

Pallavi Vemuri
11/26/2024
DS 4002
Professor Alonzi

Using Data Science to Evaluate Accuracy of Pinterest's Algorithm in Identifying Warm and Cool Tones

Overview

Our world is obsessed with beauty. From finding the perfect lipstick shade to understanding what color palette plays best with one's complexion and skin tone, the beauty industry is a 528.6 billion dollar industry that is projected to grow 8% year over year¹. Enter the 80's trend that has made a large comeback— personal color analysis. The premise is that everyone has undertones that determine their color season. This viral trend has even fueled a tourism boom for South Korea, where people flock to in order to receive a dossier of flattering colors to assist in clothing, makeup, and accessories choice. Many others have created Pinterest boards to spark inspiration on crafting their color season wardrobe “cool tone makeup looks” or “warm tone outfits.” These boards are only as useful as the Pinterest algorithm, however. For example, it may classify cool tones as warm, and vice versa. The concept of what is a cool and warm tone is also loosely defined, making it difficult for the Pinterest algorithm to be accurate. And yet, plenty of women rely on Pinterest and its algorithm for inspiration on what to wear and what makeup looks to try out; potentially resulting in women believing a shade to be warm when it is actually cool, or vice versa.

Deliverable

L'oreal's marketing team is looking into the Pinterest algorithm to see if pin recommendations for cool and warm tones are fundamentally different. L'oreal's goal is to understand consumer behavior and use it to inform a future marketing strategy involving color analysis pop-ups, and your work is critical in this endeavor.

Your mission, should you choose to accept it, is to analyze two Pinterest boards with warm and cool pins, added based on the Pinterest algorithm, and to report whether the average color of each pin is statistically different from one another by finding the average color of each pin and creating a random forest model to test whether there are significant differences in average color.

Github Repo: <https://github.com/Pallavs24/CS3-DS4002>

¹ [1] “Beauty Industry Revenue Usage Statistics”
[https://helplama.com/beauty-industry-revenue-usage-statistics/#:~:text=The%20beauty%20industry's%20net%20worth%20is%20estimated%20at%20%24528.6%20billion,in%20the%20last%2013%20years.\(accessed Nov. 26, 2024\).](https://helplama.com/beauty-industry-revenue-usage-statistics/#:~:text=The%20beauty%20industry's%20net%20worth%20is%20estimated%20at%20%24528.6%20billion,in%20the%20last%2013%20years.(accessed%20Nov.%2026,%202024).)