

LAB 1-1: CÀI ĐẶT ANACODA

1. Cài đặt chương trình Python và thư viện có sẵn trong bộ Anaconda

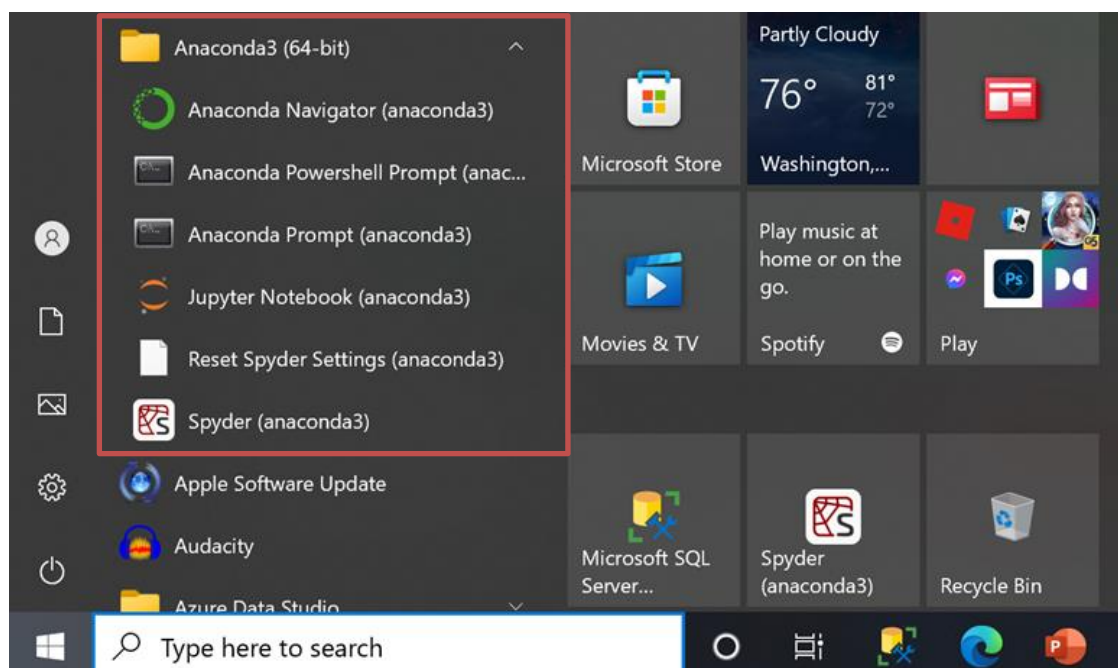
+ Cài đặt Anaconda

- **Anaconda** là một bản phân phối các ngôn ngữ lập trình Python và R cho tính toán khoa học, nhằm mục đích đơn giản hóa việc triển khai và quản lý gói. Bản phân phối bao gồm các gói khoa học dữ liệu phù hợp với **Windows**, **Linux** và **macOS**.

Hướng dẫn:

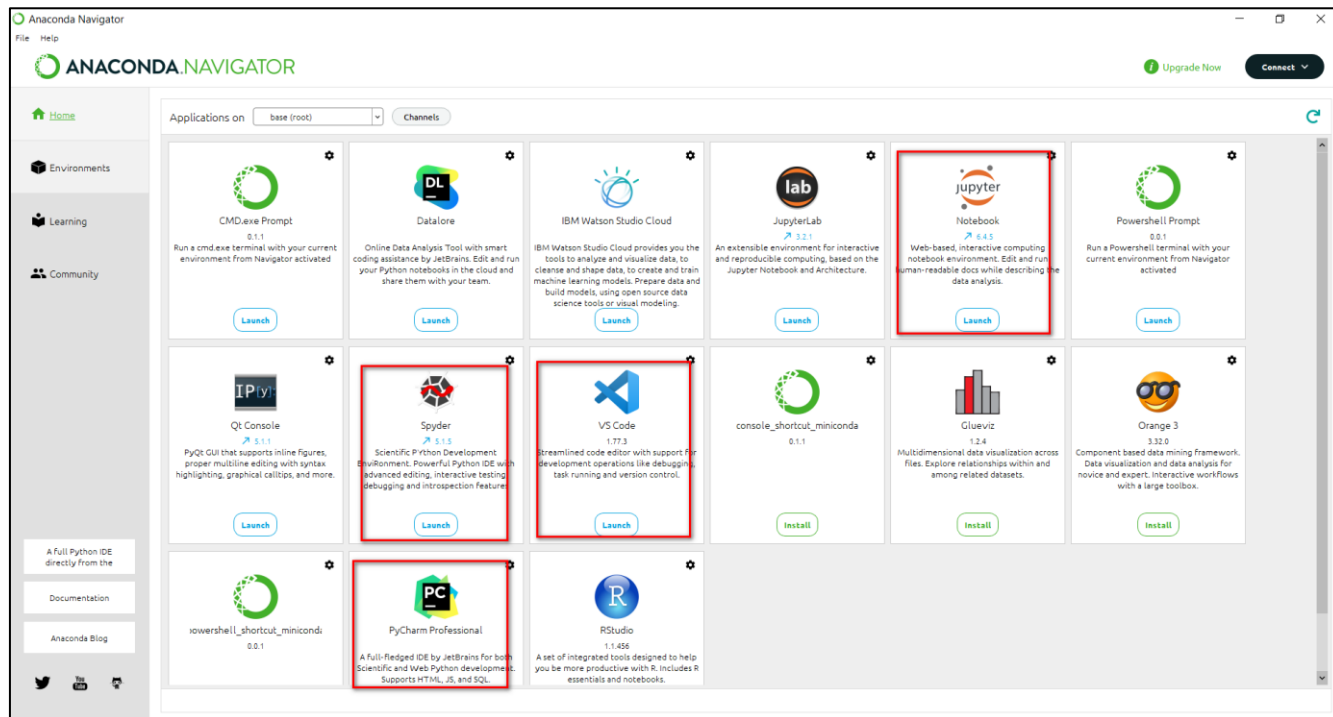
- Download gói cài đặt tại: Phiên bản hiện tại là: **Anaconda3-2022.05**

