Data cleaning & Pre-Processing (DCPP) Group Assignment

Domain: Retail/Consumer

Consumer Research: Gauge customers' view and competitiveness of graphic card category of a retail website

Group Team Members	PGID
Rahul Arora	12120036
Deeksha Sureka	12120094
Deepak Gupta	12120085
Karan Tyagi	12120012

INDEX

Executive Summary

Problem Statement:

Understanding customer behaviour and product competitiveness are an important aspect of marketing and product development functions for any organization., The problem statement considered here pertains to a leading GPU manufacturer, who wants to understand its customers' perspective as well competitive health of its graphic card product category using one of the most popular retail websites.

The goal is to build a dataset that provides insights about graphic card to the GPU manufacture's product teams and stakeholders

(Disclaimer: The working group wants to do this project as a concept testing for different use cases to maximize learning vs creating a vertical or horizontally long data set for one use case.

Proposed Solution and its workflow:

The intent of the solution developed is to explore varied data collection techniques to gather the data and conduct cleaning pre-processing activities to build a pipeline that is ready to be consumed for use by different teams to generate insights. Some of the sub-objectives are to:

- Explore various retail websites and understand complexities and possibilities of scraping the data
- Select the most impactful retail website (we have chosen AMAZON.in website, being the global and most popular retailer, to scrape)
- Define the list of attributes and product information to be extracted (Ex. Product Description, Price, etc.)
- Pre-process and clean the extracted data for better comprehension 1. https://realpython.com/beautiful-soup-web-scraper-python/
- 2. https://www.geeksforgeeks.org/difference-between-find-and-find_all-in-beautifulsoup-python/
- Use transformation techniques to create new attributes or dimension to increase usability
- Prepare the JSON file presented the scraped data in a human or machine-readable form

Challenges:

- Learning the web scraping from scratch
- Locating the required data on the web page
- No data availability in html tags
- Creating a loop to run the entire page data to scrape all products

The Chosen domain and Seed sources

As a working group we wanted to choose a domain that is relatable to our day to day lives and also provides maximum learning opportunities for individual and group learning.

We believe retail and consumer market offers products for end users

1. We use the retails apps or websites in our day to day lives to manage our living needs
2. Retail companies collect the most data about it customers so the richness in retail data is very high compared to other domains

Hence, Retail was found most relatable and actionable domain to choose for this project and we chose the biggest retail company AMAZON, being the global and known retailer in the world.

Seed URL and other URLs:
www.amazon.in
['https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=1&crid=1VR5KKO7J5XDM&qid=1652272954&rnid=3576079031&ref=sr_pg_1'
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=2&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_2',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=3&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_3',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=4&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_4',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=5&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_5',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=6&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_6',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=7&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_7',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=8&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_8',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=9&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_9',
'https://www.amazon.in/s?k=graphics+card&i=computers&rh=n%3A1375344031%2Cn%3A1375354031&dc&page=10&crid=1VR5KKO7J5XDM&qid=1652683804&rnid=3576079031&ref=sr_pg_10',]

The structured and unstructured fields from the chosen domain/internal sources

Source/Data Type	Data Attribute
Structured	Product Category – Graphic Card
Structured	Product SKU
Structured	Product Description
Structured	Available Price
Structured	Original Price
Structured (Calculated)	Discount Rate
Structured	Detail URL
Structured	Stock Tag
Unstructured	Customer Rating
Unstructured	Customer Review Title
Unstructured	No. of customer review
Unstructured	Customer review Description

Download/Scrape/collect data from all the sources

With all the decisions about domain, website and selection of data attributes, the next step was to extract the required data by crawling the amazon website from the seed links shared above (in the table)

In general, the simplest way to crawl the data is using website APIs but websites do not provide public APIs due to technical reasons and implications. So, we also didn't find any public API for amazon and used alternate ways of and scraping the data. The approach and tools are used are:

- Browsing the website to gather all the links for graphic cards from seed URL (https://www.amazon.in/)
- Used python URL handling library 'Urllib3' to fetch the URLs
- Used python web scraping library 'Beautiful soup' to extract the data
- Amazon website data structure is very complicated to scrape, so there were a lot of difficulties to locate the required data on the webpage.
- The html tags did not have data in them.
- Created a loop to run on the entire page data to scrape all products together

Using these techniques and approach, the amazon data set was seamlessly extracted by addressing the challenges at each stage.

Convert data from original sources (Webpages, pdf files, CSV files, …) to structured data fields

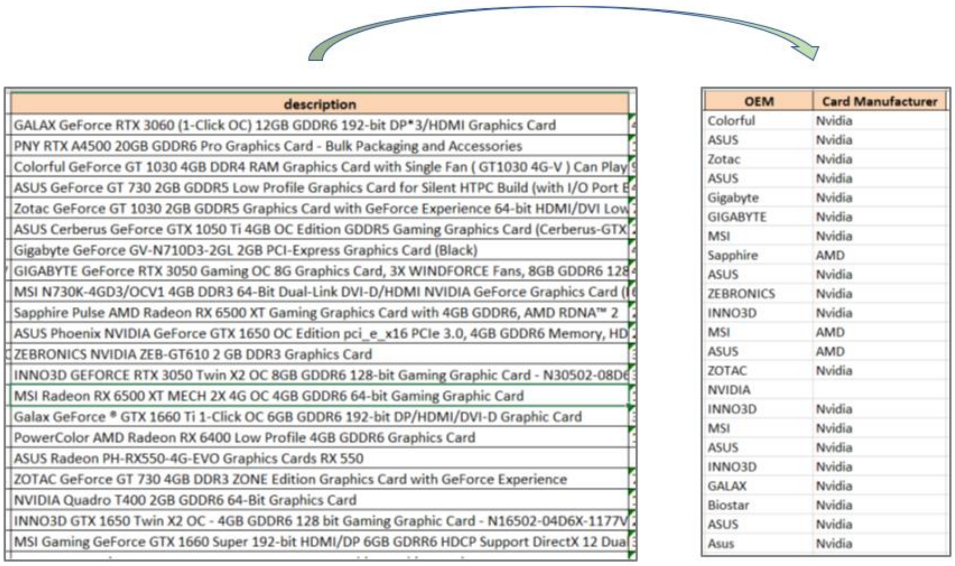
In this step, we applied the data formatting, validation methods, created new attributes and and calculated fields make the extracted data more readable and comprehensive:

