BAX-422 Final Project

"*A Study of Amazon's Product Features & Influence*"

Author:

Kexin Fu, Parth Patel, Ruihan Zhou, Yushan Liu

Mar.20th.22

Executive Summary

Amazon is among world-leading retailing services, which holds nearly 50% market share in North America. Although Amazon owns the majority of the market, its major competitor Walmart is growing up 40% in online sales annually to catch up. In order to maintain its advantage in online retailing, Amazon still needs to improve its service to boost its sales to not be overtaken by its domestic and international competitors.
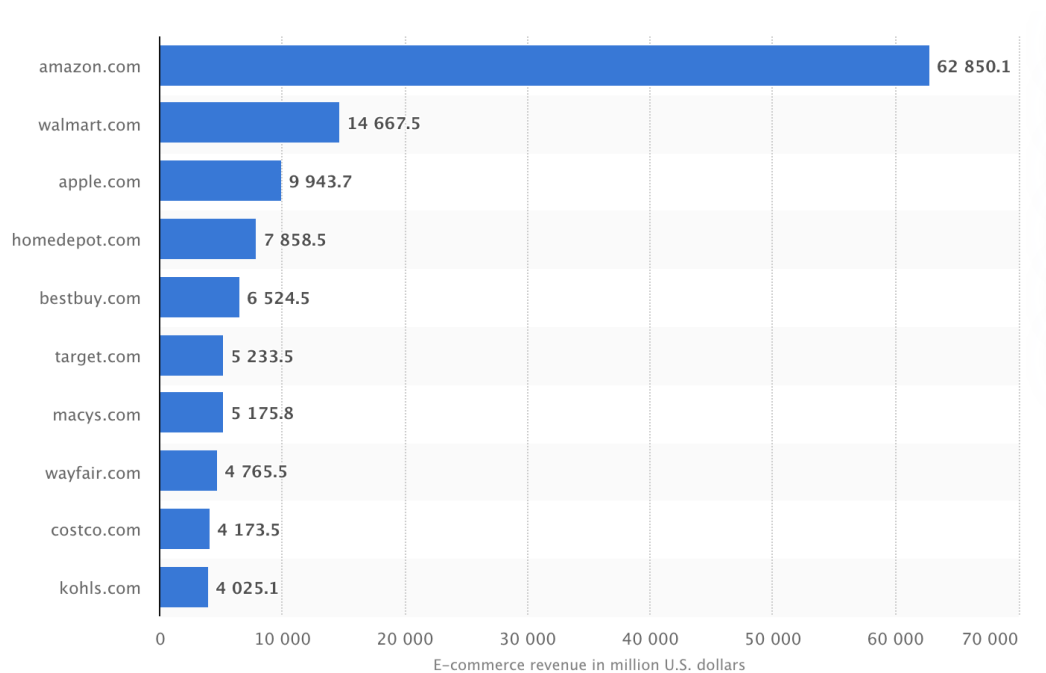
Our group wants to explore how Amazon's coupon, free delivery, prime, etc. contribute to the product sales. In this way, Amazon will understand better which aspects they need to strengthen and to pump up their sales. We decided to crawl the product information of the "office chair" on Amazon as an analysis example. Crawled content includes: product description, ranking, rating, coupon, free delivery or not, prime or not and item URL.

We used Selenium and BeautifulSoup to crawl all the top 7 pages on Amazon with the keyword office chair. A total of 383 pieces of product information are crawled. After this, we integrated all the information and loaded it into the MongoDB database. Those information could be utilized for subsequent data analysis, like clustering, regression and NLP, which will play an essential role in finding important features to increase Amazon online sales.

1.Background, Context, and Domain Knowledge

Amazon.com, Inc. is a technology company which operates in 58[1] countries up to the year of 2018. It consists of three main business segments: 1. North America, 2. International, and 3. Amazon Web Services(AWS). The first two segments provide services in retail shopping of customer products as well as subscriptions, while the third segment focuses more on providing the service of computing, storing and database. Amazon also manufactures electronic devices such as Kindle (electronic reading equipment) and Echo & Alexa devices (smart speaker with a voice controlled intelligent service).

