**Abstract:**

Customer segmentation is used by many companies to divide customers into separate groups when selling items. Our project aims at segmenting customers and predicting demand based on historical pricing data. We achieve this goal by using historical data from an anonymous store through K-Mean Clustering to divide customers into different segments. Additionally, we predict future demand using XGBoost regression accounting for factors such as inflation.

**Introduction and Question:**

Customer segmentation involves dividing customers into groups based on common characteristics. This helps in tailoring marketing strategies, improving customer service, and increasing sales. Additionally, consumer segmentation allows for companies to set prices at an optimal point for each consumer which allows for the consumer to buy a good at an ideal price for them while maximizing profits.

**Background:**

According to shopify.com, one of the largest e-commerce platforms, “Customer segmentation is the process of dividing a company’s customers into groups based on common characteristics so companies can market to each group effectively and appropriately.” (Shopify, 2023). Businesses typically segment customers based on age, gender, marital status, or location of where the customer lives. For example, many companies practice price discrimination. This involves which could run certain promotions on goods for certain segments of the population. One high-level example of price discrimination would be student pricing for sporting events. In this instance, students could be identified using machine learning based on search histories and from university lists and receive targeted advertising for the event with a promotion or other incentive to get the ideal customer to attend an event. For the purposes of our paper, we are segmenting customers in an ecommerce setting and predicting demand based on the customer’s characteristics and tastes.

**Objective:**

We intend on building a machine learning model which relies on K-Means clustering and XGBoost Regression. This will allow us to split the customer base into different segments based on company sales and profits. Additionally, we are predicting future demand for each segment using historical pricing data using machine learning.

**Methodology:**

We are using machine learning and data mining techniques to develop predictive models using historical data from an anonymous store. To segment customers, we are using K Means and hierarchical clustering. Finally, to predict future demand, we are using an XGBoost regression model. The findings will also provide insights into how different customer segments respond to price changes, aiding in more targeted and effective pricing strategies.

<https://www.shopify.com/blog/what-is-customer-segmentation#:~:text=Customer%20segmentation%20is%20the%20process,Industry>

**Features Used:**

Since most data used for predicting demand in a retail setting is proprietary, we are limited in the data we can use. In our instance, we used a dataset from an anonymous retailer with historical orders from different parts of the world.

**Customer Segmentation:**