- Formatting the data into a structured data frame
- Importing data from excel (pd.read.excel)
- Validate the attributes:
    - The number of columns extracted in the excel (head)
    - Info about data (Data type, Null – Non-Null values)

Pre-processing/Data cleaning as needed:

1. **Creating new structured data: Extract card manufacturer and OEM information from ‘description’ attribute to create separate attribute**

   (Screen shot is not showing apple to apple comparison but emphasize upon the workflow – input and output example)

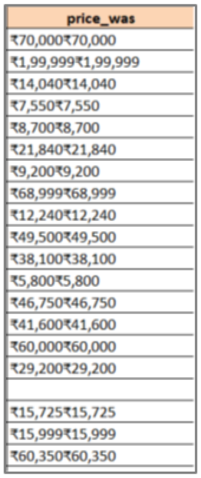


description
GALAX GeForce RTX 3060 (1-Click OC) 12GB GDDR6 192-bit DP*3/HDMI Graphics Card
PNY RTX A4500 20GB GDDR6 Pro Graphics Card - Bulk Packaging and Accessories
Colorful GeForce GT 1030 4GB DDR4 RAM Graphics Card with Single Fan (GT1030 4G-V) Can Play
ASUS GeForce GT 730 2GB GDDR5 Low Profile Graphics Card for Silent HTPC Build (with I/O Port E
Zotac GeForce GT 1030 2GB GDDR5 Graphics Card with GeForce Experience 64-bit HDMI/DVI Low
ASUS Cerberus GeForce GTX 1050 Ti 4GB OC Edition GDDR5 Gaming Graphics Card (Cerberus-GTX
Gigabyte GeForce GV-N710D3-2GL 2GB PCI-Express Graphics Card (Black)
GIGABYTE GeForce RTX 3050 Gaming OC 8G Graphics Card, 3X WINDFORCE Fans, 8GB GDDR6 128
MSI N730K-4GD3/OCV1 4GB DDR3 64-Bit Dual-Link DVI-D/HDMI NVIDIA GeForce Graphics Card (
Sapphire Pulse AMD Radeon RX 6500 XT Gaming Graphics Card with 4GB GDDR6, AMD RDNA™ 2
ASUS Phoenix NVIDIA GeForce GTX 1650 OC Edition pci_e_x16 PCIe 3.0, 4GB GDDR6 Memory, HD
ZEBRONICS NVIDIA ZEB-GT610 2 GB DDR3 Graphics Card
INNO3D GEFORCE RTX 3050 Twin X2 OC 8GB GDDR6 128-bit Gaming Graphic Card - N30502-08D6
MSI Radeon RX 6500 XT MECH 2X 4G OC 4GB GDDR6 64-bit Gaming Graphic Card
Galax GeForce ® GTX 1660 Ti 1-Click OC 6GB GDDR6 192-bit DP/HDMI/DVI-D Graphic Card
PowerColor AMD Radeon RX 6400 Low Profile 4GB GDDR6 Graphics Card
ASUS Radeon PH-RX550-4G-EVO Graphics Cards RX 550
ZOTAC GeForce GT 730 4GB DDR3 ZONE Edition Graphics Card with GeForce Experience
NVIDIA Quadro T400 2GB GDDR6 64-Bit Graphics Card
INNO3D GTX 1650 Twin X2 OC - 4GB GDDR6 128 bit Gaming Graphic Card - N16502-04D6X-1177V
MSI Gaming GeForce GTX 1660 Super 192-bit HDMI/DP 6GB GDRR6 HDCP Support DirectX 12 Dua

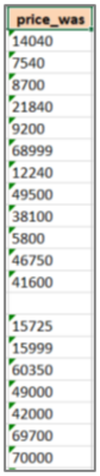
OEM	Card Manufacturer
Colorful	Nvidia
ASUS	Nvidia
Zotac	Nvidia
ASUS	Nvidia
Gigabyte	Nvidia
GIGABYTE	Nvidia
MSI	Nvidia
Sapphire	AMD
ASUS	Nvidia
ZEBRONICS	Nvidia
INNO3D	Nvidia
MSI	AMD
ASUS	AMD
ZOTAC	Nvidia
NVIDIA	
INNO3D	Nvidia
MSI	Nvidia
ASUS	Nvidia
INNO3D	Nvidia
GALAX	Nvidia
Biostar	Nvidia
ASUS	Nvidia
Asus	Nvidia

2. **Exclusions: Removal of multiple same value of original price and create a new field with single value**

   (Screen shot is not showing apple to apple comparison but emphasize upon the workflow – input and output example)



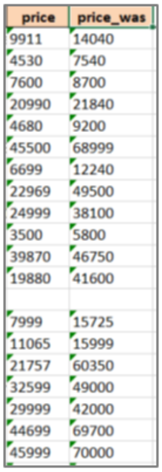
price_was
₹70,000₹70,000
₹1,99,999₹1,99,999
₹14,040₹14,040
₹7,550₹7,550
₹8,700₹8,700
₹21,840₹21,840
₹9,200₹9,200
₹68,999₹68,999
₹12,240₹12,240
₹49,500₹49,500
₹38,100₹38,100
₹5,800₹5,800
₹46,750₹46,750
₹41,600₹41,600
₹60,000₹60,000
₹29,200₹29,200
₹15,725₹15,725
₹15,999₹15,999
₹60,350₹60,350



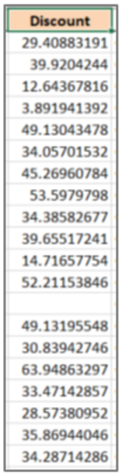
price_was
14040
7540
8700
21840
9200
68999
12240
49500
38100
5800
46750
41600
15725
15999
60350
49000
42000
69700
70000

3. **Calculated Field:**

Create a discount rate value using the 'Price' and 'Price was' columns.



