# 5 Predictive Modeling Competition: Eutropia Health Challenge

Eutropia, a city known for its advanced healthcare systems, faces unique public health challenges. The Eutropia Health Department (EHD) has collected a comprehensive health dataset from the city's population to study the factors influencing a specific disease.

# 6 Dataset Description

The dataset includes the following features from the citizens of Eutropia:

- Demographics: Age, Weight, Height, Gender
- Clinical measurements: Blood Pressure (systolic and diastolic), Cholesterol Levels, Blood Sugar Levels
- Lifestyle factors: Smoking Status, Alcohol Consumption, Physical Activity, Diet Quality
- Environmental factors: Air Quality Index, Water Quality Index
- Psychological factors: Stress Level, Sleep Quality
- Family history: Number of immediate family members with the disease
- Target Variable: Disease Status (binary: 0=no, 1=yes)

# 7 Tasks

Students will engage in the following tasks:

1. **Model Development:** Develop a model using the provided training data to predict the disease status. Explore various statistical and machine learning methods.
2. **Prediction Submission:** Submit predictions for a withheld test set. Include a brief report detailing the methodology, model choice, and insights.
3. **Evaluation:** Models will be evaluated based on metrics such as accuracy, AUC-ROC. Rankings will be announced based on performance.

The final ranking of student submissions will be based on AUC-ROC. Your report should adhere to the following structure and content guidelines:

**Data Exploration**

- **Description of the Data**: Provide an overview of the dataset characteristics.
- **Statistical Summaries**: Summarize the key statistics of the data which may include measures of central tendency and variability.
- **Visualization**: Include essential visualizations such as histograms, box plots, or scatter plots to illustrate the data distribution and any relationships between variables. Limit yourself to the most informative and relevant graphs.

### Model Development

- **Methodology**: Clearly describe the regression techniques or any other statistical methods used in your analysis.
- **Model Selection**: Discuss the rationale behind choosing specific models and their configurations.
- **Implementation Details**: Provide insights into the implementation process, including any challenges faced and how they were resolved.

### Results

- **Model Performance**: Present the findings from your model application, supported by performance metrics.
- **Interpretation of Results**: Discuss what the results imply in the context of the problem being investigated.
- **Visual Representation**: Use a limited number of charts (e.g., line plots of model accuracy or heatmaps of coefficients) to clearly represent the outcome of the modeling process.