

# Module 9: Input and Output

## 1. Read datasets from links

link : <https://data.cityofnewyork.us/City-Government/NYC-Civil-Service-Titles/nzjr-3966>  
(<https://data.cityofnewyork.us/City-Government/NYC-Civil-Service-Titles/nzjr-3966>)

In [1]:

```
import pandas as pd
```

In [2]:

```
URL = r"https://data.cityofnewyork.us/api/views/nzjr-3966/files/4b888c87-1952-4ae6-b820-df0"
civil = pd.read_excel(URL)
civil.head()
```

Out[2]:

Note:  
Hover over  
each  
column  
header for  
an  
explanation  
of each  
field.

	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6		
0	Order	Field Name	Longform Name	Description	Geocoded	Required	Data Type	Expe
1	1	Title Code	NaN	Civil Service Title Code	False	True	Plain Text	
2	2	Title Description	NaN	Name/description of the title	False	True	Plain Text	
3	3	Standard Hours	NaN	Standard weekly hours for the title	False	True	Number	
4	4	Assignment Level	NaN	The assignment level within the title	False	True	Number	

In [3]:

```
URL =r"https://data.cityofnewyork.us/api/views/nzjr-3966/rows.csv?accessType=DOWNLOAD"
civil_service = pd.read_csv(URL)
civil_service.head()
```

Out[3]:

	Title Code	Title Description	Standard Hours	Assignment Level	Union Code	Union Description	Bargaining Unit Short Name	Bargain Descrip
0	00031	HCPPA	37.0	00	23	PRINCIPAL ADMINISTRATIVE ASSOC	CWA	Communica Wrk Ame
1	00136	BEAUTICIAN	37.5	00	124	L420,L2507,L3627 DC37 HOSPITAL	DC37	District Coi
2	00486	HOME HEALTH AIDE	37.5	00	127	INSTITUTIONAL SERVICE TITLES D	DC37	District Coi
3	03647	SPECIAL ASSISTANT TO THE BOROU	35.0	00	775	MGRL TEMP - PROP NC	N/U	Non-U
4	03927	ASSISTANT SYSTEMS ANALYST (HHC	35.0	00	129	ACCOUNTING AND EDP TITLES DC37	DC37	District Coi



2. Quick Object conversion

Pandas also has the **list()** method to convert the Pandas Series into Python List.

In [4]:

```
civil_service["Title Description"].tolist()
```

Out[4]:

```
['HCPPA',  
'BEAUTICIAN',  
'HOME HEALTH AIDE',  
'SPECIAL ASSISTANT TO THE BOROU',  
'ASSISTANT SYSTEMS ANALYST (HHC',  
'ASSISTANT TO THE PRESIDENT (QN',  
'DIRECTOR OF INTERGOVERNMENTAL',  
'FOOD SERVICE MANAGER',  
'ASSISTANT TO THE PRESIDENT (ST',  
'STUDENT LEGAL SPECIALIST',  
'EXECUTIVE DIRECTOR OF ADMINIST',  
'DIRECTOR (PLANT OPERATIONS SER',  
'ASSISTANT TO THE PRESIDENT (BR',  
'SECRETARY TO ASSISTANT TO PRES',  
'SECRETARY TO THE EXCUTIVE ASSI',  
'ASSOCIATE DIRECTOR-OPERATIONAL',  
'RESEARCH AND LIAISON COORDINAT',  
'RESEARCH I TATSON ADN GOVERNMEN'.
```

Pandas Series to Pandas DataFrame

- We can convert Pandas Series to DataFrame by using **to\_frame()** method.

In [5]:

```
civil_service["Title Description"].to_frame().head()
```

Out[5]:

	Title Description
0	HCPPA
1	BEAUTICIAN
2	HOME HEALTH AIDE
3	SPECIAL ASSISTANT TO THE BOROU
4	ASSISTANT SYSTEMS ANALYST (HHC

Combine every Title Description into one string and separate by comma.

