

SEPHORA

Sephora Predict product Price

Presented By:
Wafa Albattah
Taghreed Albaiz
Leena Alshwaihi



Outline :

- Introduction
- Problem Statement
- Data set
- EDA
- Feature Engineering
- Regression Algorithms
- conclusion



Introduction :

Sephora represents one of the largest marketplaces on the world. People use its products on a daily basis to order skin care, makeup, perfumes, and even web hosting services.




Problem Statement

- analyze product based on several variables, determine what variables affect product price the most, then build a model that can predict the price of a Product.
- The goal Predict the price of product based on the features available



Dataset

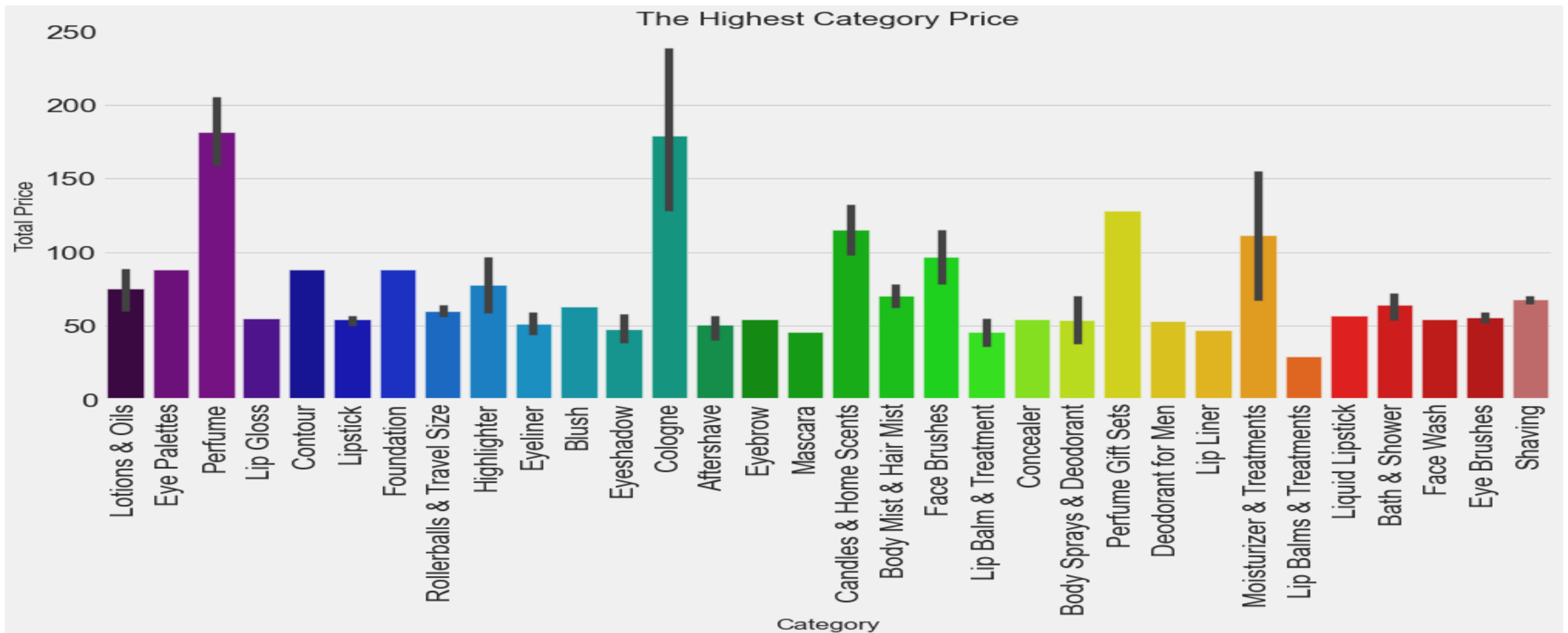
- <https://www.kaggle.com/raghadalharbi/all-products-available-on-sephora-website>
- kaggle.com
- [Sephora Website](#)
- More than 9,000 Products with their Ratings and Ingredient Lists
- Has a total of 9268 rows and 21 columns

The image shows the Sephora logo, which consists of the word "SEPHORA" in white, uppercase, sans-serif font. The logo is centered within a black horizontal band. This band is flanked by two sections of red background with a dense, sparkling glitter texture. A white curved line separates the glitter background from a dark gray background on the right side of the image.

