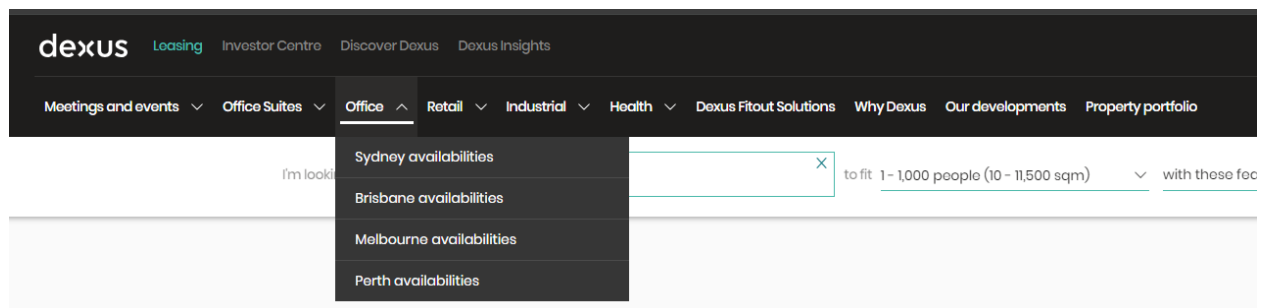# Data Scraping Job Template – 15141

⚠️ **If the resource you are scraping requires you to agree to any Terms & Conditions, please do not proceed and notify your contract manager immediately. Under no circumstances should you create a false account or fake identity.**

Description:

Please can we scrape down all availabilities for the properties listed below

Step 1: Find the property link from the website https://www.dexus.com/leasing/office

Central Place Sydney
A vibrant new business district, revitalising the face of Sydney's busiest transport interchange. This $2.6 billion investment will deliver Sydney's most significant CBD development in a decade.
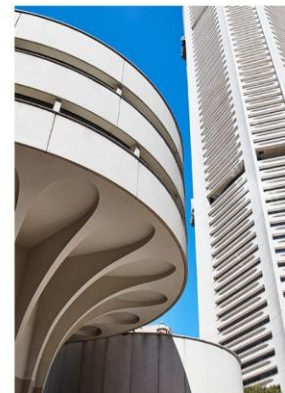
80 Collins St, Melbourne
Featuring two distinctive office towers, a boutique hotel and a contemporary, luxury retail precinct completed in 2020.

60 Collins St, Melbourne
A unique development opportunity located at the 'Paris End' of Collins Street. 60 Collins Street will deliver approximately 42,000sqm of Premium Grade office space.

25 Martin Place, Sydney
An icon of the Sydney skyline, 25 Martin Place (previously the MLC Centre) has been an Australian living landmark for over 40 years.

Step 2: Go to the property page (e.g. https://www.dexus.com/leasing/office/80-collins-street-south-tower) and go to "Availability" and scrape down all availabilities

80 Collins Street (South Tower), Melbourne

Availability Gallery Features Services Location

Availability

Find your space

Level	Space Options	Availability	Price from, pa	Outgoings, pa	Floor area	Fit-out
Part Level 37, South Tower	Part floor	Available now	POA	$189 psm	374 sqm – 31-50 people	

Property List:

80 Collins
Australia Square
130 George
130 Harris
25 Martin
1 Farrer
King Square
Waterfront
The Bond
385 Bourke
480 Queen Street
240 St George
1 Bligh Street
100 Mount
60 Castlereagh

Root URL:

https://www.dexus.com/

Job Frequency:

Daily

Output Columns:

Column name	Datatype	Xpath/location on site/number in screenshot	Example value	Comment
scrape_datetime	*datetime*	N/A	datetime.datetime.utcnow().isoformat()	*Ensure this is a datetime in ISO-8601 format, and have one value per run, evaluated at start of the script, rather than at the time of each request*
data_url	*string*		https://www.dexus.com/leasing/office/80-collins-street-south-tower	*Property page link*
Address	String	1	80 Collins Street (South Tower), Melbourne	*Address*
Level	String	2	Part Level 37, South Tower	
Space options	String	3	Part floor	
Availability	String	4	Available now	
Price from	String	5	POA	
Outgoings	String	6	$189 psm	
Floor area	String	7	374 sqm - 31-50 people	
Fitout	String	8		

Timeline:

<mark>You may complete this job any time and submit any required files to the linked Github repository within one week of accepting the job.</mark>

Please submit your code here: https://github.com/International-Data-Repository-IPD/scrape-15141

Submission Files:

       Sample.csv for sample data

A requirement.txt

scrape/ - containing all of the source code

Main file: scrape.py that will be run with a output $filename.

Job Schema/Output Format:

You should save the output csv using these settings from a pandas DataFrame:

```
encoding="utf-8",
 line_terminator="\n",
 quotechar='"',
 quoting=csv.QUOTE_ALL,
 index=False
```

Package preferences and Scraping notes

Selenium is a last resort (only purposeful obfuscation from website owner)

Generally unnecessary to run a full browser

Please provide reasoning if using Selenium

Beautiful soup

Avoid hard codes, references should be relative

Parse json from an api rather than html

Runtime Environment:

Your code will be copied form the root to/usr/src/scrape

You should feel free to modify the requirements as you need. However, you must keep the awscli dependency

You may also upload additional binaries into the repository root and reference them there.

Please do not change the Dockerfile or shell scripts in the repository as this will cause automated test failure.

python scrape.py $filename

Page access limitations (max requests / day):

If you encounter a captcha during your scrape job, please contact the job poster before continuing.

10% of website traffic max