Scraping Booking.com using Python

Challenges:

- Bypassing the bot checks
- Fetching as much data as needed
- Data Analysis

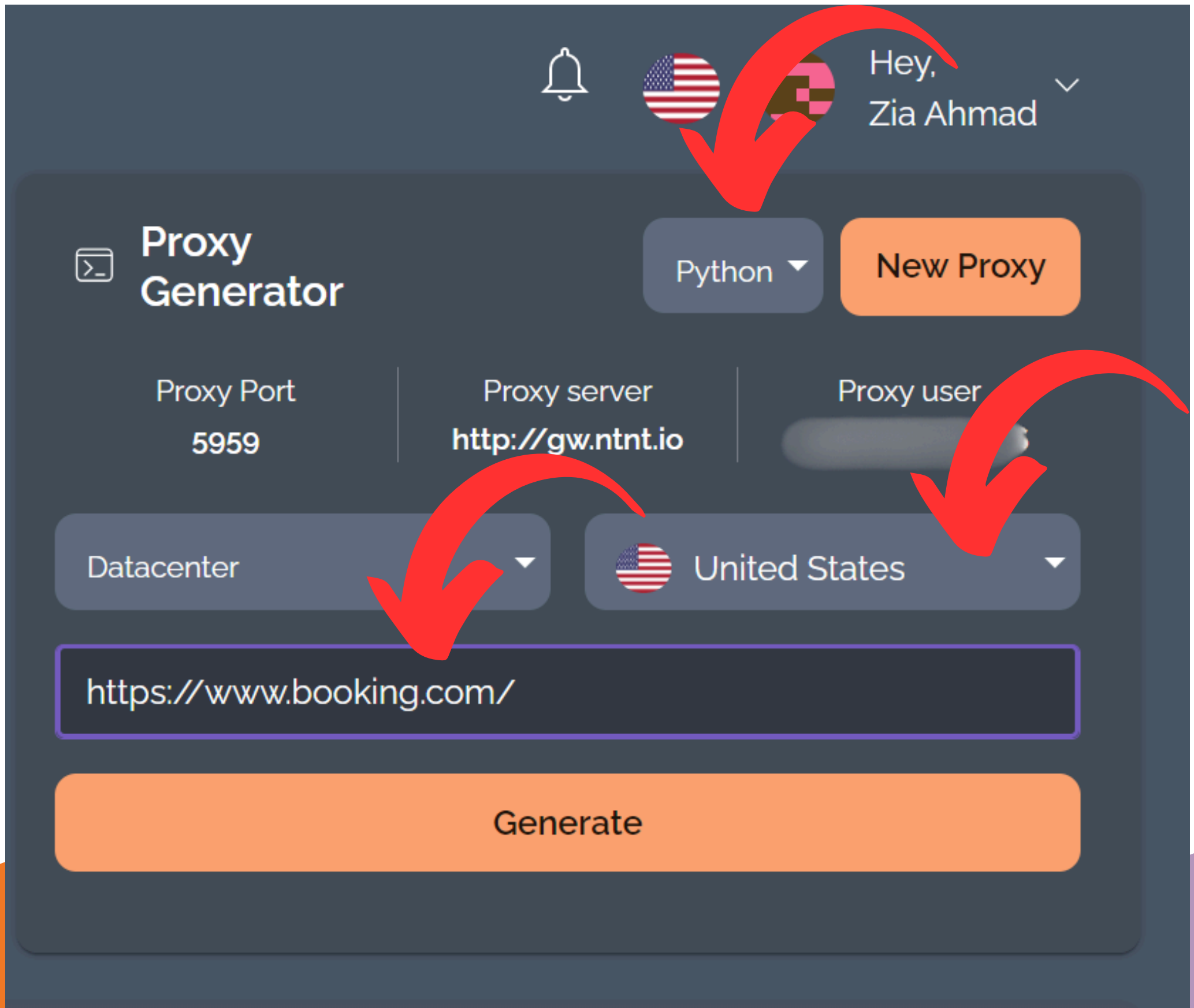
BENEFITS

- **Data Access:** Quickly gather large datasets from Booking.com.
- **Cost-Effective:** Saves money compared to manual data collection.
- **Speed:** Rapidly extract data for quick insights.
- **Competitive Advantage:** Gain insights into competitors' data.
- **Automation:** Automate data collection, saving time.