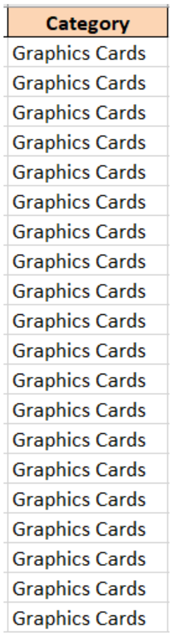
price	price_was
9911	14040
4530	7540
7600	8700
20990	21840
4680	9200
45500	68999
6699	12240
22969	49500
24999	38100
3500	5800
39870	46750
19880	41600
7999	15725
11065	15999
21757	60350
32599	49000
29999	42000
44699	69700
45999	70000



Discount
29.40883191
39.9204244
12.64367816
3.891941392
49.13043478
34.05701532
45.26960784
53.5979798
34.38582677
39.65517241
14.71657754
52.21153846
49.13195548
30.83942746
63.94863297
33.47142857
28.57380952
35.86944046
34.28714286

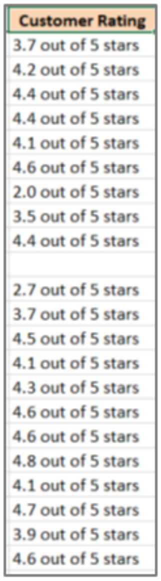
4. **Define a structured data column:**

Create a new attribute the segments the product category

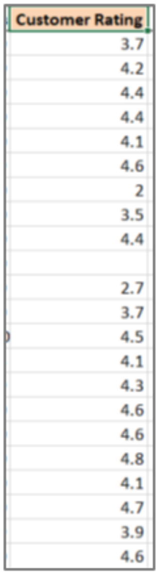


Category
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards
Graphics Cards

5. **String to Float conversion:** Change the customer rating string values to float and show only rating value

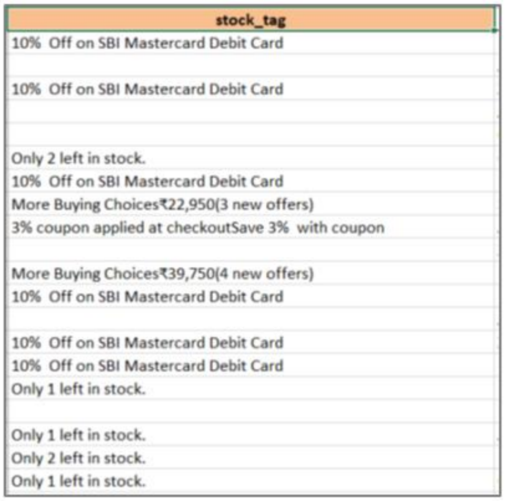


Customer Rating
3.7 out of 5 stars
4.2 out of 5 stars
4.4 out of 5 stars
4.4 out of 5 stars
4.1 out of 5 stars
4.6 out of 5 stars
2.0 out of 5 stars
3.5 out of 5 stars
4.4 out of 5 stars
2.7 out of 5 stars
3.7 out of 5 stars
4.5 out of 5 stars
4.1 out of 5 stars
4.3 out of 5 stars
4.6 out of 5 stars
4.6 out of 5 stars
4.8 out of 5 stars
4.1 out of 5 stars
4.7 out of 5 stars
3.9 out of 5 stars
4.6 out of 5 stars

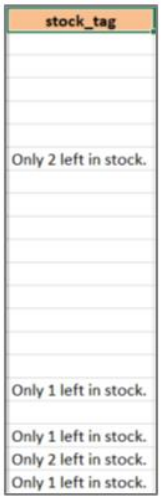


Customer Rating
3.7
4.2
4.4
4.4
4.1
4.6
2
3.5
4.4
2.7
3.7
4.5
4.1
4.3
4.6
4.6
4.8
4.1
4.7
3.9
4.6

6. **Remove irrelevant data from stock tag:** Excluded all the irrelevant data from stock tag column and replaced it with blank (For ex – 10% off on SBI Master Debit Card, has nothing no relevance with stock tag. Stock tag is to reflect the number of units left in the stock.

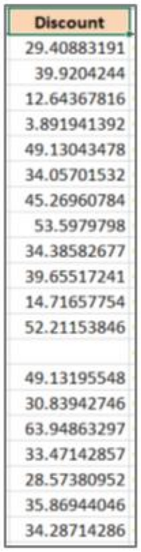


stock_tag
10% Off on SBI Mastercard Debit Card
10% Off on SBI Mastercard Debit Card
Only 2 left in stock.
10% Off on SBI Mastercard Debit Card
More Buying Choices₹22,950(3 new offers)
3% coupon applied at checkoutSave 3% with coupon
More Buying Choices₹39,750(4 new offers)
10% Off on SBI Mastercard Debit Card
10% Off on SBI Mastercard Debit Card
10% Off on SBI Mastercard Debit Card
Only 1 left in stock.
Only 1 left in stock.
Only 2 left in stock.
Only 1 left in stock.

