Loading workbooks

Let's get you started! We will provide you with all the necessary workbooks to follow along on your local computer. If you do not have Excel already installed, you can download it **here**.

Before loading data and creating basic formulas, you must learn how to load existing workbooks. If you lose progress or get stuck at any point in this course, you can load the corresponding spreadsheet to re-initialize.

How to do this exercise locally

Instructions
100 XP

Open the workbook called `1_1_loading_workbooks.xlsx` from the Workbooks folder.

You can enlarge the Excel interface by clicking on the full-screen button in the bottom right corner or by zooming out.

On the worksheet `Google Monthly Stock Price`, select all values in column `A`.

How many rows are contained in column A including the header?

- 121
- 120
- 5184804
- 43206.7

Correct! The count of all rows is 121, with the first being the data header row and the rest being the actual data points.

Working with data

Now that you know how to load existing workbooks and highlight values in a column, it's time to practice adding in data manually and by copying/pasting from another Excel file.

If you lost progress, close any open reports and load `1_2_working_with_data.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

In this exercise, we'll be creating new worksheets within our workbook. Create two worksheets, "Adding data" and "Pasting data".

To rename your sheet, double-click it and enter your new description.

In the sheet `Adding data`, manually input the below data. This data will span three columns and four rows. We recommend starting from the cell reference `A1`.

Name	Location	Orders Placed
Alexis	Belgium	52
Ciaran	Ireland	17
Miles	Canada	37

- Navigate to the sheet you created earlier, `Pasting data`.
- Open the CSV file `copying_data.xlsx` from the *Datasources* folder. **Do not close your Excel workbook.**
- Copy all data from columns `A` to `K` from the other Excel file and paste as values into your workbook.

The data looks messy; let's do some formatting to make the data easier to read.

- Set the column width for all columns to 15.
- Ensure all column headers begin with a capital letter and the rest of the column name is lowercase.

What was the runtime for The Shawshank Redemption in minutes?

142

Great work! As of June 2022, The Shawshank Redemption was ranked number 1 on IMDB with a rating of 9.3. The total film run time was 142 minutes (2 hours and 22 minutes).

Creating first formula

Great work so far - we've got ourselves more familiar with the user interface of Excel and even added data to our workbook. In this exercise, we'll be carrying out some basic arithmetic operations on our Google Stock Price data.

If you lost progress, close any open reports and load `1_3_creating_your_first_formula.xlsx` from the Workbooks folder.

How to do this exercise locally

If you do not already have it selected, navigate to the `Google Monthly Stock Price` sheet.

- Create a new column with the header "Open vs Close".
- In the second row, subtract the value in the `Open` column from the value in the `Close` column.

Copy the cell references from the second row of the `Open vs Close` column to row 121.

Set the column width for all columns to `13`.

Highlight all values in the column `Open vs Close`.

What was the average difference in open vs close price?
0.621333333

That's right! Over a 10 year period, the average difference in close price vs open price was only 0.62. Next, you'll be learning more about working with your data and how to utilize Excel's powerful built-in table features.

Working with structured cell references

Adventure Works is a global retail brand selling bikes, bike parts, and accessories. You'll support their management team as an Analyst to help them better understand their sales data.

In this exercise, we are working with their Sales table which contains all orders from 2017 to 2020. This will enable us start creating new calculated columns and using structured references.

If you lost progress, close any open reports and load `1_4_working_with_structured_references.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Open `1_4_working_with_structured_references.xlsx` from the Workbooks folder.

Create a new column "OrderToDelivery" after the delivery due date column that calculates the number of days between order and delivery for each row.

Use a structured cell reference (the column name instead of the column reference). Remember, to refer to a column name you'll need to use the `@` sign.

That doesn't look quite right! It's defaulted to a date… let's change it by changing the data type to "Number".

How many days on average does it take for Adventure Works to deliver once a customer has placed an order?

That's right! Regardless of geography, on average it takes Adventure Works six days to deliver any items ordered through their website.

