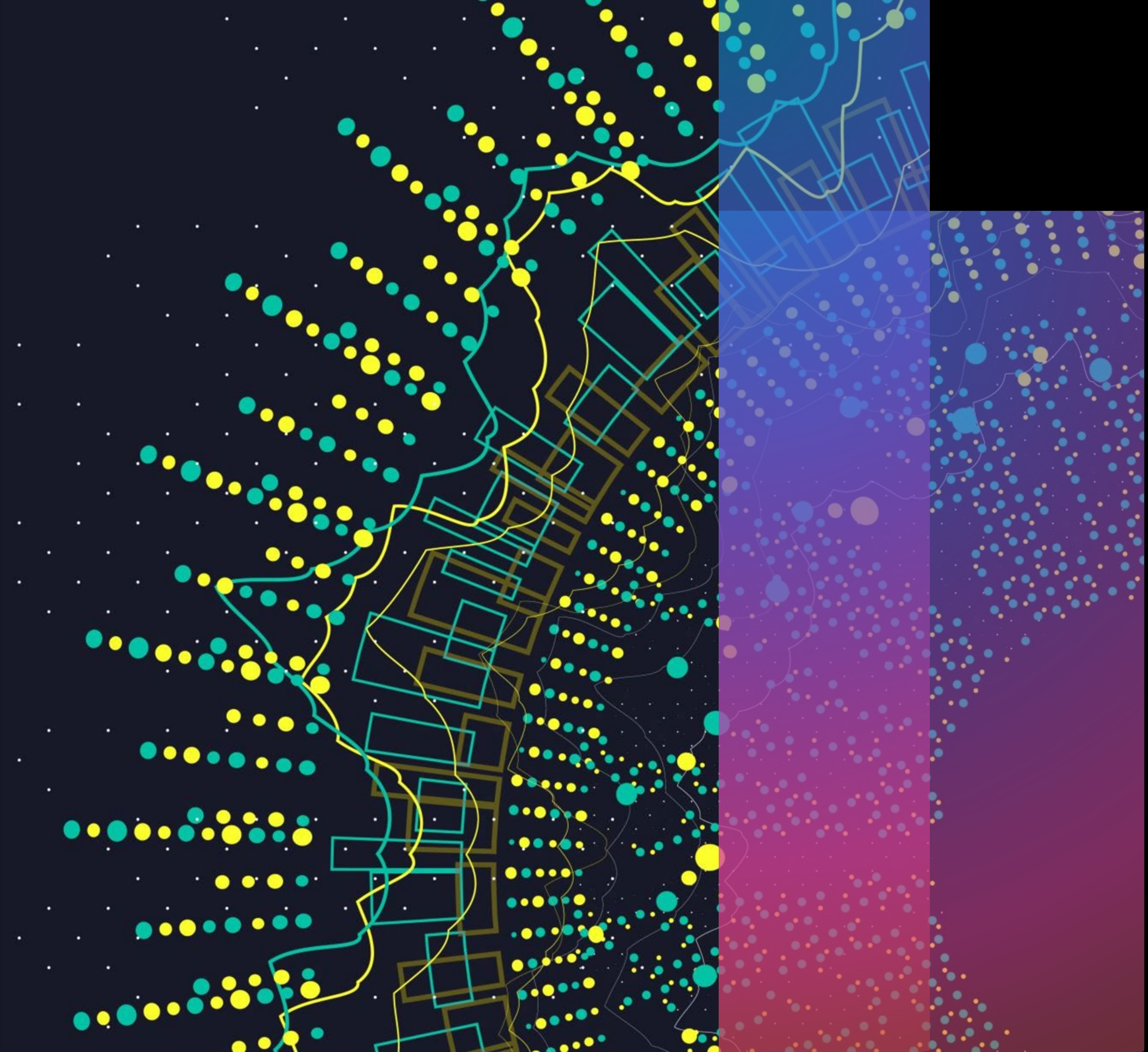


From EarthData to Action: Cloud Computing with Earth Observation Data for Predicting Cleaner, Safer Skies

Group members:

Esteban Andres Flores Calderon

Vrajesh Kumar

Kareena Zaman Eula

Sumnath Das Biraz

Ahsan Habib Niloy

Factors considered to predict smoke:

- Wind direction
- Wind speed
- Near surface air temperature
- Location and intensity of active fires
- Time of year
- Location of user (lat, long)
- Rain