In [6]:

```
", ".join(civil_service["Title Description"])
```

Out[6]:

```
"HCPA, BEAUTICIAN, HOME HEALTH AIDE, SPECIAL ASSISTANT TO THE BOROU, ASSI
STANT SYSTEMS ANALYST (HHC, ASSISTANT TO THE PRESIDENT (QN, DIRECTOR OF IN
TERGOVERNMENTAL, FOOD SERVICE MANAGER, ASSISTANT TO THE PRESIDENT (ST, STU
DENT LEGAL SPECIALIST, EXECUTIVE DIRECTOR OF ADMINIST, DIRECTOR (PLANT OPE
RATIONS SER, ASSISTANT TO THE PRESIDENT (BR, SECRETARY TO ASSISTANT TO PRE
S, SECRETARY TO THE EXECUTIVE ASSI, ASSOCIATE DIRECTOR-OPERATIONAL, RESEAR
H AND LIAISON COORDINAT, RESEARCH LIAISON ADN GOVERNMENT, DIRECTOR BOROUGH
PRESIDENT'S O, ASSISTANT TO THE PRESIDENT (BOR, ASSISTANT TO THE PRESIDENT
(BOR, FISCAL AND POLICY ANALYST (BP, CHAUFFEUR-ATTENDANT (BKLYN BOR, DEPUT
Y OPERATION SUPPORT MANAG, RESEARCH AND LIAISON SPECIALIS, CHAUFFEUR-ATTEN
DANT (BORO PRES, CHAUFFEUR-ATTENDANT (BORO PRES, RESEARCH PROJECTS COORDIN
ATOR, ASSISTANT TO THE DEPUTY MAYOR, RESEARCH PROJECTS COOR(MA)-MGR, SPECI
AL ASST TO THE DIRECTOR O, SENIOR RACKETS INVESTIGATOR (B, SUPERVISING RAC
KETS INVESTIGAT, SPECIAL ASSISTANT TO DA (QUEEN, PRINCIPAL ACCOUNTANT INVE
STIGA, EMPLOYEE HEALTH BENEFITS PROGR, CHIEF ADMINISTRATOR OF IMPARTI, DIR
ECTOR, DATA PROCESSING COOR, DIRECTOR OF BUILDING MANAGEMEN, STATISTICAL S
ECRETARY (OMB), STATISTICAL SECRETARY (OMB), DEPUTY COMMISSIONER (TAXI AN
D. GENFERAI COUNSEL (TAXI & LIMOUS. LEGISLATIVE ATDF (OFFICE OF TH. SECRETARY
```

In [7]:

```
", ".join(str(name) for name in civil_service["Title Description"])
```

Out[7]:

```
"HCPA, BEAUTICIAN, HOME HEALTH AIDE, SPECIAL ASSISTANT TO THE BOROU, ASSI
T SYSTEMS ANALYST (HHC, ASSISTANT TO THE PRESIDENT (QN, DIRECTOR OF INTERGOV
ERNMENTAL, FOOD SERVICE MANAGER, ASSISTANT TO THE PRESIDENT (ST, STUDENT LEGA
L SPECIALIST, EXECUTIVE DIRECTOR OF ADMINIST, DIRECTOR (PLANT OPERATIONS SE
R, ASSISTANT TO THE PRESIDENT (BR, SECRETARY TO ASSISTANT TO PRES, SECRETARY
TO THE EXECUTIVE ASSI, ASSOCIATE DIRECTOR-OPERATIONAL, RESEARCH AND LIAISON C
OORDINAT, RESEARCH LIAISON ADN GOVERNMENT, DIRECTOR BOROUGH PRESIDENT'S O, ASS
ISTANT TO THE PRESIDENT (BOR, ASSISTANT TO THE PRESIDENT (BOR, FISCAL AND POLI
CY ANALYST (BP, CHAUFFEUR-ATTENDANT (BKLYN BOR, DEPUTY OPERATION SUPPORT MAN
AG, RESEARCH AND LIAISON SPECIALIS, CHAUFFEUR-ATTENDANT (BORO PRES, CHAUFFEUR
-ATTENDANT (BORO PRES, RESEARCH PROJECTS COORDINATOR, ASSISTANT TO THE DEPUT
Y MAYOR, RESEARCH PROJECTS COOR(MA)-MGR, SPECIAL ASST TO THE DIRECTOR O, SENI
OR RACKETS INVESTIGATOR (B, SUPERVISING RACKETS INVESTIGAT, SPECIAL ASSISTAN
T TO DA (QUEEN, PRINCIPAL ACCOUNTANT INVESTIGA, EMPLOYEE HEALTH BENEFITS PRO
GR, CHIEF ADMINISTRATOR OF IMPARTI, DIRECTOR, DATA PROCESSING COOR, DIRECTOR
OF BUILDING MANAGEMEN, STATISTICAL SECRETARY (OMB), STATISTICAL SECRETARY (O
MB), DEPUTY COMMISSIONER (TAXI AND, GENERAL COUNSEL (TAXI & LIMOUS, LEGISLATI
VE ATDF (OFFICE OF TH. SECRETARY (OFFICE OF THE MAYOR. DEPUTY DIRECTOR FOR M
```

