**Lab 9b: Data Searching (II)**

***(Additional) Topics covered****:*

* Exhaustive searching; handling “no match” situations
* Using compound conditions
* Counting to analyze data
* Plotting using python modules
* Using nested loops

PS: In this lab, you are NOT ALLOWED to use the following functions and methods for any purpose: max(someList), sum(someList) someList.count(someItem) , or someList.index(someVal)

1. Use *File Explorer* in Windows to navigate to the folder containing the final version of file ***Searching\_Part4.py***; make a copy of the file in the same folder and rename it as ***Searching\_Part5.py***. Open file ***Searching\_Part5.py*** using *Thonny* and modify it to **request a passenger class value as input from the user and count and display the number of passengers in that input class who survived**. **YOU MUST USE EXAUSTIVE SEARCHING**. Test your program for the following input cases and make sure it produces the expected data shown in TABLE 3. (Your code should take in the first column as an input and output “Passengers:” followed by the number shown in the second column.)

|  |  |
| --- | --- |
| **TABLE3** | |
| ***Input* class value** | **Number of passengers** |
| 3 | 119 |
| 2 | 87 |
| 1 | 136 |

Next, in addition to displaying the count of matching passengers in the input class, modify your program to **compute and print the percentage of survivors and their average age rounded to two decimal places**. Test your program for the following input cases and make sure it produces the expected data shown in TABLE 4. (Your code should take in the first column as an input and output “Passengers: 119 Percentage Survived: 24.44 Average Age: 21.41” for the first input.)

|  |  |  |  |
| --- | --- | --- | --- |
| **TABLE4** | | | |
| ***Input* class value** | **Number of passengers** | **Percentage** | **Average age** |
| 3 | 119 | 24.44 | 21.41 |
| 2 | 87 | 47.28 | 26.17 |
| 1 | 136 | 62.96 | 35.96 |

Do the results in the table above validate part or all the statement: “*one of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class passengers*.” Please discuss this with a nearby classmate.

How would your program behave if you run it for input class 4? Try it. Are your getting an error? If so, why are you getting this error? Instead of an error, fix your program to display a suitable error message if the input class is not valid. **AGAIN, YOU MUST USE EXHAUSTIVE SEARCHING**. Make sure your program still produces the expected output for all valid input class values in TABLE 4 above.

When done, **save and close** file ***Searching\_Part5.py***.

**BEFORE CONTINUING: Take a screenshot of your results from Thonny when running with passenger classes 1, 2, and 3. Include this screenshot in the relevant prompt in your Lab 09 report.**

**To take a screenshot: Use the “Snipping Tool” in Windows (press ⊞ and search for “Snipping”). Then, select “New Snip”, highlight the area of your screen that you wish to screenshot, and then use the “Copy” button to place a copy of your screenshot into your Lab09 Report.**

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

1. Make a copy of file ***Searching\_Part5.py*** in the same folder and rename it as ***Searching\_Part6.py***. Open file ***Searching\_Part6.py*** using *Thonny* and make the necessary changes so that it finds the percentage of survivors by gender instead of by passenger class (i.e., percentage of women survivors and percentage of men survivors) for input file *Titanic.csv* and prints the results in MEANINGFUL statements like below, rounded to two decimal places. Note that now, the program MUST NOT take any input from the user.



To double check your program for correctness, verify that it produces the following for input file *TitanicFewerRows.csv* (instead of *Titanic.csv*).



When done, **save and close** file ***Searching\_Part6.py***.  
**BEFORE CONTINUING: Take a screenshot of your results from Thonny when running this part. Include this screenshot in the relevant prompt in your Lab 09 report.**

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

1. File ***Plots.py*** inside folder *ProgramFiles* contains a simple data visualization program that produces a bar chart using Python’s *matplotlib.pyplot* module. Open this file and run the program to see the output it produces – by the way, this is the actual average cumulative GPAs of juniors and seniors for six departments at CSBSJU from Academic Year 2017-18. Study the code carefully.

Make a copy of file ***Searching\_Part6.py*** in the same folder and rename it as ***Searching\_Part7.py***. Open file ***Searching\_Part7.py*** using *Thonny* and change it to use file *TitanicFewerRows.csv* (instead of *Titanic.csv*)as input file. Next, change the program to produce a bar chart for the percentage of survivors by gender like the ones shown below, following the example in ***Plots.py***, including meaningful labels for the x-axis and y-axis.

You will need to figure out how to store the data needed on the bar chart’s x-axis and y-axis in two lists as in sample ***Plots.py***. Note that while gender values won’t change, the percentage of survivors is likely to change depending on the data read from the input file and thus can’t be hardcoded in your program; your program should work correctly FOR ANY valid input file.

When done, test your program for correctness by verifying that it produces the bar chart below ON THE LEFT for input file *TitanicFewerRows.csv*. Next, change your program to use file *Titanic.csv* instead of *TitanicFewerRows.csv* and verify that you are getting the bar chart below ON RIGHT (notice how percentage of male survivors drops significantly now).

|  |  |
| --- | --- |
| Using input file ***TitanicFewerRows.csv*** | Using input file ***Titanic.csv*** |
|  |  |

Does this support the claim that certain groups of people, such as women, were more likely to survive than others? Please discuss this with a nearby classmate.   
  
