

This document contains all the information you will need to complete the NHS live online career challenge. You should have received the following files:

- How-to-Guide (this document)
- The data sets (Step Up Excel Data)
- Presentation (DigData–Step-Up-Presentation.pptx)
- Excel cheat sheet (excel_cheatsheet.pptx)
- R GITHUB Links ([dig-data-step-up/README.md at main · nhsbsa-data-analytics/dig-data-step-up · GitHub](#))
- Python GITHUB Links ([Digdata/README.md at main · nhsengland/Digdata · GitHub](#))

The Scenario

Turning data into insight: Building awareness of mental health prescribing

The NHS is experiencing an unprecedented surge in demand for mental health services. Long waiting times, particularly for young people are placing a strain on patients and providers alike. In parallel, the use of prescription medications, especially for managing anxiety and depression, has seen a marked increase.

This challenge offer students the opportunity to delve into datasets reflecting drug usage trends to build awareness and inform decision making using NHS data to improve patient care.

Data Analysis– designed for students who are starting their data literacy journey

Explore data on the prescription rates of common mental health medications.

Data reporting – designed for students who would like to further develop their programming skills

You are a public health officer and want to better understand antidepressant prescribing trends over time.

Instructions

You can choose to work on Challenge **A, B or C**

- **Challenge A:** Analyse and visualise one month of prescription data in excel and produce a presentation
- **Challenge B:** Using R to analyse longitudinal prescription data and produce a report
- **Challenge C:** Using Python to analyse longitudinal prescription data and produce a report

We suggest you choose challenge A if you basic are interested in basic exploratory data analysis and improving your wrangling and data visualisation skills using Excel.

We suggest you choose Challenge B or C if you want to strengthen your skills in data wrangling, data analysis and data visualisation using a programming language such as R or Python

Both paths show a data role's approach to a problem, although they cover different skills.

Each section will have the relevant information to take you from zero to hero!

Software Set up

Posit Cloud – This is the online tool we recommend for tackling this path using the R programming language. It is an online instance of the popular RStudio software. You can create a free account at <https://posit.cloud/> – simple!

Google Colab – This is an online tool we recommend for the Python Programming language. It is an online instance of the popular Jupyter notebook software. You can create a free account at [Welcome to Colaboratory - Colab](#).

Excel – If you feel more comfortable, you can choose to complete this task using Microsoft Excel spreadsheet software. If you don't have an instance of Excel on your computer you can use a free, online version of Excel by creating an account here: [Free Microsoft 365 Online | Word, Excel, PowerPoint](#)

The Data

This data is from the NSHBSA Open Data Portal and is called the 'Prescription Cost analysis' dataset.

<https://opendata.nhsbsa.net/dataset/prescription-cost-analysis-pca-monthly-data>

This data is open source, which means:

- We can share it openly
- It has no security concerns

This is a monthly dataset that describes how many medicines were prescribed across all GP Practices in England, per NHS Region, and what they cost.

This is 'real data', so actually describes real anti-depressant prescribing in England.

Data Description

For Challenge A

We have merged one month of anti-depressant prescribing data into a single dataset.

‘Step Up Excel Data Dec 2024’

This dataset now contains 6 Columns

YM: The year and month, in the format YYYYMM, where 202412 is the same as December 2024

REGION: The NHS Region. There are 7 regions in the data.

ICB: The NHS Integrated Care Board. There are 42 ICB and 1 unidentified grouping in the dataset.

DRUG: The name of the anti-depressant medicine. There are 27 of these in the data.

ITEMS: How many items were prescribed.

COST: The combined cost for all those items.

In a sentence we could describe this dataset as:

- Per English NHS ICB and for December 2024, the volume and cost of each antidepressant drug prescribed.

For Challenge B

We have merged, simplified and filtered all these monthly files into a single dataset.

'STEP_UP_REGIONAL_ANTIDEPRESSANTS.Rds'

This dataset now contains 6 columns:

***YM*:** The year and month, in the format YYYYMM, where 202401 is the same as January 2024. There are 46 year-month values in the data.

***YEAR*:** The year in the format YYYY. There are 4 years-worth of prescribing information in the dataset.

***REGION*:** The NHS Region. There are 7 regions in the data.

***DRUG*:** The name of the anti-depressant medicine. There are 32 of these in the data.

***ITEMS*:** How many items were prescribed.

***COST*:** The combined cost for all those items.

In a sentence we could describe this dataset as:

- Per English NHS Region and per year-month, the volume and cost of each antidepressant drug prescribed.

