

Air-Pollution Challenge

Iris Winkler

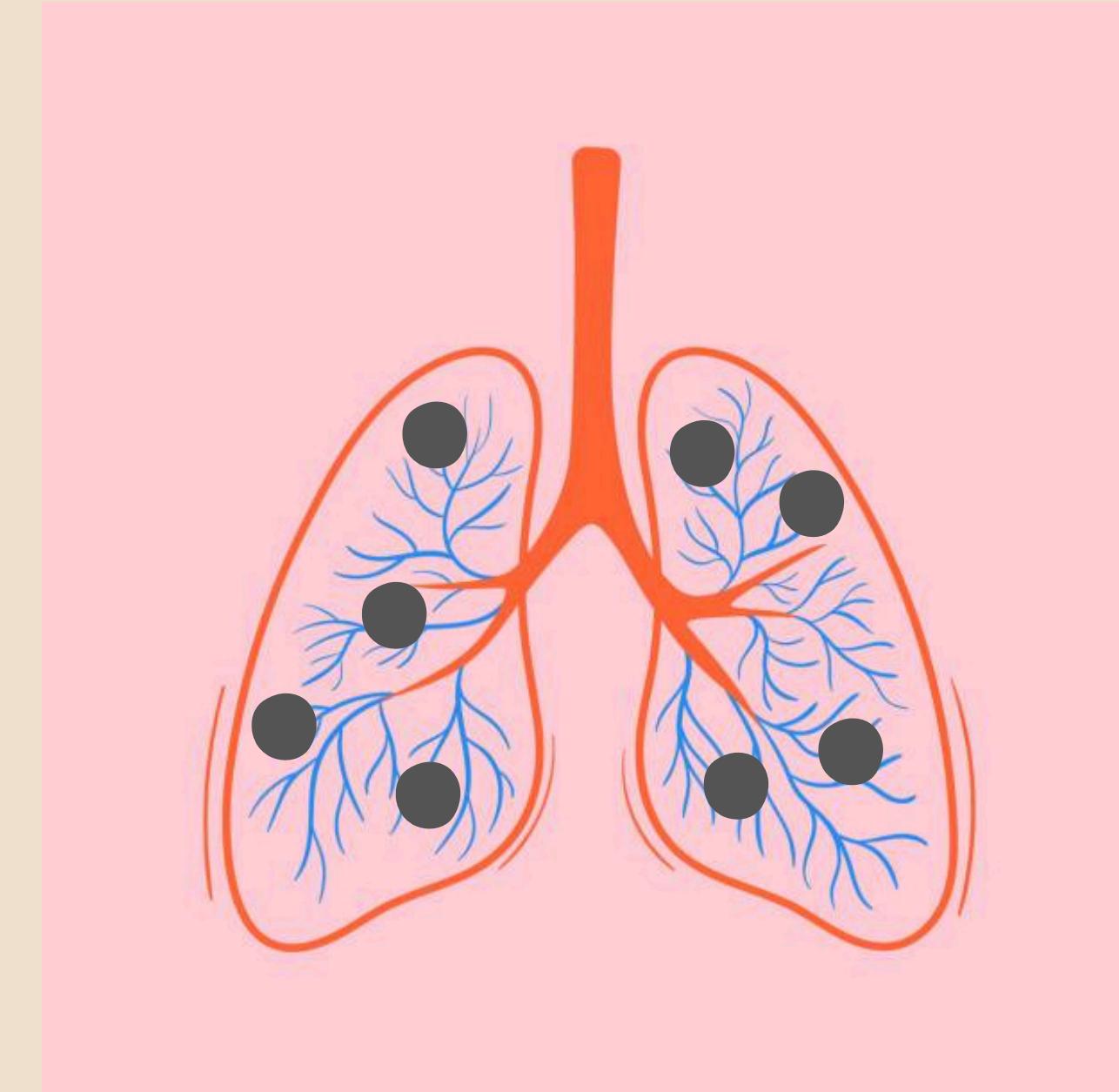
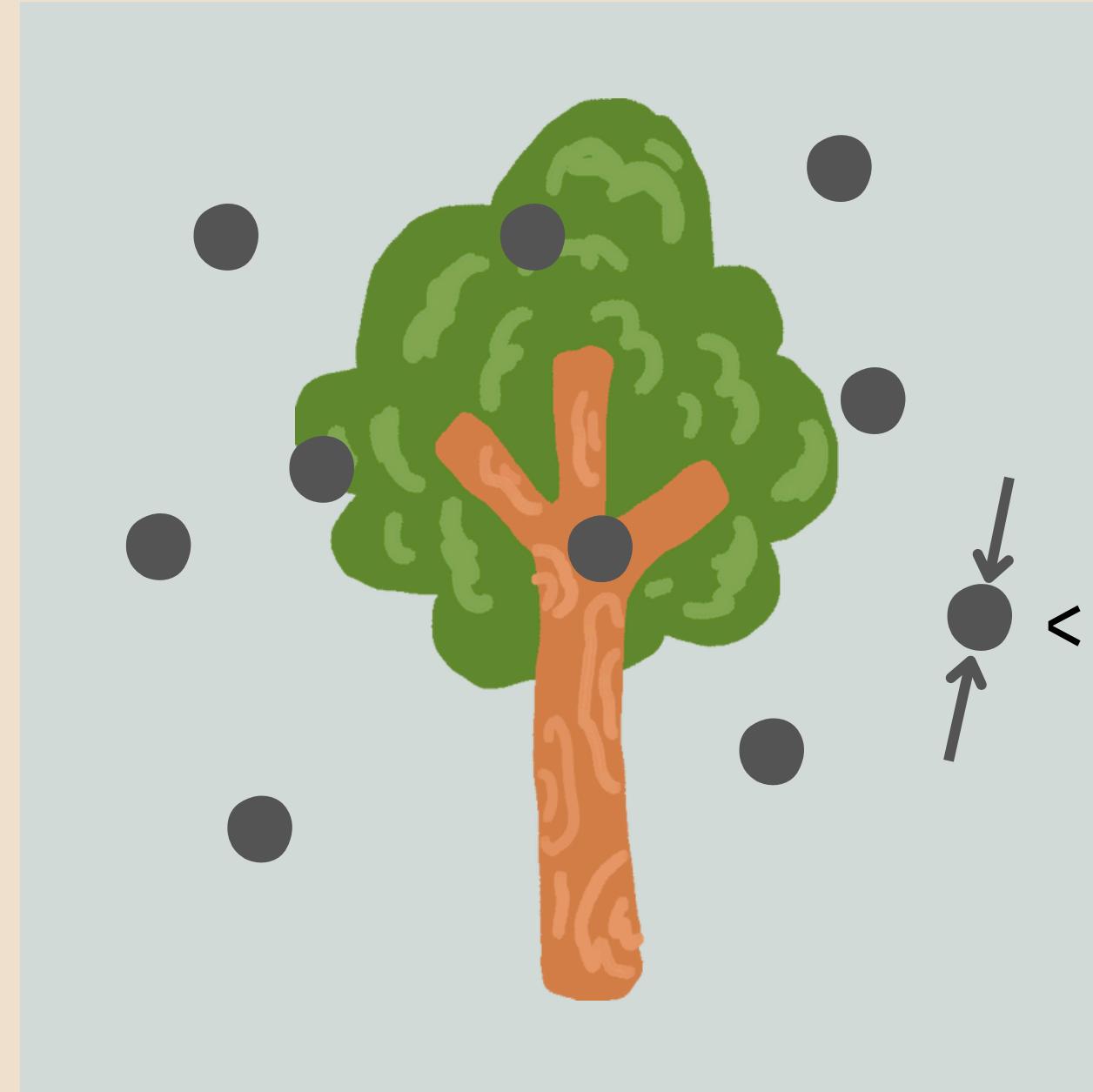
Johannes Gooth

