

# Advance Database Technologies

## Final Project Part 1 - Project Proposal

### Spring 2024

#### Project Title:

**Amazon: A2Z AmazeComparator**

#### Team Names:

- Kabir Chaturvedi ([kschatur@iu.edu](mailto:kschatur@iu.edu)) (Team Leader)
- Dhyey Joshi ([dhyejoshi@iu.edu](mailto:dhyejoshi@iu.edu))
- Vedika Halwasiya ([vhalwasi@iu.edu](mailto:vhalwasi@iu.edu))

#### Project Summary:

Our website, "A2Z AmazeComparator," leverages a comprehensive dataset that encompasses detailed information on products listed on Amazon, including product ID, name, category, pricing, ratings, and product descriptions. Designed to empower consumers and online shoppers, "A2Z AmazeComparator" aims to revolutionize the way users engage with e-commerce by enabling product comparisons, facilitating the addition of new products, and offering in-depth insights into consumer review analytics.

This platform is made for anyone looking to shop smarter online. Its easy-to-use interface offers detailed product analyses, insights into consumer reviews and sentiments, as well as information on discounts and price changes. It's the perfect tool for making well-informed purchase decisions.

#### Project Objectives:

The objective of the project "**A2Z AmazeComparator**", is designed to be a go-to online tool for shoppers looking *specifically at Amazon products*. It aims to simplify the decision-making process by *letting users compare different products side by side*. By focusing on key features, prices, user ratings, and review sentiments, it provides a comprehensive view that helps shoppers choose the best product for their needs. The tool is especially handy for those who find themselves overwhelmed by the vast options on Amazon or for *anyone looking to get the best deal without spending hours researching*. A2Z AmazeComparator stands out by offering an *easy-to-use platform* where all the important product information is available at a glance, *making shopping on Amazon a breeze*.

## Technical Stack:

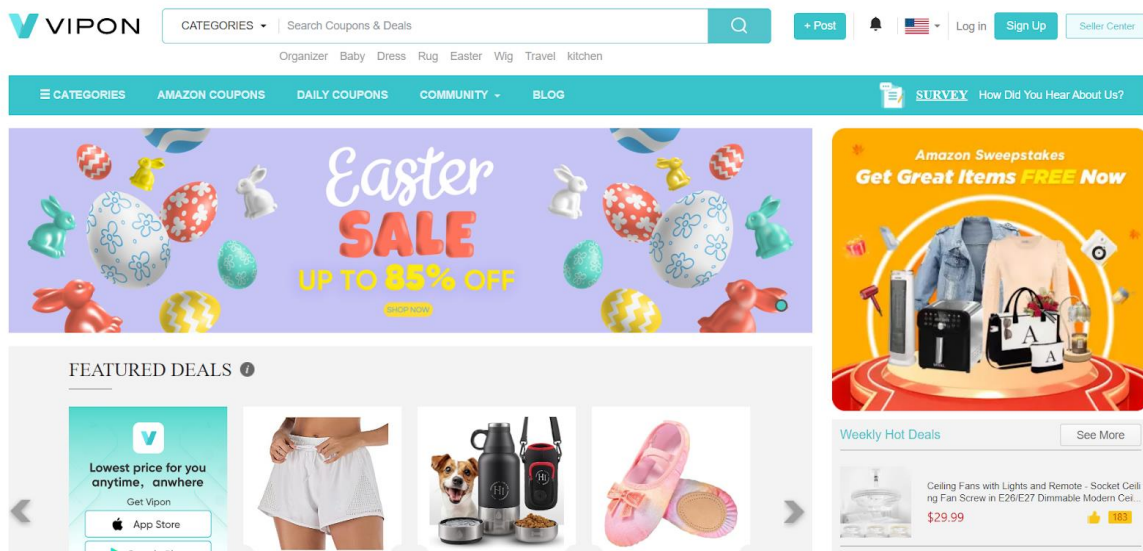
1. Frontend Development: Bootstrap, HTML, JavaScript, React.js, CSS/SCSS.
2. Backend Development: Python, Flask or Django, Pandas and Numpy.
3. Natural Language Processing (NLP): NLTK or SpaCy.
4. Data Visualization and Dashboard: Dash or Plotly.
5. Database: MySQL, SQLite.
6. Deployment and Version Control: Git / Github, AWS or any deployment platforms.
7. Data Processing Tools: Jupyter Notebook.

