Table of Contents

[General Project Information 3](#_Toc2093596)

[Business background 3](#_Toc2093597)

[Business Goals & Objectives 3](#_Toc2093598)

[a) The problem statement 3](#_Toc2093599)

[b) Why Does the Problem need to be solved 3](#_Toc2093600)

[c) How could the problem be solved manually? 4](#_Toc2093601)

[d) State of Current Solution if Any 4](#_Toc2093602)

[Scope & Deliverables 4](#_Toc2093603)

[a) Objectives 4](#_Toc2093604)

[b) Requirement Specifications 5](#_Toc2093605)

[a. Business Success Criteria & Metrics 5](#_Toc2093606)

[b. legal and security obligations 5](#_Toc2093607)

[c. Communication 5](#_Toc2093608)

[c) Major Deliverables 5](#_Toc2093609)

[Inventory of Resources 6](#_Toc2093610)

[a) Hardware 6](#_Toc2093611)

[b) Knowledge Sources & Data 6](#_Toc2093612)

[c) Personnel 6](#_Toc2093613)

[Assumptions, constraints & Risks 7](#_Toc2093614)

[a) Assumptions 7](#_Toc2093615)

[b) Constraints & Limitations 7](#_Toc2093616)

[c) RISKS & CONTINGENCIES 7](#_Toc2093617)

[d) External/internal dependencies 7](#_Toc2093618)

[e) Expectations 8](#_Toc2093619)

[Cost & Commercial Estimates 8](#_Toc2093620)

[Project Plan 8](#_Toc2093621)

[a. Timeline & Duration 8](#_Toc2093622)

[b. Executive Milestones 9](#_Toc2093623)

[Preliminary Solution Architecture 9](#_Toc2093624)

[Data Mining & Analytics Goals 9](#_Toc2093625)

[Appendix A: Glossary of Terms Used 11](#_Toc2093626)

[Appendix B: Legal and Security 12](#_Toc2093627)

Project Charter

|  |  |
| --- | --- |
| General Project Information | |
| Project Name: |  |
| Business Name: |  |
| Department Name: |  |
| Project Manager: |  |
| Date Prepared: |  |

# Business background

* Explain the business situation that drives the project. Also mention the specific industry the business operates in.
* Who is the client, what business domain the client is in.
* What does the Customer Really Want to Accomplish?
* Define the business units which are affected by the data mining project.

# Business Goals & Objectives

## The problem statement

* What business problems are we trying to address? Define the problem firstly in layman's terms and then formalize it?
* Identify the problem area (e.g., manufacturing control, CRM, business development, etc.).
* Lessons learned from a similar problem in the past. If any?
* How is the problem affecting the core company business as well as any Business units directly affected by it.

## Why Does the Problem need to be solved

* Describe the motivation for solving the problem.
* What business need will be fulfilled when the problem is solved? How is the problem affecting the Business or any of its functions.? e.g. Is this a "learning Exercise" or based on "SWOT" Analysis.
* Describe the probable benefits of the planned solution and how it may be used.
* What will the client achieve by solving the problem. Such as any capabilities it may enable.
* Just a few lines about the possible shape the solution may take. It will affect nature and overall requirements for the project deliverables. Whether the end solution is a report or a full fledge Software app.

## How could the problem be solved manually?

* This step basically tries to figure out whether what level of AI is required or if needed at all and whether the problem can be solved manually. This section is also meant to reveal critical information such as Where data may be stored, Types of features, etc.
* Describe how the problem is currently solved (if at all).
* Describe how a subject matter expert would make manual predictions.
* List out step-by-step what data you would collect, how you would prepare it and how you would design a program to solve the problem. This may include prototypes and experiments you would need to perform which are a gold mine because they will highlight questions and uncertainties you have about the domain that could be explored.
* It is recommended that this particular section is filled in by a domain matter expert with appropriate level of Data Science expertise.

## State of Current Solution if Any

* Analyze the current solution, its advantages and disadvantages, and the level to which it is accepted by the users in case the project is the result of a previous project.

# Scope & Deliverables

* What data science solutions are we trying to build?
* How is it going to be consumed by the customer?
* Identify target groups (e.g., a report for top management or an operational system to be used by naive end users).
* What are the Client expectation beyond the defined Problem?

## Objectives

Formalize the Objectives as per the previous Section's Problem Space Definition.

* [Example: Improving Detection of Fraudulent Credit Card Transaction from current x% to y%.