Carlos Duque

cgn-ds-24-2

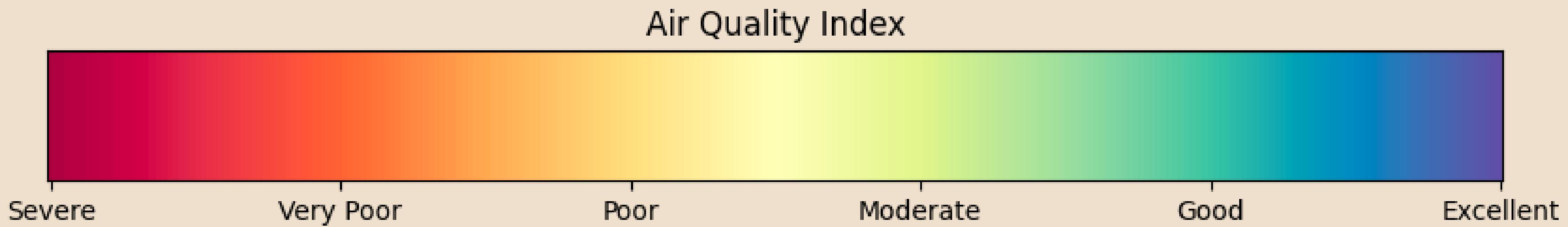


Air Pollution Issue



PM2.5 = particulate matter that is 2.5 micrometers in diameter or smaller

WHO Air Quality Index



Excellent: PM2.5 = 0-12 $\mu\text{g}/\text{m}^3$

Good: PM2.5 = 12.1-35.4 $\mu\text{g}/\text{m}^3$

Moderate: PM2.5 = 35.5-55.4 $\mu\text{g}/\text{m}^3$

Poor: PM2.5 = 55.5-150.4 $\mu\text{g}/\text{m}^3$

Very Poor: PM2.5 = 150.5-250.4 $\mu\text{g}/\text{m}^3$

Severe: PM2.5 ≥ 250.5 $\mu\text{g}/\text{m}^3$

PM2.5 Data Acquisition



Ground sensors

Our Mission

Sentinel 5P



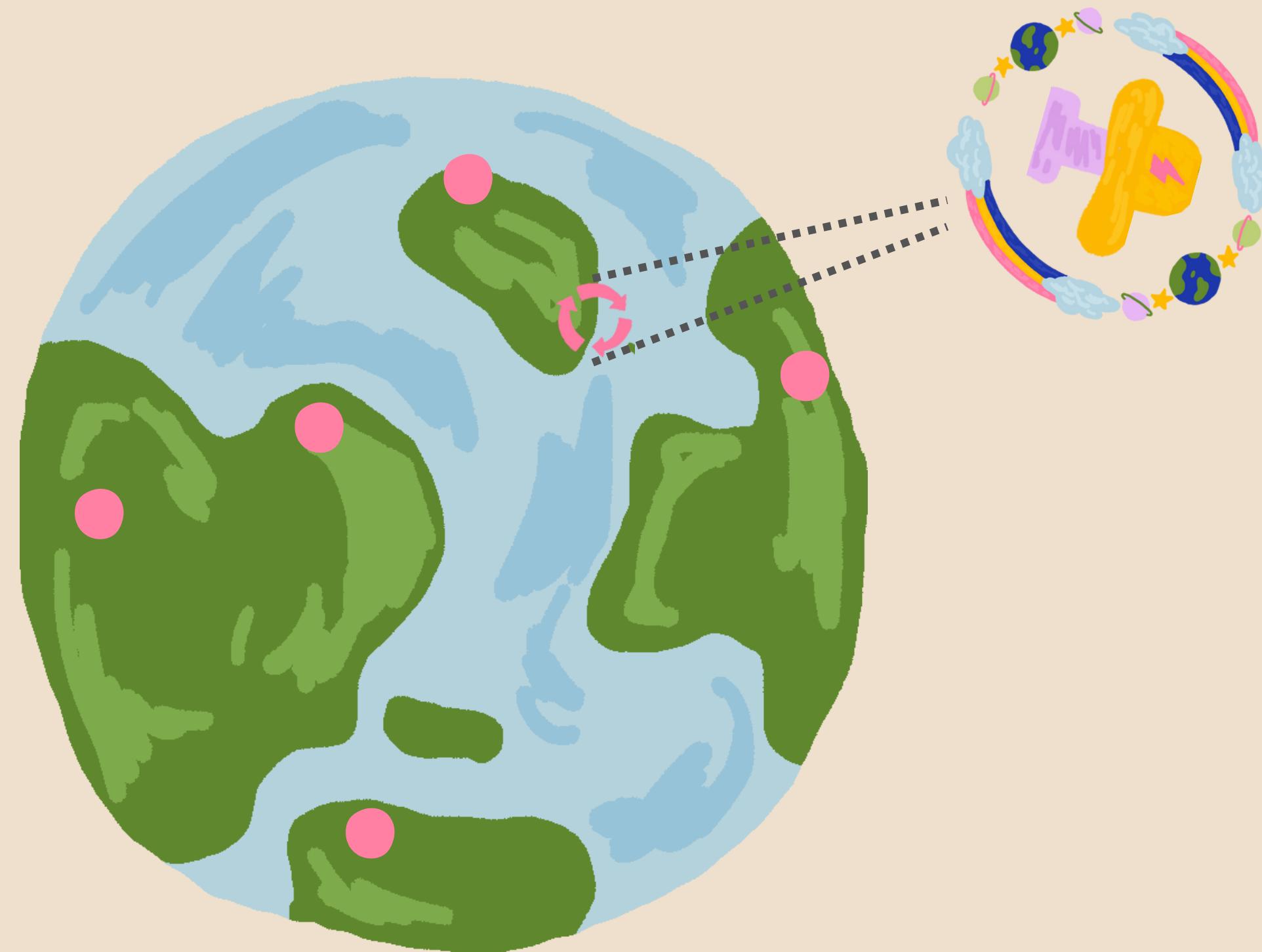
Satellite Data to predict ground-level value of PM2.5

Our Mission



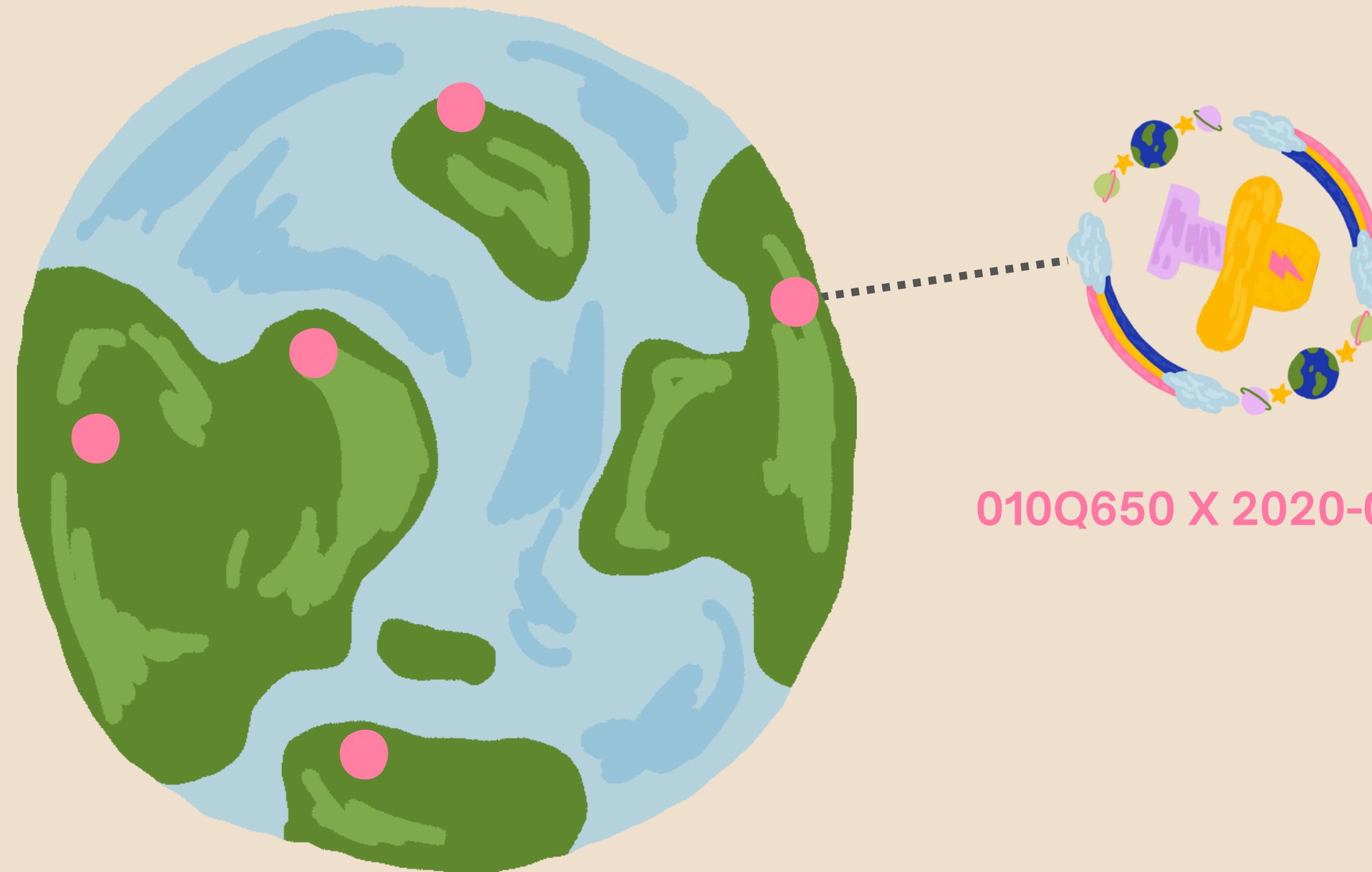
Satellite Data to predict ground-level value of PM2.5

