**ĐẠI HỌC HUẾ**

# KHOA KỸ THUẬT VÀ CÔNG NGHỆ

🙠🙟🕮🙝🙢

****

**BÁO CÁO** **ĐỒ ÁN**

**PTDL R**

**NĂM HỌC 2021-2022**

**Giảng viên hướng dẫn: HỒ QUỐC DŨNG**

**Lớp: KHDL & TTNT**

| Số phách  *(Do hội đồng chấm thi ghi)* |
| --- |

**Thừa Thiên Huế, ngày …tháng…năm.....**

**ĐẠI HỌC HUẾ**

# KHOA KỸ THUẬT VÀ CÔNG NGHỆ

🙠🙟🕮🙝🙢

****

(MẪU BÌA PHỤ)

**BÁO CÁO** **ĐỒ ÁN  
PTDL R**

**NĂM HỌC 2021-2022**

**Giảng viên hướng dẫn: HỒ QUỐC DŨNG**

**Lớp: KHDL & TTNT**

**Sinh viên thực hiện: HOÀNG TRỌNG TOÀN**

*(ký tên và ghi rõ họ tên)*

| Số phách  *(Do hội đồng chấm thi ghi)* |
| --- |

**Thừa Thiên Huế, ngày …tháng…năm.....**

# Mục Lục

# 

[**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**](#_30j0zll) **1**

[**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**](#_1fob9te) **2**

[**Mục Lục**](#_k2gbokutxygv) **3**

[**TỔNG QUAN**](#_5gulql9fq5es) **4**

[**CHƯƠNG I: THU THẬP DỮ LIỆU**](#_rrhi4r85kj08) **5**

[Nguồn dữ liệu](#_e6yqyut1io3s) 5

[Lấy Dữ liệu](#_v3moy6birk1t) 6

[**CHƯƠNG II: PHÂN TÍCH DỮ LIỆU**](#_5v5mptb2szvg) **8**

[Tổng quan dữ liệu](#_5hswtzwvx12h) 8

[Trực quan hóa dữ liệu](#_m06s3xhf1qod) 10

[Đánh giá Price](#_qfbt1gs0gbtf) 10

[Đánh giá Scrsize](#_vvx0y3foufdy) 10

[Đánh giá số lượng nhãn hàng](#_ao83epfrs4m7) 11

[Đánh giá về RAM](#_bfife5fjhpq9) 12

[Đánh giá về độ phổ biến của nhãn hàng](#_ddo15nv2hwv8) 13

[Thống kê suy diễn](#_b0yhd595zzn5) 14

[Tổng quan](#_26erlzld4wws) 14

[Liệu dung lượng ổ cứng có ảnh hưởng đến giá cả?](#_jgzudc7h62qh) 15

[Liệu nhãn hàng có ảnh hưởng đến giá cả?](#_a8rn0igmgd3j) 16

[**KIỂM TRA ĐẠO VĂN**](#_tg9jjckwuu0u) **17**

[**TÀI LIỆU THAM KHẢO**](#_dak4ucn5n2hg) **17**

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# TỔNG QUAN

Hiện nay con người đang tiến đến cách mạng 4.0 nên 1 thiết bị thông minh là nhu cầu cần thiết của mỗi chúng ta. Một chiếc laptop là 1 lựa chọn tốt cho việc hòa nhập cách mạng 4.0 hiện nay. Vậy mua laptop nên dựa trên những tiêu chí gì và những tiêu chí nào có thể ảnh hưởng đến giá cả của nó.

