1. **02.19** 🡪 Data Mining

* Identify variables from papers, Kaggle, github and other resources (to do: identify keywords from google trends)
* Download and merge them with main dataset
* Ensure homogenous formatting
* *Examine scaling strategy for ML pipeline (Normalize, Standardize, Robust Scale ? )*

1. **02.24** 🡪 Preliminary Data Analysis

* Descriptive Statistics + visualizations (“get to know the data”)
* Feature selection: basic univariate regression analyses to identify potentially important variables (reduce computational complexity), but also F statistics, correlation matrix…etc.

1. Method Design

* Identify design: one regression by year ? by multiple time periods ?
* Train various ML models (Decision Trees, XGBoost, Ridge, Lasso, Random Forest, CNNs…etc) 🡪 only keep the most performant 🡪 get SHAP values (tables + visualizations + maybe API output)
* Train forecasting with Prophet (easy) and tune model

1. Results

* SHAP values interpretation
* Deploy forecasting model (API, dashboard, or whatever we come up with)
* Setup Presentation
* *Paper Writing*
* *Setting up dashboard*