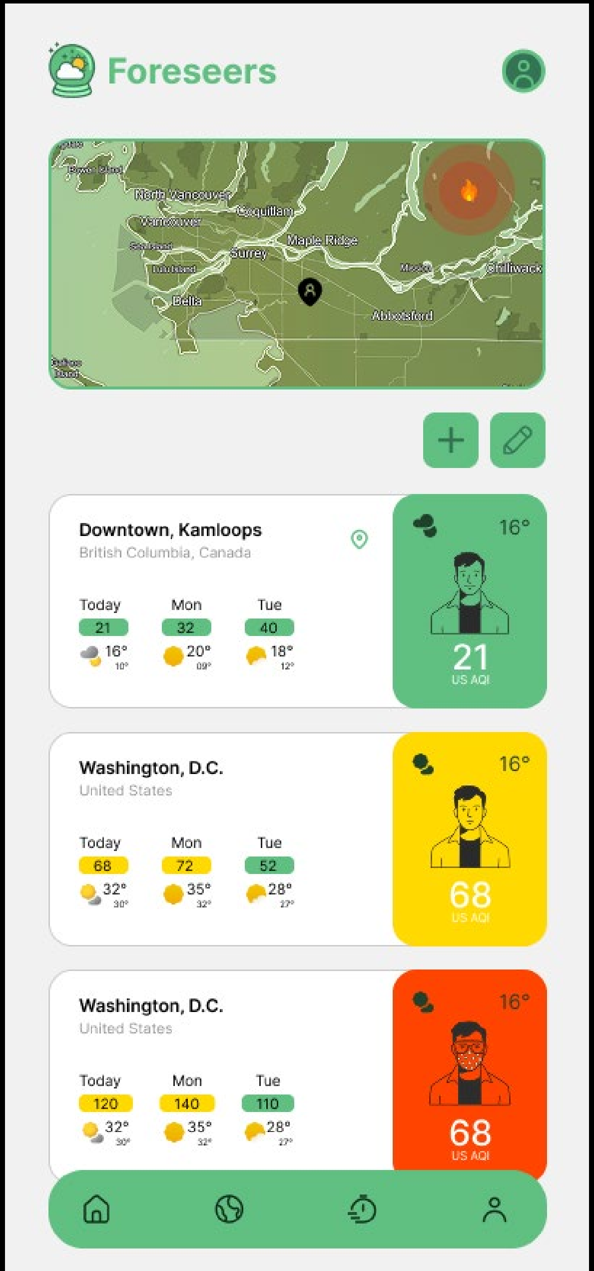
EF = Emission Coefficient of Pollutant

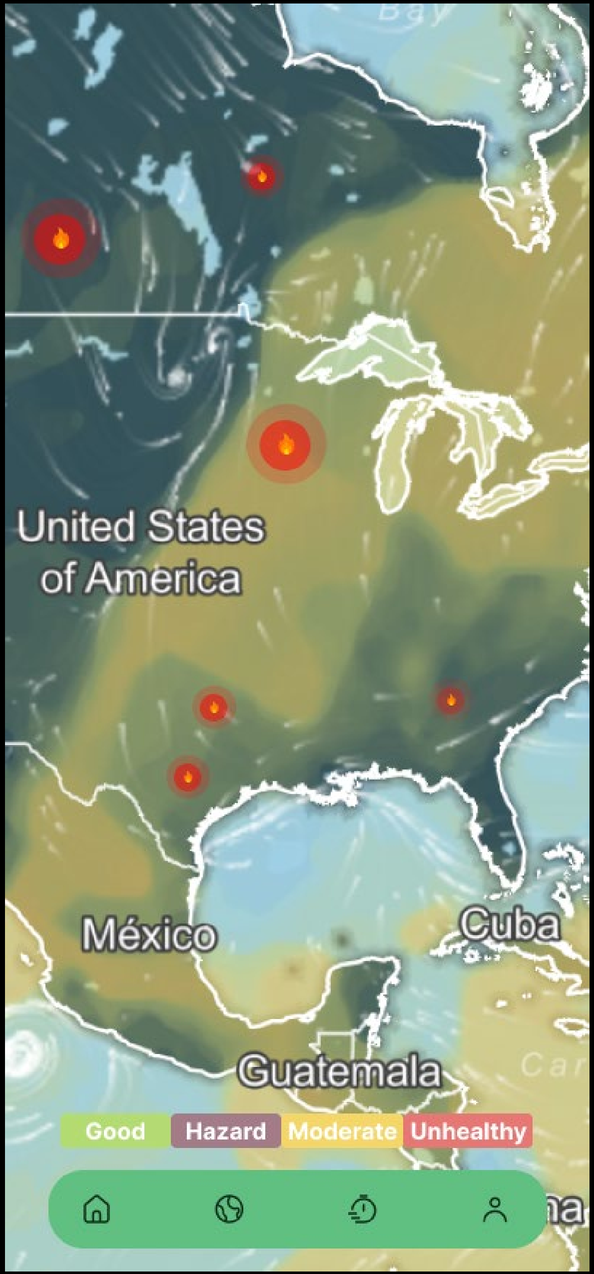
Dataset Used => FIRMS -> **Fire Radiative Power (FRP)** Archive dependent on time, latitude, and longitude, per year, per country

