# 



## Data Driven Efficiency Improvement Proposal

Project charter

**Table of Contents**

[**Executive Summary**](#_n35gxf4p29p8) **2**

[**Problem/Opportunity Statement**](#_sw6lyfvrtn02) **3**

[Market Background [optional]](#_toj33paj4qcp) 3

[Problem Definition / Market Opportunity](#_bllaijbpt7n0) 3

[Stakeholders](#_d1fsix5ykmv2) 3

[**Solution Definition**](#_kgj0at8yt9p6) **4**

[Core Solution](#_f17cltkqx86t) 4

[Deployment](#_yf6cdjvwxznr) 4

[Supporting Systems](#_wy270n6t9gl1) 4

[Service Level Agreements [optional]](#_88dble32besh) 5

[Beyond Scope](#_yyn5vp9pt3tp) 5

[**Solution Benefit**](#_6fav4os0tger) **6**

[Success Measures](#_u21tgplbmx48) 6

[Tangential benefits](#_gp6jwvff30ky) 6

[Return on Investment [optional]](#_m68pronm5muj) 6

[**Project Approach**](#_v8u3r3p5hs4i) **7**

[Project Roadmap](#_dt3puxxk7q8m) 7

[Staff and Communications [optional]](#_m771dzbe0uyo) 7

[Infrastructure [optional]](#_fwo5lvi59v93) 7

[Data](#_v0evbjlowj1f) 7

[**Appendixes [circumstantial]**](#_r5ds74owlnzr) **9**

[Initial Backlog](#_nqd6vgijxf7i) 9

[Financial Assessment](#_ijkqvmga9pr7) 9

[Assumptions and Risks](#_fejqh15fxi0f) 9

[**Document Version Control [optional]**](#_7n0inayleh42) **10**

# 

# Executive Summary

*Provide the brief pitch to those who just want a very high view. Include the 4-6 things you want people to remember.*

With the

* Who are the stakeholders?
* What is the problem or opportunity?
* What is the solution?
* How long will the solution take to develop / what will it cost?
* What is the potential and expected value of the project?
* Provide other key info per the situation.

# 

# Problem/Opportunity Statement

*Get people to clearly understand and appreciate the fundamental item you’re trying to solve. “Problem” works best if you’re looking to improve a currently poor situation. “Opportunity” works best if you’re looking to expand or generate a good situation.*

## Market Background [optional]

*Think of the intended audience. Do they need to understand the broader market situation? Or are they already experts in this field?*

* What is the overall market being served?
* What are relevant recent or upcoming trends?

## Problem Definition / Market Opportunity

*What is the underlying problem we are trying to solve? / What is the underlying opportunity that we are trying to develop?*

* Remember the true underlying problem usually differs from the presenting asks.
* List out the high-level questions that the team will investigate during the project.

## Stakeholders

*A stakeholder is anyone who is impacted by the project. Think both broad (e.g. broader society) and internal (e.g. Legal, Networking, etc). Could describe the project champion and then leave the secondary stakeholders into a table.*

* What are their needs?
* What is their authority over the project?
* How receptive are they to the project?
* Should you train them on the data science life cycle or on agile practices?

# 

# Solution Definition

*Much of this can be omitted if this is an exploratory project with no current intended productized system. Pending organizational policies, you might need architecture diagrams and product specs prior to project approval which would land nicely into appendices.*

## Core Solution

*What is the core deliverable?*

* What is the product vision?
* What is the core deliverable (usually the output of an ML model)?
* How does this deliverable solve the underlying problem?
* What is the MVP? Combine a logical set of the most promising deliverables to define the Minimal Viable Product (i.e. the first major release)

## Deployment

*Consumption focuses on the fuller value stream from the overall system or the user’s perspective. Value stream mapping and use case diagrams can help. Deployment is from the internal product’s perspective of how it will be delivered.*

* How will the output be consumed?
  + Human consumption vs into another system
  + If for human consumption, how will the user interact with the model output?
* How will the deliverable be deployed?
  + Real-time vs batch vs ad hoc?
  + API needed?
  + File delivery into a target location?
  + Into a viz such as Tableau?