Amazon's online store has other competitors demonstrated in graph (1) below.

	E-commerce revenue in million U.S. dollars
amazon.com	62 850.1
walmart.com	14 667.5
apple.com	9 943.7
homedepot.com	7 858.5
bestbuy.com	6 524.5
target.com	5 233.5
macys.com	5 175.8
wayfair.com	4 765.5
costco.com	4 173.5
kohls.com	4 025.1

Graph (1)[2]

Here, Amazon's biggest competitor in North America Walmart is growing at over 40% per year for online sale revenue and is predicted to keep growing in the future.

Since Amazon has the world-leading retailing service, we are intrigued to dive deep into how its website designs the product features and how information is displayed to encourage customer purchasing. Amazon has several marketing strategies on retail products such as personalized time-limited discount coupons to apply on certain products, free delivery and free one day delivery on eligible products, and free delivery for all products with 'prime' tag. Amazon also has enabled customer ratings and reviews to reflect information about the products. The paper decides to scrap all features mentioned above in order for future study their influence on product sales and rating.

2.Data Sources, Description of Web-Scraping Routines, and Database Design Choice

2.1 Data Source

Our team chose the Amazon website (https://www.amazon.com/) as the data source, and scraped the general search result for "office chair". The code can also be applied to all other search terms as well if we wish to study how the features affect any other products. Currently, Amazon has modified their search result display to only display 7 pages for general search terms like "office chair" or "lipstick" that actually yield millions of results. Our team figured out that only by applying a filter or inputting a very specific keyword can Amazon display more results. However, we want the code to be widely applicable without filtering or specifying too much in advance, so we stick with scraping 7 pages of "office chair", and the total result dataset contains 383 results.

2.2 Web-scraping routines

The code mainly uses selenium, BeautifulSoup and ChromeDriverManager. It starts with cleaning and generating the URL, and then we partition and extract each item as a record by searching for their "div" tag using BeautifulSoup. Then all the scraping will be done on a "item" or "product" basis, so it will not incur difficulty handling missing values and staying organized as scraping on the whole webpage will. After that, we choose 9 most important fields to scrape, which are product description, current price, original price, coupon, delivery, rating, review count, prime tag and URL. We use null values if any tags do not exist for a product. Finally, we combine them into a whole data frame as the final output.

2.3 Database Design Choice

We choose to load our dataset into MongoDB database, which is a NoSQL database that is flexible and does not require a rigid schema. All the 9 fields loaded into the database are named as: *Description, Price, Price_orig, Coupon, Delivery, Rating, ReviewCount, Prime,* and *URL.* If we view each product in JSON format, it will be presented as graph (2) below, for which we are able to see the features displayed individually for each of the 383 results we scraped. It is both clean and convenient if we wish to look up certain information for each individual item.

```
1  ▾{
2       "_id" : ObjectId("6237a5f28f0f0fb0be1c1a65"),
3       "Description" : "Amazon Basics Classic Puresoft Padded Mid-Back Office Computer Desk Chair with Armrest - Cream",
4       "Price" : "$141.36",
5       "Price_orig" : "",
6       "Coupon" : "",
7       "Delivery" : " Get it as soon as Fri, Mar 25FREE Shipping by Amazon",
8       "Rating" : "4.3 out of 5 stars",
9       "ReviewCount" : "19,320",
10      "Prime" : NumberInt(1),
11      "Url" : "https://www.amazon.com/gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pg1_1?ie=UTF8&adId=A005576830CU4
12  }
13  ▾{
14      "_id" : ObjectId("6237a5f28f0f0fb0be1c1a66"),
15      "Description" : "Office Chair Ergonomic Cheap Desk Chair Mesh Computer Chair Lumbar Support Modern Executive Adjustable
16      "Price" : "$52.99",
17      "Price_orig" : "",
18      "Coupon" : "",
19      "Delivery" : " Get it as soon as Wed, Mar 23FREE Shipping by Amazon",
20      "Rating" : "4.3 out of 5 stars",
21      "ReviewCount" : "41,247",
22      "Prime" : NumberInt(1),
23      "Url" : "https://www.amazon.com/gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pg1_1?ie=UTF8&adId=A08356713289C
24  }
```