Our Mission



Satellite Data to predict ground-level value of PM2.5

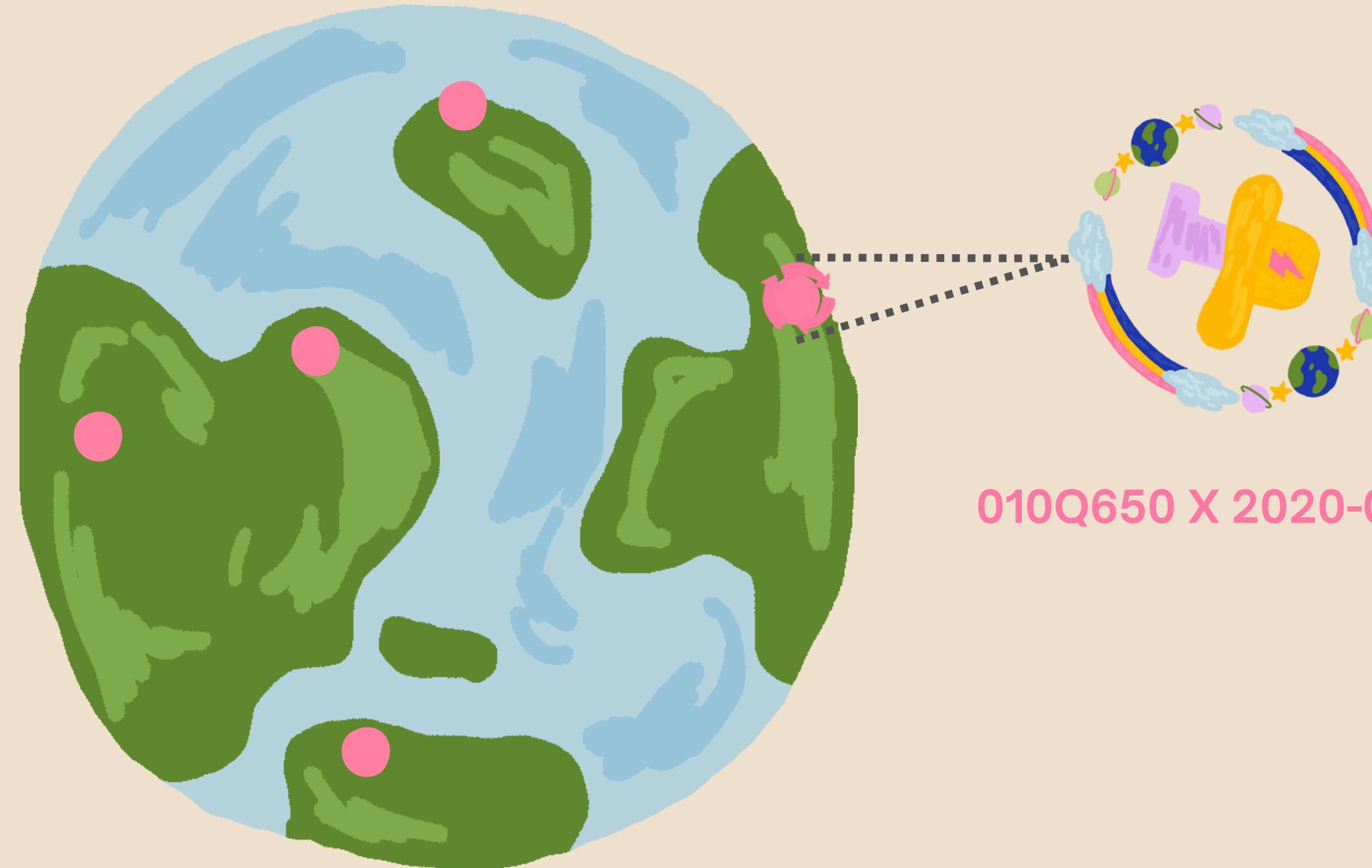
Our Mission



010Q650 X 2020-01-02

Model training

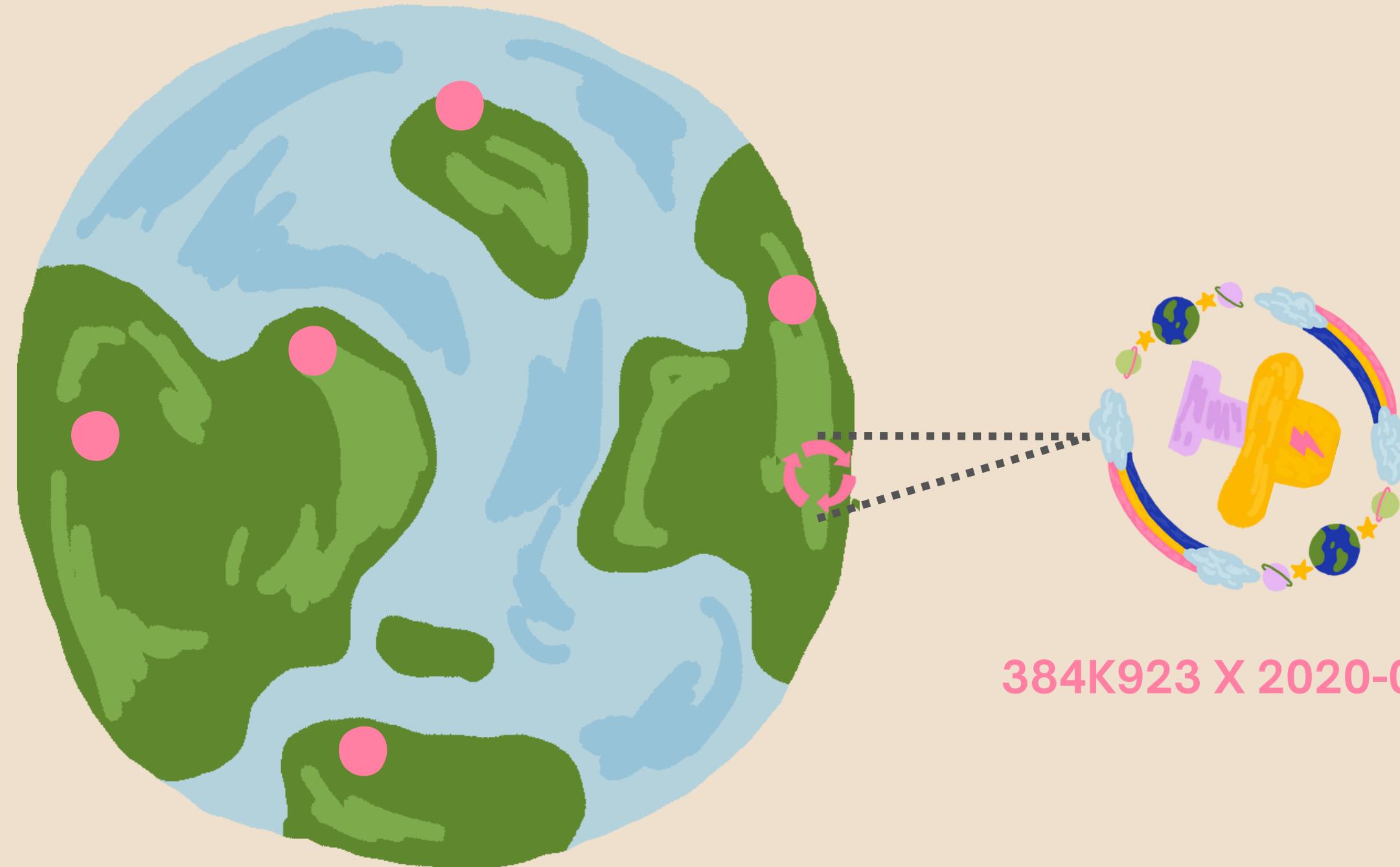
Our Mission



010Q650 X 2020-01-02

Prediction on train data

Our Mission



384K923 X 2020-01-18

Prediction on test data

Evaluation Metrics

Root-Mean-Squared-Error

Winning solution on Zindi challenge:

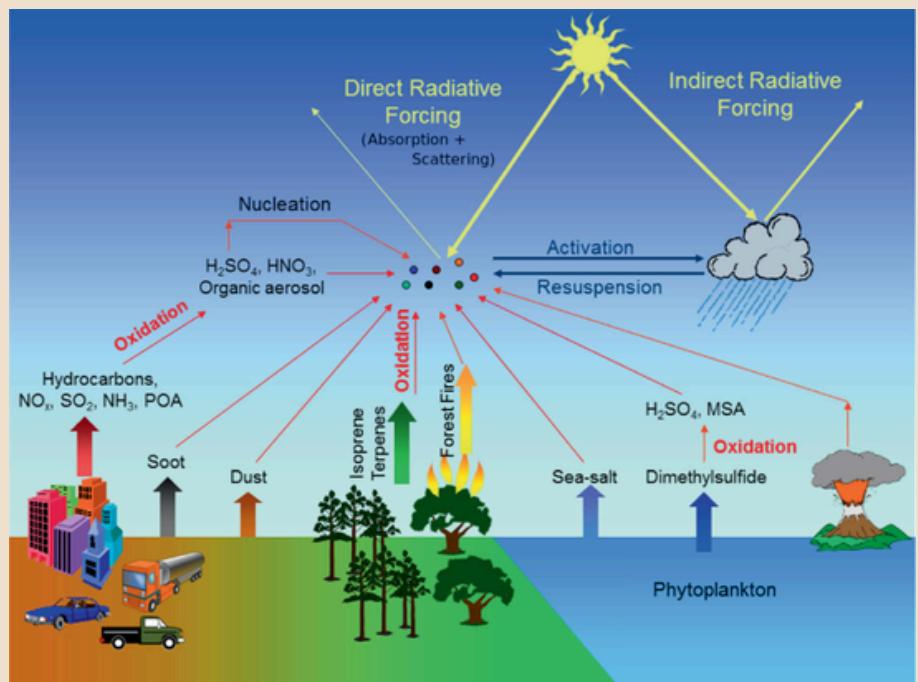
RMSE = 26.1 (Benchmark)

Zindi Ranking

RANK	USER	PRIVATE SCORE
1	 devnikhilmishra	26.09969719
2	 CoviData Team	26.27988143
3	 Klai	26.61821965
4	 Haytheem Enit	26.64522086
5	 Air_Resistance Team	26.70858184
6	 Olayinka_Fadahunsi	26.73282919

