



## Creating a Categorical Variable

## Tutorial

*Time:* 20 minutes  
*Skill level:* Medium  
*Editions:* Reader,  
Basic, Professional

### Illustrative problem

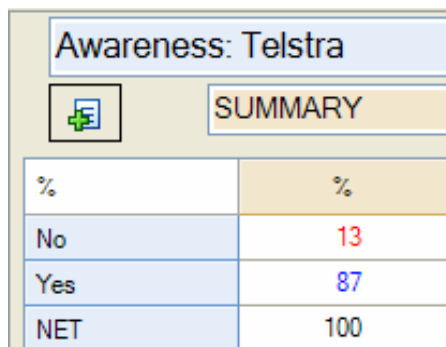
You need to combine multiple different questions and variables to create a new **Single Response – Categorical** question. For example, you may wish to segment customers according to commitment to a brand.

### Activities

1. Open Tutorial 9.Q.
2. In the **Variables and Questions** tab, right-click and select **Insert New Variable(s)** and then select **Binary – Complicated Filter...**
3. Click on the **Undefined Filter Term** inside the top right panel. Now that it is highlighted it is ready for editing.
4. From the **Variables and Questions** box select `Top of mind awareness`, in **Contains** select **Any of** and select `Telstra (Mobile Net)` from **Values**.
5. Click on **OR** at the top of the dialog box.
6. Click on the **+** button. This creates a copy of the expression that was created in step 4.
7. Select the second "Top of mind awareness" Any of "Telstra (Mobile Net)" in the **Or** tree. Observe that the expression is highlighted and recall from step 3 that highlighting means that the expression is ready for editing
8. Select `Q5. Unaided Awareness` from the **Variables and Questions** box, in **Contains** select **Any of** and select `Telstra (Mobile Net)` from **Values**.
9. Click on the **+** button again, select the second "Q5. Unaided Awareness" Any of "Telstra (Mobile Net)" in the **Or** tree.
10. Select `Telstra (Mobile Net)` from the **Variables and Questions** box, in **Contains** select **Any of** and in **Values** select `Yes`. Observe that the values of the

variable being created are previewed in the Result column of the bottom panel.

11. Change the **Name** to `awTelstra`.
12. Change the **Label** to `Awareness: Telstra`.
13. Click **OK**.
14. Click on the **Values** button in the row containing `awTelstra`. Change the **Label** `Not Selected` to `No` and the `Selected` to `Yes`.
15. Click **OK**.
16. Create a table of this variable. Your table should look like Figure 1.



%	%
No	13
Yes	87
NET	100

Figure 1. Summary of constructed binary variable

17. Following the steps immediately above, create a new binary variable with name `cuTelstra` and **Label** `Current: Telstra` (use **Company currently with**). When you do a summary, you should have 26% selected in this variable.
18. Create a variable named `prTelstra` with **Label** `Previous: Telstra`. This variable should select current Telstra users who were also with Telstra in their previous contract / arrangement. There are a variety of ways of doing this. One approach, illustrated in Figure 2, is as follows:
  - (a) use q11, q12, and q13 (which begin with 'Company for previous contract' in rows 35, 36, and 37 of the **Variables and Questions** tab) to establish if the user has *ever had* a relationship with Telstra - select these as options under **OR**;
  - (b) use q10 and q7 (in rows 31 and 34 of the **Variables and Questions** tab) to establish if they are still with their previous provider, and are currently with Telstra – select these as **AND** options, nested within the **OR**.When you view a summary of this variable, the proportion shown as **selected**

should be 19%.

