

# **MDA 720 Applied Business Analytics**

## **Capstone Project**

Topic

**Best Purchased Laptops of Previous Year**

By

**Md Redowan Amin Mollick**

# Table of Contents

- Introduction
  - 1.1 Background
  - 1.2 Objectives
  - 1.3 Scope
- Literature Review
  - 2.1 Overview of Relevant Literature
  - 2.2 Key Concepts and Definitions
- Methodology
  - 3.1 Research Design
  - 3.2 Data Collection and Sampling
  - 3.3 Data Analysis
- Results
  - 4.1 Overview of Findings
  - 4.2 Presentation of Data
  - 4.3 Analysis of Findings
- Discussion
  - 5.1 Interpretation of Findings
  - 5.2 Implications and Significance
  - 5.3 Limitations and Future Research
- Conclusion
  - 6.1 Summary of Findings
  - 6.2 Contributions to the Field
- References

# Introduction

Since childhood, in my mind I always wanted to have my own app development firm for laptops. It could be anything, gaming, lifestyle, sports biased.

So, for this reason, to get enter and plan a business I've have the full idea about the market, which specifications are mostly purchased, screen size, rom-ram capacity, gpu. Basically, that's the reason to choose this topic "Best Purchased Laptop of Previous Year".

So, for this project I first choose to work with amazon's data, but data was pretty hard to get and I got blocked. Then I choose one of the largest online market Flipcart. Flipkart Private Limited is an Indian e-commerce company, headquartered in Bengaluru, and incorporated in Singapore as a private limited company. Last year their revenue was 618.4 billion INR (US\$7.7 billion, 2022).

I got the data of last year's June laptop purchased. I will be running Text Mining to Data cleaning and processing then with those data I'll be running some data analysis.

## Literature Review

So, the data I've got from Flipcart is the laptop purchased in June 2022. So, for data cleaning and processing I'll be using Text Mining. I'll be using NLTK function in jupyter notebook(python).

Natural Language Toolkit (NLTK) is a text mining tool and a popular Python library for natural language processing (NLP). It provides a wide range of tools and resources for text processing and analysis, including tokenization, stemming, lemmatization, part-of-speech tagging, named entity recognition, sentiment analysis, and more.

NLTK also includes corpora, or large collections of text for training and testing NLP models, as well as pre-trained models for various NLP tasks. NLTK is widely used in research and industry for a variety of NLP applications, including text classification, sentiment analysis, machine translation, and information retrieval, among others.

NLTK can be used in Jupyter Notebook, as well as other Python environments, to perform text mining and analysis on textual data.

# Methodology

Data sets I got from Kreggle, they are free to use and there's no restrictions.

Designing a research project for text mining using the Natural Language Toolkit (NLTK) requires careful planning and consideration of several key factors. Here are some steps you can follow to create a solid research design:

**Define my research question:** Before I begin your text mining project, I need to have a clear research question that I want to answer. For example, I may be interested in analyzing sentiment in customer reviews or identifying patterns in political speeches. I research question will guide the rest of I project and help I determine what data to collect and how to analyze it.

**Choose my data source:** Once I have a research question in mind, I need to identify the data source(s) that I will use for your analysis. This may be a collection of texts, such as a set of social media posts or a corpus of literature. I will need to determine how to obtain this data and what format it will be in (e.g., PDF, HTML, plaintext).

**Preprocess my data:** Text data often requires preprocessing to clean and transform it into a usable format. This may involve tasks such as removing stop words, stemming or lemmatizing words, and converting the data into a machine-readable format. NLTK offers many tools for preprocessing text data, so you should familiarize yourself with these tools and use them as needed.

**Identify relevant features:** Depending on my research question, I will need to identify relevant features of the text data that you want to analyze. For example, if I am interested in sentiment analysis, one may want to focus on identifying positive or negative words in the text. If you are interested in topic modeling, one may want to identify common themes or topics in the text.

**Choose my analysis methods:** NLTK offers a wide range of tools for text analysis, including methods for sentiment analysis, topic modeling, named entity recognition, and more. I will need to choose the appropriate analysis methods for your research question and data source.

**Evaluate my results:** Once I have performed your analysis, I will need to evaluate your results to determine whether they answer my research question. This may involve visualizing the data, conducting statistical tests, or comparing your results to existing research.

By following these steps, you can design a rigorous and effective research project using the Natural Language Toolkit.

# Data Set

125%

Zoom

Category

Pivot Table

Insert

Table

Chart

Text

Shape

Media

Comment

Collaborate

Format

Organise

Sheet 1

link

0

https://www.flipkart.com/asus-rog-strix-scar-17-core-i9-12th-gen-32-gb-1-tb-ssd-windows-11-home-8-gb-graphics-nvidia-geforce-rtx-3070-ti-g733zw-l139ws-gaming-laptop/p/itm0e721469cc70e

1

https://www.flipkart.com/asus-rog-strix-scar-15-core-i9-12th-gen-32-gb-1-tb-ssd-windows-11-home-8-gb-graphics-nvidia-geforce-rtx-3070-ti-g533zx-l136ws-gaming-laptop/p/itmde8ba35232e

2

https://www.flipkart.com/hp-victus-ryzen-7-octa-core-5800h-16-gb-512-gb-ssd-windows-11-home-4-graphics-nvidia-geforce-rtx-3050-16-e0351ax-gaming-laptop/p/itm3a89e9e8d4987pid=COMG

3

https://www.flipkart.com/lenovo-ideapad-gaming-3i-ryzen-7-octa-core-5800h-5th-gen-16-gb-512-gb-ssd-windows-11-home-4-graphics-nvidia-geforce-rtx-3050-15ach6-laptop/p/itmcb01bf6c4

4

https://www.flipkart.com/lenovo-yoga-slim-7-core-i5-11th-gen-16-gb-512-gb-ssd-windows-11-home-82a300mbin-thin-light-laptop/p/itm8d670bbb74a5b7pid=COMGD8RMGDEUBYSZ&lid=LSTCOMG

5

https://www.flipkart.com/lenovo-yoga-slim-7-core-i7-11th-gen-16-gb-512-gb-ssd-windows-11-home-14i005-thin-light-laptop/p/itm4a202e05ce6327pid=COMGD8NVUWf8MF5YG&lid=LSTCOMGD

6

https://www.flipkart.com/lenovo-v15-celeron-dual-core-4-gb-256-gb-ssd-windows-10-82c30053th-thin-light-laptop/p/itmhb8f7d46586e7pid=COMMGDFCDX9JIDPC&lid=LSTCOMGDFCDX9JID

7

https://www.flipkart.com/lenovo-yoga-6-ryzen-7-octa-core-5700u-5th-gen-16-gb-1-tb-ssd-windows-11-home-13alc6-2-1-laptop/p/itm4cabce05238a77pid=COMGD8Z4UNZNHD2H&lid=LSTCO

8

https://www.flipkart.com/asus-tuf-gaming-f15-core-i5-10th-gen-8-gb-1-tb-ssd-windows-11-home-4-gb-graphics-nvidia-geforce-gtx-1650-144-hz-tx506h-hn310w-laptop/p/itm2aff3c6d83cb7pid=

9

https://www.flipkart.com/dell-inspiron-pentium-silver-8-gb-256-gb-ssd-windows-11-home-3521-notebook/p/itm8a3845e0d7c3d7pid=COMGCGFY92HFJXHTU&lid=LSTCOMGCGFY92HFJXHTUJAWJ