List down EACH Objective below:

* **Objective 1**
* **Objective 2**
* **Objective 3**
* *..........*

## Requirement Specifications

Requirements will include Clear definition of Success Criteria in terms of Qualitative and Quantitative metrics, specific communication and reporting requirements, legal and security obligations.

### Business Success Criteria & Metrics

* Describe a set of Qualitative Objectives (e.g. Increase Sales Volume) & criterion that may be mentioned by the client for a successful outcome to the project. Also Identify the project owner and who assesses the success criteria.
* Define how the results will be measured. (e.g. A/B test on a specified subset for a specified period; or comparison of performance after implementation to baseline). At least get agreement on exactly who will judge whether or not those criteria have been fulfilled.
* What is a quantifiable metric (e.g. reduce the fraction of users with 4-week inactivity)
* What is the baseline (current) value of the metric? (e.g. current fraction of users with 4-week inactivity = 60%)

### legal and security obligations

List down all Legal & security related aspects pertaining to the provisioning of the services in an Appendix and provide the link for the Appendix here.

### Communication

* How will we keep in touch? Weekly meetings?
* Who are the contact persons on both sides?
* Is there are requisite reporting involved ?

## Major Deliverables

The following table presents the major deliverables that the project’s product, service or result must meet in order for the project objectives to be satisfied.

| Major Deliverable | I Deliverable Description |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

# Inventory of Resources

A list of all resources available for the project including personnel, data (data bases, warehouses, data exports in charts, etc.), hardware, and software (data mining tools and other relevant software). Identify and involve key people in the business and their roles.

## Hardware

Identify base hardware and establish its availability for the data mining project.

## Knowledge Sources & Data

Identify data and knowledge sources, as well as their type (on line sources, experts, written documentation, etc.)

* Is data available?
* Where will the data come from?
* In What form is the data Stored?
* Check available tools and techniques.

## Personnel

This section includes individuals who will be a part of the project in any manner. This list should cover everyone including ML Engineers & Data Scientists, but also those with expert knowledge of the business problem, data managers, technical support, and others. Basically anybody who has any direct bearing on the output of this project.

* Identify systems’ administrator, database administrators, and technical support staff for further questions.
* Identify market analysts, data mining experts, and statisticians, and check their availability.
* Check availability of domain experts for later phases.
* Identify the project owner and who assesses the success criteria.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team | Name | Role | Email | Phone |
| Client Team |  | Data Admin |  |  |
| Client Team |  | Business Dev |  |  |
| Vendor Team |  | Project Lead |  |  |
| Vendor Team |  | PM |  |  |
| Vendor Team |  | Data Scientist |  |  |
| Vendor Team |  | KAM |  |  |
| .............. |  |  |  |  |
| ............. |  |  |  |  |
| ............. |  |  |  |  |

# Assumptions, constraints & Risks

## Assumptions

These may be rules of thumb and domain specific information that you think will get you to a viable solution faster. Clarify assumptions (including implicit ones) made by the project on data quality (accuracy and availability),external factors (economic issues, competitive products, technical advances), but also assumptions that lead to any of the estimates, and general constraints (legal issues, budget, timescales, and resources). Any Assumptions related to Data, Sources of Data etc.

## Constraints & Limitations

Constraints may be on the availability of resources, but may also include technological constraints, such as the size of dataset that it is practical to use for modeling.

* Check access rights to data sources (e.g., access restrictions or passwords required), and accessibility.
* There may be constraints related to operating systems, data management system, file or database format.
* Check whether relevant knowledge is accessible.
* Check budget constraints (fixed costs, implementation costs, etc.)

## RISKS & CONTINGENCIES

* Identify causes that could delay completion of the project, and prepare a contingency plan for each of them. For example, if an Internet outage in your office could pose a problem, perhaps your contingency could be to work at another office until the outage has ended.
* Identify business risks (competitors), organizational risks (e.g., if department requesting project doesn’t have funding for it), financial risks (e.g., further funding depends on initial data mining results), technical risks and other risks that depend on data and data sources (e.g., poor quality and coverage).

| **Risk** | **Mitigation** |
| --- | --- |
|  |  |
|  |  |

## External/internal dependencies

List Down any external or internal dependencies either on the business or the Vendor Side which may have a bearing of the project delivery

## Expectations

Identify users’ needs and expectations.

# Cost & Commercial Estimates

This section provides a summary of estimated & actual spending to meet the objectives of the project as described in this project charter. This summary of spending is preliminary, and should reflect costs for the entire investment lifecycle.

**(@ Azhar, We may Skip this section until later...)**

# Project Plan

### Timeline & Duration

This section should include various Phases, activities etc that will be undertaken to deliver a successful project.