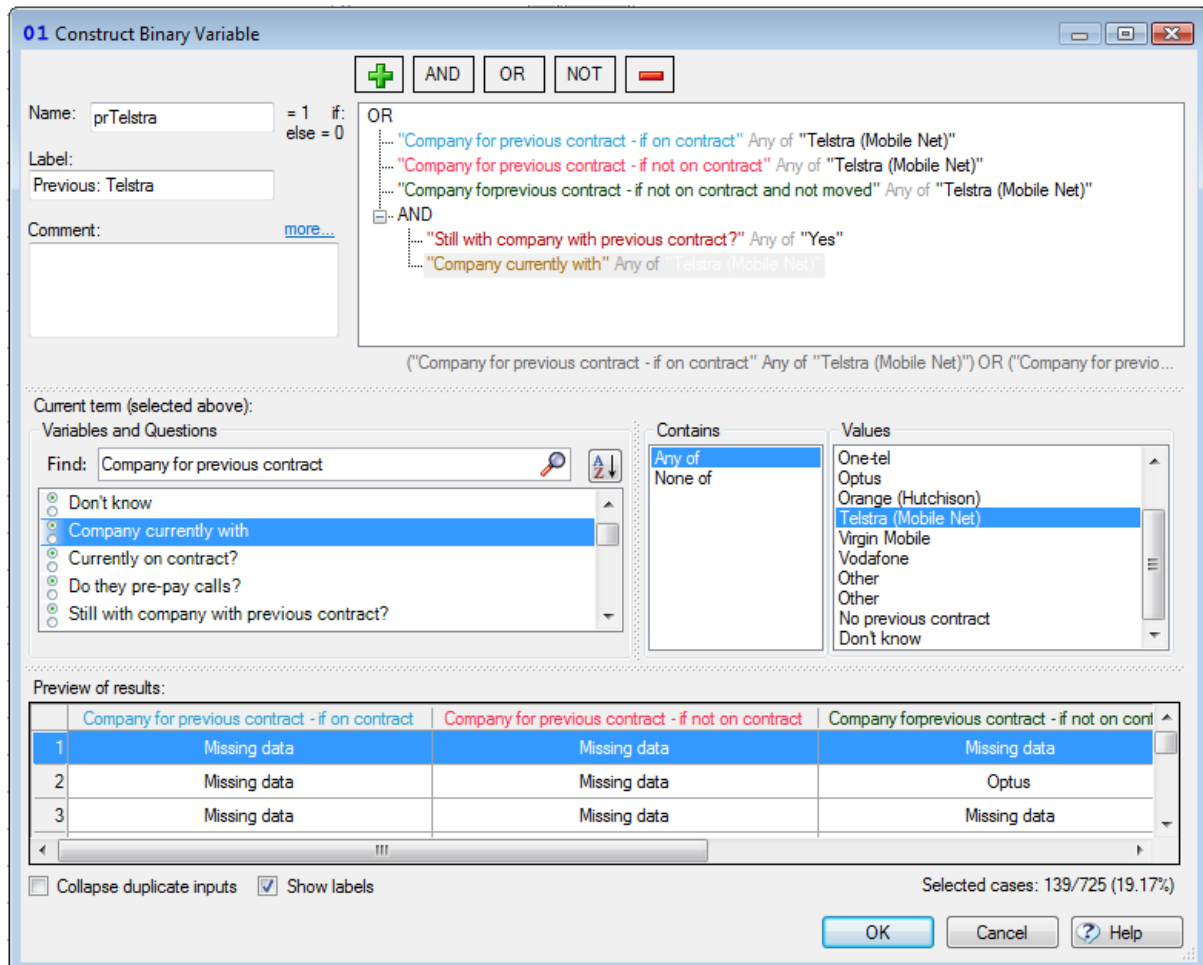


Figure 2. Creating previous contract with Telstra

19. If the three variables you have just created are not already at the top of the file, move them to the top (this is done for neatness only).

20. Order the variables with awTelstra first, prTelstra second and cuTelstra third.

21. Select all three variables, right-click and select **Insert Ready-Made Formula(s)** and then select **Multiple Response – Binary -> Single Response – Categorical...**

This will result in a warning message because the variables are not mutually exclusive, as shown in Figure 3. It is always a message you should think about – do not get into the habit of routinely ignoring this message. In the current example, the variables are not mutually exclusive because, for example, one would expect most people who are customers of Telstra to also be aware of Telstra.

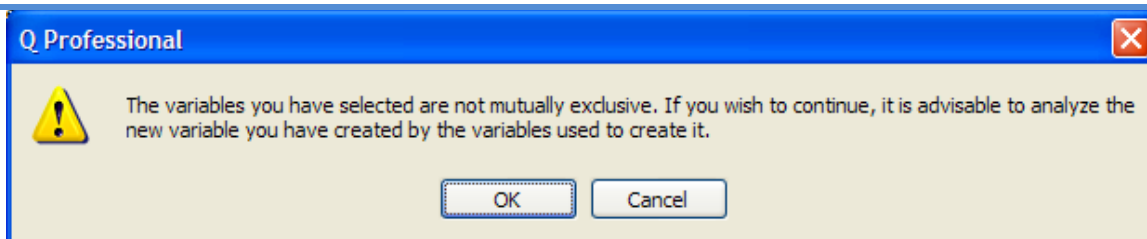


Figure 3. Warning for non-mutually exclusive variables

22. Click **OK**.

23. Name the variable `loTelstra`, give it the **Label** `Loyalty: Telstra` and click **OK**.

24. Click the variable's **Values** button.

25. In the **Label** field rename `Missing data` as `Unaware`.

26. Uncheck the **Missing Data** field for `Unaware`.

27. Tidy up the Labels to read in descending order: `Unaware`, `Aware`, `Defected` and `Customer`. Click **OK**.

28. View a summary of the newly created variable. It should look like Figure 4.

Loyalty: Telstra	
SUMMARY	
% Count	%
Unaware	11
Aware	57
Defected	6
Customer	26
NET	100

Figure 4. Telstra loyalty

29. In the **Variables and Questions** tab select `awTelstra`, `prTelstra` and `cuTelstra` and create a **Multiple Response – Binary** question, giving it a name of: `Loyalty – Not mutually exclusive`.

30. Select the new loyalty variable in the brown drop-down on the **Tables** tab.

31. Select **Total %** from **Statistics – Cells** and de-select **Column %**.

Total %	Awareness: Telstra	Previous: Telstra	Current: Telstra	NET
Unaware	0	0	0	0
Aware	57	0	0	57
Defected	5	6	0	6
Customer	25	13	26	26
NET	87	19	26	89

Figure 5. Telstra loyalty final

The result should be the table shown in Figure 5. This table shows how **Multiple Response – Binary -> Single Response – Categorical** has dealt with the non-overlapping nature of the categories. For example, it shows that anybody who has said they are a current customer of Telstra has been classified as a customer in the categorical variable (i.e., the row categories). However, if somebody indicated they are aware of Telstra, they have only been classified in the awareness category if they are not a current or former customer. The reason that Q has done this is that when faced with non-mutually exclusive categories, Q looks at the order of the variables and gives precedence to the category furthest from the top (i.e., if you place the binary variables in a different order, you will get different results when constructing a categorical variable from non-mutually exclusive categories).