LINEAR REGRESSION APPLICATION PRICE PREDICTION

▶ Raghad Aloraini

Raghad Alawad

INTRODUCTION

▶ The dataset are taken from the Google Play website, we are trying to make predictions of the price of the apps, using features such as Rating, Installs, Reviews, Category to help developers improve their work.

▶ **STEP 1**

Scrapping our data from
google play store website

▶ **STEP 2**

Cleaning & Pre-Processing,
Correlations

▶ **STEP 3**

Linear regression model

▶ **STEP 4**
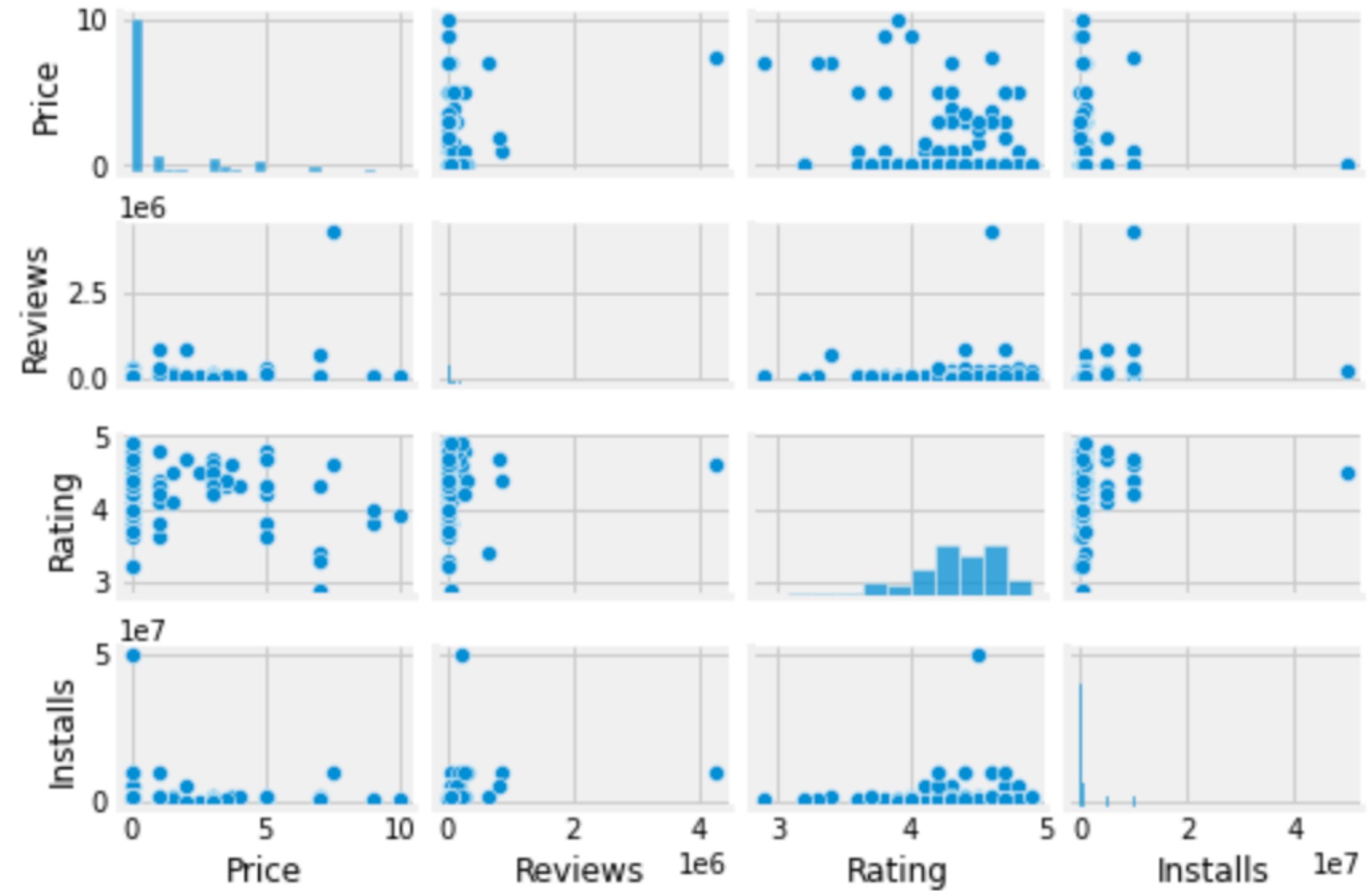
Ridge model

▶ **STEP 5**

