    <a
href="https://skills.network/?utm_medium=Exinfluencer&utm_source=Exinflue
ncer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-
Channel-SkillsNetworkCoursesIBMDA0321ENSkillsNetwork928-2022-01-01"
target="_blank">
    <img src="https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/assets/logos/SN_web_lightmode.png" width="200"
alt="Skills Network Logo" />

Collecting Job Data Using APIs

Estimated time needed: **45 to 60** minutes

Objectives

After completing this lab, you will be able to:

*   Collect job data from 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](https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBM-DA0321EN-
SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/Jobs_API.ip
ynb) 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 it on the IBM Watson studio. (If IBM Watson Cloud service
does not work in your system, follow the alternate Step 2 below)

Step2(alternate): Upload it in your SN labs environment using the upload
button which is highlighted in red in the image below:
Remember to upload this Jobs_API file in the same folder as your current
.ipynb file

<img src="https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/Upload.PNG" >

Step3: Run all the cells of the Jobs_API file. (Even if you receive an asterik sign after running the last cell, the code works fine.)

If you want to learn more about flask, which is optional, you can click on this link [here](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/FLASK_API.md.html).

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](http://api.open-notify.org/astros.json?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDA0321ENSkillsNetwork21426264-2021-01-01&cm_mmc=Email_Newsletter-_-Developer_Ed%2BTech-_-WW_WW-_-SkillsNetwork-Courses-IBM-DA0321EN-SkillsNetwork-21426264&cm_mmca1=000026UJ&cm_mmca2=10006555&cm_mmca3=M12345678&cvosrc=email.Newsletter.M12345678&cvo_campaign=000026UJ) 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/](http://open-notify.org/Open-Notify-API/People-In-Space?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-

SkillsNetworkCoursesIBMDA0321ENSkillsNetwork21426264-2021-01-01&cm_mmc=Email_Newsletter-_-Developer_Ed%2BTech-_-WW_WW-_-SkillsNetwork-Courses-IBM-DA0321EN-SkillsNetwork-21426264&cm_mmca1=000026UJ&cm_mmca2=10006555&cm_mmca3=M12345678&cvosrc=email.Newsletter.M12345678&cvo_campaign=000026UJ)

```python
import requests # you need this module to make an API call
import pandas as pd
```

```python
api_url = "http://api.open-notify.org/astros.json" # this url gives use the astronaut data
```

```python
response = requests.get(api_url) # Call the API using the get method and store the
                                 # output of the API call in a variable called response.
```

```python
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 data
                            # the variable data is of type dictionary.
```

```python

```

```python
print(data)   # print the data just to check the output or for debugging
```

{'people': [{'craft': 'ISS', 'name': 'Oleg Kononenko'}, {'craft': 'ISS', 'name': 'Nikolai Chub'}, {'craft': 'ISS', 'name': 'Tracy Caldwell Dyson'}, {'craft': 'ISS', 'name': 'Matthew Dominick'}, {'craft': 'ISS', 'name': 'Michael Barratt'}, {'craft': 'ISS', 'name': 'Jeanette Epps'}, {'craft': 'ISS', 'name': 'Alexander Grebenkin'}, {'craft': 'ISS', 'name': 'Butch Wilmore'}, {'craft': 'ISS', 'name': 'Sunita Williams'}, {'craft': 'Tiangong', 'name': 'Li Guangsu'}, {'craft': 'Tiangong', 'name': 'Li

Cong'}, {'craft': 'Tiangong', 'name': 'Ye Guangfu'}], 'number': 12, 'message': 'success'}

Print the number of astronauts currently on ISS.

```python
print(data.get('number'))
```

```
12
```

Print the names of the astronauts currently on ISS.

```python
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

```python
#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.

```python
import requests  # Ensure the 'requests' library is imported

api_url = "https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBM-DA0321EN-
SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/jobs.json"

response = requests.get(api_url)
if response.ok:
    data = response.json()
    print(data[:20])
else:
    print(f"Failed to retrieve data. Status code:
{response.status_code}")
```

