

Project Proposal

1. Name of the project and team members

Project Name: Online Shopping Platform Product Consumption Comparison

Team Members: Yiming Xiong, Haosen Guo

2. What problem are you trying to solve?

This project focuses on comparing the same product—such as the GeForce RTX 4090—sold by different manufacturers and vendors on online shopping platforms. Prices, sales volume, and buyer ratings vary widely across sellers, making it difficult for consumers to identify the best purchasing option.

The goal of the project is to develop a tool that automatically collects all listings of a single product on Newegg, extracts key information including price, sales volume, and buyer ratings, computes a final comparison score for each seller, and presents the results in a clear and standardized format.

By centralizing and standardizing information from different Newegg sellers, the system can help users make more informed and transparent purchasing decisions.

3. How will you collect data and from where?

To collect the required information, the project will use Python web-scraping tools such as BeautifulSoup, requests, and other related libraries. These will extract and filter data directly from Newegg pages, including brand name, price, sales information (when available), and buyer rating.

4. What analysis will you do and what visualizations will you create?

For analysis, I will extract RTX 4090 graphics card listings from Newegg using web scraping and clean the collected data to obtain each product's title, price, and product rating score. The analysis focuses on determining the price, sales volume, and ratings of the same model from different sellers, such as ASUS, MSI, and Nvidia.

To present the results intuitively, the project will generate a range of visualizations. Bar charts will be used to compare price differences, sales volume, and rating levels across brands. A pie chart can be used to illustrate the market share of each brand based on sales volume, helping users understand the relative popularity of different manufacturers. Additional visualizations, such as stacked bar charts or rating-distribution plots, may also be included to show how each brand performs across multiple attributes. Finally, a separate bar chart will present the combined evaluation score, making it easy to identify the best overall value.