**Project Proposal**

**Evaluating Missing Data Mechanisms for Kalman Smoothing in Water Data Imputation**

**Objective**

The goal of this project is to evaluate the effectiveness of Kalman smoothing under various missing data mechanisms: MCAR (Missing Completely at Random), MAR (Missing at Random), and MNAR (Missing Not at Random), for imputing missing water data. Additionally, Kalman smoothing will be compared to alternative imputation methods focusing on bias and accuracy to identify the most suitable approach. I will also try methods to calculate variance using the Kalman’s smoothing. I will be using a time series data from [data.world](https://data.world/epa/dwtp-data-source-water-data/workspace/file?filename=A-63xz-DWTP+data-Heberling-20160405.xlsx).

**Methods**

Simulate and analyze water data under MCAR, MAR, and MNAR conditions to evaluate their impact on Kalman smoothing's ability to preserve data trends and integrity.

Explore methods to calculate variance using Kalman smoothing.

Compare Kalman smoothing with Exponential Smoothing and a Combined Approach (Kalman + Exponential + LOCF) using metrics such as accuracy (MSE, RMSE), variance (state-space model's error covariance), and bias (average deviation between imputed and original values).

Conduct sensitivity analysis by testing varying missing data proportions (e.g., 10%, 30%, 50%) and assessing the robustness of Kalman smoothing and other methods under different missingness patterns.

**Expected Outcome**

Identify the most compatible missing data mechanism for Kalman smoothing.

Compare the performance of Kalman smoothing with other imputation techniques across different missingness scenarios.

Provide insights into variance, bias, and accuracy trade-offs, offering recommendations for effective water data imputation strategies.