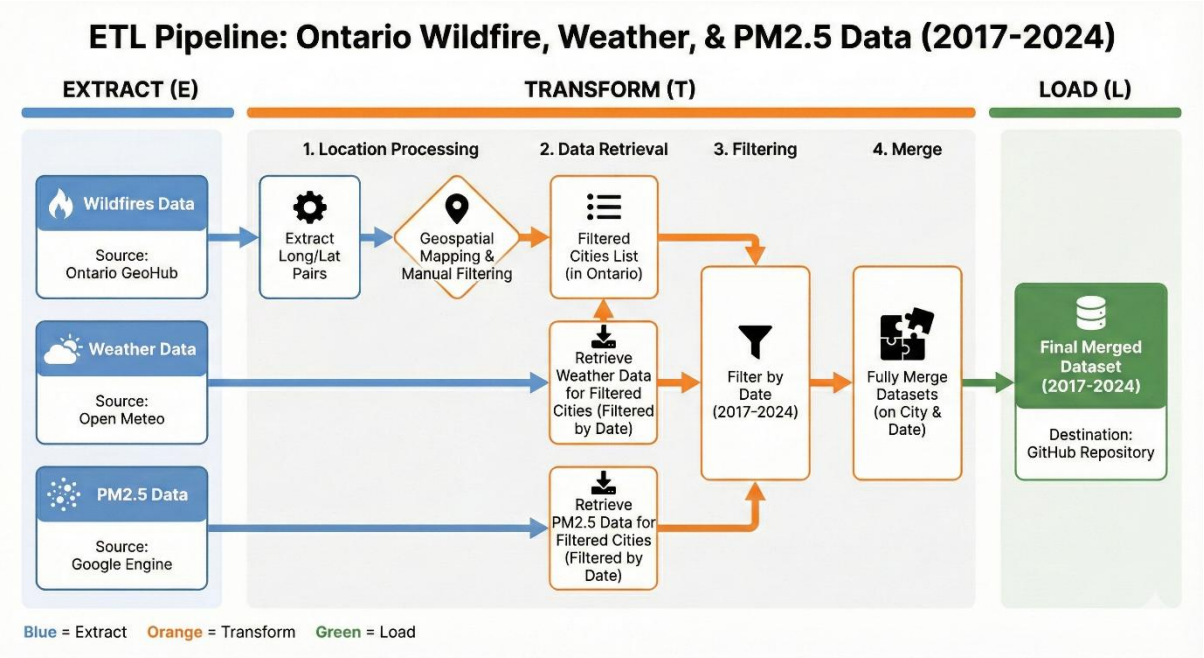
Full Integration Metadata Documentation:

Used Python Libraries: reverse_geocoder, geopandas, pandas



*** First Engineered Feature:

city: The name of the Ontario city associated with each wildfire record. Cities were assigned by taking the latitude and longitude of every fire, applying a reverse-geocoding step to retrieve the nearest populated place, then manually verifying and filtering out any locations that fall outside Ontario. This column is used to link wildfire events with their corresponding weather and PM2.5 measurements.

Needed Columns of the Datasets that were used:

1. Wildfires dataset: We got the information from: [Ontario GeoHub: Fire Disturbance Point - Data Description]

OGF_ID: NUMBER (13,0): A unique numeric provincial identifier assigned to each object

FIRE_TYPE_CODE: VARCHAR2 (3): The type of fire e.g. IFR - Inside Fire Region, OFR - Outside of Fire Region, PB - Prescribed Burn

FIRE TYPE CODE	FIRE TYPE DESCR	EXPIRY DATETIME
IFR	Inside Fire Region	
OFR	Outside Fire Region	
PB	Prescribed Burn	

FIRE_YEAR: NUMBER (4,0): The year the fire happened in.

