

Data Technician

Name:

Course Date:

Table of contents

Day 2: Task 1	3
Day 3: Task 1	4
Exercise 1: Loading and Exploring the Data.....	4
Exercise 2: Indexing and Slicing.....	5
Exercise 3: Data Manipulation.....	7
Exercise 4: Aggregation and Grouping.....	7
Exercise 5: Advanced Operations.....	8
Exercise 6: Exporting Data	9
Exercise 7: If finished early try visualising the results	10
Day 4: Task 1	10
Day 4: Task 2	11
Course Notes.....	13
Additional Information.....	14



Day 2: Task 1

It is a common software development interview question to create the below with a certain programming language. Create the below using Python syntax, test it and past the completed syntax and output below.

FizzBuzz:

Go through the integers from 1 to 100.

If a number is divisible by 3, print "fizz."

If a number is divisible by 5, print "buzz."

If a number is both divisible by 3 and by 5, print "fizzbuzz."

Otherwise, print just the number.

**Paste your completed
work to the right**

Code

```
# Iterate through numbers from 1 to 100 (inclusive)
for number in range(1, 101):

    # Check if the number is divisible by both 3 and 5 first
    # This condition must be checked before the
    individual divisibility by 3 or 5

    if number % 3 == 0 and number % 5 == 0:

        print("fizzbuzz")

    # If not divisible by both, check if divisible by 3
    elif number % 3 == 0:

        print("fizz")

    # If not divisible by 3, check if divisible by 5
    elif number % 5 == 0:

        print("buzz")

    # If not divisible by 3 or 5, print the number itself
    else:

        print(number)
```



Output

```
1 2 fizz 4 buzz fizz 7 8 fizz buzz 11 fizz 13 14 fizzbuzz 16 17
fizz 19 buzz fizz 22 23 fizz buzz 26 fizz 28 29 fizzbuzz 31 32
fizz 34 buzz fizz 37 38 fizz buzz 41 fizz 43 44 fizzbuzz 46 47
fizz 49 buzz fizz 52 53 fizz buzz 56 fizz 58 59 fizzbuzz 61 62
fizz 64 buzz fizz 67 68 fizz buzz 71 fizz 73 74 fizzbuzz 76 77
fizz 79 buzz fizz 82 83 fizz buzz 86 fizz 88 89 fizzbuzz 91 92
fizz 94 buzz fizz 97 98 fizz buzz
```

Day 3: Task 1

Download the 'student.csv', complete the below exercises as a group and paste your input and output. Although this is a group activity, everyone should have the below answered so it supports your portfolio:

Exercise 1: Loading and Exploring the Data

1. Question: "Write the code to read a CSV file into a Pandas DataFrame."
2. Question: "Write the code to display the first 5 rows of the DataFrame."
3. Question: "Write the code to get the information about the DataFrame."
4. Question: "Write the code to get summary statistics for the DataFrame."

1:

[2]:

```
import pandas as pd
```

[4]:

```
data = pd.read_excel("student.xlsx")
```

2:



```
[5]: data.head()
```

```
[5]:
```

	id	name	class	mark	gender
0	1	John Deo	Four	75	female
1	2	Max Ruin	Three	85	male
2	3	Arnold	Three	55	male
3	4	Krish Star	Four	60	female
4	5	John Mike	Four	60	female

3:

```
[8]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 35 entries, 0 to 34  
Data columns (total 5 columns):  
#   Column  Non-Null Count  Dtype  
---  ---  
0    id      35 non-null      int64  
1   name     34 non-null      object  
2   class    34 non-null      object  
3   mark     35 non-null      int64  
4   gender   33 non-null      object  
dtypes: int64(2), object(3)
```

4:

```
[7]: data.describe()
```

```
[7]:
```

	id	mark
count	35.000000	35.000000
mean	18.000000	74.657143
std	10.246951	16.401117
min	1.000000	18.000000
25%	9.500000	62.500000
50%	18.000000	79.000000
75%	26.500000	88.000000
max	35.000000	96.000000

Exercise 2: Indexing and Slicing

1. Question: "Write the code to select the 'name' column."
2. Question: "Write the code to select the 'name' and 'mark' columns."
3. Question: "Write the code to select the first 3 rows."
4. Question: "Write the code to select all rows where the 'class' is 'Four'."