10

https://www.flipkart.com/dell-inspiron-athlon-dual-core-3050u-8-gb-256-gb-ssd-windows-11-home-3525-notebook/p/itm96a8101fa53ff7pid=COMGCGFY92KXPEW&lid=LSTCOMGCGFY92KXPEW

11

https://www.flipkart.com/dell-core-i5-11th-gen-16-gb-512-gb-ssd-windows-11-home-xp9305-thin-light-laptop/p/itm401718e258f557pid=COMGBTBG43BUDZ&lid=LSTCOMGBTBG43BUDZ

12

https://www.flipkart.com/lenovo-yoga-7i-ryzen-7-octa-core-57800u-5th-gen-16-gb-512-gb-ssd-windows-10-home-14it5-2-1-laptop/p/itm3c084b43baa677pid=COMGD4Y5G53Z9QF&lid=LSTCO

13

https://www.flipkart.com/asus-ryzen-7-dual-core-7th-gen-16-gb-512-gb-hdd-512-ssd-windows-11-home-4-graphics-nvidia-geforce-rx-3050-4-gb-fa506ic-hn075w-gaming-laptop/p/itmcb73703ff25

14

https://www.flipkart.com/asus-ryzen-3-quad-core-3rd-gen-8-gb-256-gb-ssd-windows-11-home-m515da-br322ws-laptop/p/itm5c15928e928657pid=COMGCT6SPZVMCR&lid=LSTCOMGCT6SP

15

https://www.flipkart.com/asus-rog-flow-x13-2021-ryzen-9-octa-core-5900hs-5th-gen-32-gb-1-tb-ssd-windows-10-home-4-gb-graphics-nvidia-geforce-gtx-1650-120-hz-gn301qn-k4641ts-2-1-ga

16

https://www.flipkart.com/gigabyte-leopard-core-i7-11th-gen-16-gb-1-tb-ssd-windows-10-home-8-gb-graphics-nvidia-geforce-rtx-3070-240-hz-11-ug-gaming-laptop/p/itm35c5d6e2988btf7pid=COMG

17

https://www.flipkart.com/asus-vivobook-15-oled-core-3-11th-gen-8-gb-256-gb-ssd-windows-11-home-k513ea-301ws-laptop/p/itm1f81ebe4b317b7pid=COMGCGWDMTPHCH&lid=LSTCOMGCG

18

https://www.flipkart.com/realme-book-prime-core-i5-11th-gen-16-gb-512-gb-ssd-windows-11-home-cloudpro002-thin-light-laptop/p/itm41059a863fc027pid=COMGCWBEHFVWXYA&lid=LSTCO

19

https://www.flipkart.com/realme-book-prime-core-i5-11th-gen-16-gb-512-gb-ssd-windows-11-home-cloudpro002-thin-light-laptop/p/itm7d1b1a9d3579b7pid=COMGCWBEHVBYSMA6&lid=LSTCO

20

https://www.flipkart.com/realme-book-prime-core-i5-11th-gen-16-gb-512-gb-ssd-windows-11-home-cloudpro002-thin-light-laptop/p/itmce1634fe035d67pid=COMGCWBEHVEYCY7WH&lid=LSTCO

21

https://www.flipkart.com/dell-g15-core-i7-11th-gen-16-gb-512-gb-ssd-windows-11-home-4-graphics-nvidia-geforce-rx-3050-ti-165-hz-g15-5511-se-gaming-laptop/p/itm39da14bb5a4217pid=CO

22

https://www.flipkart.com/dell-ryzen-5-hexa-core-5600h-8-gb-512-gb-ssd-windows-11-home-4-graphics-nvidia-geforce-rx-3050-120-hz-g15-5515-gaming-laptop/p/itm14d1cb2c72fe7pid=COMG

23

https://www.flipkart.com/avita-satus-celeron-dual-core-4-gb-128-gb-ssd-windows-11-home-nu144inc43pn-mb-laptop/p/itmcc27c094ef28d7pid=COMGCGWTFYFFKAAAY&lid=LSTCOMGCGWTFY

24

https://www.flipkart.com/asus-vivobook-s14-oled-2022-intel-evo-core-i5-12th-gen-16-gb-512-gb-ssd-windows-11-home-s3402za-km501ws-thin-light-laptop/p/itm8d9c5d1d680c47pid=COMGCG2

25

https://www.flipkart.com/asus-vivobook-s14-oled-2022-intel-evo-core-i7-12th-gen-16-gb-512-gb-ssd-windows-11-home-s3402za-km702ws-thin-light-laptop/p/itm4086591ca87c7pid=COMGCG

26

https://www.flipkart.com/asus-zenbook-14-oled-2022-intel-evo-core-i5-12th-gen-16-gb-512-gb-ssd-windows-11-home-ux3402za-km531ws-thin-light-laptop/p/itm82ae2e0518b6547pid=COMGDG

27

https://www.flipkart.com/asus-zenbook-14x-2022-space-intel-evo-core-i7-12th-gen-16-gb-1-tb-ssd-windows-11-home-ux5401zas-kn171ws-thin-light-laptop/p/itm5ff6cd13c40e7pid=COMGDGN

28

https://www.flipkart.com/asus-vivobook-s14-oled-2022-intel-evo-core-i5-12th-gen-16-gb-512-gb-ssd-windows-11-home-s3402za-km501ws-thin-light-laptop/p/itm8d9c5d1d680c47pid=COMGCG2