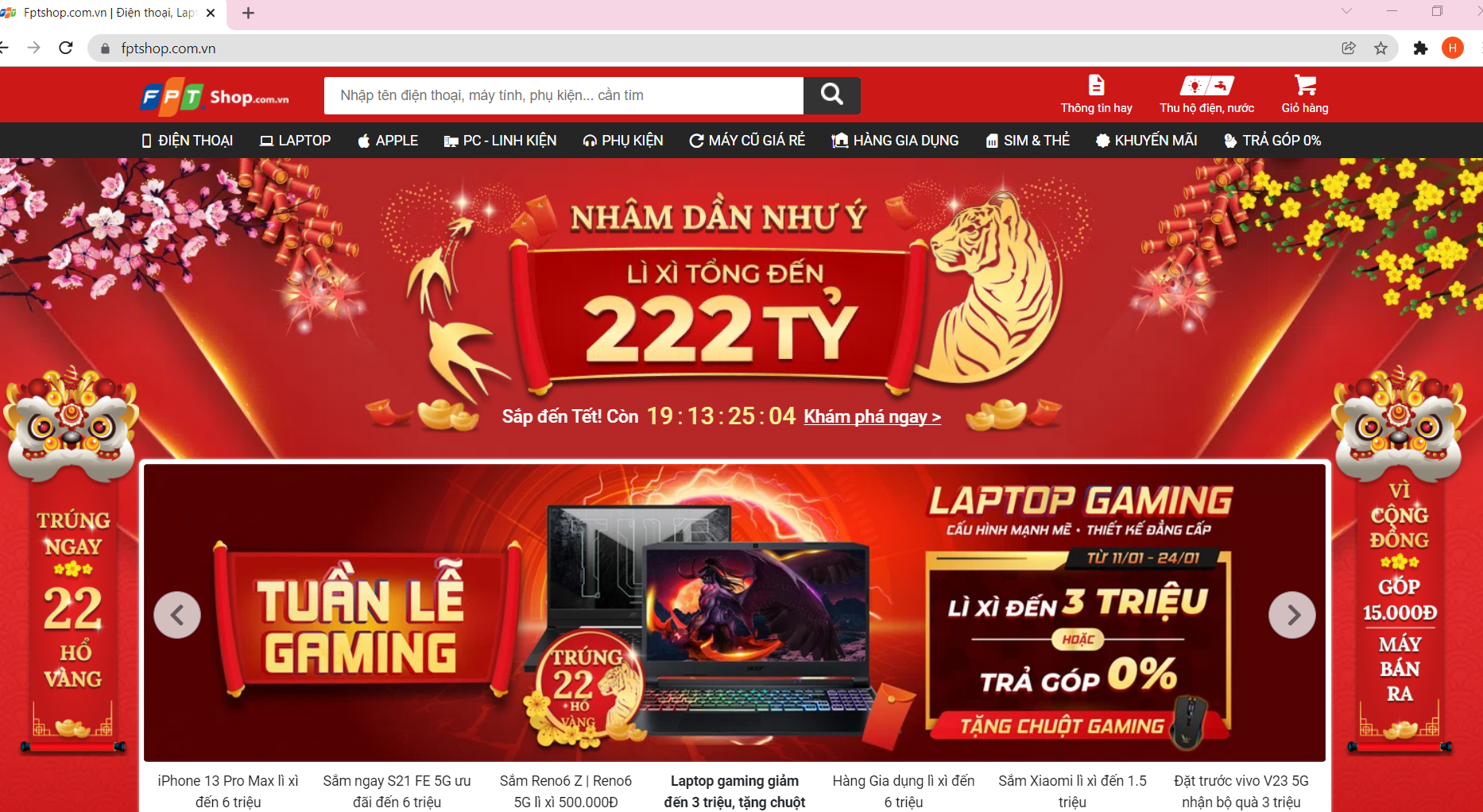


Việc sử dụng 1 số công cụ như Python, R để phân tích về giá của những chiếc laptop là điều cần thiết, có như thế ta mới chọn được 1 chiếc laptop phù hợp với công việc, kinh tế và sở thích.

# CHƯƠNG I: THU THẬP DỮ LIỆU

## Nguồn dữ liệu

Ngày nay có nhiều shop, cửa hàng, trang web mua bán về máy tính . Ở đây ta có thể chọn FPT shop để tiến hành phân tích dữ liệu.

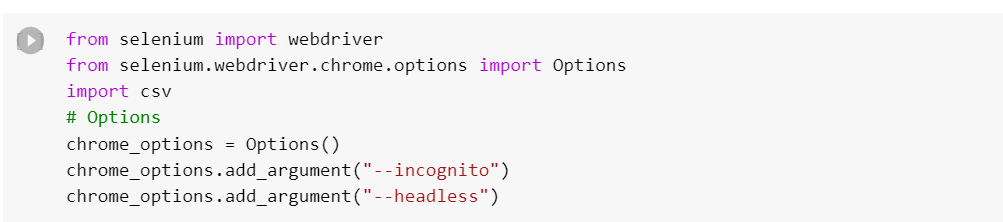


## Lấy Dữ liệu

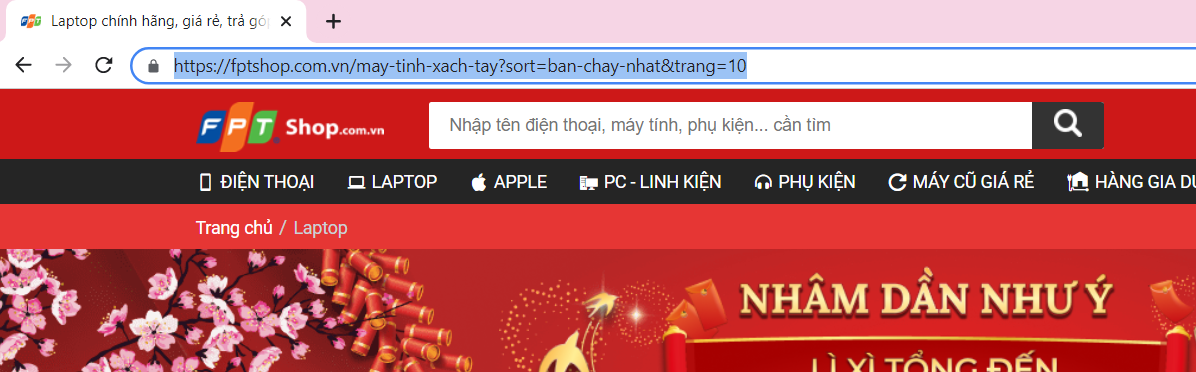
Ở chúng ta nên dùng selenium để tiến hành lấy dữ liệu vì nó là trình duyệt web tự động rất mạnh. Chọn google colab để xây dựng mã nguồn. để có thể sử dụng được selenium chúng ta cần chuẩn bị môi trường cho nó như sau:



Thay vì ở Spyder hay Pycharm chúng ta cần phải download file selenium tương ứng với phiên bản của google thì ở google colab chúng ta chỉ cần gõ mã nguồn như vậy thì tự động phiên bản selenium tương thích với google sẽ được sử dụng. Đây cũng là 1 trong những ưu điểm của nó.

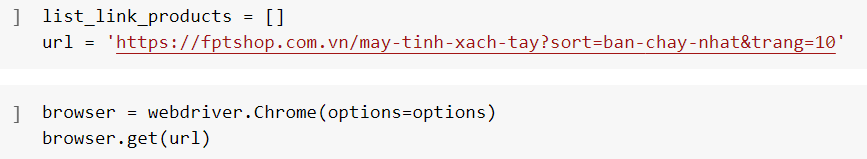
Tiếp theo ta nên chọn 1 số thư viện cần dùng trong quá trình kéo dữ liệu về đồng thời setup 1 số tham số cho thư thư viện selenium.

