



PDF và Bảng tính



Hoàn thành chương trình đào tạo Python

- Python có khả năng làm việc với các tệp PDF và tệp bảng tính.
- Trong phần này, chúng ta sẽ khám phá các thư viện cho phép chúng ta tương tác với các tệp này.
- Lưu ý: Chúng tôi khuyên bạn nên làm việc ở cùng vị trí với sổ ghi chép bài giảng, vì chúng ta sẽ tham chiếu nhiều tệp ở vị trí đó.



Làm việc với tệp CSV



Hoàn thành chương trình đào tạo Python

- CSV là viết tắt của các biến được phân tách bằng dấu phẩy và là kết quả đầu ra rất phổ biến cho các chương trình bảng tính.
- Ví dụ:
 - Tên, Giờ, Tỷ lệ
 - David, 20, 15
 - Claire, 40, 20



Hoàn thành chương trình đào tạo Python

- Lưu ý rằng mặc dù có thể xuất tệp excel và Bảng tính Google sang tệp .csv nhưng nó chỉ xuất thông tin.
- Những thứ như công thức, hình ảnh và macro không được nằm trong tệp .csv.
- Nói một cách đơn giản, tệp .csv chỉ chứa dữ liệu thô từ bảng tính.



Hoàn thành chương trình đào tạo Python

- Chúng tôi sẽ làm việc với mô-đun csv tích hợp sẵn cho Python, mô-đun này sẽ cho phép chúng tôi lấy các cột, hàng và giá trị từ tệp .csv cũng như ghi vào tệp .csv.
- Hãy nhớ rằng đây là không gian rất phổ biến cho các thư viện bên ngoài mà bạn có thể muốn khám phá.



Hoàn thành chương trình đào tạo Python

- Các thư viện khác cần xem xét:
 - Gấu trúc
 - Thư viện phân tích dữ liệu đầy đủ, có thể làm việc với hầu hết mọi loại dữ liệu dạng bảng.
 - Chạy trực quan hóa và phân tích.
 - Một trong những sở thích cá nhân của tôi, chúng tôi dạy nó trong các khóa học khoa học dữ liệu khác nhau.



Hoàn thành chương trình đào tạo Python

- Các thư viện khác cần xem xét:
 - Openpyxl
 - Được thiết kế đặc biệt cho các tập tin Excel.
 - Giữ lại nhiều chức năng cụ thể của Excel.
 - Hỗ trợ các công thức Excel.
 - python-excel.org theo dõi nhiều thư viện Python dựa trên Excel khác.



Hoàn thành chương trình đào tạo Python

- Các thư viện khác cần xem xét:
 - API Python của Google Trang tính
 - Giao diện Python trực tiếp để làm việc với Bảng tính Google.
 - Cho phép bạn trực tiếp thực hiện các thay đổi đối với bảng tính được lưu trữ trực tuyến.
 - Cú pháp phức tạp hơn nhưng có sẵn trong nhiều ngôn ngữ lập trình.



Hoàn thành chương trình đào tạo Python

- Yếu tố chung giữa tất cả các chương trình bảng tính này là chúng luôn có thể xuất sang .csv.
- Hãy khám phá các khả năng tích hợp của Python với mô-đun csv!



Làm việc với tệp PDF



Hoàn thành chương trình đào tạo Python

- PDF là viết tắt của Định dạng Tài liệu Di động và được Adobe phát triển vào những năm 1990.
- Điều quan trọng nhất cần ghi nhớ là mặc dù các tệp PDF có cùng phần mở rộng và có thể xem được trong trình đọc PDF nhưng nhiều tệp PDF không thể đọc được bằng máy thông qua Python.



Hoàn thành chương trình đào tạo Python

- Vì tệp PDF chủ yếu đóng gói và hiển thị tài liệu phẳng có bố cục cố định nên không có định dạng chuẩn nào có thể đọc được bằng máy, không giống như tệp CSV.
- Điều này có nghĩa là một tệp PDF đơn giản được quét rất khó có thể đọc được.



Hoàn thành chương trình đào tạo Python

- Các phần bổ sung vào tệp PDF như hình ảnh, bảng biểu, điều chỉnh định dạng cũng có thể khiến Python không thể đọc được tệp PDF.
- Có nhiều chương trình PDF trả phí có thể đọc và trích xuất từ các tệp này, nhưng chúng tôi sẽ sử dụng thư viện PyPDF2 mã nguồn mở và miễn phí .



Hoàn thành chương trình đào tạo Python

- Chúng tôi đảm bảo rằng PyPDF2 có thể đọc được các tệp PDF có trong tài liệu khóa học này.
- Rất tiếc, chúng tôi không thể cung cấp hỗ trợ cho các tệp PDF cá nhân của bạn nếu PyPDF2 không thể đọc được chúng.



Hoàn thành chương trình đào tạo Python

- Hãy khám phá cách làm việc với tệp PDF bằng Python.
- Hãy nhớ rằng trước tiên bạn cần cài đặt PyPDF2 tại dòng lệnh của mình:
 - `pip cài đặt PyPDF2`



PDF và CSV

Giải bài tập giải đồ