## Supporting Systems

*What are the supporting machine learning operational systems?*

* Data monitoring system
* Model monitoring system
* IT monitoring systems
* Maintenance plans

## Service Level Agreements [optional]

*Typically throw this into an appendix for detailed for contractual or mass-market products. If a non-critical, internal system deliverable and/or ad hoc models, SLAs might not be required but it’s still good to mention general availability.*

## Beyond Scope

*What should the team not do?*

* What are the anti-goals? (i.e. things you want to specifically avoid doing)
* What are beyond-scope items? (often this list is a good starting point for follow-up projects)

# 

# Solution Benefit

## Success Measures

*How will success be measured? Specific metrics are listed in* [*this post*](https://www.datascience-pm.com/9-ways-to-measure-data-science-project-performance/) *but broadly there are business, model, and system metrics.*

* Business impact metrics – Typically these are already existing metrics like financial metrics, subscriber churn, or mean-time-to-failure. Include goals if possible.
* Data science metrics – May or may not need to be defined upfront but think of your typical model performance metrics such as specificity, sensitivity, or lift.
* System metrics – Might be covered in the SLA section above.

## Tangential benefits

*Not every benefit can be measured. Include some indirect benefits*

* Organizational knowledge gained
* Team skillsets gained
* Reusable artifacts created (e.g. new data sets engineered can also be used for different use cases)

## Return on Investment [optional]

*Might need to be extensive with details in the appendix if seeking funding from an approval committee. Or could be skipped for smaller Proof of Concept or research projects*

* ROI analysis
* Cost-benefit analysis
* Break-even analysis
* CapEx vs OpEx assessment (key for some organizations)

# 

# Project Approach

## Project Roadmap

*Avoid Gantt charts. But do provide enough clarification on how the project could progress. A* [*project roadmap*](https://www.datascience-pm.com/example-roadmap/) *is recommended.*

* What is the sequence of possible deliverables?
* What are the go/no-go decisions to continue the project?
* What are the external influencers on the timelines? (e.g. an app launch date or a marketing campaign)

## Staff and Communications [optional]

*If already part of a product team that doesn’t shift with each project, then most of this can be skipped.*

* Staffing - Existing internal, Contractors, Partners, New hires?
* What are the anticipated hours or length of effort needed?
* Broader team members such as Security or Networking or Legal
* Communication – points of contact, setting up shared drives, chat groups, meetings cadences, etc
* Do project team members or stakeholders need training?

## Infrastructure [optional]

*Some teams might already admin privileges over their own servers or cloud environments. If so, this section could be skipped. Otherwise, ignoring this section could lead to significant blockers later in the project.*

* Licenses vs open-source
* Cloud vs on-prem services needed
* Requests needed such as new accounts or environments

## Data

*This is often the most unique aspect of data science vs. other project types.*

* What data do you need?
  + Do you already have the data?
  + Can you start collecting the data?
  + Can you purchase the data?
  + What metadata do you need (e.g. does the data need to be labeled?)
* Data privacy and environment
  + Classification: PII, CPNI, HIPAA, etc
  + Legal/ethical use: Can you legally or ethically use the intended data? What data can you *not* use
* How clean is this data?
  + Unless your team has already used the intended data source, exploratory data analysis is needed before understanding the data cleanliness.
* Big data considerations
  + What is data Volume (size)?
  + What is the data Variety (formats?)
  + What is the data Velocity (rate of influx?)
  + How can we process and store this data?

# 

# Appendixes [circumstantial]

*The appendices needed are highly circumstantial. Generally, every appendix used should support the statements in the body of the document.*

## Initial Backlog

* Stack rank the most promising deliverables. Often focus on the most valuable deliverables relative to their effort.
* Avoid dependencies among deliverables. But if unavoidable, note these dependencies.

## Financial Assessment

* Break-even analysis calculations
* Cash flow analysis chart
* ROI analysis calculations

## Assumptions and Risks

* What key assumptions have you made in this document?
  + What is taken as a given and beyond the project’s ability to influence?
  + What are variables that you will test as part of the project (i.e. testable hypotheses)?
* What are key dependencies?
* What are key risks? [risk assessment matrix recommended]
  + Security (e.g. how do I ensure the data or model are not compromised?)
  + Legal (e.g. will data privacy law changes impact our ability to use the intended data?)
  + Ethics (see the [Data Science Ethics](https://www.datascience-pm.com/10-data-science-ethics-questions/) post)
  + Business (e.g. market changes, stakeholder turn-over)
  + Resources (e.g. team member turn-over, shifts in funding allocation)
  + Technical (e.g. data, modeling, computing availability)

# 

# Document Version Control [optional]

|  |  |  |
| --- | --- | --- |
| Date / Version | Edited By | Changes |
|  |  |  |
|  |  |  |