SEPHORA

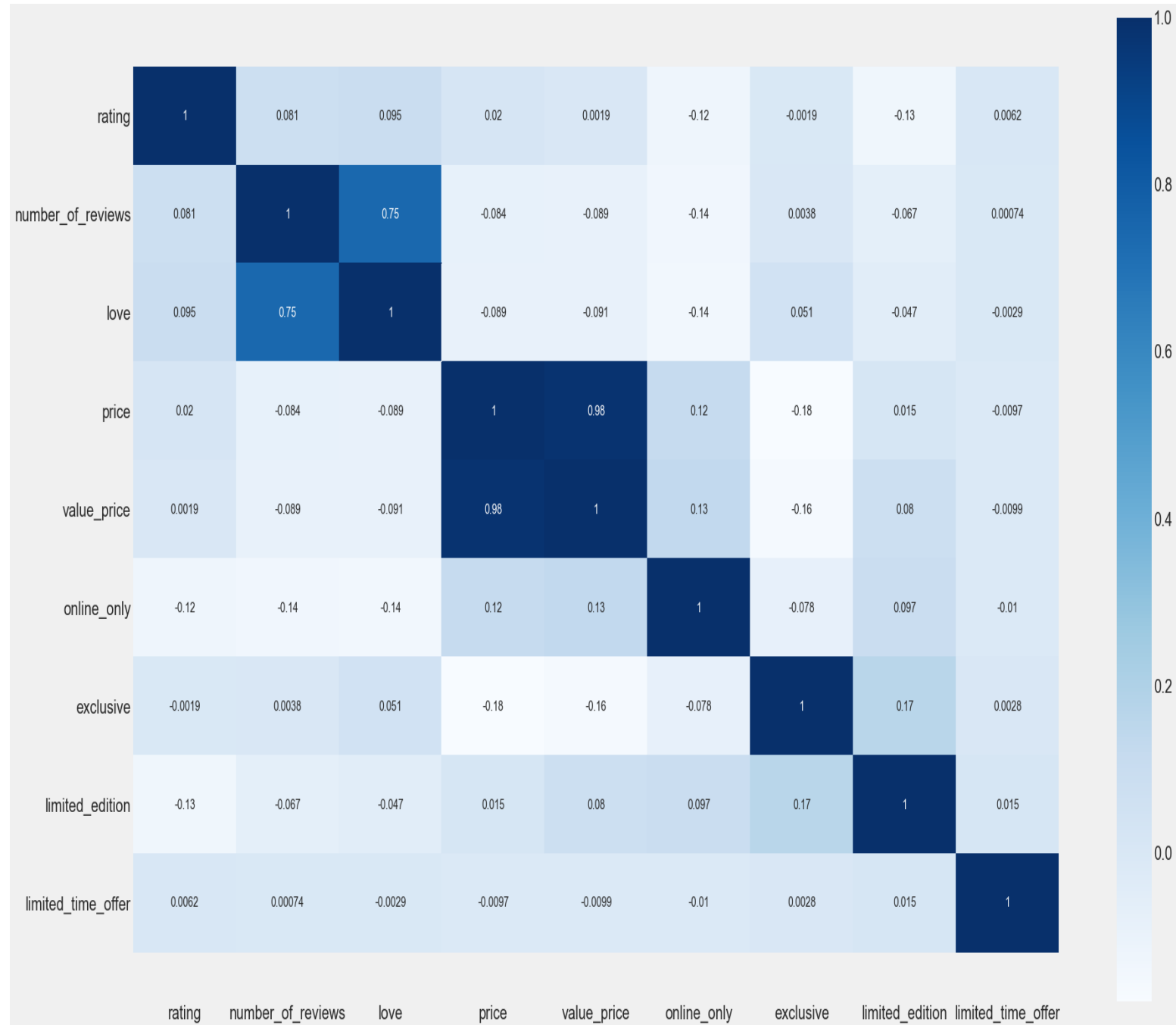
Expolarity
Data
Analysis

The highest Category based on Price

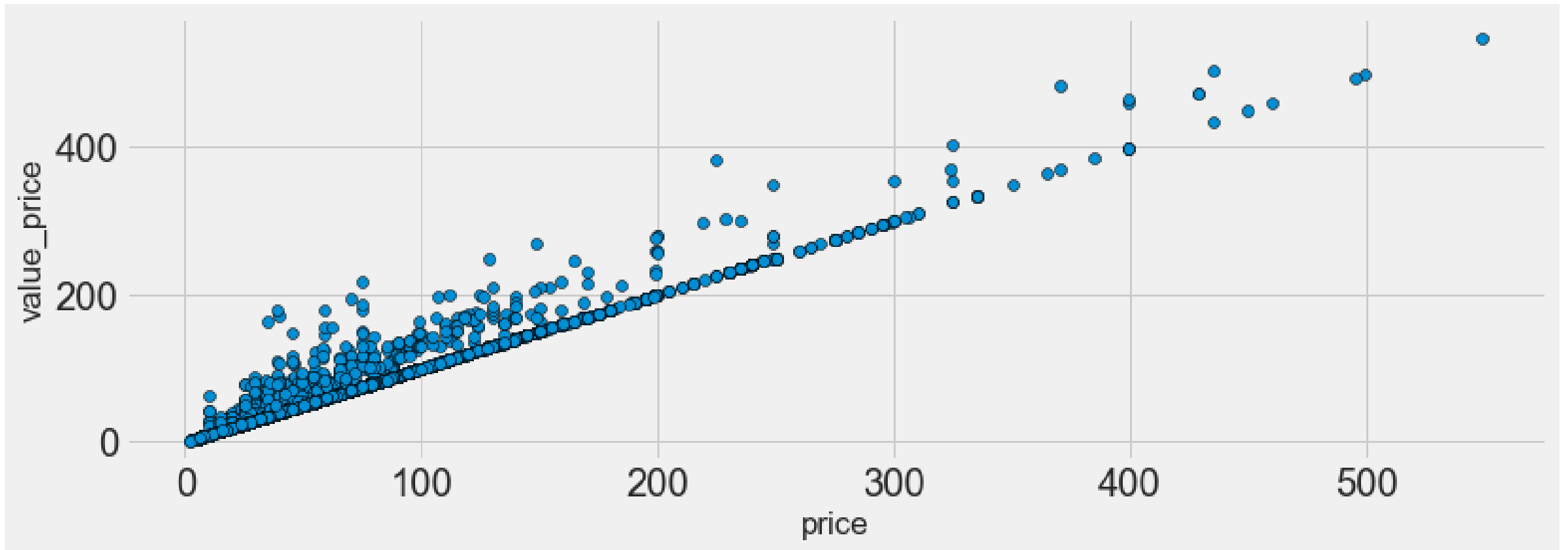


Visualization

- Value price higher coloration with price and online only




The highest price highest value price





Feature Engineering

- Create dummy variables for the categorical features
- Handle outliers in the target variable

The Sephora logo is centered on a horizontal black band. The band is flanked by two sections of red background with a dense, sparkling glitter effect. A white curved line separates the logo area from the dark grey area on the right.

SEPHORA

Regression
Algorithms

XGBoost Algorithm

R2 on Training set : 0.8195442843885937
R2 on Validation set : 0.7081371579318474

Error Table

Mean Absolute Error : 0.059941945145433245
Mean Squared Error : 0.007068365883478215
Root Mean Squared Error : 0.08407357422804275