EF = Emission Factor of Pollutant

EF(PM2.5 average at flaming stages of leaves and branches) = 10.91 g/kg (uncertainty of 3.32 g/kg)

Front End View Concept





There is a fire 20KM away
You need to evacuate from the area.

Smoke Confirmed
You need to precaution if you’r planning to go outside.

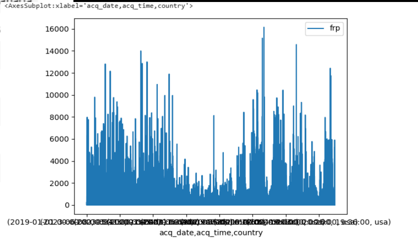
Probable Incoming Smoke
There is a probability of smoke reaching your area.

Machine Learning Attempt

acq_date	acq_time	latitude	longitude	brightness	scan	track	satellite	instrument	confidence	version	bright_t31	frp	daynight	type	country
2019-01-01	541	55.9807	-122.7407	339.4	1.1	1.0	Terra	MODIS	100	6.03	266.6	53.3	N	0	canada
	1910	55.2518	-111.7735	336.5	1.4	1.2	Terra	MODIS	90	6.03	270.4	69.0	D	0	canada
	2059	57.0958	-118.5595	310.5	1.2	1.1	Aqua	MODIS	70	6.03	269.3	20.8	D	0	canada
	2059	57.0937	-118.5797	331.4	1.2	1.1	Aqua	MODIS	82	6.03	270.0	48.9	D	0	canada
2019-01-02	2003	56.7630	-111.1588	302.8	1.0	1.0	Aqua	MODIS	54	6.03	269.9	10.0	D	0	c
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2024-12-31	1940	19.4107	-155.2601	319.7	2.2	1.5	Terra	MODIS	45	61.03	297.9	47.4	D	1	
	1940	19.4129	-155.2808	389.8	2.3	1.5	Terra	MODIS	100	61.03	322.7	638.8	D	1	
	1940	19.4023	-155.3029	398.1	2.3	1.5	Terra	MODIS	100	61.03	304.7	779.9	D	1	
	1940	19.4001	-155.2823	500.0	2.3	1.5	Terra	MODIS	100	61.03	342.8	5239.8	D	1	
	2120	33.4155	-110.8629	309.2	1.1	1.0	Aqua	MODIS	68	61.03	295.4	7.6	D	2	

1869465 rows × 14 columns

<AxesSubplot:xlabel='acq_date,acq_time,country'>



	acq_date	acq_time	country	latitude	longitude	frp	year	month	dayofyear	hour	minute
0	2019-01-01	05:41:00	canada	55.9807	-122.7407	53.3	2019	1	1	5	41
1	2019-01-01	19:10:00	canada	55.2518	-111.7735	69.0	2019	1	1	19	10
2	2019-01-01	20:59:00	canada	57.0958	-118.5595	20.8	2019	1	1	20	59
3	2019-01-01	20:59:00	canada	57.0937	-118.5797	48.9	2019	1	1	20	59
4	2019-01-02	20:03:00	canada	56.7630	-111.1588	10.0	2019	1	2	20	3
...	...	...	...	...	...	...	...	...	...	...	...
1869460	2024-12-31	19:40:00	usa	19.4107	-155.2601	47.4	2024	12	366	19	40
1869461	2024-12-31	19:40:00	usa	19.4129	-155.2808	638.8	2024	12	366	19	40
1869462	2024-12-31	19:40:00	usa	19.4023	-155.3029	779.9	2024	12	366	19	40
1869463	2024-12-31	19:40:00	usa	19.4001	-155.2823	5239.8	2024	12	366	19	40
1869464	2024-12-31	21:20:00	usa	33.4155	-110.8629	7.6	2024	12	366	21	20