The dataset

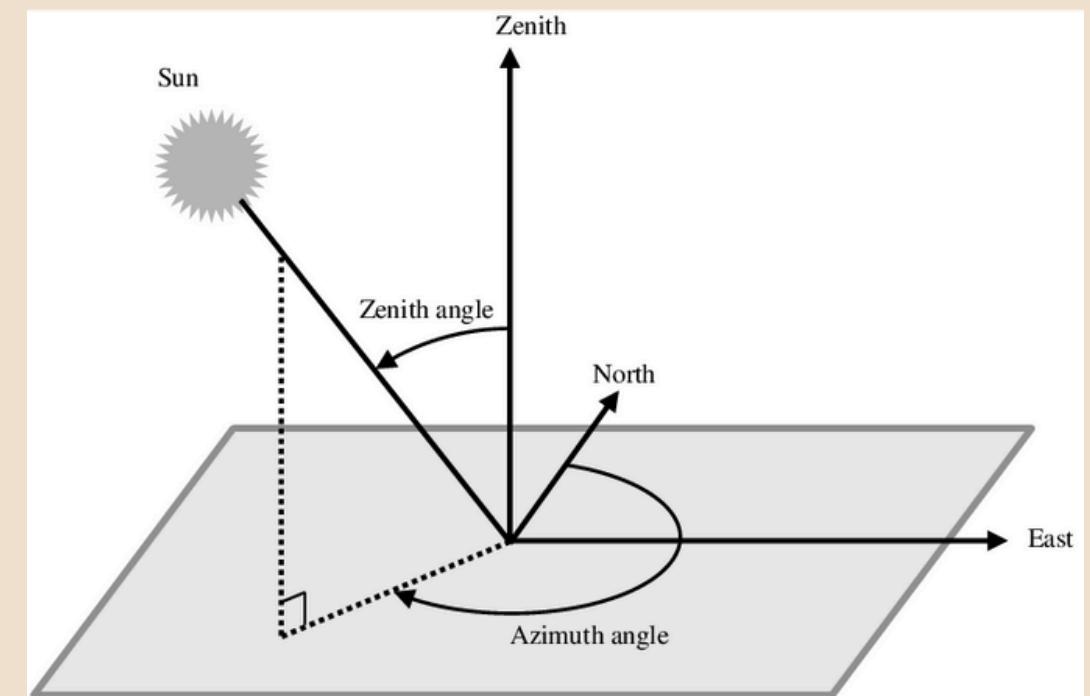
Gases in Earth's atmosphere



Cloud and wind characteristics

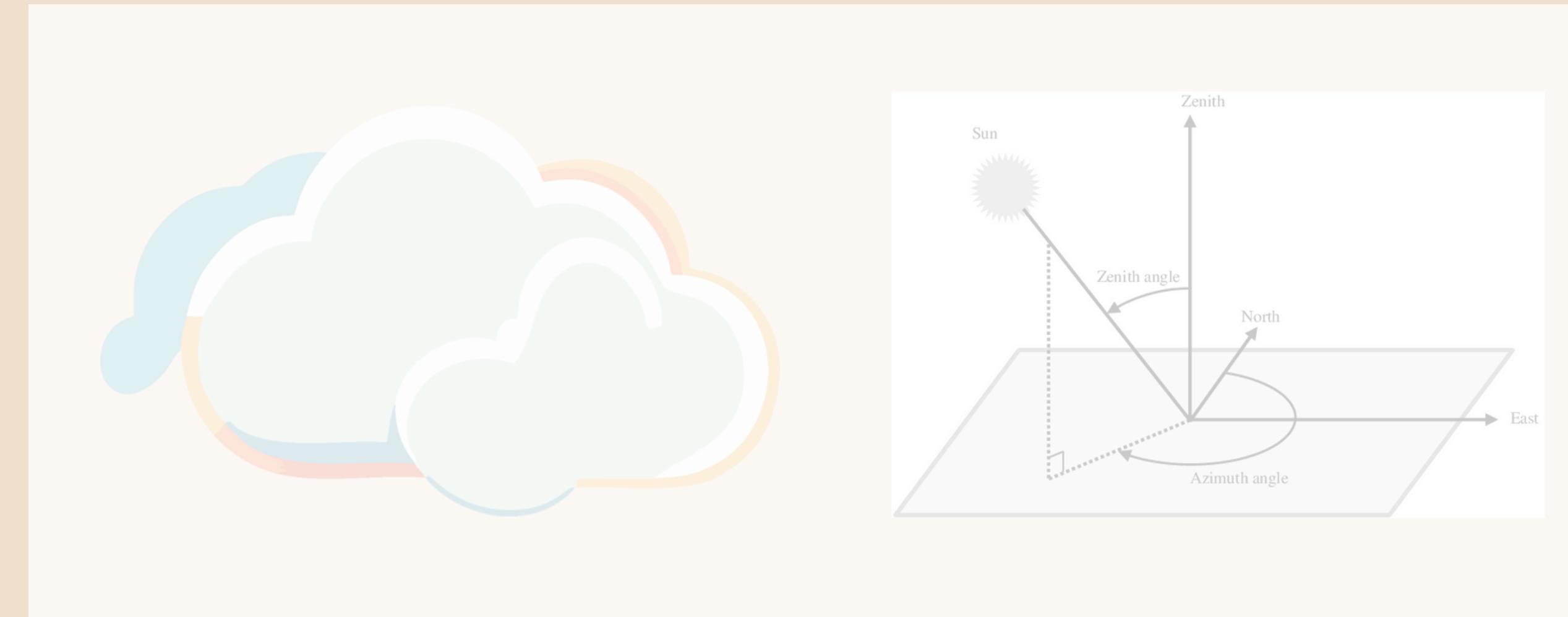
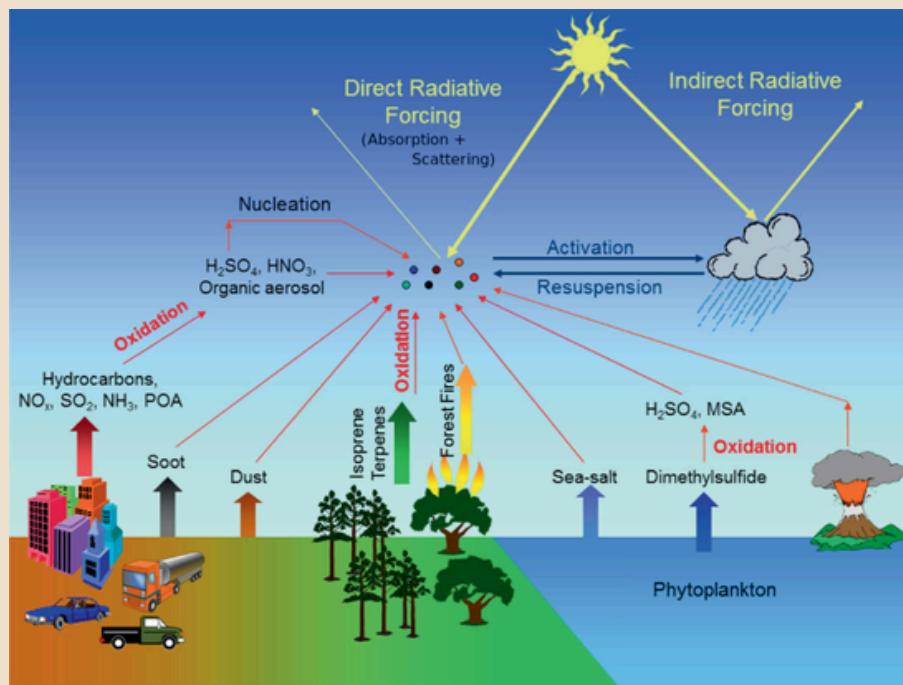


Angle data



The dataset

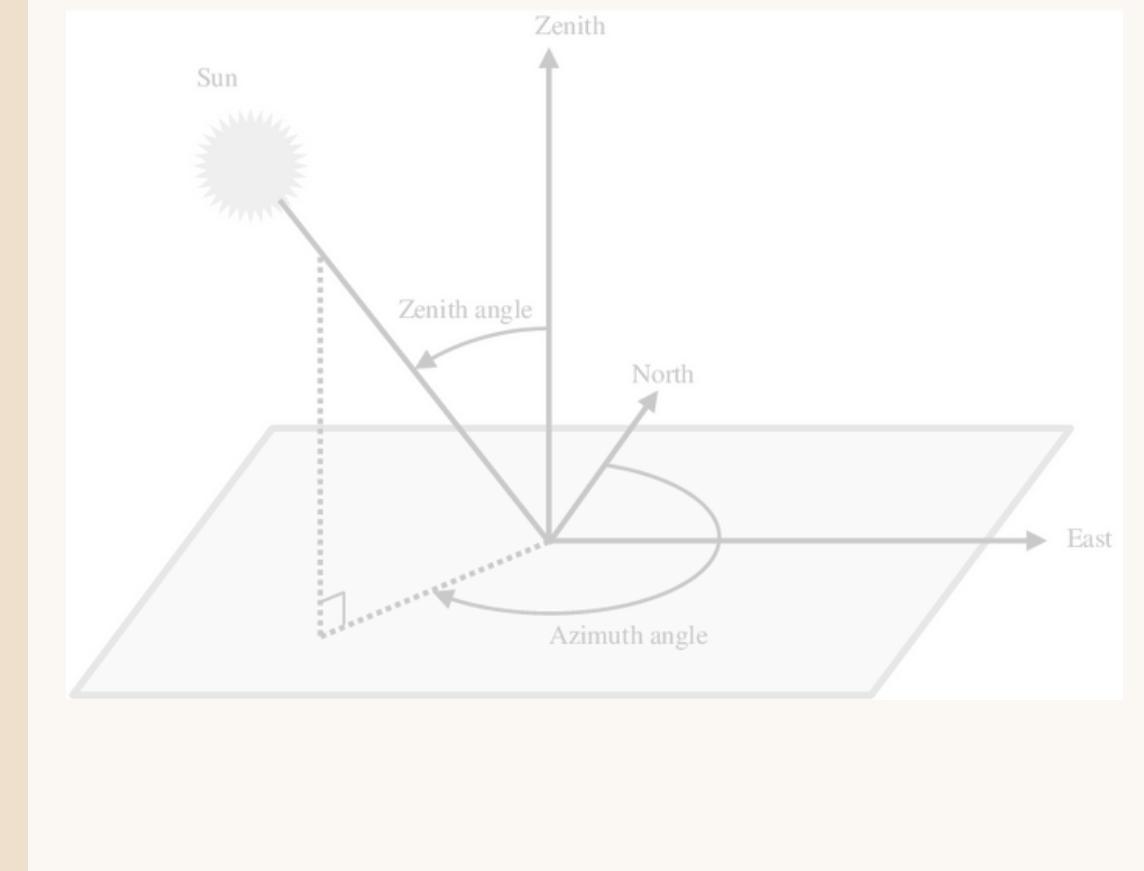
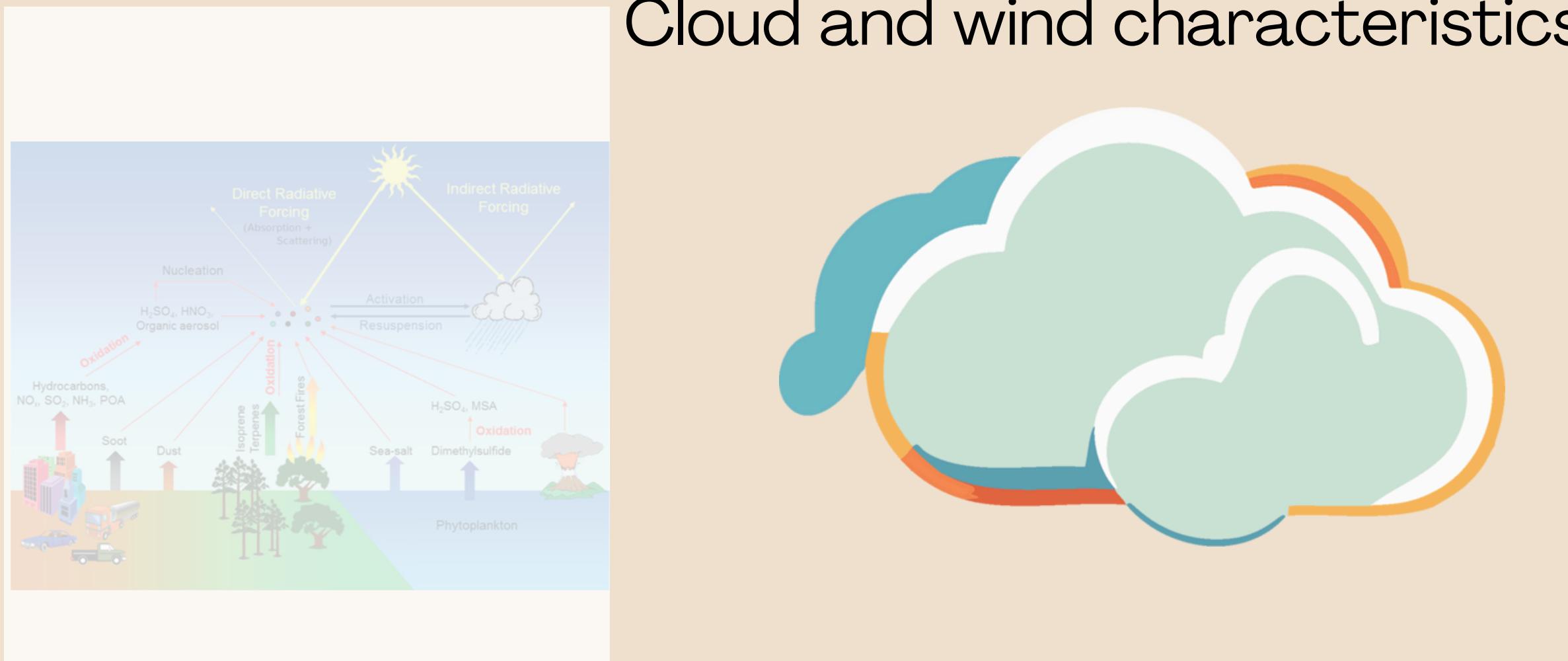
Gases in Earth's atmosphere