Bây giờ chúng ta cần đường dẫn vào trang cần lấy dữ liệu.



Lưu ý rằng chúng ta phải cho nó tải đến trang 10 để hiển thị được hết các laptop cần kéo dữ liệu.

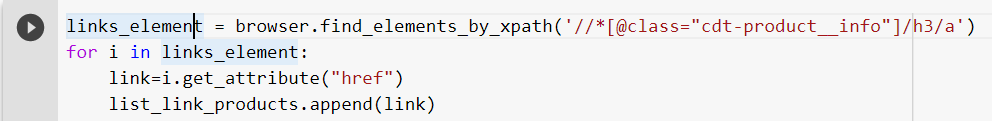
Tiến hành truy cập vào web.



Tiếp đến chúng ta cần xpath, class, … những dữ liệu nào có thể định vị được vị trí của mục tiêu cần lấy dữ liệu. ví dụ như sau

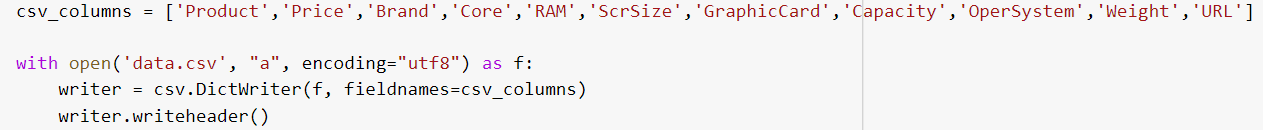


Ở đây chúng ta cần lấy link sản phẩm trong thẻ “href”, chúng ta có thể làm như sau



Ta thấy được trình duyệt sẽ truy cập vào element thông qua xpath ta đã cung cấp và từ đó lấy thẻ “href” rồi thêm vào 1 list nào đó thông qua lệnh get\_attribute và lệnh append.

Tiếp đến cần tạo các cột trong file csv như sau:



Sở dĩ chọn file csv là khi lưu ở dạng csv thì có thể loại bỏ được mọi định dạng từ đó để file dữ liệu của ta nhẹ nhất có thể vì đôi khi chúng ta kéo dữ liệu từ những trang lớn hơn thì khối lượng dữ liệu rất lớn ảnh hưởng đến việc thao tác file.

Phần còn lại của chúng ta là truy cập vào từng link trong list có sẵn, sau đó truy cập đến từng tham số chúng ta cần lấy. Lấy chúng ra rồi để vào 1 dic tên data và từ đó có thể hoàn thành việc lấy dữ liệu thông qua hàm viết dic vào file csv.

# CHƯƠNG II: PHÂN TÍCH DỮ LIỆU

## Tổng quan dữ liệu

Ở đây dữ liệu chúng ta được chia thành 10 cột sau khi qua xử lí:

Product: tên sản phầm

Price: giá

Brand: hãng

Core: chip

RAM

Scrsize: kích thước màn hình

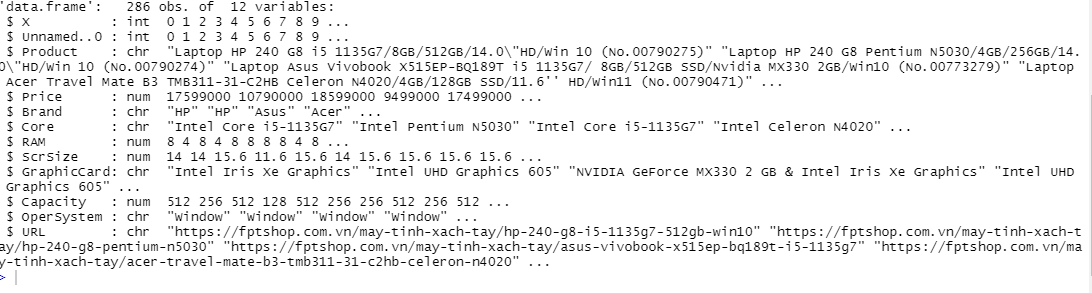
GraphicCard: Card đồ họa

Capacity: dung lượng ổ cứng

Opersystem: hệ điều hành

URL: link để truy cập sản phẩm

Thống kê các biến bằng R:



Nhận xét:

Đối với các biến như Hãng, hệ điều hành ,Ram, Chip, dung lượng ổ cứng, card đồ họa là định tính.

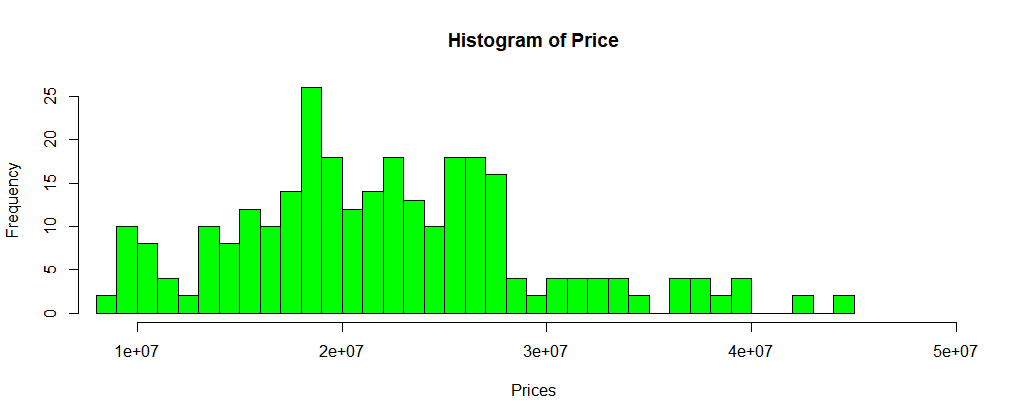
Đối với các biến như giá, độ lớn màn hình là định lượng.

