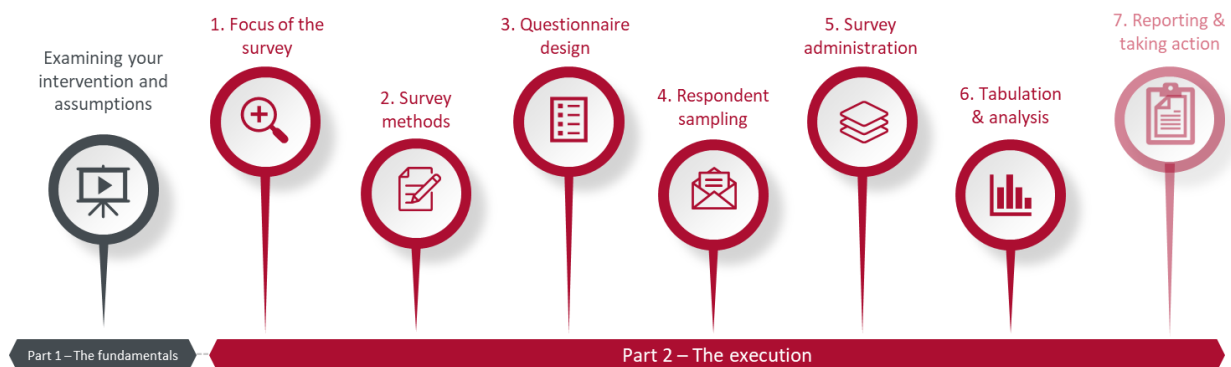


MODULE 6. ANALYZE SURVEY RESULTS



You finally have the data! There are literally hundreds of different analyses you could perform with the data you collected. But you do not have the time nor resources to “boil the ocean”. Only a few of all possible analyses will be valuable to you. *Module 6* will help you plan your analysis, execute, and extract useful insights with as much focus and efficacy as possible.

Planning the analysis

The first step is to look back at the assumptions which you listed in *Module 1*. For each assumption, you should identify which survey questions are relevant, and what kind of analyses would allow you to draw the insights that would confirm your assumption (or not). Keep in mind that there are two main types of analyses:

- Single variable analysis (or “primary analysis”): you look at the results of one given question, in terms of distribution of the answers or the average value. For instance, you can compute the Net Promoter Score, or see what percentage of respondents like a given product. This is the simplest and most superficial type of data crunching (often automatically generated by survey tools), and in some cases it can already give you important information to validate some assumptions.
- Cross tabulation (or “secondary analysis”): you look at the results of two questions at a time, i.e., you use the answers to one question, and you cross them with the answers to another question. For example, you might have computed the average satisfaction rate from a primary analysis. While this is valuable information, you might want to go deeper and look for some explanations for the result. You could calculate the average satisfaction rate by region, by product or sales person. If you see significant differences across the results, it would be an indication of what drives satisfaction among customers.

While secondary analysis typically yields the most valuable insights, you also must keep in mind that it takes more time than primary analysis, and that many, many combinations of questions are possible. Accordingly, you should carefully select which secondary analysis you want to carry out, based on what will help you test the assumptions outlined in *Module 1*. Remember that you are not simply looking for interesting findings, but rather need to focus on actionable insights. For example, seeing that satisfaction levels of product A and product B are both relatively low might be interesting, but this result alone is not actionable, as you do not know what is driving dissatisfaction, and won’t know how to improve the status quo. However, you should also be able to find out what features your customers like and dislike about for each of your products. By comparing those results, you might discover that users mostly complain about the lack of clear instructions in product A, while users of product B mostly don’t like its design. This information is not only interesting, but also actionable: it gives you two concrete levers to improve your products and increase your consumer satisfaction.

Once you are clear about this, list the primary and secondary analyses you want to perform, before launching into the analysis itself. It will serve you as a guide in the coming steps.

Preparing the data

Before you dive into the analysis, you need to prepare your data. To do so, download the raw survey data (i.e., the full list of responses) in a spreadsheet format (CSV or Excel) from the survey tool. Use common sense to check for any irregularities or inconsistencies. For example, if you asked your respondents for a percentage, you might want to review all your results for formatting consistency (some might have answered 0.5 and some 50, while both mean 50%). Some questions might have been misinterpreted or there might have been errors in entering the answers, so you will have to screen the data for outliers and correct or delete datapoints that do not make sense. If you ran a few test surveys with dummy answers, also make sure to remove those entries from the database. Checking the integrity of your data at the onset will save you time during the actual analysis and will increase the reliability of your results.

