Update 12/5

Tiền xử lí + visualize các cột theo thứ tự trong data\_df.info()

nam 1-16

đăng 17-32

hân 33-48

bửu 49-64

tú 65-80

* Mn có thể làm trên local hoặc tạo 1 file ipynb của mình trên folder này, làm xong thì copy vào file Group.ipynb
* Không cần visualize toàn bộ, chỉ chọn ra những thuộc tính mà mn thấy có ý nghĩa. Nếu thấy có thuộc tính trong phần của mình có liên hệ với một thuộc tính trong phần của ng khác thì có thể liên hệ với ng đó để đưa ra cách xử lí chung
* Ghi chú cho phần mình làm (vd vì sao làm như vậy,...) vì sau đó mình còn phải làm slide thuyết trình nữa
* Giá trị NaN categorical để biểu thị ý nghĩa 'không có' thống nhất fill là 'None' (xem file data\_description.txt để hiểu cụ thể dữ liệu)

**Tối thứ 6 (20/5) mình sẽ họp lại để làm tiếp phần model nên mn cần phải xong trước lúc đó.**

Nếu có vấn đề phát sinh hoặc có thắc mắc mn cứ nhắn lên gr :)))

Update 20/5

Mn tìm hiểu các mô hình ML để áp vào bài toán trên tập train, tối ưu tham số trên tập valid, sau đó triển khai trên tập test. Mn tìm hiểu về mô hình nào thì điền vào bảng dưới để ng khác k làm trùng mô hình. (nếu ai k nắm idea build mô hình ML thì có thể tham khảo file lab3)

**Deadline cho phần mô hình là trước tối 24/5 (thứ 3), bắt đầu từ t7-21/5**

| Nam | Đăng | Tú | Hân | Bửu |
| --- | --- | --- | --- | --- |
| Decision tree regression |  | Logistic Regression | Lasso Lars | SVM |
| Linear regression |  | Bayesian Ridge Regression | Ridge (thêm cuối file group) |  |
| Sgd regression |  |  |  |  |
| Random forest regression |  |  |  |  |