Filtering data

So far, you have learned how to work with Excel, create formulas, create calculated columns, and use structured cell references. These concepts will come in handy for the rest of this course.

In this exercise, we'll filter and sort our data in various ways to help Adventure Works better understand its sales throughout the years.

If you lost progress, close any open reports and load `1_5_filtering_data.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Filter the `Sales` table to display orders from 2020 only.

Now it's time to sort your orders by two columns in the following priority:

1. `CustomerCountry` (A-Z)
2. `OrderDate` (Oldest to Newest)

What is the order number for the first visible row in the table after the data has been filtered and sorted?

- SO63306
- SO75061
- SO63294
- SO43697

That's right! We can see that the first order in Australia for 2020 had the order number SO63294 and was made by our customer Carmen Sara.

Formatting tables

Now that we've familiarized ourselves with our data, we will apply rules to our columns for data integrity. By default, most of our columns will default to the format "General". In this exercise, we'll apply some column formatting and change the table styles.

If you lost progress, close any open reports and load `1_6_formatting_tables.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- Remove any previous filters applied from the last exercise, we want to see all orders.
- Change the sort to only sort by `OrderNo` ascending.

Format the following columns as *Currency*: `ItemCost` and `ItemPrice`.

Format the following columns as *Number* with 0 decimal places: `SalesOrderLineKey`, `OrderQuantity`, and `OrderToDelivery`.

Format all other columns as as *Text*.

Maybe it's time we change up the table style too! Navigate to *Table Design* and select any table style you like.

How many columns are formatted as text in our Sales table?

- 2
- 10
- 3

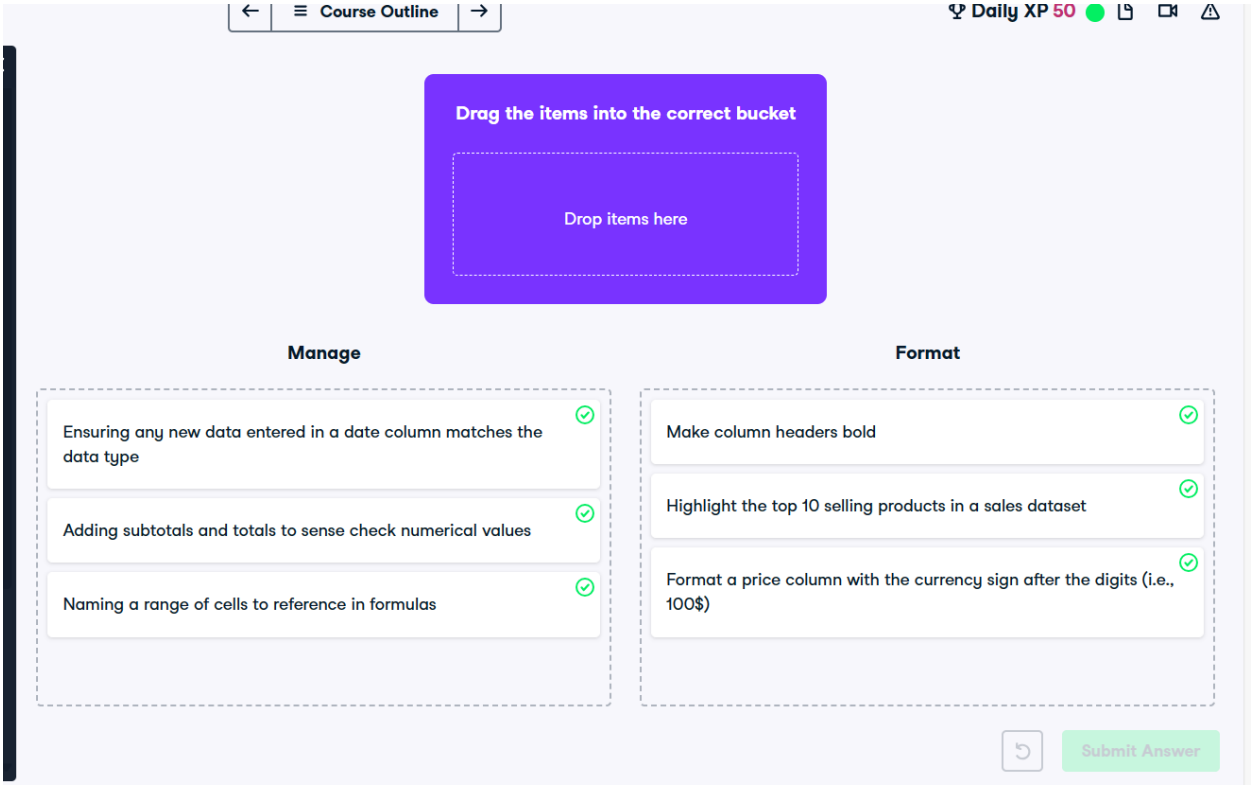
Perfect! We've just formatted 10 columns as Text, these are important dimensions that we can use to analyze our data.

