Exercise 1: Add a calculation to the data set

Task 1. Upload the American Time Use Survey data set

1. Click + New data.
2. Click Local file.
3. Click Browse.
4. Navigate to the American Time Use Survey.csv file on your hard drive.
5. Select the file.
6. Click Open.
7. Click Import (at the bottom right of the page).

Task 2. Add a calculation to the data set

1. At the bottom right of the American Time Use Survey data set, click the ellipsis.
2. Click Refine.
3. On the left, click the Actions icon.
4. Click Calculation.
5. In the blue box, type Time with Children.
6. Click the A box.
7. On the right, scroll to locate Caring for Children.
8. Click Caring for Children.
9. Click the B box,
10. On the right, scroll to locate Playing with Children.
11. Click Playing with Children.
12. Click Done.
13. Click Actions to close the Actions dialog box.
14. In the data tray at the top, click the arrow to advance the items until the Time with Children column is visible.
15. At the top left of the page, click the down arrow.
16. Beside American Time Use Survey, click the x.
17. Click Don't save.

Exercise 2: Add a data group to the data set

Task 1. Add a data group

1. At the bottom right of the American Time Use Survey data set, click the ellipsis.
2. Click Refine.
3. Click the Weekly Earnings column header.
4. In the dialog box, to the right of Weekly Earnings, click the vertical ellipsis.
5. Click Data group.

   In the Data group dialog box, Watson Analytics has created default groupings based on the analysis of the continuous data in this column. Currently the distribution of the data is equal across 5 groups. You can change the number of grouping levels and data distribution using a Custom grouping. In this case you want to create 3 groups, with a custom distribution of 0 to less than 1000, 1000 to less than 2000, and 2000 and above.
6. Delete the value in the How many box.
7. Type 3.
8. Click the value of the first distribution level.
9. Click the pencil icon.
10. Select the value by dragging the cursor across it.
11. Type 1000.
12. Click the value of the second distribution level.
13. Select the value by dragging the cursor across it.
14. Type 2000.
15. In the Name box, type Weekly Earnings Buckets.
16. Click Done.
17. Click the Next arrow (top right of the column headers) until you see the Weekly Earnings Bucket column header (last column)

Task 2. Add a data a 2nd data group

1. Click the Previous arrow (top left of the column headers) until you see the Age Range column header.
2. Click the Education Level column header.
3. In the dialog box, to the right of Education Level, click the vertical ellipsis.
4. Click Data group.

   Watson Analytics retrieves all the data values in the column. In this case there are 11 items in the Education Level column, however, you would like to see the data consolidated into three grouping levels.
5. Under Search, click 9th grade, 10th grade, 11th grade, 12th grade, and High School.
6. Click New Group.
7. In the Enter group name box, type High School or Less.
8. Click OK.
9. Click New Group.
10. In the Enter group name box type Undergraduate Studies.
11. Click OK.
12. Click Some College, Associate Degree, and Bachelor.
13. Click New Group.
14. In the Enter group name box type Post Graduate Studies.
15. Click OK.
16. Click Master, Prof Degree, and Doctoral Degree.
17. In the Enter a name box, type Education Level Groups.
18. Click Done.

19. Click the Next arrow (top right of the column headers) until you see the Education Level Groups column header (last column)
20. Place the cursor over the Education Level Groups column header.

Task 3. Save the data set and then use it to create a Discovery that includes the new data group.

21. At the top right of the page, on the toolbar, click Save.
22. At the top left of the page, beside IBM Watson Analytics, click the down arrow.
23. Beside American Time Use Survey, click the x.
24. Click American Time Use Survey.
25. In the Ask a question box, type What is the relationship between weekly earnings and education level?.
26. Press Enter.
27. Under Most relevant, click the 1st Starting point.
28. In the text area above the graph (question area), click Education Level.
29. Click Others.
30. Scroll to locate Education Level Groups.
31. Click Education Level Groups.
32. At the top left of the page, click the down arrow.
33. Beside New discovery set, click the x.
34. Click Don't save.

Exercise 3: Add a hierarchy

Task 1. Upload a data set.

