

# Tech Career Talks



Python data  
engineering skills  
for public datasets



Online



Monday 11 November



12:30pm (30 mins)



Free

Hosted by:  
Sally Sam



# Who are LearnTech?



LearnTech are a tech-specialist Independent Training Provider



We are a regulated provider, rated 'Good' by Ofsted



We have been training remotely since 2017, implementing a range of cutting-edge technologies to create engaging learning environments

# Who are LearnTech?



We work with a range of employers across the UK, large and small, to deliver specialist tech training in areas such as software development, networking, cyber security and data analysis



Free\* Tech Courses (FTCs)



★★★★★ Excellent  
Employer reviews

★★★★★ Excellent  
Apprentice reviews

98.4%  
Pass Rate

# Participation








**It is your choice if you would like to code along.**

**I would love to hear from you!**

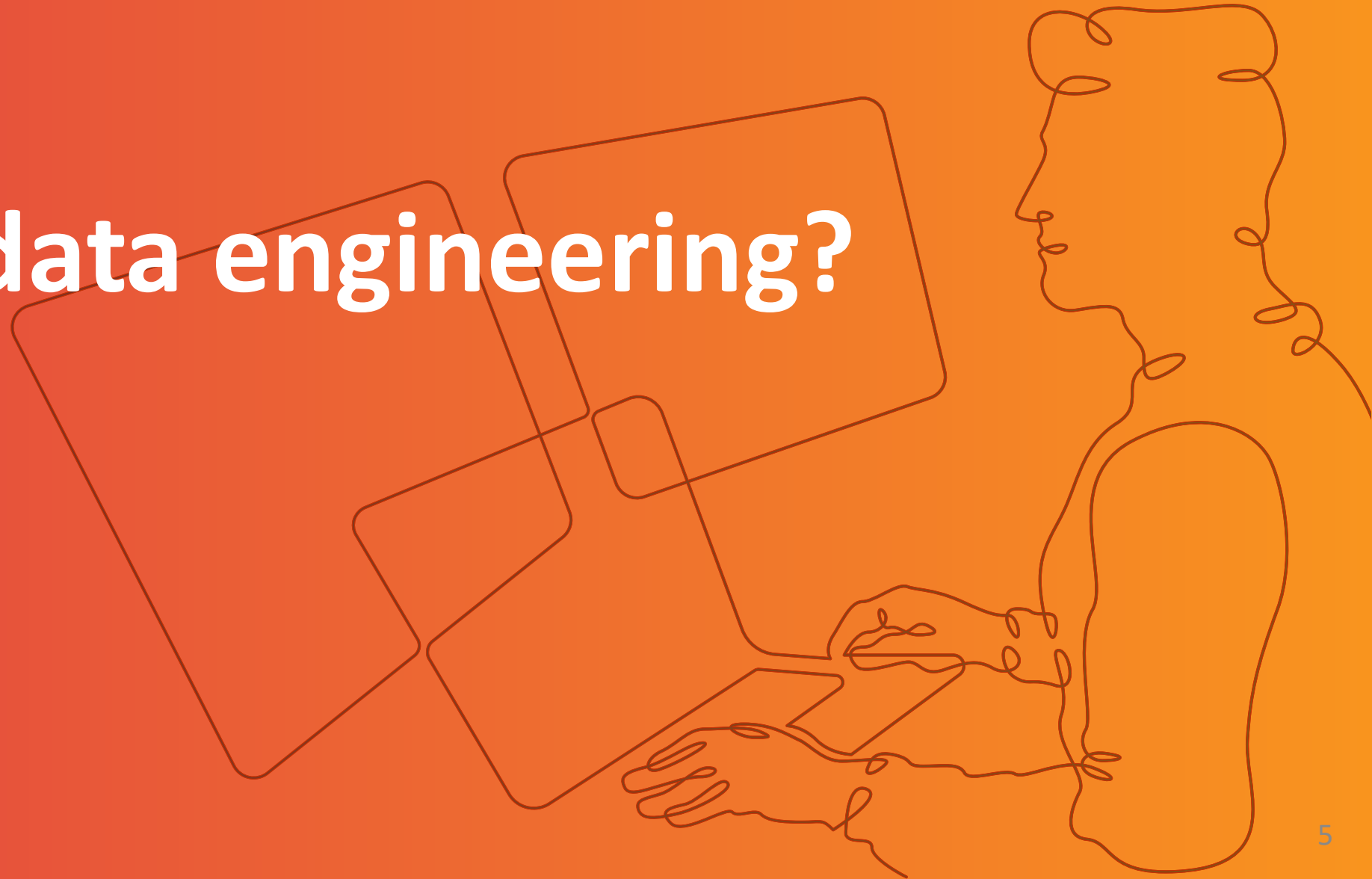
**At any time, please share comments, questions or feedback in the chat.**



For example:

-  “What do you mean by... ?”
-  “Can you repeat... ?”
-  “How did you... ?”
-  “Why is... ?”
-  “Where is... ?”

