Appendix A: Data Dictionary

Attribute Name	Full Name	Description/Units	Source
Year	Year	Identifying Feature	National Interagency Fire Center*
Month	Month	Identifying Feature	National Interagency Fire Center
Lat	Latitude	Identifying Feature	National Interagency Fire Center
Long	Longitude	Identifying Feature	National Interagency Fire Center
Occurrences	Occurrences	Number of wildfire occurrences in a given month	National Interagency Fire Center
SurfTemp	Surface Temperature	Surface temp measured in Kelvin	NASA Giovanni**
NDVI	Normalized Differential Vegetation Index	Vegetation density measure, unitless	NASA Giovanni
SurfPS	Surface Pressure	Atmospheric pressure at the Earth surface in Pa (Pascals)	NASA Giovanni
Dust	Dust	Dust Scattering, Aerosol Optical Thickness at 550 nm (dimensionless)	NASA Giovanni
SoilMoisture	Soil Moisture	Depth averaged amount of water present 0-10 cm from the surface, measured in kg /m^2	NASA Giovanni
WindVelocity	Wind Velocity	Speed of air movement at the surface, in m/s	NASA Giovanni
LAI	Leaf Area Index	Quantifies the amount of leaf material in a canopy, unitless	NASA Giovanni

Attribute Name	Full Name	Description/Units	Source
Canopy_h2o	Plant Canopy Water	Plant canopy surface water in kg m^-2	NASA Giovanni
Transpire	Transpiration	Average transpiration in W m^-2	NASA Giovanni
brightness	Brightness Temperature 21	Radiance temperature, Kelvin	NASA FIRMS***
confidence	Confidence (0-100%)	Intended to help users gauge the quality of individual hotspot/fire pixels	NASA FIRMS
bright_t31	Brightness temperature 31	Channel 31 brightness temperature of the fire pixel measured in Kelvin.	NASA FIRMS
tempmax	Maximum Temperature	Measured in Celsius	Visual Crossing Weather****
tempmin	Minimum Temperature	Measured in Celsius	Visual Crossing Weather
temp	Temperature	Measured in Celsius	Visual Crossing Weather
feelslikemax	Feels like temperature, max	Measured in Celsius	Visual Crossing Weather
feelslikemin	Feels like temperature, min	Measured in Celsius	Visual Crossing Weather
feelslike	Feels like temperature	Measured in Celsius	Visual Crossing Weather
dew	Dew Point	The temperature at which air is saturated with water vapor, Celsius	Visual Crossing Weather
humidity	Relative Humidity	Measured in %	Visual Crossing Weather
precip	Precipitation	Amount of precipitation that fell, mm	Visual Crossing Weather
preciprob	Precipitation chance	Possibility of precipitation, %	Visual Crossing Weather
precipcover	Precipitation Cover	Proportion of time for	Visual Crossing

Attribute Name	Full Name	Description/Units	Source
		which measurable precipitation was record during the time period, %	Weather
snow	Snow	Snow is the amount of new snow that has fallen in the time period, cm	Visual Crossing Weather
snowdepth	Snow Depth	Average amount of snow currently on the ground for the time period, cm	Visual Crossing Weather
windgust	Wind Gust	kph	Visual Crossing Weather
windspeed	Wind Speed	kph	Visual Crossing Weather
winddir	Wind Direction	degrees	Visual Crossing Weather
sealevelpressure	Sea Level Pressure	Atmospheric pressure at a location that removes reduction in pressure due to the altitude of the location, measured in mb	Visual Crossing Weather
cloudcover	Cloud Cover	Amount of sky that is covered by clouds, expressed as a percentage.	Visual Crossing Weather
visibility	Visibility	km	Visual Crossing Weather
solarradiation	Solar Radiation	Measures power of sun, W m^-2	Visual Crossing Weather
solarenergy	Solar Energy	Total energy of sun, MJ/(m^2)	Visual Crossing Weather
uvindex	UV Index	A value between 0 and 10 indicating the level of ultraviolet (UV) exposure	Visual Crossing Weather
severerisk	Severe Risk	Severe risk for fires	Visual Crossing Weather

Attribute Name	Full Name	Description/Units	Source
moonphase	Moon phase	A decimal value representing the current moon phase between 0 and 1 where 0 represents the new moon, 0.5 represents the full moon	Visual Crossing Weather

* The National Interagency Fire Center (NIFC) is the nation's support center for wildland fires and other emergency situations.
** Giovanni is a NASA Goddard Earth Science Data and Information Services Center (GES DISC) Distributed Active Archive Center (DISC) web application that provides a simple, intuitive way to visualize, analyze, and access Earth science remote sensing data, particularly from satellites.
*** NASA's FIRMS distributes Near Real-Time (NRT) active fire data within 3 hours of satellite observation from the MODIS and VIIRS instruments.
**** Visual Crossing is a leading provider of weather data and enterprise analysis tools to data scientists, business analysts, professionals, and academics.

Appendix B: Variable Selection

Variables chosen for Denver

Lasso Regression	***F/B Selection***
NDVI	NDVI
SurfPS	SurfPS
Dust	Dust
SoilMoisture	SoilMoisture
LAI	LAI
Transpire	Canopy_h2o
confidence	Transpire
tempmax	tempmin
tempmin	humidity
feelslikemax	preciprob
dew	snow
humidity	snowdepth
precipcover	windgust
snow	windspeed
snowdepth	winddir
windspeed	sealevelpressure
sealevelpressure	visibility
solarradiation	solarradiation
solarenergy	uvindex
uvindex	moonphase
moonphase	dew

Variables chosen for Columbus

Lasso Regression	***F/B Selection***
Dust	SurfTemp
SoilMoisture	Dust
Transpire	SoilMoisture
humidity	WindVelocity
precipprob	Canopy_h2o
snowdepth	Transpire
sealevelpressure	confidence
cloudcover	feelslikemin
moonphase	precipprob
confidence	snowdepth
WindVelocity	solarradiation

Appendix C: External Links

The Blaze Abaters GitHub.

