**Capstone Project 1**

**Thông tin chung:**

1. Họ và tên: Nguyen Que Lam
2. Mã số dữ liệu:Data 4\_Lam

**KẾT QUẢ PHÂN TÍCH**

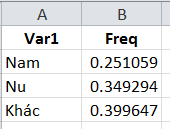
**Phần 1. Làm sạch dữ liệu**

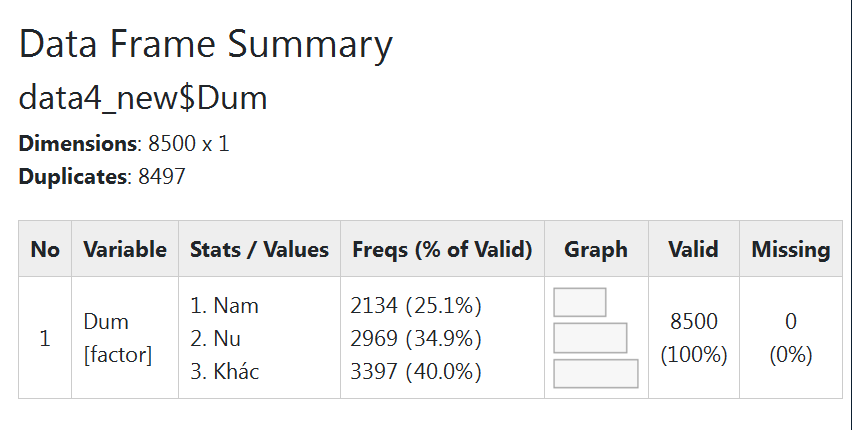
1. Import số liệu
2. Xóa dữ liệu thiếu logic (Ghi rõ biến số và dòng): Ví dụ: Biến X1, dòng 14, Biến X3, dòng 5.
3. Tìm dữ liệu bị missing (Ghi rõ biến số và dòng) và thực hiện thay thế dữ liệu bị missing (imputation)
4. Gán các thuộc tính cho biến số như label, value label, measure (Không viết báo cáo tại đây):
   * Biến số định lượng (X1, X2, X3, X4, Y): Gán numeric
   * Biến số định tính (Có tên là Dum trong file dữ liệu): Gán label (Giới tính), value label (1: Nam, 2: Nữ, 3: Khác), chuyển thành factor (lưu ý: levels = c(0, 1))

**Phần 2. Thống kê mô tả**

1. Thống kê mô tả biến số định tính:

**Bảng 1. Thống kê mô tả biến định tính**

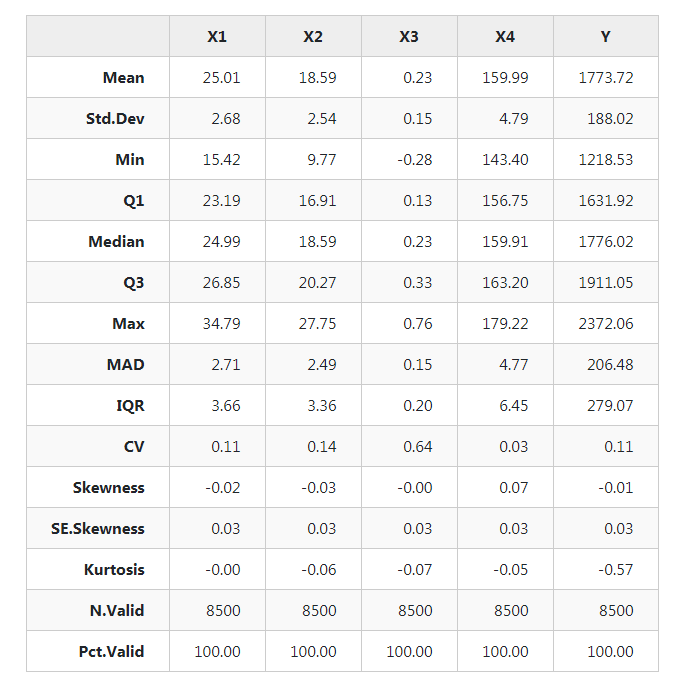




*(Insert table here)*

1. Thống kê mô tả biến số định lượng

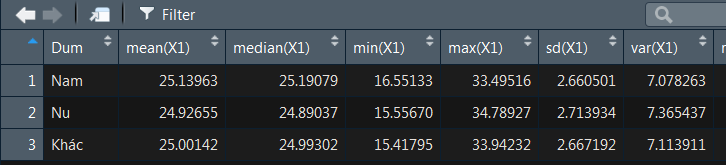
**Bảng 2. Thống kê mô tả biến định lượng**