Sheet

Sheet Name

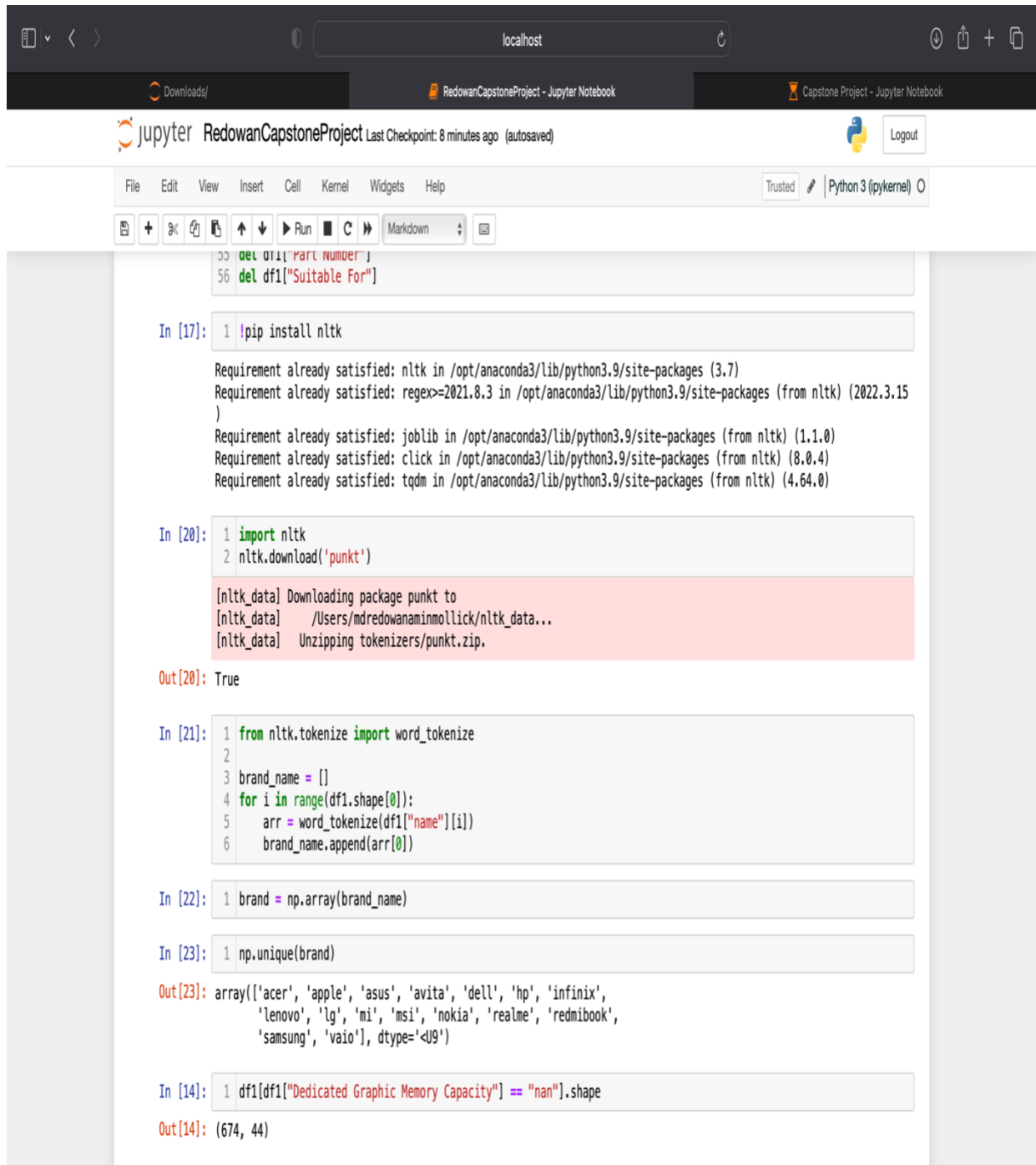
Sheet 1

Background

Duplicate Sheet

Delete Sheet

# Data cleaning and preprocessing using Text Mining



The screenshot displays a Jupyter Notebook interface with the following content:

```
del df1["Part Number"]
del df1["Suitable For"]
```

In [17]: `!pip install nltk`

Requirement already satisfied: nltk in /opt/anaconda3/lib/python3.9/site-packages (3.7)  
Requirement already satisfied: regex>=2021.8.3 in /opt/anaconda3/lib/python3.9/site-packages (from nltk) (2022.3.15)  
Requirement already satisfied: joblib in /opt/anaconda3/lib/python3.9/site-packages (from nltk) (1.1.0)  
Requirement already satisfied: click in /opt/anaconda3/lib/python3.9/site-packages (from nltk) (8.0.4)  
Requirement already satisfied: tqdm in /opt/anaconda3/lib/python3.9/site-packages (from nltk) (4.64.0)

In [20]: `import nltk`  
`nltk.download('punkt')`

[nltk\_data] Downloading package punkt to  
[nltk\_data] /Users/mdredowanaminmollick/nltk\_data...  
[nltk\_data] Unzipping tokenizers/punkt.zip.

Out[20]: True

In [21]: `from nltk.tokenize import word_tokenize`  
`brand_name = []`  
`for i in range(df1.shape[0]):`  
 `arr = word_tokenize(df1["name"][i])`  
 `brand_name.append(arr[0])`

In [22]: `brand = np.array(brand_name)`

In [23]: `np.unique(brand)`

Out[23]: array(['acer', 'apple', 'asus', 'avita', 'dell', 'hp', 'infinix',  
 'lenovo', 'lg', 'mi', 'msi', 'nokia', 'realme', 'redmibook',  
 'samsung', 'vaio'], dtype='<U9')

In [14]: `df1[df1["Dedicated Graphic Memory Capacity"] == "nan"].shape`

Out[14]: (674, 44)

localhost

Downloads/RedowanCapstoneProject - Jupyter NotebookCapstone Project - Jupyter Notebook

Jupyter RedowanCapstoneProject Last Checkpoint: 8 minutes ago (autosaved)Logout

FileEditViewInsertCellKernelWidgetsHelp

Python 3 (pykernel)

Run

Markdown

```
In [24]: 1 # Function for encoding categorical string data into integers
2
3 def encode_to_int(arr):
4     d = {}
5     for p in enumerate(arr):
6         if p[1] not in d:
7             d[p[1]] = p[0]+1
8
9     return d

In [26]: 1 req = df1["Screen Resolution"]
2 max_screen_res = list(req)
3
4 for i in range(len(max_screen_res)):
5     res = " ".join(re.split(r'\D+', max_screen_res[i])).split()
6     if len(res)==1:
7         max_screen_res[i]=int(res[0])
8     else:
9         max_screen_res[i]=min(int(res[0]),int(res[1]))
10
11 for i in range(len(max_screen_res)):
12     if max_screen_res[i]>=2160:
13         max_screen_res[i]=2160
14     elif max_screen_res[i]>=1440:
15         max_screen_res[i]=1440
16     elif max_screen_res[i]>=1080:
17         max_screen_res[i]=1080
18     elif max_screen_res[i]>=720:
19         max_screen_res[i]=720
20     else:
21         max_screen_res[i]=480

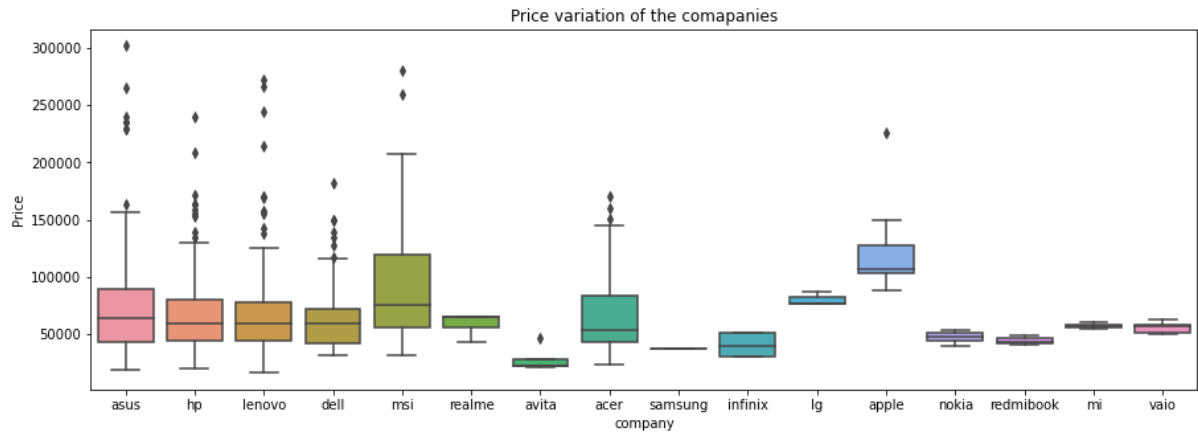
In [27]: 1 df1["max_res"] = pd.DataFrame(max_screen_res)

In [28]: 1 del df1["Screen Resolution"]

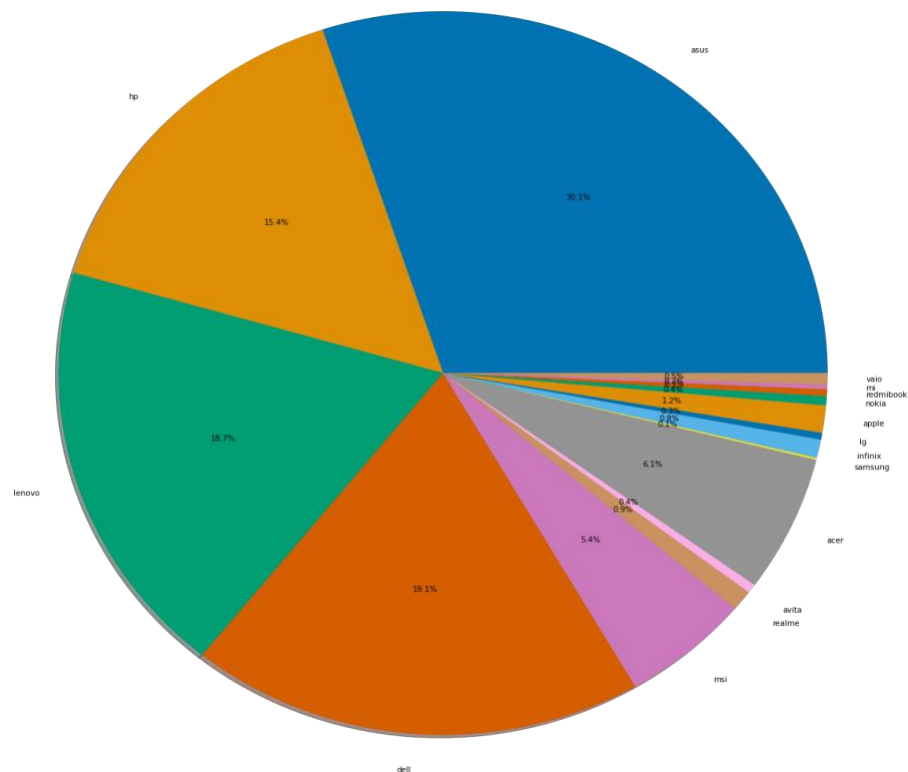
In [29]: 1 def gpu_comp(s):
2
3     if s == "nan":
```

