**Problem Statement**

Analyse the relationship between anemia diagnosis, hemoglobin (Hb) levels, and pixel color distribution (red, blue, green) to uncover patterns and insights about gender-based differences, pixel data, and Hb trends in individuals diagnosed as anaemic versus non-anaemic.

**Data Requirements**

Demographics:

Gender (Male/Female).

Biological Metrics:

Haemoglobin levels (Hb).

Diagnosis status (Anaemic: Yes/No).

Image Processing Metrics:

Percentage values for Red, Green, and Blue pixels (%Red Pixel, %Green Pixel, %Blue Pixel).

**Counts:**

Number of individuals per category (by gender and anaemia diagnosis).

**Data Collection**

**Source:**

Haemoglobin and anaemia diagnosis data collected via blood tests.

Pixel colour values derived from image processing techniques applied to blood smear samples.

Gender information collected through demographic questionnaires or records.

**Collection Methods:**

Lab instruments for haemoglobin measurements.

Medical imaging software for pixel data analysis.

Manual or automated systems for recording demographic data.

Tools Used: Medical testing devices, imaging software, and spreadsheets/databases.

Data Validation

To ensure data quality and accuracy:

**Completeness:**

Verify no missing values for Hb, pixel percentages, and anaemia diagnosis.

**Accuracy:**

Cross-check haemoglobin values against standard lab reference ranges.

Range Checks:

Confirm pixel percentages fall between 0%–100%.

Ensure Hb values are biologically realistic (e.g., Hb > 0).

**Consistency:**

Validate consistency between anaemia diagnosis and Hb values (e.g., low Hb aligns with "Yes" in Anaemic column).

Outlier Detection:

Use statistical methods to identify abnormal pixel or Hb values for review.

Data Cloning

If necessary:

Purpose: To create subsets of the dataset for testing, analysis, or model validation.

Steps:

Clone the original dataset for exploratory analysis.

Create filtered subsets (e.g., data for males only, or anemic individuals).

Ensure cloned datasets maintain integrity and are labeled properly to avoid confusion.

Tools

Power BI:

For creating visualizations and dashboards.

Excel/Google Sheets:

For initial data cleaning and validation.

Python/R:

For advanced data cleaning, statistical validation, and outlier detection.

Image Analysis Software:

Tools like ImageJ or OpenCV for deriving pixel data from medical images.

Storytelling

Key Insights from the Visuals:

Haemoglobin Trends:

The average Hb level for anaemic individuals (2.80) is significantly lower than non-anaemic individuals (11.27), clearly distinguishing the two groups.

Pixel Data and Hb Correlation:

The top bar chart shows a clear relationship between blue pixel percentage and anaemia diagnosis, with blue pixel values clustering around specific ranges for anaemic individuals.

Gender-Based Insights:

Sum of Hb and %Red Pixel by Sex:

Females show higher values for both Hb and %Red Pixel than males on average.

Anaemia Prevalence by Gender:

The stacked bar chart highlights more females being diagnosed as anaemic compared to males.

Gender Proportion:

The pie chart reveals that females (64.31%) dominate the dataset, which may influence other findings.

**Summary Card (KPI):**

The average Hb levels indicate a stark biological difference between anaemic and non-anaemic groups.

**Storyline:**

The visuals show a clear differentiation between anaemic and non-anaemic groups based on haemoglobin levels and pixel percentages.

Gender appears to play a significant role, with females both representing a larger portion of the dataset and being more frequently diagnosed as anaemic.

The patterns in pixel data (e.g., blue pixel percentages) suggest that image analysis can serve as a supplementary diagnostic tool for anaemia.