

07

Mão na massa: Extrair dados

Chegou a hora de você executar o que foi visto na aula! Para isso, execute os passos listados abaixo.

- 1) O site do [Weather Underground](https://www.wunderground.com/) (<https://www.wunderground.com/>) coleta dados sobre o clima e de tempos em tempos os grava. E para capturar os dados históricos, eles disponibilizam uma API, a [Weather API for Developers](https://www.wunderground.com/weather/api) (<https://www.wunderground.com/weather/api>).
- 2) Essa API possui alguns [planos](https://www.wunderground.com/weather/api/d/pricing.html) (<https://www.wunderground.com/weather/api/d/pricing.html>), incluindo um gratuito, onde você pode fazer até 10 chamadas por minuto, sendo até 500 chamadas por dia. Então, caso você queira, compre uma chave, que mesmo sendo gratuita, deve-ser clicar em **Purchase Key**.
- 3) Além do acesso interativo, você pode acessar os dados via URL, por exemplo:
http://api.wunderground.com/api/SUA_CHAVE/history_20160119/q/CA/San_Francisco.json
(http://api.wunderground.com/api/SUA_CHAVE/history_20160119/q/CA/San_Francisco.json).
- 4) Há um robô, desenvolvido pelo instrutor, escrito em Python, que a cada 30 minutos captura 10 JSONs do **Weather Underground**. Você pode baixá-lo [aqui](https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Coleta_Dados_Clima_v2.zip) (https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Coleta_Dados_Clima_v2.zip) e a biblioteca de utilidades pode ser baixada [aqui](https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Util.zip) (<https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Util.zip>).
- 5) Esse robô utiliza um arquivo CSV com cidades e datas, para fazer a consulta no **Weather Underground**. Você pode baixá-lo [aqui](https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Cidades_Datas_Distintas.csv) (https://s3.amazonaws.com/caelum-online-public/787-data-science-coleta-de-dados/02/Cidades_Datas_Distintas.csv).
- 6) Caso você queira executar tudo isso, o ideal seria ter um servidor que "nunca" desligasse com Python e MongoDB, criar agenda, adquirir chave da API mencionada anteriormente e colocá-la no código. Mas neste curso você receberá códigos fontes e backups para poder executar tudo local.