

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: |
| Course Date: |
|  |

**Table of contents**

[Day 2: Task 1 3](#_Toc1765240926)

[Day 3: Task 1 3](#_Toc873857632)

[Exercise 1: Loading and Exploring the Data 4](#_Toc1363711086)

[Exercise 2: Indexing and Slicing 4](#_Toc1679584900)

[Exercise 3: Data Manipulation 4](#_Toc998449442)

[Exercise 4: Aggregation and Grouping 5](#_Toc267294297)

[Exercise 5: Advanced Operations 5](#_Toc1232452443)

[Exercise 6: Exporting Data 6](#_Toc1251390377)

[Exercise 7: If finished early try visualising the results 6](#_Toc66584577)

[Day 4: Task 1 7](#_Toc1452452084)

[Day 4: Task 2 8](#_Toc934594145)

[Course Notes 8](#_Toc1227711066)

[Additional Information 9](#_Toc1836543406)

# 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 | A screenshot of a computer  Description automatically generated |

# **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."

|  |
| --- |
| A screen shot of a computer  Description automatically generated  A screenshot of a computer program  Description automatically generated |

### **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'."

|  |
| --- |
| One column selected: names  Selecting column 'name'  Two Columns selected: names and mark  Selecting 2 columns; name and mark  First 3 Rows  A screenshot of a computer  Description automatically generated  Rows where the class is ‘Four’ |

### **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."

|  |
| --- |
| New column called ‘Passed’  A screenshot of a computer  Description automatically generated  Column name change from ‘mark’ to ‘score’  A screen shot of a computer  Description automatically generated  Dropped the column ‘passed’  A screenshot of a computer  Description automatically generated |

### **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."

|  |
| --- |
| Mean mark by ‘class’ group  A screen shot of a computer  Description automatically generated  Count for number of students in each class  A screenshot of a computer program  Description automatically generated  Average Mark for each gender  A screenshot of a computer  Description automatically generated |

### **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."

|  |
| --- |
| Pivot Table  A screenshot of a computer program  Description automatically generated  New Column ‘Grade’ conditional flow and functions  A screenshot of a computer program  Description automatically generated |

### **Exercise 6: Exporting Data**

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

|  |
| --- |
| Saved csv: students\_with\_grades.csv  A screen shot of a computer program  Description automatically generatedA screenshot of a computer  Description automatically generated |

### **Exercise 7: If finished early try visualising the results**

|  |
| --- |
| A screen shot of a graph  Description automatically generated  A screenshot of a computer  Description automatically generated  A screen shot of a computer  Description automatically generated  A screen shot of a computer  Description automatically generated |

# **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

|  |
| --- |
| Read and save the ‘GDP (nominal) per Capita’ csv    Print the first 10 rows with basic information  A screenshot of a computer  Description automatically generated  Print the last 5 rows  A screenshot of a computer  Description automatically generated  Print ‘Country/Territory’ and ‘UN\_Region’ columns  A screenshot of a computer  Description automatically generated |

# **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.](https://justit831-my.sharepoint.com/:f:/g/personal/danpe_justit_co_uk/Er0ybU9i0AZKiuGaCWZyj2ABoqKD23zwLGdJf3WlaixpRA?e=QVj2Bs)

|  |
| --- |
| Number of Countries per region  A screenshot of a computer  Description automatically generated  A screen shot of a computer code  Description automatically generated  A screenshot of a computer  Description automatically generated  Plot Distribution of GDP per Capita in Europe using plotly  A screen shot of a computer  Description automatically generated  Continuation of the graph of Distribution of GDP per Capita in Europe  A screenshot of a computer program  Description automatically generated  A black rectangular object with a black stripe  Description automatically generated |

|  |
| --- |
| **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:

Python is difficult but not impossible to learn, with LLMs like Gemini, Claude, ChatGPT/Copilot learning to code can be made easy.

However, for me the learning curve is still steep in the timeframe, therefore, I’m combining my debugging skills with Ais ability to write the logic. Still the code generated by AI contains errors; mostly syntax errors, there is still human input with necessary to have functioning code.

In this stage on my learning, I’m focused on producing minimum satisfactory results and then with iterative learning and practice to improve skills and confidence with the syntax and logic of Python and its key data libraries, NumPy, Pandas, Matplotlib, Seaborn.

Over the next two months, I want to dedicate 2 hours per week to Python to build up proficiency in Python and add the use of it in other aspects of data; Python in Excel, Google sheets, linking with DBs in the cloud, using Kaggle, Freecodecamp, and Codeacademy courses. Additionally, I will complete projects incorporating these skills to add on Github.

Other cohorts have agreed to work on projects after the course to develop and showcase our skills.

**Areas I want to improve on**

* Programming Logic
* Syntax of Pandas
* Web scraping
* Statistics and visualisation (what charts to use and why)
* Blending data by connecting with Databases; SQL Server, MySQL, SQLite, MongoDB, Azure Cosmos DB, Google BigQuery, AWS Dynamo DB.

|  |
| --- |
| **Additional Information** |

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.**