# What is data engineering?



# What is data?

The Cambridge Dictionary defines data as:

- “information, especially facts or numbers, collected... and used to help decision-making,
- or information in an electronic form that can be stored and used by a computer”

For example:



Text in a document



Images



Information in a spreadsheet

# What is data engineering?

Data engineering is making data useful to others.

You may hear people mention that data has been “cleaned”.



Data engineering can involve:

- extracting useful data,
- transforming or manipulating the data into a specific layout,
- and loading or saving the data so other people can use it.

We will see an example of data engineering over the next few of slides.

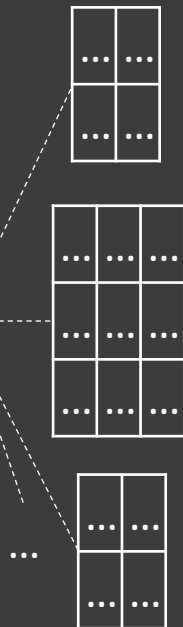
# Example of raw data table

This data table shows item reviews posted on a website.

- Quite unreadable and looks messy!
- There are also a lot of connected data tables.



review_id	date_time	text	...
05731sx	45047.89861	Not sure how long it will...	...
4829mc	44972.07292	great delivery,...	...
42259gt	45036.37292	the bottle doesn't fit...	...
b5351m	45214.60292	was ok	...
...	...	...	...



...	...
...	...

...	...	...
...	...	...
...	...	...

...	...
...	...



# Example of data engineering

How do we transform the messy looking table into a table that is more understandable? By using data engineering!

Messy looking data table with a lot of connected data tables from the previous slide.



Data engineering processes allow the data to be more readable to humans.



This is a table generated after data engineering.

- More understandable!



Name	Age	Country	Date	Review	...
Ismail al-Jazari	57	Turkey	01/05/2023	Not sure...	...
Wangari Maathai	27	Kenya	20/04/2023	the bottle...	...
Chien-Shiung Wu	43	China	15/10/2023	was ok	...
...	...	...	...	...	...

# Why is data engineering important?

To use the clean data for useful purposes such as analysis and AI. See an example below.

This is a table generated after data engineering.



Name	Age	Country	Date	answer	...
Ismail al-Jazari	57	Turkey	01/05/2023	Not sure how...	...
Wangari Maathai	27	Kenya	20/04/2023	the bottle doesn't...	...
Chien-Shiung Wu	43	China	15/10/2023	was ok	...
...	...	...	...	...	...

e.g. Can now analyse if there is a connection between age and a negative review

e.g. Can now analyse if the reviews get better or worse over time


e.g. Can predict who is most likely to buy the product and like it

And so much more!



# Learn more about data engineering through Free Tech Courses

<https://learn-tech.com/data-analysis-ftc/>

 **LearnTech**

Apprenticeships ▾ Employers ▾ Apprentices ▾ Free Tech Courses ▾ About ▾ Contact

## Data Analysis

Level 3


- Live Virtual Learning
- 16 Weeks
- Monthly Start Dates
- Worth 16 UCAS Tariff Points
- + W3Schools Certificate Programme Included
- £ Free\* / £1788 + VAT

Students will develop a sound understanding of essential data analysis techniques and tools, that can be applied effectively in a business context as part of a data analysis team. Students are taught the skills to source, collate, format, analyse and communicate data, helping drive business outcomes. Master secure data management and more!

