Preparing data

The fundamentals of data and spreadsheets in Google Sheets

2 WEEKS **45** HOURS **11** LESSONS

In this module, we focus on developing **essential data** and **spreadsheet skills**, enabling us to **effectively work with data** to **draw meaningful insights** and **make informed decisions**. By exploring data in **Google Sheets**, we will learn how to **retrieve, manipulate, analyse**, and **visualise data**.

We'll use **real-world data** and **examples** related to the **United Nations Sustainable Development Goals** to contextualise the concepts and demonstrate the practical application of using spreadsheets to solve problems.

Through **hands-on exercises, case studies**, and **a project**, we will gain **proficiency in spreadsheet operations** and develop a **strong foundation for data-driven analysis** and **decision-making**.

Module objectives

Introduction to spreadsheets

Master the **foundations** of data and spreadsheets. Learn how to **work with spreadsheets effectively**. Gain a **fundamental understanding** of **data visualisation**.

Data manipulation

Master the fundamentals of **manipulating** and **analysing** to **extract meaningful insights**. Learn how to identify and rectify data anomalies.

Introduction to statistics

Master the fundamentals of statistics in the context of **spreadsheet analysis**. Learn how to **summarise data**. Gain a fundamental understanding of **samples** and **distributions**.

Introduction to data modelling

Master **spreadsheet functions** and **control flow queries**. Learn how to **identify patterns** in data. Gain a fundamental understanding of **model accuracy** and **testing assumptions**.

Learning activities

By engaging with different types of learning activities, you will develop a **deeper understanding** of preparing data in Google Sheets and build a range of skills that will help you succeed in your **coursework** and **beyond**.

We learn by doing. We'll work on practical problem-solving and real-world projects.

Learn Watch animated videos and read practical slide decks to learn spreadsheet concepts and operations.

- **16** Animated videos
- **22** Slide decks
- **22** Reference cards

Apply Practise spreadsheet operations with step-by-step guides and apply them to real-world scenarios.

- **53** Walk-throughs
- **2** Integrated project

This integrated project spans the entire module. Each week, we'll delve into a specific part of the project and you'll have to complete corresponding multiple-choice questions (MCQs) based on that week's learning. In this way, we build on our spreadsheet skills cumulatively!

Assess Test and track your data, spreadsheet, and problem solving proficiency.

- **38** KQ assessments
- **3** MCQ assessments

Preparing data

The fundamentals of data and spreadsheets in Google Sheets

Week 1

Lesson: Data sources and access

In this lesson, we explore ways to **source the right data** and how to take **accountability** for it through data governance principles and ethics. We also look at **spreadsheets**, and more specifically **Google Sheets**, as a tool we can use to **access our data**.

- Understand how and why **sourcing appropriate data** is necessary to be able to create business value.
- Know how to **navigate Google Sheets** and **access data** through imports.
- Understand the various **principles** and **ethics** that **govern the collection and use of data**.

Lesson: An introduction to using data

In this lesson, we will be **introduced to using data in spreadsheets** as we look into **why** it is necessary. By the end of this lesson, you will be able to identify the different **data types** in spreadsheets, know when and how to make data visible, as well as know how to use a few **functions** in spreadsheets.

- List the different **types of data** and be able to describe them.
- Know how to **select and format data** in Google Sheets.
- Know how to **use spreadsheet functions** and **formulas** to do row and column calculations in Google Sheets.

Lesson: Data aggregations and descriptive statistics

In this lesson, we will learn about the different kinds of **descriptive statistics** and how to calculate them. By the end of this lesson, you will be able to **calculate central tendency** and spread as well as **use pivot tables** to summarise data.

- Understand what **descriptive statistics** are and be able to list and describe them.
- Know how to **calculate measures of central tendency and spread**.
- Know how to create and use **pivot tables** in Google Sheets to summarise data.

Lesson: An introduction to data visualisation

In this lesson, we'll take a look at **what data visualisation is** and how to choose the most **appropriate** visualisation based on our data and the story we want to tell. By the end of this lesson, you will know how to **choose appropriate visualisations**, create them in Google Sheets, and be able to **interpret** and **analyse** them.

- Understand what **data visualisation** is and why it's important.
- List and describe the **different data visualisation categories**.
- Know how to **create data visualisations in Google Sheets**.

Integrated project: Understanding the data

In this first part of the integrated project, we **investigate access to safe and affordable drinking water**. You will need to **create features**, **summaries**, and **visualisations** as per the instructions in order to **complete the compulsory case study MCQ assessment**.

- Demonstrate an understanding of **importing data** and **using formulas, functions**, and **charts** in Google Sheets.
- Demonstrate an understanding of **calculating summary statistics**, including the measures of central tendency and spread.
- Demonstrate the ability to **interpret data, summary statistics**, and **data visualisation**.

Lesson: **Data formatting**

In this lesson, we'll learn about data formatting and discover **why it is crucial to format data** in order to **guarantee** and **maintain data integrity**. Additionally, we'll learn about the **typical data types** and **structural problems** we encounter when dealing with spreadsheets.

- Understand the **importance of data quality** and **formatting**.
- Understand the various **numerical, textual, datetime**, and **conditional formatting** tools available for spreadsheets and how to use them.