Feature engineering

▶ **STEP 6**

Assumptions

CORRELLATION

LINEAR REGRESSION FEATURE ENGINEERING

▶ ## FIRST SCORE WITHOUT DUMMY VARIABLES

Linear Regression for test R^2: 0.1690647353900907

▶ ## ADDING DUMMY VARIABLES

Linear Regression for test R^2: 0.5980561751597653

▶ ## ADDING QUADRATIC TERMS TO THE DATA FRAME

Linear Regression for test R^2: 0.6020564981216501

▶ ## STEP MULTIPLICATIVE INTERACTION

Linear Regression for test R^2: 0.6080677905903289

RIDGE REGRESSION FEATURE ENGINEERING
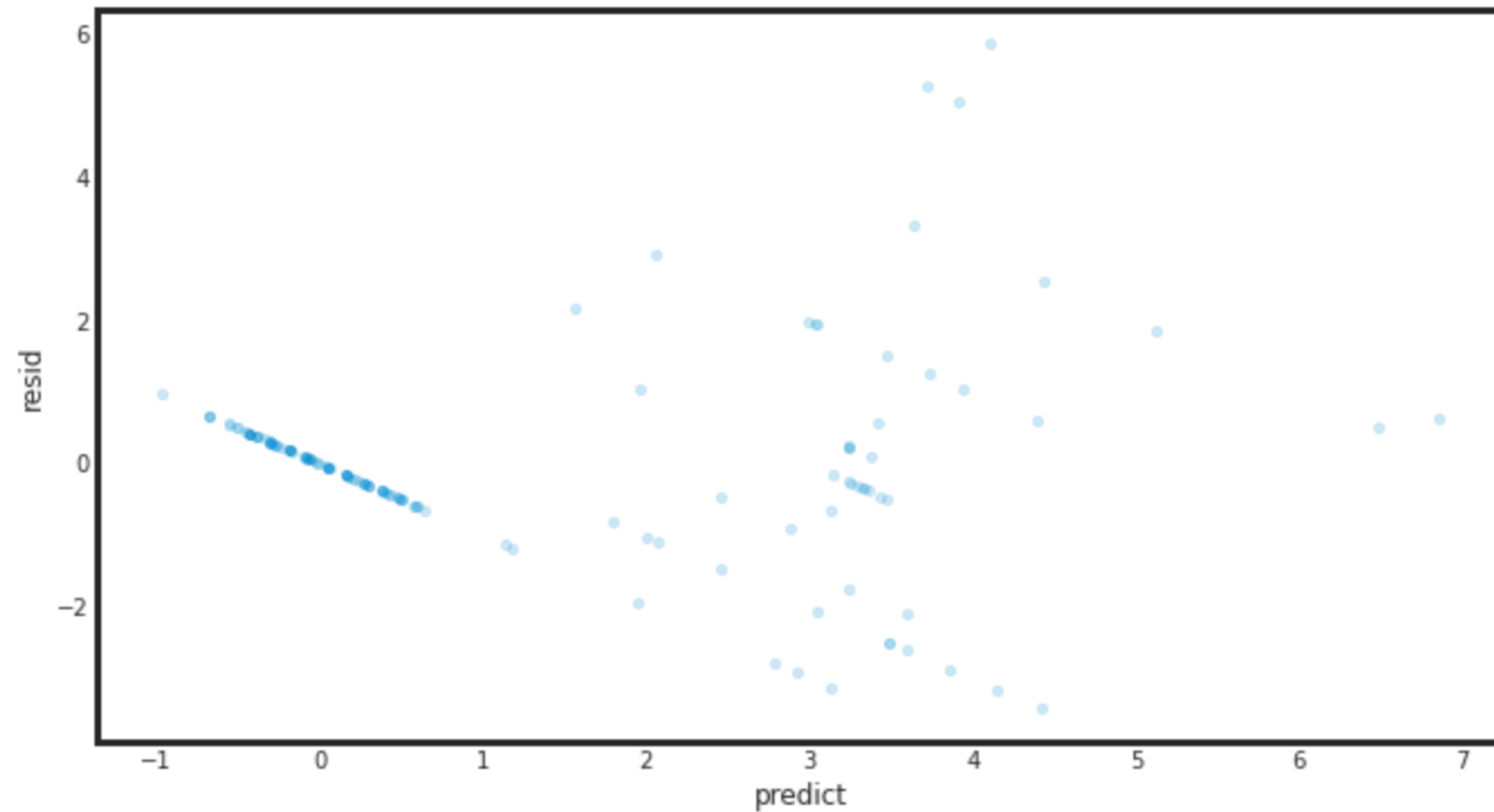
▶ FIRST SCORE