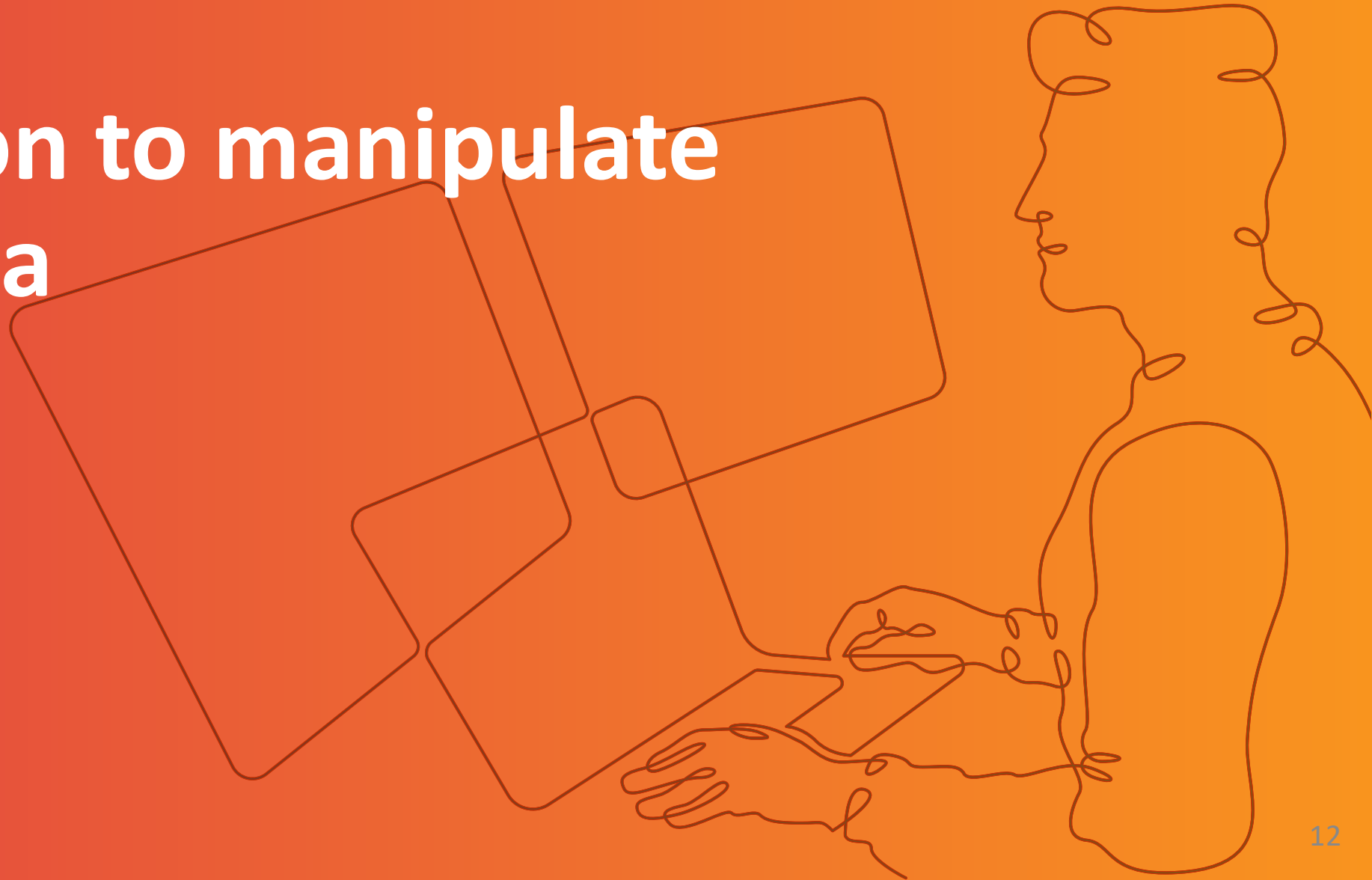
**Next Start Dates**

- Tuesday 29 October
- Tuesday 12 November
- Thursday 5 December
- Tuesday 14 January

[Apply ▸](#)



# Use Python to manipulate public data



Resources: [https://github.com/LearnTech-ssam/TechTalk\\_PythonDE](https://github.com/LearnTech-ssam/TechTalk_PythonDE)

## Scenario:

- We work as a Data Engineer at a company.
- A colleague would like to understand how many people are affected by severe flu over time.
- They would like to us to provide historic data in a specific layout.
- They will use the historic data to predict the future.

