

**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 | # Looping through 1 to 100  for num in range(1,101):  # checking if the number is divisible by 3   if num%3==0:    print("fizz")   # checking if the number is divisible by 5   elif num%5==0:      print("buzz")   #checking if the number is divisible by both 3 and 5   elif num%3==0 and num%5==0:      print("fizzbuzz")   #if none of the above condition is satisfied, prints the number   else:    print(num)  **OUTPUT:** |

# **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. Code to read a CSV file into Pandas DataFrame**    **2. Code to display the first 5 rows of the DataFrame**    **3. Code to get the information about the DataFrame.**    **4. Code to get summary statistics for the DataFrame** |

### **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. Code to select the 'name' column**    **2. Code to select the 'name' and 'mark' columns**    **3. Code to select the first 3 rows**    **4. Code to select all 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."

|  |
| --- |
| **1. Code to add a new column 'passed' that indicates whether the student passed (mark >= 60)**    **2. Code to rename the 'mark' column to 'score'**    **3. Code to drop the 'passed' column** |

### **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. Code to group the DataFrame by the 'class' column and calculate the mean 'mark' for each group**    **2. Code to count the number of students in each class**    **3. Code to calculate the average mark for each gender** |

### **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.Code to create a pivot table with 'class' as rows, 'gender' as columns, and 'mark' as values**    **2. 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. Code to sort the DataFrame by 'mark' in descending order** |

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

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

|  |
| --- |
| **Code to save the DataFrame with the new 'grade' column to a new CSV file.** |

### 

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

Histogram:

|  |
| --- |
| **Bar chart** |

# **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 Jyputer notebook
* Print the first 10 rows
* Print the last 5 rows
* Print ‘Country/Territory’ and ‘UN\_Region’ columns

|  |
| --- |
| **1. Read and save the ‘GDP (nominal) per Capita’ data to a data frame called “df”**    **2. Print the first 10 rows**    **3. Print the last 5 rows**    **4. 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.](https://justit831-my.sharepoint.com/:f:/g/personal/danpe_justit_co_uk/Er0ybU9i0AZKiuGaCWZyj2ABoqKD23zwLGdJf3WlaixpRA?e=QVj2Bs)

|  |
| --- |
| **1.To print the shape of Dataframe**     1. **To print the count and datatype**     **3.To print all column names**    **4.To print statistical summary of numerical column**    **5. To print the number of missing values**    **6.To convert a column to numeric**    **7.Creating a new column**    **8. To count number of countries in a column**    **9. To print the row with country name ‘European Union [n 1]’**    **10. To view the IMF\_Estimate column of European Union[n 1]**    **11. To calculate average GDP Per Capita for Europe**    **12. To list countries in Europe with GDP higher than UK**    **13. Average GDP Per Capita for each region**    **14. Multiple Aggregations**    **15. To get countries count per region**    **16. Richest Country that have maximum IMF\_Estimate**    **17. To find which countries below average by IMF world estimate**    **18. To find all rows with IMF estimate 0**    **19. To find the number of countries with IMF\_Estimate=0**    **20. To count how many countries having highest UN Estimate**    21. **To find the country Name with highest UN Estimate**    **22. To find which country has highest Worldbank Estimate?** |
| **23. To find which country has highest IMF Estimate** |
| **24. Listing top 5 countries**    **25. To check how many 0 values exist using NumPy**    **26. Finding the column with 0 values**    **27. Filling 0 Values by null values**    **28. To find average of Worldbank\_Estimate' and 'UN\_Estimate' column**    **29. Finding mean excluding null values**    **30. checking missing values**    **31. Viewing all rows with any missing values**    **Visualisation**  **Histogram**  **1.Histogram for Numeric columns**    **2.Histogram for selected columns** |
| 1. **Histogram with bins mentioned** |
| **4.Aggregation for WorldBank\_Estimate**    **5. Bins size**    **6.Bin size if not given any number**    **7.Histogam plot** |
| **8. Histogram with different bins and size** |
| **Correlation HeatMap**  1.    **2. To get correlation heat map using Matplotlib and Seaborn** |
| **3**. |
| 1. **Formatting the heatmap** |
| 1. **Formatting with colour**       **6**. |
| **Bar Plot**   1. **Returning first 5 rows**     **2.Bar plot with X and Y axis**      **3.Bar Plot with x and Y axis swapped**      **4**. |
| **5. Bar Plot** |
| **Scatter Plot** |
| **Boxplot and Outliers**  **1.Box Plot**      **2. Filter the dataframe if the value greater than 50000** |
| **3. Box Plot**      **4.Box Plot**      **5. Filter the dataframe if the value greater than 100000** |
| **6.Average of UN\_Estimate**    **7.Shape of the Dataframe**    **Creating Another Dataframe**  **1.Create a new dataframe called data excluding 5 countries with highest UN estimate**    **2.Displaying first 5 rows**    **3. Shape of Dataframe**    4**.**    **5.Box Plot**      **Removing Outliers**  1**.**    **2**.    **3**.    4**.**    **5**.    **6.Filter based on upper boundary and lower boundary**    **7.First 5 rows after filtering**    **8.Shape of filtered dataframe**    **9.Average of UN\_Estimate column**    **10.Average of WorldBnk\_Estimate column**    **11.Average of IMF\_Estimate column of filtered dataframe**    **12.Average of IMF\_Estimate column of original Dataframe** |

|  |
| --- |
| **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:

|  |
| --- |
| **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.**