`Ridge Regression val R^2: 0.531`

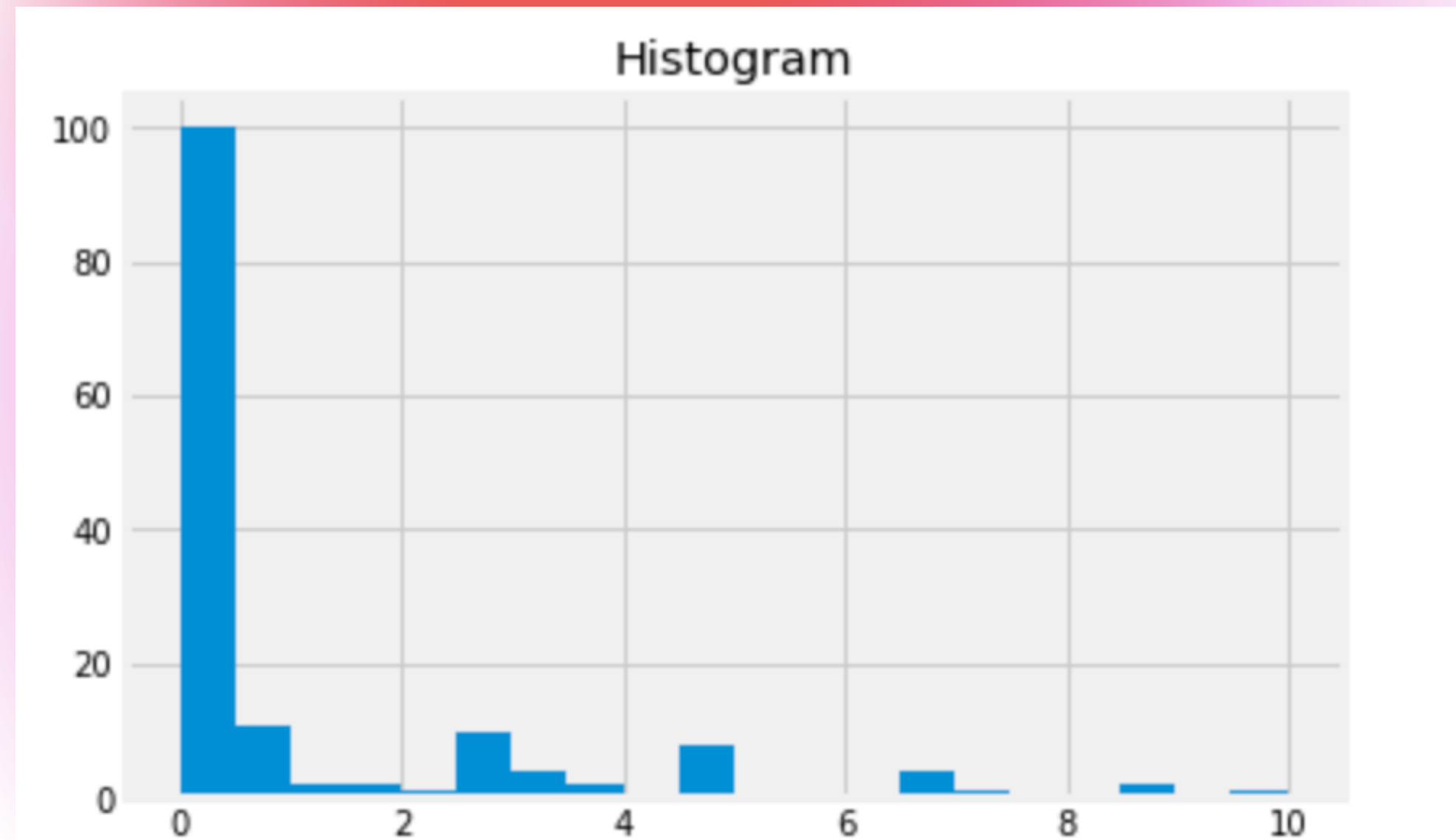▶ ADDING QUADRATIC TERMS TO THE DATA FRAME