Lesson: **Data cleaning**

In this lesson, we are introduced to the **concepts of data cleaning** and **data integrity** and the crucial role they play in **enhancing** and **maintaining data quality**. We will also see how we can implement these concepts practically in Google Sheets.

- Understand the concept of **data cleaning** and how it **contributes to accurate** and **reliable data analysis**.
- Know about **common data issues** that can arise and the data **cleaning techniques** that can be used to address them.
- Understand the role that **data integrity** plays is **maintaining the quality of data**.

Week 2

Lesson: **Samples and distributions**

In this lesson, we will introduce the concept of **probability** and delve into how to **apply the empirical probability** approach to data. We will explore the concept of **sample and population data** and how sample sizes affect **confidence intervals** and the **margin of error**.

- Understand the basic concepts and terminology of **probability theory**.
- Be familiar with the **empirical probability approach and how to apply** it to data in Google Sheets.
- Understand the concept of **sample and population data** and how increasing the **sample size** affects the **confidence interval** and **margin of error**.

Lesson: **Spreadsheet functions**

In this lesson, we'll cover a **wide range of spreadsheet function**s. By the end of this lesson, you will be able to employ techniques such as VLOOKUP, regular expressions, and logical functions to sort, filter, and format data to make it more **accessible** and **meaningful**.

- Understand the **importance of spreadsheet functions** and how they can be leveraged to solve problems.
- Understand **logic functions** and how to use them to solve control flow problems.
- Identify and define various spreadsheet functions that can be used to **analyse text data**, including **regular expressions**, and know how to use them.

Integrated project: **Transforming the data**

In this second part of the integrated project, we extend our **investigation on access to safe and affordable drinking water**. You will need to c**reate features**, **summaries**, and **visualisations** as per the instructions in order to **complete the compulsory case study MCQ assessment**.

- Demonstrate the ability to use **formulas** and **charts** in Google Sheets.
- Demonstrate the ability to **interpret data, summary statistics**, and **data visualisation**.
- Demonstrate the ability to **use spreadsheet functions** to **transform** and **clean** data.

Lesson: **Identifying patterns**

In this lesson, we'll explore the **line of best fit, correlation coefficients, R-squared values** and **heatmaps**. By the end of this lesson, we will be able to use these tools to **interpret relationships, identify data patterns** and **make predictions**.

- Understand the concept of the **line of best fit** and its role in **identifying trends** and **relationships** between variables.
- Evaluate correlation coefficients to **determine the strength of the linear relationship** between two variables and **interpret the R-squared value** to evaluate goodness of fit.
- Utilise scatter plots, conditional formatting, and heatmaps to **visualise relationships** between variables, **identify patterns**, and **evaluate the line of best fit**.

Lesson: **Accuracy**

In this lesson, we'll learn about **model accuracy metrics** and how to use these to **assess our models**. We'll uncover common challenges to model accuracy, such as bias, variance, overfitting, and underfitting, and explore strategies to address these obstacles. Finally, we will learn how to use **polynomial trend lines** to find the best fit for our data. By the end of this lesson, you will be able to assess the accuracy of models and apply techniques to improve the fit of models to our data.

- Know how to **calculate** and **interpret residuals, residual plots**, and **accuracy metrics** for evaluating model performance.
- Understand **common challenges to model accuracy**, including **bias, variance, overfitting**, and **underfitting**, and learn **strategies to address them**.
- Understand and **evaluate polynomial trend lines** and explore how this relates to **overfitting** and **underfitting**.

Lesson: **Drawing and testing assumptions**

In this lesson, we will explore the **fundamental concepts of hypothesis testing** and its importance in statistical analysis. By the end of this lesson, you will be able to **formulate hypotheses** and test their **validity** using various statistical tests.

- Know how to **formulate** the **null** and **alternative hypotheses** and how **statistical significance** and **level of significance** can influence the results of a hypothesis test.
- Know how to **calculate** the **test statistic** and **critical value** for a **t-test, z-test**, and **Kolmogorov-Smirnov** test in Google Sheets, and how to use these metrics to reject or fail to reject a null hypothesis.
- Know the different **types of errors** in hypothesis testing.

MCQ: **Preparing data**

Test your knowledge of **the line of best, the line of best fit equation**, and **drawing and testing assumptions**.

- Demonstrate an understanding of **trend and relationships** in data and **interpreting** and **evaluating** model performance.
- Demonstrate an understanding of **hypothesis testing**.

Module summary

Throughout this module, we've embarked on a journey to **cultivate essential data and spreadsheet skills**. We've armed ourselves with the **expertise needed to efficiently work with data** and **derive meaningful insights**. Our exploration focused on spreadsheets, delving into data retrieval, manipulation, analysis, and visualisation to unlock their true potential.

By completing this module, we've not only become **proficient in spreadsheet operations** but also laid a **strong foundation** for **data-driven analysis** and **decision-making** that will be key in our data career path.

What's next?

We'll take these foundational skills forward as we **explore other data tools and technologies**, and expand on our capabilities in **data-driven analysis and decision-making**.

Remember, the data landscape is vast and ever-evolving. **Stay curious, keep practising**, and **apply the knowledge and skills you've gained in real-world scenarios**. Whether you're aiming to shape policies for sustainable development, streamline business strategies, or drive innovation, **your proficiency in handling data will always be a priceless asset**.