**Laptop Price Modeling**

**Problem Statement**

The increasing complexity and power of laptops makes their pricing challenging. Without appropriate pricing, companies can make potential customers shy away or lose out on revenue. Understanding consumer desires and what factors may have the greatest impact on their purchasing decisions will allow a company to better price their products and learn what components to focus on. This knowledge is vital so that companies do not waste resources pursuing functionality consumers do not find useful or miss out on great pricing strategy.

I hope to create a machine learning model that can group laptops by their specifications and their price. As we create new laptops, we can then use these groups to help us place our laptop and the pricing that should go along with it.

**Data Wrangling**

The dataset used was scrapped by a user on Kaggle and contains approximately 900 rows and 23 columns. Keeping these rows intact and making sure that the data is usable is key to any good analysis. There are some columns that we may not find very useful in our analysis so some initial exploration will help us reduce these columns. There was also some missing data that I needed to look deeper into and will discuss in a later section.

Ultimately, I decided that the data should focus on the laptop specifications themselves so I dropped the ratings, star ratings, and review columns as these columns would not provide us with technical data. I also removed the columns for the old price and the discount amount as I am only focusing on the latest price. With that, we were able to begin exploring the data in earnest.

**Exploratory Data Analysis**

I started looking at features that may be useful for predicting pricing and the ways the different categories interact with each other. This helped me consider what other processing I wanted to do on the data and how I wanted to approach the modeling portion. First, I looked at the price and how pricing was distributed across all the laptops (**Figure 1**). I found the median was at around $825.42 with a few laptops reaching prices of over $4,000. I then followed that with a heatmap that helped look at whether there were any correlations for our numerical measurements with price in the data because that information would be the most helpful for deciding which features to prioritize during the process. Through that process, I found that there was some correlation between the RAM size, SSD size, and graphics card size (**Figure 2**) with price.

A picture containing chart

Description automatically generated

Figure 1. Median price of laptops

A picture containing chart

Description automatically generated

Figure 2 Heatmap of correlations for some features

Following these two analyses, I wanted to break down the information further and look at the categorical information that was collected. What I found was that for the most part, prices were distributed evenly across the board and not many features really stood out in terms of pricing. This led me to believe that it was a combination of features that led to certain prices for the laptops. The trend of laptops with larger storage space being more expensive was shown in the boxplots and was consistent with the positive correlation we saw in the heatmap. The most interesting boxplot was one made of the brands, and we saw a few brands really jump out. Namely, Alienware and Apple really stood out from other brands in terms of pricing (**Figure 3**) so knowing the branding of a laptop could really impact the pricing.

Chart, box and whisker chart

Description automatically generated

**In-Depth Processing**

**Modeling**

**Results and Discussion**