

ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ





# BÁO CÁO

# ĐỒ ÁN

**Học kỳ I, năm học 2021 – 2022**

### **Học phần:** Phân tích dữ liệu với R

***Đề tài:*** Phân tích dữ liệu sản phẩm giày nam trên Lazada

**Số phách**

*(Do hội đồng chấm thi ghi)*

Thừa Thiên Huế, ngày …tháng…năm....



ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ





# BÁO CÁO

## ĐỒ ÁN

**Học kỳ I, năm học 2021 – 2022**

### **Học phần:** Phân tích dữ liệu với R

***Đề tài:*** Phân tích dữ liệu sản phẩm giày nam trên Lazada

###### 

###### Giảng viên hướng dẫn: Hồ Quốc Dũng

###### Sinh viên thực hiện: Nguyễn Chính Phi Hùng

###### Lớp: KHDL&TTNT

**Số phách**

*(Do hội đồng chấm thi ghi)*

Thừa Thiên Huế, ngày …tháng…năm.....

**MỤC LỤC:**

[I. MÔ TẢ VỀ DỮ LIỆU. 5](#_Toc92294709)

[**1. Cách lấy dữ liệu: 5**](#_Toc92294710)

[**2. Số lượng dòng dữ liệu: 6**](#_Toc92294711)

[**3. Chuyển đổi giá trị: 6**](#_Toc92294712)

[II. PHÂN TÍCH THÔNG KÊ MÔ TẢ SỐ LIỆU. 7](#_Toc92294721)

[**1. Tính các chỉ số thống kê. 7**](#_Toc92294722)

[**2. Vẽ các biểu đồ để trực quan hóa dữ liệu. 7**](#_Toc92294723)

[a. Biểu đồ Dot Plost: 7](#_Toc92294724)

[b. Biểu đồ Histogram : 7](#_Toc92294726)

[c. Biểu đồ Pie Charts : 8](#_Toc92294727)

[d. Biểu đồ Bar Plot: 9](#_Toc92294728)

[e. Biểu đồ Box plosts : 10](#_Toc92294729)

[f. Nhận xét : 11](#_Toc92294730)

[III. Ý tưởng phân tích số liệu từ Kaggle 11](#_Toc92294731)

[**1.** **Import các thư viện :** **11**](#_Toc92294732)

[**2.** **Đọc các file dữ liệu :** **11**](#_Toc92294733)

[**3.** **Làm sạch dữ liệu:** **12**](#_Toc92294734)

[**4.** **Phân tích:** **13**](#_Toc92294735)

[a.Phân tích đơn biến: 13](#_Toc92294736)

[b.Phân tích đa biến: 15](#_Toc92294737)

[**5.** **Kết luận:** **16**](#_Toc92294739)

# I. MÔ TẢ VỀ DỮ LIỆU.

## 1. Cách lấy dữ liệu:

-Sử dụng Visual Studio Code để crawl dữ liệu trên lazada:

Text

Description automatically generated

Data đã được tải về:

A screenshot of a computer

Description automatically generated with medium confidence

## 2. Số lượng dòng dữ liệu:

Có 40 dòng dữ liệu

## 3. Chuyển đổi giá trị:

### **-**Cột đánh giá: kiểu số

### -Cột đánh giá tích cực : kiểu số

### -Cột giá: kiểu số

### -Cột giao hàng đúng hẹn: kiểu số

### -Cột phí vận chuyển: kiểu số

### -Cột sale: kiểu số

### -Cột phản hồi: kiểu số

### Text Description automatically generated

# II. PHÂN TÍCH THÔNG KÊ MÔ TẢ SỐ LIỆU.

## 1. Tính các chỉ số thống kê.

**-**Tính toán các chỉ số thống kê mô tả cho cột giá

Chart

Description automatically generated with low confidence

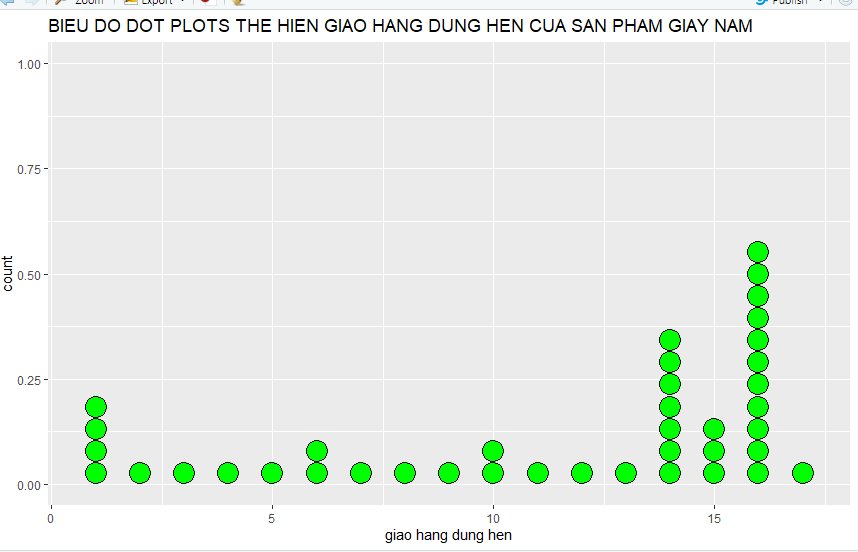
## 2. Vẽ các biểu đồ để trực quan hóa dữ liệu.

