

Social Issues in Egypt Dataset Analysis

Analytical Workflow

1. Data Understanding

- Identify all available files and the content of each one.
- Determine the type of each variable (numerical – categorical – textual).
- Define potential relationships between files (e.g., Governorate and Year as keys).

2. Initial Data Exploration

- Display the first few rows of each file.
- Check the number of rows and columns.
- Inspect data types for each column.
- Detect missing or abnormal values.

3. Data Merging

- Identify common columns across datasets (e.g., Governorate, Year, Age Group).
- Combine the files into one unified DataFrame for analysis when appropriate.
- Ensure no duplicates appear after merging.

4. Data Cleaning

- Handle missing values (remove or impute).
- Fix or standardize column names if needed.
- Convert incorrect data types (e.g., strings to numbers).
- Remove duplicate rows.



5. Data Preparation

- Create new columns if needed for analysis (e.g., rates, percentages).
- Encode categorical values into numerical form when necessary.
- Address outliers if they affect accuracy.

6. Descriptive Analysis

- Calculate mean, median, min, max for key variables.
- Analyze distribution of education (literacy) levels.
- Analyze incidence of health issues, begging, addiction, divorce.
- Count records by category or group (Governorate, Year, Age Group).

7. Analytical Questions

- Formulate key questions to explore relationships in the data.
- Investigate correlations between variables (e.g., illiteracy vs health outcomes).

8. Data Visualization

- Create charts to illustrate key findings.
- Compare variables visually.
- Use bar charts, pie charts, histograms, line charts, or heatmaps as needed.

9. Insights & Findings

- Extract the most important results from the analysis.
- Highlight key observations and detected patterns.

10. Recommendations

- Suggest improvements or actions based on the findings.
- Connect insights to project goals and possible interventions.



Question Analysis

- What is the relationship between education (literacy) level and health outcomes?
- Do areas with higher illiteracy have worse health indicators?
- How is begging distributed across governorates and years?
- Is there a correlation between addiction rates and divorce rates?
- What is the average rate of each issue per governorate?
- Does governorate-level education influence addiction or begging rates?
- What percentage of population/groups falls into each risk category?
- How do indicators vary across urban vs rural areas (if available)?
- Is there a link between reported begging cases and poverty/unemployment indicators?
- What are the most influential factors affecting health and social vulnerability?
- Do regions with better education achievements show lower social issues?
- Are there noticeable differences across age groups?
- What is the overall distribution of each issue across the country?
- Does one social issue tend to co-occur with another (e.g., addiction + divorce)?
- How do education, health, and social vulnerability interact together?