## Project Description:

### Usefulness:

This website makes it easy to choose between products by letting you compare them quickly, saves you time, helps you understand what other buyers think, lets you find deals that suit your needs, is great for *people who likes to compare product before making a deal and want to know all the details, and overall, helps you shop smarter and with more confidence.*

There are very few websites that offer similar features, but they are not specifically for Amazon products. Instead, they are based on trader websites that allow users to search for products and buy them at lower prices, such as Vipon, found here: <https://www.myvipon.com/>

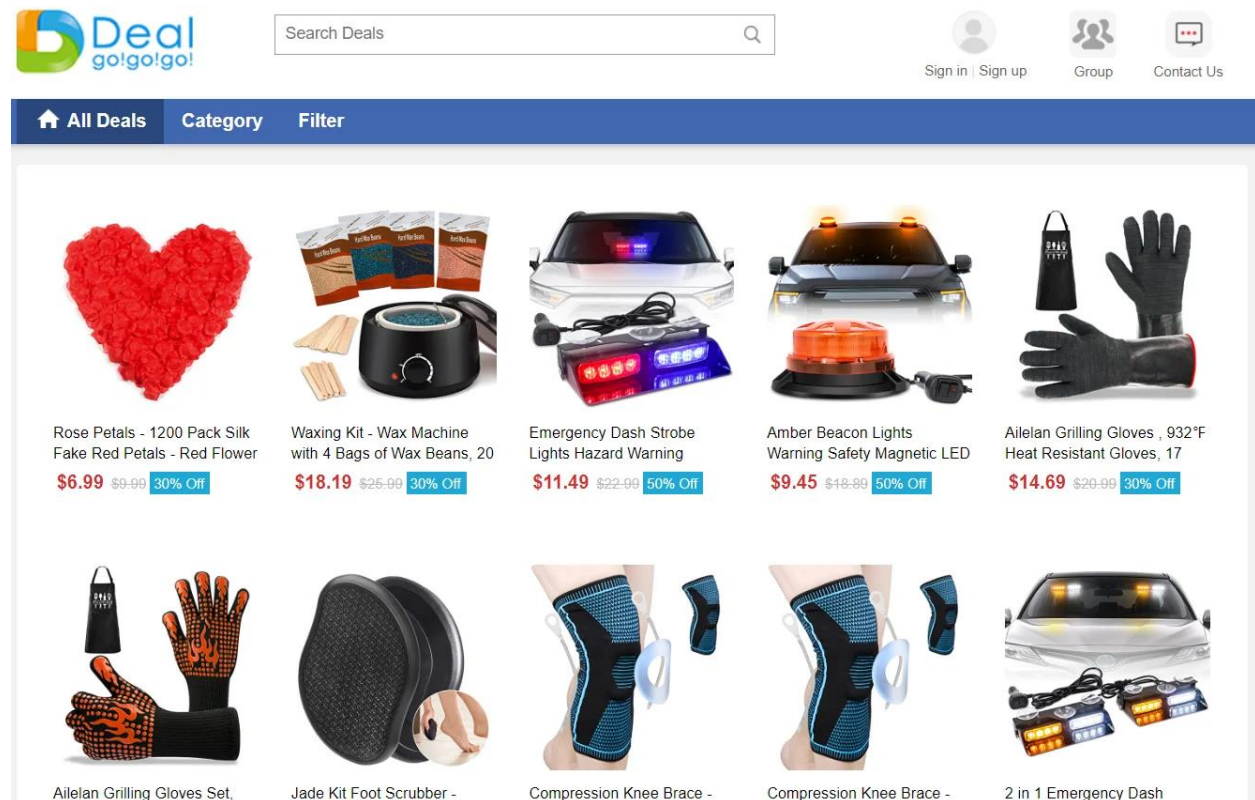


The website, named "Vipon," is one of the oldest and most popular sites for finding Amazon deals, offering some of the best sales available. However, the existing website we discovered online only

showcases the best deals, lacking features for comparison, review sentiment analysis, or dashboards for visualizing trending products with maximum discounts. Our website aims to fill this gap by providing all these exclusive features to users, making it an easy-to-use comparison tool.

Another website was developed by DealGoGoGo

<https://www.dealgogogo.com/>



The website named "Deal Go Go Go" hosts a vast array of product categories but is considered unsafe to use, as many of the products displayed are fake and not accurately represented. This website offers an additional filtering feature that allows buyers to easily navigate through the categories to find what they want. However, this website has a non-user-friendly UI, and its server stability is questionable. The features we will provide on our website are far more user-friendly and advanced than those of any existing websites, including "Deal Go Go Go."

The targeted stakeholders are smart online shoppers who want a simple way to compare Amazon products. Our features, like side-by-side comparisons and review analysis, help users quickly understand which product is the best for them. This means less guesswork and more confident shopping decisions.

