# Business Understanding

## Background

South Africa face ongoing challenges related to health and living conditions, which include:

* Limited access to clean water and sanitation facilities.
* Unequal access to healthcare facilities across different provinces.
* High child and maternal mortality rates in disadvantaged communities.
* Gaps with both immunization coverage and HIV prevention behavior.
* Education and literacy levels influence healthcare awareness.

The government and non-profit organizations need to improve public healthcare outcomes, but data-driven insights are needed to guide these interventions, allocate resources effectively and identify the high-risk areas.

## Business Problem and Objectives

**Business Problem:**

Despite numerous past interventions, vulnerable populations in South Africa still experience poor healthcare outcomes due to systematic issues like sanitation, lack of clean water and healthcare access. Stakeholders require reliable, data-driven insights to understand the relationships between living conditions, healthcare access and healthcare risks.

**Business Objectives:**

* Provide actionable insights that stakeholders can use to target interventions more effectively.
* Discover the key factors that are currently contributing to poor health outcomes, for example clean water, sanitation, literacy etc.
* Develop interactive dashboards and reports that make complex data easy to understand.
* Identify high-risk regions and communities based on demographic and health indicators.

## Stakeholders and Their Requirements

|  |  |  |
| --- | --- | --- |
| **Stakeholder** | **Requirements** | **Example KPI** |
| Government (Department of Health) | Identify provinces with the high health risks and resource gaps. | Reduction in child mortality rates (%). |
| NGOs/NPOs | Prioritize regions for campaigns (HIV prevention, sanitation, immunization) | Increase in immunization coverage (%). |
| Healthcare Providers and Clinics | Understand the population needs and healthcare access issues. | Clinic to population ratio. |
| Researchers/Academics | Access to structured datasets and insights for policy studies. | Number of published findings. |

## Success Criteria

**Analytical Success:**

* Build models (classification, clustering, association rules) that achieve acceptable performance (for example, prediction accuracy > 70%).
* Identify meaningful clusters with similar healthcare challenges.
* Extract meaningful rules (for example, low literacy + no sanitation = high ARI risk).

**Business Success:**

* Insights led to practical recommendations for healthcare and community interventions.
* Dashboards highlight priority areas for intervention (geographic or demographic).
* Stakeholders confirm results are useful for guiding decision-making.

## Inventory of Resources

* **Datasets:** 13 provided datasets (access to healthcare, literacy, immunization etc.).
* **Tools:** R and R Markdown (for analysis and modelling) and Power BI (for dashboards).
* **Team Resources:** Group of 5 students.
* **Infrastructure:** RStudio, PowerBI, central repository and task manager via GitHub.

## Risks, Assumptions and Constraints

* **Risks:**
  + Missing or incomplete data in the datasets.
  + Regional disparities (like rural vs urban) may skew insights.
  + Time limitations for full data cleaning and modelling.
* **Assumptions:**
  + Provided datasets are representative of the population.
  + Stakeholders will use the generated insights to improve healthcare policy.
* **Constraints:**
  + Project is limited to the provided datasets, as there is no primary data collection.

## Data Mining Goals and Success Criteria

* Apply clustering to group regions by their risk factors.
* Apply classification models to predict child mortality or healthcare access.
* Apply association rule mining to uncover hidden patterns between literacy, sanitation and disease.
* Validate and assess models by making use of accuracy, F1-score or Silhouette index.

## Importance of CRISP-DM

Using CRISP-DM provides a structured, industry standard approach to guide this project and its progress.

* Makes the process repeatable and transparent for stakeholders.
* Ensures systematic alignment from problem definition to deployment.
* Encourages iteration (revisiting earlier steps).