Kết luận: Ở đây mọi biến đều có thể ảnh hưởng đến giá cả ngoại trừ link và tên sản phẩm.

Tương đối phổ biến là kích cỡ 13-14 inch.

## Trực quan hóa dữ liệu

### Đánh giá Price



Ở đây dùng histogram để vẽ biểu đồ phân bố của giá cả.

Nhận xét:

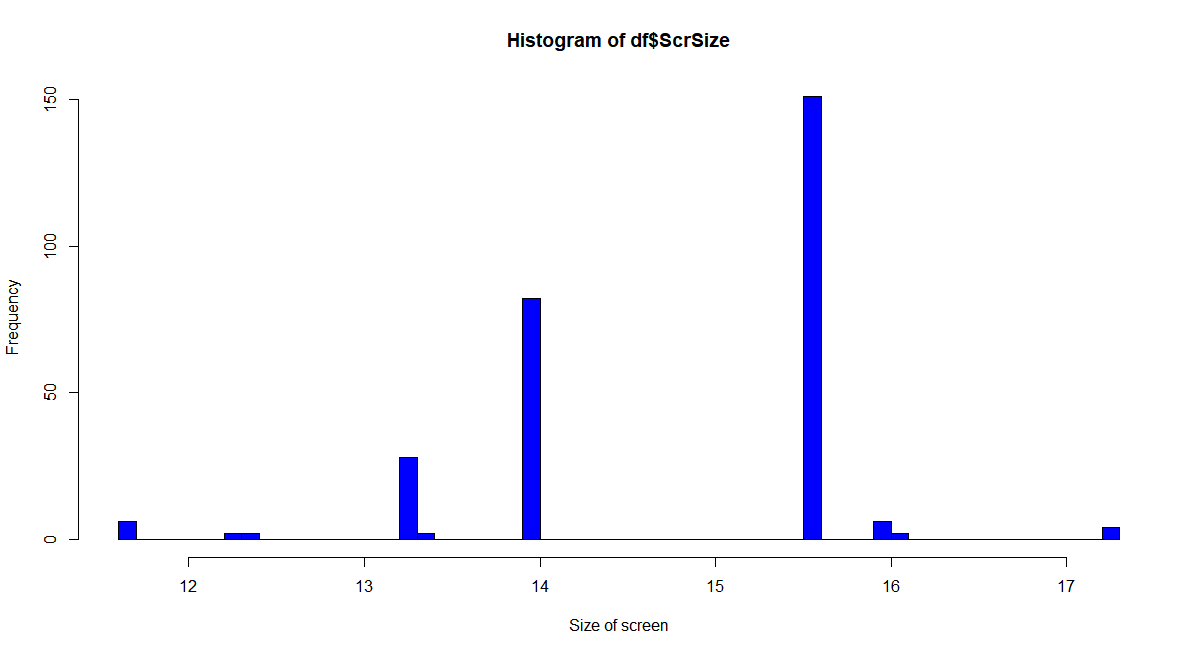
Giá laptop phổ biến giá động từ 8 - 45tr

Giá laptop phổ biến nhất rơi vào khoảng từ 15-28 tr

Giá laptop trước ngưỡng gái 10tr có số lượng khá it.

Giá laptop cực kì thưa thớt khi vượt ngưỡng 40tr.

### Đánh giá Scrsize



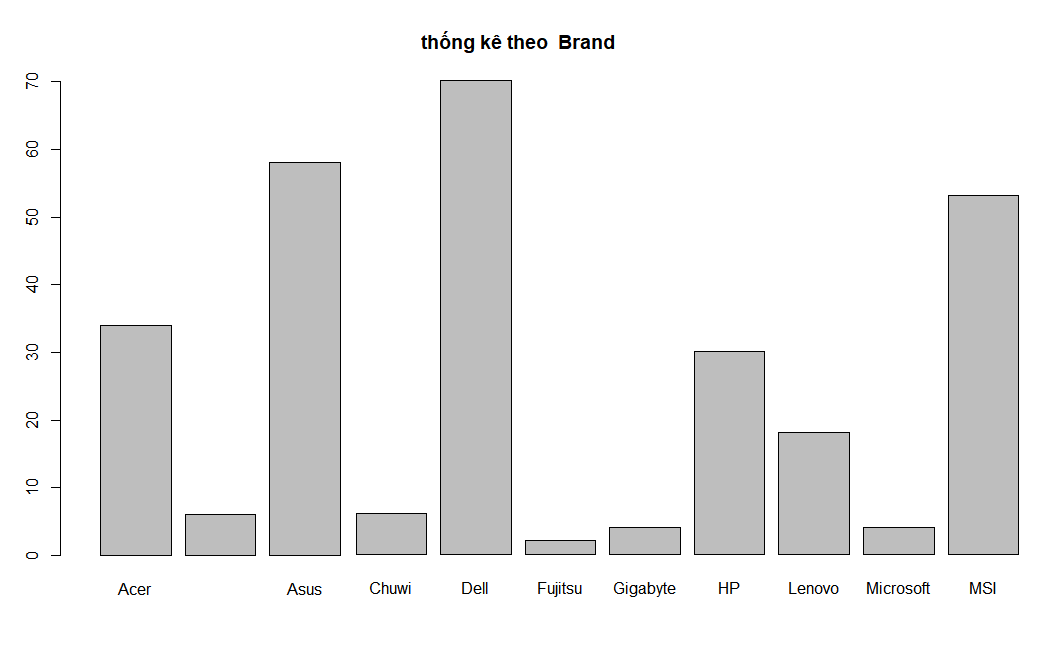
Dùng histogram để vẽ vùng phân bố của kích cỡ màn hình:

Nhận xét:

Dễ dàng nhận thấy rằng trước ngưỡng 13 và sau ngưỡng 16 có số lượng laptop rất ít, nó chỉ chiếm số rất nhỏ,

Vùng phân bố rộng nhất của laptop rơi cào cỡ màn hình từ 13-16.2 inch.

