

## Weekly Challenge 4 Group 7

### **Problem statement:**

XYZ Company is exploring an environmentally friendly practice and aims to analyze the relationship between raw materials and finished products. The main goal is to determine whether the relationship impacts the profitability of envelopes, while maintaining competitive pricing.

We began by cleaning up the original dataset, BSS\_Retail\_Data, while working the data set, we ran into high amounts of outliers per column. We handled the outliers by replacing and filling with the mean of each column, however there were some that still had high numbers of outliers which we decided to drop due to them not having sufficient data to be relevant. After preparing the dataset, we then merged our external data sources to align by the appropriate date. Both data sources were imported from FRED using the web scraping feature, this provided recycled prices for wood and plastic based on specific dates. To focus only on envelopes, we dropped all non-envelope products from the SKU column in this dataset. This allowed us to accurately compare projected profits for envelopes using recycled materials.

### **Hypothesis:**

Envelopes produced using recycled wood will have higher profit margins compared to those that are made with recycled plastic, this is due to the lower cost of goods sold (COGS) associated with recycled wood.

### **Key Insights:**

Analyzing the histograms for “cogs”, “cogs-Plastic\_Price”, and “cogs-Wood\_Price shows the distribution and the variability of each cost. The recycled plastic shows higher and more variable costs. This suggests that recycled wood could be a more stable and profitable option.

The line plot shows that wood prices are much more stable with a slight downward trend, while the plastic prices are more volatile and fluctuate more. Even though plastic seems to be cheaper, it is much more volatile and can impact profit margins. This chart also suggests that wood may be the best option.

### **Dashboard description:**

From our interactive histogram we see that plastic and wood cogs tends to have negative values, while the cogs of products at the moment start at zero. And from the scatterplot we were able to see there is no distinctive relationship between price and salesdate.

LINK : <http://192.168.4.217:8501>