**Tổng quan**

Robert là một thanh niên người Mỹ năm nay 25 tuổi, anh vừa kết thúc nghĩa vụ quân sự. Robert giờ muốn theo học ngành Hệ thống quản lý và Công nghệ thông tin (Management Systems and Information Technology).

Hãy dùng kĩ năng data của bạn để giúp tìm ngôi trường tốt nhất cho Robert. Robert sẵn sàng chuyển tới bất kỳ đâu trên nước Mỹ để sinh sống, tuy nhiên anh ý có một vài tiêu chí chọn trường như sau: 1) an toàn (tỷ lệ tội phạm thấp), 2) ở đô thị -- vì Robert thích cuộc sống ở thành phố, và 3) start-ups – ngôi trường phải ở trong một khu vực mà xếp hạng cao về tinh thần doanh nhân (entrepreneurialism), vì anh ý có kế hoạch sẽ tìm kiếm 1 công việc thực tập tại 1 startup trong khi theo học.

Hãy giúp Robert rút gọn danh sách tìm kiếm xuống thành một danh sách những trường tiềm năng để anh ý có thể tiếp tục nghiên cứu kĩ hơn trước khi đưa ra lựa chọn.

**Nhiệm vụ cụ thể của bạn**

1. Xuất ra 1 danh sách những trường mà thỏa mãn tất cả các tiêu chí của Robert
2. Xếp hạng các trường từ cao xuống thấp theo các tiêu chí đó.
3. Sử dụng Python để import và xử lý dữ liệu. Sử dụng Jupyter notebook và markdown trong jupyter notebook để trình bày bài làm. Dùng Markdown là cách viết tài liệu nhanh gọn và rõ ràng để chia sẻ các phân tích giữa các thành viên trong các team analytics. Hãy tham khảo video này để học cách dùng markdown trong jupyter notebook: <https://www.youtube.com/watch?v=uVLzL5E-YBM>

Sau khi làm xong, gửi lại file .ipynb

Các tiêu chí chọn trường:

1. Phải ở khu vực thành thị (urban/metropolitan area).
2. Phải ở 1 thành phố mà xếp hạng 75th percentile hoặc cao hơn trong Dữ liệu “metro\_startup\_ranking” (tức là trong top 25% có rank cao nhất, với rank 1 là cao nhất).
3. Phải ở 1 thành phố mà có thấp hơn 50th percentile vể tổng số vụ phạm tội (total number of crime).
4. Có chương trình học về lĩnh vực “Computer And Information Sciences And Support Services”.

Gợi ý:

1. Đọc data dictionary để hiểu ý nghĩa các bảng, cột (chỉ tập trung đọc những phần mình quan tâm ở dictionary thôi).
2. Xóa các cột không cần thiết.
3. Tìm các cột phù hợp để join các bảng với nhau.
4. Thực hiện data cleaning và chuẩn hóa data nếu cần để có thể join các bảng với nhau.
5. Tạo 1 bảng summary variable thể hiện số lượng các vụ phạm tội ở khu vực của các trường, từ đó ta có thể so sánh các trường theo tổng số vụ phạm tội.
6. Tạo ra 1 phương pháp để xếp hạng các trường theo đánh giá tất cả các tiêu chí của Robert.