Random Forest

R2 on Traing set : 0.9589680768507084

R2 on Validation set : 0.7219987992161072

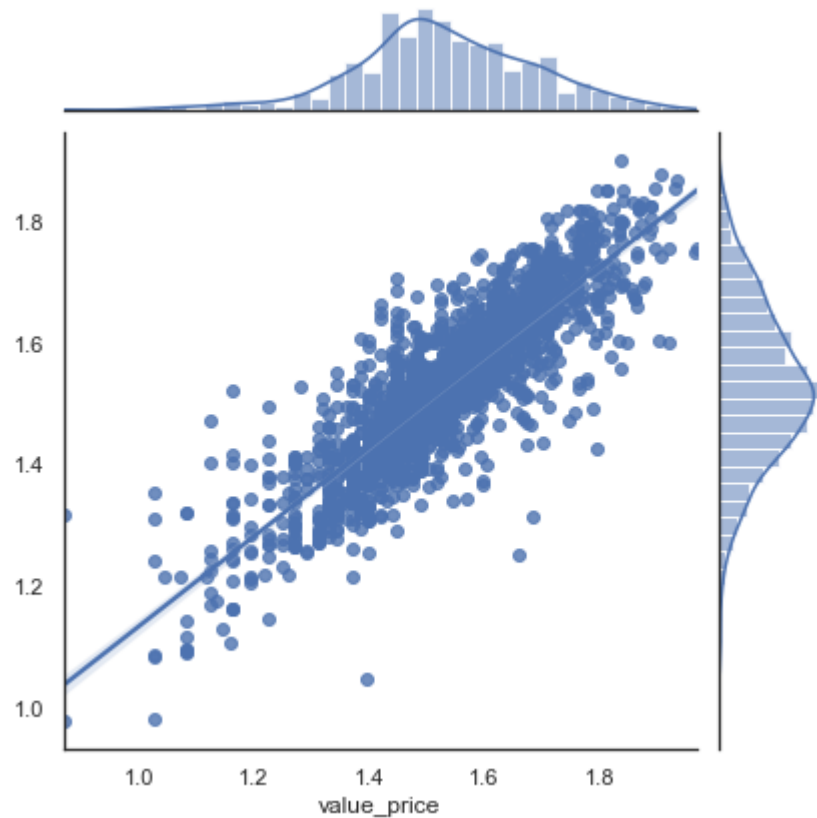
Error Table

Mean Absolute Error : 0.055895392402975703

Mean Squared Error : 0.006732663155277563

Root Mean Squared Error : 0.08205280711394074

Ridge Algorithm



R2 on Traing set : 0.804965637330968

R2 on Validation set : 0.7476897832258586

Error Table

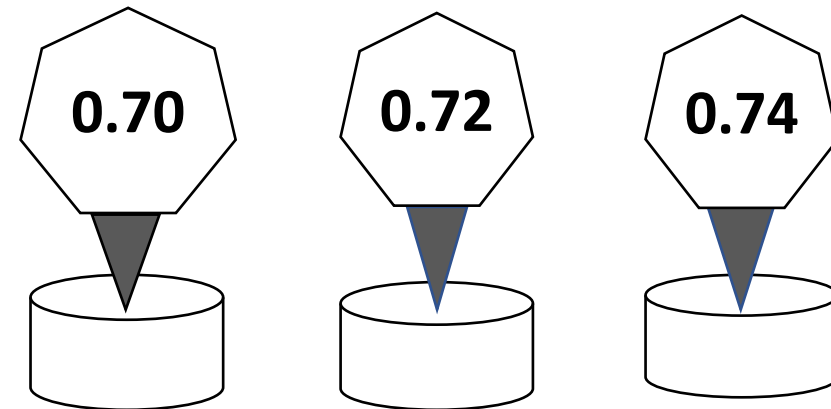
Mean Absolute Error : 0.055101920084600915

Mean Squared Error : 0.0061104761252303895


Root Mean Squared Error : 0.078169534508211

Conclusion :

Results of the R^2 (validation) for the three models:



Ridge Regression Algorithm has the best results!

The Sephora logo is centered on a black horizontal band. The background of the entire slide is a vibrant red with a dense pattern of small, shimmering gold and white confetti. A large, smooth, light gray curved line sweeps from the bottom left towards the right, separating the logo area from the text area.

SEPHORA

Thank you!

Any Questions?