Graph (2)

If presented in table format, the results will be displayed as graph (3). In table format, the information is displayed in a more holistic manner when being compared to the JSON format. We can easily see the range and general characteristics of all products we scraped in one table.

amazon_office_chairs > Description

_id	Description	Price	Price_orig	Coupon	Delivery	Rating	ReviewCount	Prime	Url
6237a5f28f0f0f...	Amazon Basics...	$141.36			Get it as soon …	4.3 out of 5 sta…	19,320	1	https://www.a...
6237a5f28f0f0f...	Office Chair Er...	$52.99			Get it as soon …	4.3 out of 5 sta…	41,247	1	https://www.a...
6237a5f28f0f0f...	Ticova Ergono...	$259.99	$299.99$299.99	Save $60.00 w...	Get it as soon …	4.3 out of 5 sta…	2,330	1	https://www.a...
6237a5f28f0f0f...	FlexiSpot OC3...	$199.99			Get it Thu, Mar...	4.5 out of 5 sta…	643	0	https://www.a...
6237a5f28f0f0f...	NEO CHAIR Of...	$68.98	$104.98$104.98		Get it Wed, Ma...	4.2 out of 5 sta…	12,563	0	https://www.a...
6237a5f28f0f0f...	Etasker Ergono...	$122.81	$169.99$169.99		Get it as soon …	4.8 out of 5 sta…	29	1	https://www.a...
6237a5f28f0f0f...	Office Chair, Er...	$85.99			FREE delivery	4.2 out of 5 sta…	851	1	https://www.a...
6237a5f28f0f0f...	Executive Offic...	$159.99		Save 5% with …	Get it as soon …	4.4 out of 5 sta…	308	1	https://www.a...
6237a5f28f0f0f...	Steelcase Amia...	$849.00			Get it Wed, Ma...	4.5 out of 5 sta…	72	0	https://www.a...
6237a5f28f0f0f...	Furmax Office ...	$51.99			FREE delivery	4.3 out of 5 sta…	4,555	1	https://www.a...
6237a5f28f0f0f...	Sytas Office Ch...	$109.99	$149.99$149.99		Get it as soon …	4.4 out of 5 sta…	308	1	https://www.a...
6237a5f28f0f0f...	ERGOUP Ergo...	$989.99		Save 5% with …	Get it as soon …	4.4 out of 5 sta…	6	1	https://www.a...
6237a5f28f0f0f...	Office Chair Co...	$84.99			Get it Mon, Ma...	4.3 out of 5 sta…	4,372	0	https://www.a...
6237a5f28f0f0f...	Steelcase Serie...	$396.33			Get it as soon …	4.4 out of 5 sta…	1,187	1	https://www.a...
6237a5f28f0f0f...	X Chair X1 Tas...	$699.99			Get it as soon …	4.0 out of 5 sta…	11	1	https://www.a...
6237a5f28f0f0f...	High Back Ga...	$69.99	$84.99$84.99	Save 5% with …	Get it as soon …	4.5 out of 5 sta…	12,210	0	https://www.a...
6237a5f28f0f0f...	Amazon Basics...	$107.87			Get it as soon …	4.2 out of 5 sta…	1,573	1	https://www.a...
6237a5f28f0f0f...	BestOffice PC ...	$59.99	$72.88$72.88	Save 5% with …	Get it Fri, Mar ...	4.3 out of 5 sta…	30,327	0	https://www.a...
6237a5f28f0f0f...	OFIKA Home ...	$126.88		Save 5% with …	Get it Wed, Ma...	3.9 out of 5 sta…	2,390	0	https://www.a...
6237a5f28f0f0f...	OFIKA Office C...	$129.90			Get it Wed, Ma...	3.8 out of 5 sta…	42	0	https://www.a...
6237a5f28f0f0f...	HON Ignition ...	$403.99			FREE Shippin...	4.3 out of 5 sta…	1,240	1	https://www.a...
6237a5f28f0f0f...	Amazon Basics...	$169.99			Get it as soon …	4.4 out of 5 sta…	20,287	1	https://www.a...
6237a5f28f0f0f...	Ergonomic Off...	$189.99		Save $30.00 w...	Get it as soon …	4.4 out of 5 sta…	301	1	https://www.a...
6237a5f28f0f0f...	Varier Variable...	$599.00			Get it as soon …	3.3 out of 5 sta…	3	1	https://www.a...
6237a5f28f0f0f...	HOMREST Erg...	$229.99		Save $12.00 w...	Get it Thu, Mar...	4.2 out of 5 sta…	70	0	https://www.a...
6237a5f28f0f0f...	SAMOFU Ergo...	$439.99		Save $40.00 w...	Get it as soon …	4.5 out of 5 sta…	77	1	https://www.a...
6237a5f28f0f0f...	Office Chair, H...	$167.28			Get it as soon …	4.4 out of 5 sta…	14	1	https://www.a...
6237a5f28f0f0f...	Gabrylly Ergon...	$289.50	$329.50$329.50		Get it as soon …	4.4 out of 5 sta…	6,792	1	https://www.a...
6237a5f28f0f0f...	Serta Ergonom...	$313.16			Get it as soon …	4.0 out of 5 sta…	1,814	1	https://www.a...
6237a5f28f0f0f...	Lucklife Ergon...	$99.99		Save $40.00 w...	Get it as soon …	3.7 out of 5 sta…	168	1	https://www.a...

