**Price Predictions**

**Scenario:** You are tasked with predicting the prices of kids' resale items based on the data obtained from our Shopify store. The attached Excel sheet contains information about thousands of clothing items and accessories, and we would like you to focus on **active listings** [Filter for active listings and remove archived listings]**.** Specific criteria is provided for analysis, including Brand, Category, Condition, and potential considerations based on the item title.

**Instructions:**

1. **Dataset:** Utilize the provided Excel sheet data from Shopify, paying attention exclusively to active listings.

[products\_export\_1-2.csv](https://prod-files-secure.s3.us-west-2.amazonaws.com/4355f174-e99a-439e-8bc0-9d34fb43bba0/eaabc096-0a5e-4a99-9214-a222198ce311/products_export_1-2.csv)

1. **Input Variables:**
   * Brand
   * Condition (Cleaned Values: Excellent, Very Good, Like New, Brand New with Tags, New without Tags, Good, Play)
   * Item Type (There will be about 10-18 cleaned values)
   * Age/Size (Optional, evaluate if needed)
2. **Output Variable:**
   * Price
3. **Tasks:**

a. **Data Exploration:**

* + Draw 3-4 scatter plots to visually explore the relationships between input variables and the output variable.
  + Properly label the axes and differentiate conditions with various colors or markers.

b. **Data Cleaning:**

* + Remove archived listings to consider only "Active Listings" in line with customer instructions.
  + Clean the "Condition" column based on the provided final values.

c. **Outlier Detection and Normalization:**

* + Identify and remove outliers from the dataset.
  + Provide an explanation of your approach to outlier detection and removal.
  + Normalize the data after outlier removal.

d. **Feature Engineering:**

* + Enhance the dataset by incorporating criteria such as 'contains' from the item title.
  + **Example:** For the item type "Pants," consider the following 'contains' criteria:
    - Joggers
    - Jeans
    - Leggings
  + If the title contains "Joggers," create a new feature indicating the item type as "Joggers."

e. **Regression Analysis:**

* + Choose an appropriate regression analysis model for predicting prices.
  + Split the dataset into training and testing sets.
  + Train the model on the training set and evaluate its performance on the testing set.
  + Assess the model's accuracy and discuss any improvements or adjustments you would make.

f. **Prediction:**

* + Make price predictions for a sample of items from the dataset.
  + Compare the predicted prices with actual prices to evaluate the model's effectiveness.