## Data engineering steps:

1. Agree data with colleague
2. Extract and transform data to the required layout
3. Check no errors have been made
4. Save the data and discuss with colleague

# If you are coding along

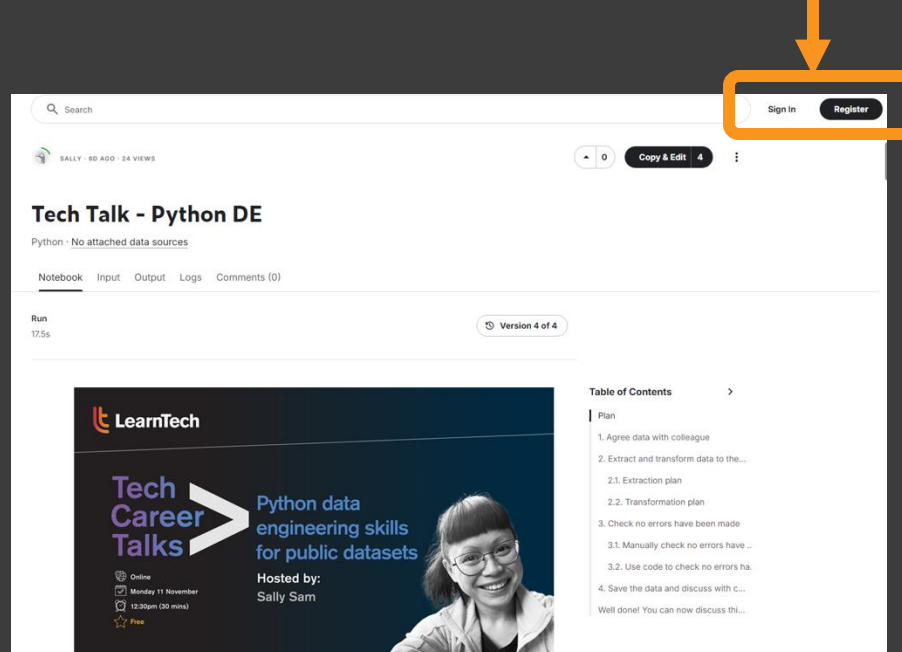


# If you are coding along: sign into Kaggle

Resources: [https://github.com/LearnTech-ssam/TechTalk\\_PythonDE](https://github.com/LearnTech-ssam/TechTalk_PythonDE)

Kaggle: <https://www.kaggle.com/code>

- **Step 1.** Sign in or create a free account using an email address



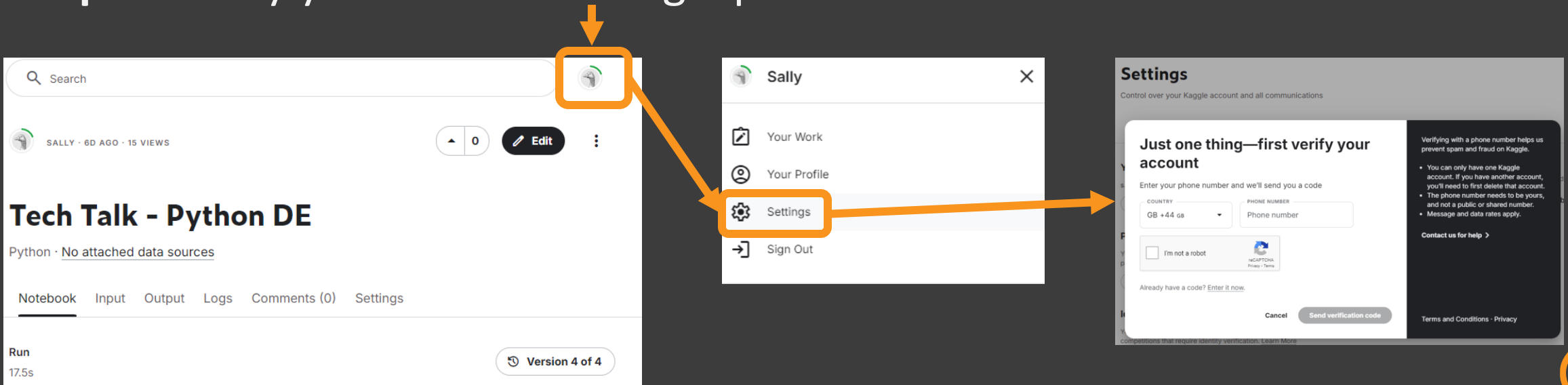
# If you are coding along: how to use internet through Kaggle

Resources: [https://github.com/LearnTech-ssam/TechTalk\\_PythonDE](https://github.com/LearnTech-ssam/TechTalk_PythonDE)

Kaggle: <https://www.kaggle.com/code>

To be able to use the internet through Kaggle:

- **Step 2.** Verify your account using a phone number




The first screenshot shows a Kaggle notebook titled "Tech Talk - Python DE". In the top right corner, the user's profile icon is highlighted with an orange box. An arrow points from this box to the second screenshot, which shows the user's profile dropdown menu. In this menu, the "Settings" option is highlighted with an orange box. Another arrow points from this box to the third screenshot, which shows the "Settings" page. On the "Settings" page, the section titled "Just one thing—first verify your account" is highlighted with an orange box. This section prompts the user to enter their phone number and send a verification code. To the right of this section, there is a list of bullet points explaining why verification is required.

