Python Programming (Basic-Intermediate)

Module 4 - IO

File operation

```python
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```python
f = open('/content/drive/MyDrive/AIS_DG/gpl-3.0.txt')
content = f.read()
f.close()
print(content)
```

```
                    GNU GENERAL PUBLIC LICENSE
                       Version 3, 29 June 2007

 Copyright (C) 2007 Free Software Foundation, Inc. <https://fsf.org/>
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

                            Preamble

  The GNU General Public License is a free, copyleft license for
software and other kinds of works.

  The licenses for most software and other practical works are design
to take away your freedom to share and change the works.  By contrast
the GNU General Public License is intended to guarantee your freedom
share and change all versions of a program--to make sure it remains f
software for all its users.  We, the Free Software Foundation, use th
GNU General Public License for most of our software; it applies also
any other work released this way by its authors.  You can apply it to
```

```python
f.closed
```

```
True
```

```
f.mode
```

```
'r'
```

```
f.name
```

```
'/content/drive/MyDrive/AIS_DG/gpl-3.0.txt'
```

```python
f1 = open('/content/drive/MyDrive/AIS_DG/superstore.data', mode='rb')
x = f1.read()
f1.close()
x
```

```python
with open('/content/drive/MyDrive/AIS_DG/gpl-3.0.txt') as f:
  content = f.read()
  print(content)
```

```
                GNU GENERAL PUBLIC LICENSE
                   Version 3, 29 June 2007

 Copyright (C) 2007 Free Software Foundation, Inc. <https://fsf.org/>
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

                        Preamble

  The GNU General Public License is a free, copyleft license for
software and other kinds of works.

  The licenses for most software and other practical works are desigr
to take away your freedom to share and change the works.  By contrast
the GNU General Public License is intended to guarantee your freedom
share and change all versions of a program--to make sure it remains f
software for all its users.  We, the Free Software Foundation, use th
GNU General Public License for most of our software; it applies also
any other work released this way by its authors.  You can apply it tc
```

Reading line by line

```python
filename = '/content/drive/MyDrive/AIS_DG/gpl-3.0.txt'

with open(filename) as file_object:
  ind = 1
  for line in file_object:
    print(ind,' ', line)
    ind += 1
```

```
1                    GNU GENERAL PUBLIC LICENSE

2                       Version 3, 29 June 2007

3

4    Copyright (C) 2007 Free Software Foundation, Inc. <https://fsf.c

5    Everyone is permitted to copy and distribute verbatim copies

6    of this license document, but changing it is not allowed.

7

8                            Preamble

9

10    The GNU General Public License is a free, copyleft license for
```

Reading a list of lines from a file

```python
filename = '/content/drive/MyDrive/AIS_DG/gpl-3.0.txt'

with open(filename) as file_object:
  contents = file_object.readlines()

print('Number of Lines: ', len(contents))
print(contents)
```

```
Number of Lines:  674
['                    GNU GENERAL PUBLIC LICENSE\n', '
```

Create and write a file

```python
with open('/content/drive/MyDrive/AIS_DG/programming.txt','w') as f:
  f.write('I love programming.')
```

```python
!cat /content/drive/MyDrive/AIS_DG/programming.txt
```

I love programming.

File position

```python
pos = 0
with open('/content/drive/MyDrive/AIS_DG/gpl-3.0.txt') as f:
  print(f.readline())
  pos = f.tell()
  print(pos, ' ', f.readline())
  pos = f.tell()
  print(pos, ' ', f.readline())
  pos = f.tell()
  print(pos, ' ', f.readline())
  print('=== Seek to 47 from the beginning ===')
  f.seek(47, 0)
  pos = f.tell()
  print(pos, ' ', f.readline())
```

