



**Institute** of  
**Data**

---

2022



# Data Science and AI

---

## Documentation

---



# Why Document things?

- People easily forget what was said or agreed which can lead to misunderstanding, confusion and wasted time
- Trying to achieve an objective without documentation creates a very high risk of failure, job loss and reputational damage
- The quality of your thinking is demonstrated by the quality of your written documentation
- Other people will duplicate your work in the future if there is no documentation to rely on
- Great documentation results in project approval, promotion and remains an artefact of your success for years to come



# Remembering Your Objective

## DEFINE, DESIGN, DELIVER

- You are the solution
- **Define** the issue (problem or opportunity?)
- **Design** the solution
- **Deliver** the results



# 3Ds mapping to Documentation

Typical Document Sections	3xDs stage
Background & Objectives	Define
Assumptions	Define
User Stories and Requirements	Define
Open Questions/Out of Scope	Define
Designs, Mockups, Flow Charts etc	Design
Tasks	Deliver
Results and Conclusion	Define



# Writing Style - SAW!

## Short

- Sentences - use full stops. Avoid using “and”
- Paragraphs - break them up. Use bold, sub-headings & lists

## Active voice

- Active (subject performs action): *“Customers believed their account was suspended when they saw the alert”*
- Passive (subject receives action): *“It was believed by customers that their account was suspended when they saw the alert”*

## We focused

- Generally refer to “we” and not “I”



# When to Document

Waiting until the end of a project to write things up doesn't work.  
So record and document continuously:

- At the start of a project
- When stakeholder feedback has been obtained
- During project meetings
- If a project decision has been taken - record it!
- At the completion of a project
- After the results of a project have been obtained



# What to Document

- Project briefs - capture verbal briefs in writing and email
- **Project plans - continually maintain**
- Meeting minutes - attendees, agenda, decisions, tasks
- Designs - wire frames, flow charts, screen designs, architecture
- **User requirements - [persona, feature, reason]**
- Functional requirements - technical detail
- Tasks - person and due date
- Presentations - beginning, middle & end





# Where to Document

- Confluence
- MS Office
- Google Docs



# How to Document

- Use logical structure
- Always have a beginning, middle and an end
- Use a [project document template](#)
- Remember to use a SAW
- Remember to Define, Design and Deliver.



Institute of  
Data

Define



# Document Macro

Note: The following are the candidate sections of the document. They are presented here for guidance. Questions in each section could be used as possible aspects to cover. Some questions may not be applied to each project. On the other hand, additional information may be needed.

## Problem statement

- What is the problem or the opportunity that the project is investigating?
- Why is this problem valuable to address?
- What is the current state (e.g. unsatisfied customers, lost revenue)?
- What is the desired state?
- Has this problem been addressed by other research projects? What were the outcomes?



# Industry and domain

## Industry/ domain

- What is the industry/ domain?
- What is the current state of this industry? (e.g. challenges from startups)
- What is the overall industry value-chain?
- What are the key concepts in the industry?
- Is the project relevant to other industries?

## Stakeholders

- Who are the stakeholders? (be as specific as possible)
- Why do they care about this problem?
- What are the stakeholders' expectations?



# Business and Data Questions

## Business question

- What is the main business question that needs to be answered?
- What is the business value of answering this question? (quantify value and make necessary assumptions)
- What is the required accuracy? What are the implications of false positives or false negatives?

## Data question

- What is the data question that needs to be answered?
- What is the data required to answer the question?



# Data

## Data

- Where was the data sourced?
- What is the volume and attributes of the data?
- How reliable is the data?
- What is the quality of the raw data?
- How was this data generated?
- Is this data available on an ongoing basis?



# Data Science Process

## Data analysis

- What data pipeline was to wrangle the raw data?
- What are the highlights of the Exploratory Data Analysis (EDA)?
- Is the pipeline reusable? (for example, to process future data?)
- What are the intermediary data structures used (if any)?

## Modelling

- What are the main features used?
- Did you find any interesting interactions between features?
- Is there a subset of features that would get a significant portion of your final performance? Which features?
- How did you select features?
- What feature engineering techniques used?
- What are the models used?
- How long does it take to train your model?
- What are the tools used? (cloud platform, for example)
- What are the model performance metrics?
- Which model was selected?

## Outcomes

- What are the main findings and conclusions of the data science process?





# Answers

## Data answer

- Was the data question answered satisfactorily?
- What is the confidence level in the data answer?

## Business answer

- Was the business question answered satisfactorily?
- What is the confidence level in the business answer?



# Recommendations

## Response to stakeholders

- What are the overall message and recommendations to the stakeholders?



# Solution

## End-to-end solution

- What is the overall end-to-end solution to use the model developed in the project?



# References

- Where are the data and code used in the project? (show a simplified list of main items: notebooks, datasets, exported models)
- What are the resources used in the project? (libraries, algorithms, etc)



Institute of  
Data

Deliver

# How to deal with time estimates?

## “How long will it take?”

- Answer a question with a question - “When do you need it?”
- If an executive gives you a poorly defined request and demands to know how long it will take. You can respond by saying:
  - “I can design & deliver something that will meet your objectives fast, but I’ll need to define the project in more detail, then I can give you an exact timeframe....
  - “So tell me, what is the primary objective?”
- Projects should never be longer than 6-8 weeks
- If projects get too long then break them into phases
- A 3-month project is a warning that project definition might be poor or ill-conceived.



# Tasks

## Tasks

*List the tasks and ensure that every task has a person assigned to the task and the due date of the task is noted. Break large tasks into smaller ones.*



# Results & Conclusions

## Results and Conclusions

1. *Revisit the Background and Objectives of the project and report on whether and to what extent the project achieved the objectives.*
2. *Draw conclusions and demonstrate insight.*
3. *Present options for next steps and*
4. *make a recommendation about which option to follow.*





# Fail to plan, plan to fail

Typical Document Sections	3xDs stage
Background & Objectives	Define
Assumptions	Define
User Stories and Requirements	Define
Open Questions/Out of Scope	Define
Designs, Mockups, Flow Charts etc	Design
Tasks	Deliver
Results and Conclusion	Define





Institute<sup>of</sup>  
Data

# Presentations and Review