Sign up to:

https://portal.netnut.io/

Pick your preferred Programming Language, Location & URL



After hitting Generate button Copy this Code & Paste into the Coding Environment

Proxy Generator

Python

New Proxy

Here is a basic CURL request to our Datacenter network. Copy and paste the code below into your cmd (win users) or terminal (mac users).

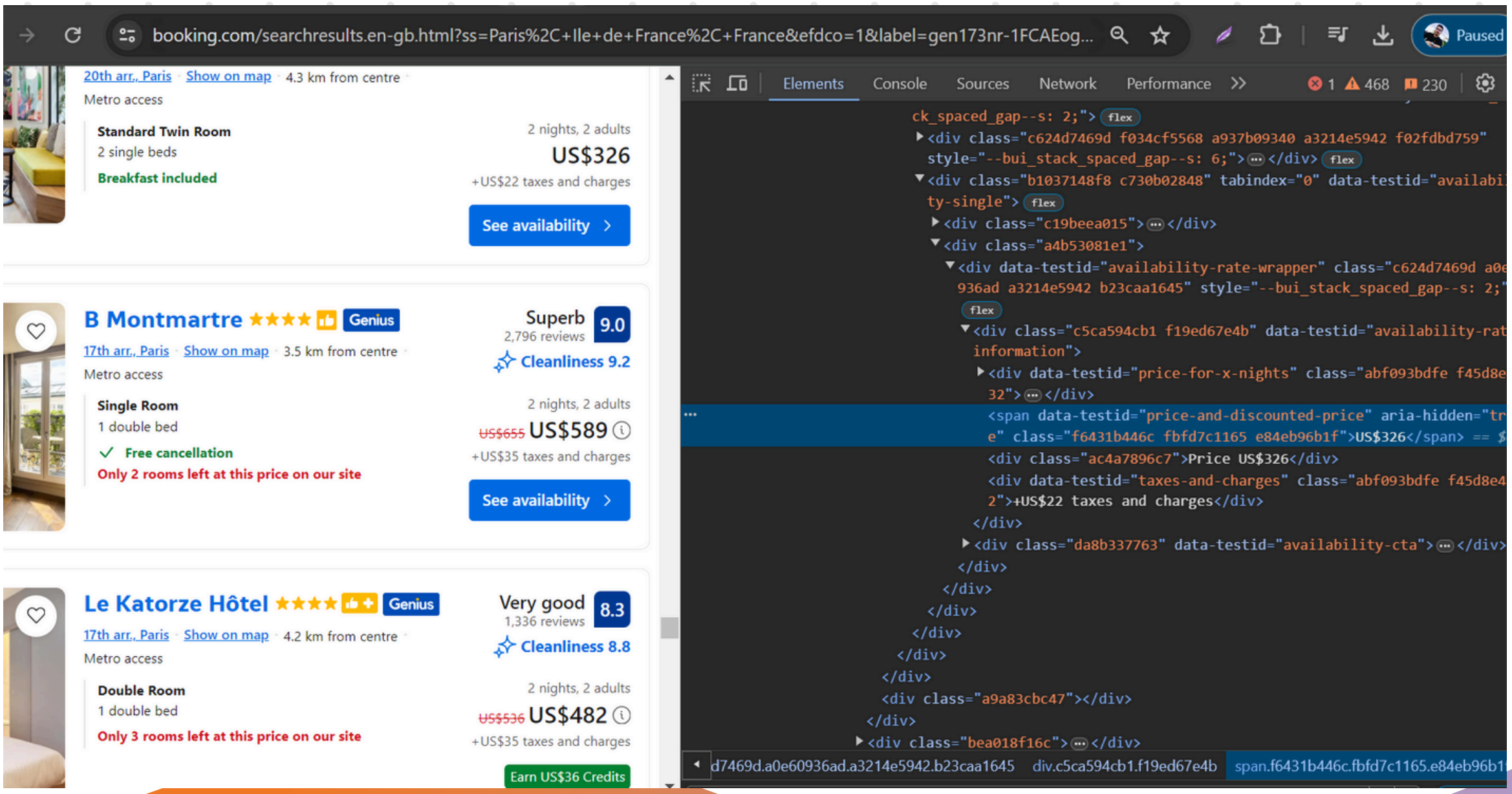
Copy code

```
    # to use this example you must have
'requests' installed: pip install requests
import requests

username = '
password = '
server = 'gw.ntnt.io'
```