To get the most out of your open-ended questions, you will also need to “code” the answers to analyse the results quantitatively. This means you will have to go through the responses, identify categories for the most common answers, and assign each response to the relevant category/ categories – essentially turning an open-ended question into close-ended. Concretely, in your data spreadsheet you should add a column for each category (including an “Other” category for less common answers) and put a “1” where applicable, like in the following example:

<i>Please describe what you like the most about spending your vacation in Brazil</i>	Weather	Sun	People	Food	Beach	Relaxing	Cheap flights	Other
I love the sunny weather and the beautiful beaches	1	1			1			
For me, it's all about the people, the food and the relaxed atmosphere			1	1		1		
I have many friends here so I love visiting them!								1
I love the sun and spending time on the beach and in the pool		1			1			
Cheap flights and the beach is beautiful					1		1	
I just enjoy doing very little and relax on the beach					1	1		
The climate, the food and the people, the sun, the beach... I love everything about it!	1	1	1	1	1			
Flights are cheap and the hotels, too							1	1

Depending on the survey tool and the specific format in which your raw data is exported, you might need to manually code multiple-choice questions (close-ended questions where people could select more than one answer). Some survey tools already format the answers as shown above, with one column per answer option (with 1 if selected and 0 if not selected), while others combine all selected answers into a single column, as follows:

<i>What aspects of our product did you like the most? Pick up to 3 options</i>
Price; Ease of use
Performance; Ease of use; Durability
Price; Performance

This makes it tricky to analyse this question (e.g., calculate what percentage of respondents picked each individual option). Therefore, you should manually reformat the question by doing the same process described for open-ended questions (except that you already have the categories defined).

Finally, for questions with continuous numerical answers (e.g., age or income), it is useful to group answers into ranges to facilitate comparisons, creating a new column for the range, like in the following

example:

<i>What is your age?</i>	Age ranges
24	20 to 29
18	Below 20
46	40 or above
32	30 to 39
28	20 to 29
25	20 to 29
38	30 to 39

Running the analysis

The most accessible and easy-to-use tool for data analysis is undoubtedly Excel's pivot table functionality. Some of you might already be familiar with pivots, some maybe not. But the good news is that you don't have to be an Excel guru to use this tool and perform a solid analysis. Although it might take some time and practice at the beginning to learn how pivot tables and charts work, it is a worthwhile investment: you can use this tool for many different purposes, and it will save you a lot of time and frustration.

Although an introduction to Excel pivot tables is beyond the scope of this training, you can find plenty of great tutorials online. For example, the following 3-part video series (each about 15 minutes long) is an excellent resource to get you started:

Part 1: <https://www.youtube.com/watch?v=9NUjHBNWe9M>

Part 2: <https://www.youtube.com/watch?v=g530cnFfk8Y>

Part 3: <https://www.youtube.com/watch?v=FyggutiBKvU>

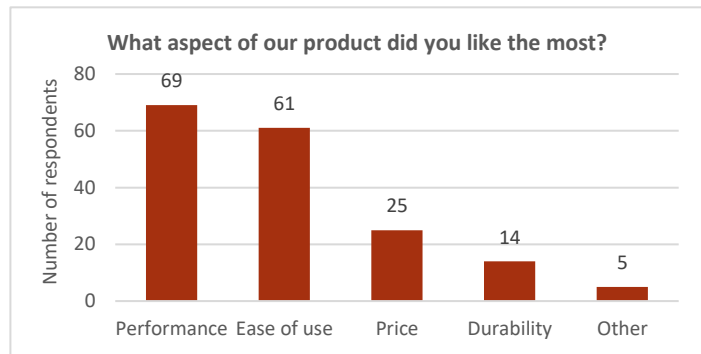
Given you probably have several analyses to perform, it is important to keep your spreadsheet organized. For example, you can create different tabs for primary and secondary analyses, or organize them by topic (e.g. one tab per topic). It is also advisable to always put titles above each table or chart, to easily identify what analysis you ran. Do not take shortcuts here, as these simple steps will allow you (and other people) to navigate your Excel document. And taking care of this while you are doing the analysis is more efficient than "fixing it" at the end.

It is best to start by analysing the profile of the respondents, running primary analysis on profiling questions (e.g., gender, age, location, and any other relevant characteristic). This will also allow you to assess the representativeness of your sample. Next, you can move on to performing the primary and secondary analyses you identified earlier. Many survey tools will automatically do the primary analyses for you but remember that if you cleaned up your data significantly, you should re-run the analysis yourself on the clean dataset. As you start seeing the results, you might identify the need for further analyses (e.g., to dig deeper into an unexpected result). This is perfectly normal but remember to always keep your goal in mind: the assumptions you want to test.