### Đánh giá số lượng nhãn hàng



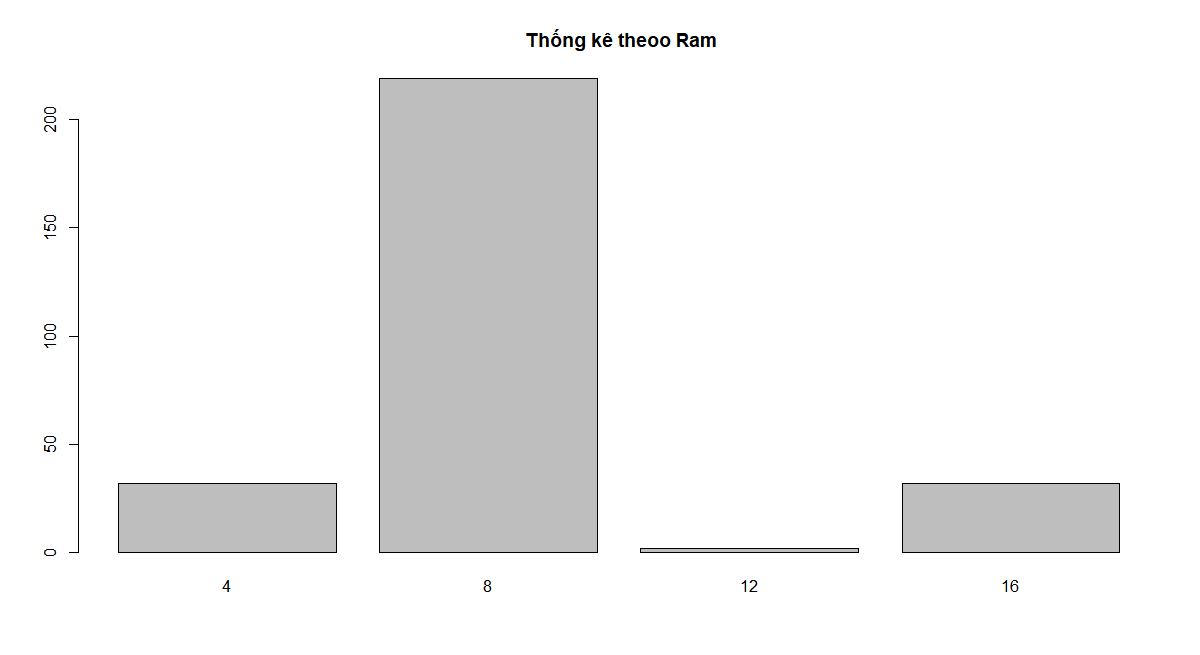
Nhận xét:

Ở đây có thể thấy được số lượng máy của các hãng trong shop dẫn đầu là các nhãn hàng Dell, Asus, MSI trong đó Dell dẫn đầu.

Nhòm tiếp theo gồm Acer, HP, Lenovo tới số lượng máy mỗi hãng giao động trong khoảng từ 20- dưới 40 máy.

Phần còn lại là các máy xuất hiện cực ít với mỗi nhãn hàng Chuwi, Fuijtsu, Gigabity, Microsoft dưới 10 máy.