Manage vs. Format

You have been introduced to some new features in Excel that help you with your data preparation tasks - the ability to manage data better and carry out more advanced data formatting. Now it's time to identify which tasks in Excel relate to which feature.

Instructions
100XP

Classify each description based on which data preparation process it applies to.



Congratulations! Making the most of Excel's data management and additional formatting features can help you take your data preparation to the next level, ensuring that your dataset is as clean and organized as possible for any subsequent analysis or visualization. Let's now see some of these features in action!

Naming ranges

Naming specific cell ranges within your data can be extremely useful. If there is a particular column or row of information that you will use regularly for formulas or calculations throughout your workbook, it can definitely be easier and save time if you have given that particular range an easy reference to use.

We've reverted our dataset to before we applied our table styles, we'll now add some names to some important ranges within our data.

How to do this exercise locally

Instructions

100 XP

Open `2_1_naming_ranges.xlsx` from the Workbooks folder.

Numerical variables, like `ItemPrice`, are more likely to be used in future calculations and formulas.

- Select cells with values in them from the `ItemPrice` column.

Name the range "item_price" by using the _Define Name_ feature in Excel. Make sure that the name refers to the data from cell `E$2` to `E$60399`. This will ensure the header isn't included in the range.

Repeat the range naming process for the `OrderQuantity` column, using the the name "quantity". The range you want to be naming is `C2:C60399`.

- In column `R`, create a new header called "OrderLinePrice".
- In cell `R2`, create a new formula that multiples our two ranges: `item_price` and `quantity`.

What is the average Order Line price across the entire dataset? Round to 2 decimal places.

520.32

Nicely done! Similar to how you would create a calculated column in a table, you've applied the same logic here without the need for a formatted table.

Totalling up

Carrying out quick calculations of numerical values within our dataset is a great way to sense-check that things are in order and ready for much deeper analysis later on. We don't even need to start creating formulas by hand - Excel provides us with features that make it easy for us to check our data.

The management team would like to understand whether there is a significant difference between each country's order to the delivery period. In this exercise, we'll use Sub-Totals to work this out.

If you lost progress, close any open reports and load `2_2_totalling_up.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Remove the `OrderLinePrice` column you created in the previous exercise.

Duplicate the worksheet and rename it "Sales Subtotals".

We want to use the *Subtotal* feature in Excel, but before we do this we will need to sort our data.

- Add filter options to each column.
- Sort the `CustomerCountry` by *Ascending* order.

Now let's add a subtotal row. Using the *Subtotal* feature, configure the new row to apply for each change in the `CustomerCountry` and to show the average `OrderToDelivery` time for each group of values in the `CustomerCountry`. Include a summary at the bottom of the dataset.

On the left-hand side of the page we should see the numbers `1`,`2`,`3` next to each other. This indicates that some new groupings have been added to the page and shows you the levels you can expand the dataset by. If you click on the `2` it should collapse the data so that we only see the average order to delivery time for each country.

