Mô tả công việc

# 1. Mô tả công việc

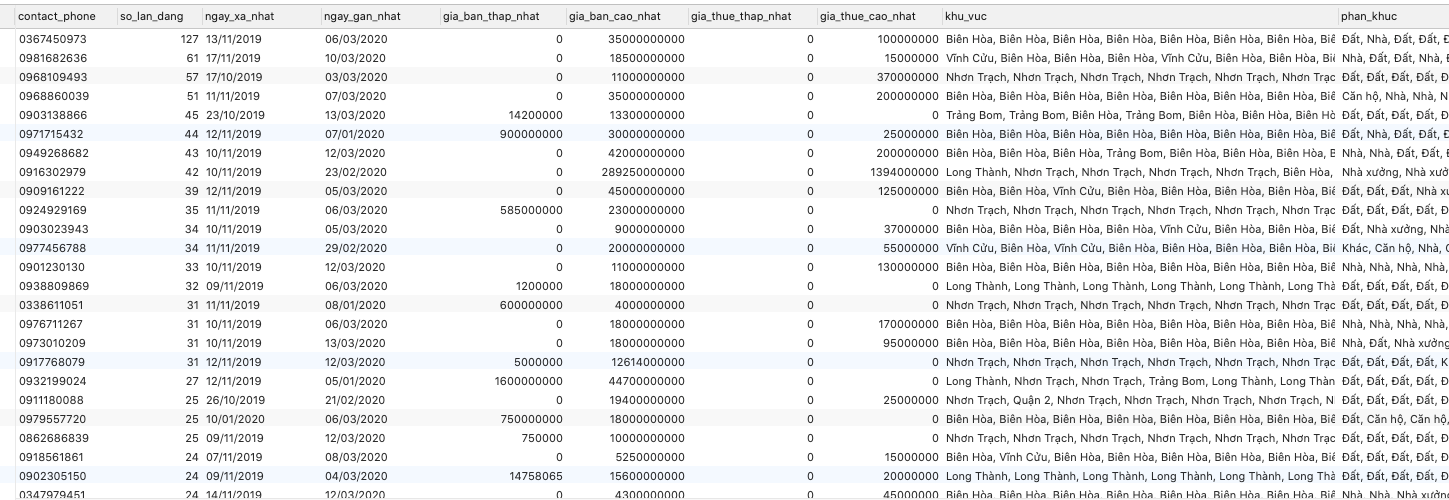
Các bạn sẽ làm tiếp phần xác định nhà chính chủ nhưng lần này ngoài việc xét content của từng bài đăng (code trước các bạn đã làm thành công), số điện thoại trong nội dung cũng được đưa vào xem xét.

Cụ thể, số điện thoại sẽ được xem xét dựa vào các **rule** sau:

|  |
| --- |
| - Các sđt nào đăng nhiều lần thì khả năng cao là môi giới (VD: n > 20)  - Các sđt nào chỉ đăng cùng 1 nội dung nhưng nhiều lần thì khả năng cao là chính chủ  - Các sđt nào đăng 1 lần thì khả năng cao là chính chủ |

**Ví dụ:**

Nhìn hình bên dưới, các bạn có thể thấy số điện thoại 0367550973 xuất hiện 126 lần trong DB nên ta có thể cho rằng số điện thoại này của môi giới.



# 2. Mô tả file

File **‘data\_1.json’** và file kết quả **‘result2.json’** được lấy tại link drive sau: <https://drive.google.com/drive/u/1/folders/1-KqkjPR331MEvYmGh-PvwZwIVZpwaCRg>. Mỗi bài đăng (post) sẽ gồm:

- id bài đăng

- content bài đăng

- contact\_phone của người rao bài nếu có

Các bạn dùng dữ liệu này để test thuật toán nhé.