Capstone Project 1 - School violence in the United States
Data Wrangling

1. Dropped information:
    a. Columns: age_shooter2, gender_shooter2, race_ethnicity_shooter2, shooter_relationship2, shooter_deceased2, deceased_notes2, deceased_notes1, weapon_source
        i. The columns about a second shooter had only 2 entries, leaving 105 empty values in each.
        ii. Didn't feel that deceased_notes1 (how the shooter died) would add much value to later analysis
        iii. Weapon_source (where the student got the weapon) was mostly missing and I don't believe it will add a lot to later analysis
    b. Row for the school shooting at Vereen School on November 5, 2015
        i. This row was responsible for the only missing value in many columns (school demographics, shooter information, and time).
2. Missing values:
    a. 22 missing values for the age of the shooter
        i. Replaced with the median age of shooters
        ii. Chose median because there is a wide range of ages (from 6 to 53)
    b. Various numbers of missing values in columns state, school_type, shooting_type, gender_shooter1, race_ethnicity_shooter1, ulocale, day_of_week, and city
        i. Replaced all missing entries with string 'unknown' in preparation for converting to categorical variables.
    c. 2 missing times of the incidents
        i. For Stellar Leadership Academy, I replaced the missing value with 1:30 PM. I was able to find this rough approximation in a local news article written about the event.
        ii. For Redland Middle School, I replaced the missing value with 11:00 AM. There was no time reported in any of the news articles, but they all mentioned that it happened during a student's class. So I chose the middle of the typical school day as the time.
    d. Various number of missing values for racial/ethnic demographics of the school
        i. Replaced missing counts for Rebound High School and Success Academy based on online reports.
        ii. Assuming all remaining missing ethnic counts (hawaiian native and two or more ethnicities) are actually 0
    e. Missing values in staffing and lunch
        i. Replaced all missing values in staffing and lunch with 999 and 9999 respectively
            1. These represent unknown values while maintaining the columns as numeric dtypes

3. Correcting dtype of DataFrame columns
    a. Categorical
        i. Converted state, school_type, shooting_type, gender_shooter1, race_ethnicity_shooter1, shooter_relationship1, weapon, ulocale, day_of_week, and city
            1. shooter_relationship1 - There were 20 unique responses in this column that I re-grouped into student, former student, student relation, not a student, and unknown
            2. weapon - There were 20+ unique responses in this column as well so I regrouped them into generic type of gun (revolver, handgun (semi-automatic), shotgun, rifle, unknown
            3. shooting_type - Consolidated responses into fewer categories (targeted, accidental, indiscriminate, suicide, and unclear)
    b. Datetime
        i. Combined date and time columns into a new column labeled date_time. Then converted date_time column
    c. Numeric
        i. Replaced kindergarten(KG) and pre-kindergarten(PK) in low_grade column to numeric values 0.5 and 0 respectively.
        ii. Then converted both low_grade and high_grade