When done, **save and close** file ***Searching\_Part7.py***.

**BEFORE CONTINUING: Take a screenshot of your two plots, similar to those above. Include these screenshots in the relevant prompt in your Lab 09 report.**

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

1. Make a copy of file ***Searching\_Part7.py*** in the same folder and rename the copy as ***Searching\_Part8.py***. Open file ***Searching\_Part8.py*** using *Thonny* and make the necessary changes so that it produces a bar chart for the **percentage of survivors by passenger class**. Please use *TitanicFewerRows.csv* instead of *Titanic.csv* as an input file in your program initially.

Again, the percentage of survivors will likely change depending on the data read from the file and thus can’t be hardcoded in your program. To test your program for correctness, first run your program using file *TitanicFewerRows.csv* and make sure it produces the bar chart below ON THE LEFT. Then, modify your program to use file *Titanic.csv* instead of *TitanicFewerRows.csv* and verify that it produces the bar chart below ON THE RIGHT.

|  |  |
| --- | --- |
| Using input file ***TitanicFewerRows.csv*** | Using input file ***Titanic.csv*** |
|  |  |

When done, **save and close** file ***Searching\_Part8.py***.  
**BEFORE CONTINUING: Take a screenshot of your two plots, similar to those above. Include these screenshots in the relevant prompt in your Lab 09 report.**

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

1. As the number of label values on the x-axis increases (such as having 30 passenger class values instead of just 3), your program will get more complex requiring many more counters as well as branches in the outer *if* statement inside the loop used to distinguish among the label values. One way to simplify your program is to loop over the label values on the x-axis, and for each label, loop over the data in order to count survivors and total matches for that specific label. This way, instead of having a separate branch in the outer *if* statement for each label value on the x-axis, only one branch is needed along with two counters (one for the total matches and another for survivors) provided you remember to reset them back to zero for each label value. Note that you’ll still need an *if* statement to distinguish the survivors from the rest.

This solution uses a nested loop; there are definitely other ways to arrive at the same outcome and probably more efficiently, but we are more interested in simplicity at this stage.

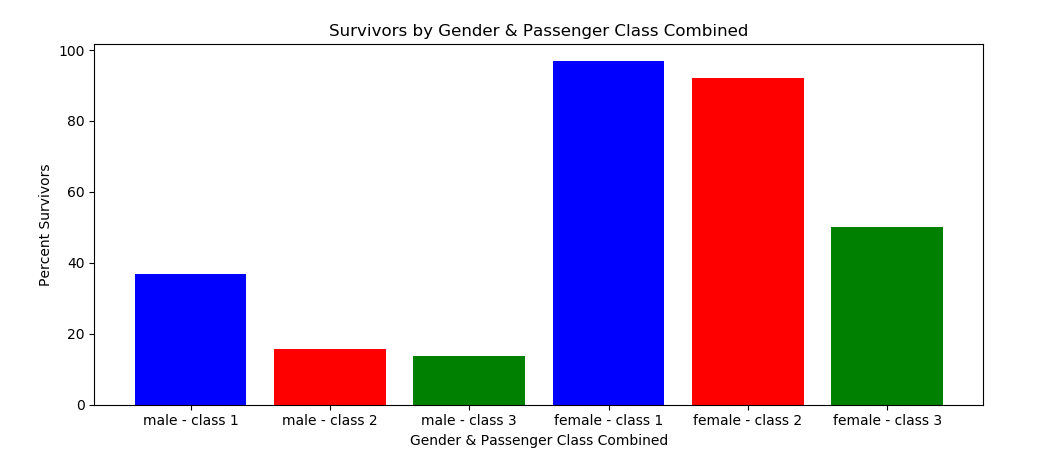
Make a copy of file ***Searching\_Part8.py*** in the same folder, and rename the copy as ***Searching\_Part9.py***. Open file ***Searching\_Part9.py*** using *Thonny* and make the necessary changes so that it produces the same bar chart from the previous step only using the nested loop logic described in this part.

When done, **save and close** file ***Searching\_Part9.py***.

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

1. Make a copy of file ***Searching\_Part9.py*** in the same folder, and rename the copy as ***Searching\_Part10.py***. Open file ***Searching\_Part10.py*** using *Thonny* and use the same logic from the previous part to produce a bar chart for the percentage of survivors by gender and passenger class COMBINED.

You will now need to use nested loops to go over both lists (gender values list and passenger class values list) to create the labels for the x-axis before looping over the data to count (using yet another nested loop). The expected outcome is shown below. Have fun!



When done, **save and close** file ***Searching\_Part10.py***.

**CHECKPOINT: Before you move on, get this part checked off with your TA or Lab Instructor!**

***Submission Instructions:***

The directions for submitting this lab are quite different than other labs because this lab will also be used for course assessment purposes. You will be submitting your Lab09.zip as usual, but also completing a report summarizing your results. See the “Lab09 Final Report” assignment for complete details. **You will NOT RECEIVE CREDIT for this lab if you do not complete the report portion.** You should take **time to reflect** when working on your report. **If necessary, your score will be adjusted based on the quality of your final report.**

Make sure your completed exercises AND Lab Report upload are checked off by a TA.