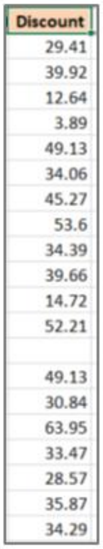


stock_tag
Only 2 left in stock.
Only 1 left in stock.
Only 1 left in stock.
Only 2 left in stock.
Only 1 left in stock.

7. **Rounding off discount column to 2 decimal places**

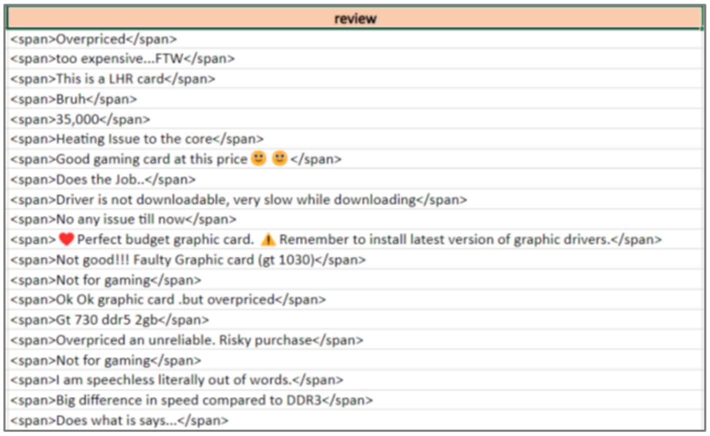


Discount
29.40883191
39.9204244
12.64367816
3.891941392
49.13043478
34.05701532
45.26960784
53.5979798
34.38582677
39.65517241
14.71657754
52.21153846
49.13195548
30.83942746
63.94863297
33.47142857
28.57380952
35.86944046
34.28714286

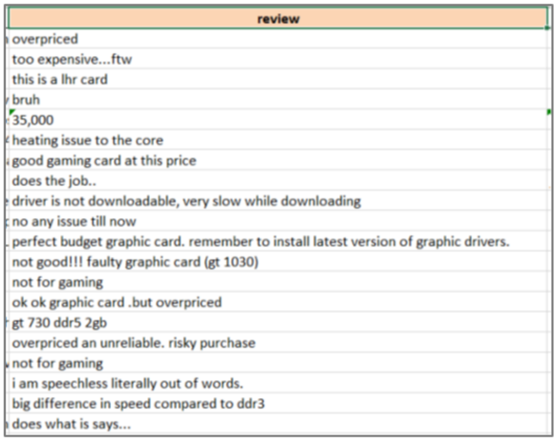


Discount
29.41
39.92
12.64
3.89
49.13
34.06
45.27
53.6
34.39
39.66
14.72
52.21
49.13
30.84
63.95
33.47
28.57
35.87
34.29

8. **Removing <span> and </span> tags and emojis**

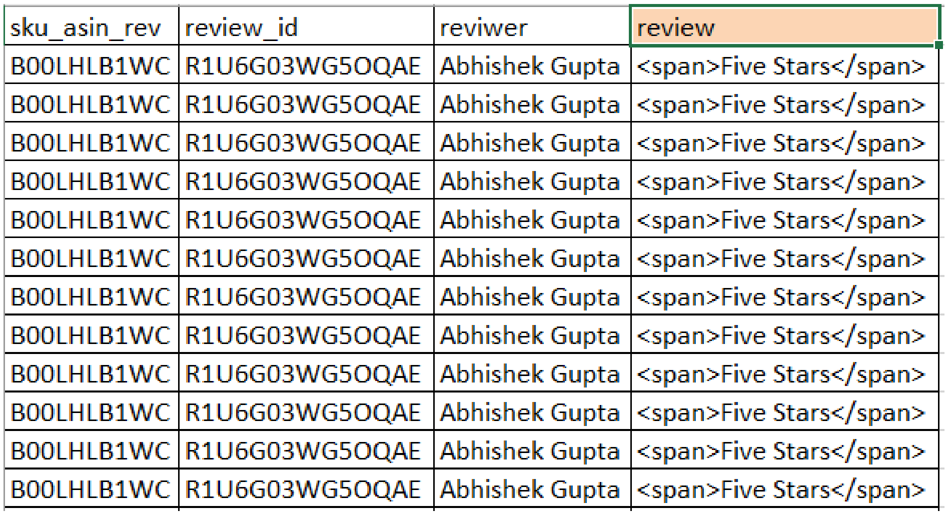


review
<span>Overpriced</span>
<span>too expensive...FTW</span>
<span>This is a LHR card</span>
<span>Bruh</span>
<span>35,000</span>
<span>Heating Issue to the core</span>
<span>Good gaming card at this price 🙂 🙂 </span>
<span>Does the Job..</span>
<span>Driver is not downloadable, very slow while downloading</span>
<span>No any issue till now</span>
<span>❤️ Perfect budget graphic card. ⚠️ Remember to install latest version of graphic drivers.</span>
<span>Not good!!! Faulty Graphic card (gt 1030)</span>
<span>Not for gaming</span>
<span>Ok Ok graphic card .but overpriced</span>
<span>Gt 730 ddr5 2gb</span>
<span>Overpriced an unreliable. Risky purchase</span>
<span>Not for gaming</span>
<span>I am speechless literally out of words.</span>
<span>Big difference in speed compared to DDR3</span>
<span>Does what is says...</span>



