Scenario

For this project, you will assume the role of data engineer working for an international financial analysis company. Your company tracks stock prices, commodities, forex rates, inflation rates. Your job is to extract financial data from various sources like websites, APIs and files provided by various financial analysis firms. After you collect the data, you extract the data of interest to your company and transform it based on the requirements given to you. Once the transformation is complete you load that data into a database.

Project Tasks

In this project you will:

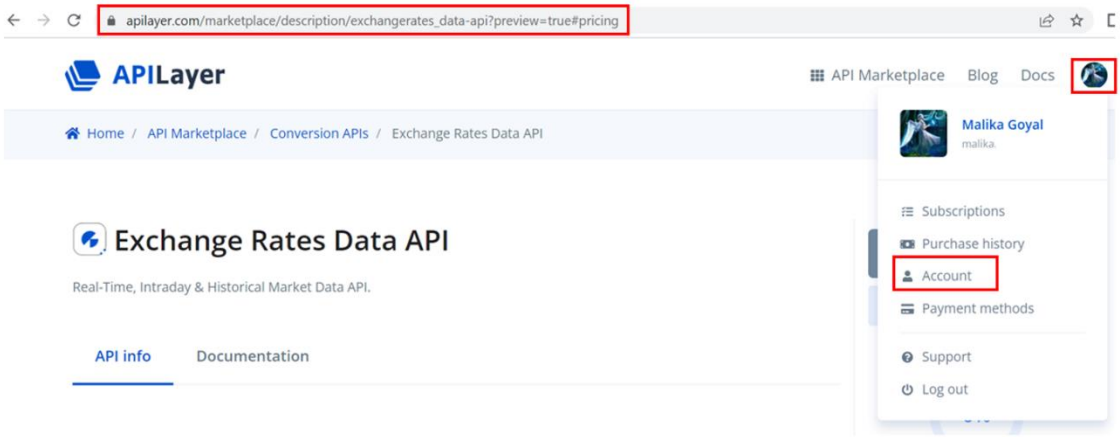
- Collect data using web scraping.
- Collect data using APIs
- Download files to process.
- Read csv, xml and json file types.
- Extract data from the above file types.
- Transform data.
- Save the transformed data in a ready-to-load format which data engineers can use to load the data.

Before script, please install Python & below libraries in your system

- BeautifulSoup4
- html5lib
- requests
- pandas

Using ExchangeRate-API we will extract currency exchange rate data. Use the below steps to get the access key and to get the data.

- Open the url : https://exchangeratesapi.io/ and click on Get Free API Key.
- Subscribe for Free plan and Sign-in with the Google Account.
- Once the account is created you will be redirected to https://apilayer.com website.
- Now, click on the user icon and click Account as shown below:



- Scroll down and you will get the API Key section. Copy the API key & save it in text file.
- Goto exchangerates_data – Click on Free Sunscibe

Assignment 1

Job 1 - Extract data using Web Scrapping

The Wikipedia webpage https://en.wikipedia.org/wiki/List_of_largest_banks provides information about largest banks in the world by various parameters. Scrape the data from the table **'By market capitalization**' and store it in a JSON file.

Gather the contents of the above webpage in text format using the requests library and assign it to the variable **data**.

Question 1 : After performing the above task, Print out the output of the following line. What will be the output

data[101:124]

Question 2 : Using the contents and beautiful soup load the data from the By market capitalization table into a pandas dataframe. The dataframe should have the country Name and Market Cap (US$ Billion) as column names.

Sample Code:

```
data = pd.DataFrame(columns=["", ""])
for row in soup.find_all('tbody')[4].find_all('tr'):
    col = row.find_all('td')
    if len(col) > 0:
            #Process – Add your code Here
```

Question 3: Display the first five rows using head

Question 4: Convert Data frame into json file

Job 2 – Extract API Data

Question 1: Using the requests library call the endpoint given code below and save the data to variable Called **Data**

url = "https://api.apilayer.com/exchangerates_data/latest?base=EUR&apikey= *********"

#Make sure to change ******* to your API key.

data = json.loads(requests.get(url).text)

Get value of INR Value

Question 2: Using the data gathered turn it into a pandas dataframe. The dataframe should have the Currency as the index and Rate as their columns. Get Total Rows Count

Question 3: Using the dataframe save it as a CSV name exchange_rates_1.csv

Assignment 2

Job 1 – Extract Data from Local Files

Use the attached files from below link.

Files

1. bank_market_cap_1.json

2. bank_market_cap_2.json

3. exchange_rates.csv

ETL Process:

Question 1: Create function to Extract information from json file

Question 2 : Define the extract function that finds JSON file `bank_market_cap_1.json` and calls the function created above to extract data from them. Store the data in a `pandas` dataframe. Use the following list for the columns.

columns=['Name','Market Cap (US$ Billion)']

```
def extract():
    # Write your code here
```

Question 3: Load the file **exchange_rates.csv** as a dataframe and find the exchange rate for British pounds with the symbol GBP, store it in the variable exchange_rate.

Transform:

Question 4: Using exchange_rate find the exchange rate of USD to GBP. Write a transform function that

1. Changes the `Market Cap (US$ Billion)` column from USD to GBP
2. Rounds the Market Cap (US$ Billion)\` column to 3 decimal places
3. Rename `Market Cap (US$ Billion)` to `Market Cap (GBP$ Billion)`

Load

1. Create a function that takes a above dataframe and load it to a csv named `bank_market_cap_gbp.csv`.

Once you are done with the assignment, please upload the results in below link.

https://forms.office.com/r/jqM60h8MXm