# Results & Visualizations

## > Companies Available in Market

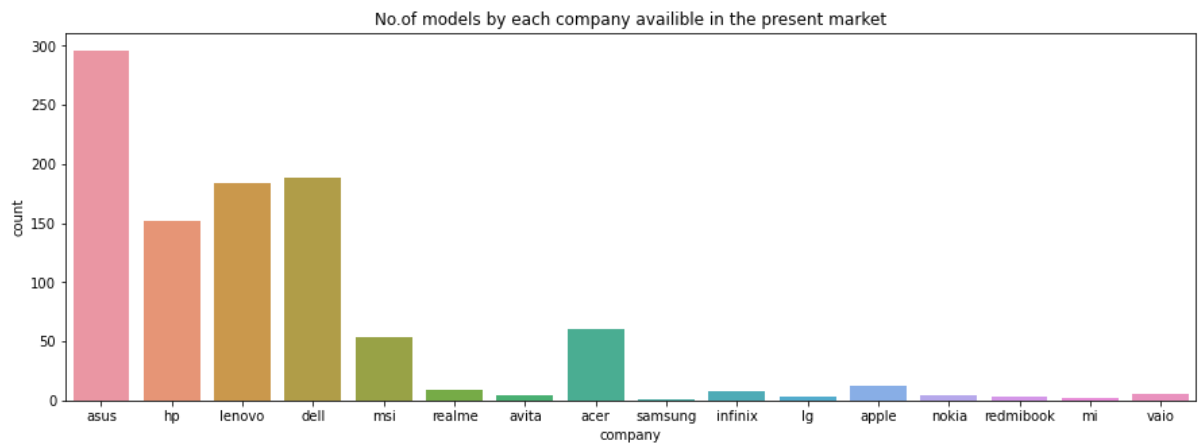


Here, we can see the companies available in the market with the competitive price points.