[{'Id': 0, 'Job Title': 'Digital Media Planner', 'Job Experience Required': '5 - 10 yrs', 'Key Skills': 'Media Planning| Digital Media', 'Role Category': 'Advertising', 'Location': 'Los Angeles', 'Functional Area': 'Marketing , Advertising , MR , PR , Media Planning', 'Industry': 'Advertising, PR, MR, Event Management', 'Role': 'Media Planning Executive/Manager'}, {'Id': 1, 'Job Title': 'Online Bidding Executive', 'Job Experience Required': '2 - 5 yrs', 'Key Skills': 'pre sales| closing| software knowledge| clients| requirements| negotiating| client| online bidding| good communication| technology', 'Role Category': 'Retail Sales', 'Location': 'New York', 'Functional Area': 'Sales , Retail , Business Development', 'Industry': 'IT-Software, Software Services', 'Role': 'Sales Executive/Officer'}, {'Id': 2, 'Job Title': 'Trainee Research/ Research Executive- Hi- Tech Operations', 'Job Experience Required': '0 - 1 yrs', 'Key Skills': 'Computer science| Fabrication| Quality check| Intellectual property| Electronics| Support services| Research| Management| Human resource management| Research Executive', 'Role Category': 'R&D', 'Location': 'San Francisco', 'Functional Area': 'Engineering Design , R&D', 'Industry': 'Recruitment, Staffing', 'Role': 'R&D Executive'}, {'Id': 3, 'Job Title': 'Technical Support', 'Job Experience Required': '0 - 5 yrs', 'Key Skills': 'Technical Support', 'Role Category': 'Admin/Maintenance/Security/Datawarehousing', 'Location': 'Washington DC', 'Functional Area': 'IT Software - Application Programming , Maintenance', 'Industry': 'IT-Software, Software Services', 'Role': 'Technical Support Engineer'}, {'Id': 4, 'Job Title': 'Software Test Engineer -hyderabad', 'Job Experience Required': '2 - 5 yrs', 'Key Skills': 'manual testing| test engineering| test cases| web testing| web technologies', 'Role Category': 'Programming & Design', 'Location': 'Boston', 'Functional Area': 'IT Software - QA & Testing', 'Industry': 'IT-Software, Software Services', 'Role': 'Testing Engineer'}, {'Id': 5, 'Job Title': 'Opening For Adobe Analytics Specialist', 'Job Experience Required': '5 - 7 yrs', 'Key Skills': 'adobe experience manager| digital| digital marketing| adobe analytics| data analytics| aem| adobe| sitecatalyst', 'Role Category': 'Programming & Design', 'Location': 'Seattle', 'Functional Area': 'IT Software - Application Programming , Maintenance', 'Industry': 'IT-Software, Software Services', 'Role': 'System Analyst'}, {'Id': 6, 'Job Title':