- Some gases found in the air such as CO₂ or CO signal the potential presence of PM2.5.
- Other Gases like CH₄ are not directly related to PM2.5 and can be neglected!

The dataset

Cloud and wind characteristics

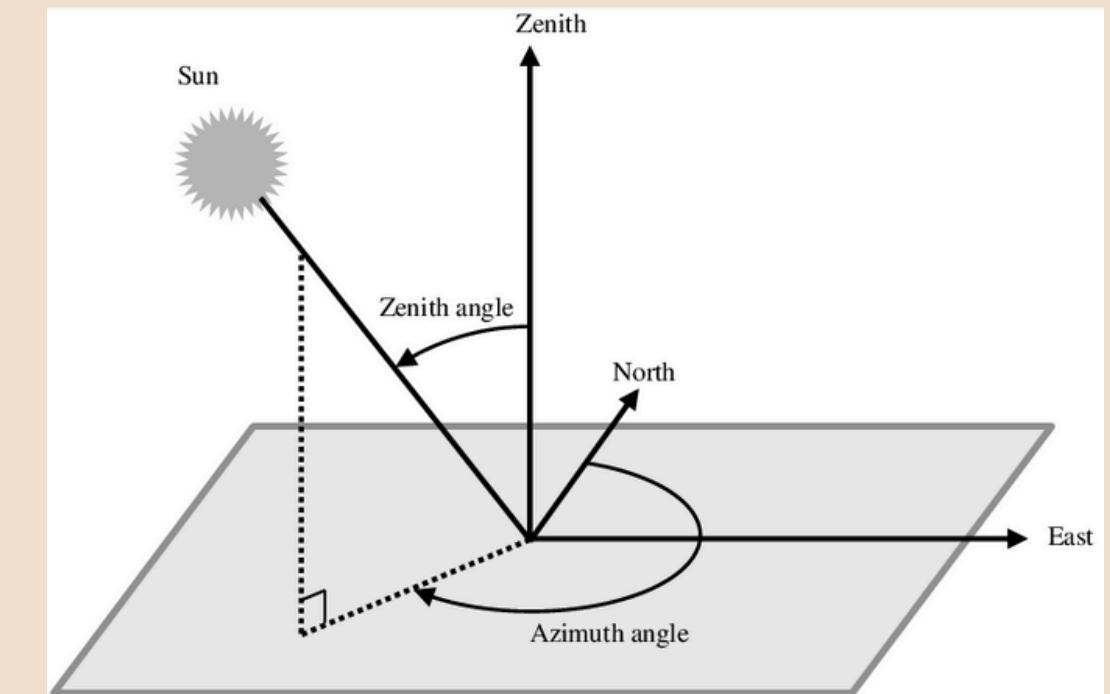


- Variables like cloud fraction or cloud height can affect the concentration levels of PM2.5.
- Wind measurements like wind speed or wind direction partially explain how PM2.5 particles propagate.

The dataset



Angle data



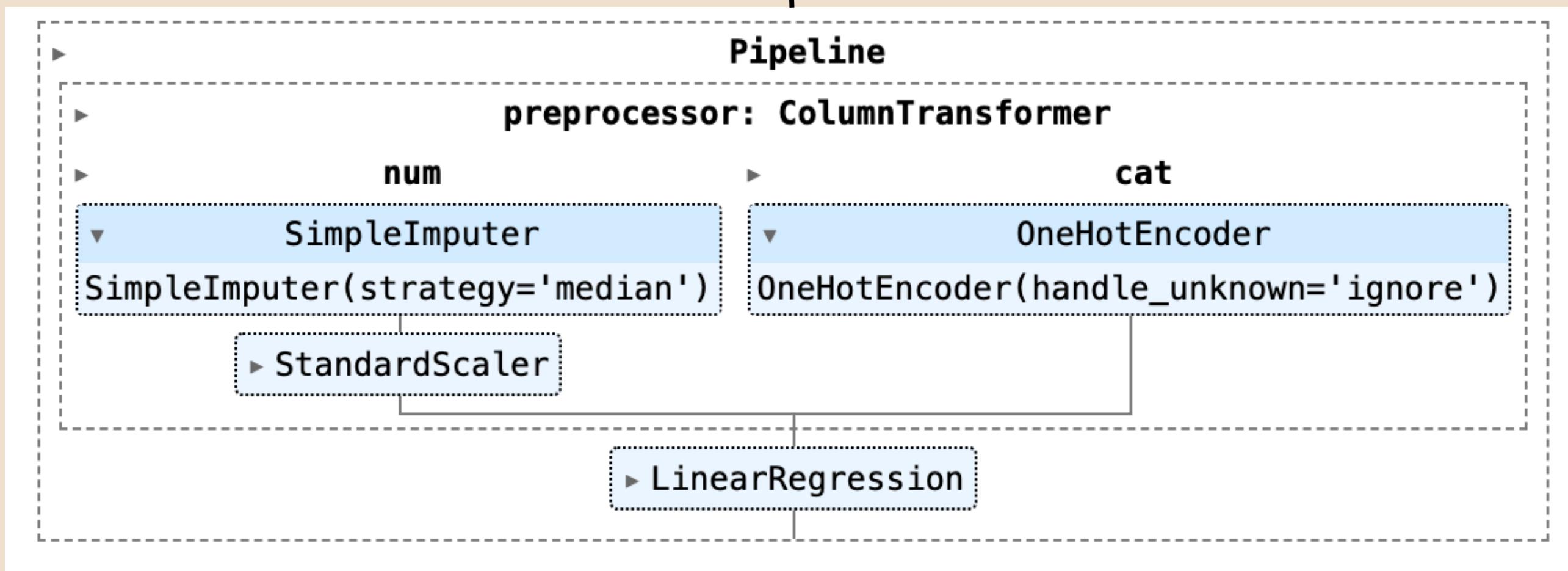
The angle data encode important features that can be related to geographical and directional data.