### Đánh giá về RAM



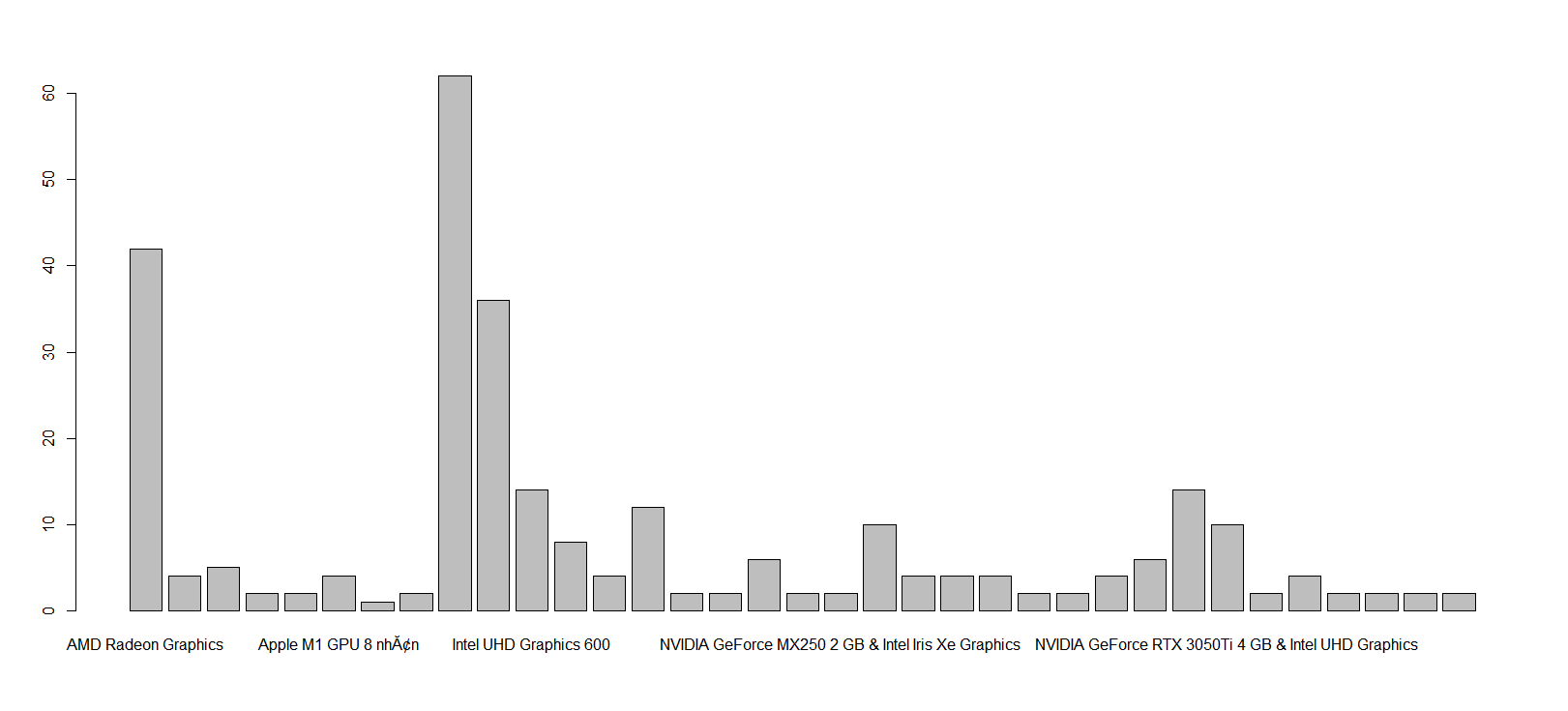
Nhận xét:

Số lượng máy có 8Gb Ram có số lượng rất lớn với trên 200 máy.

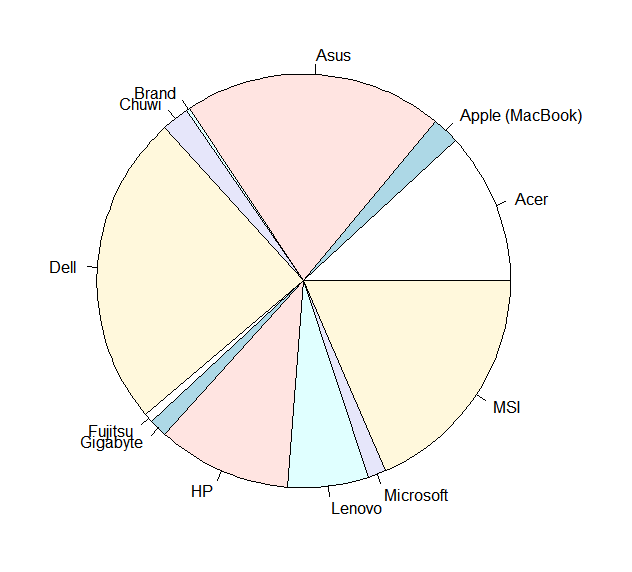
Số lượng máy có 4Gb Ram và 16Gb Ram có số tương đương nhau với trên 30 máy.

Số lượng máy có 12Gb Ram chiếm số lượng cực kì ít.

Ở đây ta có thêm biểu đồ thể hiện mức độ phổ biến của các loại card đồ họa.