### **a. Biểu đồ Dot Plost.**

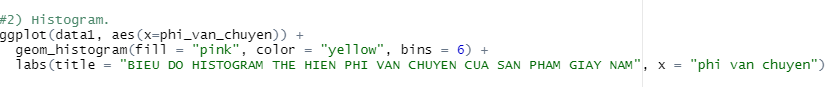
Sử dụng thư viện ggplot2 để vẽ biểu đồ Dot Plost của cột dữ liệu giao hàng đúng hẹn:

### 

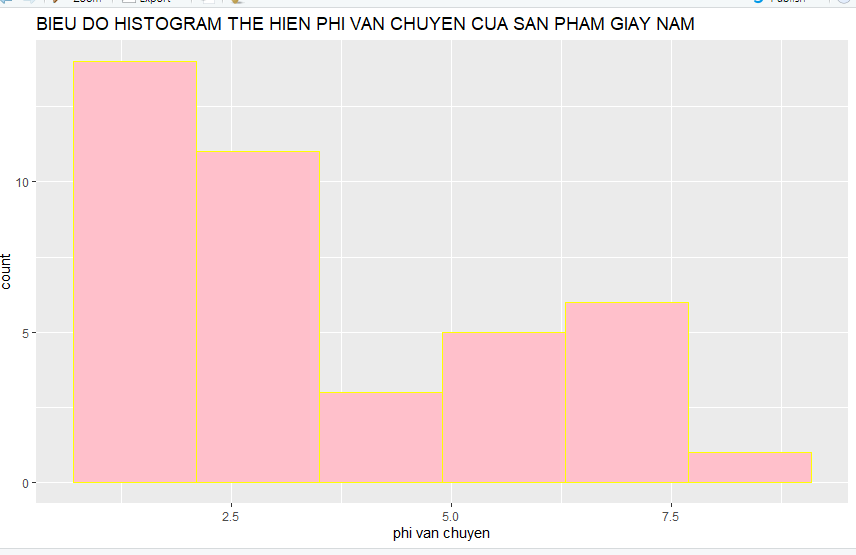
Kết quả biểu đồ:



### **b. Biểu đồ Histogram :**

****

Kết quả của biểu đồ Histogram:



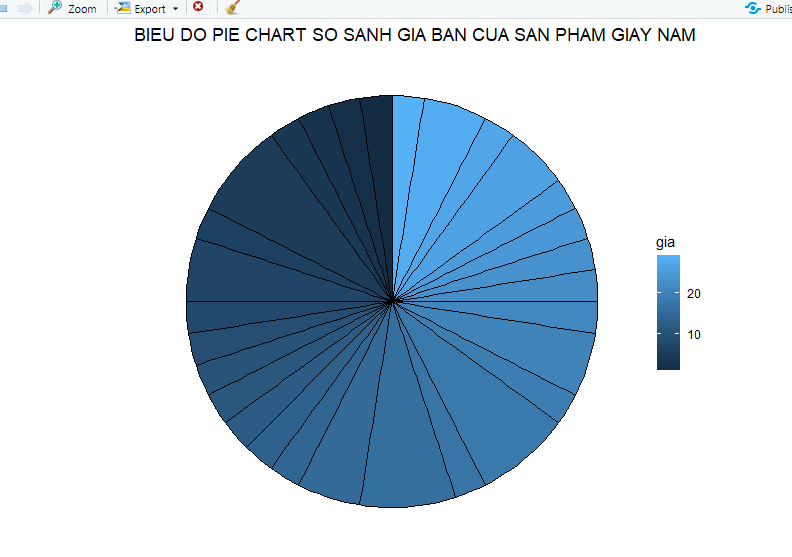
### **c. Biểu đồ Pie Charts :**

Vẽ biểu đồ Pie charts so sánh giá bán của các sản phẩm:

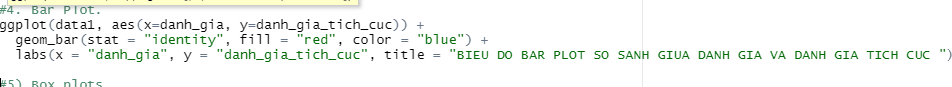
A picture containing text

Description automatically generated

Kết quả của biểu đồ:



### **d. Biểu đồ Bar Plot:**

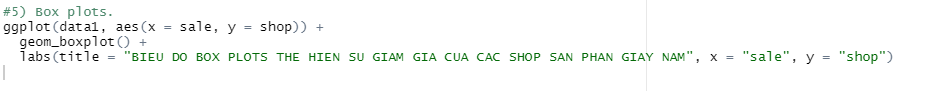


Vẽ biểu đồ:

Chart, bar chart

Description automatically generated

### **e. Biểu đồ Box plosts :**

****

Vẽ biểu đồ :

Calendar

Description automatically generated

### **f. Nhận xét:**

-Biều đồ dos plost: thể hiện số phần trăm giao hàng đúng hẹ của sản phẩm

-Biểu đồ histogram: thể hiện rõ phí vận chuyển của các sản phẩm

-Biểu đồ pie chart: thể giá bán của các sản phẩm

-Biểu đồ bar plot: thể hiện các đánh giá của khách hàng cho sản phẩm

-Biểu đồ box plots: thể hiện sản phẩm hạ giá của mỗi shop

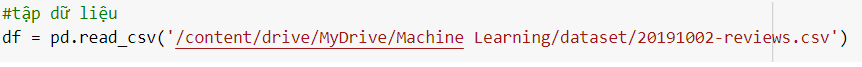
# III. Ý tưởng phân tích số liệu từ Kaggle :

1. **Import các thư viện :**

Text

Description automatically generated

1. **Đọc các file dữ liệu :**



1. **Làm sạch dữ liệu:**

-Đầu tiên loại bỏ các cột không cần thiết:

Table

Description automatically generated with low confidence

-Tiếp đến xử lí các mục trùng lặp, bị thiếu và các mục có giá trị không tự nhiên:

Table

Description automatically generated

Sau khi xử lý làm sạch dữ liệu, tập dữ liệu bao gồm 25.847 hàng và 4 cột.

1. **Phân tích**
2. **Phân tích đơn biến:**

Text

Description automatically generated

Biểu đồ:

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Phân tích :

* Hơn 8.000 khách hàng của Lazada đã mua hàng cho danh mục sản phẩm Ổ cứng gắn ngoài.
* Hơn 10.000 khách hàng của Lazada thực hiện giao dịch trên Android.
* Hơn 12.000 khách hàng của Lazada đưa ra đánh giá 5 cho các sản phẩm họ đãmua. Khách hàng của Lazada có xu hướng hài lòng với sản phẩm họ mua với khoảng điểm mức độ phù hợp từ 30 - 40.

1. **Phân tích đa biến:**

A picture containing company name

Description automatically generated

Biểu đồ:

Chart, radar chart, line chart

Description automatically generated

Logo

Description automatically generated with low confidence

Biểu đồ:

Chart, line chart

Description automatically generated

## ****Kế****

1. **Kết luận:**

* Khách hàng mua hàng trong danh mục sản phẩm Máy tính xách tay, TV thông minh và TV kỹ thuật số có xu hướng đưa ra xếp hạng tốt và hài lòng với chất lượng sản phẩm họ đã mua.
* Mặc dù các giao dịch mua hàng diễn ra trên nhiều thiết bị di động khác nhau, nhưng hầu hết các giao dịch mà khách hàng của Lazada thực hiện đều có xếp hạng cao và có mức độ phù hợp cao với sản phẩm họ đã mua.

ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ



PHIẾU ĐÁNH GIÁ ĐỒ ÁN/TIỂU LUẬN/BÀI TẬP LỚN

**Học kỳ I, năm học 2021 - 2022**

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1** | **Cán bộ chấm thi 2** |
| **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT1:**  Bằng số: .........................................  Bằng chữ: ....................................... | **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT2:**  Bằng số: .........................................  Bằng chữ: ....................................... |

Điểm kết luận: ...........................................................................................................

Bằng số:.....................................................................................................................

Bằng chữ: .................................................................................................................

*Thừa Thiên Huế, ngày tháng năm 2021*

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1**  *(Ký và ghi rõ họ và tên)* | **Cán bộ chấm thi 2**  *(Ký và ghi rõ họ và tên)* |