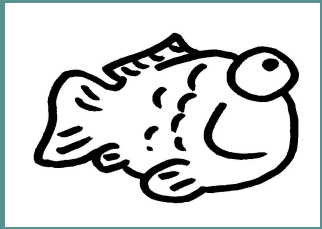
AirQo Ugandan Air Quality Forecast Challenge



Maike, Jonas, Michael

Content

- Background
- EDA
- Prediction Model
- Results
- Next Steps

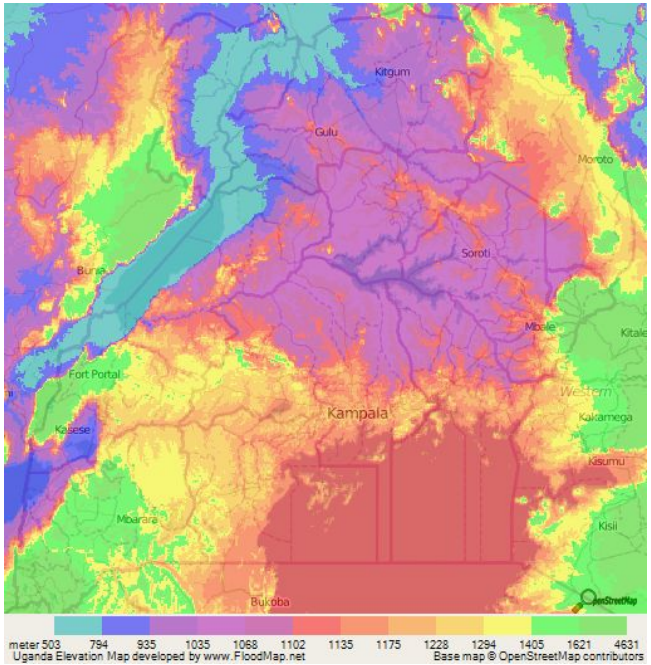
Background

About Uganda:

- Located in central eastern Africa north of Lake Victoria
- Mostly tropical climate (northern regions less humid)
- Depending on agriculture

Task:

- **AirQo: Predict the air quality one day in the future using weather data from the past 5 days**
- **Goal: Best RMSE**



https://zindi.africa/competitions/airqo-ugandan-air-quality-forecast-challenge/

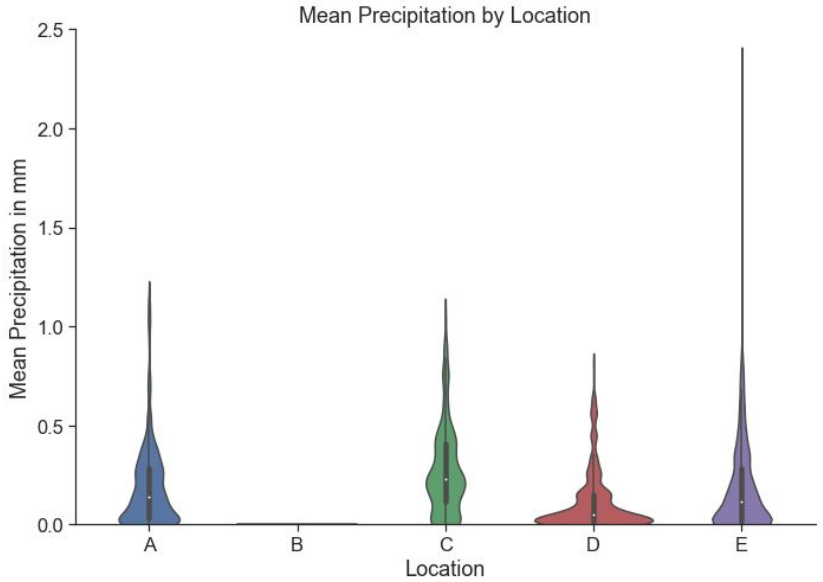
Weather Sensors

Weather Data
from 5 Tahmo-Weather-Stations

- Temperature (°C)
- Precipitation (mm)
- Relative Humidity
- Wind Direction (degrees)
- Wind Speed (m/s)
- Atmospheric Pressure (kPa)