1:



```
[9]: data["name"]
```

```
[9]: 0    John Deo
      1    Max Ruin
      2    Arnold
      3    Krish Star
      4    John Mike
      5    Alex John
      6    My John Rob
      7    Asruid
      8    Tes Qry
      9    Big John
     10    Ronald
     11    Recky
     12    Kty
     13    Bigy
     14    Tade Row
     15    Gimmy
     16    Tumyu
     ..
```

2:

```
[12]: data[["name", "mark"]]
```

```
[12]:
```

	name	mark
0	John Deo	75
1	Max Ruin	85
2	Arnold	55
3	Krish Star	60
4	John Mike	60
5	Alex John	55
6	My John Rob	78
7	Asruid	85

3:

```
[13]: data[:3]
```

```
[13]:
```

	id	name	class	mark	gender
0	1	John Deo	Four	75	female
1	2	Max Ruin	Three	85	male
2	3	Arnold	Three	55	male

```
[ ]:
```

4:

```
[14]: data[data["class"]=="Four"]
```

```
[14]:
```

	id	name	class	mark	gender
0	1	John Deo	Four	75	female
3	4	Krish Star	Four	60	female
4	5	John Mike	Four	60	female
5	6	Alex John	Four	55	male
9	10	Big John	Four	55	female
15	16	Gimmy	Four	88	male

Exercise 3: Data Manipulation

1. Question: "Write the code to add a new column 'passed' that indicates whether the student passed (mark >= 60)."
2. Question: "Write the code to rename the 'mark' column to 'score'."
3. Question: "Write the code to drop the 'passed' column."

1:

```
[15]: data['passed'] = data["mark"] >= 60
```

```
[16]: data.head()
```

```
[16]:
```

	id	name	class	mark	gender	passed
0	1	John Deo	Four	75	female	True
1	2	Max Ruin	Three	85	male	True
2	3	Arnold	Three	55	male	False
3	4	Krish Star	Four	60	female	True
4	5	John Mike	Four	60	female	True

2:

```
[20]: data = data.rename(columns={"mark": "score"})
```

```
[21]: data.head()
```

```
[21]:
```

	id	name	class	score	gender	passed
0	1	John Deo	Four	75	female	True
1	2	Max Ruin	Three	85	male	True
2	3	Arnold	Three	55	male	False
3	4	Krish Star	Four	60	female	True
4	5	John Mike	Four	60	female	True

3:

```
[23]: data.drop("passed", axis=1, inplace=True)
```

```
[24]: data.head()
```

```
[24]:
```

	id	name	class	score	gender
0	1	John Deo	Four	75	female
1	2	Max Ruin	Three	85	male
2	3	Arnold	Three	55	male
3	4	Krish Star	Four	60	female

Exercise 4: Aggregation and Grouping

1. Question: "Write the code to group the DataFrame by the 'class' column and calculate the mean 'mark' for each group."
2. Question: "Write the code to count the number of students in each class."



3. Question: "Write the code to calculate the average mark for each gender."

1:

```
[27]: grouped_data = data.groupby('class')
[28]: grouped_data['score'].mean()

[28]: class
      Eight    79.000000
      Fifth    78.000000
      Five     80.000000
      Four     68.750000
      Nine     41.500000
      Seven    77.600000
      Six      82.571429
      Three    73.666667
      Name: score, dtype: float64
```

2:

```
[31]: data.groupby('class')['id'].count()

[31]: class
      Eight     1
      Fifth     1
      Five      2
      Four      8
      Nine      2
      Seven    10
      Six       7
      Three     3
      Name: id, dtype: int64
```

3:

```
[32]: data.groupby('gender')['score'].mean()

[32]: gender
      female    77.312500
      male      71.588235
      Name: score, dtype: float64
```

[]:

Exercise 5: Advanced Operations

1. Question: "Write the code to create a pivot table with 'class' as rows, 'gender' as columns, and 'mark' as values."
2. Question: "Write the code to create a new column 'grade' where marks ≥ 85 are 'A', 70-84 are 'B', 60-69 are 'C', and below 60 are 'D'."
3. Question: "Write the code to sort the DataFrame by 'mark' in descending order."

1:




```
[33]: data.pivot_table(index='class', columns='gender', values='score')
```

```
[33]: gender  female  male
```

	class	
	female	male
Eight	NaN	79.0
Fifth	NaN	78.0
Five	NaN	80.0
Four	63.8	77.0
Nine	65.0	18.0
Seven	81.4	73.8

2:

```
[37]: import numpy as np
```

```
[38]: # List of conditions (evaluated in order)
```

```
conditions = [  
    data['score'] >= 85,  
    data['score'] >= 70,  
    data['score'] >= 60  
]  
# List of corresponding values for each condition  
choices = ['A', 'B', 'C']  
data['grade'] = np.select(conditions, choices, default='D')
```

```
[39]: data.head()
```

```
[39]:
```

	id	name	class	score	gender	grade
0	1	John Deo	Four	75	female	B
1	2	Max Ruin	Three	85	male	A

3:

```
[40]: data.sort_values(by='score', ascending=False)
```

```
[40]:
```

	id	name	class	score	gender	grade
32	33	Kenn Rein	Six	96	female	A
11	12	Recky	Six	94	female	A
31	32	Binn Rott	Seven	90	female	A
10	11	Ronald	Six	89	female	A
24	25	Giff Tow	Seven	88	male	A
15	16	Gimmy	Four	88	male	A
14	15	Tade Row	NaN	88	male	A
13	14	Bigy	Seven	88	female	A
12	13	Kty	Seven	88	female	A

