**Learning Log: Start a data analysis checklist**

**Instructions**You can use this document as a template for the learning log activity: Consider how data analysts approach tasks. Type your answers in this document, and save it on your computer or Google Drive.

We recommend that you save every learning log in one folder and include a date in the file name to help you stay organized. Important information like course number, title, and activity name are already included. After you finish your learning log entry, you can come back and reread your responses later to understand how your opinions on different topics may have changed throughout the courses.

To review detailed instructions on how to complete this activity, please return to Coursera: [Learning Log: Start a data analysis checklist](https://www.coursera.org/learn/analyze-data/supplement/yq5wr/learning-log-start-a-data-analysis-checklist).

| **Date:** <enter date> | **Course/topic:** Course 5: Analyze data to answer questions | | |
| --- | --- | --- | --- |
| **Learning Log:** Start a data analysis checklist | | |
| **Create your checklist** | Start a high-level checklist of tasks you need to complete during analysis using the steps you have already learned:  **Ask**   * *State the problem.* * Our statement of the problem is our focus. * Try to see the bigger picture! * Never set the sail alone!   **Prepare**   * What metrics to measure? * What factors should be taken into account? * Where is the data located ? * If the data will be moved, how it will be stored and what are the needed security measures to protect that data.   **Process**   * Using proper tools to find incorrect and incomplete data. * Removing inconsistencies in data. * Identifying whether your data is biased.   **Analyze**   * Performing different calculations to get additional metrics. * Combining additional data attributes from a variety of sources to get a more comprehensive story. * Create different views for the data. Like tables with your results, filter and pivot them. * Make it visual if possible! Charts tell more than a thousand words.   **Share**   * Making better decisions. The feedback will help to answer the questions that initially were not thought of. * Making more informed decisions. Feedback will not be merely critic, but also suggestions and additional information on the matter. * Improve the general outcome. From one angle, the decision will most likely be more informed and better, but also the transparency will grant that there is more support to the findings.   **Act**   * How can the feedback received during the sharing phase (step 5) be used to meet the stakeholder’s needs and expectations? * What potential solutions to the outlined problem could there be? * Is this problem worth solving? (Yes, that is also a potential outcome) | | |
| **Reflection:** | Write 3-5 sentences (60-100 words) answering the following questions: | | |
| **Questions and responses:** | * Did making a checklist help you better understand the data analysis process?   *I found that making a checklist helped me better understand the data analysis process. By breaking down the process into a series of clear, concise steps, I was able to gain a deeper understanding of the different stages involved and the tasks that need to be completed at each stage.*   * How can you adapt your checklist to fit each new project?   *I can adapt data analysis checklist to fit each new project by considering the unique requirements and goals of the project. This can involve modifying the steps in your checklist to reflect the specific data sources, analysis techniques, and reporting formats that are required for the project.* | | |

## **Step 1: Ask - Understand the problem**

It is always important to understand what even seems to be the problem or the question. Making an assumption or not understanding fully the problem will lead to wrong conclusions and will result in wrong actions. Identifying the problem is naturally also one of the hardest tasks. Like A. Einstein stated:

'If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.'

So what would help to identify the problem? The following actions should help:

* State the problem. This will become your first cornerstone. If it should change over the course of time, it is very natural. The more we know the wiser we are. All the problem statements should be measurable, clear and concise.
* Our statement of the problem is our focus. Everything else should be an afterthought and avoided.
* Try to see the bigger picture! Take a step back and see the whole situation in context. And context is crucial here. Different settings can give different meanings.
* Never set the sail alone and make sure you fully understand the collaborators' expectations. This means, get people involved, get their views and interest. Once that is there, make it also clear what they expect! Be open in all the conversation and do not play telephone game\*, meaning try to communicate to everyone in the same manner and if possible in the same conversation.

\*Chinese whispers (Commonwealth English) or telephone (North American English) is an internationally popular children's game. Players form a line or circle, and the first player comes up with a message and whispers it to the ear of the second person in the line.

* Questions to ask:

1. What are the stakeholders stating as their problems?
2. How can the stakeholders questions be resolved? (this is a bit reverse engineering)
3. Is the stated problem really the root cause?

## **Step 2: Prepare - What do I need?**

Once there is an understanding of the problem, one can think about how to solve this. Time to decide what data needs to be collected in order to answer the questions and how to organize it so that it is useful. One should think about the following aspects:

* What metrics to measure? (\*Metrics are quantitative measurements). In answering this question, there might be a need to answer also sub-questions (e.g., Is our time-to-market competitve for product X? If not, what process improvements would help?).
* What factors should be taken into account?
* Where is the data located (files, database, external system, internal system)?
* If the data will be moved, how it will be stored and what are the needed security measures to protect that data.