- https://github.com/Blaze-Abaters/ML-Models

Dashboards. These are interactive dashboards that provide additional visualizations for our data.

- The Blaze Abaters Data Dashboard: https://t.ly/w-a_
- The Blaze Abaters Fire Map: https://tamu.maps.arcgis.com/apps/dashboards/8f183504120b433ea87a87a0cc462e38

Appendix D: Communication with end-users

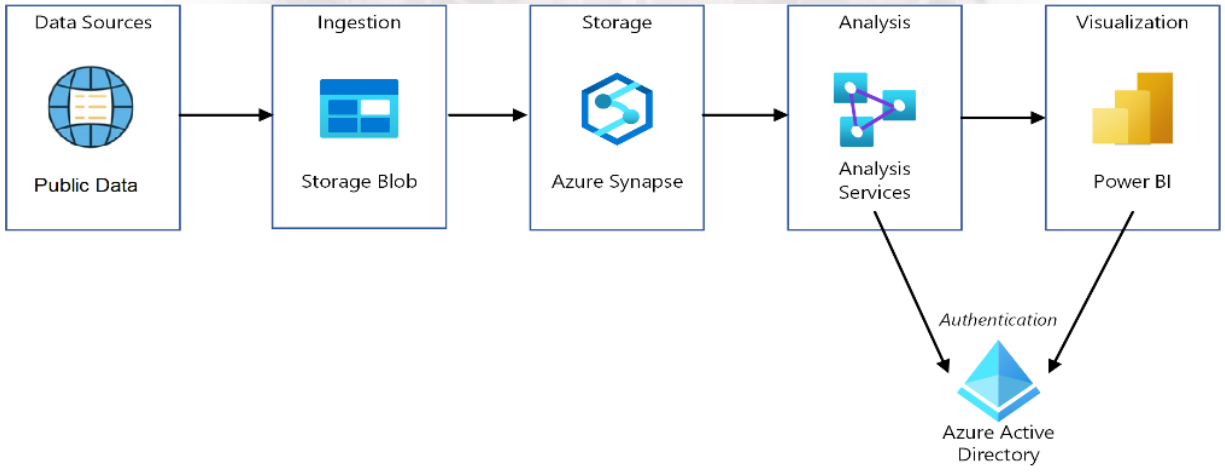
One of the most important problems when it comes to fighting fires is effectively communicating to end users the data necessary to make decisions and execute action plans. For this reason, we have created a data warehouse that collects the information in real time and organizes that information in a database for its later use in the prediction models and finally shows the findings in a dashboard accessible from any device with Internet connection.

In this document we will first explain the technology used to create the Data Warehouse, its architecture, and visualization of the findings in a dashboard and then we will explain the displayed values.

I. Data Warehouse

In order to store, transform and visualize the data, we have created a Data Warehouse (DW) using Microsoft Azure. This DW is built with a robust infrastructure that allows for easy scalability, with which it will be possible to work with hundreds of thousands of data, even millions, in an automated way (that is, using automated pipelines). That means that the dashboard updates by itself instantly and shows the information for each month that has already completed (e.g. at the time of writing this document, the last month shown is April 2023).

The DW architecture is as follows:



The DW has automated pipelines which extract the data from public repositories and files (raw data), then some cleaning methodologies are applied to obtain a new dataset which will be used by ML models and visualized through a Power BI dashboard. Also, data security is managed by Azure Active Directory, that assure only auth users access to the Main Dashboard.

II. Dashboard using Power BI

This dashboard is linked with an Azure data warehouse, so it is automatically capable of displaying insights from machine learning models.

The dashboard contains 2 pages:

- Dashboard and,
- Future Research.

In addition, each page contains a bookmark that, when is clicked, opens a website that shows an interactive heat map for each region (Denver and Columbus) with the number of occurrences for each one. Also, we consider that the dashboard and its charts are intuitive, using a standard color palette, with maroon (#700000) for Denver values and gray (#BCB9B9) for Columbus values.

1. Dashboard

This page contains 6 graphs and 2 cards that show an exploratory data analysis (EDA) of the main characteristics collected and used by the data models to generate predictions.

In addition to the graphs, there are 2 filters: one is to select the Columbus and Denver regions and other is to select the years. It is possible to select a single element, several or all at the same time. Whenever a single selection or combination of items is made, the data loaded in the charts and cards is filtered.

The 6 graphs displayed on the dashboard are the following:

a. **Occurrences by Month:** shows the average number of occurrences for each month, with a different line for each region.
b. **Brightness and Confidence by Month:** Displays the average brightness and confidence by month based on the selected region(s).
c. **Soil Moisture vs Surface Temp:** Shows the net values of each variable for each region.
d. **Average Temperature °F:** Shows the averages of minimum, maximum and temperature values, in °F, of the selected regions and years.
e. **Average Feels Like °F:** Shows the averages of minimum feels like, maximum feels like and feels like, in °F, for the selected regions and years.
f. **Humidity by Region:** This graph has a detail: in addition to showing the % humidity for each region, when the mouse is hover over the graph, we can see the % relative humidity, as well as the averages of Dew, Dust and Wind Velocity for the selected region.

On the other hand, the 2 graphs show the % of Cloud Cover and Visibility.

2. Future Research

Page 2 contains 5 possible lines of future research, as well as 2 scatter graphs considering the data of the Moonpahse vs Brightness and Humidity vs Transpire variables, respectively, for each region.