Collecting Job Data Using APIs

Estimated time needed: **30** minutes

Objectives

After completing this lab, you will be able to:

- Collect job data using Jobs API
- Store the collected data into an excel spreadsheet.

> **Note: Before starting with the assignment make sure to read all the instructions and then move ahead with the coding part.**

Instructions

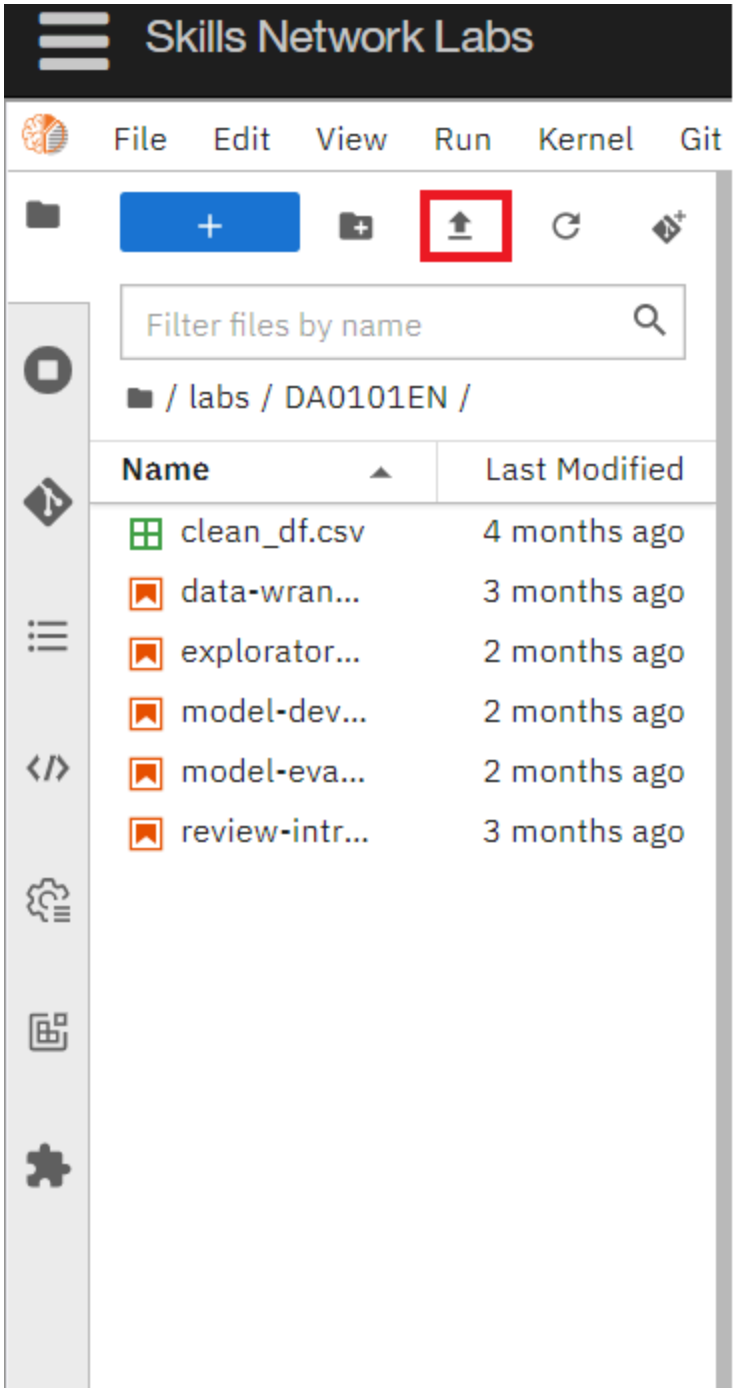
To run the actual lab, firstly you need to click on the Jobs_API notebook link. The file contains flask code which is required to run the Jobs API data.

Now, to run the code in the file that opens up follow the below steps.

Step1: Download the file.

Step2: Upload the file into your current Jupyter environment using the upload button in your Jupyter interface. Ensure that the file is in the same folder as your working .ipynb file.

Step 2: If working in a local Jupyter environment, use the "Upload" button in your Jupyter interface to upload the Jobs_API notebook into the same folder as your current .ipynb file.



Step3: Open the Jobs_API notebook, and run all the cells to start the Flask application. Once the server is running, you can access the API from the URL provided in the notebook.

If you want to learn more about flask, which is optional, you can click on this link here.

Once you run the flask code, you can start with your assignment.

Dataset Used in this Assignment

The dataset used in this lab comes from the following source: https://www.kaggle.com/promptcloud/jobs-on-naukricom under the under a **Public**

Domain license.

> Note: We are using a modified subset of that dataset for the lab, so to follow the lab instructions successfully please use the dataset provided with the lab, rather than the dataset from the original source.

