**Case Study #1 Answer:**

Nhật xét tập dữ liệu:

- Dữ liệu gồm 55 biến và 10.000 dòng dữ liệu

- Dữ liệu NA nằm tại các biến:

+ emp\_title, emp\_length, debt\_to\_income, annual\_income\_joint, verification\_income\_joint, debt\_to\_income\_joint, months\_since\_last\_delinq, months\_since\_90d\_late, months\_since\_last\_credit\_inquiry, num\_accounts\_120d\_past\_due.

+ trong đó có một số biến quan trọng về mặt lý thuyết tài chính, do vậy việc thiếu dữ liệu có thể ảnh hưởng đến tính chính xác của độ tương quan giữa biến đó với lãi suất.

- debt\_consolidation là mục đích vay nhiều nhất cho thấy đa phần khách hàng có nhu cầu vay lâu dài và tăng trưởng mức vay, khách hàng có thể vay cho kinh doanh cá nhân, hoặc tập thể.

- debt\_consolidation, credit\_card, house, small\_business là những mục đích vay có lãi suất vay cao nhất. Nhưng khách hàng có nhu cầu credit\_card, house, small\_business vay với mức lãi cao không nhiều.

- Phương thức giả ngân phổ biến nhất là tiền mặt

- Giải ngân bằng tiền mặt mang lại nhiều khoản vay có lãi suất cao hơn phương thức khác

- Khách hàng vay tiền chủ yếu có tỷ lệ debt\_to\_income từ 4% đến 30%. Tuy nhiên mức lãi vay cao hay thấp lại phân bổ đều cho các mức tỷ lệ debt\_to\_income. Cho thấy mức lãi vay cao hay thấp có mức phụ thuộc thấp vào tỷ lệ debt\_to\_income.