https://repo.anaconda.com/archive/Anaconda3-2022.05-Windows-x86_64.exe

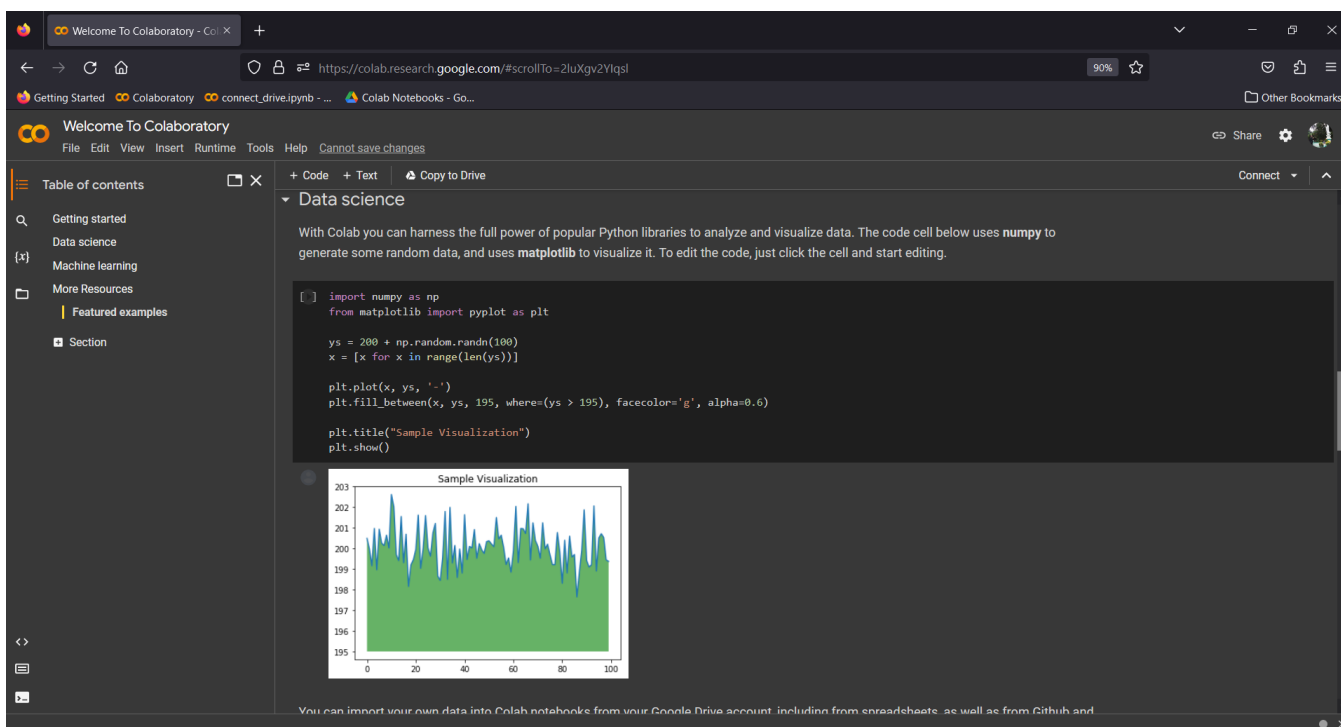


- Trong bộ Anaconda3(64 bit) có thể sử dụng sau để soạn thảo code

- Chương trình **Spyder**
- Chương trình **VS Code**
- Chương trình **Jupyter**
- Chương trình **PyCharm**



2. Có thể dùng COLABORATORY (online), viết trực tiếp trên giao diện



3. Tìm hiểu về các thư viện thông dụng trong Python

a. Thư viện NumPy

- **NumPy** là một gói xử lý (Processing Package) phổ biến của **Python**. **NumPy** làm phong phú ngôn ngữ lập trình Python với các cấu trúc dữ liệu mạnh mẽ để tính toán hiệu quả các mảng và ma trận đa chiều. **NumPy** không chỉ là một gói module để xử lý mảng mà nó còn cung cấp khả năng quản lý mảng cực kỳ vượt trội.

b. Thư viện SciPy

- Thư viện hữu ích này bao gồm các module cho đại số tuyến tính, tích hợp, tối ưu hóa và thống kê. Chức năng chính của nó được xây dựng dựa trên NumPy, vì vậy các mảng của nó sử dụng thư viện này. SciPy hoạt động hiệu quả cho tất cả các loại dự án lập trình khoa học (khoa học, toán học và kỹ thuật). Nó cung cấp các quy trình số hiệu quả như tối ưu hóa số, tích hợp và các quy trình khác trong module con. Tài liệu phong phú giúp làm việc với thư viện này thực sự dễ dàng.

c. Thư viện Pandas

Pandas là một thư viện được tạo ra để giúp các nhà phát triển làm việc với dữ liệu "labeled" và "relational" một cách trực quan. Nó dựa trên hai cấu trúc dữ liệu chính: "Chuỗi" (một chiều, giống như danh sách các mục) và "Khung dữ liệu" (hai chiều, giống như một bảng có nhiều cột). Pandas cho phép chuyển đổi cấu trúc dữ liệu thành các đối tượng DataFrame, xử lý dữ liệu bị thiếu và thêm / xóa các cột khỏi DataFrame, đưa vào các tệp bị thiếu và vẽ dữ liệu bằng biểu đồ hoặc hộp biểu đồ. Đây là điều bắt buộc phải có để xử lý dữ liệu, thao tác và trực quan hóa.

d. Thư viện Matplotlib

Đây là một thư viện khoa học dữ liệu tiêu chuẩn giúp tạo ra các trực quan hóa dữ liệu như biểu đồ và biểu đồ hai chiều (biểu đồ, biểu đồ phân tán, biểu đồ tọa độ phi Descartes). Matplotlib là một trong những thư viện vẽ biểu đồ thực sự hữu ích trong các dự án khoa học dữ liệu - nó cung cấp một API hướng đối tượng để nhúng các biểu đồ vào ứng dụng.

e. Thư viện Keras

Keras là một thư viện tuyệt vời để xây dựng mạng nơ-ron và mô hình hóa. Nó rất dễ sử dụng và cung cấp cho các nhà phát triển một mức độ mở rộng tốt. Thư viện tận dụng các gói khác (**Theano** hoặc **TensorFlow**) làm phụ trợ của nó.