[illegible]



## Other Amazon Sales Dataset on kaggle: Structured / Normalized / Require slight Pre-Processing

<https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data/data>

Index	Order ID	Date	Status	Fulfillment	Sales Channel	Service	Style	SKU	Category	Size	ASIN	Courier	Str	Qty	Currency	Amount	Ship City	Ship State	Ship Postal Code	Ship Count	Promotion	B2B
2	0 405-8078784-5731	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	SET389	SET389-KR Set	S		B09KXVB07Z			0	INR	647.62	MUMBAI	MAHARAS	400081	IN		FALSE
3	1 171-9198151-1101	4/30/2022	Shipped	Merchant	Amazon.in	Standard	JNE3791	JNE3791-K kurta	3XL		B09K3WF1	Shipped		1	INR	406	BENGALURU	KARNATAK	560085	IN	Amazon Pl	FALSE
4	2 404-0687676-7272	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3371	JNE3371-K kurta	XL		B07WV4V9	Shipped		1	INR	329	NAVIMUM	MAHARAS	410210	IN	IN Core Fr	TRUE
5	3 403-9615377-8133	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	J0341	J0341-DR- Western D	L		B099NRC77B			0	INR	753.33	PUDUCHE	PUDUCHE	605008	IN		FALSE
6	4 407-1069790-724C	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3671	JNE3671-T Top	3XL		B098714B	Shipped		1	INR	574	CHENNAI	TAMIL NAD	600073	IN		FALSE
7	5 404-1499804-457E	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	SET264	SET264-KR Set	XL		B09YK7XD	Shipped		1	INR	824	GHANASA	UTTAR PR	201102	IN	IN Core Fr	FALSE
8	6 408-5748498-6855	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	J0095	J0095-SET Set	L		B09C9H4K	Shipped		1	INR	653	CHANDIG	CHANDIG	160036	IN	IN Core Fr	FALSE
9	7 406-7807733-3785	4/30/2022	Shipped	Merchant	Amazon.in	Standard	JNE3405	JNE3405-K kurta	S		B081W4X4	Shipped		1	INR	399	HYDERAB	TELANGAN	500032	IN	Amazon Pl	FALSE
10	8 407-5443024-5923	4/30/2022	Cancelled	Amazon	Amazon.in	Expedited	SET200	SET200-KR Set	3XL		B08151Z7	Cancelled		0			HYDERAB	TELANGAN	500006	IN	IN Core Fr	FALSE
11	9 402-4393761-8311	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3461	JNE3461-K kurta	XXL		B08832F5	Shipped		1	INR	365	CHENNAI	TAMIL NAD	600041	IN		FALSE
12	10 407-5633625-697C	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3160	JNE3160-K kurta	S		B07K3YQ1	Shipped		1	INR	685	CHENNAI	TAMIL NAD	600073	IN		FALSE
13	11 171-4638481-6326	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3500	JNE3500-K kurta	XS		B0981170	Shipped		1	INR	364	NOIDA	UTTAR PR	201303	IN		FALSE
14	12 405-5513694-814E	4/30/2022	Shipped	Merchant	Amazon.in	Standard	JNE3405	JNE3405-K kurta	XS		B091XCMY	Shipped		1	INR	399	Amravati	MAHARAS	444606	IN	Amazon Pl	FALSE
15	13 408-7955686-3083	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	SET182	SET182-KR Set	S		B095H594	Shipped		1	INR	657	MUMBAI	MAHARAS	400053	IN		FALSE
16	14 408-1298370-192C	4/30/2022	Shipped	Merchant	Amazon.in	Standard	J0351	J0351-SET Set	L		B09C5S0Y	Shipped		1	INR	771	MUMBAI	MAHARAS	400053	IN	Amazon Pl	FALSE
17	15 403-4965861-952C	4/30/2022	Shipped	Merchant	Amazon.in	Standard	PINE3368	PINE3368-kurta	6XL		B09P9Y95	Shipped		1	INR	544	GUNTAKA	ANDHRA P	515801	IN	Amazon Pl	FALSE
18	16 406-9379318-6555	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3721	JNE3721-K kurta	XXL		B099FC76	Shipped		1	INR	329	JAIPUR	RAJASTHA	302020	IN	IN Core Fr	FALSE
19	17 405-9013803-8005	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3405	JNE3405-K kurta	XL		B081W76C	Shipped		1	INR	399	NEW DELH	DELHI	110074	IN		FALSE
20	18 402-4030358-5833	4/30/2022	Shipped	Merchant	Amazon.in	Standard	JNE3697	JNE3697-K kurta	XXL		B098133P	Shipped		1	INR	458	Gurgaon	HARYANA	122004	IN	Amazon Pl	FALSE
21	19 405-5957858-1051	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	SET254	SET254-KR Set	XS		B0983DDF	Shipped		1	INR	886	BENGALUR	KARNATAK	560017	IN		FALSE
22	20 405-0607769-071E	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3795	JNE3795-K kurta	3XL		B09HMXV4	Shipped		1	INR	517	TIRUCHIR	TAMIL NAD	620018	IN	IN Core Fr	FALSE
23	21 404-8494550-586C	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	SET345	SET345-KR Set	M		B08KXV4B	Shipped		1	INR	666	BENGALUR	KARNATAK	560040	IN	IN Core Fr	FALSE
24	22 171-1305077-281C	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	JNE3373	JNE3373-K kurta	L		B082W7G7	Shipped		1	INR	376	HYDERAB	TELANGAN	500072	IN		FALSE
25	23 404-6019946-2905	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	SET291	SET291-KR Set	M		B099NK5YG			0	INR	570.48	pune	MAHARAS	411044	IN		FALSE
26	24 402-3384087-4003	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	MEN5002	MEN5002-kurta	L		B08YYDYN	Shipped		1	INR	499	TEZPUR	ASSAM	784001	IN	IN Core Fr	FALSE
27	25 405-8191138-517E	4/30/2022	Shipped	Merchant	Amazon.in	Standard	NW030	NW030-TF Set	XS		B08G2RQ2	Shipped		1	INR	582	RANCHI	JHARKHAN	834002	IN	Amazon Pl	FALSE
28	26 403-9230474-9657	4/30/2022	Shipped	Merchant	Amazon.in	Standard	JNE3415	JNE3415-K kurta	3XL		B082W8JX	Shipped		1	INR	299	BILASPUR	CHHATTIS	495001	IN		FALSE
29	27 408-3484251-6901	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	PINE2199	PINE2199-kurta	4XL		B08LD2W5	Shipped		1	INR	459	PUNE	MAHARAS	411052	IN		FALSE