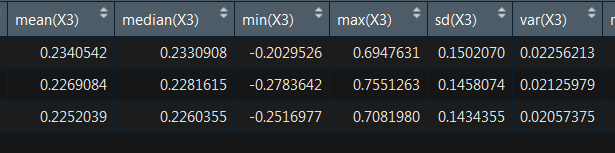
*(Insert table here)*

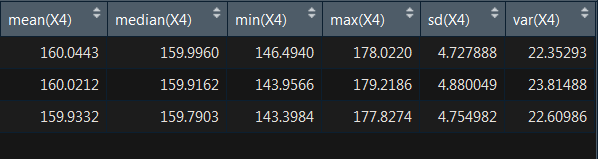
1. Thống kê bảng chéo biến định được mô tả phía trên theo tất cả các biến định lượng

**Bảng 3. Bảng thống kê chéo**







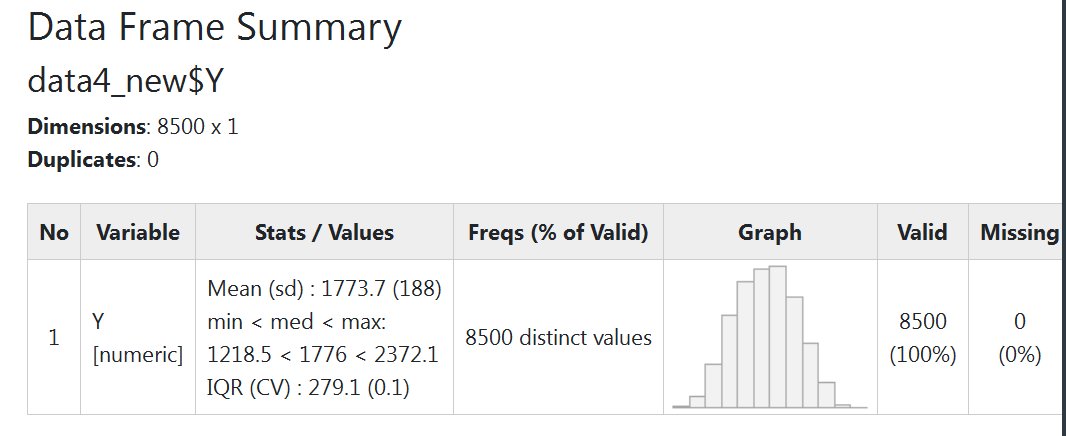


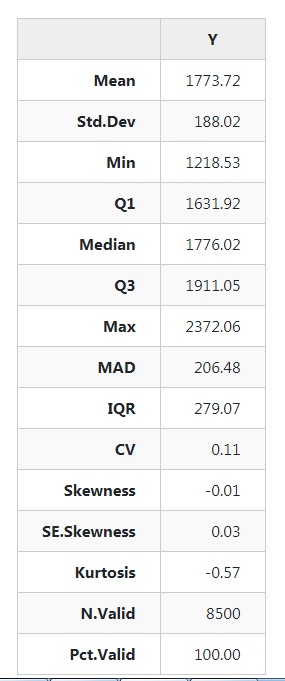
*(Insert table here)*

1. Thống kê tứ phân vị biến phụ thuộc:

**Bảng 4. Bảng thống tứ phân vị**



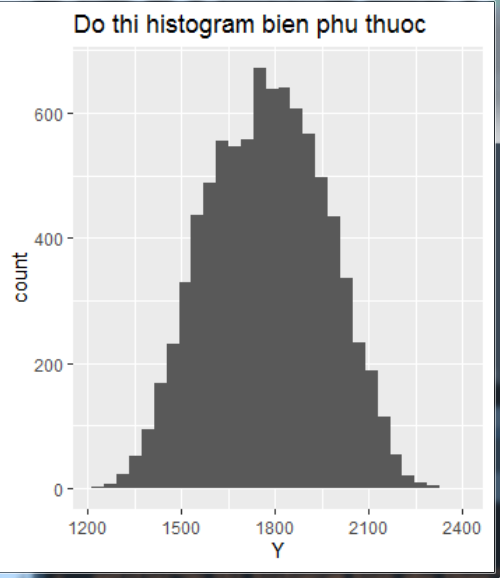




*(Insert table here)*

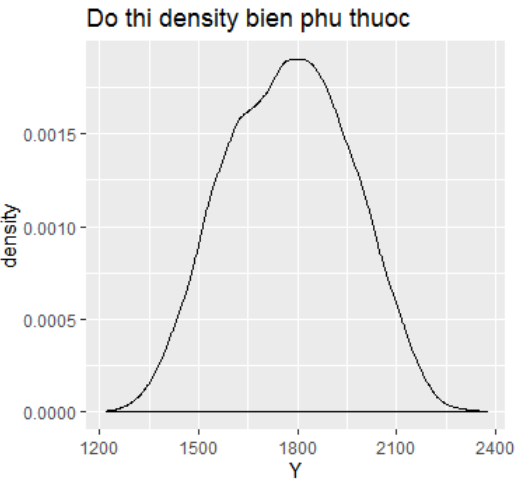