### 3. Export DataFrame into CSV file

We use `pd.to_csv()` method to export DataFrame. In this method, there are a few parameters that will help us a lot:

- **path** : The filename we have to save. Always remember to include **.csv** at the end of the string
- **columns** : Selected column names to export.
- **index** : By default it is set to False. If True, the index will be removed in the CSV file.
- **encoding** : "utf-8" is the most popular character encoding. If you encounter error when encoding, try to set this value.

In [8]:

```
civil_service.head(3)
```



Out[8]:

	Title Code	Title Description	Standard Hours	Assignment Level	Union Code	Union Description	Bargaining Unit Short Name	Bargain Descrip
0	00031	HCPPA	37.0	00	23	PRINCIPAL ADMINISTRATIVE ASSOC	CWA	Communica Wrk Ame
1	00136	BEAUTICIAN	37.5	00	124	L420,L2507,L3627 DC37 HOSPITAL	DC37	District Coi
2	00486	HOME HEALTH AIDE	37.5	00	127	INSTITUTIONAL SERVICE TITLES D	DC37	District Coi

In [9]:

```
civil_service.to_csv("Civil Service NYC.csv",index=False, columns=["Title Description","Sta
```

There will be no output when running this code. However, we can see a file is generated.

 all\_stocks\_byr.csv  
 Civil Service NYC.csv

## 4. Install XLRD Library

This library is used to read Excel files using Pandas. It should already be included in your Anaconda packages

```
(base) C:\Users\SAM>conda install xlrd
Collecting package metadata: done
Solving environment: failed
```

## 5. Read Excel Files into Pandas

If the Excel file contains only one spreadsheet, Pandas will create the first spreadsheet by default.

When we have multiple spreadsheets, we need to specify which sheet we want to read.

In [10]:

```
pd.read_excel("data/Data - Single Worksheet.xlsx")
```

Out[10]:

	First Name	Last Name	City	Gender
0	Brandon	James	Miami	M
1	Sean	Hawkins	Denver	M
2	Judy	Day	Los Angeles	F
3	Ashley	Ruiz	San Francisco	F
4	Stephanie	Gomez	Portland	F

## Multiple spreadsheets

In [11]:

```
pd.read_excel("data/Data - Multiple Worksheets.xlsx", sheet_name=0)
```

Out[11]:

	First Name	Last Name	City	Gender
0	Brandon	James	Miami	M
1	Sean	Hawkins	Denver	M
2	Judy	Day	Los Angeles	F
3	Ashley	Ruiz	San Francisco	F
4	Stephanie	Gomez	Portland	F

In [12]:

```
pd.read_excel("data/Data - Multiple Worksheets.xlsx", sheet_name=1)
```

Out[12]:

	First Name	Last Name	City	Gender
0	Parker	Power	Raleigh	F
1	Preston	Prescott	Philadelphia	F
2	Ronaldo	Donaldo	Bangor	M
3	Megan	Stiller	San Francisco	M
4	Bustin	Jieber	Austin	F

We can also choose the sheets based on the sheet's name. For example, the second sheet is "Data 2"

In [13]:

```
pd.read_excel("data/Data - Multiple Worksheets.xlsx", sheet_name="Data 2")
```

Out[13]:

	First Name	Last Name	City	Gender
0	Parker	Power	Raleigh	F
1	Preston	Prescott	Philadelphia	F
2	Ronaldo	Donaldo	Bangor	M
3	Megan	Stiller	San Francisco	M
4	Bustin	Jieber	Austin	F

## Importing Relational Databases (MySQL)

It is quite common to store our data in external relational databases. Python provides a quick and easy way to import data from databases using external packages sqlalchemy. If you wish to import PostgreSQL or MySQL, you would need to install *psycopg2* or *pymysql* module.

