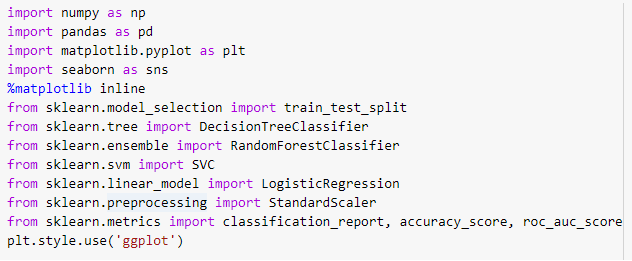
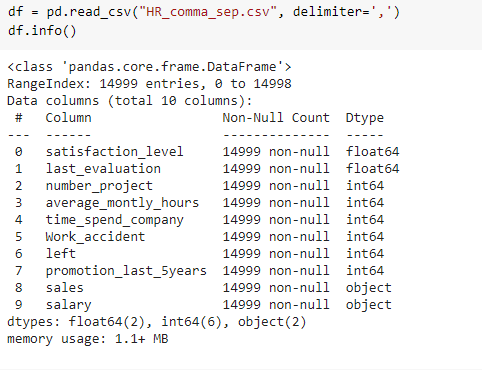
Nhập các thư viện

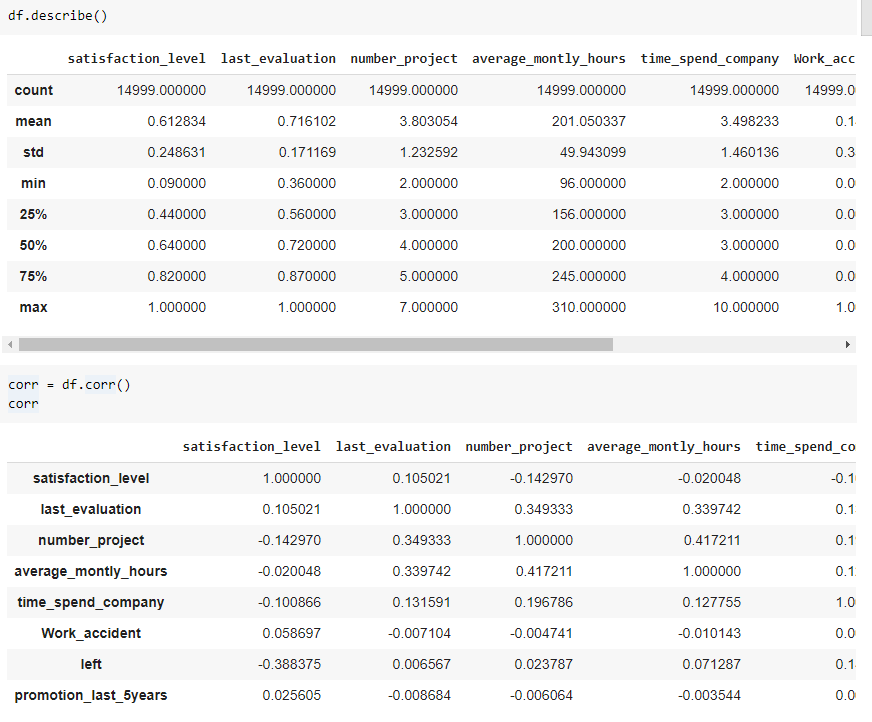


Tải dữ liệu từ file csv vào khung dữ iệu

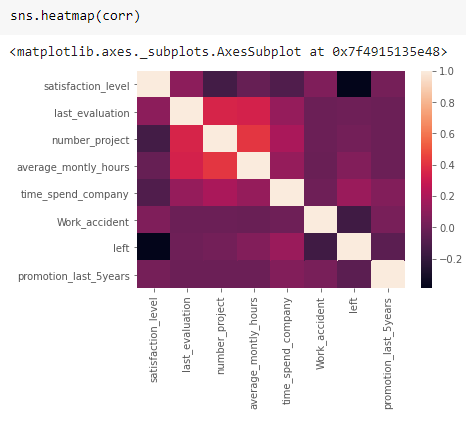


Thống kê dữ liệu cơ bảng bằng describe().

Dùng corr() để biết mối tương quan giữa các số liệu của dữ liệu



Vẽ biểu đồ trực quan

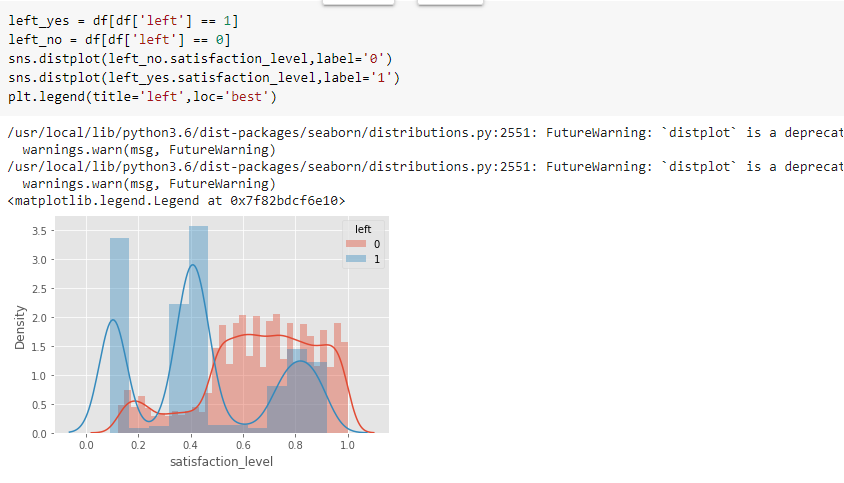




Nhận xét: Một số lượng lớn nhân viên với mức lương thấp hơn so với những người khác. Bộ phận kinh doanh có số lượng nhân viên rời công ty nhiều nhất, tiếp theo là bộ phận kỹ thuật và hỗ trợ.



