**Data Set (Agriculture Science)**

**Introduction:**

These datasets (phenotypic and genotypic data) can be used in the fields of genetics and plant breeding. Molecular breeding strategies such as genome-wide association studies (GWAS) and genomic prediction have revolutionized crop improvement by enhancing selection accuracy and genetic gains. Based on these datasets, it is possible to associate the phenotype of the DH lines from the Germplasm Enhancement of Maize (BGEM) and their testcross hybrids, helping to elucidate the genetic basis of oil content and fatty acid composition, as well as to predict superior hybrids and breeding populations. The phenotypic data set represent the field perfomance of the BGEM lines and testcross hybrids for oil content (oil, %) and ten fatty acids traits: palmitic acid (16:0, %), palmitoleic acid (16:1, %), stearic acid (18:0, %), oleic acid (18:1, %), linoleic acid (18:2, %), linolenic acid (18:2, %), arachidic acid (20:0, %), gadoleic acid (20:1, %), docosanoic acid (22:0, %) and tetracosanoic acid (24:0, %). On the other hand, the genotypic data set are GBS markers.

**Problem Statement:**

The goal of this research is to leverage the phenotypic and genotypic datasets from the Germplasm Enhancement of Maize (BGEM) to identify and understand the genetic factors influencing oil content and fatty acid composition in maize. The datasets, which include field performance data for oil content and ten specific fatty acid traits, as well as genotypic data obtained through Genotyping-by-Sequencing (GBS), will be used to:

1. **Identify genetic markers** associated with oil content and fatty acid composition, particularly focusing on palmitic acid (16:0), palmitoleic acid (16:1), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), linolenic acid (18:3), arachidic acid (20:0), gadoleic acid (20:1), docosanoic acid (22:0), and tetracosanoic acid (24:0).
2. **Develop genomic prediction models** to predict superior hybrids and breeding populations based on genotypic and phenotypic data, enhancing the efficiency of breeding programs and increasing selection accuracy.

This study aims to contribute to the development of molecular breeding strategies that can accelerate maize improvement for desired oil and fatty acid profiles.