

Data Analytics with SAS UK

A guide to delivery



Introduction

This project has been co-developed with the support of SAS UK and JMP, and provides free downloadable industry-standard data analytics software to all TechFuture Classroom schools.

When you decide to use this project, you can download the software from the link on the Resources section of the course page and then email sue@thetechpartnership.com for the licence code.

There are two parts to this project. In the first, students have to solve a problem using large datasets and the JMP software. The problem focuses on an illness outbreak in school. They have to identify the source of the illness, and what it is. To do this, they carry out a set of tasks based upon analysis of the dataset provided. Everything they need for completion of this task is provided in the e-learning module.

The second task is more challenging. Here the students use a second dataset with JMP to identify the correct medication that will help patients with the disease. They have to recognise the issues around the mutation of the virus and its association with blood group types. *There is cross-curricular relevance of this project with the science curriculum, specifically biology.*

The e-learning module has been created for students to achieve a set of six awards as they work through the videos and activities. These awards are:

- > The **JMP Expert** award for getting started with JMP
- > The **Data Analyser** award for understanding how to use the dataset to identify relative numbers of male and female students who are ill
- > The **Data Comparer** award for identifying which year group is most affected by the illness.
- > The **Data Explorer** award for further exploration of the data to identify more factors that help to identify the cause of the illness
- > The **Creative Presenter** award for using graphs effectively to identify trends in the data
- > The **Investigator** award for accurately identifying the cause of the illness.

The videos provided in the e-learning module use the Titanic passenger list to demonstrate how to use JMP. More specific support, using the actual datasets that have been created for the main challenge, is available through downloadable PDF documents that explain how to use JMP to solve the problems along the way. These are available both inside the e-learning module and on the course page under Resources.

Brought to you by



When all the main activities using JMP are completed, students access the SAS Challenge document which uses a second dataset to identify which are the best medicines to deal with an outbreak of the illness (guinea pig flu).

By the end of the e-learning module, students will have solved the problem and identified what the illness was that caused so many students to become ill, and how it started. There is an Open Badge available, automatically awarded, to all students who complete the e-learning module. See below for more information.

Delivery timings

The first part of his project can be delivered in the classroom over two 50 minute lessons (approx.), when using the JMP software and the e-learning module. By the end of this, students will have gained skills in using JMP and have solved the first challenge – which illness is causing the problem.

Students need access to the JMP software and the e-learning, preferably individually but it is also possible to complete this project in small groups. Individual access, though, is required for students to achieve the open badge (see below).

The second part of the challenge, where they have to identify the most effective medicines to deal with another outbreak will take an additional two lessons. This is a task where JMP is used again with the second dataset, and requires students to make decisions about which aspects of JMP will help them to solve it. This could be used as an extension task.

Provided datasets

There are three downloadable datasets in the Resources area of the site. These are in a zipped folder, which needs to be unzipped to access the datasets. **Note, these are specifically built for JMP and will not work in other software.**

Students need the *Outbreak Data.jmp* dataset for the first task.

The *Extended Outbreak Data.jmp* is needed for the task where students solve the SAS Challenge. They also need the *Greenhill High nurses report.jmp* for additional information about symptomology and timings.

Gaining badges in the Tech Partnership Badge Academy

There are two badges for students to earn as a result of completing this project. The first badge is automatically awarded for completion of the e-learning module. This is the **SAS Data Analyst** badge. The badges can be seen on the course page.

The second badge is a teacher-verified badge. This is the **SAS Data Analyst Expert** badge. Here you will be able to issue a code to students when you feel they have completed the larger task, to use a new dataset independently or in groups to solve the challenge of which medicines should be prescribed to treat the illness outbreak.



Steps in the e-learning content

The following table shows the steps in the e-learning and where assessment points and additional resources are available.

