

## PRÉ- REQUISITOS

- Python 3

## BIBLIOTECAS/FERRAMENTAS QUE USAREMOS

- pandas (pip install pandas)
- jupyter notebook (pip install notebook)

## PARA CONFIGURAR UM AMBIENTE VIRTUAL (venv)

<https://pythonacademy.com.br/blog/python-e-virtualenv-como-programar-em-ambientes-virtuais>

## CRIAR PASTA PARA O PROJETO E ADICIONAR ARQUIVOS CSV (DATASETS)

- nomes\_africanos\_1.csv
- nomes\_africanos\_2.csv

Arquivos oriundos de: <https://www.slavevoyages.org/resources/names-database>

## INSTALAÇÃO

```
pip install great_expectations
```

```
great_expectations --version
```

## TROUBLESHOOTING (UBUNTU 20)

ERROR: Cannot uninstall 'ruamel-yaml'. It is a distutils installed project and thus we cannot accurately determine which files belong to it which would lead to only a partial uninstall.

```
pip install great_expectations --ignore-installed ruamel.yaml
```

## INICIAR O GREAT EXPECTATIONS E EXECUTAR A CONFIGURAÇÃO INICIAL

```
cd <sua_pasta_do_projeto>
```

### CRIAR DATA CONTEXT

```
great_expectations --v3-api init
```

Let's create a new Data Context to hold your project configuration.

Great Expectations will create a new directory with the following structure:

OK to proceed? [Y/n]: Y

### CONECTAR DATASOURCE

```
great_expectations --v3-api datasource new
```

What data would you like Great Expectations to connect to?

1. Files on a filesystem (for processing with Pandas or Spark)
2. Relational database (SQL)

1

What are you processing your files with?

1. Pandas
2. PySpark

1

Enter the path (relative or absolute) of the root directory where the data files are stored.

data

*Um notebook será aberto: "datasource new"*

- **Alterar célula:** datasource\_name = "nomes\_africanos"
- Executar notebook

**Deletar arquivo:** great\_expectations/uncommitted/datasource\_new.ipynb

### EDITAR EXPECATIONS SUITE

```
great_expectations --v3-api suite new
```

How would you like to create your Expectation Suite?

1. Manually, without interacting with a sample batch of data (default)
2. Interactively, with a sample batch of data
3. Automatically, using a profiler

3

Which data asset (accessible by data connector "default\_inferred\_data\_connector\_name") would you like to use?

1. nomes\_africanos\_1.csv
2. nomes\_africanos\_2.csv

: 1

Name the new Expectation Suite [nomes\_africanos\_1.csv.warning]: **nomes**

When you run this notebook, Great Expectations will store these expectations in a new Expectation Suite "nomes" here:

file:///media/joamila/HD  
Joamila/GreatExpectations/nomes\_africanos/great\_expectations/expectations/nomes.json

Would you like to proceed? [Y/n]: Y

## CONTATO INICIAL COM PROFILE E DATA DOCS

- Um notebook será aberto
- Comentar as colunas importantes na célula: *ignored\_columns*
- Executar notebook
- A página html que abre é o Data Docs.

## EDITAR AS EXPECTATIVAS

great\_expectations --v3-api suite edit *nomes*

How would you like to edit your Expectation Suite?

1. Manually, without interacting with a sample batch of data (default)
2. Interactively, with a sample batch of data

2

A batch of data is required to edit the suite - let's help you to specify it.

Which data asset (accessible by data connector "default\_inferred\_data\_connector\_name") would you like to use?

1. nomes\_africanos\_1.csv

2. nomes\_africanos\_2.csv

1

Editar notebook seguindo o glossário:

[https://docs.greatexpectations.io/docs/reference/glossary\\_of\\_expectations](https://docs.greatexpectations.io/docs/reference/glossary_of_expectations)

[https://legacy.docs.greatexpectations.io/en/latest/reference/glossary\\_of\\_expectations.html](https://legacy.docs.greatexpectations.io/en/latest/reference/glossary_of_expectations.html)

Obs: mudar o valor de *expectation\_suite\_name* = "nomes1" para que as atualizações de deletar sejam refletidas no data docs.

## CRIAR CHECKPOINT PARA VALIDAR NOVOS DADOS

```
great_expectations --v3-api checkpoint new checkpoint1
```

Fazer mudanças no notebook que irá abrir, se necessário em:

- data\_asset\_name:
- expectation\_suite\_name:

Executar notebook e verificar data docs

**TUTORIAL CONCLUÍDO!!!**