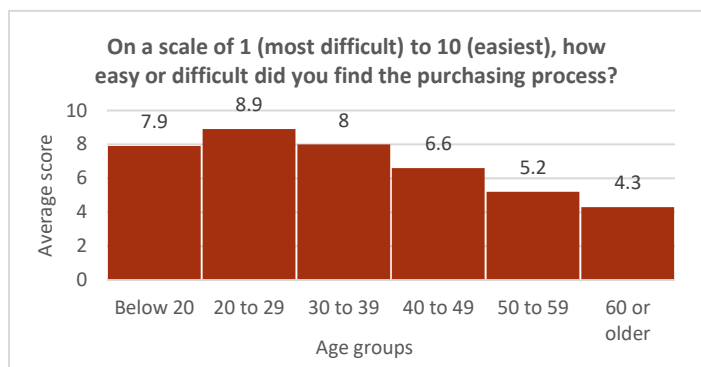
Visualizing the results

Besides the numerical analysis (tables), it is useful to visualize results through charts, which are often easier to understand. There are many ways to visualize data, but some charts are best suited for a given purpose. Picking one visualization or another can make a big difference in how data is interpreted. Here are the most common chart types and how to use them:

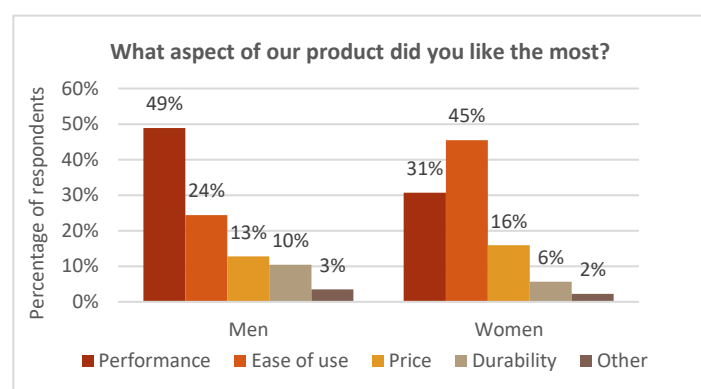
Column charts are used to compare quantity, frequency, or other values across different categories. They are often used in primary analysis, to display the number (or percentage) of respondents who gave each answer. If you have many categories, or categories with long names, it is best to use bar charts, which are the same but organized vertically rather than horizontally.



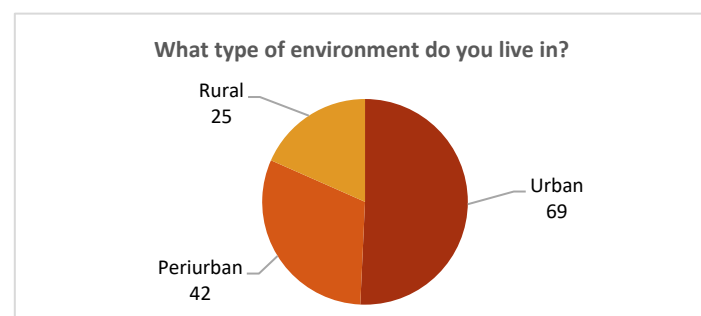
Histograms are like bar charts, but each column represents a group defined by a continuous numerical variable (rather than discrete qualitative categories), divided into ranges, and sorted in increasing or decreasing order.



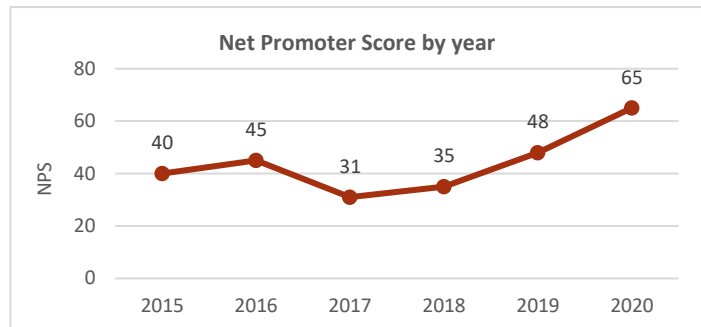
Clustered column charts are another variation on column charts used to compare the breakdown of a variable across categories. They can be useful to visualize secondary analysis, but keep in mind that if you have too many categories, they can become hard to read. Also, it is usually best to display the breakdown in percentages rather than absolute numbers, to facilitate comparisons across groups of different sizes.



Pie charts are used to show the breakdown of a single variable into mutually-exclusive categories, where the total adds up to 100%. Given the human mind is not very good at assessing angles and areas, pie charts should only be used when plotting few categories of data.



Line charts are used to show changes in value over time (e.g., months or years), for one or more variables. They are useful to identify trends over time. While they are very useful to display operational and financial data (e.g., number of customers or revenue), they are less commonly used when analysing survey data.



Scatter plots are used to visualize the correlation (or lack thereof) between two numerical variables, one positioned on the horizontal axis and one on the vertical axis. Each dot represents one datapoint (i.e., one respondent). Scatter plots can also help you identify clustering trends and spot data anomalies or outliers.



Besides choosing the best chart type to visualize your results, keep in mind the following best practices:

- Include a chart title
- Display axis titles and make sure units are clear
- Show data labels (except if you have a lot of datapoints, e.g., with scatterplots)
- Avoid coloured backgrounds and special effects (3-D, shadows, etc.)
- Use a simple and consistent colour scheme (e.g., if you have multiple charts comparing two groups, always use the same colour for each group)
- Arrange data intuitively and logically (e.g., largest to smallest).