

Ambiente Data Science

Primeiramente precisamos configurar nosso ambiente para podermos trabalhar com **Data Science**!! E precisaremos de **3 ferramentas** instaladas em nossa máquina:

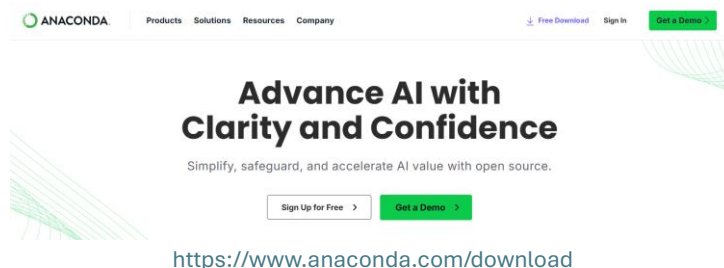
- Linguagem de programação **Python**.
- IDE de uso preferencial.
- **Librarys Python** necessárias.

Plataforma Anaconda



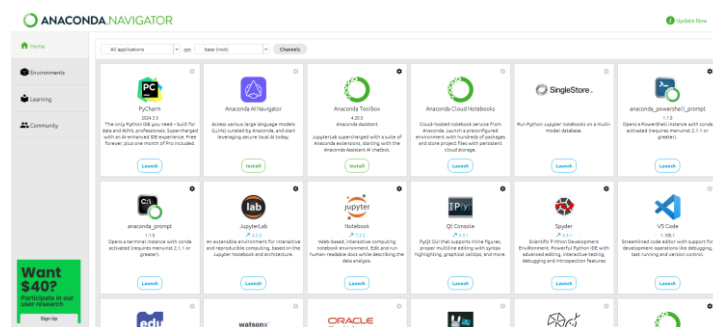
Basicamente se trata de um **tollkit**, onde será responsável por instalar **TUDO** para nós! Logo será instalado o **Python**, uma **IDE** (que no nosso caso será o **Jupyter**), e todas as **Librays Python** necessárias que precisamos ter para trabalhar com **Data Science** e **Machine Learning**!

Dito isso, segue o link para a instalação do **Anaconda**:

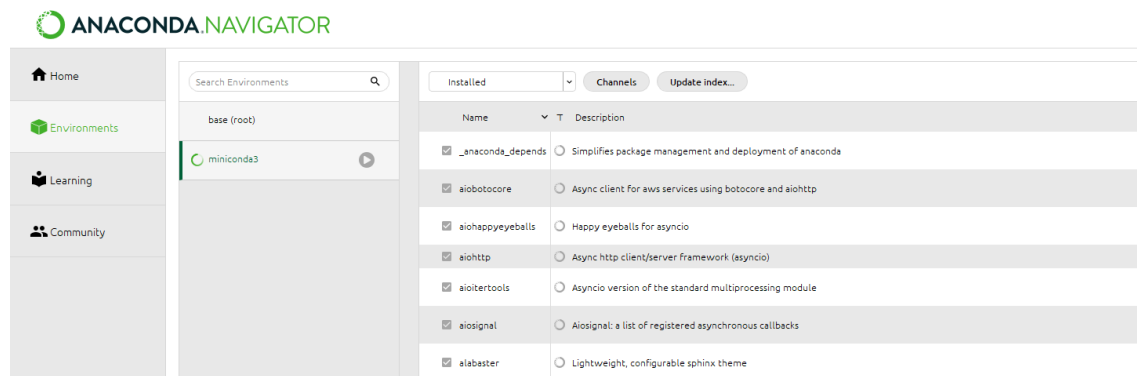


Entendendo melhor a plataforma Anaconda

Após a instalação do **Anaconda**, veremos que foi instalado um software chamado **Anaconda Navigator**, onde podemos instalar outros softwares ou iniciar caso já estiver instalado! E nesse caso iremos utilizar a **IDE JupyterLab**!!



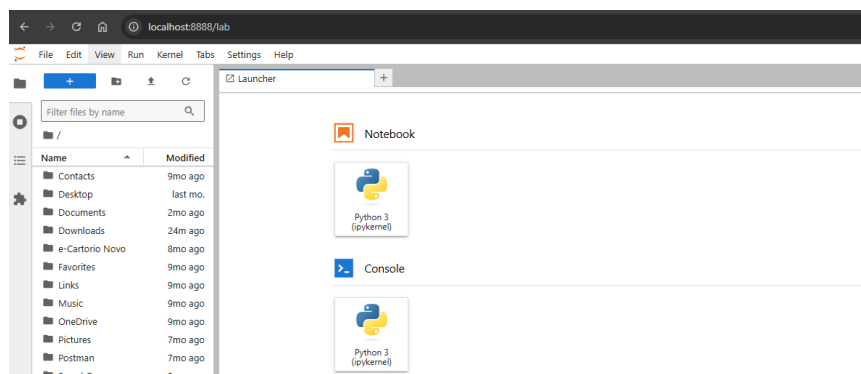
Mas antes de iniciar o **JupyterLab**, vamos entender um pouco mais do **Anaconda Navigator**!! E se notarmos bem, temos algumas opções lado esquerdo, entre eles está o **environment** do **Python**!!