1869465 rows × 11 columns

```python
import pandas as pd
import numpy as np


frp_ds["year"] = frp_ds["acq_date"].dt.year
frp_ds["month"] = frp_ds["acq_date"].dt.month
frp_ds["dayofyear"] = frp_ds["acq_date"].dt.dayofyear

frp_ds["hour"] = frp_ds["acq_time"].apply(lambda t: t.hour)
frp_ds["minute"] = frp_ds["acq_time"].apply(lambda t: t.minute)


X = frp_ds[["latitude", "longitude", "month", "dayofyear", "hour"]]


y = frp_ds["frp"]

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

from sklearn.ensemble import RandomForestRegressor

model = RandomForestRegressor(
    n_estimators=100,
    random_state=42,
    n_jobs=-1
)
model.fit(X_train, y_train)

from sklearn.metrics import mean_squared_error, r2_score

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse:.3f}")
print(f"R² Score: {r2:.3f}")
```

```
Mean Squared Error: 36060.164
R² Score: 0.450
```

Model Accuracy based on R^2 value: **0.45** (mildly weak)

Main Problem: Does not consider other factors like vegetation density or temperature

Learning algorithm used: Random Forest Regressor

We tried to use other datasets (MERRA-2) to feed in information such as temperature, humidity but were unsuccessful

Parameter(s):
T2M MERRA-2 Temperature at 2 Meters (C)
QV2M MERRA-2 Specific Humidity at 2 Meters (g/kg)
PRECTOTCORR MERRA-2 Precipitation Corrected (mm/hour)
ALLSKY_SFC_SW_DWN CERES SYN1deg All Sky Surface Shortwave Downward Irradiance (MJ/hr)
WS2M MERRA-2 Wind Speed at 2 Meters (m/s)
WD2M MERRA-2 Wind Direction at 2 Meters (Degrees)
PS MERRA-2 Surface Pressure (kPa)
WSC MERRA-2 Corrected Wind Speed (Adjusted For Elevation) (m/s)
Message(s):
Corrected Wind Speed has a custom surface implemented: Airport: flat rough grass (airportgrass)
-END HEADER-

YEAR	MO	DY	HR	T2M	QV2M	PRECTOTCORR	ALLSKY_SFC_SW_DWN	WS2M	WD2M	PS	WSC
2025	9	7	0	19.11	12.17	1.53	-999.0	0.35	157.9	95.1	0.41
2025	9	7	1	18.67	12.4	0.82	-999.0	0.34	162.7	95.05	0.43
2025	9	7	2	17.95	12.6	2.59	-999.0	0.45	146.7	95.09	0.71
2025	9	7	3	17.46	12.28	2.94	-999.0	0.58	120.1	95.15	0.75
2025	9	7	4	17.34	12.1	2.61	-999.0	0.53	110.8	95.18	0.68
2025	9	7	5	17.11	11.96	3.17	-999.0	0.57	148.5	95.22	0.72
2025	9	7	6	16.73	11.74	4.33	-999.0	0.67	163.5	95.26	0.86
2025	9	7	7	16.51	11.52	2.36	-999.0	0.59	153.9	95.27	0.7
2025	9	7	8	16.27	11.39	1.88	-999.0	0.54	142.5	95.28	0.66
2025	9	7	9	16.01	11.3	1.23	-999.0	0.57	145.7	95.28	0.73
2025	9	7	10	15.79	11.25	1.25	-999.0	0.59	149.5	95.3	0.75
2025	9	7	11	15.57	11.16	1.25	-999.0	0.54	158.2	95.33	0.7
2025	9	7	12	15.33	11.04	1.51	-999.0	0.54	168.3	95.34	0.66
2025	9	7	13	15.17	10.94	2.03	-999.0	0.48	174.1	95.34	0.6
2025	9	7	14	15.13	10.91	2.09	-999.0	0.4	168.4	95.35	0.49
2025	9	7	15	15.17	10.87	1.76	-999.0	0.41	162.9	95.35	0.46
2025	9	7	16	15.35	10.85	1.68	-999.0	0.57	153.9	95.39	0.6
2025	9	7	17	15.61	10.83	2.02	-999.0	0.83	147.8	95.45	0.89