**Settings**  
Control over your Kaggle account and all communications

**Just one thing—first verify your account**  
Enter your phone number and we'll send you a code

COUNTRY: GB +44 GB | PHONE NUMBER: Phone number

☐ I'm not a robot | 

Already have a code? [Enter it now.](#)

[Cancel](#) [Send verification code](#)

Verifying with a phone number helps us prevent spam and fraud on Kaggle.

- You can only have one Kaggle account. If you have another account, you'll need to first delete that account.
- The phone number needs to be yours, and not a public or shared number.
- Message and data rates apply.

[Contact us for help >](#)

Terms and Conditions · Privacy



# If you are coding along: create new notebook

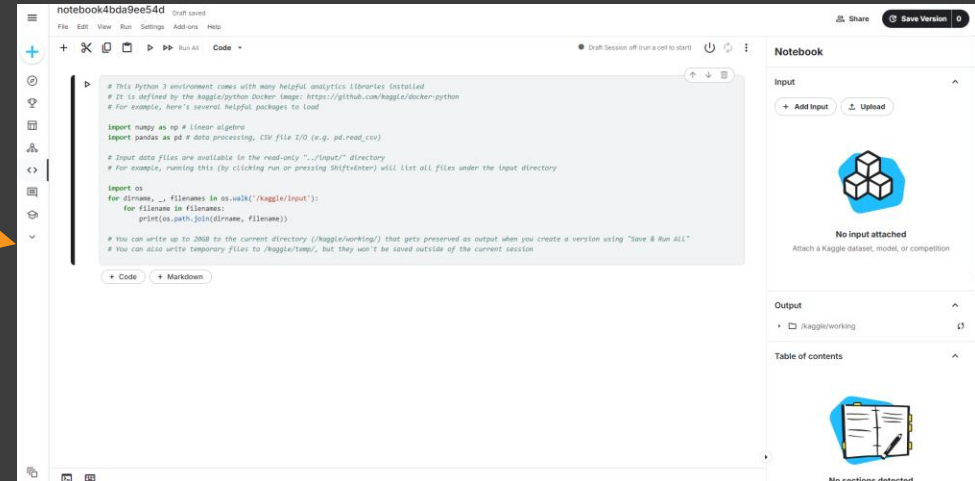
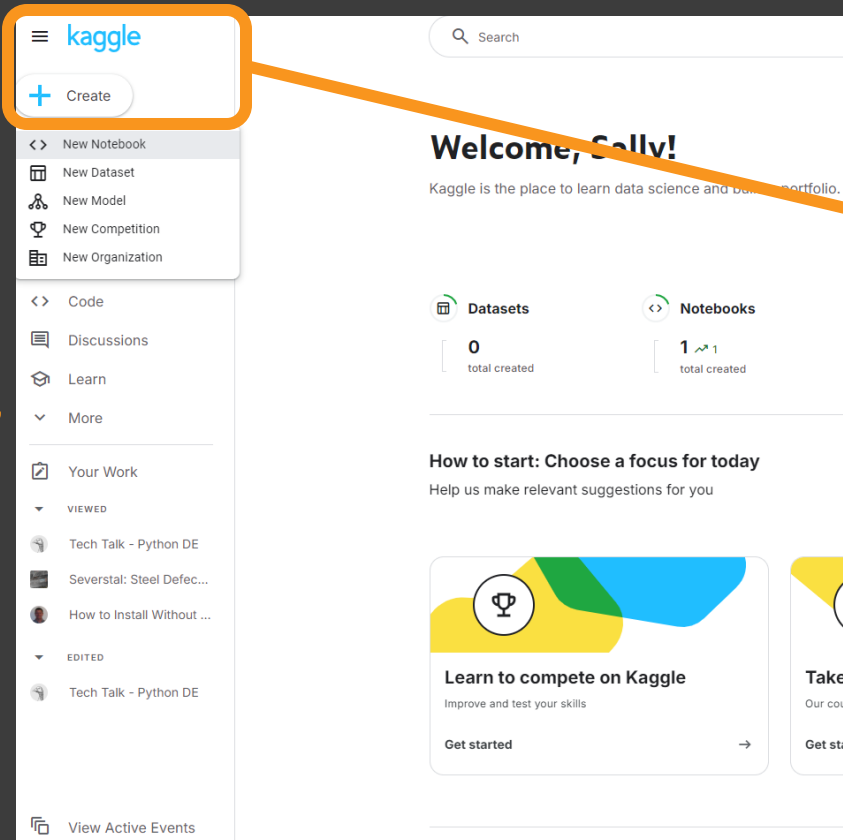
Resources: [https://github.com/LearnTech-ssam/TechTalk\\_PythonDE](https://github.com/LearnTech-ssam/TechTalk_PythonDE)

