### README - Medical Follow-up Indicators & Diabetes Prediction

## Project Overview

This project focuses on analyzing and predicting medical follow-up indicators related to diabetes. The dataset contains historical data on various medical tests and consultations for diabetic patients in France (2010-2023). The objective is to visualize trends and use linear regression to predict future values.

## Dataset Description

The project utilizes multiple datasets related to medical follow-ups:

* Type1\_Diabetes.xlsx: Contains data on the prevalence of Type 1 diabetes.
* Prevalence\_Generale.xlsx: Includes general statistics on diabetes monitoring.
* Doc\_globale\_des\_analyses.xlsx: Aggregated dataset of medical follow-ups.
* Medical Follow-up Datasets (with yearly evolution):
  + HbA1c.xlsx: Blood glucose monitoring.
  + Creatininemie.xlsx: Kidney function test.
  + Microalbuminurie.xlsx: Another kidney function test.
  + Lipides.xlsx: Lipid level check.
  + Dentaire.xlsx: Dental consultation follow-up.
  + Cardio.xlsx: Cardiological check-up.
  + Ophtalmo.xlsx: Ophthalmology consultation.

Workflow & Implementation

Steps

### Data Loading & Cleaning

* Each dataset is loaded from the "Graphiques" sheet.
* The first few rows containing unnecessary headers are removed.
* Missing values are handled, and data types are converted.
* The top 10 and bottom 10 regional rankings are filtered out.

### Data Fusion

* The datasets containing medical indicators over the years are merged on the Année column.
* This results in a single dataset with multiple columns representing different indicators.

### Data Visualization

* A line plot is generated to display trends in medical follow-ups between 2010-2023.

### Linear Regression Prediction

* A linear regression model is applied to predict the future values of medical indicators.
* The model is trained using Scikit-learn to predict trends for upcoming years.

How to Run the Code?

Install dependencies (if not already installed):  
pip install pandas matplotlib scikit-learn

Run the script in a Jupyter Notebook or Python environment:  
python analysis.py

Expected Outputs:

* + Cleaned datasets
  + Line chart of trends (2010-2023)
  + Predictions of medical follow-ups for future years

## Example Output

* Graph: Trend of HbA1c, Creatininémie, etc.
* Predicted Values: Future medical checkup frequencies.

## To Improve in the Future

* Use more advanced machine learning models (e.g., Random Forest, Neural Networks).
* Add more external factors (e.g., air pollution, socio-economic factors).
* Improve the visualization dashboard.