What is the average order to delivery time for customers in Germany? Round to 2 decimal places.

6.08

Awesome! That is actually the highest average order to delivery time across the six different countries that Adventure Works sell to. This has enabled us to sense check out data, and we can see that regardless of country we aim to delivery to our customers within 6 days of their initial order.

Custom formats and validation

Excel allows users to format the data within sheets in many different ways. It even allows us to create our own custom formats if we can't find the type we want from Excels prepared formats. As well as formatting the data, we can also ensure the integrity of the data within columns and rows. This helps when sharing workbooks with others, or when required to enter new data that might be needed into existing columns or rows.

If you lost progress, close any open reports and load `2_3_custom_formats_and_validation.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- We no longer require our `Sales Subtotals` worksheet, so delete this from your workbook.
- Format all the data in your `Sales` worksheet as a table.

In our sheet, we have two columns that represent monetary values. But they are not formatted this way. Let's fix that.

- Select `ItemCost` and navigate to the custom format window.

We want our `ItemCost` values to be shown to zero decimal places, with a comma as the thousands separator, and the currency (dollar) sign shown after the value. For example, something like `1,000 $`.

- Go to the *Custom* section within the custom format window.
- In the *Type* section, using "#" to represent the numbers of our value, set the format type we desire along with the currency sign.

That looks much better. Do the same for the `ItemPrice` values, using the same type of custom format.

In the future we will be sharing the file with others and they may add new rows to the dataset. We want to ensure that we have applied some validation to ensure that `OrderQuantity` has the value of at least 1.

- Select the column `OrderQuantity` and navigate to the *Data Validation* settings menu.

Using the *Data Validation* feature, create the following criteria for the `OrderQuantity` column:

- Only allow whole numbers greater than the value `0`.
- Display an error message that informs the user if the value they entered does not meet the column criteria.

If you entered the value "0" in cell C60400, would you get an error message - Yes or No?

Yes!

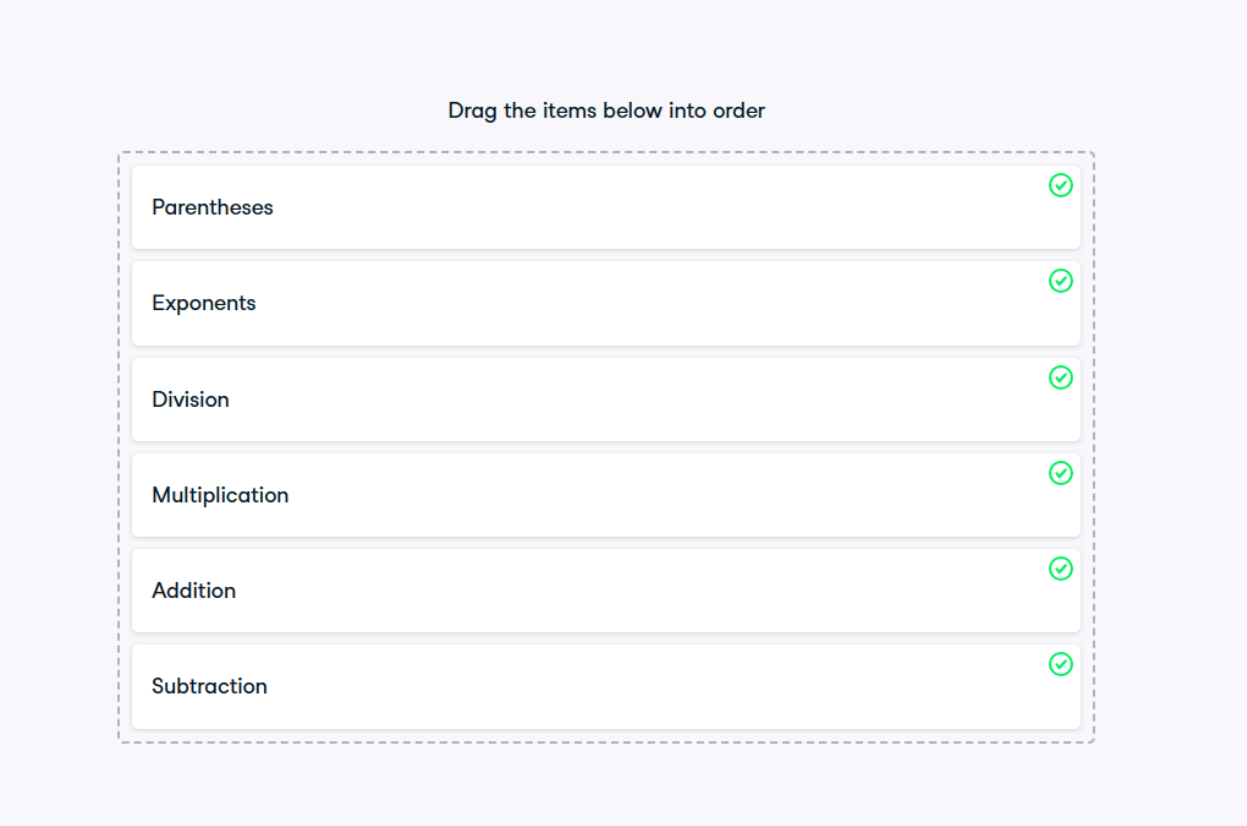
Exactly! You cannot have an order quantity set to 0. Therefore, with the data validation rule we have set we can see the error message appears, and we can only enter values that meet the criteria for the column. This feature is a great way of ensuring that after we have prepared our data, if new data is entered it will fit our required criteria and will therefore not require extra check or data cleaning steps.