For Challenge C

We have merged, simplified and filtered all these monthly files into a single dataset.

'BSA_ODP_PCA_REGIONAL_SUMMARY.csv'

The data set matches that used in the R pathway.

Your Job

Your job is to examine and analyse the data provided to:

- Understand national and regional prescribing volumes and costs
- Understand national and regional prescribing trends
- Understand monthly and annual trends
- And finally, maybe even predict future monthly anti-depressant prescribing volumes

We want to address questions like:

- What are the top 10 prescribed anti-depressants nationally?
- How many anti-depressant items were prescribed nationally?
- What are the contrasts between antidepressant prescribing volumes and costs?

Below is a list of tasks you would typically go through to analyse the prescribing data

- If you are completing this challenge in R or Python, please refer to the provided .html file for a rundown of the tasks required to analyse the data!

- If you are completing this challenge in Excel, please refer to the excel_cheatsheet.ppt for hints and tips on how to complete the tasks in Excel!

Challenge A: Step by Step Guide

1. Open the file in Excel
2. Examine the data

Exercise 1 Transforming and Aggregating Data

Use the functions available in Excel to aggregate the supplied data to answer the below questions:

Question 1: Calculate the nationally top 10 prescribed anti-depressants sorted from biggest from smallest.

Question 2: Calculate the national cost of Mirtazapine prescribing

Question 3: Calculate what is the spend of Sertraline hydrochloride prescribing in the Midlands region?

Question 4: Calculate how many anti-depressant items were prescribed nationally?

Exercise 2 Visualising Data

Creating plots/visualisations allows us to identify patterns in the data quickly.

Try to create the following plots:

Question 1: Create a horizontal bar chart of the top 5 most prescribed drugs in December 2024, arranged in order.

Question 2: Create a vertical bar chart showing the total cost of Sertraline hydrochloride prescribing in the Northeast region.

Question 3: Create two vertical bar charts for comparison. First, create a bar chart showing the total antidepressant prescribing (items) by type of antidepressant nationally. Second, create a bar chart that shows the total antidepressant prescribing cost by type of antidepressant nationally. Experiment with different ordering of items in your chart, largest to smallest, alphabetically, etc Which works best visually, and which works best for comparison?

Question 4: Create two pivot tables for comparison. First, create a pivot table showing the total antidepressant prescribing per region. Second, create a pivot table showing the antidepressant prescribing cost per region.

Question 5: Create two horizontal bar charts for comparison. First, create an ordered bar chart showing the 10 most prescribed antidepressants (items). Second, create an ordered bar chart that shows antidepressants with the greatest total prescribing cost.

Exercise 3 Data Metrics and Insight

It's now time to start working on our presentations. An important requirement for a data role is to be able to present your work. Create a PowerPoint presentation to present your analysis and results to the key stakeholders.

You can use any of the charts you have created so far or include any additional you have chosen to create from the data.

Your challenge was to explore drug usage to build awareness and inform decision making using NHS data to improve patient care. You need to articulate to a non-technical audience what the data is telling us about antidepressant prescribing in England.

- Start by describing what each chart is telling us and compare the charts against each other.
- Describe some regional variation and contrasts.
- By now you may have noticed there are contrasts between antidepressant prescribing volumes and costs, explore the implications of this with your intended audience.
- Highlight and describe any differences between the charts.

Try to be concise with your presentation; Health Care professionals are often short on time so identifying key messages you want them to take on board is key, as is using language that is accessible to all.

YOU REACHED THE END OF CHALLENGE A! CONGRATULATIONS!

We hope you learned something, and we have inspired you to pursue an analytical role in the NHS

Challenge B: Step by Step Guide

For the R coding version of the challenge follow along with the R Step by Step guide and follow along with the exercises, [dig-data-step-up/README.md at main · nhsbsa-data-analytics/dig-data-step-up · GitHub](#)

Challenge C: Step by Step Guide

For the Python coding version of the challenge follow along with the Python Step by Step guide and follow along with the exercises, [Digdata/README.md at main · nhsengland/Digdata · GitHub](#)

Completed your challenge Path

Once complete please send your work to info@digdata.online stating your full name, university year group, and university name and you will receive your virtual work experience certificate within 1 month.