1. Vẽ đồ thị histogram, density và đồ thị box – plot biến phụ thuộc

**Hình 1. Đồ thị histogram biến phụ thuộc**



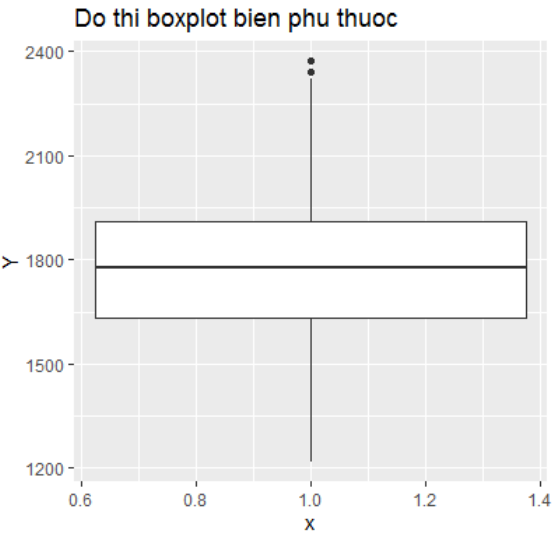
*(Insert figure here)*

**Hình 2. Đồ thị density biến phụ thuộc**



*(Insert figure here)*

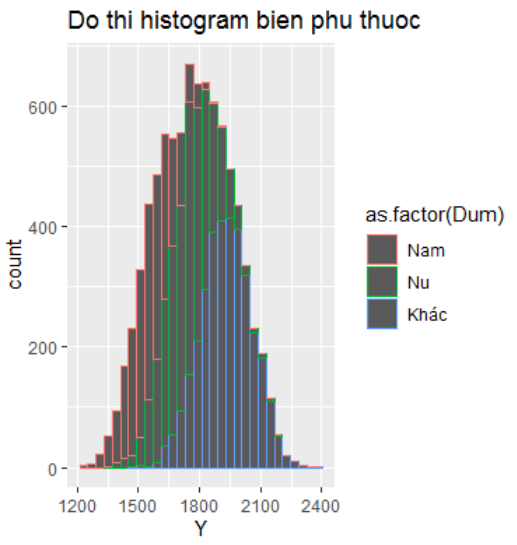
**Hình 3. Đồ thị box - plot biến phụ thuộc**



*(Insert figure here)*

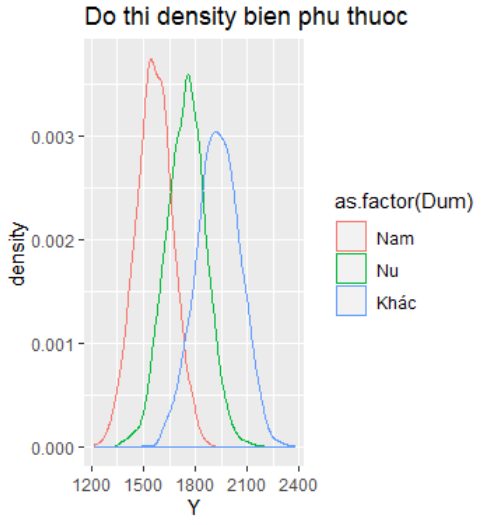
1. Vẽ đồ thị density, histogram và đồ thị box – plot biến phụ thuộc sử dụng colour = Dummy