### **Questions to ask yourself in this step:**

1. What needs to be figured out how to solve this problem?
2. What would help to measure the outcome of any change to the problematic area?
3. What research is needed?
4. Where is the information held?

## **Step 3: Process - Make it usable!**

When we start using the data, it might be a combination from different sources or it might not be of the highest quality. A process known as data cleaning is the fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. What we aim to achieve is clean data. And to tell the truth, that is a science on its own. There are plenty of tools, theories, and methods to use, but let's keep everything basic here. Data Cleaning does not require fancy tools or words, a simple Spreadsheet program (yes, that is Excel) will suffice. Although my preference lies with others (Pandas!👌). So during this step one might:

* Using proper tools to find incorrect and incomplete data.
* Removing inconsistencies in data. Sometimes there might be duplicated entries.
* One of the most important aspects to keep in mind - identifying whether your data is biased. Essentially, data that is biased will not be representative of the population or phenomenon of study, our issue we are trying to solve.

### **Questions to ask yourself in this step:**

1. Is the data source trustable and data quality high?
2. What data errors or inaccuracies could occur within given dataset.
3. What is the best possible answer to the problem being solved?
4. How to clean the data so the information is more consistent? (e.g. replace values with mean values, et cetera)

## **Step 4: Analyze - Tell me the story!**

Next up is to make some conclusions based on the trustable data. Data Analyses is a skill that takes time to master, but over time the patterns will emerge faster and methods one uses will develop. Main concept is to think analytically about your data, be critical and be creative. There might be a need to sort and format the data to make it easier to process, make a Pivot table, or create awesome graphs! Remember it is a story that must unfold. Further processing might include:

* Performing different calculations get get additional metrics.
* Combining additional data attributes from a variety of sources to get a more comprehensive story.
* Create different views for the data. Like tables with your results, filter and pivot them.
* Make it visual if possible! Charts tell more than a thousand words.

### **Questions to ask yourself in this step:**

1. What story is my data telling me?
2. Why can’t it be done?
3. Will X (e.g. time, money, manpower or expertise) allow us to solve the issue?
4. How will my data help me solve this problem?
5. Who needs my company’s product or service?
6. What type of person is most likely to use it?

## **Step 5: Share - Get different views**

One thing still to remember, whatever we do, we are biased. So as the next step, get additional opinions about the findings. This will significantly help to improve the results and ensure that main aspects were taken into account. As the are many ways to share the finding each person has their preference and so does each company. However, many studies reassure that with clear and enticing visuals of the analysis results, the story is better understood. (A Good article on this https://hdsr.mitpress.mit.edu/pub/zok97i7p/release/3). The tools do not really matter here, it can be Tableau, Excel or even good old paper and pencil! But take this as a chance to show the stakeholders how their problem was solved. Sharing will certainly help with:

* Making better decisions. The feedback will help to answer the questions that initially were not tought of.
* Making more informed decisions. Feedback will not be merely critic, but also suggestions and additional information on the matter.
* Improve the general outcome. From one angle, the decision will most likely be more informed and better, but also the transparency will grant that there is more support to the findings.

### **Questions to ask yourself in this step:**

1. How can I make what I present to the stakeholders engaging and easy to understand?
2. What would help me understand this if I were the listener?
3. What makes a data visualisation good?

## **Step 6: Act - We know the problem, Let's solve it!**

No analysis conclusion should remain to collect dust on a shelf! Rather some action should be taken. Taken the results and depending on the problem statement, recommendations for further actions can be made. And once the recommendations are ready, the actual decision can be made! Not necesarrily is the conductor of the analysis the one to make a decision, it could also mean providing the decision-makers(stakeholders) with recommendations based on the findings so they can make data-driven decisions. But the key here is data-driven decisions.

### **Questions to ask yourself in this step:**

1. How can the feedback received during the sharing phase (step 5) be used to meet the stakeholder’s needs and expectations?
2. What potential solutions to the outlined problem could there be?
3. Is this problem worth solving? (Yes, that is also a potential outcome)

And done you are!

* These are the six steps that Google has outlined for Data Analysis. And they do help to structure the thinking when conducting Data Analysis. To define structured thinking then actually it is breaking the (data analysis) process into smaller, manageable parts. According to Googlers, this process involves four basic activities:

1. Recognizing the current problem or situation.
2. Organizing available information.
3. Revealing gaps and opportunities.
4. Identifying your options.

But with that, let's go and make some data-driven decisions!