                    GNU GENERAL PUBLIC LICENSE

47   Version 3, 29 June 2007

94

95   Copyright (C) 2007 Free Software Foundation, Inc. <https://fsf.o

=== Seek to 47 from the beginning ===
47   Version 3, 29 June 2007

os package

```python
import os
os.listdir('/content/drive/MyDrive/AIS_DG/')
```

```
['Gov_Expenditure_EDU.xls',
 'btc.json',
 'movies.csv',
 'p1.txt',
 'retail.csv',
 'gpl-3.0.txt',
 'Flight_airlines.csv',
 'Telco-Churn.csv',
 'superstore.data',
 'Superstore.xlsx',
 'coding.gif',
 'bank-data.csv',
 'Flight_flights.csv',
 'lib',
 'cc.json',
 'cc1.json',
 'programming.txt']
```

```
os.rename('/content/drive/MyDrive/AIS_DG/programming.txt',
          '/content/drive/MyDrive/AIS_DG/p1.txt')
os.listdir('/content/drive/MyDrive/AIS_DG/')
```

```
['Gov_Expenditure_EDU.xls',
 'btc.json',
 'movies.csv',
 'p1.txt',
 'retail.csv',
 'gpl-3.0.txt',
 'Flight_airlines.csv',
 'Telco-Churn.csv',
 'superstore.data',
 'Superstore.xlsx',
 'coding.gif',
 'bank-data.csv',
 'Flight_flights.csv',
 'lib',
 'cc.json',
 'cc1.json']
```

```
import glob
glob.glob('/content/drive/MyDrive/AIS_DG/*.csv')
```

```
['/content/drive/MyDrive/AIS_DG/movies.csv',
 '/content/drive/MyDrive/AIS_DG/retail.csv',
 '/content/drive/MyDrive/AIS_DG/Flight_airlines.csv',
 '/content/drive/MyDrive/AIS_DG/Telco-Churn.csv',
 '/content/drive/MyDrive/AIS_DG/bank-data.csv',
 '/content/drive/MyDrive/AIS_DG/Flight_flights.csv']
```

```
os.environ
```

```
environ{'SHELL': '/bin/bash',
        'NV_LIBCUBLAS_VERSION': '12.2.5.6-1',
        'NVIDIA_VISIBLE_DEVICES': 'all',
        'COLAB_JUPYTER_TRANSPORT': 'ipc',
        'NV_NVML_DEV_VERSION': '12.2.140-1',
        'NV_CUDNN_PACKAGE_NAME': 'libcudnn8',
        'CGROUP_MEMORY_EVENTS': '/sys/fs/cgroup/memory.events /var/cc
        'NV_LIBNCCL_DEV_PACKAGE': 'libnccl-dev=2.19.3-1+cuda12.2',
        'NV_LIBNCCL_DEV_PACKAGE_VERSION': '2.19.3-1',
        'VM_GCE_METADATA_HOST': '169.254.169.253',
        'HOSTNAME': 'd75fff63881f',
        'LANGUAGE': 'en_US',
        'TBE_RUNTIME_ADDR': '172.28.0.1:8011',
        'GCE_METADATA_TIMEOUT': '3',
        'NVIDIA_REQUIRE_CUDA': 'cuda>=12.2 brand=tesla,driver>=470,dr
        'NV_LIBCUBLAS_DEV_PACKAGE': 'libcublas-dev-12-2=12.2.5.6-1',
        'NV_NVTX_VERSION': '12.2.140-1',
        'COLAB_JUPYTER_IP': '172.28.0.12',
        'NV_CUDA_CUDART_DEV_VERSION': '12.2.140-1',
        'NV_LIBCUSPARSE_VERSION': '12.1.2.141-1',
```

Reading CSV (Pandas DataFrame)

```python
import pandas as pd
df = pd.read_csv('/content/drive/MyDrive/AIS_DG/Telco-Churn.csv')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
```

```
 14  StreamingMovies   7043 non-null    object
```

```python
df_chunks = pd.read_csv('/content/drive/MyDrive/AIS_DG/Telco-Churn.csv',
                        iterator=True, chunksize=700)
```

```python
for d in df_chunks:
    print(max(d.index))
```

```
699
1399
2099
2799
3499
4199
4899
5599
6299
6999
7042
```

```python
x1 = df_chunks
print(x1.shape)
x1.head()
```

```python
x2 = df_chunks.get_chunk()
print(x2.shape)
x2.head()
```

```python
df_chunks.
```