Order of operations

In math as in Excel, the order of operations tells you the correct sequence of steps needed to evaluate your expression.

Instructions
100XP

Organize the order in the correct sequence, according to how Excel handles mathematical operations.



Congratulations! Not only have we refreshed our high school math knowledge, we also have a good understanding of the way in which Excel processes are formulas.

Summarizing sales

By now you'll be familiar with the general structure of a formula and will have the capabilities to perform basic arithmetic operations.

In this exercise, we'll start working with aggregate functions and summarize key metrics on our data including the minimum, maximum and average price of our items.

If you lost progress, close any open reports and load `2_4_summarizing_sales.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Write the following list of descriptions in the following cells:

- `S2`: "Lowest Item Price"
- `S3`: "Highest Item Price"

- `S4`: "Average Item Price"
- `S5`: "Total Sales"
- `S6`: "Total Profit"
- `S7`: "Profit Margin Ratio"

In cell `T2`, calculate the lowest item price across all sales.

In cell `T3`, calculate the highest item price across all sales.

In cell `T4`, calculate the average item price across all sales.

Format all three cells you've created to have the dollar currency to 2 decimal places.

What is the average item price for all sales at Adventure Works?

- $486.09
- $3,578.27
- $2.29

That's right! For all sales from 2017 to 2020, our average item price was $486.09. We can take this a step further in future exercises!

Calculating total sales

In the previous exercise, we started by calculating some basic statistics on our data.

Each row of our data indicates an line item in a given invoice, so to calculate total sales we need to understand the total for each row by looking at quantity multiplied by the item price. This will give us a better understanding of our total sales.

If you lost progress, close any open reports and load `2_5_calculating_total_sales.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- Create a new column between `E` and `F` called "OrderLineSales".
- Order line sales can be calculated as the multiplication of order quantity and item price.

In cell `U5`, calculate the total sales for all rows based on our newly created column `OrderLineSales`.

What is the total amount of sales the company has generated?

31,426,180 $

Great work! In 4 years Adventure Works has generated over 31 million dollars in revenue. In the next exercise, we'll be using the columns you've created to calculate the profit margin ratio.

