# New to Data Science? One way of structuring your projects.

[Chris Gudde](https://medium.com/@chris.gudde)Follow

Feb 13

## Making the most of OSEMN as a new Data Scientist.

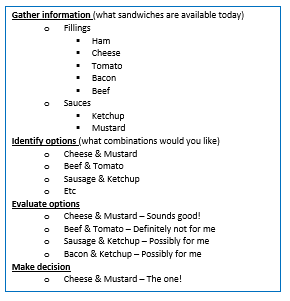
As a new Data Scientist you are likely to be told that in a Data Science role following a clear structure (like OSEMN) when producing your work is absolutely key, and you should start practising now. The impression you may get is that this is mainly to help you when you enter the world of work, some generic point in the future; however not so — the main reason to start now is **not** to practice for the future but it will help you a huge amount to be successful in your current activities. The intention of this blog is to give some tips that I found useful in structuring my projects.

## Decision Making Processes

For the purpose of this explanation I’ve used very simple decision-making process that we all use every day even if we do not realise it. This process notionally has 4 steps:

* Gather information
* Identify options
* Evaluate options
* Make decision

Decision making processes such as the one shown here or OSEMN can seem likely alien concepts however we use them every day, we just do not think of them exactly in this structure. An example in a sandwich shop would be as follows:

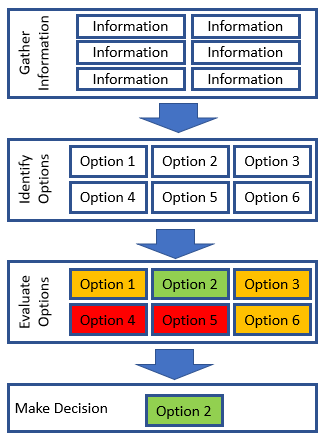


The sandwich shop menu



Sandwich you’ve been dreaming of?

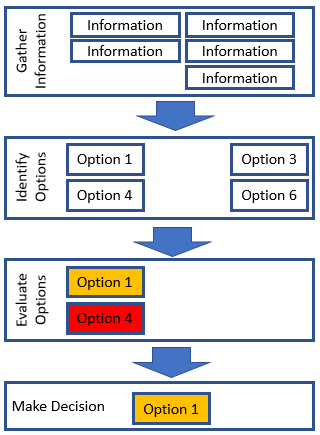
The natural assumption of how to tackle this process is very linear; completing the entirety of each stage in order.



Basic decision making process

## Process Fallibility

Sounds good, you’ve assessed all the information, identified all options — then evaluated them all and made a selection. However, the reality of the situation is more likely to be the following:



Basic process — not going so well

With the time pressures of your course / role / life the want is to move on to the next stage and it can be very challenging to thoroughly complete each stage — it is very easy to seize upon the items that you already have, which by their very nature are likely to be the ones most comfortable for you to understand and work with and move on to the next stage.

There are several flaws in the reality of this linear model:

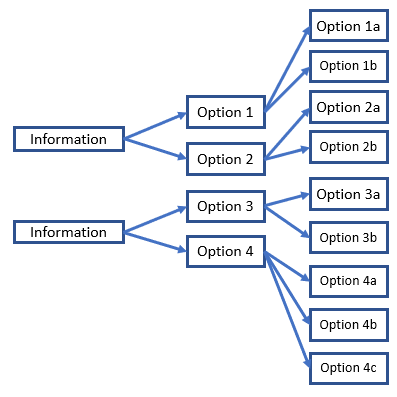
1) Failure to identify optimal solution

As highlighted in the example above if not all options are identified/evaluated then it is quite possible that a sub-optimal option will be selected because the more optimal solution is simply not known about.

2) Lack of knowledge of information/options to gather

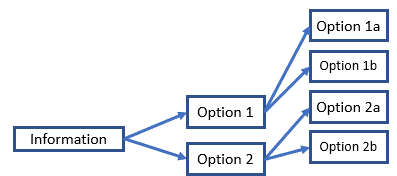
It is normal to enter into a project with far from perfect information, it is extremely challenging to gather ‘all’ (whatever all actually means) information at the start because its only when the problem is being explored that many of the options, areas of information required are identified.

Failing to gather information / options at an early stage has a disproportionate impact on the outcome of the project. Expanding on the over simplified model above, in reality there are likely to be many more options (think of the sandwich example) where there will be many sub-options. If in the sandwich shop they don’t usually have Turkey, if a cursory decision making process is used this may not be identified which will eliminate a whole genre of sandwiches from the process.