Steps in the e-learning content	Resources
Students watch a video that explains big data and how SAS UK uses it for predictive analytics	
From the ' What is JMP? ' icon students find out what the JMP software is with the first of a set of video tutorials	JMP download
Tech Quiz – students' understanding of JMP is assessed with the first quiz. <i>First award is gained for successful completion of the quiz.</i>	
From the ' Analysing data ' tab students find out what the problem is they have to solve. The school is on lockdown because a virus is spreading. They have to find the cause. Three more icons allow students to explore possible causes – food poisoning, writer's rash and guinea pig flu.	Downloadable datasets for students to work with in the JMP software
' Who is sick? ' screen has two videos that explain how to take the datasets and analyse them to see if girls or boys are more likely to become ill. Students create distribution graphs and tables to identify numbers of male and female students who are ill.	Data Analyser document to support, downloadable from both the e-learning module and the course page Resources section.
Tech Quiz – students' understanding of distribution graphs and tables is tested in this second quiz. <i>Second award is gained for successful completion of the quiz.</i>	
From the ' Comparing Data ' icon on the menu screen, students now go to the next task where they compare data from the datasets to identify which year group is most affected by the illness.	
' Which Year Group? ' screen has another video to identify more about the data, using a graph to compare year groups.	Compare data document downloadable from the e-learning module and the Resource section of the course page.
Tech Quiz – students' understanding of distribution graphs is assessed with this third quiz. <i>The third award is gained for successful achievement of this quiz.</i>	
From the ' Exploring Data ' icon on the menu screen, students are taken to two more videos to identify the possible cause on the ' Who do you suspect? ' screen. In these videos, they learn about filtering the data and creating more graphs to visually represent what is happening in Year 8.	Exploring data document downloadable from the e-learning module and the Resource section of the course page.
Tech Quiz – students' understanding of the deeper exploration of the Year 8 data is assessed with this fourth quiz. <i>The fourth award is gained for successful achievement of this quiz.</i>	
From the Graphical Creations icon on the menu screen, students are taken to the ' Can you solve it? ' screen where	Graphical creations document downloadable from the e-learning module and the



Steps in the e-learning content	Resources
they watch another video that explains how to create a scatterplot matrix to pick up links in the data set.	Resource section of the course page.
Tech Quiz – students' understanding of the scatterplot matrix and what it shows is assessed with this fifth quiz. <i>The fifth award is gained for successful achievement of this quiz.</i>	
Investigation recap goes back over the information found so far. They are given a Disease Guide as a reminder of the three possible diseases that could be responsible for the illness.	Disease Guide document downloadable from the e-learning module and the Resource section of the course page.
Tech Quiz – this final assessment checks that students have identified the correct illness (guinea pig flu) as a result of their data analysis. <i>The sixth award is gained for successful achievement of this quiz.</i>	

The SAS Challenge

When students have completed the e-learning, the SAS Challenge opens up and they can engage with further use of JMP with a new dataset, this time working out which is the best medicine to recommend to deal with a further outbreak of the virus. This requires a more independent approach but is supported with a guide that is downloadable from the e-learning module, and an information chart downloadable from the Resources area.

Students use the additional dataset in JMP and make decisions about the analysis to find out the best medicines, and their quantity, to be sent to a hospital to deal with a disease outbreak. This requires understanding the link between guinea pig flu and mutation with blood proteins A and B – which causes more severe symptoms in those patients with blood groups A, B and AB – and the Rhesus D antigen (which makes a blood group either positive or negative). Those who are Rhesus D positive demonstrate secondary symptoms after 55 to 65 minutes, whilst Rhesus D negative patients demonstrate secondary symptoms after a few minutes (up to 10 minutes). These pieces of information are needed to solve this challenge.

The answers for this can be found in the Resources area in the document SAS Challenge Solution. This should be for teacher use only!

This task could be used as an extension task.

If you have questions or queries about this project, contact sue@thetechpartnership.com and we will do our best to respond within 48 hours.