**Hình 4. Đồ thị histogram biến phụ thuộc**



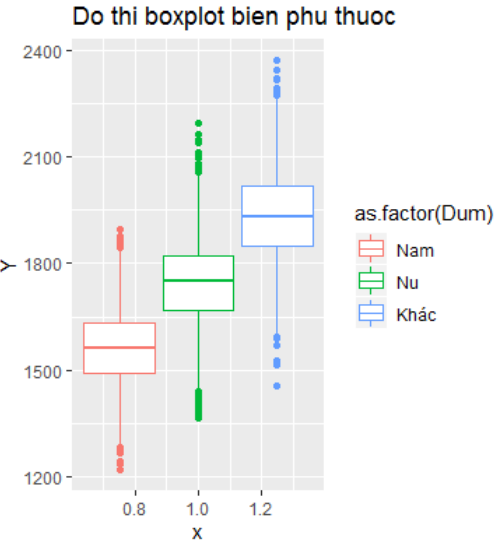
*(Insert figure here)*

**Hình 5. Đồ thị density biến phụ thuộc**



*(Insert figure here)*

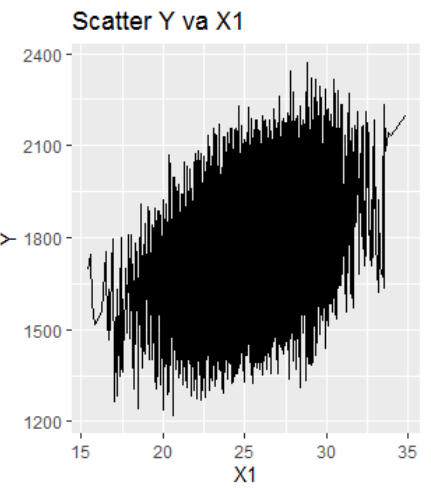
**Hình 6. Đồ thị box - plot biến phụ thuộc**



*(Insert figure here)*

1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1

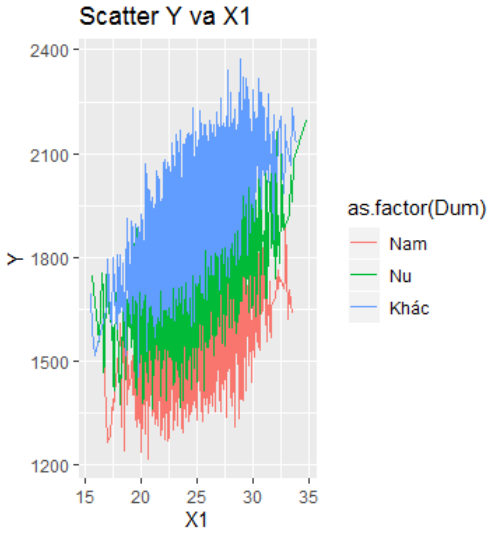
**Hình 7. Đồ thị scatter giữa biến phụ thuộc Y và biến X1**



*(Insert figure here)*

1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  sử dụng colour = Dummy

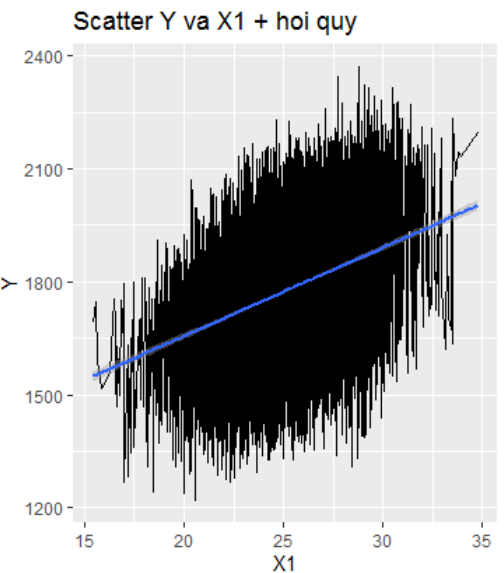
**Hình 8. Đồ thị scatter giữa biến phụ thuộc Y và biến X1**



*(Insert figure here)*

1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  + vẽ đường hồi quy

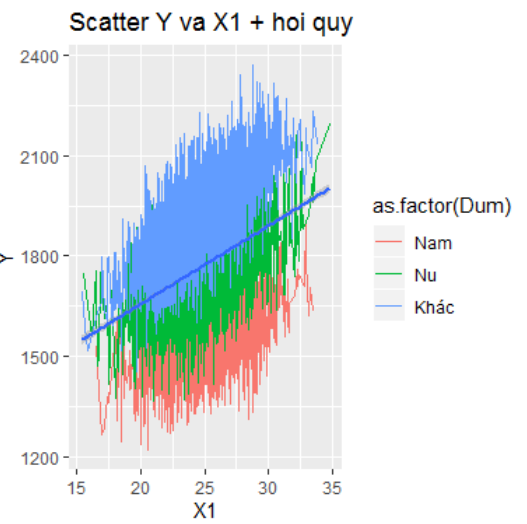
**Hình 9. Đồ thị scatter và đường hồi quy giữa biến phụ thuộc Y và biến X1**