'Sales- Fresher-for Leading Property Consultant', 'Job Experience Required': '0 - 0 yrs', 'Key Skills': 'channel partners| real estate| negotiation| property| Residential Sales', 'Role Category': 'Retail Sales', 'Location': 'Seattle', 'Functional Area': 'Sales , Retail , Business Development', 'Industry': 'Real Estate, Property', 'Role': 'Sales Executive/Officer'}, {'Id': 7, 'Job Title': 'Opportunity For Azure Devops Architect For Hyderabad Location', 'Job Experience Required': '9 - 14 yrs', 'Key Skills': 'TFS| Azure| Git| VSTS| Docker| DynaTrace| Splunk| AWS', 'Role Category': 'Programming & Design', 'Location': 'Detroit', 'Functional Area': 'IT Software - Application Programming , Maintenance', 'Industry': 'IT-Software, Software Services', 'Role': 'Technical Architect'}, {'Id': 8, 'Job Title': 'BDE- New Jersey', 'Job Experience Required': '2 - 7 yrs', 'Key Skills': 'Bde', 'Role Category': 'Institutional Sales', 'Location': 'Austin', 'Functional Area': 'Sales , Retail , Business Development', 'Industry': 'IT-Software, Software Services', 'Role': 'Sales Executive/Officer'}, {'Id': 9, 'Job Title': 'Technical Support/ Product Support', 'Job Experience Required': '1 - 5 yrs', 'Key Skills': 'technical support| support services| application support| communication skills| voice based process| semi voice| semi voice process', 'Role Category': 'Voice', 'Location': 'Austin', 'Functional Area': 'ITES , BPO , KPO , LPO , Customer Service , Operations', 'Industry': 'IT-Software, Software Services', 'Role': 'Associate/Senior Associate -(Technical)'}, {'Id': 10, 'Job Title': 'Executive Assistant To Chairman', 'Job Experience Required': '5 - 10 yrs', 'Key Skills': 'secretary| executive assistant| ea', 'Role Category': 'Corporate Planning/Consulting/Strategy', 'Location': 'Los Angeles', 'Functional Area': 'Strategy , Management Consulting , Corporate Planning', 'Industry': 'Courier, Transportation, Freight , Warehousing', 'Role': 'Corporate Planning/Strategy Manager'}, {'Id': 11, 'Job Title': 'SEO Executive', 'Job Experience Required': '1 - 6 yrs', 'Key Skills': 'website| web analytics| xml| link building| google analytics| case studies| seo| social media| tracking| maintaining', 'Role Category': 'Admin/Maintenance/Security/Datawarehousing', 'Location': 'New York', 'Functional Area': 'IT Software - eCommerce , Internet Technologies', 'Industry': 'IT-Software, Software Services', 'Role': 'Webmaster'}, {'Id': 12, 'Job Title': 'Workflow Coordinator', 'Job Experience Required': '2 - 7 yrs', 'Key Skills': 'operations| workflow| tat| monitoring| mts| email', 'Role Category': 'Operations', 'Location': 'San Francisco', 'Functional Area': 'ITES , BPO , KPO , LPO , Customer Service , Operations', 'Industry': 'BPO, Call Centre, ITeS', 'Role': 'Process Flow Analyst'}, {'Id': 13, 'Job Title': 'Oracle IDAM', 'Job Experience Required': '3 - 7 Years', 'Key Skills': 'Oracle IDAM|OIM|OAM', 'Role Category': 'Programming & Design', 'Location': 'Washington DC', 'Functional Area': 'IT Software - Application Programming , Maintenance', 'Industry': 'IT-Software / Software Services', 'Role': 'Software Developer'}, {'Id': 14, 'Job Title': 'Looking Facebook /social Media Manager For our Digital Marketing Ageny', 'Job Experience Required': '2 - 4 yrs', 'Key Skills': 'digital marketing| seo| social media marketing| content marketing| email marketing| sem| google analytics| campaign management| facebook| media management', 'Role Category': 'Online/Digital Marketing', 'Location': 'Washington DC', 'Functional Area': 'Marketing , Advertising , MR , PR , Media Planning', 'Industry': 'Advertising, PR, MR, Event Management', 'Role': 'Social Media Marketing Manager'}, {'Id': 16, 'Job Title': 'Tooling & Sampling', 'Job Experience Required': '2

