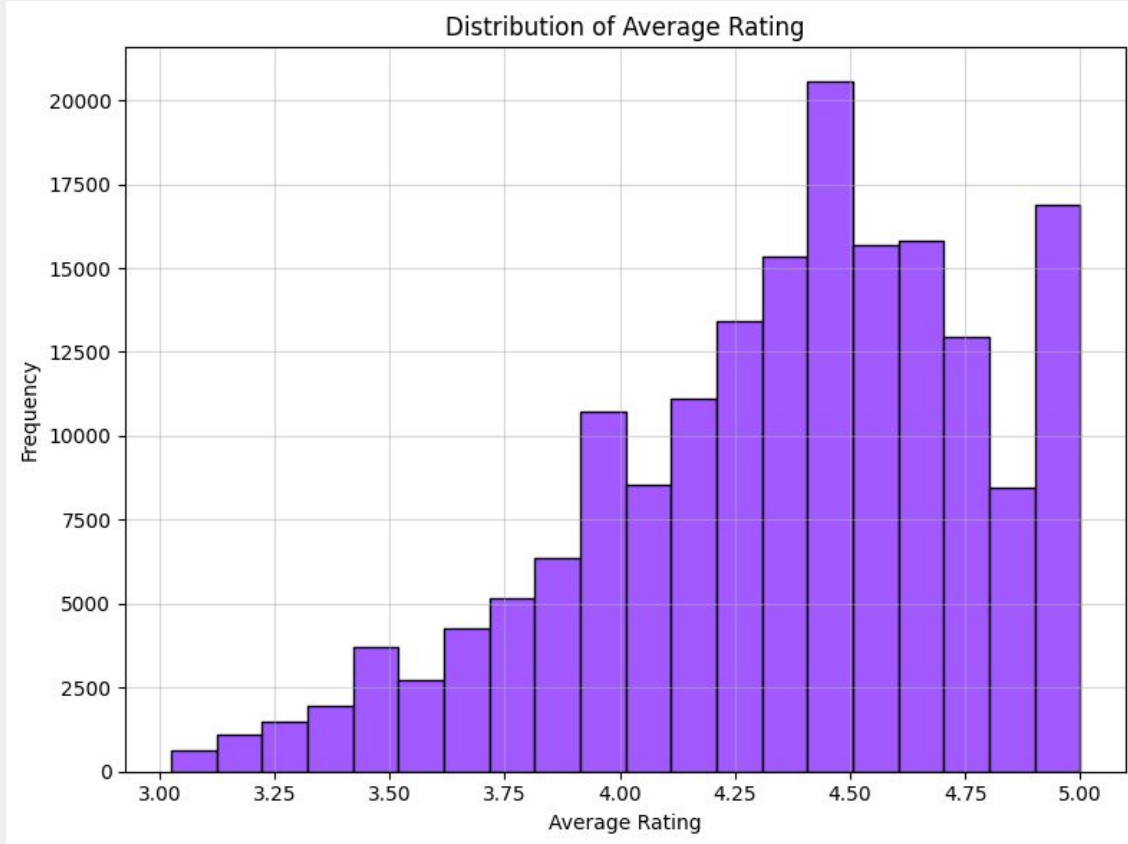
Statistical Output- Ratings

$R^2=.050$

Variable	Coefficient	std_err	P>|t|
const	4.4577	0.004	0
Headline Length	-0.0001	4.16E-05	0.006
Price	0.0002	1.34E-05	0
Subscribers	-6.28E-06	7.91E-07	0
Reviews	-7.12E-05	1.55E-05	0
Comments	0.001	9.01E-05	0
lectures	0.0003	3.91E-05	0
Content Length (Minutes)	4.90E-05	4.11E-06	0
Revenue	-8.11E-07	5.15E-08	0
English	-0.0776	0.003	0
Software	-0.1604	0.003	0
Business	-0.1254	0.003	0

Statistical Output - Revenue

$R^2=.227$

Variable	Coefficient	Standard Error	P-Score
Constant	10490	870.743	0
Title Length	61.4188	6.491	0
Headline Length	62.4641	2.654	0
Price	71.1796	8.765	0
Subscribers	4.0067	0.045	0
Reviews	-44.2191	0.914	0
Comments	369.3753	5.242	0
lectures	43.2617	2.322	0
Content Length (Min)	0.741	0.2045	0.002
Average Rating	-2879.8381	182.019	0
English	5045.3171	163.318	0
Software	2900.213	191.388	0
Business	1195.6989	207.439	0
Media	1047.3328	360.33	0.004

Statistical Output - Linear Regression of Revenue From Avg Rating R^2=.003

Variable	Coefficient	Standard Error	P-Score
Constant	36920	883.49	0
Average Rating	-3339.1942	201.714	0

Conclusion

- Importance
  - This information is important for Udemy Instructors seeking to optimize their courses for increased profit or popularity.
- Limitations:
  - We excluded all rating values < 3 (and rounded to closest .25)
  - Eliminated revenues of zero and >125,000 as well as logged revenue values <6.
  - This left us with 176,000 rows out of the original 209,000.
  - To more efficiently analyze the data, we chose to merge related categories.
  - In predicted vs residuals chart- capped ratings at 5 Maximum