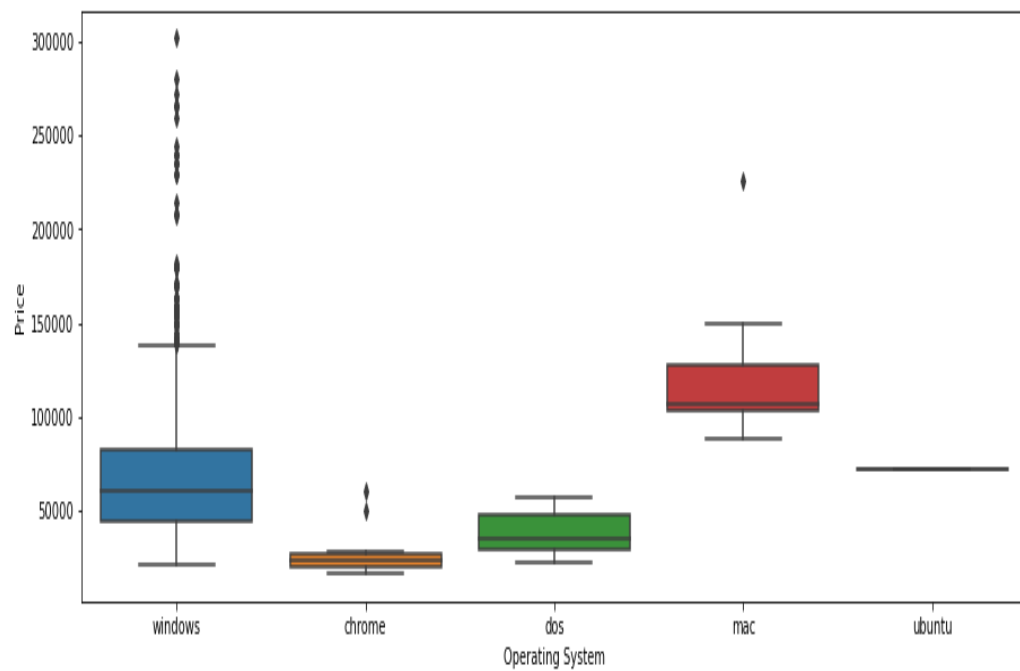




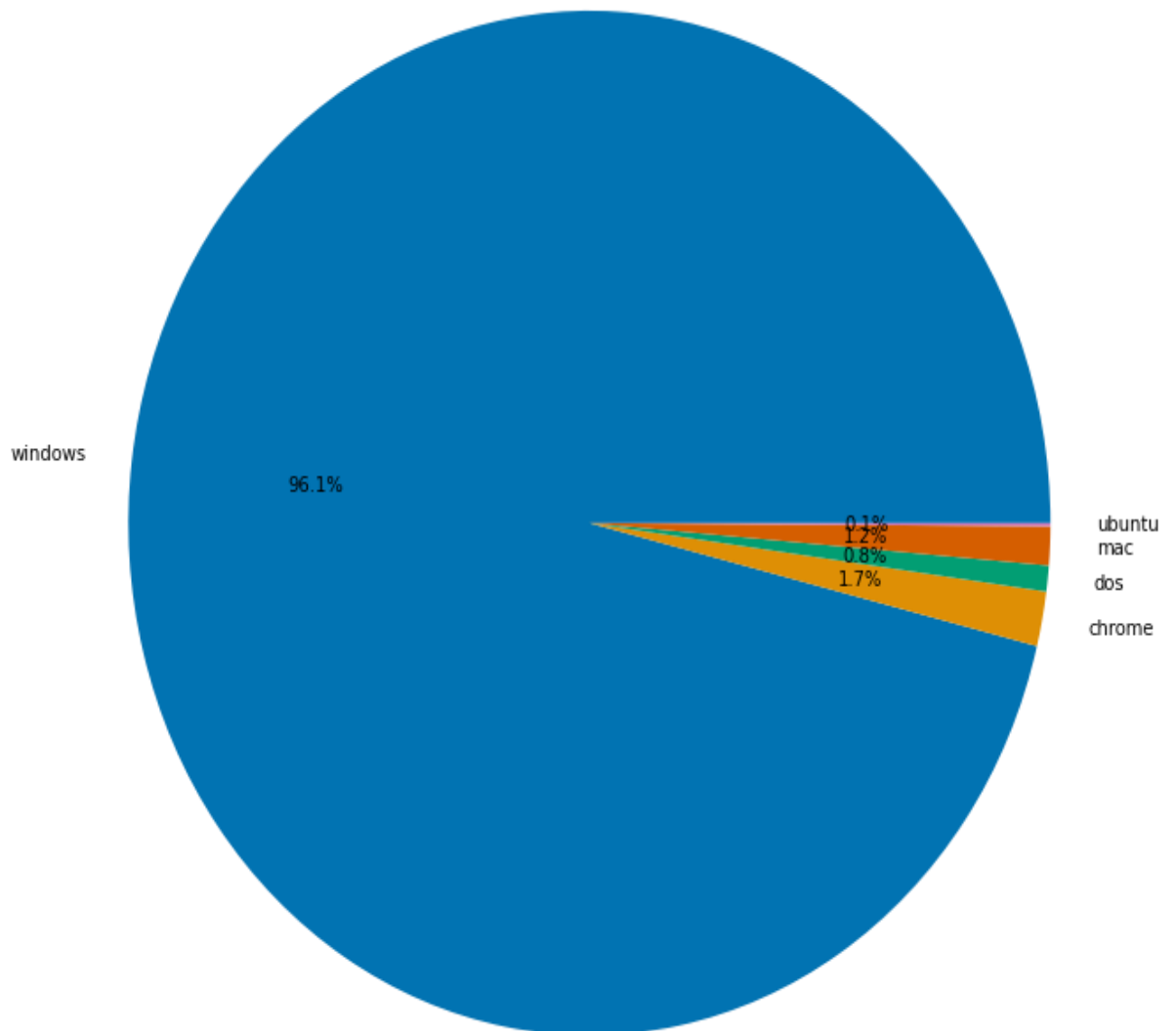
## ➔ Number of Models Available in Present Market



## ➔ OS Price Variation

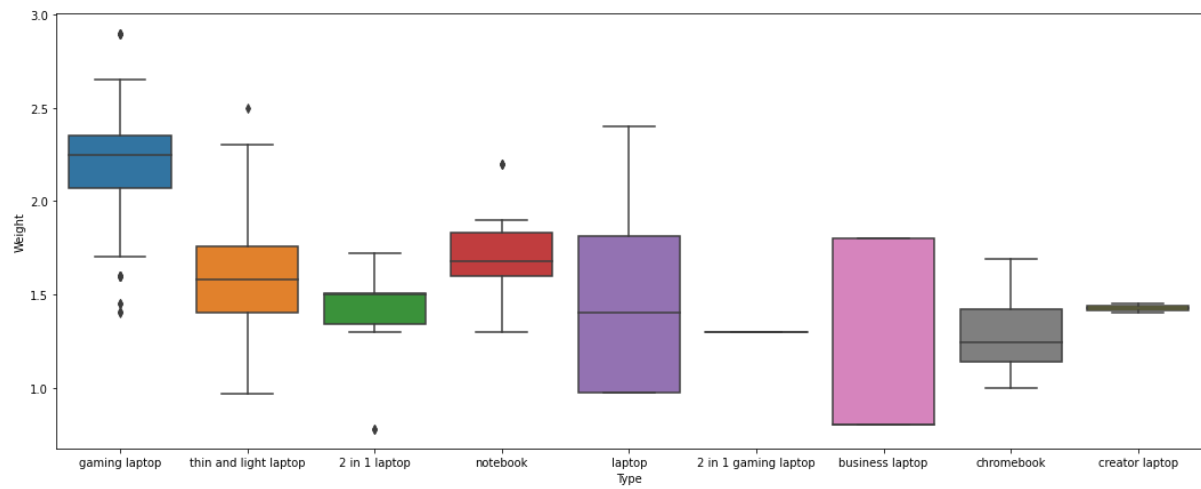


➔ Most Available OS in Market Place

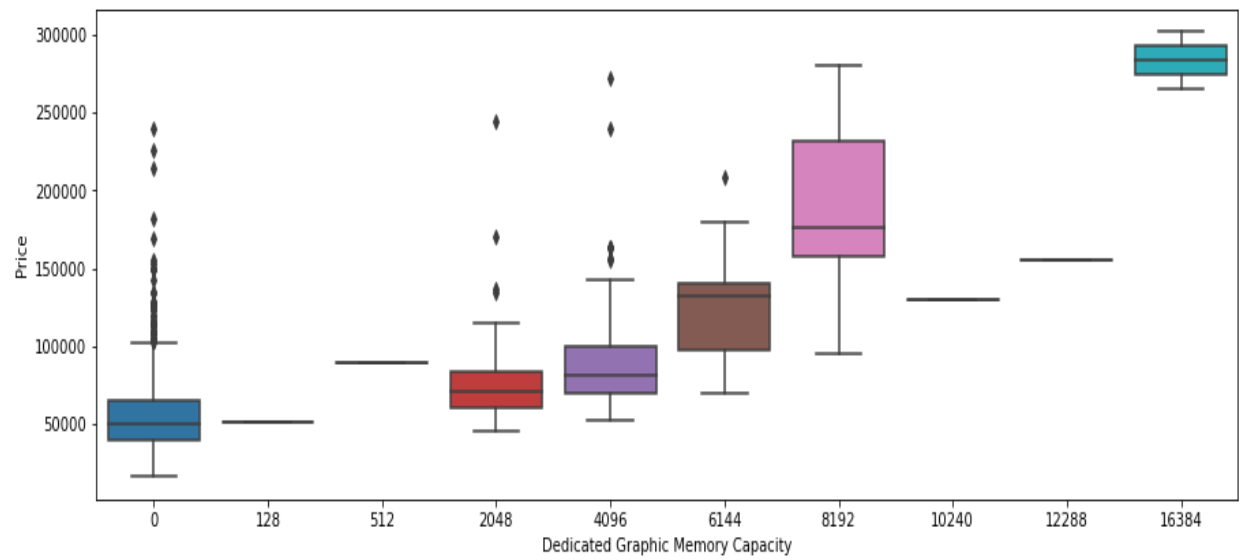


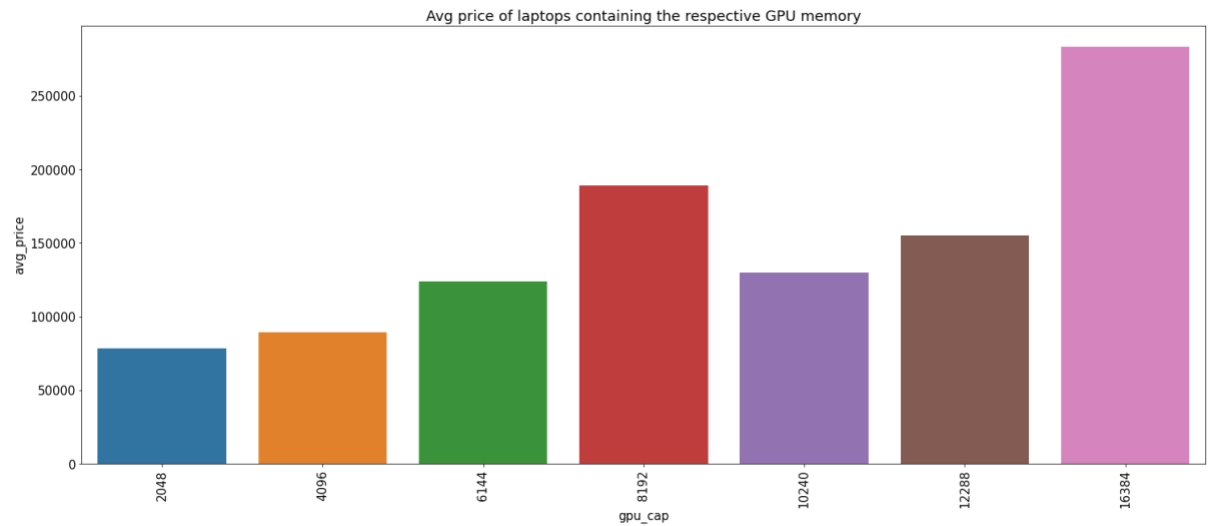
We could see that, the MAC OS, or apple laptops are generally higher priced than the other ones, while Windows OS is the most available in the market with a lot of price variation