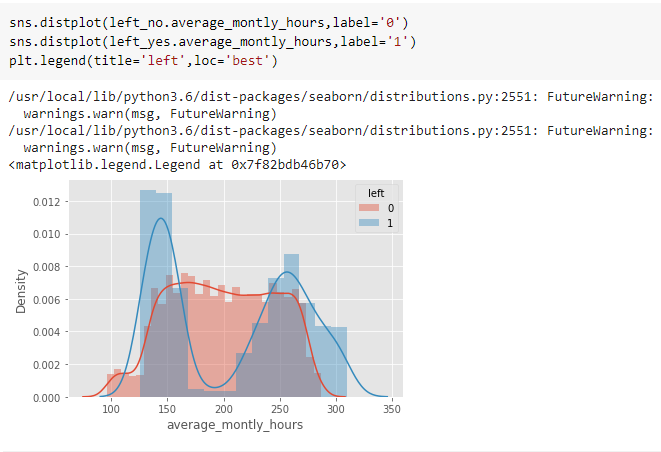
Người có 2 dự án rời công ty đang ở mức báo động. Việc rời công ty có xu hướng tăng ngay trong năm thứ 3. Một phần lớn nhân viên đã rời công ty sau 3 năm làm việc tại đó. Có một xu hướng giảm với con số hợp lý cho đến năm thứ sáu.



Nhận xét: Những nhân viên có mức độ hài lòng với công việc thấp hơn đã rời khỏi tổ chức ở tất cả các bộ phận. Một số nhân viên có mức độ hài lòng với công việc cao hơn cũng rời đi.

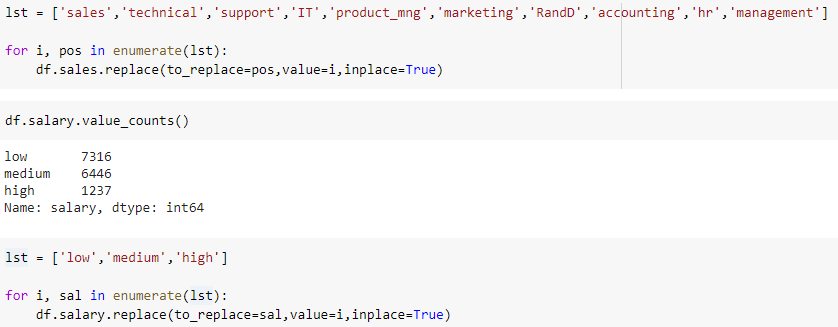


Nhận xét: Những người làm theo giờ cao hơn đã rời khỏi tổ chức ở tất cả các phòng ban.

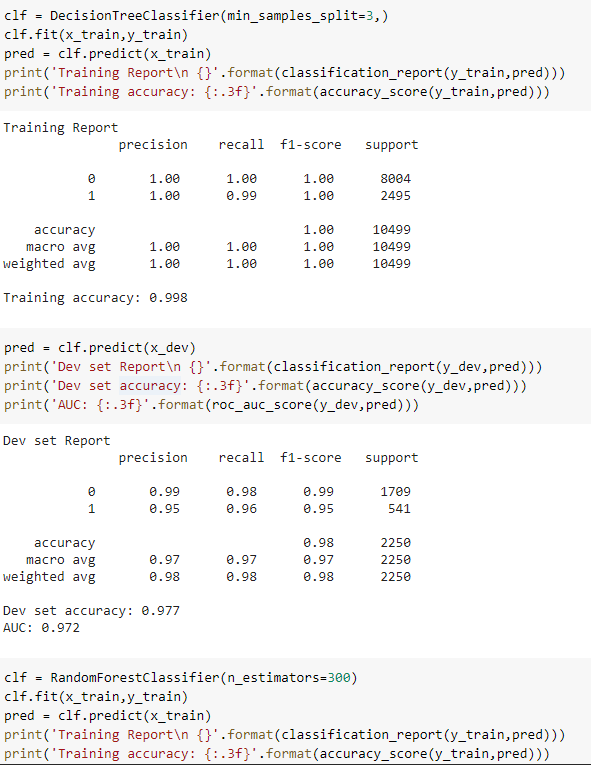


Nhận xét: Số người làm việc khoảng 150 giờ và số người làm việc khoảng 250 giờ còn lại nhiều hơn những người khác. Sự phân bố có hai đỉnh ở những vùng này.

Tiền xử lý dữ liệu



Tạo cây quyết định



Phân lớp dữ liệu