`Ridge Regression val R^2: 0.526`

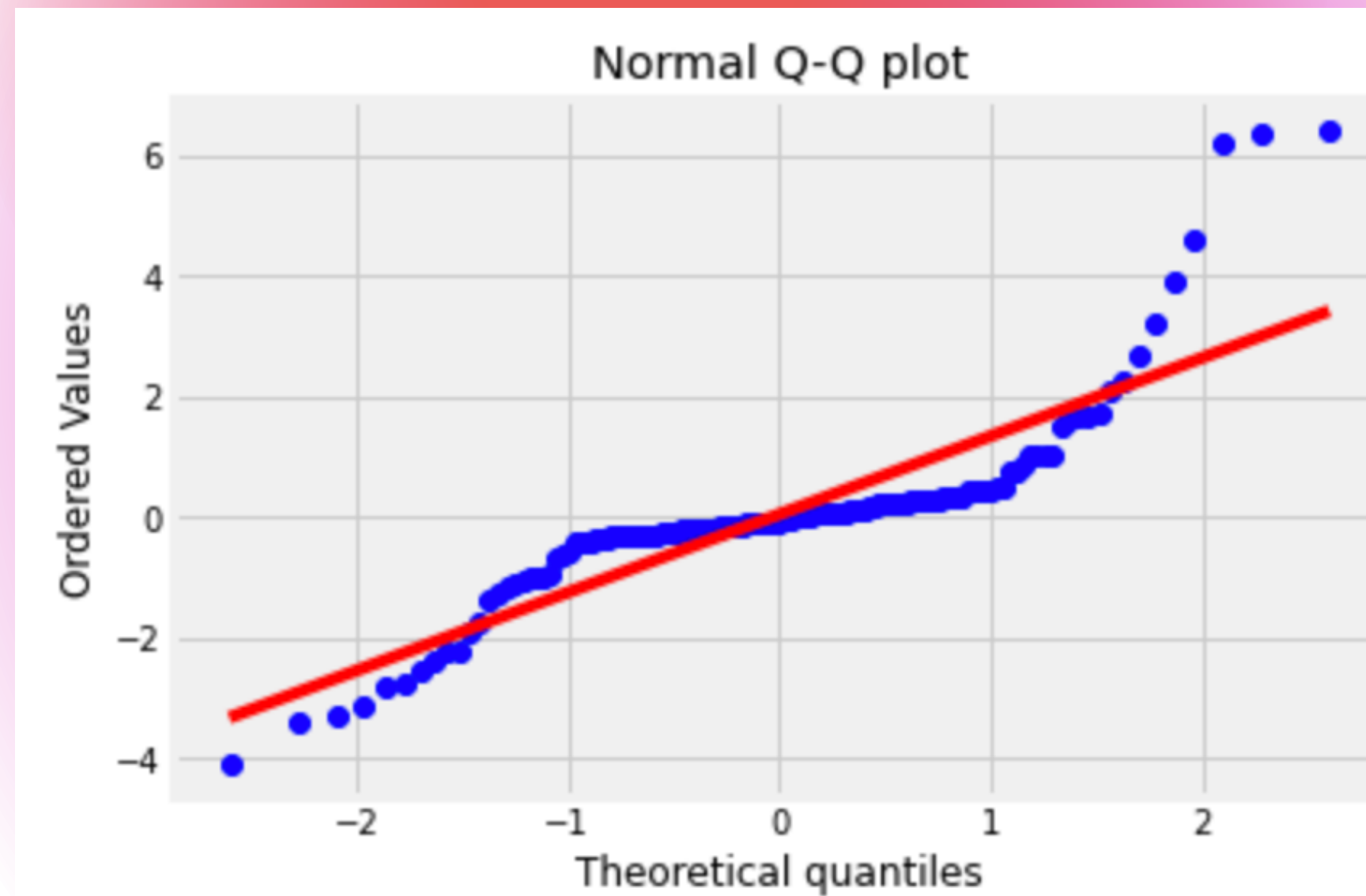▶ STEP MULTIPLICATIVE INTERACTION

`Ridge Regression val R^2: 0.540`

ASSUMPTION: RESIDUAL PLOT

ASSUMPTION: HISTOGRAM

ASSUMPTION: Q-Q PLOT

CONCLUSION

▶ THERE IS A RELATIONSHIP BETWEEN PRICE AND RATINGS

▶ IF THE PRICE OF THE APP INCREASE THE RATINGS WILL DECREASE

▶ THE DATA IS LIMITED ITS BIAESD TO ZERO

THANK YOU !