

## Analytics Worksheet 1

1. Name the stages of Analytics that a computer scientist typically conducts?

Accessing Data, PreProcessing Data, Visualisation, Analysis

2. Name two examples of data pre-processing.

Checking for outliers

Sorting Data

3. An airport has a self service check in. What pre-processing problems might exist here that a computer scientist may need to account for?

Errors in entering data resulting in missing data or outliers

4. Why would sorting data be useful for computer scientists?

It allows computer scientists to identify minimum and maximum values easily.

It makes it very easy to identify the most frequently occurring item.

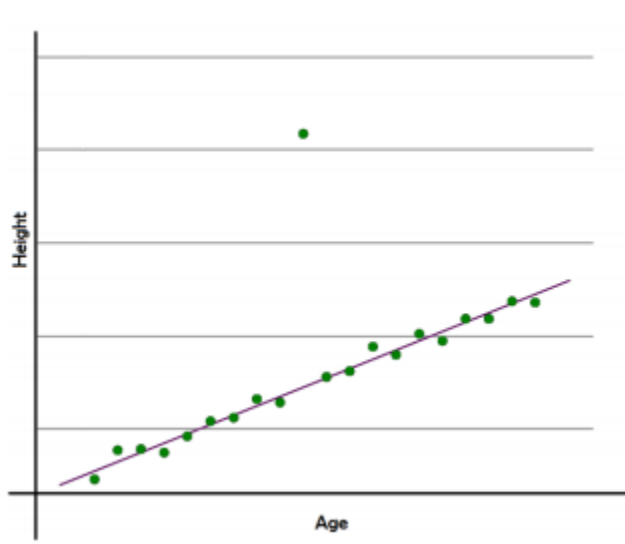
5. Suggest (you may have to research this) what is a possible outlier value for a heart rate?

Any heart rate below 40 or above 200

6. We had an example of an outlier being the age 200 entered. The easiest way to deal with this is to delete this piece of data. Suggest an alternative approach to removing this outlier

You could replace it with the average age of the rest of the data.

7. The visualisation below suggests an outlier. Identify this outlier and suggest a solution for dealing with this outlier in pre processing.



The outlier is the only data point that is not along the graph line.

A solution would be to delete the data.

8. Name the extension of a text file and a CSV file.

.txt

.csv

9. Give one reason why computer scientists often prefer CSV files.

A csv file can be large and therefore hold a very large amount of data.  
Python can read and write to csv files very quickly

10. When using the `open()` method, what is the first argument?

The first argument is the name of the file.

11. When using the `open()` method, what is the second argument? List and describe three options for the second argument.

The second argument is either of the following  
r - to read from the file  
w - to write to the file  
a - to append to the file

12. What does CSV stand for?

Comma separated values

13. What variable type is returned when reading a CSV file using the `open()` method?

The open method reads the file contents as a string.