Kaggle: <https://www.kaggle.com/code>

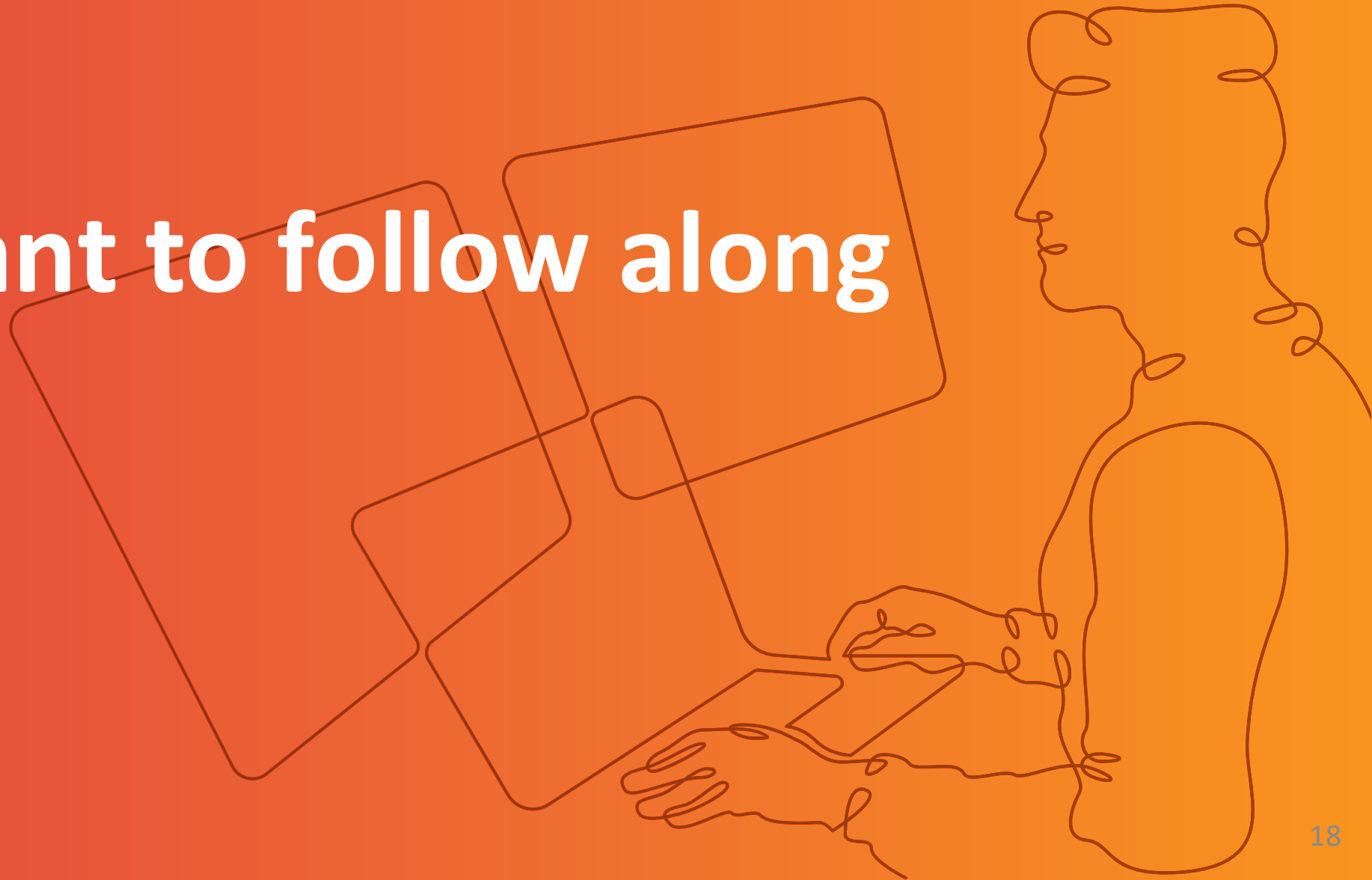
Click  
“+ Create”

then

click  
“New  
Notebook”