## Identifying all the options

If not all the information is gathered then an entire section could be eliminated, in this example not identifying piece of information two will eliminate 5/9ths of the options.



Identifying significantly less options

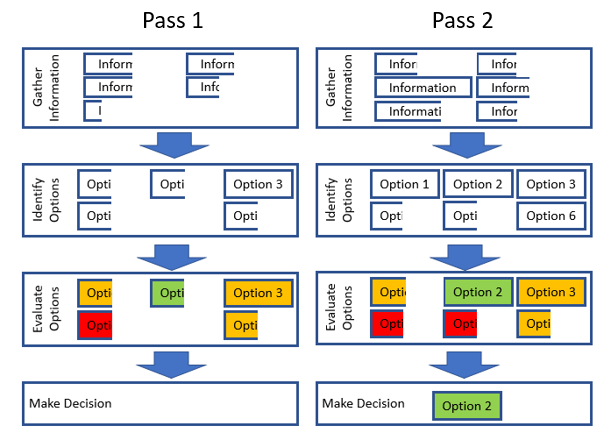
## So, what is the solution?!

An iterative process — this may seem like an obvious answer however without structuring how you complete this process it can end up taking a huge amount of time and be very inefficient. If a process doesn’t feel natural the what reliable happens is the user deviates or takes short cuts; a decision-making process is not effective if it is not user friendly.

Firstly, what are we trying to achieve in the decision-making process? We are not trying to get all the information on all topic areas, we are not trying to get all the complete detailed options we are trying to get an **overview** of all the areas/options to determine if there are worth taking further.

The solution in this case is to take “Passes” a through the decision process, with each pass exposing some additional detail, any number of passes can be completed. In each pass some or all of the steps in the process are completed as required.

If we consider work to equal the area of the boxes shown then the amount of work can be reduced significantly by taking an approach where only some of the information in each area is gathered, a reduced level of detail is gathered on each option. Then when the information / options are evaluated a decision can be made, ‘red’ options are discarded so no further detail is gathered, amber options may have additional detail gathered and green have all the detail gathered.

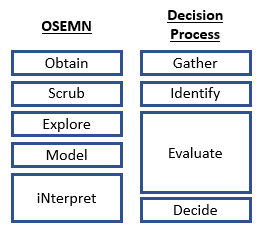


Decision process following an iterative approach

As can be seen in the diagram above all the options have been identified and assessed but a significantly reduced amount of work has been completed.

## Applying this to OSEMN

Roughly matching the decision process shown and OSEMN as follows.



OSEMN and Decision Process comparison

The main takeaway of this blog is the iterative part, OSEMN is inherently an iterative process — one which not all steps need to be completed in each pass. This is how I approached using OSEMN in a recent project.

I find it really important to understand why I am doing each pass of the OSEMN process otherwise it is easy to over or under complete each pass. I laid out clear objectives for each pass, so that I knew what I was trying to achieve.

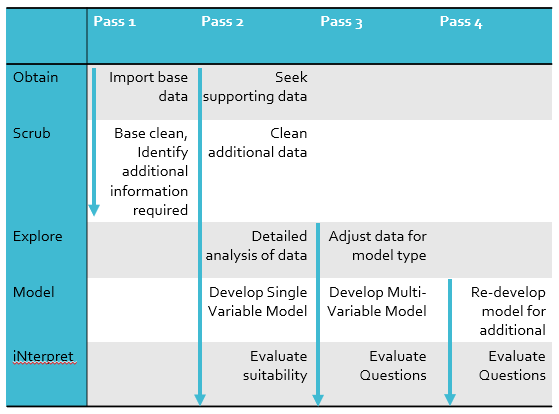
Objective for each pass:

1) Identify what additional information is required to make the dataset usable.

2) Clean and explore the data. Produce a bench mark model. Evaluate if model provides sufficient accuracy to answer questions (likely no — but if yes stop, we don’t want to complete more unnecessary work).

3) Further explore the data to ensure is appropriate for expanded scope model. Develop expanded scope model.

4) Re-work model to answer additional questions, or to provide greater detail for existing questions. Confirm project answers questions sufficiently or explains why they have not been answered.



Using OSEMN with an iterative approach

## Conclusion

Using a decision process such as OSEMN to framework your projects from the very beginning of your data science journey is not only great practice for your career but will also help you significantly in completing your projects effectively.

When using OSEMN (or others process) use the framework iteratively and be clear about the structure you are going to use.

Hopefully you found this helpful and your feedback would be useful!