Feature engineering

In-house preprocessing

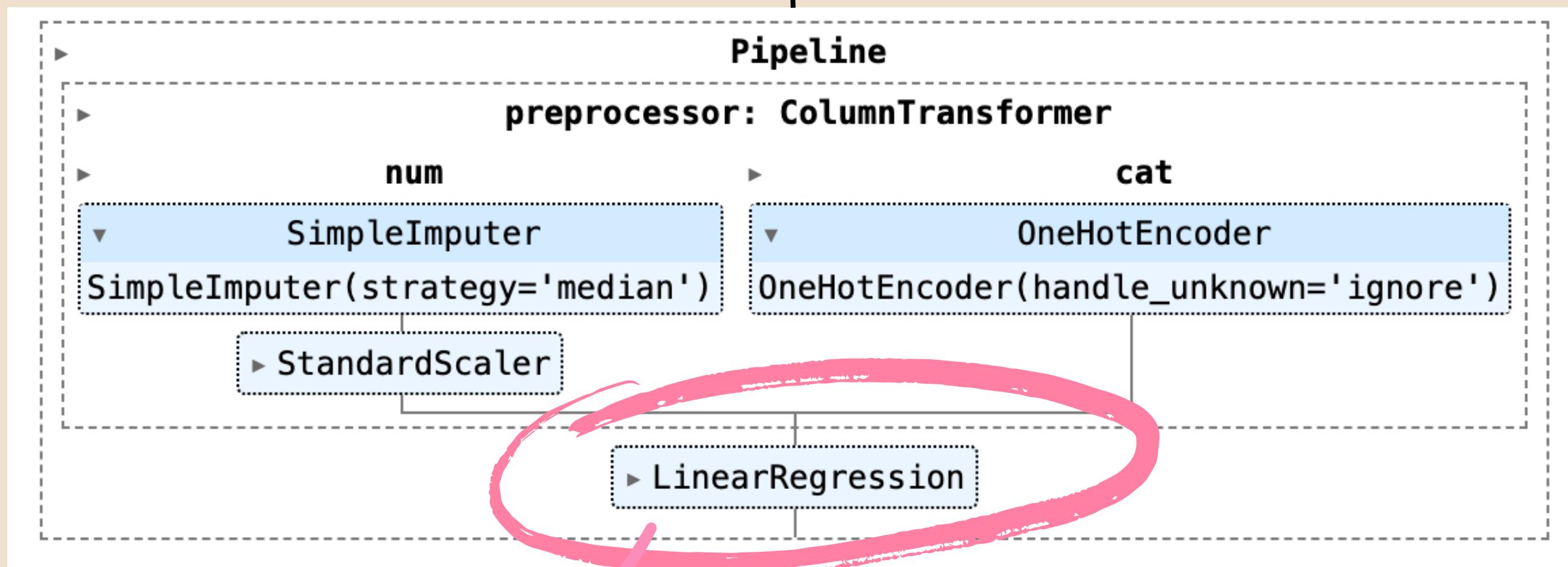
Feature engineering

In-house preprocessing



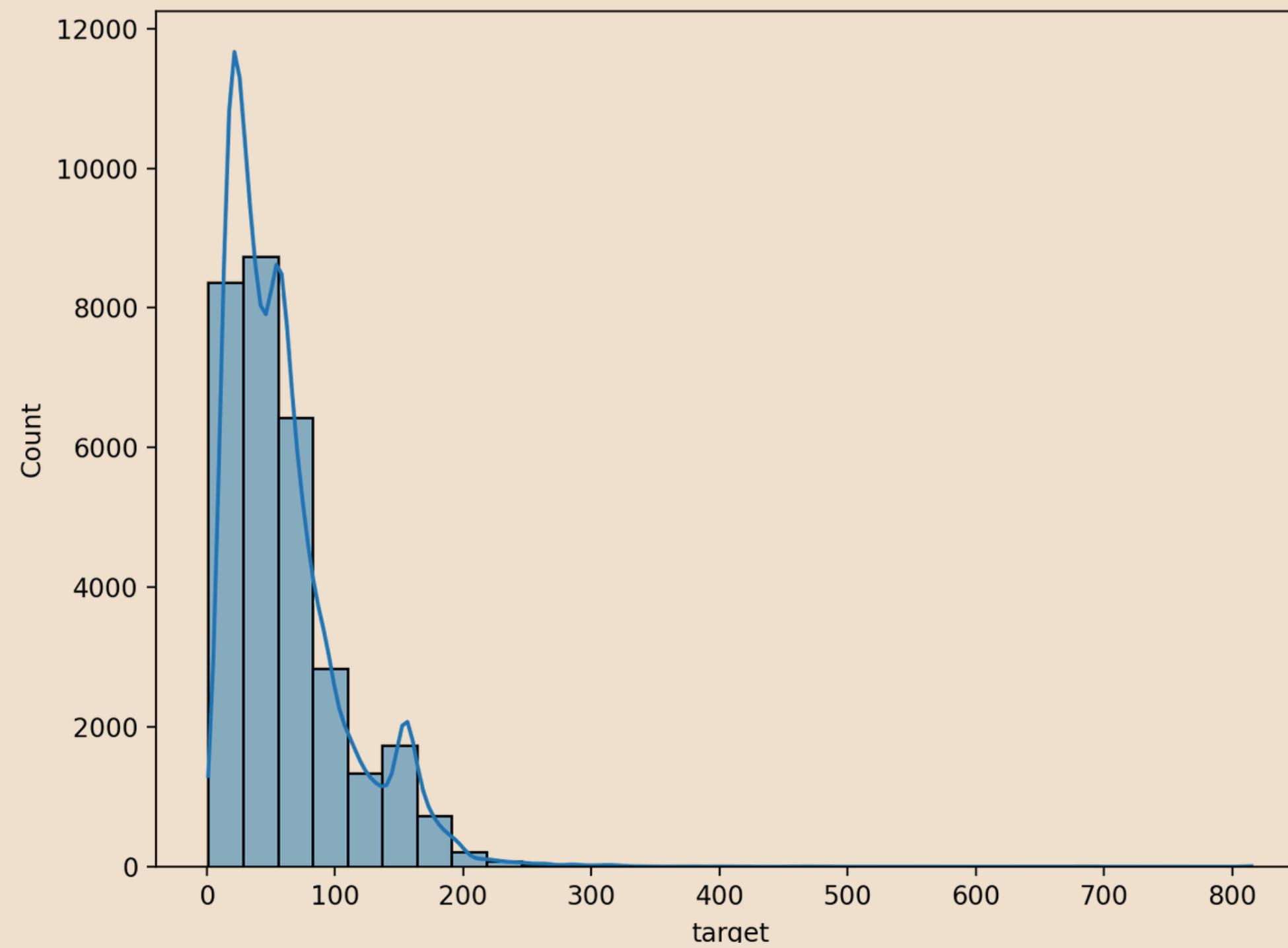
Feature engineering

In-house preprocessing

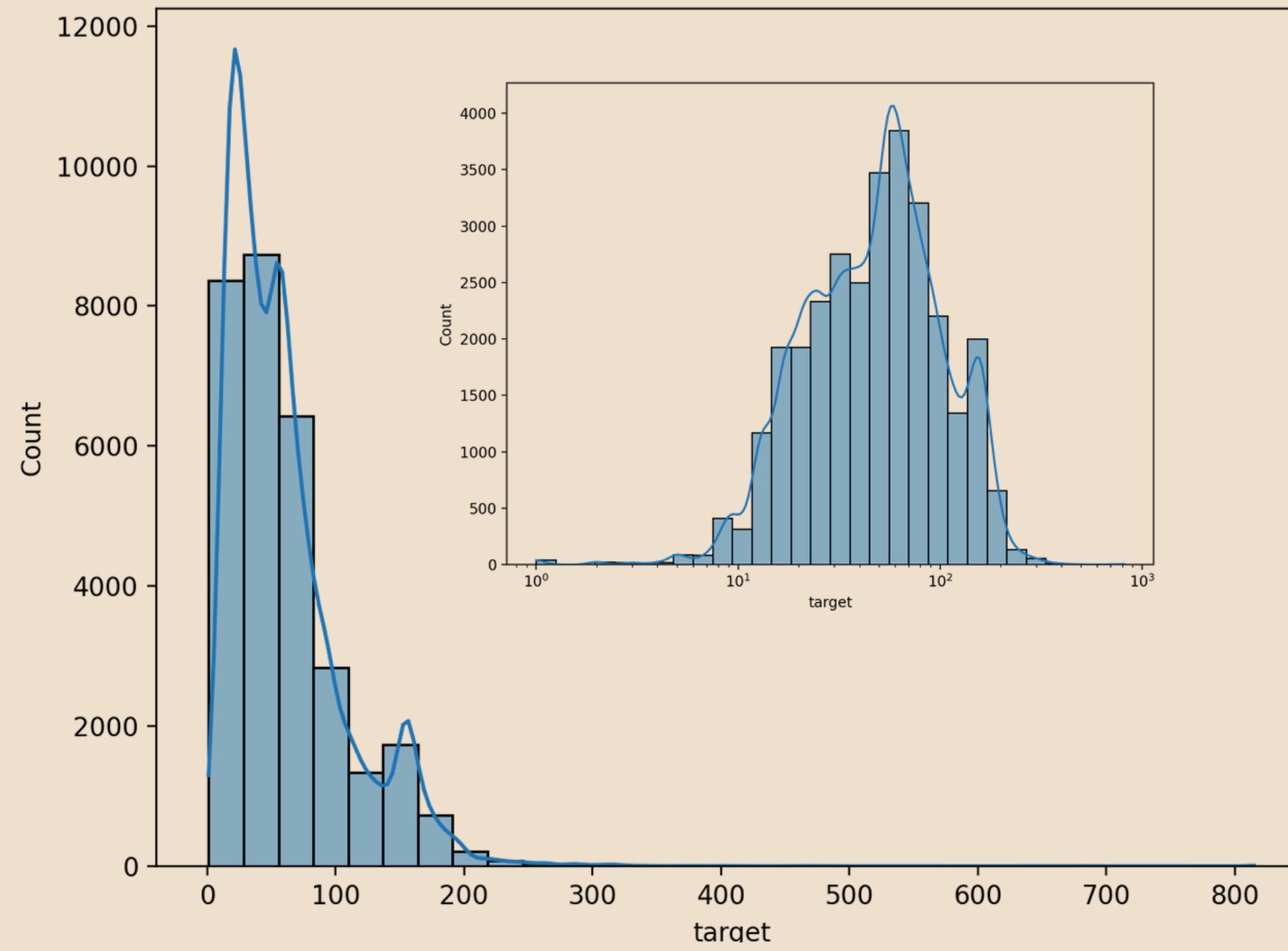


Grid search (best parameter)

