# Social Issues in Egypt Dataset Analysis

## Project Overview:

This project focuses on analyzing five major social issues in Egypt: Illiteracy, Women & Divorce Issues, Homelessness, Addiction, and Public Health challenges.
The main idea of the project is fully inspired by **Egypt's Vision 2030** and the **Sustainable Development Goals (SDGs)**, as these social issues directly affect national development, youth stability, and community well-being.

We aimed to explore how these issues have evolved over recent years, understand their root causes, and visualize the findings in an organized and interactive way using Power BI.
Our analysis contributes to several strategic goals, including Quality Education, Gender Equality, Good Health & Well-Being, No Poverty, and Sustainable Cities & Communities.

Our work included gathering diverse datasets, conducting a public survey, cleaning and preparing the data, building a robust data model, and designing full dashboards in both light and dark themes.
To complete this project, our team followed a structured data-analysis workflow to ensure accuracy, reliability, and meaningful insights aligned with Egypt's national development priorities.

### 1- Data Collection

At the beginning of the project, we spent a significant amount of time collecting datasets related to the five selected issues.
The data came from different online sources, government portals, open-data websites, research reports, and statistical platforms.

Because the topics were highly diverse, we had to explore many sources until we found data that supported our analysis.

We also designed and distributed a **survey** to collect public opinions regarding the five issues.
The survey helped us understand people's awareness levels, perceptions, and which issues they believe are most urgent.

## 2- Data Cleaning

Each team member was responsible for cleaning one of the issues' datasets.

We used several tools during the cleaning phase:

- Excel
- Power Query
- Python

The cleaning process included:

- Handling missing and inconsistent values
- Standardizing columns
- Fixing incorrect data types
- Removing duplicates
- Preparing the data for modeling and visualization

This step took effort because the datasets were collected from multiple different sources, each with different formats and structures.

## 3- Data Modeling

After cleaning the data, we built a professional data model using Power BI.

The modeling phase included:

- Designing a **star schema**
- Building fact and dimension tables
- Establishing proper relationships
- Creating calculated measures and KPIs
- Structuring the model to support drill-down and filtering

Good modeling ensured that the dashboards were accurate, efficient, and easy for users to explore.

## 4- Data Visualization

This was one of the most detailed and creative stages of the project.

We built a full Power BI reporting system, including:

### ● Main Navigation Page

A central home page linking all dashboards inside the project.

### ● Overview Dashboard

A comprehensive, high-level dashboard summarizing all five issues together.
It provides cross-issue comparisons, multi-year trends, and general patterns.

### ● KPI Dashboard

A dedicated dashboard for the most important indicators across all issues, displaying:

- Top KPIs
- Governorate comparisons
- Yearly trends
- Highlighted anomalies
  This dashboard allows users to understand the situation at a glance.

### ● Individual Issue Dashboards

Each issue has a dedicated dashboard with KPIs, insights, trends, and charts:

- Illiteracy / Education
- Women & Divorce
- Homelessness / Begging
- Addiction
- Healthcare

### ● Survey Dashboard

Analyzes the responses collected from our public survey, offering direct insights into citizens' perspectives.

### ● Light & Dark Modes

All dashboards were designed in both Light and Dark themes for better UI flexibility.

### ● Enhanced Interactivity

We applied several advanced BI techniques:

- Full drill-down and drill-through
- Hover-based tooltips
- Dynamic filters & slicers
- Natural-language Q&A visual
- Buttons & bookmarks for smooth navigation

### ● Embedded Awareness Videos

Each dashboard contains an image linking to a related YouTube awareness video, supporting social impact and increasing user engagement.

### ● Row-Level Security (RLS)

Ensures that each user only sees the data they are authorized to access.

## 5-K-Means Clustering Model

To enhance the analytical depth of the project, we developed an **Unsupervised Machine Learning model (K-Means Clustering)** that groups the governorates into meaningful clusters based on:

- Risk levels
- Social indicators
- Health metrics
- Education levels
- Addiction and homelessness patterns

The clustering results were visualized through:

- **Maps** to show geographical distribution
- **Cluster summary tables** presenting the characteristics of each group

This unsupervised model helped reveal hidden patterns, identify **high-risk governorates**, and provide insights that support **data-driven decision-making**.

# Analytical Questions Explored

In which year did divorce cases reach their highest level?

What is the total number of divorce cases?

How have total divorce cases changed over the years?

What is the distribution of divorce cases between urban and rural areas?

What is the divorce rate per 1000 people?

Which governorate has the highest divorce cases?

How are divorce cases distributed across governorates?

In which year did homelessness issues peak?

Which governorate has the highest homelessness rate?

How are homelessness cases distributed across age groups?

What are the main causes of homelessness?

What is the average number of homelessness cases?

Which location reports the highest daily income among homeless individuals?

How do estimated homelessness cases vary across governorates?

What is the average number of teachers in the education system?

What is the overall illiteracy rate?

How do literacy, dropout, and enrollment trends change across years?

Where does the education system fall behind the most?

What is the gender gap in educational attainment?

What is the secondary enrollment rate?

What is the primary enrollment rate?

What is the youth literacy rate?

What is the average number of schools?

What is the total number of students?

How do primary dropout rates change over issues and years?

What is the most frequent societal issue?

In which year did societal issues reach their peak?

Which governorate has the highest prevalence of issues?

What is the highest homelessness year?

What is the total number of addiction cases?

What is the highest divorce year?

What is the overall illiteracy rate?

What is the obesity rate?

What is the addiction relapse rate?

What is the total number of divorce cases?

What is the total number of students?

What is the depression rate?

What issue do people consider the most widespread?

What is the most common cause of societal problems according to people?

What issue represents the highest stability concern?

Which issue has the highest family impact?

Which issue is considered the most psychologically damaging?

How are individuals distributed by gender?

Which issue most affects each age group?

Which age group is most affected by addiction?

What is the overall expected worsening trend of issues?

How are people distributed across age groups?

# Overall Summary

This project provided a comprehensive look into five of the most critical social challenges affecting Egypt today.
By bringing together data from multiple official sources, public perceptions collected through our survey, and advanced analytics such as K-Means clustering, we were able to uncover real patterns, trends, and risk indicators that would otherwise remain hidden.

Our work does not only visualize the current situation ,it aligns with Egypt's Vision 2030 and the UN Sustainable Development Goals, highlighting issues that directly impact education quality, public health, family stability, community safety, and overall social well-being.

The interactive dashboards, modeling techniques, and comparative insights aim to support better understanding and more informed decision-making. Ultimately, this project demonstrates how data analytics can play a powerful role in identifying challenges, raising awareness, and guiding future solutions that contribute to sustainable development in Egypt.