➔ Market Place Laptop Weights Variations



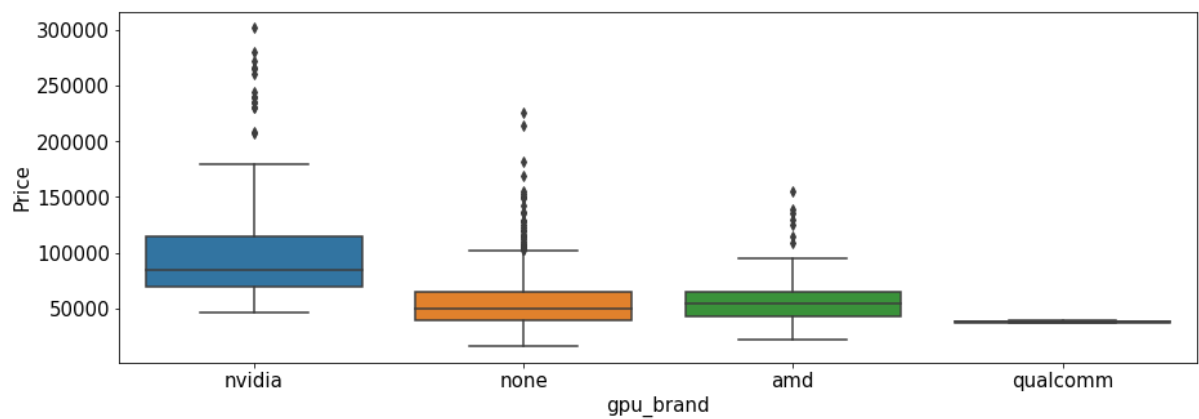
## ➔ GPU Memory Capacity and Price Variactions



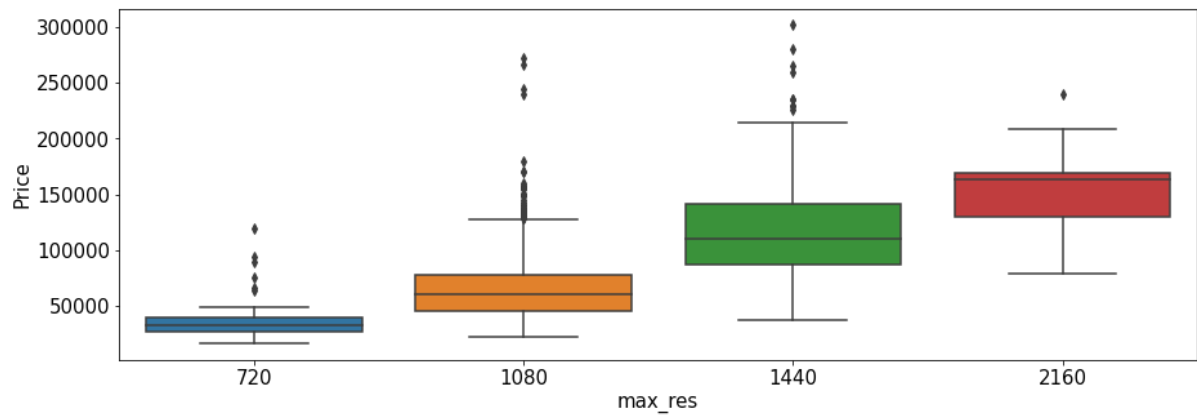


Note that, the GPU capacity isn't a very great indicator of the Graphics performance of a laptop as the no. of cores and specific model actually matters in terms of performance.