# If you want to follow along



# If you want to follow along

Resources: [https://github.com/LearnTech-ssam/TechTalk\\_PythonDE](https://github.com/LearnTech-ssam/TechTalk_PythonDE)

## Materials

- [data](#)
  - [2024-10-16 RAW England flu bed occupation 2023-24.csv](#) is a csv version of the raw data.
  - [2024-10-16 England flu bed occupation 2023-24.csv](#) is a csv of the cleaned data, after data engineering processes.
- [images](#) contains images shown in this repository.
- [README.md](#) is the file you are currently reading.
- [Tech\\_Talk\\_Python\\_DE.html](#) is a html version of the .ipynb file and is screen reader compatible.
- [Tech\\_Talk\\_Python\\_DE.ipynb](#) contains Python code to clean the raw data and save it as the cleaned csv. The file was created to be used in Kaggle Notebooks.

“In” are code cells that have run.

“Out” are the outputs of any code cells.

The number in square brackets, e.g. [3], is the order the cell was run.

### 2.1.3. Using the `read_excel` function in pandas, read that data from

- Only read in the “Flu” sheet
- Skip the first 13 rows
- Assign the first two rows as the header
- Use the `head` function to show the first 5 rows of the table on the screen. Change the number to

```
In [3]: raw_data = pd.read_excel(
        io = excel_link,
        sheet_name = "Flu",
        skiprows = 13,
        header = [0,1]
    )

    print("A snippet of the raw data:")
    raw_data.head(5)
```

A snippet of the raw data:

```
Out[3]:
```

	Unnamed: 0_level_0	Unnamed: 1_level_0	Unnamed: 2_level_0	Unnamed: 3_level_0	Unnamed: 4_level_0	2023-11-20 00:00:00	2023-11-21 00:00:00	2023-11-22 00:00:00
	Unnamed: 0_level_1	NHS England Region	NHS England Region	Code	Name	G&A flu beds	CC flu beds	G&A flu beds
0	NaN	-	NaN	-	ENGLAND (All Acute Trusts)	143.0	6.0	133.0

# Let's code!




# What next?



# Learn more about data engineering through Free Tech Courses

<https://learn-tech.com/data-analysis-ftc/>

 Apprenticeships ▾ Employers ▾ Apprentices ▾ Free Tech Courses ▾ About ▾ Contact

## Data Analysis

Level 3


- Live Virtual Learning
- 16 Weeks
- Monthly Start Dates
- Worth 16 UCAS Tariff Points
- + W3Schools Certificate Programme Included
- £ Free\* / £1788 + VAT

Students will develop a sound understanding of essential data analysis techniques and tools, that can be applied effectively in a business context as part of a data analysis team. Students are taught the skills to source, collate, format, analyse and communicate data, helping drive business outcomes. Master secure data management and more!

**Next Start Dates**






- Tuesday 29 October
- Tuesday 12 November
- Thursday 5 December
- Tuesday 14 January

Apply ▸



# FREE\*

**Kickstart your tech career with our FREE\* online qualifications**

-  Live virtual learning
-  8, 12 and 16 week courses
-  Optional Cisco & Other Certifications
-  Monthly start dates
-  W3Schools Certificate Programme Included (Level 3s)
-  Free subject to eligibility



