Shootout 2017 Data Dictionary

Understanding Wildfires

This document lists the datasets provided to accompany the 2017 SAS Analytics Shootout Competition. For each dataset provided in the Shootout package, a description of the file structure and variables are listed here.

Provided Data:

1. wildfire_events.sas7bdat
2. drought_Severity.sas7bdat
3. weather_by_stations.sas7bdat
4. storm_events.sas7bdat
5. wildfire_narratives.sas7bdat
6. Land_Cover_by_County.csv
7. Population_by_County_2001_to_2009.csv
8. Population_by_County_2010_to_2015.csv
9. NOAA_Zones_to_Counties.csv
10. US_weather_stations_counties.csv
11. drought_severity_scoring.sas7bdat
12. weather_station_scoring.sas7bdat
13. storm_events_scoring.sas7bdat
14. US_Counties.sqmiles.csv

Descriptions

1. **Wildfire_events.sas7bdat:** Wildfire events in the US, by county/zone, from 2001 to 2015. Contains 16 columns and 4,956 rows
   - **EPISODE_ID:** ID for the wildfire episode. Episodes may span across multiple counties/zones
   - **EVENT_ID:** ID for the wildfire event. Events ID's are unique, meaning there is a single Event ID per wildfire location in county/zone
   - **STATE:** State of the wildfire event
   - **MONTH_NAME:** name of the month in which the wildfire occurred
   - **EVENT_TYPE:** specifies the event type, which in this dataset are all "Wildfire"
   - **CZ_TYPE:** flag to specify whether the location is in a zone (Z) or county (C)
   - **CZ_FIPS:** ID for the county/zone. The same CZ_FIPS may be used for county/zones in different states. The combination of State AND CZ_FIPS will provide a unique ID across the entire dataset
   - **CZ_NAME:** name of the county/zone

- **BEGIN_DATE_TIME:** beginning date and time of the wildfire event
- **END_DATE_TIME:** ending date and time of the wildfire event
- **INJURIES_DIRECT:** injuries that were a direct result of the wildfire
- **INJURIES_INDIRECT**: injuries that were an indirect result of the wildfire. An example of this may be an emergency response vehicle involved in an accident while on the way to the wildfire location
- **DEATHS_DIRECT:** deaths that were a direct result of the wildfire
- **DEATHS_INDIRECT:** injuries that were an indirect result of the wildfire
- **DAMAGE_PROPERTY:** approximate damage to property in USD
- **DAMAGE_CROPS:** approximate damage to crops in USD

2. **Drought_Severity.sas7bdat:** Weekly reports for droughts in the US by different severity levels, by county, for 2001 to 2015. Contains 12 columns and 2,520,477 rows.
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **county:** name of the county
    - **state:** name of the state
    - **NONE:** percentage of county with no drought conditions
    - **D0_D4:** percentage of county in D1 "Abnormally Dry drought condition"
    - **D1_D4:** percentage of county in D2 "Moderate drought condition"
    - **D2_D4:** percentage of county in D3 "Severe drought condition"
    - **D3_D4:** percentage of county in D4 "Extreme drought condition"
    - **D4:** percentage of county in D5 "Exceptional drought condition"
    - **validStart:** start date of drought week
    - **validEnd:** end date of drought week

3. **weather_by_stations.sas7bdat:** Selected weather variables reported by date, by weather station, for 2001 to 2015. Contains 11 columns and 13,122,253 rows.
    - **STATION_ID:** ID of the weather station
    - **Date:** date of the weather data
    - **State:** state of the weather station
    - **County:** county of the weather station
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **STATION_Elevation:** elevation in meters of the weather station
    - **TMAX_F:** maximum temperature in degrees Fahrenheit
    - **TMIN_F:** minimum temperature in degrees Fahrenheit
    - **PRCP_in:** precipitation in inches
    - **SNWD_in:** snow depth in inches
    - **SNOW_in:** snowfall in inches