The sqlalchemy and pymysql modules have already been installed for you in your computer device.

Import create\_engine module from sqlalchemy to connect to the external database server and then, use the following credentials:

- Username: tempuser
- Password: temp201&
- Server: research.fsktm.um.edu.my
- Port: 3306
- Database: eis2016

Select all columns and rows from table query\_introtor.

In [14]:

```
from sqlalchemy import create_engine
# Create the connection engine
engine = create_engine('mysql+pymysql://tempuser:temp201&@research.fsktm.um.edu.my:3306/eis2016')

# Connect the engine to the database server
con = engine.connect()

# Run SQL command
rs = con.execute('SELECT * FROM query_introtor')
```

In [15]:

```
import pymysql
import pandas as pd
```

In [16]:

```
# Store the data in a data frame
df = pd.DataFrame(rs.fetchall())

#Assign the column names
df.columns = rs.keys()

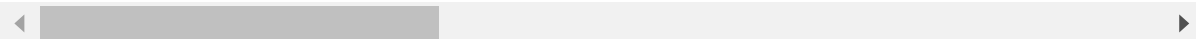
#Close connection
con.close()

df.head()
```

Out[16]:

	applid	fullname	callname	initial1	post_code	dept_code	aliasid	alias	status
0	1	Abdullah Gani	Prof. Dr.	ABG	P	STK	2	Gani, Abdullah	
1	230	Nazean Jomhari	Dr.	NJ	SL	SE	11	Jomhari, N.	
2	549	Shahaboddin Shamshirband	Dr.	None	VSL	STK	15	Shamshirband, Shahaboddin	
3	65	Miss Laiha Mat Kiah	Prof. Dr.	ML	P	STK	16	Kiah, Miss Laiha Mat	
4	221	Nor Badrul Anuar Jumaat	Assoc. Prof. Dr.	NBA	SL	STK	17	Anuar, Nor Badrul	

5 rows × 41 columns



## 6. Export Excel file

- Create Excel Writer object
- Add the DataFrame into the ExcelWriter object
- Save the Excel Writer object into .xlsx file



In [17]:

```
civil_service.head()
```

Out[17]:

	Title Code	Title Description	Standard Hours	Assignment Level	Union Code	Union Description	Bargaining Unit Short Name	Bargain Descrip
0	00031	HCPPA	37.0	00	23	PRINCIPAL ADMINISTRATIVE ASSOC	CWA	Communica Wrk Ame
1	00136	BEAUTICIAN	37.5	00	124	L420,L2507,L3627 DC37 HOSPITAL	DC37	District Coi
2	00486	HOME HEALTH AIDE	37.5	00	127	INSTITUTIONAL SERVICE TITLES D	DC37	District Coi
3	03647	SPECIAL ASSISTANT TO THE BOROU	35.0	00	775	MGRL TEMP - PROP NC	N/U	Non-U
4	03927	ASSISTANT SYSTEMS ANALYST (HHC	35.0	00	129	ACCOUNTING AND EDP TITLES DC37	DC37	District Coi

In [18]:

```
title = civil_service[["Title Code", "Title Description"]]  
union = civil_service[["Union Code", "Union Description"]]
```

1. Create Excel Writer object

In [19]:

```
civilExcel = pd.ExcelWriter("Civil Service.xlsx")
```

2. Add the DataFrame to Excel Writer object

In [20]:

```
title.to_excel(civilExcel, sheet_name="Title")
```

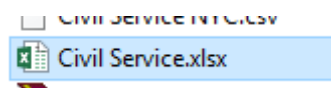
In [21]:

```
union.to_excel(civilExcel, sheet_name="Union")
```

3. Export the ExcelWriter object into .xlsx file

In [22]:

```
civilExcel.save()
```



32	30	0527A	RESEARCH PROJECTS COOR(MA)-MGR
33	31	05301	SPECIAL ASST TO THE DIRECTOR O
34	32	05306	PARK BOROUGH COMMISSIONER (PAR
35	33	05322	SENIOR RACKETS INVESTIGATOR (B