Profit margin ratio

Great work so far! We're going to take it up a new level by combining two calculations using parenthesis and calculating our first percentage value. The output that we calculate here will help Adventure Works better understand their total profits and profit margin.

If you lost progress, close any open reports and load `2_6_profit_margin_ratio.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- Create a new column between `E` and `F` called "OrderLineCost".
- Order line cost can be calculated as the multiplication of order quantity and item cost.

- Create a new column between `G` and `H` called "OrderLineProfit".
- Order line profit can be calculated as `OrderLineSales` subtract `OrderLineCost`.

In cell `W6`, calculate the total profit for all rows based on our newly created column `OrderLineProfit`.

- In cell `W7`, calculate the profit margin ratio which divides total profit by total sales.
- Format this value as a percentage to one decimal place.

What is the profit margin ratio for all sales?

42.1%

Great job! Adventure Works has a really high profit margin ratio, this is great news for the leadership team.

To the left, to the left

What does the following Excel formula return?

```
=LEFT("DataCamp Rocks", 4)
```

Answer the question

50XP

Possible Answers

Select one answer

- Rocks

  `PRESS1`

- DataCamp

  `PRESS2`

- Data

  `PRESS3`

- ocks

  `PRESS4`

That's right! In the next few exercises we'll be utilizing various functions to format, clean and summarize our data.

Product profiles

Keep up the good work! Data doesn't always come formatted as we require, so here, we'll focus on formatting a new column by applying a function to an existing column. Additionally, we'll create a unique identifier that can be used for future analysis.

If you lost progress, close any open reports and load `3_1_product_profiles.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- Create a new column between `O` and `P` called "Country".
- Write a formula that converts the customer country into uppercase.

- Wait… why is it only showing the formula? Well that's because Excel thinks your formula is text.
- Convert the column to the data type *General*.

- Create a new column between `A` and `B` called "OrderID".
- Write a formula that takes the last 5 characters from `OrderNo`.
- Make sure that the column data type is *General*.

For OrderID: 43697, what is the total Order Line Profit? Answer to 2 decimal places

1,406.98 $

That's right! On that one order AdventureWorks made almost 1500 dollars profit. The subsequent orders also appeared to make a good profit compared to the average profit which is 219 dollars. That sounds like a huge win!

Profit performance

Date functions can be powerful when preparing and transforming data ready for analysis, so we'll utilize functions such as `YEAR` and `MONTH` in this exercise.

If you lost progress, close any open reports and load `3_2_profit_performance.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

- Create a new column "OrderYear" next to `OrderDate` that returns the year of the order.
- Make sure that the column data type is *Number* and formatted to `0` decimal places.

Create a new column "OrderMonth" next to `OrderYear` that returns the month of an order.

What is the average order line profit for 2020 orders?

147 $

That's right! The average order line profit for orders placed in 2020 were 147 $ which is below the average we've seen across all four years. It's something the management team should continue to monitor to see if any actions need to happen as a result.

Formatting with functions

Finally, we'll finish this chapter by applying some rounding functions to format our cells and ensure they are easy to read.

If you lost progress, close any open reports and load `3_3_formatting_with_functions.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Unfilter your data from the previous exercise, so we can see all orders.

Next to the summary statistics we created earlier, let's apply some rounding to our lowest and highest item prices.

In cells `AB2` and `AB3` apply a **rounding up** formula to `0` digits.

In cell `AB4`, apply a standard **rounding** formula to `1` digit.

What is the profit margin ratio when rounded to one digit?

40.0%

That's right! The total profit margin for AdventureWorks was 40% when rounded to 1 digit. Still a great achievement!

Do you know the meaning of data viz?

Which of these statements about data visualization is FALSE?

Answer the question
50XP

Possible Answers

Select one answer

- Enhances understanding but not retention of information.

  `PRESS1`

- Data visualization is the graphical representation of data using charts, graphs, and other visual tools.

  `PRESS2`

- It enables you to present complex data to identify outliers.

  `PRESS3`

- Critical part of the decision-making process in data-driven organizations.

