

# Data Academy HLT

Tasks	GLH
Core Learning - Pandas fundamentals	+150 minutes
Portfolio Task - Pandas projects	120 - 240 minutes
Flipped Learning - Data Visualisation	30 minutes
Enrichment - More Pandas concepts	60 minutes
Technology - Cloud Computing	60 minutes
Wellbeing - Sleep	60 minutes
Soft Skills - Active listening	60 minutes
Employability - What to expect in interview	60 minutes



# Core Learning 1 – Dirty Dataset part 1

- 1- Import the file named Data\_Analyst\_Sample.csv in Pandas
- 2- Explore the file with at least 3 common exploration methods in Pandas
- 3- How many rows and columns the data file has?
- 4- Which column names have missing values and how many?
- 5- Remove missing values from columns that contain Pandas object datatype
- 6- Replace missing values by zero in columns that contain numbers
- 7- Verify that no missing values are in the data file

Don't forget to import your python libraries before using them! They need to be imported only once in your script

# Marking Criteria – Core Learning 1

	Pass	Merit	Distinction
Code	<ul style="list-style-type: none"><li>Imported the Pandas library properly</li><li>Used at least one exploration command</li><li>Was able to either delete or replace some missing values</li></ul>	<ul style="list-style-type: none"><li>Imported the Pandas library properly</li><li>Used at 2 exploration commands</li><li>Attempts to delete all rows with missing values</li><li>Attempts at replacing all missing values</li></ul>	<ul style="list-style-type: none"><li>Imported the Pandas library properly</li><li>Used at 3 exploration commands</li><li>Was able to delete all rows with missing values</li><li>Was able to replace all missing values</li></ul>

# Core Learning 2 - Dirty dataset part 2



1- Open the "errors.csv" file with Excel to check the errors associated to the "Campaign name".

2- Correct the errors that appear in the Data\_Analyst\_Sample.csv imported with the Pandas into a Python script:

2.1 Count how many errors there are in the dataset (you might want to use a loop for this and using the .contains() method to find the errors)

2.2 Use the .replace() method to correct the errors. Verify the errors have been corrected

Example of a technical interview question!

Help:  
Use the Pandas session slides and what you learnt in the Python foundation sessions to help you

# Marking Criteria – Core Learning 2

	Pass	Merit	Distinction
Code	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas</li><li>• Attempts at counting and correcting some errors</li></ul>	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas</li><li>• Able to count and correct at least 1 error</li></ul>	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas</li><li>• Able to count and correct all errors</li></ul>

# Core Learning 3 - Exploratory Data Analysis (EDA)

30-60 mins



Use the Pandas library to perform the 3 essential steps of any EDA using the file named Superstore\_data\_sample.csv:

## 1-Data Exploration:

Import your file in Pandas and explore the file with key exploration commands and graphs

## 2-Data Cleaning:

Remove any rows that have missing values from columns that would be an issue to identify (such as name of a customer, name of a product etc).

Replace missing values by zero if justified (give your reasons).

## About the data:

The data contains information about products that has been sold from 2014 to 2017 and their customers at a superstore in the US.





# Core Learning 3 - continued

## 3. Data Analysis & Visualisation:

- How many categories and sub-categories of products there are?
- Which products were sold the most and the least?
- Which products were sold the most in the different US regions?
- Which States sold the most?
- Which products were the most successful each year?
- Which customers ordered the most and what was the total value of their purchases?
- How many customers brought products each year?
- How many customer segments there are in the dataset?

Use graphs to help you answer to questions!

Important note: Questions are here to guide you do some analysis. Pick 2-3 questions to do your analysis (see marking). You can come up with your own questions to answer as well based on the dataset. Use what you have seen in the course and slides to help you do some analysis and plot some results.

# Marking Criteria – Core Learning 3

	Pass	Merit	Distinction
Code	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas and explore the file.</li><li>• Attempts at doing some cleaning</li><li>• Attempts at doing some analysis and plot some data (answered at least to 1 question related to the product and one question related to the customers)</li></ul>	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas and explore the file.</li><li>• Was able to do some cleaning</li><li>• Was able at performing some analysis and plot some data (answered at least to 2 questions related to the product and one question related to the customers)</li></ul>	<ul style="list-style-type: none"><li>• Was able to import the dataset in pandas and explore the file.</li><li>• Was able to do some cleaning</li><li>• Was able at performing some analysis and plot some data answered at least to 3 questions related to the product and 2 questions related to the customers)</li></ul>





# Portfolio Tasks -

Portfolio  
Projects

Make a Gradebook With  
Python & Pandas

<https://realpython.com/pandas-project-gradebook/>

Time-series with Pandas

<https://www.dataquest.io/blog/tutorial-time-series-analysis-with-pandas/>

Data Science Project Using  
Numpy, Pandas, and  
Matplotlib

<https://python.plainenglish.io/beginners-data-science-project-using-numpy-pandas-and-matplotlib-c4608e196436>

Remember to  
upload to GitHub





# Flipped Learning

## Flipped Learning

Use the following resources to support you ahead of next weeks session on the following concepts:

- What makes a good visualisation?
- How can I create a story of my data?