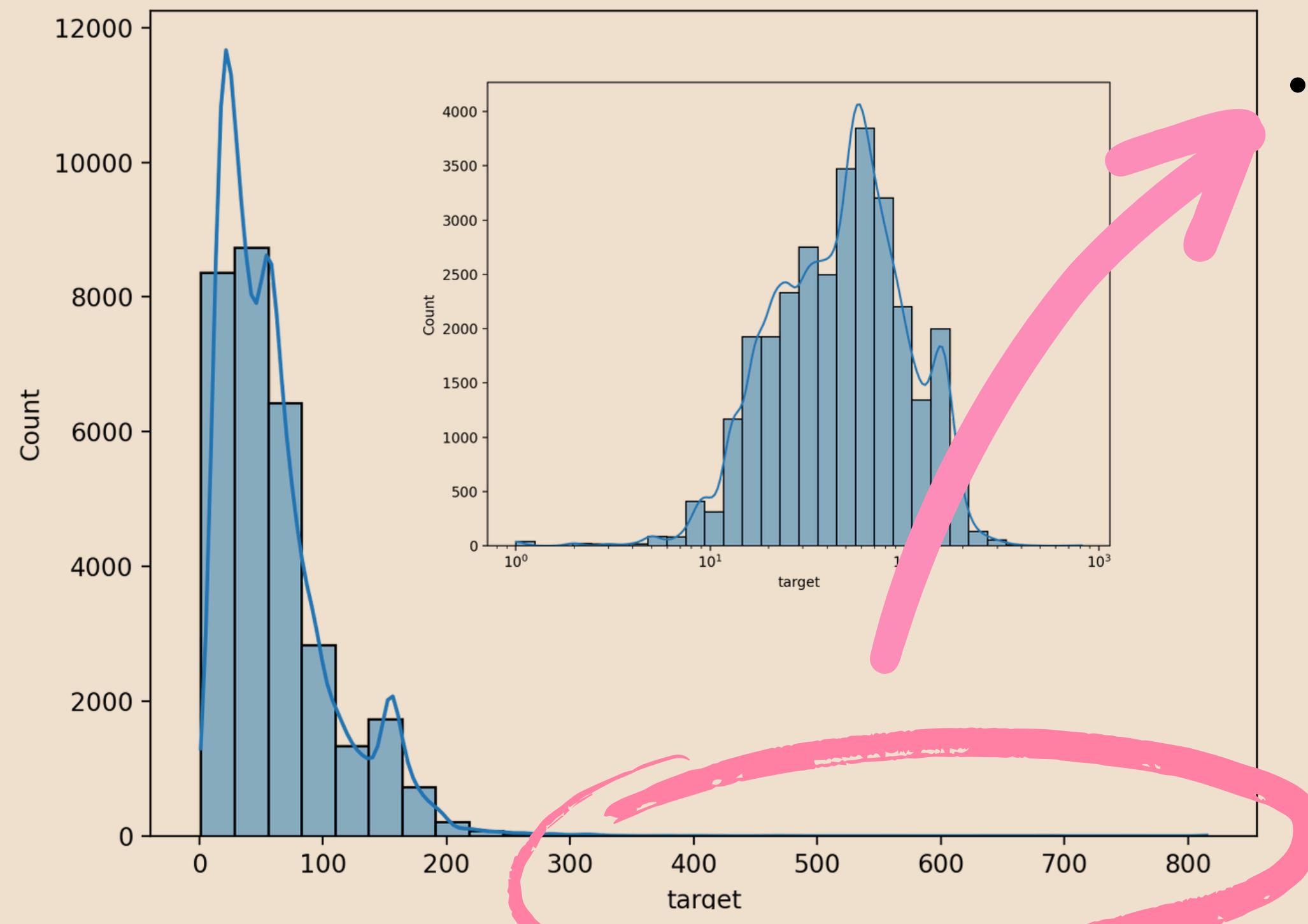
Feature engineering



Feature engineering

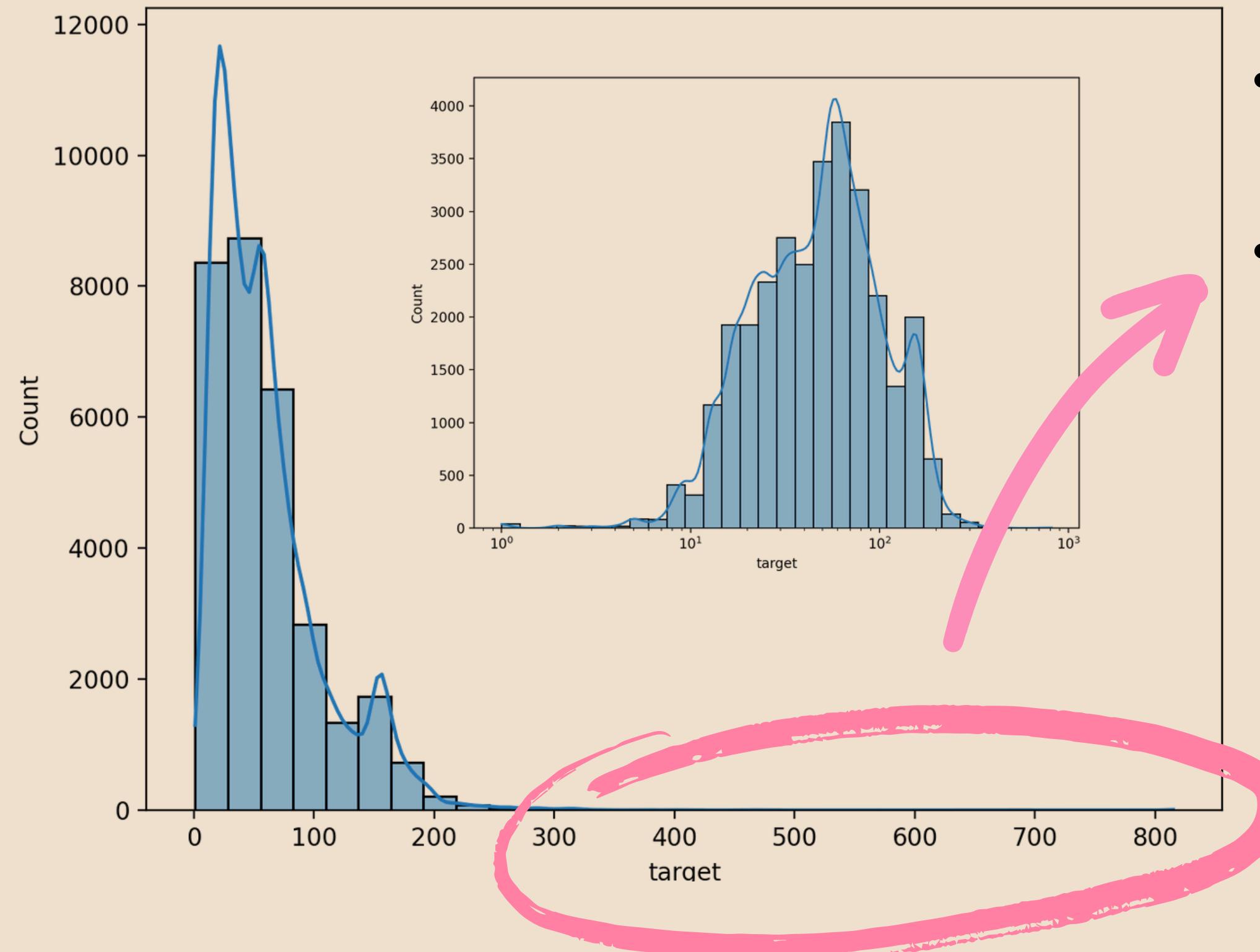


Feature engineering



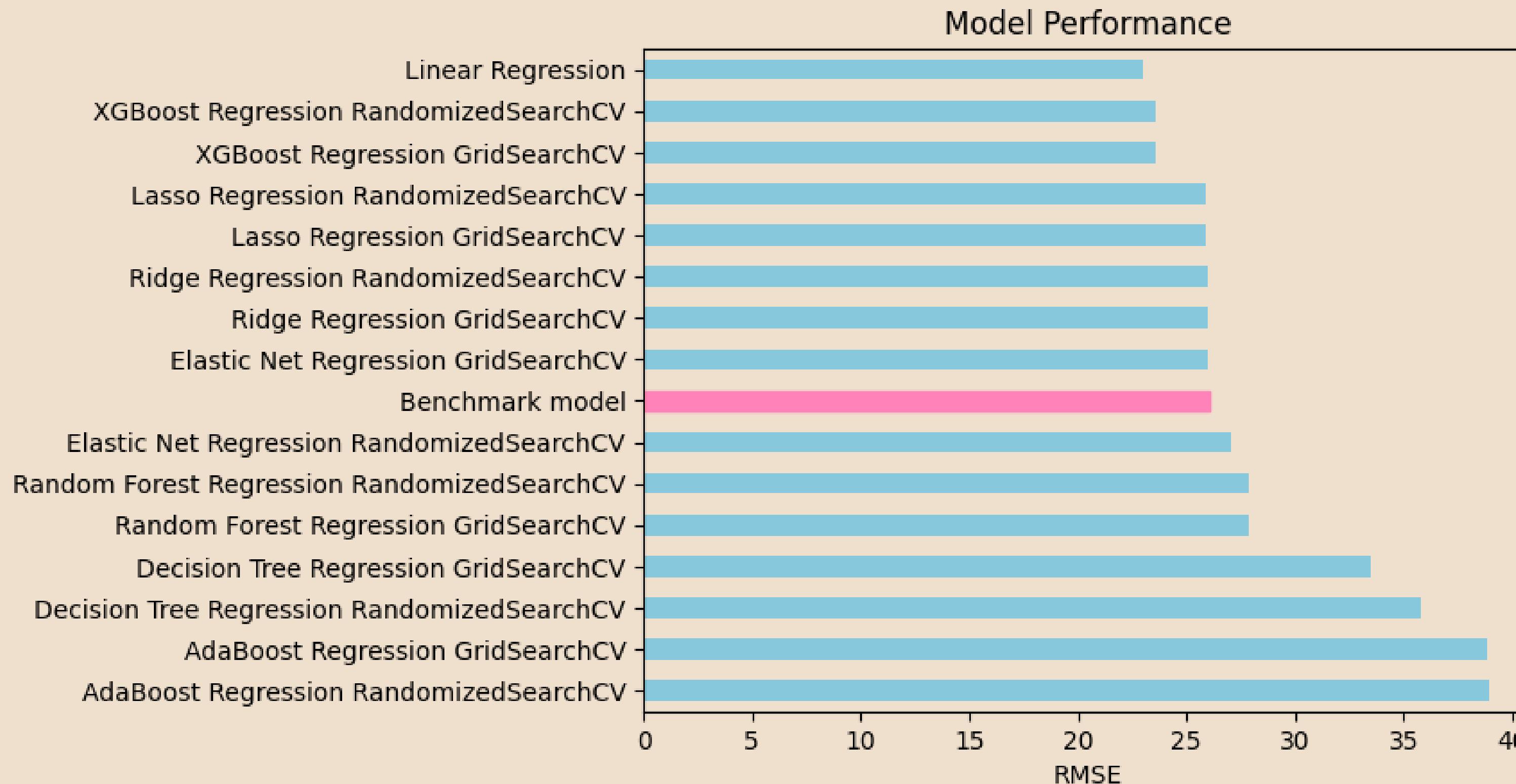
- We have a highly skewed target distribution.

Feature engineering



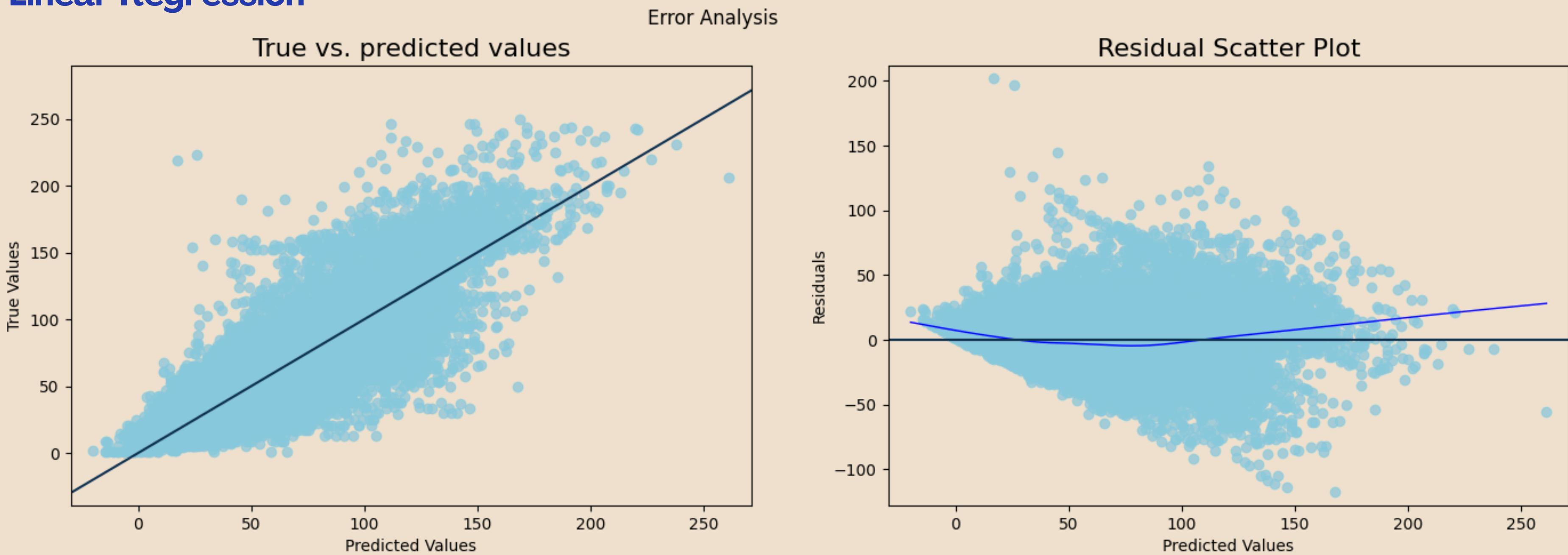
- We have a highly skewed target distribution.
- It is convenient to avoid the tail so we use a target threshold.

Model comparison



Best model

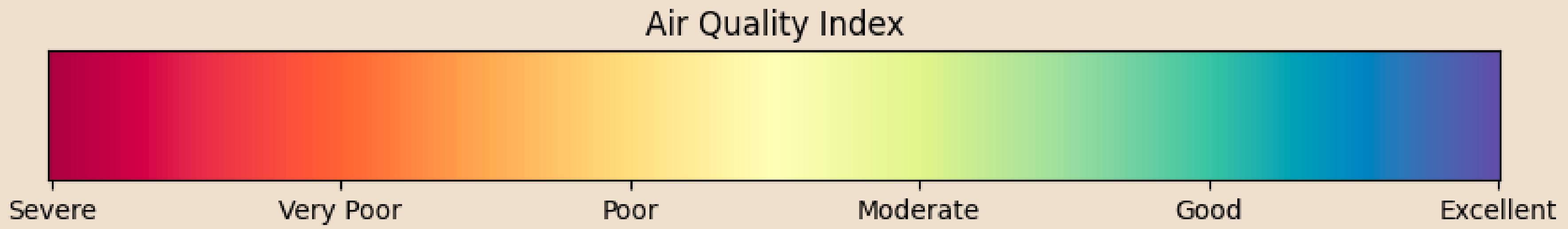
Linear Regression



RMSE: 22.99

Adj. R²: 0.72

