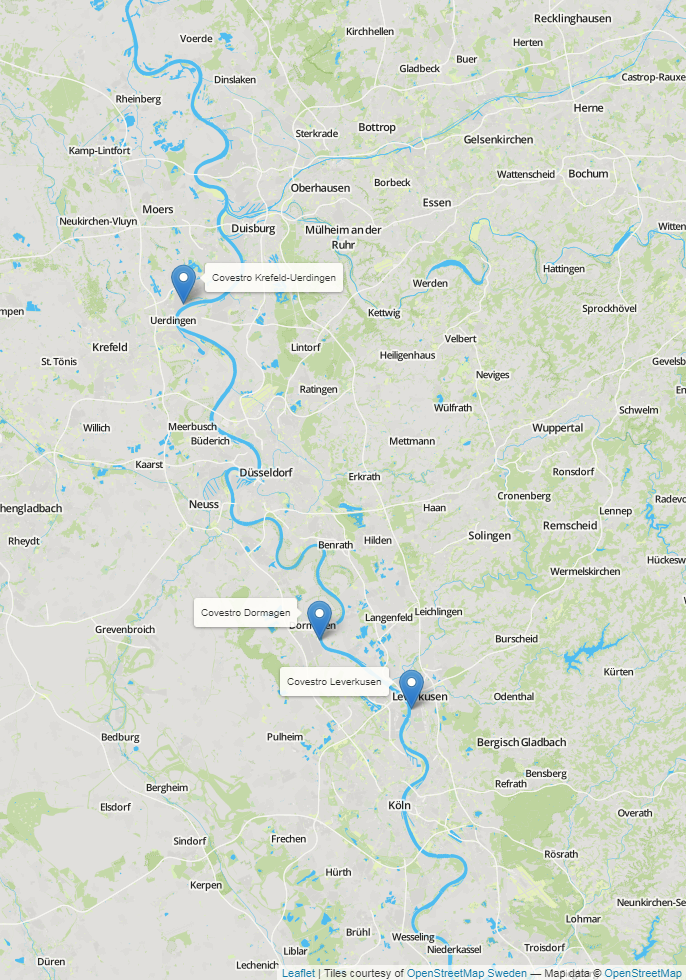
**Rhine Water Level Use Case**

**Description:**

Covestro operates world scale production lines for a diverse set of polymers, demanding a large and predictable inflow of raw materials to produce large volumes of finished goods. These volumes cannot solely be transported on land, which leads to a high dependency on barge and ship shipments along the Rhine River near Leverkusen, Germany (NOTE: the locations along the Rhine of interest to Covestro are highlighted in the image below). Additionally, large amounts of the river water are used in different process steps (e.g. as cooling medium).



**Pain Point:**

Low / high water levels lead to:

* Shortages of raw materials, which can delay and/or prevent the production of finished goods.
* Prevention of the distribution of finished goods, which creates inventory buildup and customer dissatisfaction.
* Prevention of Covestro being able to use the water in the production process.

**Data Provided:**

The provided time series data consists of the following for the years of 1965-2012:

* Water levels of the Rhine River and its side arms in Germany.
* Discharge volume of the Rhine River and its side arms in Europe.
* Precipitation values of the Rhine region.

Additional weather data can be found at: <https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/>.

You are welcome to use the provided weather data, additional data from the link, or other open source weather data as you see fit.

Every 3.5 months, 16 days of data are missing. More details on the data can be found below.

**Goal and Evaluation Metric:**

Predict the delta of the Rhine water level at Duesseldorf and Cologne (station ID’s 6335050 and 6335060) of the first, second, and third day of each missing period compared to the last day prior to the missing period. Performance of the model will be evaluated using R2. The R² formula should be calculated as follows:

Where:

**WL**  water level at the given time point

**t0**  the last day available prior to the missing period

**t<i>**  the ith day of the missing time period

**Submission File:**

The submission file should contain the predicted delta to the reference value for each date and station id in the to\_predict.csv file. The submission should be a CSV file entitled “team**#**submission.csv”, where **#** is your assigned team number, and should follow the template format provided for the chosen use case. The file should be sent to the following email address: [Chemalytix@covestro.com](mailto:Chemalytix@covestro.com).

**Data Details**

**river\_stations.csv:**

Missing values: -999

|  |  |
| --- | --- |
| Variable | Description |
| GRDC-No. | station ID of the river station of measurement |
| River | river name |
| Station | station name |
| Country | country |
| Latitude (DD) | latitude |
| Longitude (DD) | longitude |
| Catchmentarea | area covered |
| Altitude | altitude of the station |
| Nextdownstreamstation | station ID of the next downstream station |

**weather\_stations.csv:**

Missing values: -999

|  |  |
| --- | --- |
| Variable | Description |
| STATIONS\_ID | station ID of the weather station of measurement |
| LATITUDE | latitude |
| LONGITUDE | longitude |
| City | city name |
| STATE | state |

**river\_data.csv:**

Missing values: -999

|  |  |  |
| --- | --- | --- |
| Variable | Unit | Description |
| station\_no. |  | station ID of the station of measurement |
| date | YYYY-mm-dd | measurement date |
| discharge | m³/s | daily average of the discharge rate |
| water\_level | cm | daily average of the water level |

**weather\_data.csv:**

Missing values: -999

|  |  |  |
| --- | --- | --- |
| Variable | Unit | Description |
| RS | mm | daily precipitation height |
| RSF | numeric code | precipitation form  0 = no precipitation (conventional or automatic measurement), relates to WMO code 10  1 = only rain (before 1979)  4 = unknown form of recorded precipitation  6 = only rain; only liquid precipitation at automatic stations, relates to WMO code 11  7 = only snow; only solid precipitation at automatic stations, relates to WMO code 12  8 = rain and snow; liquid and solid precipitation at automatic stations, relates to WMO code 13  9 = error or missing value or no automatic determination of precipitation form, relates to WMO code 15 |
| SH\_TAG | cm | height of snow pack |
| NSH\_TAG | cm | fresh snow depth |
| MESS\_DATUM | YYYY-mm-dd | measurement date |
| STATIONS\_ID |  | weather station ID |

**to\_predict.csv:**

Data points to be predicted.

|  |  |  |
| --- | --- | --- |
| Variable | Unit | Description |
| station\_no. |  | station ID of the station to be predicted |
| date | YYYY-mm-dd | date of the measurement to be predicted |

**team#submission.csv:**

Sample submission file representing the expected format of the submission file.

|  |  |  |
| --- | --- | --- |
| Variable | Unit | Description |
| station\_no. |  | station ID of the station of predicted measurement |
| date | YYYY-mm-dd | measurement date |
| delta | cm | predicted delta |