- 7 yrs', 'Key Skills': 'Help Desk| Customer Service| customer relationship| access controls| Asset Management| Vendor Management', 'Role Category': 'Engineering Design', 'Location': 'Washington DC', 'Functional Area': 'Engineering Design , R&D', 'Industry': 'Industrial Products, Heavy Machinery', 'Role': 'Senior Design Engineer'}, {'Id': 17, 'Job Title': 'Account Manager', 'Job Experience Required': '3 - 7 yrs', 'Key Skills': 'Report Generation| Policies| analytical| procedures| Analysts| Data Entry| analytical skills', 'Role Category': 'Accounts', 'Location': 'Washington DC', 'Functional Area': 'Accounts , Finance , Tax , Company Secretary , Audit', 'Industry': 'IT-Software, Software Services', 'Role': 'Accounts Manager'}, {'Id': 18, 'Job Title': 'Magento Developer', 'Job Experience Required': '2 - 7 yrs', 'Key Skills': 'Copyright| Email| jQuery| XML| Javascript| CakePHP| Customer service| Open source| Facebook| CSS3', 'Role Category': 'Programming & Design', 'Location': 'Washington DC', 'Functional Area': 'IT Software - eCommerce , Internet Technologies', 'Industry': 'Textiles, Garments, Accessories', 'Role': 'Software Developer'}, {'Id': 19, 'Job Title': 'Looking For Trained Freshers in MS Dynamics CRM- Hyderabad', 'Job Experience Required': '0 - 3 yrs', 'Key Skills': 'C#| MS Dynamics CRM| Microsoft Dynamics| ms dynamics', 'Role Category': 'Programming & Design', 'Location': 'Washington DC', 'Functional Area': 'IT Software - Application Programming , Maintenance', 'Industry': 'IT-Software, Software Services', 'Role': 'Software Developer'}, {'Id': 20, 'Job Title': 'Job Openings Kotak Life/ Max Life/ Aditya Birla Health for Sales Mgr', 'Job Experience Required': '3 - 7 yrs', 'Key Skills': 'sales manager| business development manager| relationship manager| brokers| insurance sales| channel sales| direct sale| direct marketing| sales| Marketing Executive| Territory Sales Manager', 'Role Category': 'Retail/Personal Banking', 'Location': 'Washington DC', 'Functional Area': 'Financial Services , Banking , Investments , Insurance', 'Industry': 'Insurance', 'Role': 'Sales Officer'}]

Calling the function for Python and checking if it works.

```python
get_number_of_jobs_T("Python")
```

```
    ---------------------------------------------------------------------
------

    NameError                                 Traceback (most recent call
last)

    Cell In[44], line 1
    ----> 1 get_number_of_jobs_T("Python")


    Cell In[9], line 5, in get_number_of_jobs_T(technology)
          2 def get_number_of_jobs_T(technology):
```

```
      3
      4     #your code goes here
----> 5     return technology,number_of_jobs


NameError: name 'number_of_jobs' is not defined
```

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

```python
#def get_number_of_jobs_L(location):

    #your code goes here
    #return location,number_of_jobs

def get_number_of_jobs_L(data, Location):
    # Count the number of jobs where the location matches
    number_of_jobs = sum(1 for job in data if Location.lower() in
job['Location'].lower())

    return Location, number_of_jobs

# Replace 'data' with your actual job data source before calling the
function
data = data = response.json()
[{'Location': 'Los Angeles'}, {'Location': 'Washington DC'}, ...]  #
Example job data
Location, num_jobs = get_number_of_jobs_L(data, "New York")
print(f"Location: {Location}, Number of Jobs: {num_jobs}")
```

```
Location: New York, Number of Jobs: 3226
```

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

```python
def get_number_of_jobs_L(data, Location):
    # Count the number of jobs where the location matches
    number_of_jobs = sum(1 for job in data if Location.lower() in
job['Location'].lower())

    return Location, number_of_jobs

# Replace 'data' with your actual job data source before calling the
function
```

```
data = data = response.json()
[{'Location': 'Los Angeles'}, {'Location': 'Washington DC'}, ...]  # Example job data
Location, num_jobs = get_number_of_jobs_L(data, "Los Angeles")
print(f"Location: {Location}, Number of Jobs: {num_jobs}")
```

```
    Location: Los Angeles, Number of Jobs: 640
```

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**.

