



Lecture – Housekeeping

- ❑ The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly.
 - ❑ Please review Code of Conduct (in Student Undertaking Agreement) if unsure
- ❑ No question is daft or silly - **ask them!**
- ❑ There are Q&A sessions midway and at the end of the session, should you wish to ask any follow-up questions.
- ❑ Should you have any questions after the lecture, please schedule a mentor session.
- ❑ For all non-academic questions, please submit a query: www.hyperiondev.com/support

Lecture Objectives

1. Composition of XML and JSON files
2. CSV files
3. Manipulation of files

Introduction

- Before we can do any analysis, we need to obtain the data which can come from different sources
- In this lecture, we will cover two types of data formats which you will need to be familiar with:
 1. XML files
 2. JSON files

XML files

- eXtensible Markup Language (XML) files are used to transmit data over the internet
- XML files use a hierarchical structure consisting of elements, attributes, and text content.
- Elements can be nested within each other, forming a tree structure that represents the relationship between the data components

XML files

```
<menu>
  <meal type="Breakfast">
    <name>Belgian Waffles</name>
    <total>$5.95</total>
    <description>Two of our famous Belgian Waffles with plenty of
    real maple syrup</description>
  </meal>
  <meal type="Lunch">
    <name>Strawberry Belgian Waffles</name>
    <total>$7.95</total>
    <description>Light Belgian waffles covered with strawberries
    and whipped cream</description>
  </meal>
</menu>
```

JSON files

- JSON stands for JavaScript Object Notation.
- The syntax for a JSON object looks like a Python dictionary. It is a key-value structure enclosed with curly braces.
- The key-value structure is a data structure that uses an associative array of keys, where each key is associated with only one value in a collection

JSON files

```
{
  "books": [{
    "title": "Homegoing",
    "author": "Yaa Gyasi",
    "year": "2002",
    "edition": 2
  },
  {
    "title": "The river between",
    "author": "Ngugi wa Thiong'o",
    "year": "2004"
  }
]
```


CSV files

- Comma Separated Values (CSV) files are used for exchanging large chunks of data between systems
- CSV files are widely supported. Many spreadsheet programs and programming languages support CSV files

CSV file structure

```
A, B, C, D  
1, 2, 3, 4  
4, "5,3", 6, 7.3
```

- Two-dimensional structure containing rows of data
- Each row contains cells
- The cells within a row are separated by commas
- Note the value with the double quotation marks: we want the computer to disregard the comma inside the quotation marks and read it as one value



Questions and Answers

Questions around Data Formats