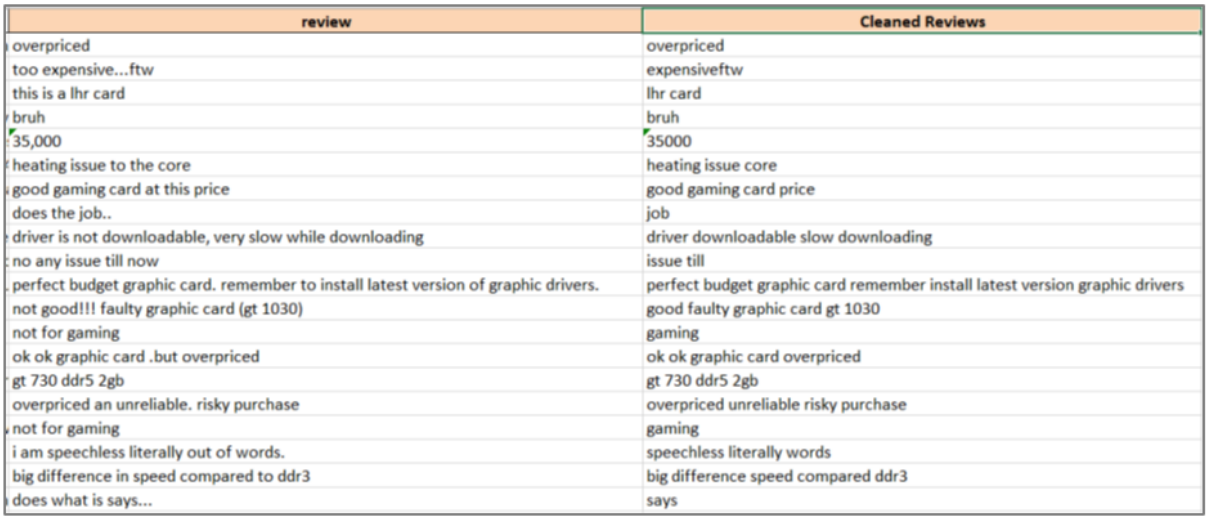
review
overpriced
too expensive...ftw
this is a lhr card
bruh
35,000
heating issue to the core
good gaming card at this price
does the job..
driver is not downloadable, very slow while downloading
no any issue till now
perfect budget graphic card. remember to install latest version of graphic drivers.
not good!!! faulty graphic card (gt 1030)
not for gaming
ok ok graphic card .but overpriced
gt 730 ddr5 2gb
overpriced an unreliable. risky purchase
not for gaming
i am speechless literally out of words.
big difference in speed compared to ddr3
does what is says...

9. Removing Duplicates



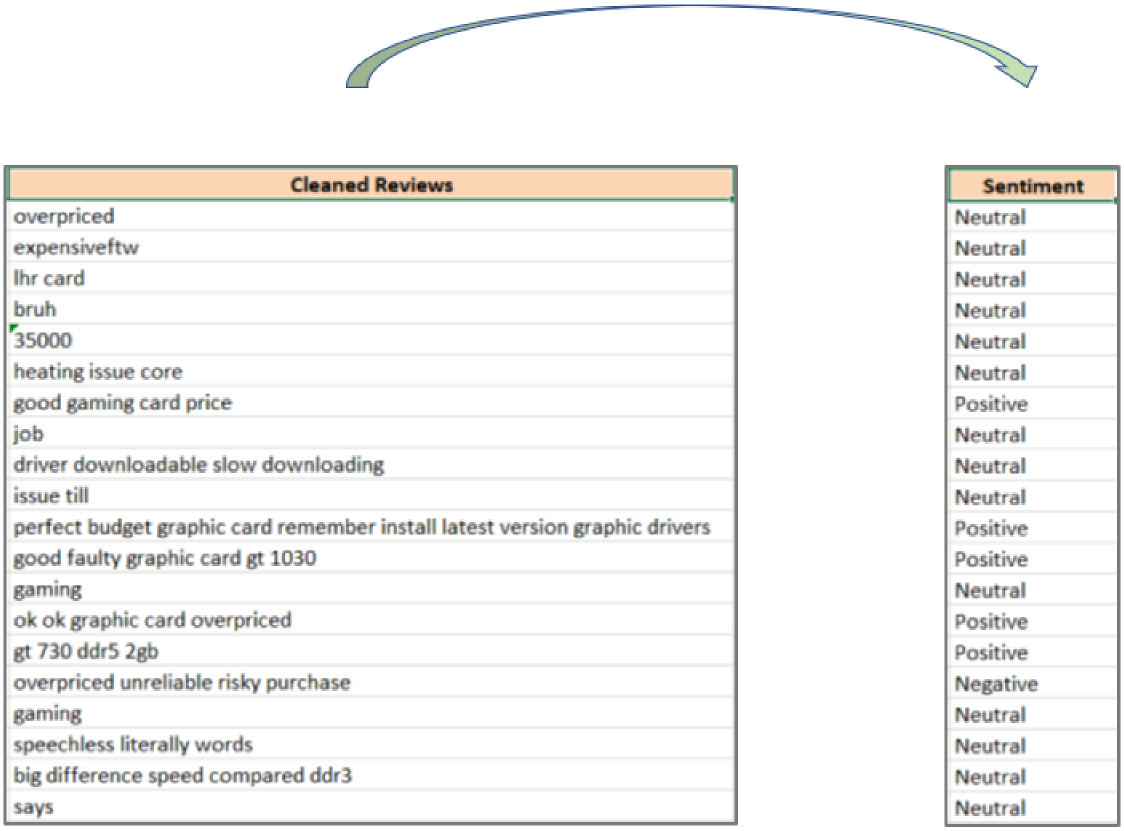
sku_asin_rev	review_id	reviwer	review
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>
B00LHLB1WC	R1U6G03WG5OQAE	Abhishek Gupta	<span>Five Stars</span>