Reading an Excel file

```python
superstore = pd.read_excel('/content/drive/MyDrive/AIS_DG/Superstore.xls
superstore.head()
```

```
   Row ID       Order ID Order Date  Ship Date     Ship Mode Custo
0       1  CA-2013-152156 2014-11-09 2014-11-12  Second Class    CG
1       2  CA-2013-152156 2014-11-09 2014-11-12  Second Class    CG
2       3  CA-2013-138688 2014-06-13 2014-06-17  Second Class    DV
```

```
3        4  US-2012-108966 2013-10-11 2013-10-18   Standard Class     SC
4        5  US-2012-108966 2013-10-11 2013-10-18   Standard Class     SC

     Customer Name      Segment          Country            City  ...  \
0     Claire Gute       Consumer  United States         Henderson  ...
1     Claire Gute       Consumer  United States         Henderson  ...
2  Darrin Van Huff     Corporate  United States       Los Angeles  ...
3   Sean O'Donnell      Consumer  United States   Fort Lauderdale  ...
4   Sean O'Donnell      Consumer  United States   Fort Lauderdale  ...

   Postal Code    Region     Product ID          Category Sub-Category
0        42420     South  FUR-BO-10001798        Furniture    Bookcases
1        42420     South  FUR-CH-10000454        Furniture       Chairs
2        90036      West  OFF-LA-10000240  Office Supplies       Labels
3        33311     South  FUR-TA-10000577        Furniture       Tables
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Post Cod
0	1	CA-2013-152156	2014-11-09	2014-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	424
1	2	CA-2013-152156	2014-11-09	2014-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	424
2	3	CA-2013-138688	2014-06-13	2014-06-17	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	900
3	4	US-2012-108966	2013-10-11	2013-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	333
4	5	US-2012-108966	2013-10-11	2013-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	333

5 rows × 21 columns

```
superstore_file = pd.ExcelFile('/content/drive/MyDrive/AIS_DG/Superstore
```

```
sn = superstore_file.sheet_names
```

```
d_excel = pd.read_excel(superstore_file, sheet_name=sn)
```

```
d_excel['People']
```

```
            Person      Region
0      Anna Andreadi       West
1       Chuck Magee        East
2     Kelly Williams    Central
3  Cassandra Brandow      South
```

	Person	Region
0	Anna Andreadi	West
1	Chuck Magee	East
2	Kelly Williams	Central
3	Cassandra Brandow	South

Reading JSON

```
import json

f = open('/content/drive/MyDrive/AIS_DG/btc.json')

content = json.load(f)

f.close()
print(content)
```

```
{'id': 'bitcoin', 'symbol': 'btc', 'name': 'Bitcoin', 'current_price':
```

```
d = json.load(open('/content/drive/MyDrive/AIS_DG/cc.json'))
d[:2]
```

```
[{'id': 'bitcoin',
  'symbol': 'btc',
  'name': 'Bitcoin',
  'image': 'https://assets.coingecko.com/coins/images/1/large/bitcoin
  'current_price': 42365,
  'market_cap': 830056440968,
  'market_cap_rank': 1,
  'fully_diluted_valuation': 888834064145,
```

```
    'total_volume': 13680664375,
    'high_24h': 42743,
    'low_24h': 41750,
    'price_change_24h': 580.57,
    'price_change_percentage_24h': 1.38946,
    'market_cap_change_24h': 10178363134,
    'market_cap_change_percentage_24h': 1.24145,
    'circulating_supply': 19611293.0,
    'total_supply': 21000000.0,
    'max_supply': 21000000.0,
    'ath': 69045
```

```python
pd.read_json('/content/drive/MyDrive/AIS_DG/cc.json')
```

```
             id   symbol            name  \
0       bitcoin      btc         Bitcoin
1      ethereum      eth        Ethereum
2        tether     usdt          Tether
3   binancecoin      bnb             BNB
4        solana      sol          Solana
..          ...      ...             ...
95         bonk     bonk            Bonk
96         usdd     usdd            USDD
97  oasis-network    rose   Oasis Network
98    klay-token     klay          Klaytn
99    frax-ether   frxeth      Frax Ether

                                              image   current_price
0   https://assets.coingecko.com/coins/images/1/la...   42365.000000
1   https://assets.coingecko.com/coins/images/279/...    2272.860000
2   https://assets.coingecko.com/coins/images/325/...       1.000000
3   https://assets.coingecko.com/coins/images/825/...     306.920000
4   https://assets.coingecko.com/coins/images/4128...      97.380000
```

	id	symbol	name	image	current_price	market
0	bitcoin	btc	Bitcoin	https://assets.coingecko.com/coins/images/1/la...	42365.000000	830056
1	ethereum	eth	Ethereum	https://assets.coingecko.com/coins/images/279/...	2272.860000	273004
2	tether	usdt	Tether	https://assets.coingecko.com/coins/images/325/...	1.000000	960106
3	binancecoin	bnb	BNB	https://assets.coingecko.com/coins/images/825/...	306.920000	471802

Read/write pickles