- **Go the Booking.com**
- **Select Destination & related fields**
- **Click Search button**

- **Copy the URL after search**
- **Go to Inspect**
- **Copy the classes of the targeted values**



Importing Libraries

```
import requests
from bs4 import BeautifulSoup
```

Setting up the Proxy

```
username = 'ziaahmad-dc-US'
password = 'r8RNn8Qw6rDjcdM'
server = 'gw.ntnt.io'
port = '5959'
proxy = {'http': f'http://{username}:{password}@{server}:{port}'}

headers = {'User-Agent': 'Chrome/58.0.3029.110'}
url = "https://www.booking.com/searchresults.en-gb.html?ss=Paris"
response = requests.get(url, proxies=proxy, headers=headers)
```

Code for Data Mining

```
soup = BeautifulSoup(driver.page_source, 'html.parser')
post_cards = soup.find_all("div", "c624d7469d a0e60936ad a3214e5942 b0db0e8ada")
data = []
for i in post_cards:
    dct = {}
    try:
        dct['Hotel Name'] = i.find(class_="f6431b446c a15b38c233").get_text()
        dct['Price'] = i.find(class_="f6431b446c fbfd7c1165 e84eb96b1f").get_text()
        dct['Rating'] = i.find(class_="a3b8729ab1 d86cee9b25").get_text().split()[-1]
        dct['Room Type'] = i.find(class_="abf093bdfe e8f7c070a7").get_text()
        data.append(dct)
    except:
        print("Didn't find the Text")

pd.DataFrame(data)
```

Paste the classes that we copied at Step 5

Sample Output

Hotel Name	Price	Rating	Room Type
Royal Madeleine Hotel & Spa	US$789	8.5	Classic Double Room
HotelF1 Paris Saint Ouen Marché Aux Puces	US$165	6.4	Side-Car Room
Grand Hotel Dore	US$412	8.2	Double Room
Hôtel Aiglon	US$516	8.7	Cosy Double Room
Gardette Park Hotel	US$535	8.1	Standard Double Room
Hotel Elysees Opera	US$682	8.5	Superior Double Room
NH Paris Gare de l'Est	US$585	7.7	Standard Room
La Belle Ville	US$550	8.7	Double Room
Holiday Inn Express Paris-Canal De La Villette...	US$461	8.1	Standard Room
Hotel Residence Foch	US$568	8.2	Traditional Double or Twin Room
Grand Hotel Francais	US$467	8.5	Superior Room
Hôtel Korner Etoile	US$501	8.3	Twin Room
Best Western Bretagne Montparnasse	US$389	8.3	Superior Queen Room - Non-Smoking
Hôtel Le Daum	US$544	8.4	Comfort Double Room
Hotel de L'Esperance	US$729	8.5	Junior Suite
Hotel Armoni Paris	US$447	8.1	Standard Double Room
hotelF1 Paris Porte de Châtillon	US$185	7.3	Side-Car Room
Le Katorze Hôtel	US$624	8.3	Double Room
Hotel le 18 Paris	US$360	7.7	Comfort Double Room
Kyriad Paris 18 - Porte de Clignancourt - Mont...	US$331	8.1	Double Room

Summary

- Create an account
- Generate the script for proxy at Netnut.io
- Copy the desired URL
- Open any Python IDE
- Paste the script that we copied from netnut
- Edit the URL
- Mine the Data According to the need

Scraping Booking.com using Python

Hit Follow!

Save for later!

netnut.io