

Team hermosillobazan

Basilio Bazan III

J. Adolfo Hermosillo

The Question to Answer

How do storm events and wildfires in the United States compare when considering the time of the year, location differences, and damage caused to agriculture?

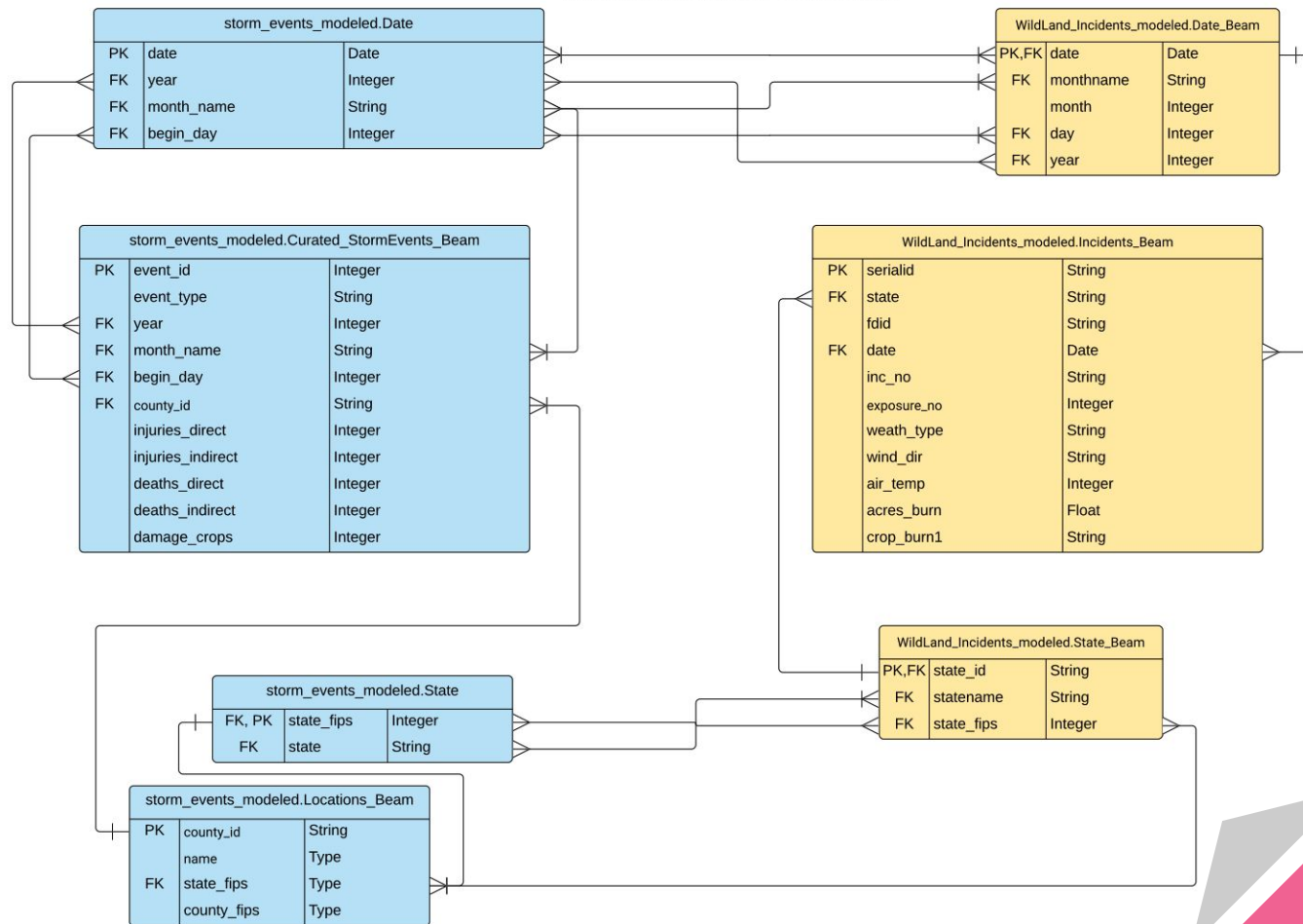
- Time by months
- Location by states
- Damage by reported instances of crop damage



Our Datasets

- Storm Events by the National Oceanic Atmospheric Administration (NOAA)
 - <https://public.enigma.com/browse/collection/national-climatic-weather-center-storm-events/3d0d8bdf-b885-48b9-a3b1-8a8441fc0131>
- Wild Land Incidents provided by the National Fire Incident Reporting System of the U.S Government's Federal Emergency Management Agency.
 - <https://public.enigma.com/browse/collection/nfirs-wild-land-incidents/952d1599-65cc-4374-afa8-b07f6cdb2f98>





Modeled Datasets

Beam Pipelines

Storm Events

- Created primary key for Locations table.
- Capitalized location names to keep consistency.
- Removed duplicate records.

