Acea Smart Water Analysis (Petrignano)

**Problem Statement Worksheet**

Forecast the depth to groundwater of an aquifer located in Petrignano, Italy. Also forecast the underground(aquifers) water level and water flow, for each day of the year.

**Context**

The wells field of the alluvial plain between Bastia Umbra and Petrignano is fed by three underground aquifers separated by low permeability septa.

The groundwater levels are influenced by the following parameters:

* Rainfall
* depth to groundwater
* temperatures
* drainage volumes

Solution will help Petrignano water plant; in suggesting best recommendation from its prediction of Under Ground water level depth. With advance prediction, it helps the plant manage its supply.

**Criteria of success**

In order to do so, we must determine what features influence the amount of water in that waterbody, and in which way those features affects it.

To handle daily consumption, Petrignanowater Plant need to forecast the underground(aquifers) water level and water flow, for each day of the year.

**Scope of solution scope**

Will build a model to predict the amount of water in Water body: Petrignano. How rainfall, temperature, drainage volume, river hydrometer impact depth of ground water What is the future depth to groundwater of a well belonging to the aquifer in Petrigrano over the next few weeks. As I’m using time series analysis so I can predict future analysis.

**Data science approach**

## First I check data is stationary or not to do so I’m conducting [Augmented Dickey-Fuller (ADF)](https://www.kaggle.com/andreshg/timeseries-analysis-a-complete-guide#2.3.1). For EDA I’m using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

## I shall be using ARIMA, Auto-ARIMA model for Time Series analysis and LSTM recurrent neural network to model our business problem. Using Facebook Prophet  for forecasting time series data.

**Stakeholders to provide key insight.**

Petrignano water plant of Acea Group

**Key data sources**

Although the dataset contains multiple water bodies, we will only be looking at the Aquifer\_Petrignano.csv file.

[https://www.kaggle.com/c/acea-water-prediction/data/](https://www.kaggle.com/c/acea-water-prediction/data/%20)

CSV file: - Aquifer\_Petrignano.csv

**Approach:**

# Multiple steps will be taken to build a predictive model for this project as well as to analyze the resulting predictions.

# The csv file is imported, merged, and cleaned via Python. Missing values and outliers will be handled appropriately based on specific factors.

# The cleaned dataset will be explored visually in order to find interesting trends/correlations in the data.

# A model to be determined will be employed on the dataset.

**Deliverables:**

The final draft of the project will be presented in the form of a slide deck and formal project report. Jupyter Notebooks will be delivered detailing each step taken and code written for the analysis of the project. A GitHub repository for the project will be created as well.