## ➔ GPU Brands and Their Costs

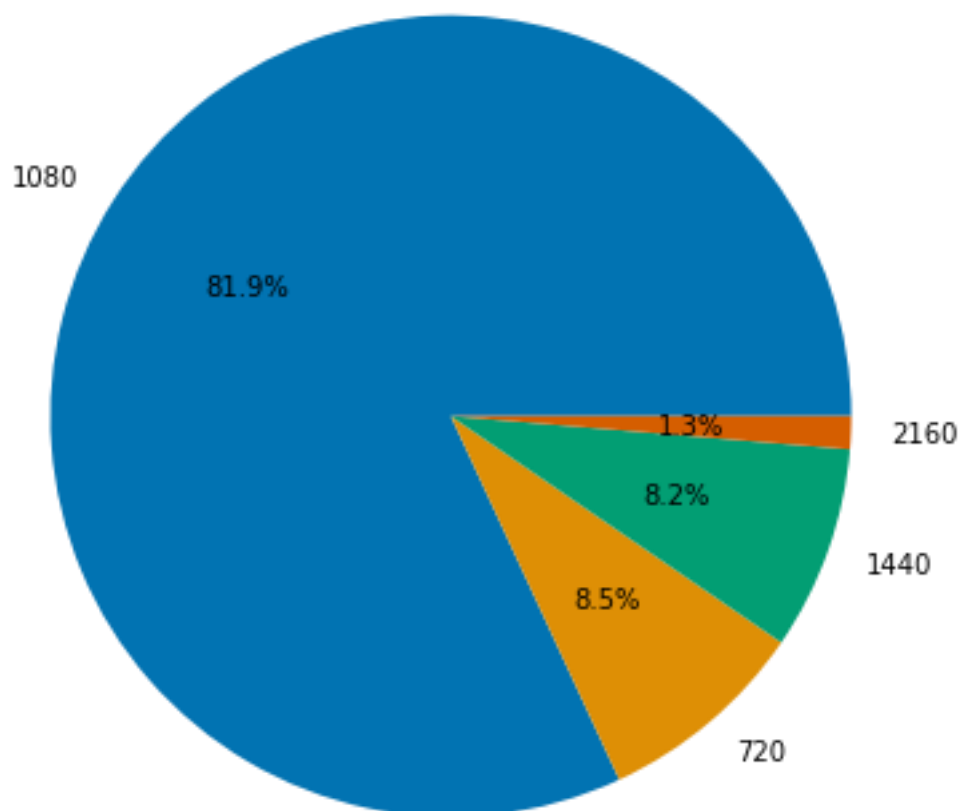


➔ Screen resolution and price variation

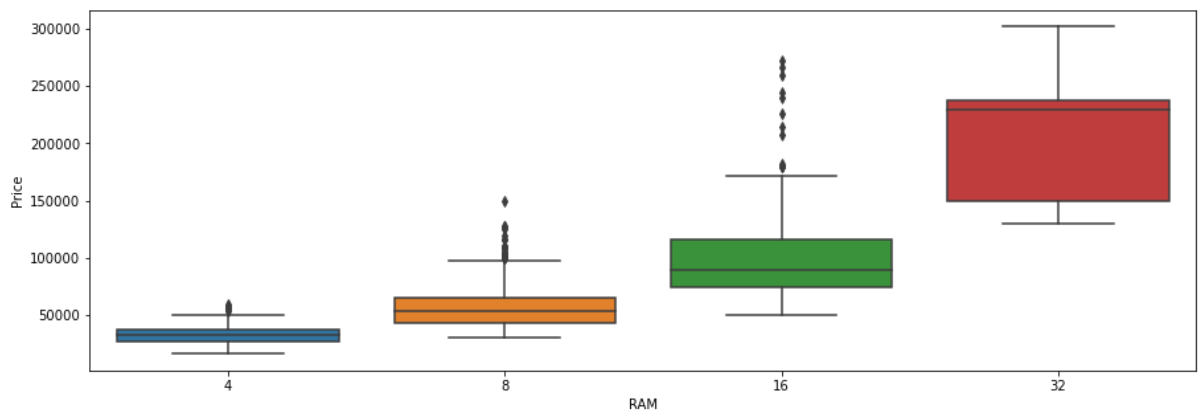


➔ Distribution of the available screen resolution in laptops

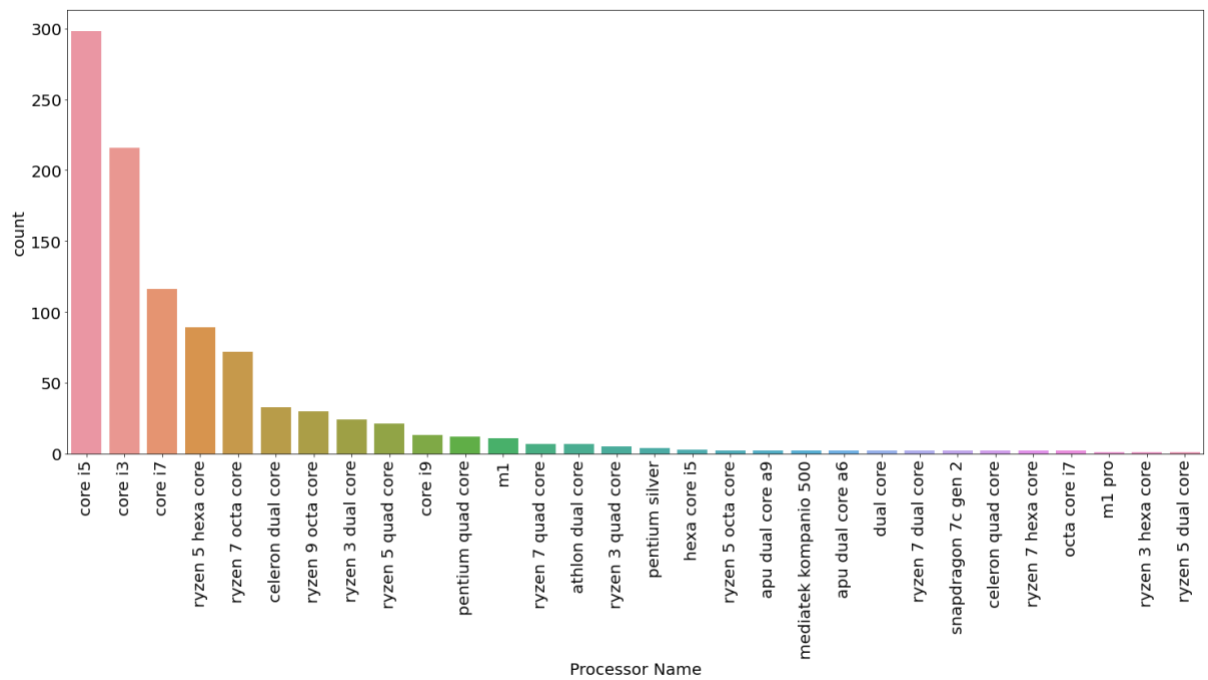
Distribution of the available screen resolution in laptops



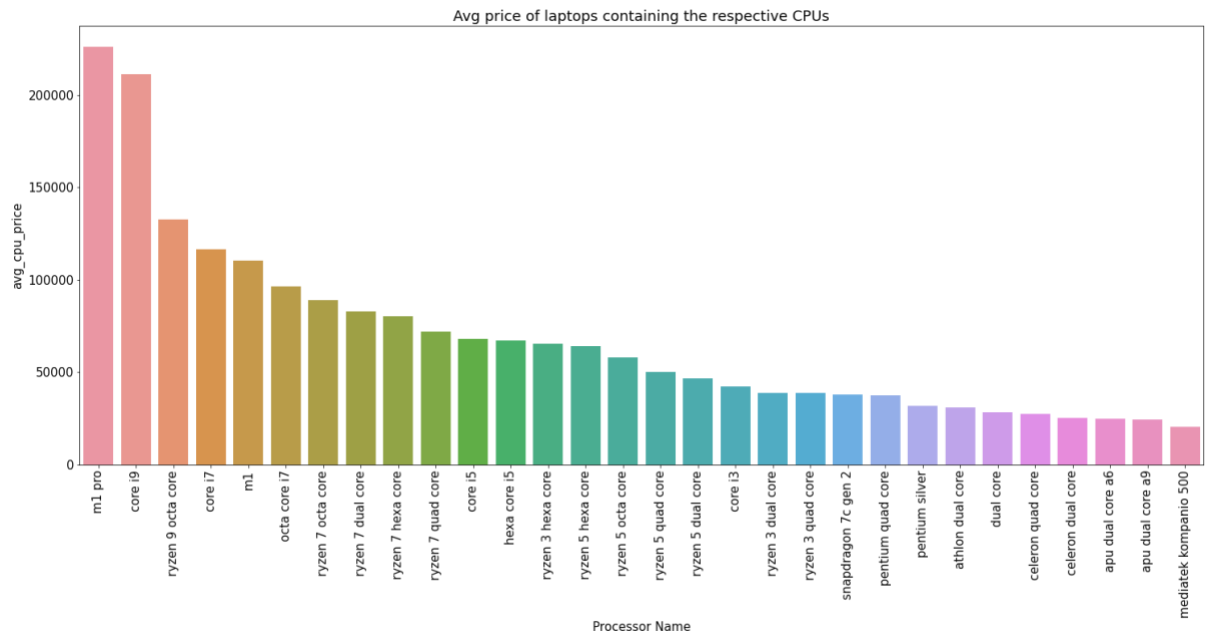
➔ Amount of RAM and it's price variation



