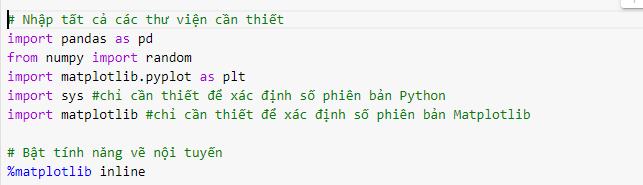
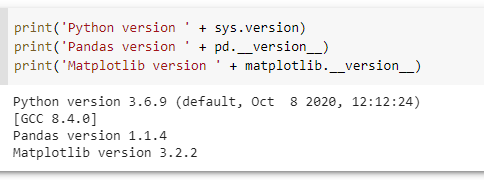
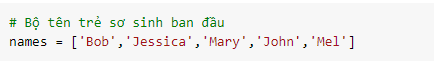
Khai báo tất cả các thư viện



Xuất ra thông tin các phiên bản



Tạo dữ liệu



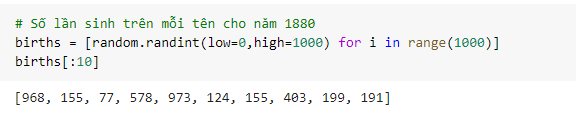
random.seed(500): Tạo seed ngẫu nhiên gồm 500 phần tử

random\_names = Chọn một tên ngẫu nhiên từ danh sách tên và thực hiện điều này n lần.

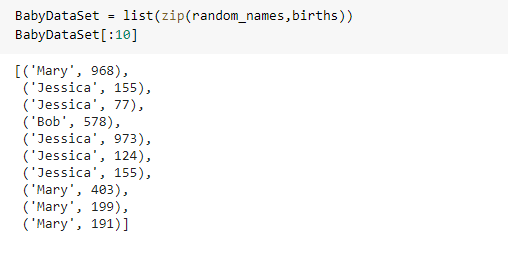
for i in range(n) - Lặp lại cho đến khi i bằng n, tức là 1,2,3, .... n.



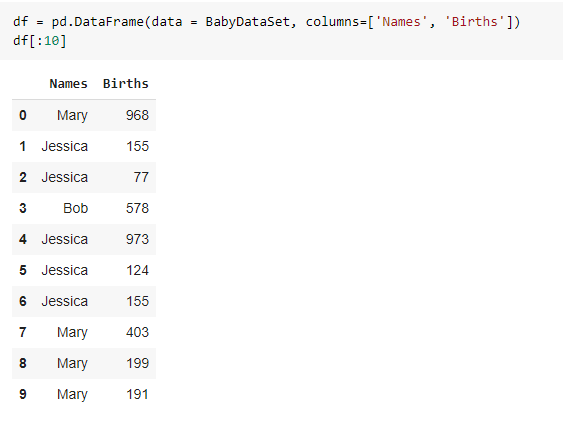
Tạo một số ngẫu nhiên từ 0 đến 1000



Hợp nhất **names** và tập dữ liệu **births** bằng hàm **zip**.



Tạo dataframe



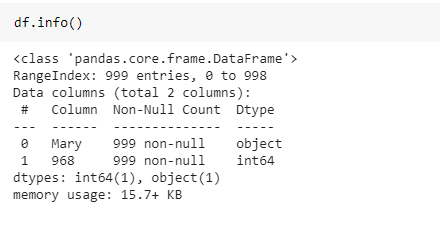
Xuất ra file txt

Các tham số duy nhất chúng tôi sẽ sử dụng là **index** và **header**. Đặt các tham số này thành False sẽ ngăn không cho xuất tên index và header.



chuyển nó vào vị trí của tệp văn bản

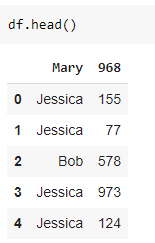




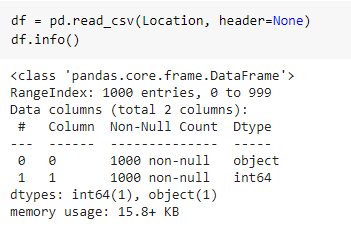
Thông tin cho biết:

* Có 999 bản ghi trong tập dữ liệu
* Có một cột tên là Mary với giá trị 999
* Có một cột tên là 968 với 999 giá trị
* Trong số hai cột, một cột là số, cột kia không phải là số

Xem dữ liệu 5 dòng đầu



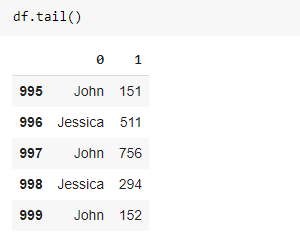
Chúng tôi sẽ chuyển tham số **header** cho hàm read\_csv và đặt nó thành **None** (có nghĩa là null trong python).