The original dataset is a csv. We have converted the csv to json as per the requirement of the lab.

Warm-Up Exercise

Before you attempt the actual lab, here is a fully solved warmup exercise that will help you to learn how to access an API.

Using an API, let us find out who currently are on the International Space Station (ISS).
The API at http://api.open-notify.org/astros.json gives us the information of astronauts currently on ISS in json format.
You can read more about this API at http://open-notify.org/Open-Notify-API/People-In-Space/

In [52]:
```
import requests # you need this module to make an API call
import pandas as pd
```

In [53]:
```
api_url = "http://api.open-notify.org/astros.json" # this url gives use the astrona
```

In [54]:
```
response = requests.get(api_url) # Call the API using the get method and store the
                                 # output of the API call in a variable called respo
```

In [55]:
```
if response.ok:              # if all is well() no errors, no network timeouts)
    data = response.json()   # store the result in json format in a variable called
                             # the variable data is of type dictionary.
```

In [56]:
```
print(data)   # print the data just to check the output or for debugging
```

```
{'people': [{'craft': 'ISS', 'name': 'Oleg Kononenko'}, {'craft': 'ISS', 'name': 'Ni
kolai Chub'}, {'craft': 'ISS', 'name': 'Tracy Caldwell Dyson'}, {'craft': 'ISS', 'na
me': 'Matthew Dominick'}, {'craft': 'ISS', 'name': 'Michael Barratt'}, {'craft': 'IS
S', 'name': 'Jeanette Epps'}, {'craft': 'ISS', 'name': 'Alexander Grebenkin'}, {'cra
ft': 'ISS', 'name': 'Butch Wilmore'}, {'craft': 'ISS', 'name': 'Sunita Williams'},
{'craft': 'Tiangong', 'name': 'Li Guangsu'}, {'craft': 'Tiangong', 'name': 'Li Con
g'}, {'craft': 'Tiangong', 'name': 'Ye Guangfu'}], 'number': 12, 'message': 'succes
s'}
```

Print the number of astronauts currently on ISS.

In [57]:
```
print(data.get('number'))
```

```
12
```

Print the names of the astronauts currently on ISS.

In [58]:
```
astronauts = data.get('people')
print("There are {} astronauts on ISS".format(len(astronauts)))
print("And their names are :")
for astronaut in astronauts:
    print(astronaut.get('name'))
```

```
There are 12 astronauts on ISS
And their names are :
Oleg Kononenko
Nikolai Chub
Tracy Caldwell Dyson
Matthew Dominick
Michael Barratt
Jeanette Epps
Alexander Grebenkin
Butch Wilmore
Sunita Williams
Li Guangsu
Li Cong
Ye Guangfu
```

Hope the warmup was helpful. Good luck with your next lab!

Lab: Collect Jobs Data using Jobs API

Objective: Determine the number of jobs currently open for various technologies and for various locations

Collect the number of job postings for the following locations using the API:

- Los Angeles
- New York
- San Francisco
- Washington DC
- Seattle
- Austin
- Detroit

In [59]:
```
#Import required libraries
import pandas as pd
import json
```

Write a function to get the number of jobs for the Python technology.

> Note: While using the lab you need to pass the **payload** information for the **params** attribute in the form of **key value** pairs.

Refer the ungraded **rest api lab** in the course **Python for Data Science, AI & Development** link

The keys in the json are

- Job Title
- Job Experience Required
- Key Skills
- Role Category
- Location
- Functional Area
- Industry
- Role

You can also view the json file contents from the following json URL.

In [60]:
```
api_url="http://127.0.0.1:5000/data"
import requests

def get_number_of_jobs_T(technology):
    api_url = "http://127.0.0.1:5000/data"
    params = {"Key Skills": technology}
    response = requests.get(api_url, params=params)
    if response.ok:
        data = response.json()
        return technology, len(data)
    else:
        return technology, 0
```

Calling the function for Python and checking if it works.

In [61]:
```
get_number_of_jobs_T("Python")
```

Out[61]:
```
('Python', 1173)
```

Write a function to find number of jobs in US for a location of your choice

In [62]:
```
def get_number_of_jobs_L(location):
    import requests  # Import requests library for API calls

    # API endpoint
    api_url = "http://127.0.0.1:5000/data"

    # Query parameters with the chosen location
    params = {"Location": location}
```

```
    # Make a GET request to the API
    response = requests.get(api_url, params=params)

    if response.ok:
        # Parse the response JSON
        data = response.json()

        # Return the location and the count of jobs found
        return location, len(data)
    else:
        # Handle failed API requests
        print(f"Error: {response.status_code} - {response.text}")
        return location, 0
```

Call the function for Los Angeles and check if it is working.

In [63]:
```
location, number_of_jobs = get_number_of_jobs_L("Los Angeles")
print(f"There are {number_of_jobs} jobs in {location}.")#your code goes here
```

```
There are 640 jobs in Los Angeles.
```

Store the results in an excel file

Call the API for all the given technologies above and write the results in an excel spreadsheet.

