**No-Code AI Workshop**

**Prompt examples to use with “Data Analyst” GPT**

**Prompt for Exploratory Data Analysis**

You are an expert dendrochronologist and you are going to analyze the attached Excel workbook. This workbook has three sheets:

* observed\_data: This sheet contains the observed seasonal precipitation data for the Sava River Basin (SRB).
* PDSI\_cells\_data: This sheet includes data from the PDSI (Palmer Drought Severity Index) cells located within a 450 km radius around the SRB. The data, extracted from tree rings, contain annual information covering a period longer than two millennia.
* coordinates: This sheet provides the ID and location coordinates (latitude and longitude) for each of the PDSI cells.

I want you to conduct an exploratory data analysis and present me with the most important information I need to know about the data.

**Prompt for Data Plotting**

Create some graphs. I want to see the behavior of the precipitation in the observed period, I also want you to use the coordinate information to paint a map showing the location of the 249 PDSI cells.

**Prompt for Data Plotting**

Can we analyze PDSI data trends?

**Prompt for Model creation and Performance evaluation**

Remember that you are an expert dendrochronologist. Let's do the paleo reconstruction of precipitation in the Sava River basin.

* The goal is to create and train a regression model that uses the PDSI cells’ information (independent variables) for the observed period (1876-2012) to predict precipitation (dependent variable). Once the model is built and validated, it can be used to reconstruct precipitation for the period before the observed data (years 0 to 1875).
* Please create a dataset with the observed data and the 249 PDSI cells. With that dataset try different machine learning models to do a regression using 5 or 10-fold cross validation.
* Create a table with the accuracy metrics for each model you try. You can use the metrics you consider more relevant.
* With the best performing model in the metrics, make a plot of the predicted precipitation in the validation set compared to the observed data in the validation set.
* Construct another plot comparing the predictions for all years of the observed period with the observed data.
* Finally, apply the model to the PDSI cells of the period prior to the observed period and constructs a plot of the reconstructed precipitation in that period.

**Prompt for Export results**

Export the results to a csv file with three columns year, observed\_P, and predicted\_P. The year should start at 0 and go to 2012. For years where there are no observed values, the observed\_P fields for the corresponding years should be empty in this resulting file.

**Prompt for Feature Engineering**

Run a feature engineering process to automatically select the PDSI cells that work best with the model and try to improve the metrics you obtained previously.

**Prompt for Model optimization**

Since the model performance did not improve with feature engineering, can you try to improve it through hyperparameter optimization?