### Đánh giá về độ phổ biến của nhãn hàng

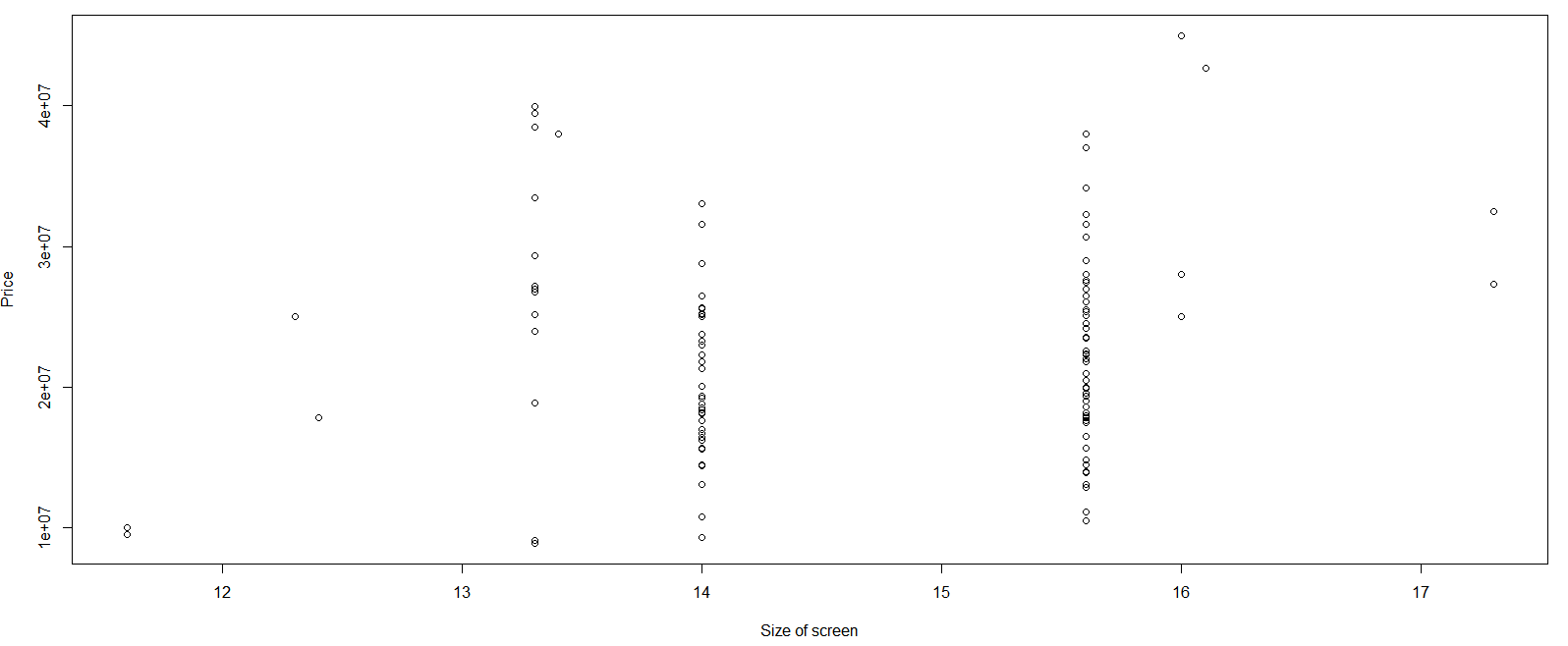


Nhận xét:

Độ phổ biến của 3 nhãn hàng MSI, Dell, Asus là lớn nhất

Độ phổ biến trung bình thuộc về 3 nhãn hàng Acer, HP, Lenovo

Đánh giá phân bố của giá cả theo kích cỡ màn hình



Nhận xét:

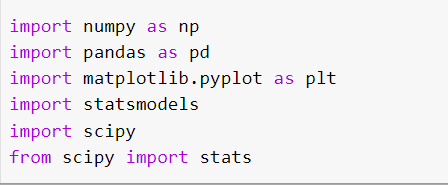
Phổ biến nhất vẫn là kích cỡ màn 14- dưới 16 inch(10-40tr).

Tương đối phổ biến là kích cỡ 13-14 inch (8-40tr).

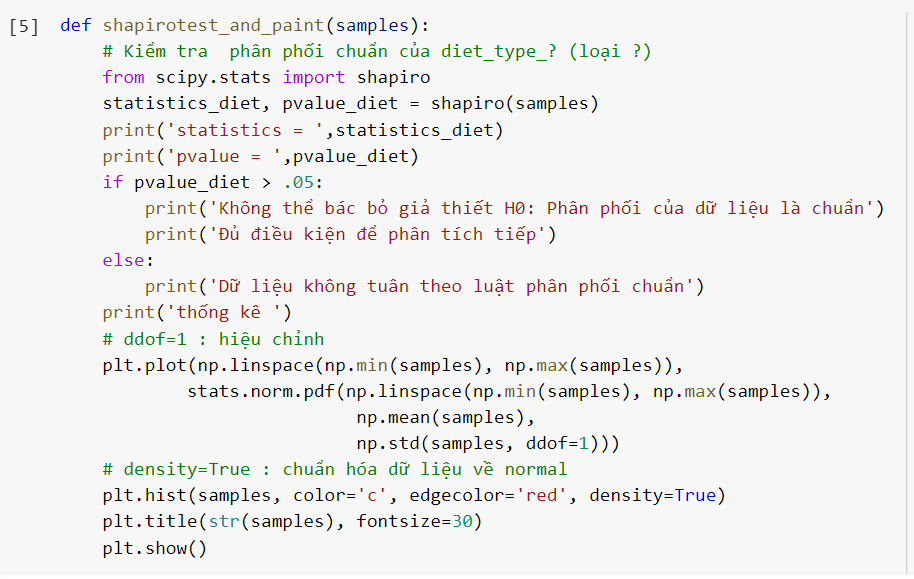
## Thống kê suy diễn

### Tổng quan

Ở đây chúng ta dùng Python vì nó cung cấp rất nhiều thư viện hữu ích cho việc này. Chúng ta cần gọi ra những thư viện sau.

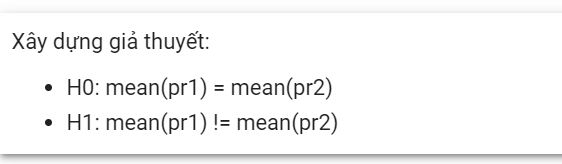


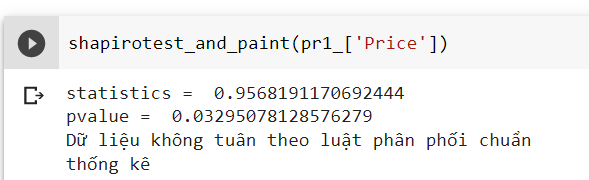
Thiết lập hàm để kiểm định phân phối chuẩn để thực hiện ANOVA.

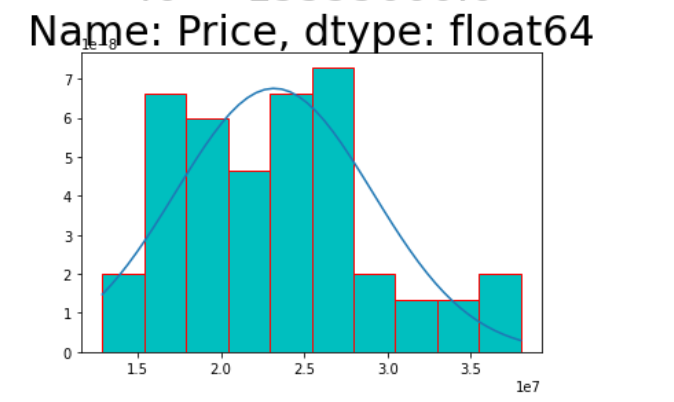


Sau khi đọc và làm sạch các dữ liệu chúng ta tiến vào phân tích.