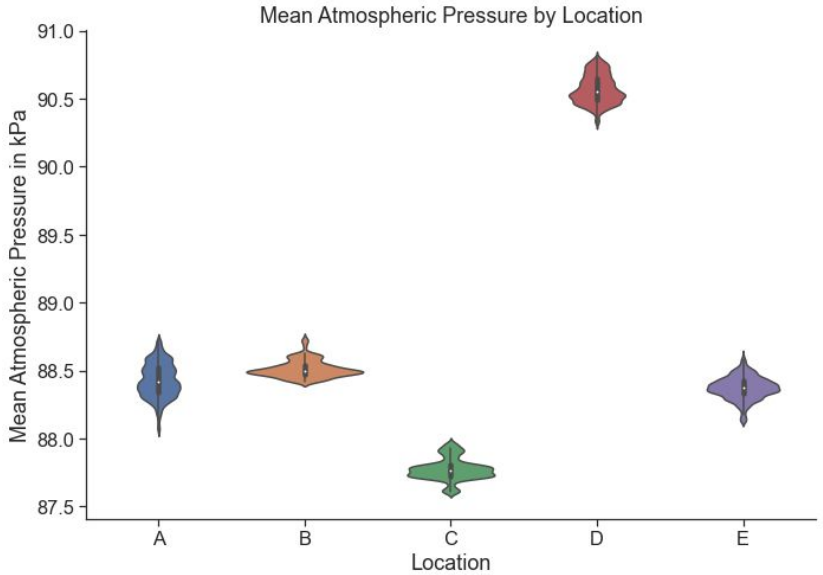
Source: Tahmo.org

Weather Data by Location



- No rain in location B
- Mostly raining in location C

Weather Data by Location

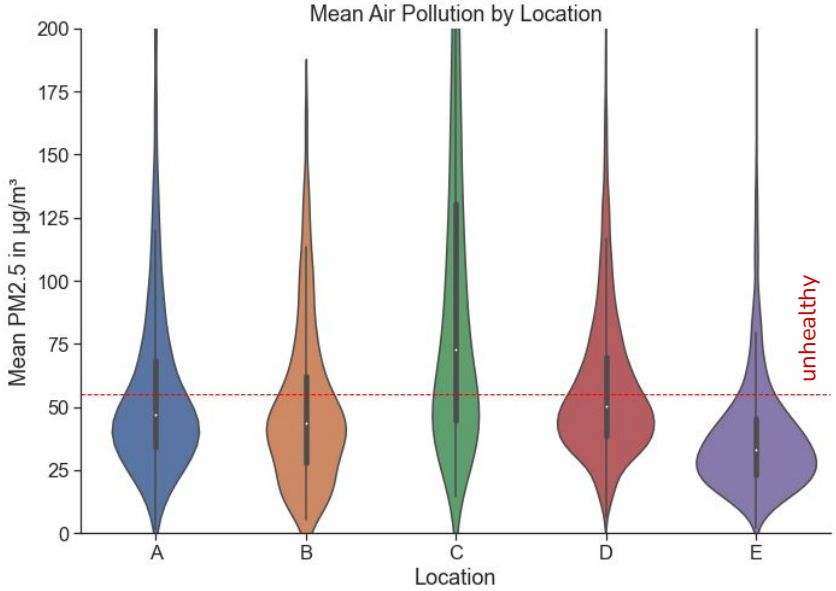


- Location D located far lower than other locations
- Location C mostly in low-pressure area