4. **storm_events.sas7bdat:** storm events in the US, by county/zone, from 2001 to 2015. Contains 15 columns and 881,599 rows.
    - **EPISODE_ID:** ID for the storm episode. Episodes may span across multiple counties/zones
    - **EVENT_ID:** ID for the storm event. Events ID's are unique, meaning there is a single Event ID per storm location in county/zone
    - **STATE:** state of the storm event
    - **EVENT_TYPE:** specifies the storm type
    - **CZ_TYPE:** flag to specify whether the location is in a zone (Z) or county (C)
    - **CZ_FIPS:** ID for the county/zone
    - **CZ_NAME:** name of the county/zone
    - **BEGIN_DATE_TIME:** beginning date and time of the storm event
    - **END_DATE_TIME:** ending date and time of the storm event
    - **INJURIES_DIRECT:** injuries that were a direct result of the storm
    - **INJURIES_INDIRECT**: injuries that were an indirect result of the storm
    - **DEATHS_DIRECT:** deaths that were a direct result of the storm
    - **DEATHS_INDIRECT:** injuries that were an indirect result of the storm
    - **DAMAGE_PROPERTY:** approximate damage to property in USD
    - **DAMAGE_CROPS:** approximate damage to crops in USD

5. **wildfire_narratives.sas7bdat:** text descriptions of wildfire episodes/events in the US from 2001 to 2015. Contains 4 columns and 4,956 rows.
    - **EPISODE_ID:** ID for the wildfire episode. Episodes may span across multiple counties/zones
    - **EPISODE_NARRATIVE:** text description of the wildfire episode
    - **EVENT_ID:** ID for the wildfire event. Events ID's are unique, meaning there is a single Event ID per wildfire location in county/zone
    - **EVENT_NARRATIVE:** text description of the event episode

6. **Land_Cover_by_County.csv:** Percentage of land cover types in the US by county. Contains 20 columns and 3,109 rows.
    - **STATE_NAME:** state of county
    - **STATE_ID:** ID for state
    - **ST_CO_ID:** ID for state and county
    - **County:** county name
    - **11:** percentage of county with Open Water- areas of open water, generally with less than 25% cover of vegetation or soil.
    - **12:** percentage of Perennial Ice/Snow- areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover

- **21:** percentage of Developed, Open Space- areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes
- **22:** percentage of county with Developed, Low Intensity- areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units
- **23:** percentage of county with Developed, Medium Intensity -areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units
- **24:** percentage of county with Developed High Intensity-highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover
- **31:** percentage of county with Barren Land (Rock/Sand/Clay) - areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover
- **41:** percentage of county with Deciduous Forest- areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change
- **42:** percentage of county with Evergreen Forest- areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green foliage
- **43:** percentage of county with Mixed Forest- areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover
- **52:** percentage of county with Shrub/Scrub- areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions
- **71:** percentage of county with Grassland/Herbaceous- areas dominated by gramanoid or herbaceous vegetation, generally greater than 80% of total

vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing

- **81:** percentage of county with Pasture/Hay-areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation
- **82:** percentage of county with Cultivated Crops -areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled
- **90:** percentage of county with Woody Wetlands- areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water
- **95:** percentage of county with Emergent Herbaceous Wetlands- Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water

7. **Population_by_County_2001_to_2009.csv:** Population in the US, by county, by year (taken in July), for 2001 to 2009. Contains 11 columns and 3,143 rows.
    - **State:** state of the county
    - **County_name:** name of the county
    - **2001:** population estimate in 2001
    - **2002:** population estimate in 2002
    - **2003:** population estimate in 2003
    - **2004:** population estimate in 2004
    - **2005:** population estimate in 2005
    - **2006:** population estimate in 2006
    - **2007:** population estimate in 2007
    - **2008:** population estimate in 2008
    - **2009:** population estimate in 2009

8. **Population_by_County_2010_to_2015.csv:** Population in the US, by county, by year (taken in July), for 2010 to 2015. Contains 9 columns and 3,220 rows.
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **State:** state of the county
    - **County_name:** name of the county **2010:** population estimate in 2010
    - **2011:** population estimate in 2011

- **2012:** population estimate in 2012
- **2013:** population estimate in 2013
- **2014:** population estimate in 2014
- **2015:** population estimate in 2015

