Objective

This document outlines the approach for the business understanding phase, the first milestone, guided by the CRISP-DM methodology. The primary goal is to clearly define the project from a business perspective, ensuring alignment with organisational goals and stakeholder needs before any data collection or analysis begins

Problem Statement:

The problem the business faces is to leverage data from various health and social indicators to gain actionable insights for a government or an organisation. The project will aim to address critical public health and social challenges, such as reducing child mortality, improving access to healthcare or enhancing vaccination rates. The problem is often complex, requiring a data-driven approach to identify key contributing factors and vulnerable populations

Business Objective

The project has multiple objects each driven by the need for accurate data.

- Identify key drivers of specific health outcomes
- Pinpoint underserved or at-risk population based on geographical, social or economic data
- Inform strategic decision-making for resource allocation, program development and policy implementation.
- Create a robust data foundation that can be used for ongoing monitoring and evaluation of public health initiatives

Stakeholders

Stakeholders that will be interested in this project are those who have a vested interest in the project. People who are directly affected by the public health issues the project aims to address. For starters, the government would be a stakeholder, because the government's civilians are affected. The stakeholders are as followed:

- Government health officials: They need the data to enforce policy and monitor program effectiveness

- Non-profit program managers: They require insights to optimise the delivery of aid and services.
- Field workers and frontline staff: They need easy-to-understand information to guide their day-to-day operations and outreach efforts.
- Data analysts/scientists: The project team responsible for the execution.
- Donors/funding organisations: They need to see evidence of impact to justify their investment

The stakeholders will have varied requirements. Interviews or workshops would be required. Questions may include:

- Specific questions for example: “Which region have the lowest immunisation rates”.
- What their desired results would be
- What timeline do they desire and what is their budget
- How they would prefer their data to be used, with other words, data privacy and the security concerns

Inventory of Resources

There is mostly 4 resource topics to consider, namely: The datasets, software and tools, personnel and finally the computing resources

Datasets

The primary data resources for this project are the provided datasets, which include:

- Access to health care
- Child mortality
- HIV behaviour
- Immunisation
- Toilet facilities
- Water
- Covid-19 Prevention
- Maternal Mortality
- Literacy
- ARI Symptoms

- Anthropometry
- Infant and young child feeding

Software and Tools

The project only requires 2 tools, but proficiency is expected in these fields. R, a programming language used for statistical computing and data visualisation, will be the primary tool for data analysis, statistical modelling and scripting. Power BI, primarily used for business intelligence and data visualisation, is the tool which is going to be used for creating interactive dashboards and visualisation to present findings to the stakeholders

Personnel

The core project team will consist of data analysts and data scientists with expertise in the CRISP-DM methodology, R programming, data visualisation and domain knowledge in public health or social issues.

Computational Resources

The team will require sufficient computing power to handle the datasets, which may involve using local machines or cloud-based platforms for data processing and analysis.

Risk, assumptions and constraints

A thorough analysis and understand of potential risks, key assumptions and project constraints is crucial for effective project management and to set realistic expectations

Risks

There is various risk involving this project, but the main 4 would be: Data quality issues, stakeholder disengagement, technical challenges and ethical and privacy concerns. To discuss said risk in further detail:

- Data Quality issues: The provided datasets may contain significant missing value, inaccurate, or inconsistencies which could impact the reliability o the analysis.
- Stakeholder disengagement: A lack of consistent feedback or engagement from key stakeholders could lead to misaligned project outcomes.
- Technical Challenges: The integration of disparate datasets may prove more complex than anticipated, leading to delays.

- Ethical and Privacy concerns: Handling sensitive public health data requires carful adherences to data privacy regulations, which could impose restrictions on data usage and sharing.

Assumptions

- Data Sufficiency: It is assumed that the provided datasets contain sufficient information to address the business problem and achieve the stated objectives
- Resource availability: We assume that all necessary software, tools and computational resources are readily available to the project team.
- Stakeholder buy-in: We assume that stakeholders have a strong interest in the project's success and will provide timely feedback.

Constraints

The project must be completed within the specified timeline and budget, creating a fixed timeline and budget constraint. A data access limitation constraint is created due to the analysis being constrained to the provided datasets and external data sources may not be available or accessible. The project is limited to using R or analysis and Power BI for visualisation as specified in the project outline

Success Criteria

Success will not be measured solely by the accuracy of a model but by its impact on the organisation's mission. The success criteria will be defined in collaboration with the stakeholders and may include measurable outcomes, the adoption of the data, the efficiency and the actionability. To go further into the criteria:

- Measurable outcomes: An example would be a 10% increase in childhood immunisation rates in a targeted region within an year of implementing the project's recommendations.
- Adoption: The Power BI dashboards and reports are actively used by at least 80% of identified stakeholders.
- Efficiency: The insights provided lead to a 15% more efficient allocation of resources.
- Actionability: The project's findings directly influence and guide the launch of a new public health program.

Importance of CRISP-DM

The CRISP-DM methodology is fundamentally important to the success of this project because it provides a structured, repeatable and well-defined roadmap from start to finish. Rather than a linear, one-way process, it functions as an iterative cycle that ensures every technical step is directly linked to the initial business objectives. This structured approach helps to mitigate common project risks, such as solving the wrong problem or producing an analysis that is not actionable. By beginning with a thorough “Business Understanding phase, the methodology forces a deep dive into stakeholder needs and success criteria, preventing the team from wasting time on technically elegant but irrelevant insights. Each subsequent phase, from data understanding and preparation to modelling and evaluation, builds upon this foundation, allowing for continuous feedback and refinement. Ultimately, CRISP-DM ensures the project’s output is not just a report or a dashboard, but a tangible solution that directly contributes to the organisation’s mission and provides a clear path for ongoing monitoring and improvement.