

First, I've downloaded all datasets, importing csvs with `pd.read_csv()`. Then get the JSONs from the provided APIs as dictionaries. Up next, I've used `pd.DataFrame.from_dict()` to turn them into dataframes. Later I've merged all dataframes into one using `pd.concat()`.

After that I've made a visual analysis of the dataframe, using `pd.head()`, `pd.describe()`, checking unique values on every column and checking for NaN values. For dimensionality reduction, I've decided to remove redundant columns (such as min/max wind speed and temperatures) while keeping its averages columns, and I've also removed unnecessary columns such as country name, city name, reporter's name, continent (they're all Europe), targetRelease (only one value), day of report, city ID and facilityName.

Then I proceed to encode `eptrSectorName` and `pollutant`, adjusting it to the challenge's requirements. I've removed duplicated feature columns, leftovers from the encoding step. After that, I've declared the X and y variables, split them into training and testing for scoring check. I've decided to use K Nearest Neighbors as a classification model for the challenge, using a pipeline to scale it and GridSearch for hyperparameters tuning. Then I've processed the test data as the train data, trained the model, made predictions and necessary output files.