Machine Learning Model Proof of Concept Usage

```
[63]: FEATURES = ["latitude","longitude","month","dayofyear","hour"]

sample = pd.DataFrame(
    [[54.0685, -115.9707, 1, 12, 20]],
    columns=FEATURES
)

pred = model.predict(sample)[0]
print(pred)

47.014000000000024
```

	acq_date	acq_time	country	latitude	longitude	frp	year	month	dayofyear	hour	minute
20	2019-01-05	18:47:00	canada	51.2238	-122.1109	99.5	2019	1	5	18	47
21	2019-01-07	20:21:00	canada	52.5133	-115.2146	0.0	2019	1	7	20	21
22	2019-01-10	19:05:00	canada	52.8514	-118.0757	20.2	2019	1	10	19	5
23	2019-01-10	20:52:00	canada	54.3379	-118.7933	142.0	2019	1	10	20	52
24	2019-01-11	04:39:00	canada	54.7290	-115.7212	65.4	2019	1	11	4	39
25	2019-01-11	18:15:00	canada	42.8011	-80.1078	10.4	2019	1	11	18	15
26	2019-01-12	20:40:00	canada	54.0685	-115.9707	54.8	2019	1	12	20	40
27	2019-01-12	20:40:00	canada	54.0664	-115.9860	28.7	2019	1	12	20	40
28	2019-01-12	20:40:00	canada	54.0778	-117.4422	16.1	2019	1	12	20	40

Mass of Pollutant (smoke mass) = FRP * EF(emission factor) -> consider time is 1 for standard purposes

Mass of Pollutant (smoke mass PM2.5) = 47.014 * 10.91 g/kg = 513.22 g of PM2.5 in the area

Considering each pixel from the MODIS dataset is 1*10^6 m^2 and considering 1m of height

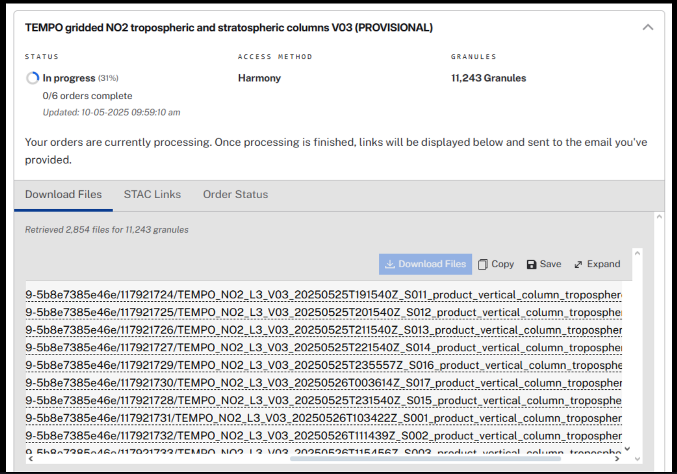
Values lesser than 35 ug/m^3 of PM2.5 is considered healthy

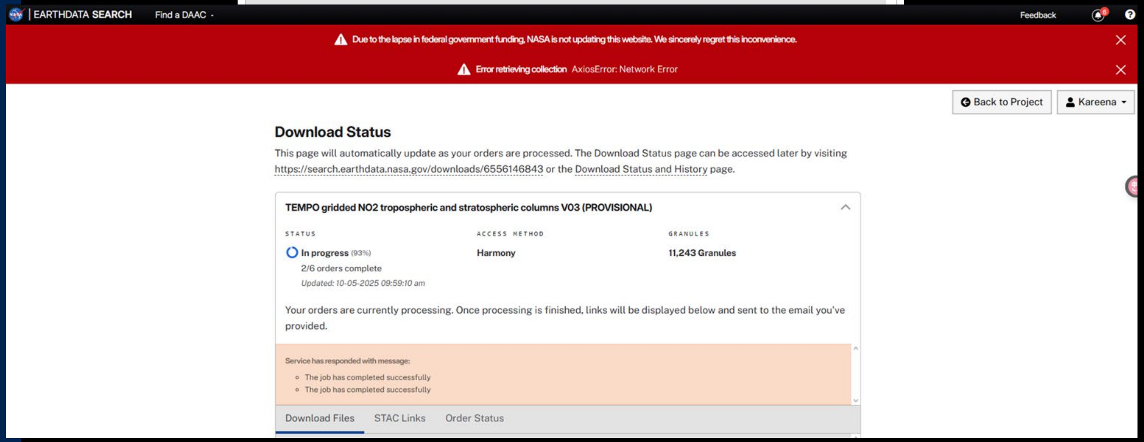
= 47.014 * 10.91 g/kg = 513.22 g of PM2.5 in the area
= 513.22/(1*10^6) = 5.1322*10^-4 g/m^3
= 5.1322*10^-4*10^-6 = 5.1322*10-10 ug/m^3 in a single second or in 24 hrs...
= 5.1322*10-10 = 123.17*10^-10 ug/m^3 in 24 hrs (which is accordingly not a serious hazard, which makes sense since frp isnt as high as other value entries) + consider other factors since this prediction might be wrong