Air Pollution by Location

particulate matter smaller than 2.5 micrometers in diameter

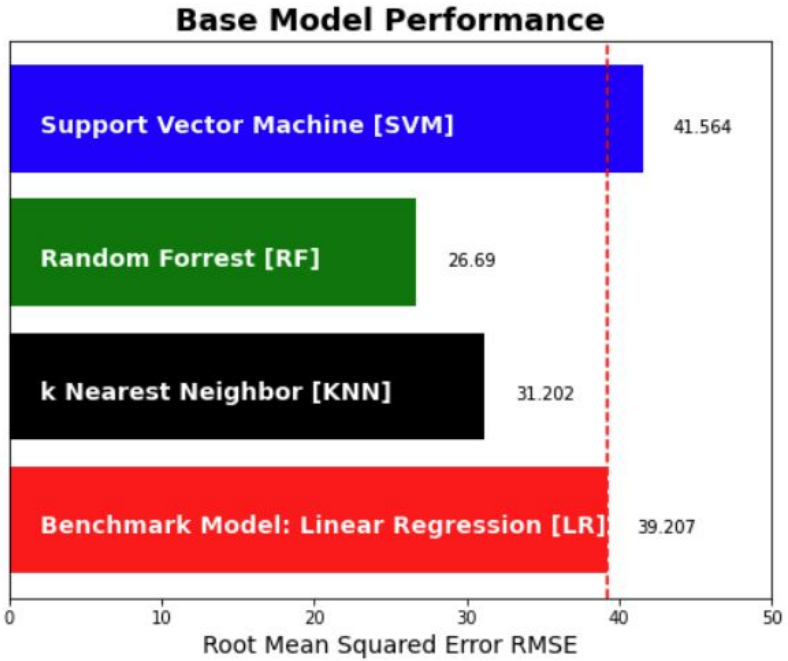
Health Concern	$PM_{2.5}$ (μgm^{-3})	Precautions
Good	0 - 12	None
Moderate	13 - 35	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	36 - 55	Sensitive groups should reduce prolonged or heavy exertion
Unhealthy	56 - 150	Everyone should reduce prolonged or heavy exertion, take more breaks during outdoor activities
Very Unhealthy	151 - 250	Everyone should avoid prolonged or heavy exertion, move activities indoors or reschedule
Hazardous	250 +	Everyone should avoid all physical activities outdoors.



Analysis of Base Model

Definition Base Model:

- ❏ Simplest Model
- ❏ Only Descriptive Statistics Values w/o Time Series
- ❏ No Further Parameter Tuning

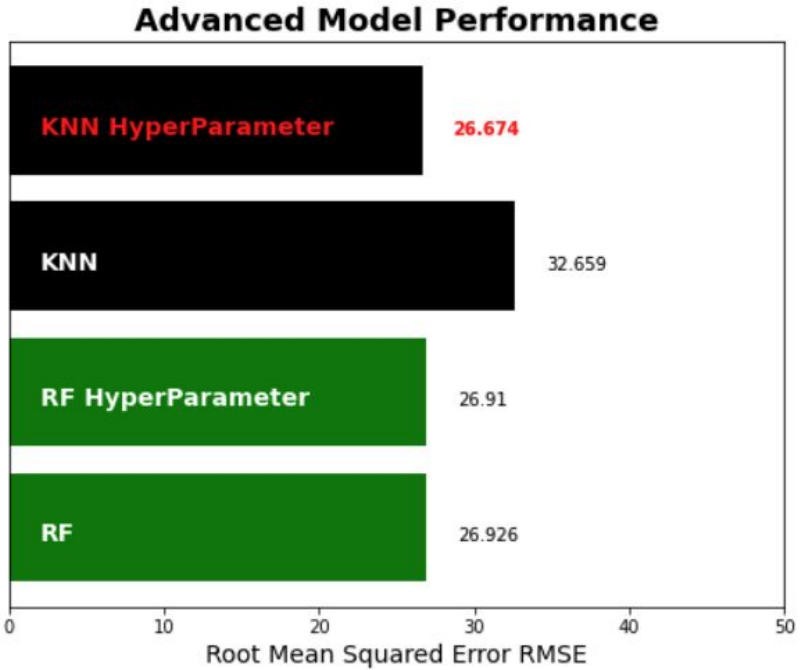


- **RF- & KNN Regressor better than base model LR → Both will be taken into account for further investigations**
- SVM worse than LR → Regressor will not be considered furthermore