10.Removing Punctuation and stop words



review	Cleaned Reviews
overpriced	overpriced
too expensive...ftw	expensiveftw
this is a lhr card	lhr card
bruh	bruh
35,000	35000
heating issue to the core	heating issue core
good gaming card at this price	good gaming card price
does the job..	job
driver is not downloadable, very slow while downloading	driver downloadable slow downloading
no any issue till now	issue till
perfect budget graphic card. remember to install latest version of graphic drivers.	perfect budget graphic card remember install latest version graphic drivers
not good!!! faulty graphic card (gt 1030)	good faulty graphic card gt 1030
not for gaming	gaming
ok ok graphic card .but overpriced	ok ok graphic card overpriced
gt 730 ddr5 2gb	gt 730 ddr5 2gb
overpriced an unreliable. risky purchase	overpriced unreliable risky purchase
not for gaming	gaming
i am speechless literally out of words.	speechless literally words
big difference in speed compared to ddr3	big difference speed compared ddr3
does what is says...	says

11.Creating Sentiment scoring and bands using sentiment intensity analyser:



Cleaned Reviews	Sentiment
overpriced	Neutral
expensiveftw	Neutral
lhr card	Neutral
bruh	Neutral
35000	Neutral
heating issue core	Neutral
good gaming card price	Positive
job	Neutral
driver downloadable slow downloading	Neutral
issue till	Neutral
perfect budget graphic card remember install latest version graphic drivers	Positive
good faulty graphic card gt 1030	Positive
gaming	Neutral
ok ok graphic card overpriced	Positive
gt 730 ddr5 2gb	Positive
overpriced unreliable risky purchase	Negative
gaming	Neutral
speechless literally words	Neutral
big difference speed compared ddr3	Neutral
says	Neutral

Observations/ Insights and Analysis on the data collected

After pre-processing and cleaning the data, our data was ready for generating early insights and we decided to look at multiple segments about the product category using different metrics and slicer –

- Competitor product listing analysis (Using Card Manufacturer and OEM Attribute)
- Top products with high discounts (How do we compare with competition on price offering?)
- Top Customer Rated Products
- Top products with highest number of reviews
- Sentiment scoring and band on reviews given

Few Early insights:

86% of OEM listings on Amazon are of NVIDIA so AMD holds a very low share

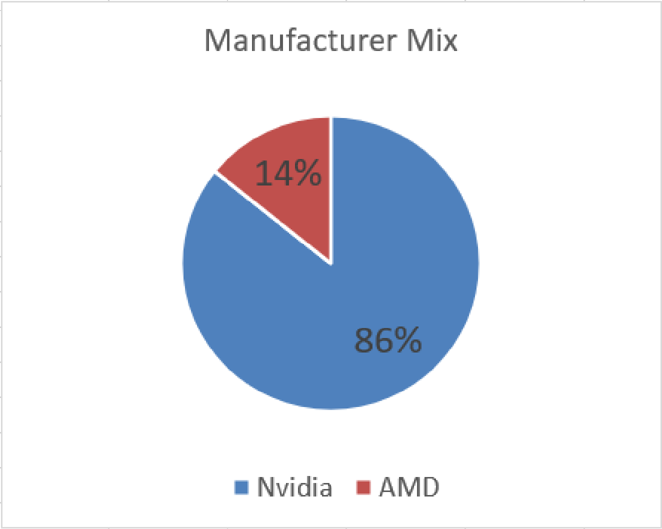
Similar OEMs (Gigabyte and AMD), have higher discount offerings by NVIDIA than AMD

Comparatively, AMD ASUS has higher customer rating (4.56) than NVIDIA ASUS (4.42) and of course, NVIDIA has higher number of OEM types than AMD

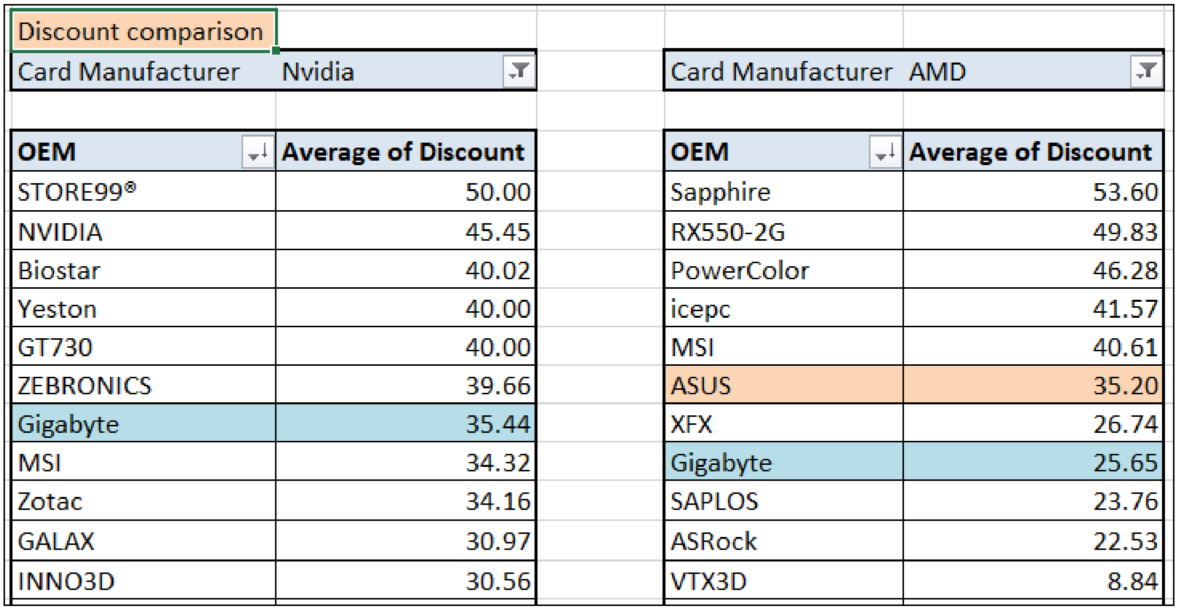
The numbers of reviews received by NVIDIA OEM (Gigabyte – 11962, ASUS – 10886) is much higher than the AMD OEM (Gigabyte – 9, ASUS – 128)