FIRE_GENERAL_CAUSE_CODE: VARCHAR2 (3): The determined cause or origin of the forest fire

FIRE GENERAL CAUSE CODE	FIRE GENERAL CAUSE DESCR	EXPIRY DATETIME
IDF	INDUSTRIAL (forest industry)	
IDO	INDUSTRIAL (other)	
INC	INCENDIARY	
LTG	LIGHTNING	
MIS	MISCELLANEOUS	
REC	RECREATION	
RES	RESIDENT	
RWY	RAILWAY	
UNK	UNKNOWN	

FIRE_WEATHER_INDEX: NUMBER (4,1): No Description

FIRE_RESPONSE_OBJ_CODE: VARCHAR2 (3): No Description

FIRE RESPONSE OBJ CODE	FIRE RESPONSE OBJ DESCR	EXPIRY DATETIME
FUL	Full	
MDP	Modified Under Prescription	
MNP	Monitored Under Prescription	
MOD	Modified	
MON	Monitor	
NR	Not Recorded	
PRO	Protection	
SUP	Supression	

FIRE_START_DATE: DATE: The date that the Forest Fire was known or estimated to have started. The Aviation and Forest Fire Management Section (AFFM) uses the MM/DD/YYYY format to represent this date.

FIRE_OUT_DATE: DATE: The Date that the forest fire was declared to be out. The Aviation and Forest Fire Management Section (AFFM) uses the MM/DD/YYYY format to represent this date

FIRE_FINAL_SIZE: NUMBER (7,1): The field calculated final size of the area burned by the forest fire in hectares, as determined by the fire crewboss. For large fires, major green areas and large lakes are removed from the total area.

2. Weather dataset: We got the information from Open-Meteo: API Documentation

date: The day on which the weather variables (temperature, wind speed, humidity, etc.) were recorded. The weather data is provided at a daily resolution, so this column identifies the exact date each set of measurements applies to.

precipitation_sum (mm): Sum of daily precipitation (including rain, showers and snowfall)

rain_sum (mm): Sum of daily rain

snowfall_sum (cm): Sum of daily snowfall

relative_humidity_2m_mean (%): Average relative humidity at 2 meters above ground

relative_humidity_2m_max (%): Maximum relative humidity at 2 meters above ground

relative_humidity_2m_min (%): Minimum relative humidity at 2 meters above ground

dew_point_2m_mean (Â°C): Average dew point temperature at 2 meters above ground

shortwave_radiation_sum (MJ/mÂ²): The sum of solar radiation on a given day in Megajoules

sunshine_duration (s): The number of seconds of sunshine per day is determined by calculating direct normalized irradiance exceeding 120 W/m^2, following the WMO definition. Sunshine duration will consistently be less than daylight duration due to dawn and dusk.

temperature_2m_mean (Â°C): Average air temperature at 2 meters above ground

temperature_2m_max (Â°C): Maximum air temperature at 2 meters above ground

temperature_2m_min (Â°C): Minimum air temperature at 2 meters above ground

wind_speed_10m_mean (km/h): Average wind speed and gusts on a day at 10 meters above ground

wind_speed_10m_max (km/h): Maximum wind speed and gusts on a day at 10 meters above ground

winddirection_10m_dominant (Â°): Dominant wind direction at 10 meters above ground
vapour_pressure_deficit_max (kPa)

et0_fao_evapotranspiration (mm): Daily sum of ET_0 Reference Evapotranspiration of a well watered grass field. ET_0 Reference Evapotranspiration of a well watered grass field. Based on FAO-56 Penman-Monteith equations ET_0 is calculated from temperature, wind speed, humidity and solar radiation. Unlimited soil water is assumed. ET_0 is commonly used to estimate the required irrigation for plants.

3. PM2.5 air pollutant dataset:

city: The Ontario city associated with the PM2.5 measurement. Cities were determined by reverse-geocoding the wildfire coordinates and keeping only locations within Ontario.

date: The specific day for which the PM2.5 concentration was recorded. The dataset is daily, so each entry represents one day of air-quality data for a given city.

year: The year extracted from **date**.

lat: The latitude coordinate of the city's reference point used when retrieving PM2.5 data.

lon: The longitude coordinate of the city's reference point used when retrieving PM2.5 data.

pm25: The daily average concentration of fine particulate matter (PM2.5) for that city on the given date.

4. Engineered Features:

Will be filled when I get the data from the team