The dataset we're using stands out from others on the market because of the following reasons:

- **Our dataset is primarily focused on products listed on Amazon**, including details such as product ID, name, category, pricing information, ratings, rating count, and product descriptions. **It has more unnormalized structure, where information about products is combined into single rows, potentially leading to data redundancy if the same product appears multiple times.** On the other hand, the second dataset from Kaggle, "Amazon Sale Report," is a **sales report containing details** such as order ID, purchase date, order status, fulfillment, sales channel, SKU, product category, amount, shipping information, and B2B indicators. This **dataset seems to have a more normalized structure** aimed at tracking sales transactions, which could include multiple products per order.
- Both datasets include categorical and numerical data, but they are applied to different types of analysis. Our dataset contains more descriptive data about products, while the other one (Amazon Sales Report) is transactional and includes dates, numerical identifiers, and financial figures. The **structure of our dataset**, focusing on product details and reviews, lends itself more to product analysis but **requires more preprocessing to handle redundant data effectively if the same product is listed multiple times.**

### ***Target User Group:***

The targeted stakeholders are ***smart online shoppers*** who want a simple way to compare Amazon products. Our features, like side-by-side comparisons and review analysis, help users quickly understand which product is the best for them. This means less guesswork and more confident shopping decisions.

### **Dataset:**

For this project we have chosen to utilize a comprehensive dataset from Kaggle titled “[Amazon Sales Data](#)”. This dataset includes detailed record of product listings and customer reviews across various categories on Amazon, featuring 1465 entries with 16 distinct attributes such as product name, category, price, discount rate, rating, and user review.

This dataset enables us to extract and compare key product features, pricing, ratings, and sentiment analysis of reviews – essential components for our “A2Z AmazeComparator”. It is specifically curated to assist users in navigating the extensive product options on Amazon, simplifying the decision-making process by providing a side-by-side comparison of products.

This aligns perfectly with our project’s aim to enhance shopping experience by offering a user-friendly platform where important product information is readily accessible, thereby saving time and effort in finding the best deals on Amazon.

### **Sharing:**

We have set up a GitHub repository to store the code for the project. We also use Microsoft Teams to chat and host team meetings.

GitHub repository: [Click Here!](#)

### **Group Contribution:**

<i>Name</i>	<i>Tasks</i>	<i>Average Time Spent (per milestone)</i>
All three team members	Brainstorming of project ideas	3 hours

All three team members	Researching existing applications & finding a right database	More than 2 days
All three team members	Inspect database and what further processing is needed. Planed how we will use the database to create an end-to-end application	4 hours
All three team members	Researched tools and technologies required to build the application such as React.js / Flask/ Python libraries, such as Dash etc.	2.5 hours
All three team members	Project proposal draft collaboration	5 hours