**Building Price Estimator Model for Lavender Lane case study**.

**Scenario**

A retail company in Ghana has requested for a price estimator on the website to help its customers have an estimated price of the products in stock to aid them have a smooth customer buying experience. Due to this, the marketing manager contacted its team of website developers to help with this feature. The website developers advised that they do not have machine learning engineers nor any expertise in machine learning on their team. Hence recommended that they onboard data professionals with these expertise to help since they would need data to work on. Rest-assuredly, they will add the new feature on the website but for functionality, they will count on the machine learning model built at the back end. The marketing manager then relayed this information to the CEO who requested that the team recruits Data Analysts, Data Scientists and Machine Learning Professionals ASAP. The team wants to be the leading company in the field to accelerate the change in the industry. The team arranged and included in the budget onboarding these professionals. You saw the advertisement and applied for the role. At the interview, you mentioned how you built a predictive model to estimate prices based on some key features provided in the dataset. In your explanation, you mentioned that the dataset used was a fashion dataset which was very close to what the team needed. This made you stand out in the first stage of the interview. The team has requested that you showcase this project at the second stage of the interview and explain your reason for choosing the model.

**About the company**

Lavender Lane is a new retail company in Ghana committed to giving its customers the best customer experience. They have in stock products from top brands like Adidas, Nike, Zara, H&M and Gucci. As part of the company’s mission in Ghana, it seeks to create more employment for the locals in the country hence employ locals to run its business. Lavender Lane is growing steadily and has moved into digital space. Lavender Lane estimates that in ten years’ time, it would be a fashion hub for West Africa and beyond. Recent technological advancements have made the fashion industry very competitive but with innovation and creativity, Lavender Lane knows that it will strive to be the best in the industry. As a result, it has onboarded Data Analysts and Data Scientists on the team to help with market research, and insights to make data-driven decisions for the company.

**MODEL CREATED BY**: Angelica Adjei-Kwarteng.

**Date: 15th November, 2024**

This is my capstone project on machine learning as part of my formal training at Blossom Academy.

Data source: <https://www.kaggle.com/datasets/bhanupratapbiswas/fashion-products/discussion/450302>

**Project Objective**: To build a price estimator model.

Jupyter Notebook analysis and model explained.

1. Imported all relevant libraries.
2. Loaded the dataset into the notebook.
3. Conducted some exploratory analysis:
4. Inspected data.
5. Performed some data transformation: a. changed some datatypes

b. no missing values and no duplicates in dataset hence perfect for building the model.

c. dropped product id column and renamed User ID and Rating columns.

d. created a new Rating column from the Rating Scale column.

e. included visualizations to inform stakeholders and gain better insights.

f. performed hypothesis testing to explain that there was no significant difference in ratings across the brands in stock.

g. Since our features were categorical data, they needed to be converted to numeric data to work with. Note that our target variable is price.

h. Conversion was done using one-hot encoding on features.

i. Brand and category features were merged leaving products pivoted.

4. Split datasets into training and testing datasets.

5. Machine learning models used: Decision Tree, Random Forest and Linear Regression.

6. Trained model and made price predictions.

7. Evaluated the model. Linear Regression model is the best Model.

8. Created a sample visualization of errors from our models. (optional)

The smaller the error, the more efficient the model will be.

Thank you!