### Liệu dung lượng ổ cứng có ảnh hưởng đến giá cả?

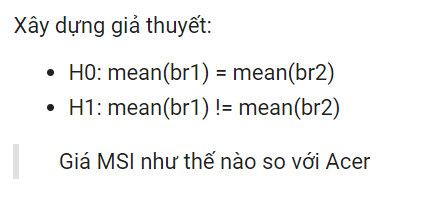


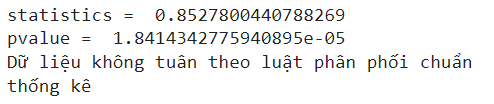


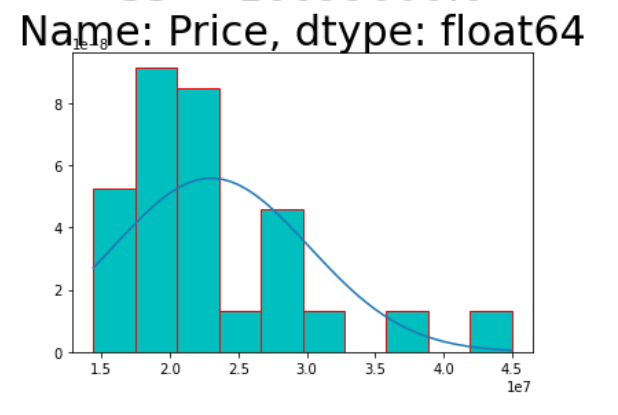


Như vậy không thể tiến hành phân tích anova do mẫu không phải phân phối chuẩn.

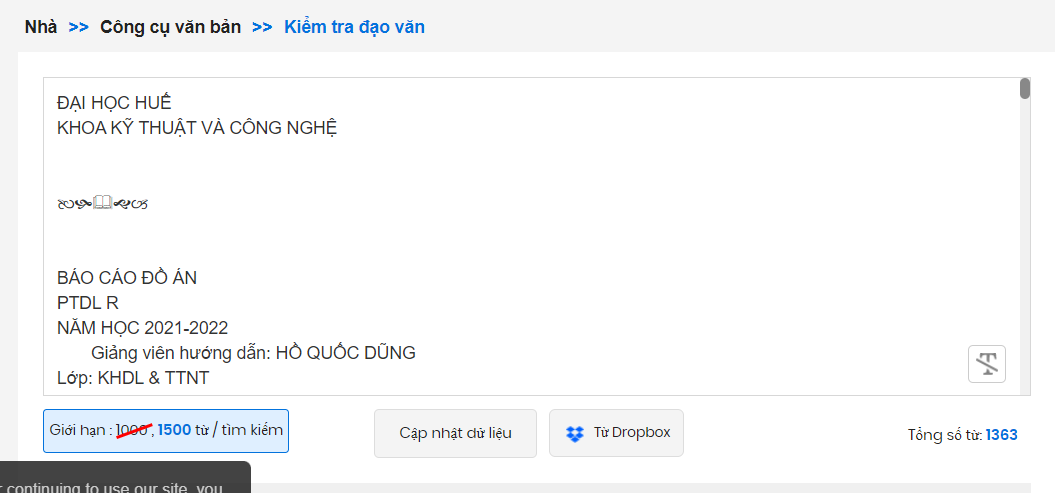
### Liệu nhãn hàng có ảnh hưởng đến giá cả?







# KIỂM TRA ĐẠO VĂN





# TÀI LIỆU THAM KHẢO

Trang 5: [link](https://images.search.yahoo.com/images/view;_ylt=AwrOueSYSt5hnUcveoqJzbkF;_ylu=c2VjA3NyBHNsawNpbWcEb2lkAzIwMjBmY2U2MDY5MGNiNjExNjExZWM5M2RkOWY0YjJhBGdwb3MDMTA5BGl0A2Jpbmc-?back=https%3A%2F%2Fimages.search.yahoo.com%2Fsearch%2Fimages%3Fp%3Dlaptop%26type%3DE211US826G91508%26fr%3Dmcafee%26fr2%3Dpiv-web%26nost%3D1%26tab%3Dorganic%26ri%3D109&w=1600&h=900&imgurl=uwaterloo.ca%2Farts-computing-newsletter%2Fsites%2Fca.arts-computing-newsletter%2Ffiles%2Fuploads%2Fimages%2F4-laptops-dan-02.jpg&rurl=https%3A%2F%2Fuwaterloo.ca%2Farts-computing-newsletter%2Fwinter-2016%2Ffeature%2Frecommended-laptops-students&size=181.2KB&p=laptop&oid=2020fce60690cb611611ec93dd9f4b2a&fr2=piv-web&fr=mcafee&tt=Recommended+Laptops+for+Students+%7C+Arts+Computing+Office+Newsletter+%7C+University+of+Waterloo&b=61&ni=21&no=109&ts=&tab=organic&sigr=_YL.GmK.3wf.&sigb=NxuQNPpLgzC9&sigi=e8HXqJfNKdPe&sigt=FmXNv7LKubO0&.crumb=xken5IF36Y5&fr=mcafee&fr2=piv-web&type=E211US826G91508)