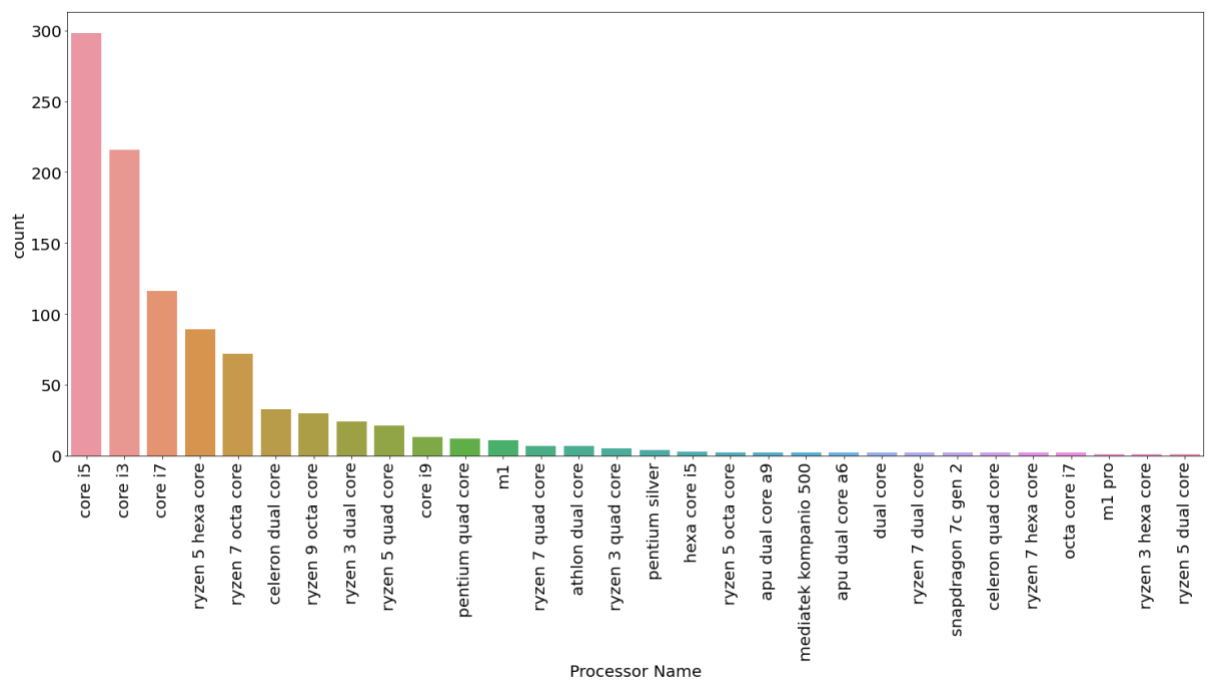
➔ Most frequent CPU processors which are used in the market (May-June 2022)



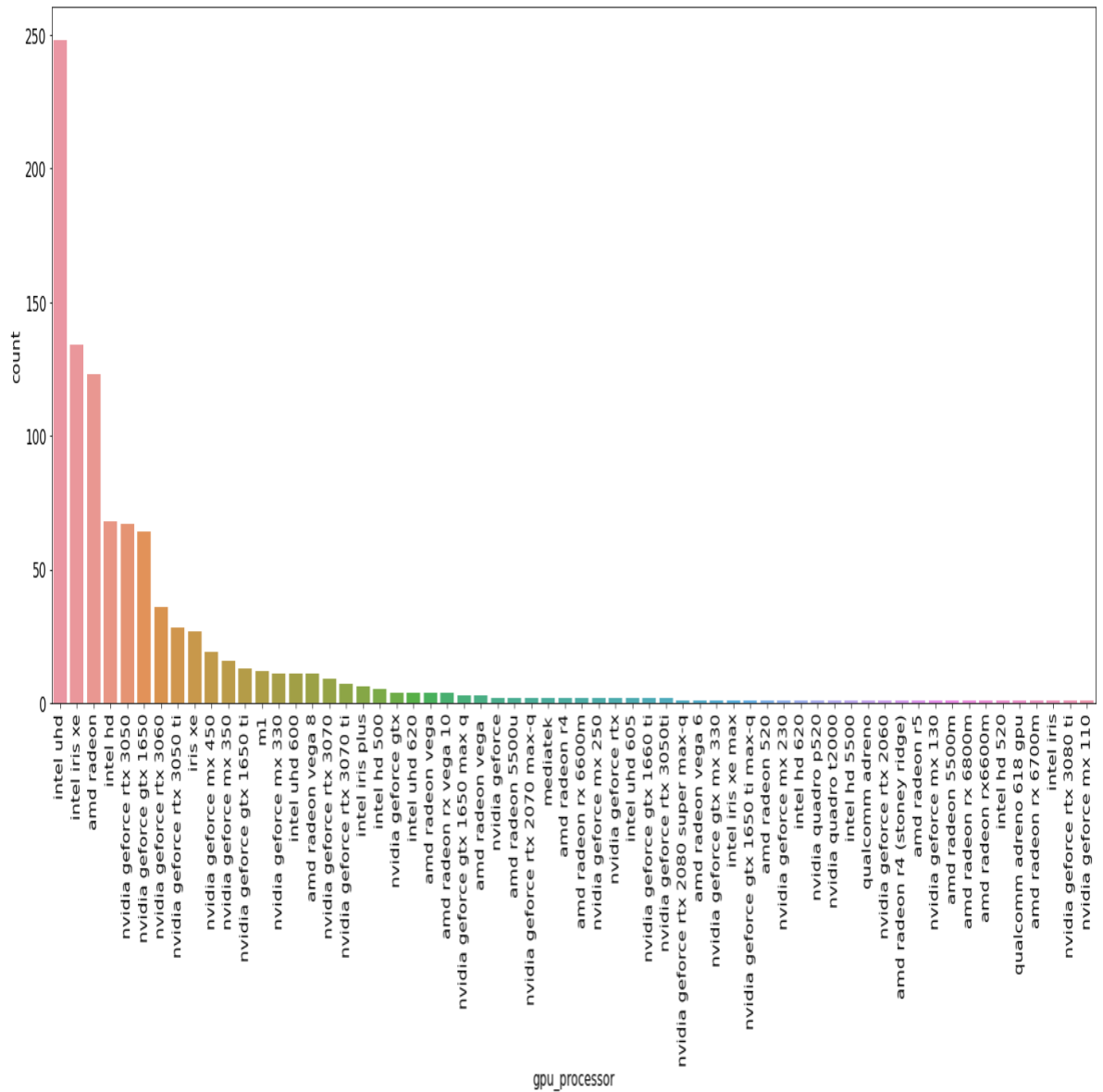
➔ Price variation of CPU processors which are available in the market (July-May 2022)



➔ Most frequent GPU processors which are used in the market (July-May 2022)



➔ Price variation of GPU processors which are available in the market (July-May 2022)





# Conclusion

After text mining and data visualization we can see which laptops are being purchased more and basis on that we got the ideas on which platform we should work and make a target market.

This will help provide context for our findings and remind readers of the goals of our project. This could include patterns or trends that emerged, significant differences between groups of texts, or unexpected results that we did not anticipate.

As example we can take away from this project is that, if we wanted to make a software or a laptop based app then we should go for windows based platform first, where gpu memory is moderately high with higher resolution screens.

# Reference

Data Collected from –

<https://www.kaggle.com/code/sreevaatsavbavana/market-analysis-of-laptops-june-2022/notebook>

Coding Help from –

<https://www.youtube.com/watch?v=HiOtQMcl5wg>

<https://github.com/AlexTheAnalyst>

[https://www.udemy.com/course/learning-python-for-data-analysis-and-visualization/?ranMID=39197&ranEAID=E7g%2FxD%2FWGM&ranSiteID=E7g\\_xDr\\_WGM-ijq56neXAHECQAOWf6bAXw&LSNPUBID=E7g%2FxD%2FWGM&utm\\_source=aff-campaign&utm\\_medium=udemyads](https://www.udemy.com/course/learning-python-for-data-analysis-and-visualization/?ranMID=39197&ranEAID=E7g%2FxD%2FWGM&ranSiteID=E7g_xDr_WGM-ijq56neXAHECQAOWf6bAXw&LSNPUBID=E7g%2FxD%2FWGM&utm_source=aff-campaign&utm_medium=udemyads)