<!--

```
from openpyxl import Workbook        # import Workbook class from module openpyxl
wb=Workbook()                        # create a workbook object
ws=wb.active                         # use the active worksheet
ws.append(['Country','Continent'])   # add a row with two columns 'Country' and 'Continent'
ws.append(['Eygpt','Africa'])        # add a row with two columns 'Egypt' and 'Africa'
ws.append(['India','Asia'])          # add another row
ws.append(['France','Europe'])       # add another row
wb.save("countries.xlsx")            # save the workbook into a file called countries.xlsx
```

-->

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

```python
import requests
import pandas as pd

# List of technologies to search for
technologies = ["Python", "JavaScript", "Java", "C++", "Ruby", "Go", "Swift", "Kotlin"]

# Collect job counts for each technology
```

```
job_counts = []
for tech in technologies:
    tech_name, num_jobs = get_number_of_jobs_L(data, tech)
    job_counts.append({"Technology": tech_name, "Number of Jobs":
num_jobs})

# Convert to a DataFrame
job_counts_df = pd.DataFrame(job_counts)

# Write the results to an Excel spreadsheet
output_file = "job_counts.xlsx"
job_counts_df.to_excel(output_file, index=False)

print(f"Job counts successfully written to {output_file}")
```

```
    Job counts successfully written to job_counts.xlsx
```

Import libraries required to create excel spreadsheet

```python
import requests
import pandas as pd
from openpyxl import Workbook  # Required for writing Excel files
```

Create a workbook and select the active worksheet

```python
# Create a new workbook and select the active worksheet
wb = Workbook()
ws = wb.active
ws.title = "Job Counts"

# Write the results to an Excel spreadsheet
output_file = "job_counts.xlsx"
job_counts_df.to_excel(output_file, index=False)

print(f"Job counts successfully written to {output_file}")
```

```
    Job counts successfully written to job_counts.xlsx
```

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

```python
# Function to count number of jobs for a given location
def get_number_of_jobs_L(data, Location):
    number_of_jobs = sum(1 for job in data if Location.lower() in
job['Location'].lower())
    return Location, number_of_jobs

# Collect job counts for each location
Location_counts = []
for loc in Location:
    loc_name, num_jobs = get_number_of_jobs_L(data, loc)
    Location_counts.append({"Location": loc_name, "Number of Jobs":
num_jobs})

# Convert to a DataFrame
Location_counts_df = pd.DataFrame(Location_counts)

# Create a new workbook and select the active worksheet
wb = Workbook()
ws = wb.active
ws.title = "Job Counts by Location"

# Write the results to an Excel spreadsheet
output_file = "job_counts_by_location.xlsx"
Location_counts_df.to_excel(output_file, index=False)

print(f"Job counts by location successfully written to {output_file}")
```

    Job counts by location successfully written to
job_counts_by_location.xlsx

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


```python
# Write the results to an Excel spreadsheet
output_file = "job-postings.xlsx"
Location_counts_df.to_excel(output_file, index=False)
print(f"Job postings successfully written to {output_file}")
```

    Job postings successfully written to 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

```python
import requests
import pandas as pd
from openpyxl import Workbook  # Required for writing Excel files

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

# API URL
api_url = "https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBM-DA0321EN-
SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/jobs.json"

# Make the API request
response = requests.get(api_url)
if response.ok:
    data = response.json()
else:
    print(f"Failed to retrieve data. Status code:
{response.status_code}")
    data = []  # Exit early if data retrieval fails

# Function to count number of jobs for a given technology
def get_number_of_jobs_T(data, technology):
    number_of_jobs = sum(1 for job in data if technology.lower() in
job['Job Title'].lower())
    return technology, number_of_jobs

# Collect job counts for each technology
tech_counts = []
for tech in technologies:
    tech_name, num_jobs = get_number_of_jobs_T(data, tech)
    tech_counts.append({"Technology": tech_name, "Number of Jobs":
num_jobs})
```

```
# Convert to a DataFrame
tech_counts_df = pd.DataFrame(tech_counts)

# Create a new workbook and select the active worksheet
wb = Workbook()
ws = wb.active
ws.title = "Job Counts by Technology"

# Write the results to an Excel spreadsheet
output_file = "job-postings.xlsx"
tech_counts_df.to_excel(output_file, index=False)

print(f"Job counts by technology successfully written to {output_file}")
```

```
    Job counts by technology successfully written to job-postings.xlsx
```

Author

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