If you do not know how create excel file using python, double click here for **hints**.

Create a python list of all technologies for which you need to find the number of jobs postings.

In [64]:
```
# List of technologies to query
technologies = [
    "C", "C#", "C++", "Java", "JavaScript",
    "Python", "Scala", "Oracle", "SQL Server",
    "MySQL Server", "PostgreSQL", "MongoDB"
]
```

Import libraries required to create excel spreadsheet

In [66]:
```
!pip install openpyxl
from openpyxl import Workbook
```

```
Collecting openpyxl
  Downloading openpyxl-3.1.3-py2.py3-none-any.whl (251 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 251.3/251.3 kB 27.3 MB/s eta 0:00:00
Collecting et-xmlfile (from openpyxl)
  Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
Installing collected packages: et-xmlfile, openpyxl
Successfully installed et-xmlfile-1.1.0 openpyxl-3.1.3
```

Create a workbook and select the active worksheet

In [67]:
```
# Step 1: Import Workbook from openpyxl
from openpyxl import Workbook

# Step 2: Create a new workbook
wb = Workbook()

# Step 3: Select the active worksheet
ws = wb.active

# Step 4: Rename the worksheet
ws.title = "Job Postings"

# Step 5: Save the workbook to verify
wb.save("example_openpyxl.xlsx")
```

Find the number of jobs postings for each of the technology in the above list. Write the technology name and the number of jobs postings into the excel spreadsheet.

In [68]:
```
import requests
from openpyxl import Workbook

# List of technologies
technologies = [
    "C", "C#", "C++", "Java", "JavaScript",
    "Python", "Scala", "Oracle", "SQL Server",
    "MySQL Server", "PostgreSQL", "MongoDB"
]

# Function to get job postings for a technology
def get_number_of_jobs_T(technology):
    api_url = "http://127.0.0.1:5000/data"  # Replace with your API endpoint
    params = {"Key Skills": technology}  # Pass the technology as a parameter
    response = requests.get(api_url, params=params)
    if response.ok:
        data = response.json()  # Parse the response JSON
        return technology, len(data)  # Return the technology and number of jobs
    else:
        print(f"Error: {response.status_code} - {response.text}")
        return technology, 0  # If the request fails, return 0 jobs

# Collect job postings data
job_postings = [get_number_of_jobs_T(tech) for tech in technologies]
```

Save into an excel spreadsheet named **job-postings.xlsx**.

In [73]:
```
# Write results to Excel
wb = Workbook()
ws = wb.active
ws.title = "Job Postings"
ws.append(["Technology", "Job Postings"])
for tech, count in job_postings:
```

```
    ws.append([tech, count])
wb.save("job-postings.xlsx")
```

In the similar way, you can try for below given technologies and results can be stored in an excel sheet.

Collect the number of job postings for the following languages using the API:

- C
- C#
- C++
- Java
- JavaScript
- Python
- Scala
- Oracle
- SQL Server
- MySQL Server
- PostgreSQL
- MongoDB

In [74]:
```
import requests
from openpyxl import Workbook

# List of programming languages
languages = [
    "C", "C#", "C++", "Java", "JavaScript",
    "Python", "Scala", "Oracle", "SQL Server",
    "MySQL Server", "PostgreSQL", "MongoDB"
]

# Function to get job postings for a language
def get_number_of_jobs(language):
    api_url = "http://127.0.0.1:5000/data"  # Replace with your API endpoint
    params = {"Key Skills": language}  # Pass the language as a query parameter
    response = requests.get(api_url, params=params)
    if response.ok:
        data = response.json()  # Parse the response JSON
        return language, len(data)  # Return the language and the count of jobs
    else:
        print(f"Error: {response.status_code} - {response.text}")
        return language, 0  # Return 0 jobs if the request fails

# Collect job postings data
job_postings = [get_number_of_jobs(lang) for lang in languages]

# Create an Excel workbook
wb = Workbook()
ws = wb.active
ws.title = "Programming Languages"
ws.append(["Language", "Job Postings"])
```

```
for lang, count in job_postings:
    ws.append([lang, count])
wb.save("job-postings-languages.xlsx")
```

Authors

Ayushi Jain

Other Contributors

Rav Ahuja

Lakshmi Holla

Malika

<!--## Change Log

<!--| Date (YYYY-MM-DD) | Version | Changed By | Change Description | | ----------------- | ------ | ---------------- | -------------------------------- | | 2022-01-19 | 0.3 | Lakshmi Holla | Added changes in the markdown | | 2021-06-25 | 0.2 | Malika | Updated GitHub job json link | | 2020-10-17 | 0.1 | Ramesh Sannareddy | Created initial version of the lab |--!>