Wild Land Incidents

- Created primary key for date and split into month, day, year.
- Matched Storm Events State Format by adding Full state names to State table.



Airflow DAG

- Help to track down errors happening on which task
- Split modeled table construction including beam pipelines.

```
date_beam = BashOperator(  
    task_id='date_beam',  
    bash_command='python /home/jupyter/airflow/dags/transform_Date_single.py')  
  
date_dataflow = BashOperator(  
    task_id='date_dataflow',  
    bash_command='python /home/jupyter/airflow/dags/transform_Date_cluster.py')  
  
state_beam = BashOperator(  
    task_id='state_beam',  
    bash_command='python /home/jupyter/airflow/dags/transform_State_single.py')  
  
state_dataflow = BashOperator(  
    task_id='state_dataflow',  
    bash_command='python /home/jupyter/airflow/dags/transform_State_cluster.py')  
  
incidents_beam = BashOperator(  
    task_id='incidents_beam',  
    bash_command='python /home/jupyter/airflow/dags/transform_Incidents_single.py')  
  
incidents_dataflow = BashOperator(  
    task_id='incidents_dataflow',  
    bash_command='python /home/jupyter/airflow/dags/transform_Incidents_cluster.py')  
  
create_staging_dataset >> create_modeled_dataset >> load_2010 >> load_2011 >> split  
split >> create_date >> date_beam >> date_dataflow  
split >> create_state >> state_beam >> state_dataflow  
split >> create_incidents >> incidents_beam >> incidents_dataflow
```

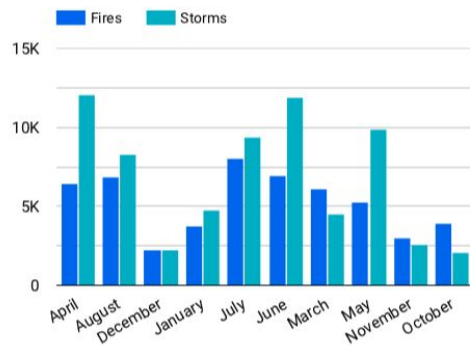
Cross-Dataset Queries

1. Storms and Fire events per month
2. Crops damaged by Storms and Fire events
3. Storms and Fire events per state

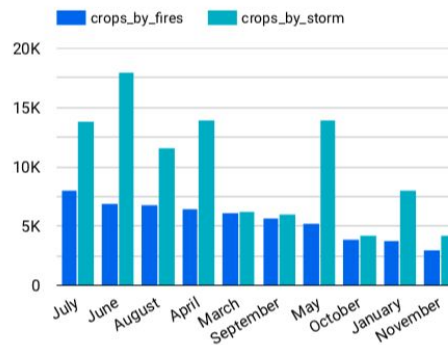
[LIVE DEMO]



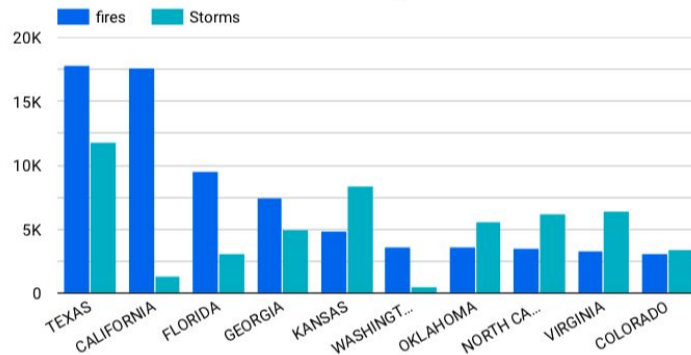
Fires and storms per month



Crops destroyed by fires and storms



Fires and storms per state



Future Improvements

1. Better matching datasets
2. More years
3. More countries
4. Human casualties on fires

