



# PREDICTING CONSUMER PRODUCTS' PRICES AND PERFORMANCE

May, 2021

---

Damilola Efuwape  
Springboard

## Abstract

*Price and demand are generally believed to move in opposite directions. Changes in price usually cause a movement along the demand curve, but not a complete shift. Understanding how a potential price change could affect product sales and vice-versa is very important in developing good strategies to improve sales and increase profit. This project will use patterns of known changes in prices of certain products and quantities demanded to predict future prices and quantity sales based on price changes.*

## Problem Statement

What quantities of consumer products are likely to be sold after anticipated price changes?

## Specifications

The project will pay close attention to the "Sales Amount" and "Quantity Sold" columns to determine a pattern of price changes of goods and quantity sold for better analysis and prediction. This project will help manufacturers and management of retail stores track, analyse, assess and determine stock prices of products. Alternative questions that will be answered include:

1. What are the purchase patterns of certain products by certain customers?
2. How does price changes affect demand for substitute products?
3. What other factors cause a change in demand and quantity demanded?
4. What relationships exist between price and quantity demanded?

## Data Collection

A CSV-formatted retail store sales transaction data with one hundred and thirty-one thousand, seven hundred and six rows and eight columns from Kaggle will be used for this project. The data usability is 9.4. Below is the link to the dataset:

<https://www.kaggle.com/marian447/retail-store-sales-transactions>

## Data Processing, Preparation and Feature Engineering

Although the data does not have any missing values, regular review of the data will be done to check for other defects. This will be done through data profiling and data mining. In order to not deface the data, outliers will be capped with percentile.

## Machine Learning: Model Selection

Some regression models and the Random Forest model will be used for the project. Train-Test Split will be used to test the model performance and cross-validation will be used for performance validation.

The results, along with a report and presentation slide will be uploaded to GitHub upon completion of the project.