10/05

12/05

02/06

05/06

System Development

Completed

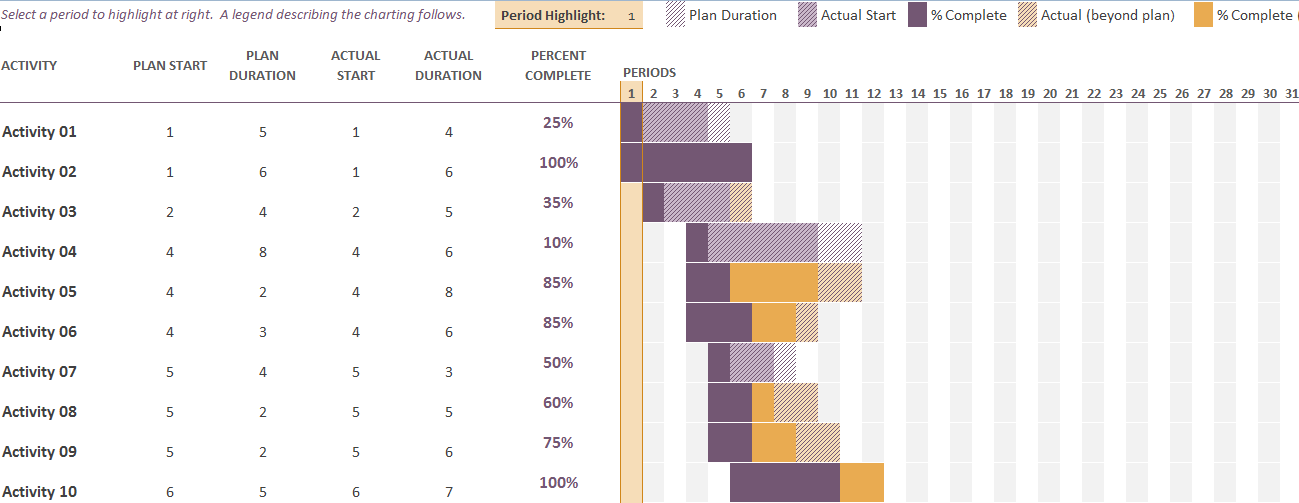
Developed Prototype

Requirements Analysis

Completed

Project Plan Completed

**OR**



### Executive Milestones

The table below lists the high-level Executive Milestones of the project and their estimated completion timeframe divided across various Phases.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Phase 1** | | **Phase 2** | | **Phase 3** | | **Phase 4** | |
| Milestone | Milestone Timeline & Duration | Milestone | Milestone Timeline & Duration | Milestone | Milestone Timeline & Duration | Milestone | Milestone Timeline & Duration |
| Activity 1 | | Activity 1 | |  | |  | |
| Activity 2 | | Activity 2 | |  | |  | |
| Activity 3 | |  | |  | |  | |
| Activity 4 | |  | |  | |  | |
| Activity 5 | |  | |  | |  | |
|  | |  | |  | |  | |

# Preliminary Solution Architecture

This section is technical and ought to be filled in by a domain matter expert.

* Data Engineering Design: This should try to address design requirements related to ETL, storage, ingestion, logging, tools and techniques that will be used. Whether on premises or cloud solution will be used. How will the Data movement will be handled such as using Flume, Kafka, Airflow or Nifi.
  + Also What type of Analytics resources may be put to use in the solution.
    - HDI/Hive/R/Python for feature construction, aggregation and sampling
    - AzureML for modeling and web service operationalization
    - TensorFlow or Pytorch or another choice of frameworks.
    - Whether Spark or Mapreduce will be used or something different altogether such as Apache Flink/Storm.
* Data Quality Management: This should explain the current level of data sanity and what may be done with the data to get it in required shape. Whether some/detailed aggregations, scaling, transformations are needed etc.
* How will the Solution by visualized.
* How will the solution by deployed and it will go into production. How it may fit within the existing deployed solution.
  + How will the customer use the model results to make decisions?
  + Make a 1 slide diagram showing the end to end data flow and decision architecture.

## Data Mining & Analytics Goals

This section may present a more technical aspects of Goals & Objectives from ML and Data mining perspectives incorporating CRISP-DM phases and steps such as:

* Feature Engineering
* Feature Selection and Extraction
* Exploratory Data Analysis techniques
* Model Building and Evaluation etc.

# Appendix A: Glossary of Terms Used

1. **Business Terminology**
2. **Data Science Related Terminology**

# Appendix B: Legal and Security