Based on customer reviews, the graphic card category has positive sentiments across all products

1. Manufacturer Mix:



2. Discount comparison – NVIDIA Vs AMD



Discount comparison

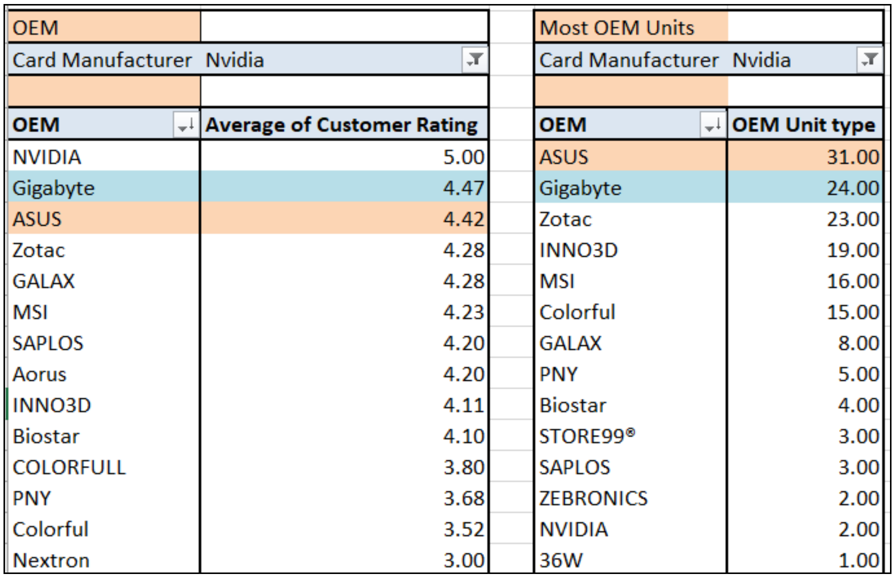
Card Manufacturer Nvidia

OEM	Average of Discount
STORE99®	50.00
NVIDIA	45.45
Biostar	40.02
Yeston	40.00
GT730	40.00
ZEBRONICS	39.66
Gigabyte	35.44
MSI	34.32
Zotac	34.16
GALAX	30.97
INNO3D	30.56

Card Manufacturer AMD

OEM	Average of Discount
Sapphire	53.60
RX550-2G	49.83
PowerColor	46.28
icepc	41.57
MSI	40.61
ASUS	35.20
XFX	26.74
Gigabyte	25.65
SAPLOS	23.76
ASRock	22.53
VTX3D	8.84

3. Top 15 products with average customer rating and OEM unit type by NVIDIA Vs AMD:



OEM

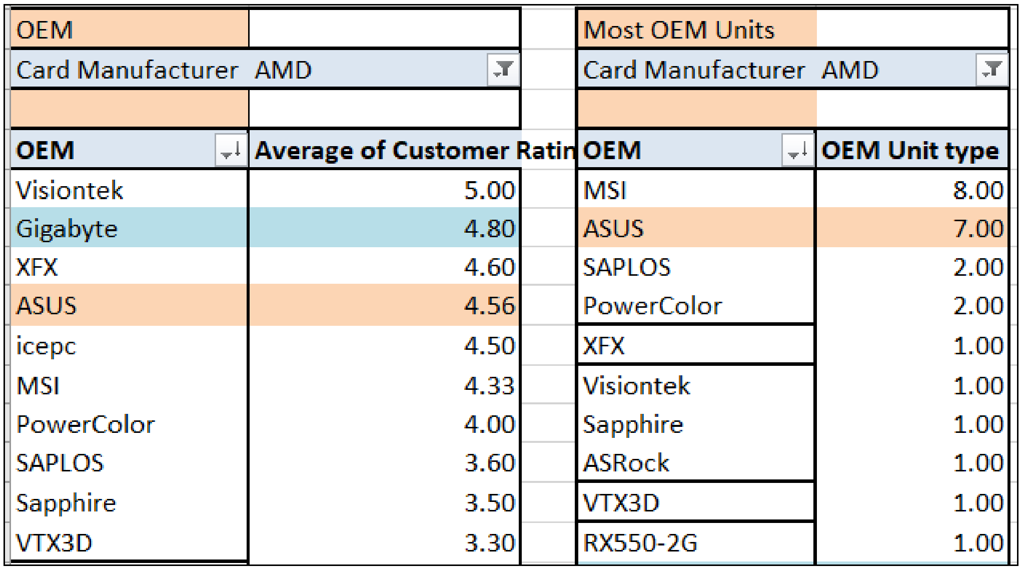
Card Manufacturer Nvidia

OEM	Average of Customer Rating
NVIDIA	5.00
Gigabyte	4.47
ASUS	4.42
Zotac	4.28
GALAX	4.28
MSI	4.23
SAPLOS	4.20
Aorus	4.20
INNO3D	4.11
Biostar	4.10
COLORFULL	3.80
PNY	3.68
Colorful	3.52
Nextron	3.00

Most OEM Units

Card Manufacturer Nvidia

OEM	OEM Unit type
ASUS	31.00
Gigabyte	24.00
Zotac	23.00
INNO3D	19.00
MSI	16.00
Colorful	15.00
GALAX	8.00
PNY	5.00
Biostar	4.00
STORE99®	3.00
SAPLOS	3.00
ZEBRONICS	2.00
NVIDIA	2.00
36W	1.00