9. **NOAA_Zones_to_Counties:** crosswalk table between zones (Z) and counties (C) in the US. Contains 11 columns and 3,681 rows.
    - **ST:** abbreviated state name
    - **CZ_FIPS:** ID for the county/zone. The same CZ_FIPS may be used for county/zones in different states. The combination of State AND CZ_FIPS will provide a unique ID across the entire dataset
    - **Zone_Name**: text name of the zone
    - **State_Zone**: combination of state abbreviation and CZ_FIPS
    - **Zone_Long**: longitude of the geographic center of zone
    - **Zone_Lat**: latitude of the geographic center of zone
    - **State_Name:** full text for state name
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **County:** name of the county
    - **CO_Lat:** latitude of the geographic center of county
    - **CO_Long:** longitude of the geographic center of county
    -
10. **US_weather_stations_counties.csv:** locations of US weather stations. Contains 7 columns and 54,211 rows.
    - **STATION_ID:** ID of the weather station
    - **STATION_LAT:** latitude of the weather station
    - **STATION_LONG:** longitude of the weather station
    - **STATE_NAME:** full name of state the weather station is in
    - **County:** name of the county the weather station is in
    - **State:** abbreviated name of the state
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)

11. **drought_severity_scoring.sas7bdat:** Population in the Weekly reports for droughts in the US by different severity levels, by county, for the scenario year. Contains 13 columns and 167,388 rows.
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **county:** name of the county
    - **state:** name of the state
    - **NONE:** percentage of county with no drought conditions

- **D0_D4:** percentage of county in D1 “Abnormally Dry drought condition”
- **D1_D4:** percentage of county in D2 “Moderate drought condition”
- **D2_D4:** percentage of county in D3 “Severe drought condition”
- **D3_D4:** percentage of county in D4 “Extreme drought condition”
- **D4:** percentage of county in D5 “Exceptional drought condition”
- **StartDay:** start of drought by day
- **StartMonth:** start of drought by month
- **End Day:** end of drought by day
- **End Month:** end of drought by month

12. **weather_station_scoring.sas7bdat:** Selected weather variables reported by date, by weather station, for the scenario year. Contains 12 columns and 963,657 rows.
    - **STATION_ID:** ID of the weather station
    - **DAY:** day of month of the weather data
    - **Month:** month of the weather data
    - **State:** state of the weather station
    - **County:** county of the weather station
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID’s are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **STATION_Elevation:** elevation in meters of the weather station
    - **TMAX_F:** maximum temperature in degrees Fahrenheit
    - **TMIN_F:** minimum temperature in degrees Fahrenheit
    - **PRCP_in:** precipitation in inches
    - **SNWD_in:** snow depth in inches
    - **SNOW_in:** snowfall in inches

13. **storm_events_scoring.sas7bdat:** storm events in the US, by county/zone, for the scenario year. Contains 18 columns and 51,635 rows.
    - **EPISODE_ID:** ID for the storm episode. Episodes may span across multiple counties/zones
    - **EVENT_ID:** ID for the storm event. Events ID’s are unique, meaning there is a single Event ID per storm location in county/zone
    - **Begin Day:** beginning day of month of the storm event
    - **Begin Time:** beginning time of the storm event
    - **End Day:** ending day of month of the storm event
    - **End Time:** ending time of month of the storm event
    - **Month Name:** month of the storm event
    - **STATE:** state of the storm event
    - **EVENT_TYPE:** specifies the storm type
    - **CZ_TYPE:** flag to specify whether the location is in a zone (Z) or county (C)

- **CZ_FIPS:** ID for the county/zone
- **CZ_NAME:** name of the county/zone
- **INJURIES_DIRECT:** injuries that were a direct result of the storm
- **INJURIES_INDIRECT**: injuries that were an indirect result of the storm
- **DEATHS_DIRECT:** deaths that were a direct result of the storm
- **DEATHS_INDIRECT:** injuries that were an indirect result of the storm
- **DAMAGE_PROPERTY:** approximate damage to property in USD
- **DAMAGE_CROPS:** approximate damage to crops in USD

14. **US_Counties.sqmiles.csv:** County names, ID's, and area in Square Miles. Contains 6 columns and 3,221 rows.
    - **County_Name:** name of the county
    - **State_Name:** name of the state
    - **State_FIPS:** id of the State
    - **County_FIPS:** id of the County
    - **ST_CO_ID:** ID for the state and county. ST_CO_ID's are 5 digits; the first two represent the state FIPS (ID) and the last three represent the county FIPS (ID)
    - **SQMI:** area of the county in square miles