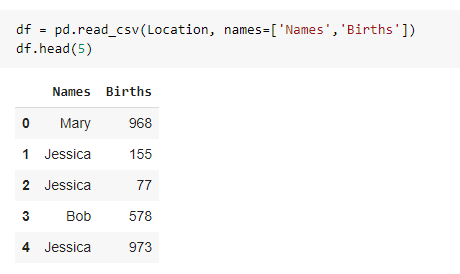
Thông tin bây giờ cho biết:

* Có 1000 bản ghi trong tập dữ liệu
* Có một cột tên 0 với 1000 giá trị
* Có một cột tên là 1 với 1000 giá trị
* Trong số hai cột, một cột là số, cột kia không phải là số

Xem dữ liệu 5 dòng cuối



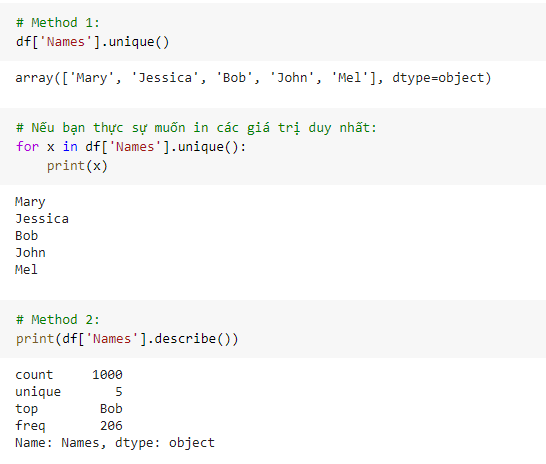
Đặt tên cụ thể cho cột



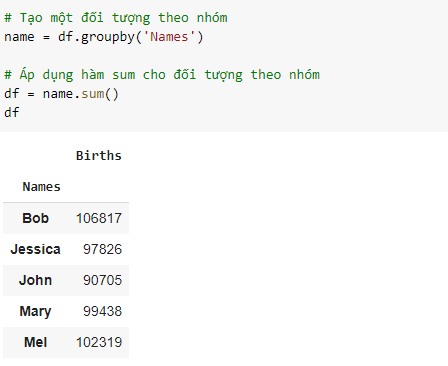
Xóa tệp txt



Xác minh cột "Names" vẫn chỉ có năm tên duy nhất.

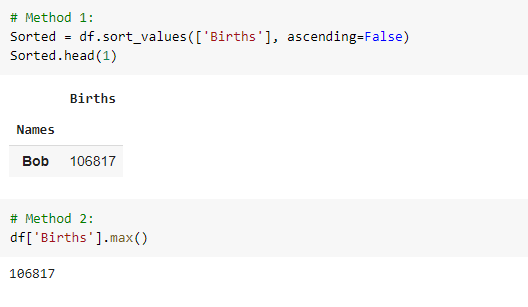


Tổng hợp dữ liệu để mỗi tên chỉ xuất hiện một lần. Thực hiện bằng cách sử dụng hàm groupby



Tìm tên phổ biến nhất hoặc tên có tỷ lệ sinh cao nhất. Có 2 cách:

* Sắp xếp khung dữ liệu và chọn hàng trên cùng
* Sử dụng thuộc tính max() để tìm giá trị lớn nhất



Vẽ biểu đồ cột **Births** và gắn nhãn biểu đồ để hiển thị cho người dùng cuối điểm cao nhất trên biểu đồ. Kết hợp với bảng, người dùng cuối có một bức tranh rõ ràng rằng **Bob** là tên em bé phổ biến nhất trong tập dữ liệu.