1. Click + New data.
2. Click Local file.
3. Click Browse.
4. Navigate to the CPG data.xlsx file on your hard drive.
5. Select the file.
6. Click Open.
7. Click Import (at the bottom right of the page).

Task 2. Identify auto-generated hierarchies in the data set

1. At the bottom right of the CPG data data set, click the ellipsis.
2. Click Refine.
3. In the left pane, click the Actions button.
4. Click Unhide all.
5. Scroll down to, and then click, Sub Category - Product.

Watson Analytics has automatically created this hierarchy based on its initial analysis of the data during import. It has two levels from the data set, one based on the Sub-Category column and the other based on the Product column. You will delete this hierarchy and manually recreate it in a Discovery, later in the exercise.

6. Click Delete Hierarchy.

You will determine if Watson Analytics has auto-generated any other hierarchies in this data set.

7. Scroll the list of columns in the Actions list.

You can identify a column having a compound name, Year (Date) - Day (Date). This is a cue that this column represents a hierarchy.

8. Click Year (Date) - Day (Date)

This is in fact a hierarchy, consisting of three levels, Year (Date), Month (Date), and Day (Date). You can modify the hierarchy by adding levels or by renaming the hierarchy.

9. Beside Year (Date), click the dash icon.

You can also remove levels. You will not make any changes to the hierarchy at this point.

10. Click Cancel.
11. In the left pane, click the Actions button.

You can view the hierarchical structure of the data prior to using it in a Discovery.

12. At the top right of the page, beside the last column header, click the Next arrow to scroll the columns until you reach the Year (Date) - Day (Date) column.
13. Click the Year (Date) - Day (Date) column header.

You can see the hierarchical structure of the data values, including the years level, months level, and days level.

You will save the data set so that you can use it to create a Discovery

14. On the toolbar, click Save.
15. At the top left of the page, click the down arrow.
16. Beside CPG data, click the x.

Task 3. Create a Discovery based on a data set that includes a hierarchy.

1. Click the CPG data data set.
2. In the Ask a question about your data box, type What is the trend of Sales over Year by Category?.
3. Press Enter.
4. Click the first visualization under Most Relevant

In this case, Watson Analytics has automatically used the hierarchy in a visualization. Year (Date) - Day (Date) appears in the x axis of the graph, and by default data is displayed at the highest level in the hierarchy, which is the year level (2012 and 2013)

If you want to analyze the data at a lower level of granularity, for example at the month or day level, the hierarchy lets you drill down, first on the year data values in the graph, and subsequently on the month data values.

5. Right-click 2012.
6. Click Go down.
7. Right-click January.
8. Click Go down

You can also navigate back up the hierarchy.

9. Right-click 1.
10. Click Go up.
11. Right-click January.
12. Click Go up.

Task 4. Add a hierarchy to a Discovery.

At this point in your analysis, you know that there is another natural relationship in your data, between Category, which is currently what you are analyzing, Sub Category, and Product. You would like to see values at a lower level of granularity, so you will attempt to drill down on Category.

13. Right-click Oral.

Notice, that there is no option to drill down, even though you know a relationship exists between these values. When you did the import, Watson Analytics detected a portion of this relationship and automatically created a hierarchy but it was not complete. You could have modified the hierarchy as part of the data set, but for the purposes of showing you how to add a hierarchy in Discovery, it was deleted in the data set. You will add the hierarchy now.

14. At the bottom left of the page, click Actions, and then click Hierarchy.
15. In the Name box, type Products.
16. Click Add level.
17. Click Category.
18. Click Add level.
19. Click Sub Category.
20. Click Add level.
21. Click Product.
22. Click Done.

To use the hierarchy you must add it to the visualization, replacing an existing item. In this case, you will replace Category with the Products hierarchy.

23. In the question section, click Category.

24. Click Others.
25. Scroll to locate Products.
26. Click Products

Now you can drill down on values at the Category level, to see values at the Sub Category, and continue on down to values at the Product level.

27. Right-click Oral.
28. Click Go down.
29. Right click Toothpaste.
30. Click Do down.
31. At the top left of the page, click the down arrow.
32. Beside New discovery set, click the x.
33. Click Don't save.