OEM

Card Manufacturer AMD

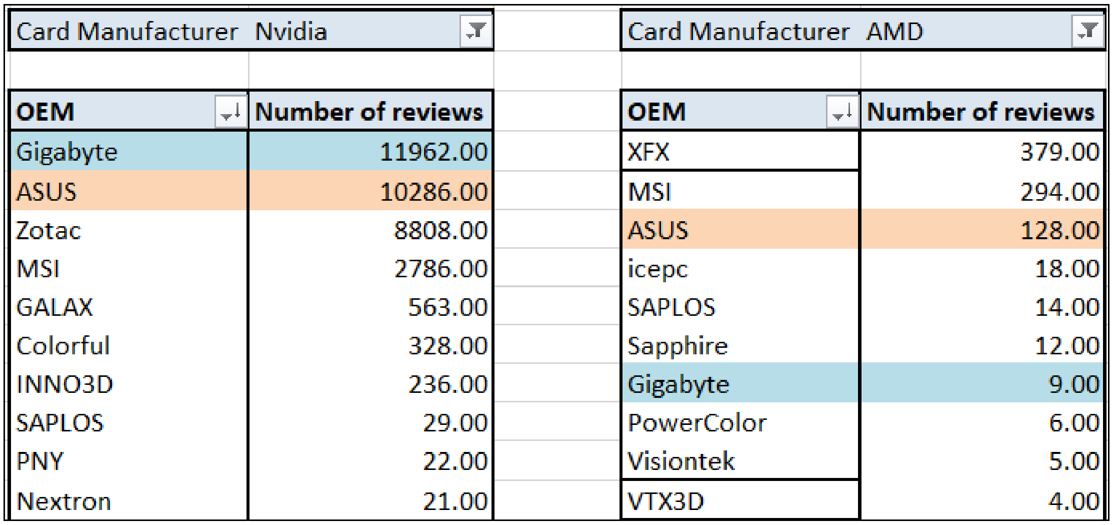
OEM	Average of Customer Ratin
Visiontek	5.00
Gigabyte	4.80
XFX	4.60
ASUS	4.56
icepc	4.50
MSI	4.33
PowerColor	4.00
SAPLOS	3.60
Sapphire	3.50
VTX3D	3.30

Most OEM Units

Card Manufacturer AMD

OEM	OEM Unit type
MSI	8.00
ASUS	7.00
SAPLOS	2.00
PowerColor	2.00
XFX	1.00
Visiontek	1.00
Sapphire	1.00
ASRock	1.00
VTX3D	1.00
RX550-2G	1.00

4. Most Reviews by OEM – NVIDIA Vs AMD



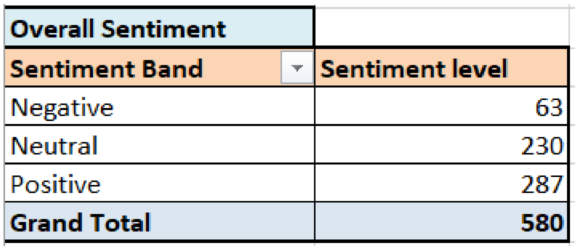
Card Manufacturer Nvidia

OEM	Number of reviews
Gigabyte	11962.00
ASUS	10286.00
Zotac	8808.00
MSI	2786.00
GALAX	563.00
Colorful	328.00
INNO3D	236.00
SAPLOS	29.00
PNY	22.00
Nextron	21.00

Card Manufacturer AMD

OEM	Number of reviews
XFX	379.00
MSI	294.00
ASUS	128.00
icepc	18.00
SAPLOS	14.00
Sapphire	12.00
Gigabyte	9.00
PowerColor	6.00
Visiontek	5.00
VTX3D	4.00

5. Overall Sentiment – Overall graphic card category has positive sentiment the most in customer reviews across all products



Overall Sentiment	
Sentiment Band	**Sentiment level**
Negative	63
Neutral	230
Positive	287
Grand Total	**580**

Strategy to enhance the data with crowd sourcing methods

Crowd sourcing is “an online, distributed problem-solving and production model” – Daren Brabham (2008)

In the last few years, crowd sourcing has emerged as one of the strong ways of engaging with large group of people in generating ideas for complex business problem in tech landscape.

In context of retail domain, it is vital for any e-commerce company to create a seamless app and website experience for its customers. Hence, one of the most beneficial areas for e-commerce companies to engage in crowdsourcing is software testing.

While this can be done by an in-house team and contractors but with the recent growth seen in the unique number of mobile devices and websites, there is a problem at hand with big retailers to test these things at scale to provide great customer experience.

Strategically,

1. **Crowdsourced Software testing:** Engaging with crowdsourcing-based software testing companies is a great way to deal with this problem at scale with same speed and convenience as In-house set ups. There is high likeliness to be successful with this approach as the crowdsourcing-based software testers are paid based on the bugs they find vs the time spent on testing the software.

2. **Crowd voting:** Engage users in crowd voting through some suggestion forums and online communities to gauge feedback on specific areas – app experience, product availability experience, delivery experience

3. **Crowd focus groups:** Use traditional ways of gathering crowd ideas are geo centric focus groups and telephonic interviews on specific business problems

4. **Open Crowd work:** Leverage literature and research work in field of data science (A lot can be referenced through sites like Kaggle)

We would like to leverage these approaches of crowdsourcing as per the business needs

References and Sources used for this Assignment

1. https://realpython.com/beautiful-soup-web-scraper-python/

2. https://www.geeksforgeeks.org/difference-between-find-and-find_all-in-beautifulsoup-python/

3. https://www.amazon.in/

4. https://www.w3resource.com/pandas/series/series-str-split.php#:~:text=split()%20function%20is%20used,split().

5. https://www.w3resource.com/python-exercises/nltk/index.php

6. https://kanoki.org/2019/11/12/how-to-use-regex-in-pandas/

7. https://www.browserstack.com/guide/inspect-element-in-chrome#:~:text=One%20of%20the%20easiest%20ways,%2C%20Sources%2C%20and%20other%20tools.

8. https://github.com/deepak-gupta-isb/DCPP

-