Exercise 6: Exporting Data

1. Question: "Write the code to save the DataFrame with the new 'grade' column to a new CSV file."

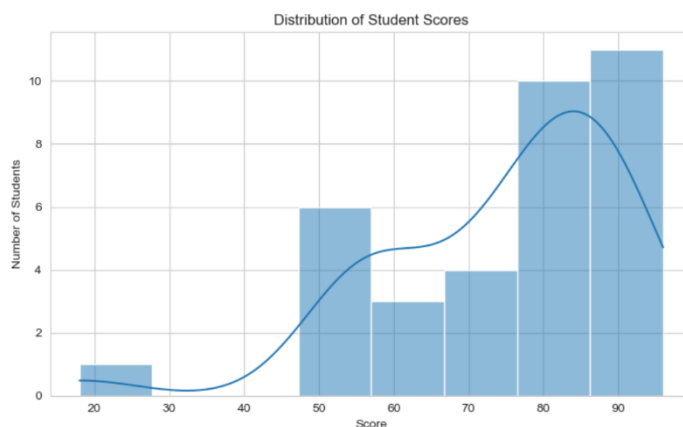
```
[41]: data.to_csv("student_complet.csv")
```

```
[ ]:
```

```
|
```

id	name	class	score	gender	grade
1	John Deo	Four	75	female	B
2	Max Ruin	Three	85	male	A
3	Arnold	Three	55	male	D
4	Krish Star	Four	60	female	C
5	John Mike	Four	60	female	C
6	Alex John	Four	55	male	D
7	My John Rob	Fifth	78	male	B
8	Asruid	Five	85	male	A
9	Tes Qry	Six	78		B
10	Big John	Four	55	female	D
11	Ronald	Six	89	female	

Exercise 7: If finished early try visualising the results



Day 4: Task 1

Using the 'GDP (nominal) per Capita.csv' which can be downloaded from the shared Folder, complete the below exercises and paste your input and output. Work individually, but we will work and support each other in the room.

- Read and save the 'GDP (nominal) per Capita' data to a data frame called "df" in Jupyter notebook
- Print the first 10 rows
- Print the last 5 rows
- Print 'Country/Territory' and 'UN_Region' columns

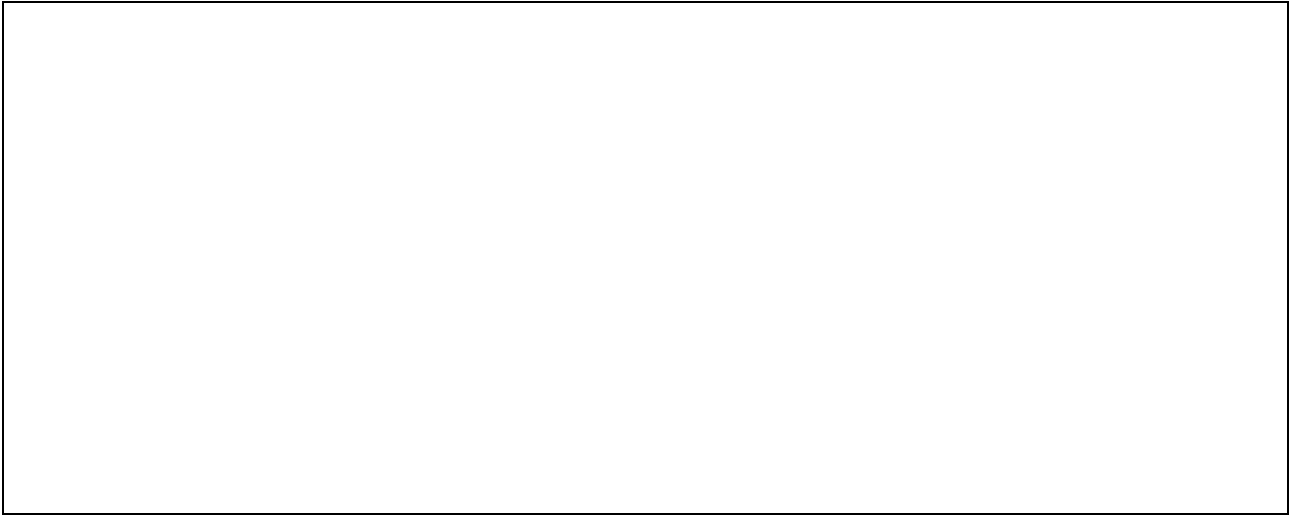
Day 4: Task 2

Back with 'GDP (nominal) per Capita'. As a group, import and work your way through the Day_4_Python_Activity.ipynb notebook which can be found on the shared Folder. There are questions to answer, but also opportunities to have fun with the data – paste your input and output below.

Once complete, and again as a group, work with some more data and have some fun – there is no set agenda for this section, other than to embed the skills developed this week. Paste your input and output below and upon return we'll discuss progress made.

[Additional data found here.](#)





Course Notes

It is recommended to take notes from the course, use the space below to do so, or use the revision guide shared with the class:



We have included a range of additional links to further resources and information that you may find useful, these can be found within your revision guide.

END OF WORKBOOK

Please check through your work thoroughly before submitting and update the table of contents if required.

Please send your completed work booklet to your trainer.