Graph (3)

3. Advantages of Database Implementation and Business Value Created

3.1 Database Implementation

The database implementation we have chosen (MongoDB) allows for easier access and retrieval of data and is fully accessible on the cloud. This will help for scalability purposes and information sharing across geographies. Additionally MongoDB is more cost effective than competitor solutions as it allows users to specify the instance size they currently need before database creation. This eliminates the need for excessively large databases that would only increase costs. Additionally, MongoDB allows users to increase the size of their database when the need arises, allowing them to scale at the appropriate time. Finally, MongoDB has powerful in-built querying capabilities, allowing users to filter, aggregate or combine data without the need for complex joins and other commands that may be too technical for some users.

3.2 Business Value Created by Database Choice

In this scenario, we are looking at how different aspects of an office chair product listing on Amazon contributes to its price and rating. However, this can be applied to any number of product types and listings. This type of database implementation can be useful for many independent sellers and small businesses on Amazon, who have a small product offering but do not have the ability to run analytics on the performance of their listings. With their products stored on MongoDB, these sellers are able to easily access their data on their local systems. Additionally, they will be able to analyze what aspects of their listing are the most effective in driving sales and improving their rating. Finally, due to the cost effective nature of MongoDB, these sellers will be able to operate nimbly, without too much additional overhead. This is

crucial, as independent sellers on Amazon have more constrained budgets than the large corporations that also list products on the website.

4. Summary

In a nutshell, we devised a code which can scrape Amazon website's general search results using tools like Selenium and BeautifulSoup. Office chair is our interested industry in this case. We gathered relevant data including their product description, current price, original price, coupon, delivery, rating, review count, prime tag and URL. After that, we loaded the data into the MongoDB database which allows flexibility and scalability in business scenarios. On the one hand, with the information we will be able to find out general key factors influencing sales and review numbers using methods like regression, clustering and statistical tests. The analysis will shed light on how an office chair product could thrive in the immense e-commerce industry. On the other hand, the dataset itself can provide individual sellers with an opportunity to access their data and competitors' data in their local systems, and seek solutions on how to improve their own product attractiveness.