Conner Simmering
Tyler Rostenbach

Visualizing Amazon Bestseller Data
Final Project Proposal – INFO 4602

Introduction

Sales figures are some of the most important sets of data in any retail business. This is because investing in sales data analysis could lead to trend insights, future sales, and increased revenue. Our final project will consist of sales data analysis for the largest online retailers in the world, Amazon.com. Through computer visualization, we will reveal pertinent trends to Amazon's business, such as their best selling products, best selling categories, and what kinds of products are predicted to generate the most revenue based on the number of reviews.

Data

Our data consists of cross-category sales data extracted from Amazon.com. Typical sales data is unavailable to the public, but by using the Amazon API and an Amazon product research tool called JungleScout (https://www.junglescout.com/), we've scraped sales data from Amazon's top product categories. Product categories we're selected from https://www.amazon.com/Best-Sellers/zgbs and the top 20 listings we're scraped from each category. The final sales dataset is a CSV consisting of the product name, brand, price, category, rank, sales, revenue, number of reviews, rating, ASIN, and link for the top 20 products in Amazon's most popular categories. Sales and revenue data are predicted using estimations provided by JungleScout using their closed source algorithm.

Conner Simmering
Tyler Rostenbach

Analysis

Cross examining product categories in Amazon is something that's rarely done. This is because different categories have different markets, and the comparison is often useless from a marketing perspective. However, an interesting data point that's not available the public and not readily available to us is: what is the top selling, most valuable item on Amazon? In our visualization, this will be revealed. We will also compare review numbers, revenue per review, and profit margins to visualize the best categories that a vendor should pursue when selling products on Amazon.

Platform

We plan on creating D3 visualizations that will live on a bootstrap front-end. We're choosing D3 because we can bring the data to life using web standards and browser capabilities without tying ourselves to a proprietary framework. Additionally, D3 provides powerful visualization components and DOM manipulation procedures.

Evaluation

The goal of our visualizations is to gain insight, and in evaluating them, we will use our intuition to measure the insight gained. Additionally, we will cross exam our raw data in order to make sure we're getting sanitary results. For example, we can visualize a revenue per revue by doing mathematical operations in D3 and then checking for the same results in Excel.