E vale destacar de forma bem resumida que o **environment** do **Python** funciona como uma **instância isolada do Python**, com sua **própria versão do Python e dos pacotes**. Ou seja, dessa forma podemos trabalhar com vários projetos do **Python** com suas versões isoladas, sendo assim não afetando outros projetos!!

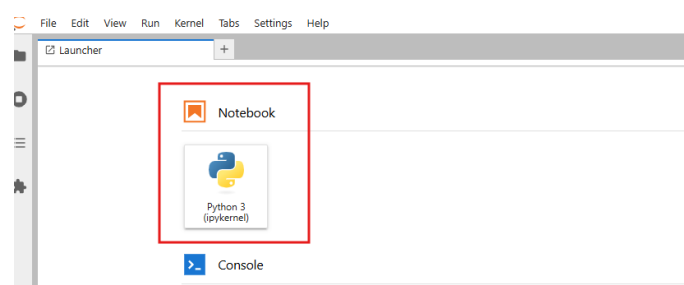
Iniciando o JupyterLab

Agora sim, vamos iniciar a nossa IDE do **JupyterLab**, e quando for aberto pelo Anaconda, teremos a seguinte cara!

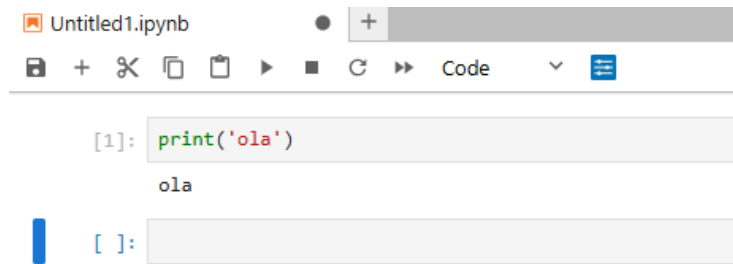


Veja que será aberto uma **interface web**, mas não se engane!! A **IDE** rodará em sua máquina local, tanto que a url está apontando para o **localhost**!

Veja que podemos iniciar criando 2 tipos de arquivos, **Tipo Notebook** e **Tipo Console**, e no nosso caso será do **Tipo Notebook**.

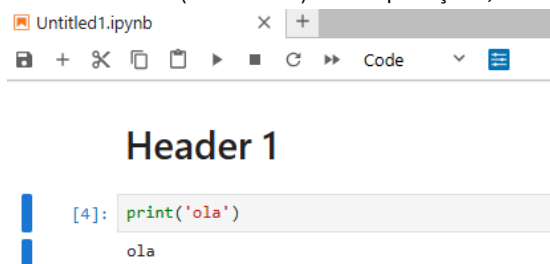


E logo quando abrimos o arquivo, veremos algo diferente, pois se trata de um arquivo de **células**, e é nessas **células** onde colocaremos o nosso **código python** que será executado!! Como podemos ver a seguir:



The screenshot shows a Jupyter Notebook window titled 'Untitled1.ipynb'. The toolbar includes icons for saving, adding, deleting, copying, pasting, running, and other actions. Below the toolbar, there is a code cell labeled '[1]:' containing the Python code `print('ola')`. The output of this cell is 'ola'. Below the first cell, there is a second cell labeled '[]:' which is currently empty.

Esse formato de arquivo que estamos fazendo pelo **Jupyter**, em colocar **código python** em **células**, pode ser um pouco confuso no início, mas ele é essencial quando trabalhamos com **Data Science**!! Isso porque:

- **Temos mais Interatividade**
 - No **Jupyter**, você pode rodar célula individualmente, o que te permite:
 - Testar pequenos trechos de código, sem precisar rodar tudo de novo.
 - Fazer análises passo a passo.
 - Ver resultado da execução por blocos separados.
 - **Mistura de código e explicação**
 - Com o **Jupyter**, podemos ter:
 - Células de código (**Python**)
 - Células de texto (**Markdown**) com explicações, fórmulas e até mesmo **gráficos**
- 
- The screenshot shows a Jupyter Notebook window titled 'Untitled1.ipynb'. The toolbar is visible. Below the toolbar, there is a Markdown cell containing the text 'Header 1'. Below the Markdown cell, there is a code cell labeled '[4]:' containing the Python code `print('ola')`. The output of this cell is 'ola'.
- **Perfeito para Data Science**, pois precisamos explicar o nosso raciocínio junto com os dados!
- **Visualização de dados**
 - Gráficos do **matplotlib** ou tabelas **pandas** aparece direto abaixo da célula, sem precisar abrir janelas extras.
- **Exploração de dados em tempo real**
 - Como você trabalha de forma interativa, podemos:
 - Testar hipóteses
 - Ajustar filtros e ver na hora.
 - Repetir análises sem precisar reiniciar tudo.