Analysis of Advanced Model

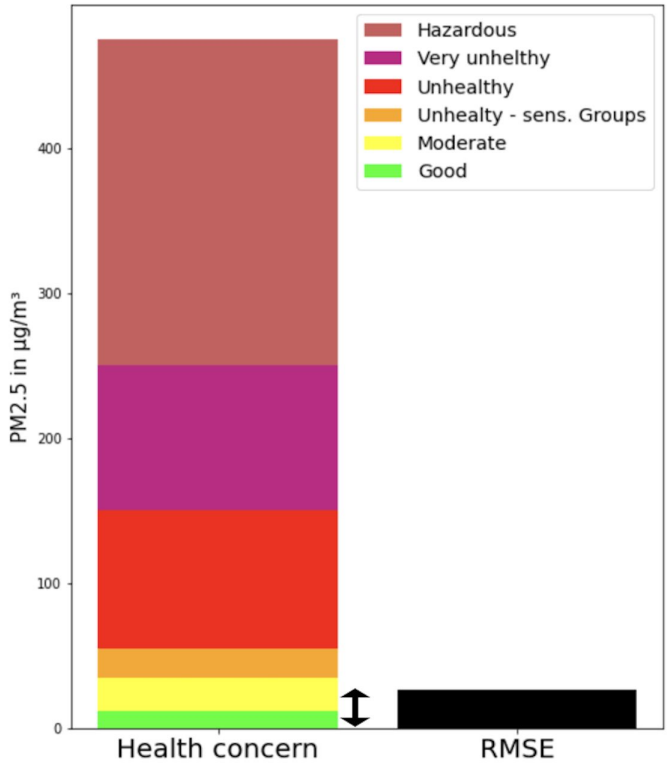
Definition Advanced Model:

- ❏ Descriptive Statistics Values and Time Series
- ❏ Including Feature Engineering
- ❏ Including Hyper Parameter Tuning



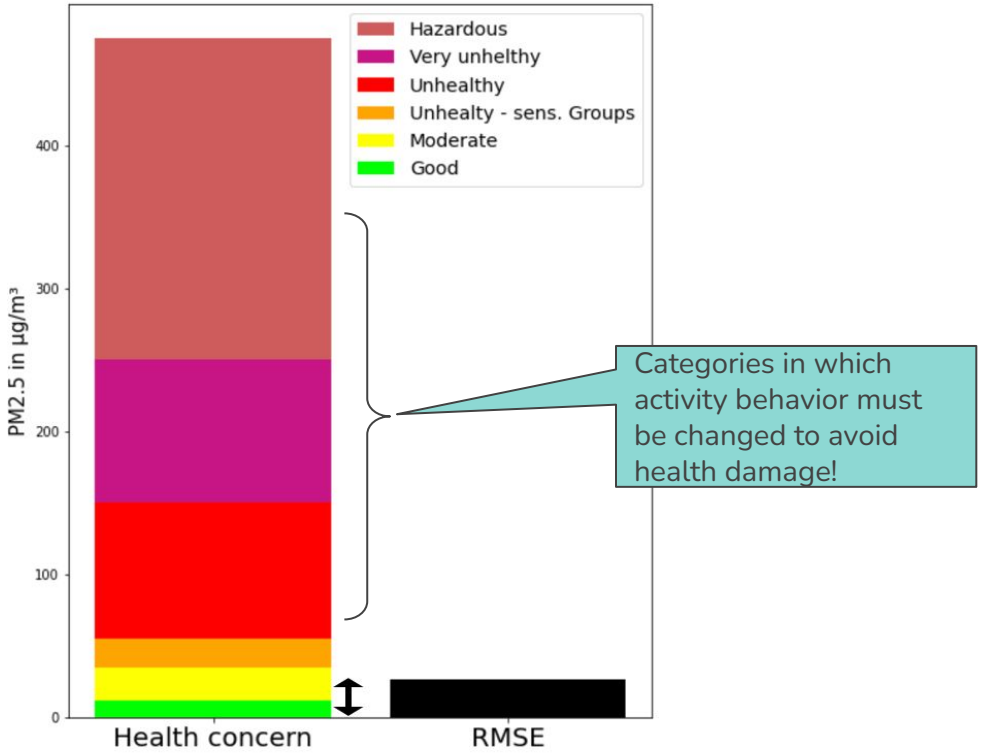
- **Parameter Tuning at KNN with big effect, KNN best model with RMSE: 26.67**
- Parameter Tuning at RF does not achieve noteable improvement
- Overall, consideration time data does not improve RMSE