Watch YouTube videos;

Data visualisation theory - <https://www.youtube.com/watch?v=loYuxWSsLNc>

Tableau - <https://www.youtube.com/watch?v=6xv1KvCMF1Q>



# Enrichment

Use the following resources to explain concepts in more detail:

Data plotting with Pandas- <https://realpython.com/pandas-plot-python/>

Fast, Flexible, Easy and Intuitive: How to Speed Up Your Pandas Projects-  
<https://realpython.com/fast-flexible-pandas/>

Combining Data in Pandas With-  
<https://realpython.com/pandas-merge-join-and-concat/>



# Technology

This week, our focus is on **Cloud Computing**.

Cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping you lower your operating costs, run your infrastructure more efficiently, and scale as a business needs change.

The main benefits cloud computing offers are:

- Reduced costs
- Improved speed
- Increased productivity
- Higher performance
- Reliability

For more information about Cloud Computing, visit the following links:

Microsoft Azure:

<https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing/#benefits>

Amazon Web Services (AWS):

<https://www.youtube.com/watch?v=mxT233EdY5c>

Quick overview of cloud computing:

[https://www.youtube.com/watch?v=M988\\_fsOSWo](https://www.youtube.com/watch?v=M988_fsOSWo)



# Wellbeing

This week, our focus is on **Sleep**.

We spend approximately a third of our lives asleep. Sleep is an essential and involuntary process, without which we cannot function effectively. It is as essential to our bodies as eating, drinking and breathing, and is vital for maintaining good mental and physical health. Sleeping helps to repair and restore our brains, not just our bodies.

Everyone needs sleep, but many of us have problems with it.

You may :

- Find it hard to fall asleep, stay asleep or wake up earlier than you'd like to
- Have problems that disturb your sleep, such as panic attacks, flashbacks, nightmares or psychosis
- Find it hard to wake up or get out of bed
- Often feel tired or sleepy – this could be because you're not sleeping enough, not getting good quality sleep or because of health problems
- Sleep a lot – which could include sleeping at times when you want, or need, to be awake.

To find out more about sleep and things you can do to improve your quality of sleep, visit the following websites:

<https://www.mind.org.uk/information-support/types-of-mental-health-problems/sleep-problems/about-sleep-and-mental-health/>

<https://www.nhs.uk/every-mind-matters/mental-health-issues/sleep/>

<https://www.sleepfoundation.org/mental-health>



# Soft Skills

This week, our focus is **Active Listening**.

Listening is one of the most important skills you can have. How well you listen has a major impact on your job effectiveness and on the quality of your relationships with others. Research suggests that we only retain about 25 -50% of a conversation.

By becoming a better listener, you can improve your productivity, as well as your ability to influence, persuade and negotiate. What's more, you'll avoid conflict and misunderstandings. All of these are necessary for workplace success!

There are many methods and tips that are available which can support us with becoming a better listener. Use the links below to explore some of these methods and find one which could work for you.

<https://www.mindtools.com/CommSkill/ActiveListening.htm>

<https://www.forbes.com/sites/womensmedia/2012/11/09/10-steps-to-effective-listening/?sh=1214bf8d3891>

<https://www.thebalancecareers.com/active-listening-skills-with-examples-2059684>



# Employability

Most of us are nervous when being interviewed for a job but what if there was a method that you could apply which could support you when answering questions in interview. By using the STAR method you can confidently answer questions knowing you have provided all the relevant details.

The STAR method is an interview technique that gives you a straightforward format you can use to tell a story by laying out the Situation, Task, Action, and Result.

- **Situation:** Set the scene and give the necessary details of your example.
- **Task:** Describe what your responsibility was in that situation.
- **Action:** Explain exactly what steps you took to address it.
- **Result:** Share what outcomes your actions achieved.

For further information about the STAR method, please use the following links:

<https://www.themuse.com/advice/star-interview-method>

<https://uk.indeed.com/career-advice/interviewing/star-technique>

<https://www.reed.co.uk/career-advice/star-method-what-you-need-to-know/>