Implementation

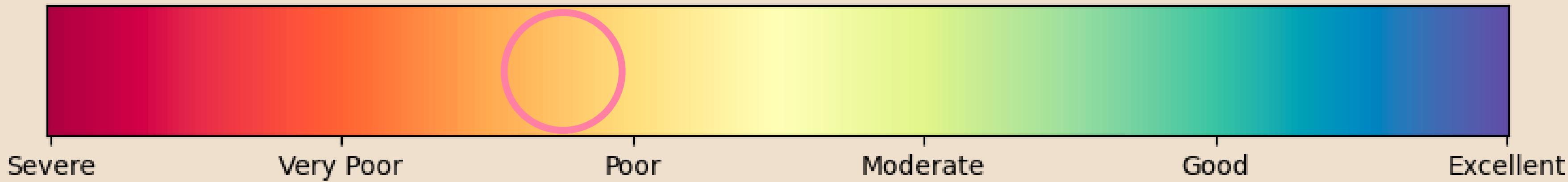


Predication of air quality at every time & every location

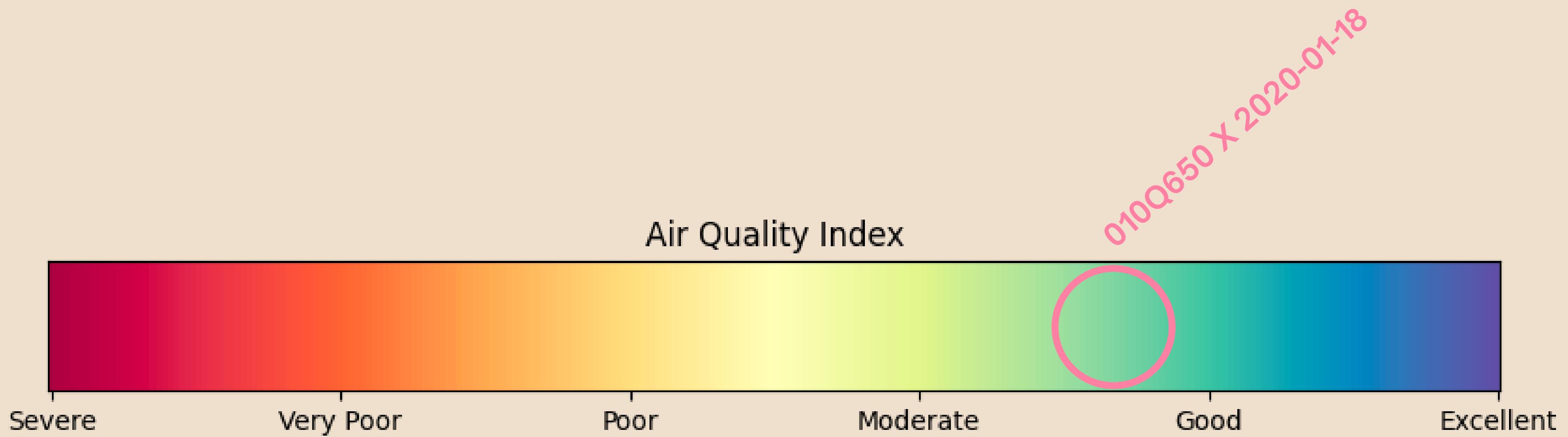
Implementation

010Q650X2020-01-02

Air Quality Index



Implementation



Summary

01. Successful prediction of air quality index by satellite data → beats benchmark model
02. Fast model allows real-time prediction → implemented as app on company's homepage
03. Contribution to unified world-wide air quality information and support of environmental awareness & protection → open access



Thanks