Category Range Vs. RMSE

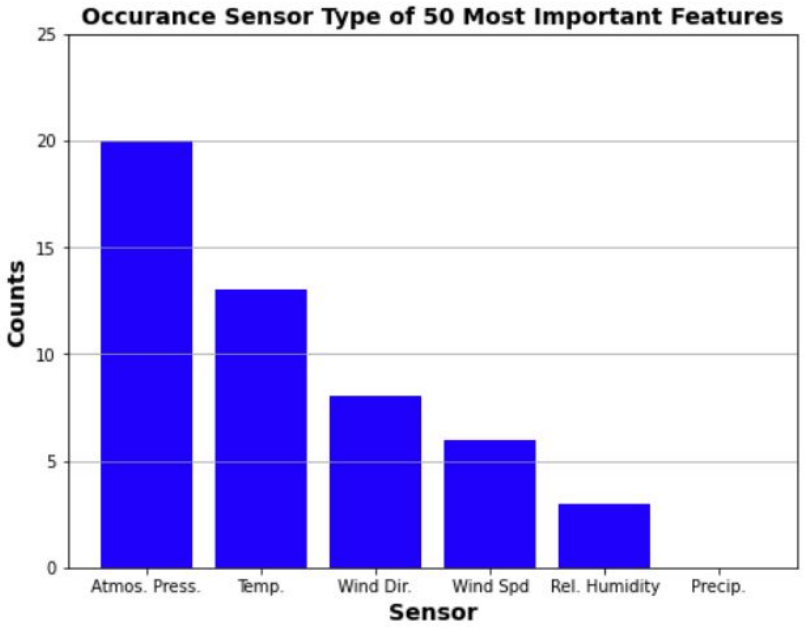


- RMSE within healthy categories (good, moderate range) relatively big but..
- in the unhealthy range good to distinguish!

Category Range Vs. RMSE



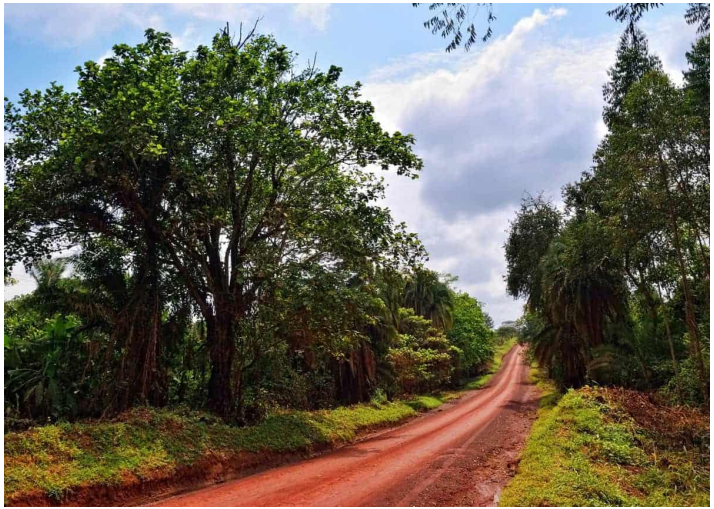
Feature Importance



- Atmospheric pressure, temperature and wind direction are most important features
- Not relevant is precipitation

Next Steps

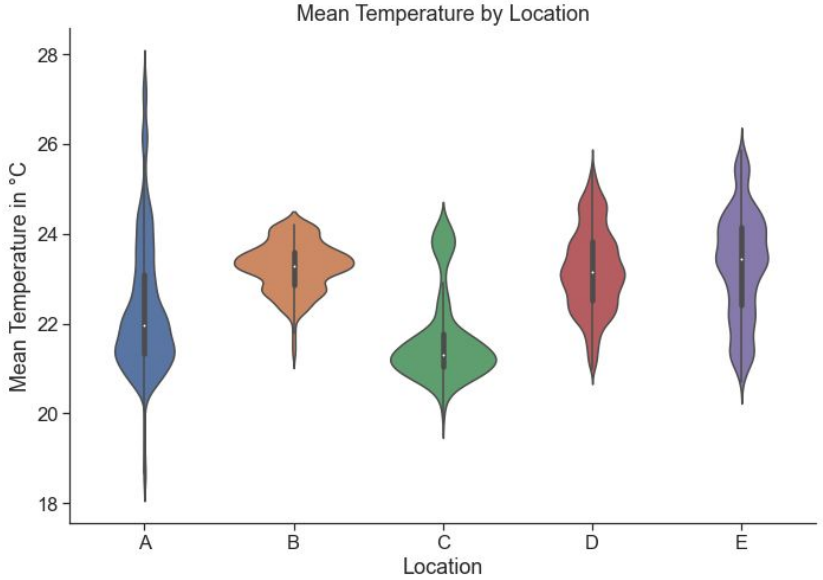
- Choose more suitable method to replace the missing values.
- Use other regression models that may be more appropriate.
- Improve feature engineering.
- Take other data into account.