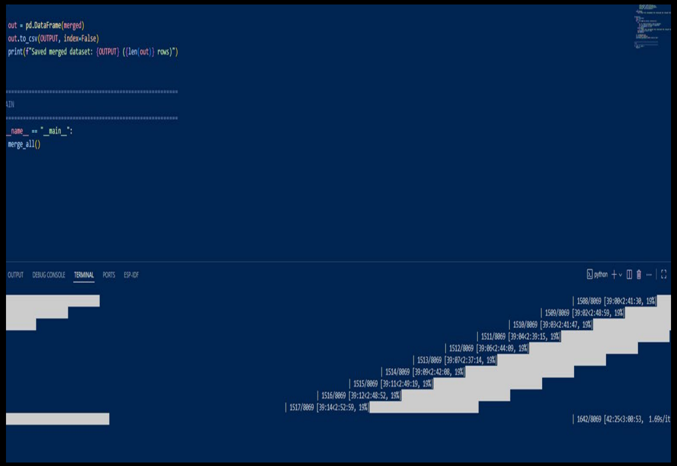
Trying to Improve Machine Learning Dataset

```python
def _parse_power(j):
        "wind_dir_deg": first("WD10M"),
        "precip_mm": first("PRECTOTCORR"),
    }


def fetch_weather(lat, lon, date):
    """Try NASA POWER, fallback to Open-Meteo."""
    ds = date.strftime("%Y%m%d")
    url = (f"{POWER_API}?parameters={POWER_PARAMS}&community={POWER_COMMUNITY}"
           f"&latitude={lat}&longitude={lon}&start={ds}&end={ds}&format={POWER_FMT}")
    for _ in range(MAX_RETRIES):
        try:
            r = requests.get(url, timeout=10)
            r.raise_for_status()
            j = r.json()
            if "properties" in j:
                return _parse_power(j)
        except Exception:
            time.sleep(1)
    # fallback: Open-Meteo
    try:
        om = ("https://archive-api.open-meteo.com/v1/archive"
              f"?latitude={lat}&longitude={lon}"
              f"&start_date={date:%Y-%m-%d}&end_date={date:%Y-%m-%d}"
              "&hourly=temperature_2m,wind_speed_10m,wind_direction_10m,precipitation")
```







Tried downloading data to add to FIRM dataset for ML model (code on left) but datasets are just too big to download or work with. Also tried to divert our analysis objective to NO2 air concentration in area for air quality (right image), but same problem.

We tried using AWS servers to process the data, but we couldn't get it to work since the data never fully downloaded and couldn't be accessed. Limited availability of other datasets online

Key Takeaways and Possible Future Improvements

- We need to feed more data that will allow us to **better predict FRP within an area at a certain time** by using other factors such as **vegetation and temperature**
- In the future, the prediction of fire smoke can be used to predict i**f the smoke will reach the user soon based on wind speed and direction,** as well as distance between the user and the fire source predicted.
- Datasets can be used to create machine learning models, but large datasets (such as NASAs) require a lot of time to **not only download but also organize for proper machine learning.**
- This project has a tremendous impact in locations where wildfires are common, such as British Columbia in Canada, where **people can be informed of possible nearby fires, allowing them to prepare for the day** (i.e. using face masks, etc.)
- The app could also be used to automatically **inform fire departments** of possible fires to prepare them and help extinguish fires **more effectively and faster.**

Our Work -> https://github.com/EstebanF1612/NASA-SpaceAppsChallenge2025-TeamForeseers