**Web Design  
& Development**



**Cyber  
Security**



**Data  
Analysis**



**Coding  
Practices**

# Free Tech Courses

## Eligibility

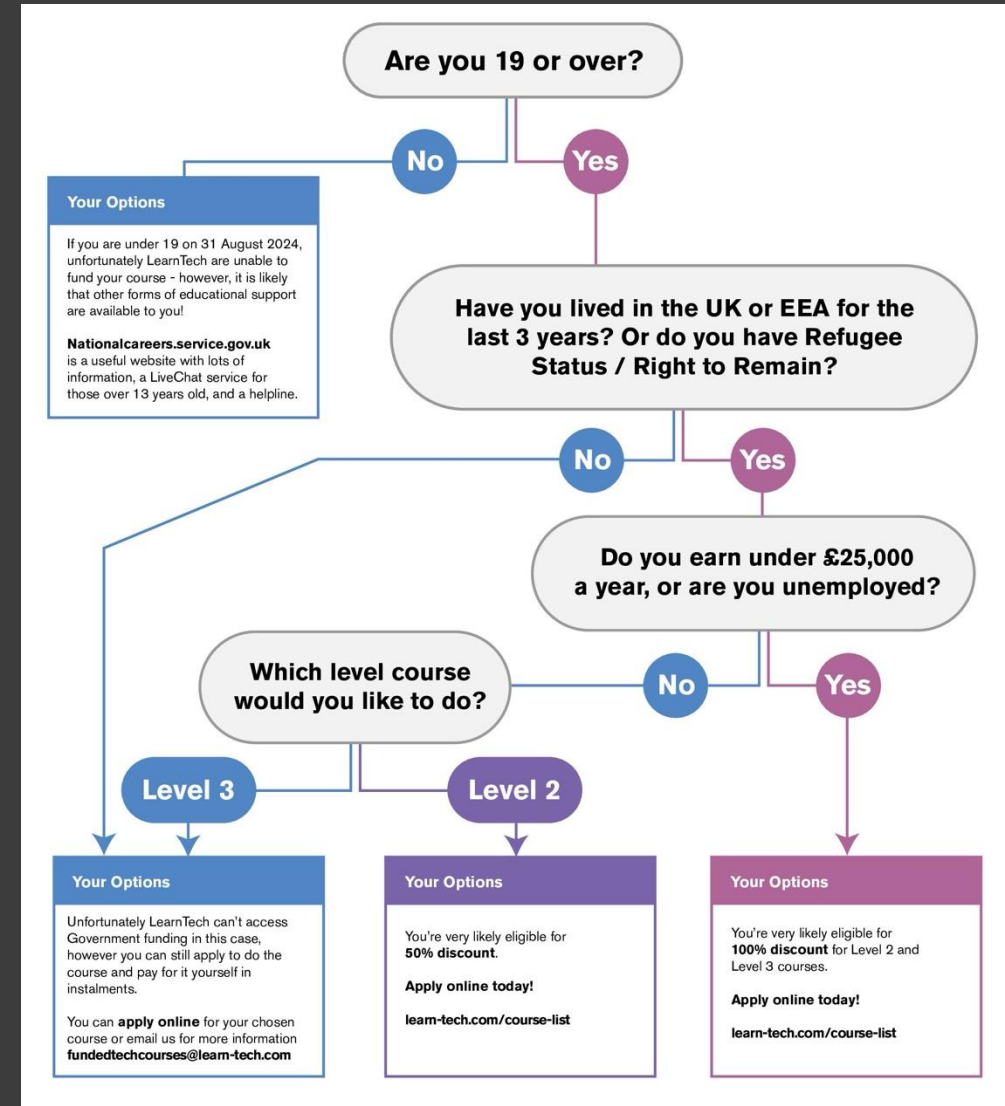
### Step 1 – Check Your Postcode Eligibility

Please enter your postcode below to check if your area is eligible for funding...

If you live in London, the Combined Authorities of Greater Manchester, West Midlands, South or West Yorkshire, unfortunately we are currently unable to access government funding for you to do this course for free.

See the full list of courses here:

<https://learn-tech.com/course-list/>





# Next Tech Career Talk

Next session will be...



**LearnTech**

**Tech Career Talks** 

**Create an AI model  
in 30 minutes  
with Python**

Hosted by:  
**Sally Sam**

 Online  
 Monday 25 November  
 12:30pm (30 mins)  
 Free



For more details and recordings visit  
[learn-tech.com/news-events](https://learn-tech.com/news-events)

# Connect with us

Keep up-to-date with the latest news and celebrations!



@LearnTechUK



@LearnTechUK



@LearnTechUK



@Learn\_Tech\_

Recordings are posted on social and our website

If you would like more information on data engineering or AI, you can email:

**Sally Sam**

Technical Trainer Mentor

[s.sam@learn-tech.com](mailto:s.sam@learn-tech.com)

# Tech Career Talks



Python data  
engineering skills  
for public datasets



Online



Monday 11 November



12:30pm (30 mins)



Free

Hosted by:  
**Sally Sam**