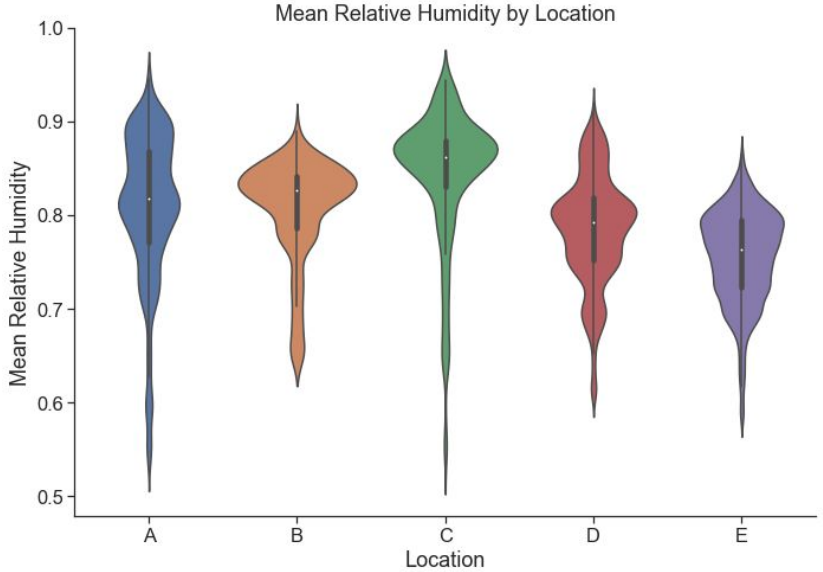
Thank you for your attention!

Attachment

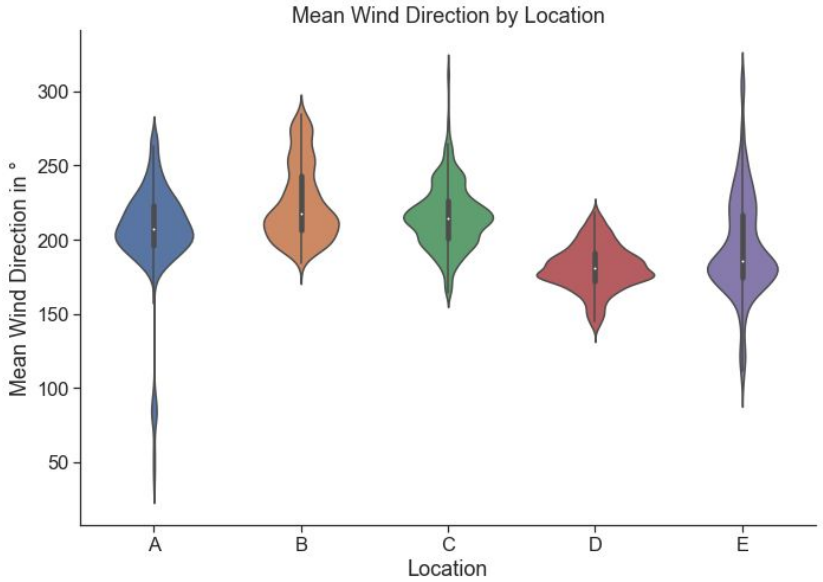
Weather Data by Location



Weather Data by Location



Weather Data by Location



Weather Data by Location