*(Insert figure here)*

1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  sử dụng colour = Dummy + đường hồi quy

**Hình 10. Đồ thị scatter và đường hồi quy giữa biến phụ thuộc Y và biến X!\**

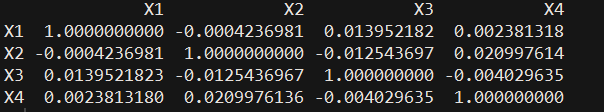


*(Insert figure here)*

**Phần 3. Phân tích ma trận tương quan:**

1. Ma trận tương quan giữa các ***biến độc lập định lượng***

**Bảng 5. Bảng ma trận tương quan**



*(Insert table here)*

|  |
| --- |
| Nhận xét tóm gọn về hệ số tương quan: |

* **X1 và X2: hệ số tương quan âm <0:** có nghĩa khi X1 tăng thì X2 giảm và ngược lại khi X1 giảm thì X2 tăng: Tuy nghiên hệ số gần bằng một nên ta có thể nói X1 và X2 không tương quan nhau
* **X1 và X3,X4: hệ số tương quan dương >0:** có nghĩa là X1 tăng thì X3, X4 cùng tăng nhưng X3 sẽ tăng cao hơn so với mức tăng X4. Do mức tương quan của X1 và X3 cao hơn so với X4
* **X2 và X3: hệ số tương quan âm <0:** có nghĩa khi X2 tăng thì X3 giảm và ngược lại khi X2 giảm thì X3 tăng:
* **X2 và X4:** **hệ số tương quan dương >0: có nghĩa** khi X2 tăng thì X3 cũng tăng
* **X3 và X4: hệ số tương quan âm <0:** có nghĩa khi X3 tăng thì X4 giảm và ngược lại khi X4 giảm thì X3 tăng: Tuy nghiên hệ số gần bằng một nên ta có thể nói X3 và X4 không tương quan nhau

**Phần 4. Phân tích hồi quy:**

**4.1. Mô hình 1: Tác động của biến độc lập định lượng tới biến phụ thuộc**

* **Phương trình:**

**Y = 0 + 1X1 + X + …+ nXn**

**Bảng 6. Kết quả hồi quy**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

**4.2. Mô hình 2: Tác động của biến độc lập định tính và định lượng tới biến phụ thuộc (không xuất hiện biến tương tác)**

* **Phương trình:**

**Y = 0 + 1Dum + X + X + …+ kXn**

**Bảng 7. Kết quả hồi quy**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

**4.3. Mô hình 3: Tác động tương tác của biến độc lập định tính và định lượng tới biến phụ thuộc**

* **Phương trình:**

**Y = 0 + 1(Dum\*X1) + Dum + X + X +…+ kXn**

**Bảng 8. Kết quả hồi quy**

*(Insert table here)*

* **Viết phương trình kết quả hồi quy:**

*(Insert here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

* **Kiểm định đa cộng tuyến:**

**Bảng 9. Hệ số VIF**

*(Insert table here)*

|  |
| --- |
| Nhận xét tóm gọn kết quả kiểm định: |

* **Kiểm định phương sai của sai số thay đổi:**

**Bảng 10. Kết quả kiểm định phương sai của sai số thay đổi**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn kết quả kiểm định:** |

* **Kiểm định phân phối chuẩn của sai số** (**nâng cao**):

**Bảng 11. Kết quả kiểm định phân phối chuẩn của sai số thay đổi**

*(Insert table here)*

* Tính biến số được đặt tên là **Y\_fitted** (giá trị dự đoán của Y) và **Residuals** (sai số của mô hình): Không báo cáo ở đây
* Vẽ đồ thị scatter: Y\_fitted vs Y và Y\_fitted vs. Residuals

**Hình 11. Đồ thị scatter Y\_fitted và Y**

*(Insert figure here)*

**Hình 12. Đồ thị scatter Y\_fitted và Residuals**

*(Insert figure here)*

**KẾT THÚC**

***Chúc mừng các em đã hoàn thành capstone project!!!***