  `PRESS4`

That's right! This statement is FALSE because not only does data visualization help enhance your understanding of data, but it can also help you retain that information.

Visualizing categories

The operations team wants to understand which products perform best, so let's visualize average sales across product categories. Excel has a variety of options to represent categorical data effectively, so we'll use column or bar charts to convey this information. These charts are powerful tools, enabling concise and visually impactful data presentations.

How to do this exercise locally

Instructions

100 XP

- Close any files that you've currently got open.
- Open the workbook called `3_4_visualizing_categories.xlsx` from the Workbooks folder.

Navigate to the `By Product Subcategory` sheet.

From the *Insert* tab, create a *2D Column* or *2D Bar* chart to visualize `Average Sales` by `Product Subcategory`.

Change the title of the chart to "Average Sales".

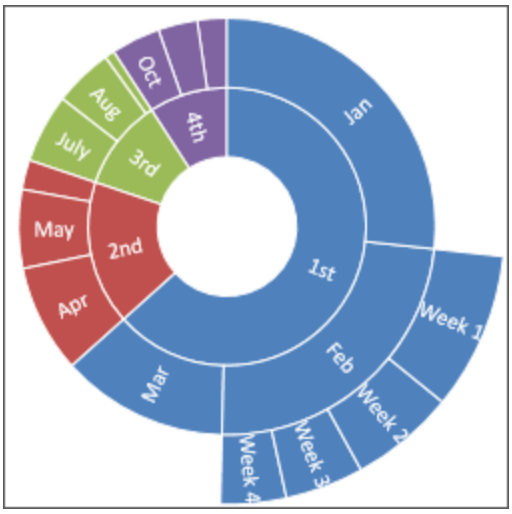
Which category has the tallest column/bar and therefore has the highest average sales?

Mountain Bikes

That's right! Mountain Bikes generate the highest amount of sales on average compared to any other category. This might be expected as they are typically more expensive that clothing / accessories. But this chart enables us to easily compared to our other bike types.

Capturing parts of a whole

The management team wants to explore how their orders vary from country to country and state to state. We'll utilize a sunburst chart to visualize sub-categories within a larger category. This powerful chart type allows us to visually represent the interconnectedness and proportions of our data clearly and concisely. Sunburst charts are a type of treemap.



If you lost progress, close any open reports and load `3_5_capturing_parts_to_a_whole.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Navigate to the *By Location* sheet.

Create a *Sunburst chart* to visualize `Orders` by `Country` and `State`.

- Change the title of the chart to "Orders by Country and State".
- Re-size the chart to make it easier to read.

Which state in Australia has had the highest number of orders?

New South Wales

That's right! New South Wales had almost 6,000 orders making it the state with the most orders across Australia. In our next exercise, we'll be exploring how to look at data over time.

Capturing a trend

Visualizing data over time can help find trends in sales performance. To achieve this, we will utilize line charts. However, a crucial step will be editing the data source, which we'll explore during the exercise. Line charts offer a powerful means to represent trends and fluctuations, providing valuable insights into sales patterns.

If you lost progress, close any open reports and load `3_6_capturing_a_trend.xlsx` from the Workbooks folder.

How to do this exercise locally

Instructions

100 XP

Navigate to the *By Date* sheet.

Create a line chart to visualize `Total Sales` by `OrderYear` and `OrderMonth`. It may not look right initially, but don't worry - we'll fix that in the next step.

Let's start by opening the *Select Data Source* pop-up.

Let's update the data source to:

- `Total Sales` as the *Legend entries (series)*.
- `Order Year` and `Order Month` as the *Horizontal (category) axis label*. (Do not select the column headers!)

- Change the title of the chart to "Total sales over time".
- Re-size the chart to make it easier to read.

Which month and year had the highest number of total sales? I.e. 1 2020

5 2020

That's right! May 2020 had the highest sales for the entire dataset - this could have been due to the increased number of people at home during lockdowns.