```python
import glob

superstore.to_pickle('/content/drive/MyDrive/AIS_DG/superstore.data')
print(glob.glob('/content/drive/MyDrive/AIS_DG/*.data'))
```

```
['/content/drive/MyDrive/AIS_DG/superstore.data']
```

```python
df1 = pd.read_pickle('/content/drive/MyDrive/AIS_DG/superstore.data')
df1.head()
```

```
   Row ID      Order ID Order Date  Ship Date      Ship Mode Custo
0       1  CA-2013-152156 2014-11-09 2014-11-12   Second Class    CG
1       2  CA-2013-152156 2014-11-09 2014-11-12   Second Class    CG
2       3  CA-2013-138688 2014-06-13 2014-06-17   Second Class    DV
3       4  US-2012-108966 2013-10-11 2013-10-18  Standard Class    SC
4       5  US-2012-108966 2013-10-11 2013-10-18  Standard Class    SC

      Customer Name    Segment        Country            City  ...  \
0       Claire Gute   Consumer  United States        Henderson  ...
1       Claire Gute   Consumer  United States        Henderson  ...
2   Darrin Van Huff  Corporate  United States      Los Angeles  ...
3     Sean O'Donnell   Consumer  United States  Fort Lauderdale  ...
4     Sean O'Donnell   Consumer  United States  Fort Lauderdale  ...

   Postal Code  Region     Product ID         Category Sub-Category
0        42420   South  FUR-BO-10001798        Furniture    Bookcases
1        42420   South  FUR-CH-10000454        Furniture       Chairs
2        90036    West  OFF-LA-10000240  Office Supplies       Labels
3        33311   South  FUR-TA-10000577        Furniture       Tables
4        33311   South  OFF-ST-10000760  Office Supplies      Storage
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Post Cod
0	1	CA-2013-152156	2014-11-09	2014-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	424
1	2	CA-2013-152156	2014-11-09	2014-11-12	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	424
2	3	CA-2013-138688	2014-06-13	2014-06-17	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	900
3	4	US-2012-108966	2013-10-11	2013-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	333
4	5	US-2012-108966	2013-10-11	2013-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	333

```python
import pickle
```

```python
df1.to_pickle('test.data')
```

Read file from URL

```python
weather = pd.read_csv('http://fastdata.in.th/AIS/weather_daily_darksky.c
weather.head()
```

```
name 'pd' is not defined
```

API call

```python
import requests

resp = requests.get('https://api.coingecko.com/api/v3/coins/markets/?vs_
print(resp.status_code)
crypto_data = resp.json()
print(crypto_data)
```

```
200
[{'id': 'bitcoin', 'symbol': 'btc', 'name': 'Bitcoin', 'image': 'https
```

Activity

Add a function to myutils.py

Function Name: load_current_weather

Description: extract data from TMD Weather API and append it to file specified in the argument. If not exist, create a new file.

Import and test the function.

TMD Weather URL: https://data.tmd.go.th

```python
# work here
```

```python
import requests

url = "https://data.tmd.go.th/api/Weather3Hours/V2/?uid=api&ukey=api1234

response = requests.request("GET", url)

data = response.text
```

```python
data
```

```
'<?xml version="1.0"?>\n<Weather3Hours Version="2.0"><Header><Title>We
```

Where is API DATA?

```python
from bs4 import BeautifulSoup
```

```python
soup = BeautifulSoup(data, 'xml')
```

```python
list_stations = soup.find_all('Stations')
```

```python
len(list_stations)
```

1

```python
list_stations[0]
```

<Stations/>

There's no Data Sensei

```python
list_stations[0].find('StationNameThai').text
```

```python
dt1 = [s.find('DateTime').text for s in list_stations]
sn1 = [s.find('StationNameThai').text for s in list_stations]
at1 = [s.find('AirTemperature').text for s in list_stations]
```

```python
d = pd.DataFrame({'Date':dt1,'StationName': sn1, 'AirTemperature': at1})
d.head()
```

```
Empty DataFrame
Columns: [Date, StationName, AirTemperature]
Index: []
```

Date	StationName	AirTemperature

```
arg = '/content/drive/MyDrive/AIS_DG/result.csv'
d.to_csv(arg,mode = 'a',header=False)
```

```
pd.read_csv(arg).shape
```