1. What were the main objectives of your analysis?
2. Which specific insights did you uncover from the dataset?
3. What data visualization techniques did you use?
4. How did you preprocess and clean the dataset?
5. Did you encounter any challenges during the analysis?
6. What are the key findings or conclusions from your analysis?
7. How can the results of this analysis be interpreted or utilized?
8. The main objective of my analysis was to develop visual charts based off the data points found in the dataset that recorded all U.S. military fatalities and their categorical breakdown. I also created a PowerPoint slide to present my graphs and include commentary. Instead of a specific area of focus, I wanted to show a generalized observation of trends that may or may not be interrelated with the others. Some trends were directly related to specific conflicts, economic activity, military force restructuring, military program implementation, medical treatment and so on.
   1. I generated a line plot from 1980-2010 showing the year-by-year changes in total military strength. For example, in the 1980s, total military strength was an estimated 2.4 million and suffered a steep decline in the 1990s to its lowest level in 1999 which reached an estimated 1.5 million and then rose slightly in the following decade due to the war on terror.
   2. I created 3 doughnut charts to illustrate different snapshots of force distribution (Active Duty, Selected Reserve, Full Time Guard and Reserve. I used 1980, 1990 and 2010 as reference. I did observe that between 1980 and 2010, the percent of the active duty force shrunk from 95% in 1980 to 84.9% in 2010. Inversely, there was a several percentage point increase for Guard and Reserve.
   3. When I summarized all of the U.S. Military Fatalities, and created a doughnut chart to represent their portion, what caught my eye was that fatalities caused by accidents represented 52.10% of all fatalities. This was followed by illnesses at 17.83%, Self-Inflicted at 14.36%, Hostile Action at 10%, homicide at 4.84% and terrorist attacks at 0.87%.
   4. I created line charts to represent each of these categories using a ratio of per 100,000 to ensure accuracy due to fluctuating total force numbers. While it was observed that accidents have represented over half of all fatalities, there has been a significant decrease in the recorded time period. In 1980, accident deaths were just over 70 per 100,000, sharply declining in the following 15 years and bottoming out at around 30 accident deaths per 100,000 before climbing again to 40 per 100,000 before declining again to 20 per 100,000 in 2010.
   5. The illness line chart, using the same ratio, showed a sharp decrease from 1981 to 1999, going from 22 to 11, then rising again to 18 in 2006-2010.
   6. The self-inflicted chart has shown 4 major spikes, 1983 at 11 per 100,000, 1985-1988 at 13, 1996 at 15 and then heavily spiking from 2004-2010 topping out at 18 per 100,000.
   7. The hostile action chart showed two major events, the 1983 Beirut Barracks Bombing and the War on Terror.
   8. The Homicide chart shows 9 year decrease from 8 per 100,000 before sharping rising again peaking at 5.5 around 1992, then sharply declining to less than 2 by 1998. From 1998 to 2010, there has been a slight increase with a sharp spike in 2008-2009 going from 3 to 5 before sharply dropping again to 2.5.
   9. The terrorist action chart shows several spikes. The most significant spike was in 1983 with the Beirut U.S. Embassy Bombing, the spike peaked at almost 12 per 100,000 before dropping back to near zero. Three more smaller spikes occurred most likely due to the Oklahoma City Bombing, the USS Cole Bombing in 2000 and the September 11th attacks.
   10. I wanted to also explore if there are any decrease on a decade-by-decade basis. I created a chart that looked at all accidents per 100,000. I used 1980-1989, 1990-1999 and 2000-2010. For each time period I calculated if there was a decrease. Using the information from the previous line chart, we know that there was a decrease in accident fatalities, but I wanted to know how much and by what percent. My results showed that from 1980 to 1989, there was over a 4% decrease, from 1990-1999 there was a 2.4% decrease and from 2000-2010, there was a 1% decrease.
   11. I used the same type of formula for self-inflicted deaths. There was a small 2-4% decrease from 1980 to 1989, from 1990-1994, it increased by 5% before sharply decreasing from 1995-1999 by nearly 6%. After the year 2000, increases started to occur again, 1% initially from 2000-2004 and then by over 5% from 2005-2010.
   12. The total deaths chart, using the same calculation, did not show much noteworthy data as there were small single percentage decreases year on year until the 2000-2004 time period where deaths spiked unsurprisingly due to the war on terror.
9. I used various data visualization techniques in my analysis. Due to the prevalence of very large and very small numbers, I had to be creative on how to present the data that made sense. For the total military strength, because of the 40 year time period, I shaped the graph very long on the x axis and short on the 4 axis. I also limited the y axis ticks to only show the numbers in the range. For example, total military strength fluctuated from a peak of 2.4 million to a low of 1.5 million. The chart would look silly if I had the y axis start at 0 so instead I started it just under and over the max and min ranges. To show a visually aesthetic and informative breakdown of all major categories of fatalities, I created a ring chart where the categories are shown slightly separated, color coded, labeled and with their percentages identified. For the 6 line graphs I created to show data trends over time, I had to be creative as to ensure my line charts represented the number accurately. To achieve this, I ratioed the death category number against the calendar year and total military strength for that year. To show more exaggeration in the graph, I compressed it to be narrow, placing the focus less on the actual numbers but the direction of the line itself to illustrate the data. My final set of charts were bar charts, but this time I wanted to show a positive or negative change over a time period, between 5 and 10 years. For example, the bar chart showing the average annual decrease of accidents per 100,000 soldiers illustrated, between the 3 10 year time periods, 1980-1989, 1990-1999 and 2000-2010, showed a percentage drop of accidents year over year. From 1980-1989, accident deaths, as a percentage, dropped by over 4 percent. From 1990-1999, they dropped over 2% and from 2000-2010, they dropped by slightly over 1%. The self inflicted bar chart, using the same calculation, this time over 5 year periods from 1980 – 2010, showed minimal percentage decreases in the 1980s, a 5% rise in the early 1990s, a sharp drop of over 6% in the late 1990s and then an upward and final spike of 5 percent in the late 2000s.
10. The dataset I obtained was already structured fairly well, there were no missing numbers, the column names were very straightforward. The only formatting issue I had as I was creating my graphs from the data were that there was some unseen leading or trailing whitespaces from which I had to strip using the .strip method. I created a clean.py file and imported all of my modules. I then defined a function called file\_import and set the parameter to the filepath of the dataset is formatted as a .csv file. In this function, I created a variable called df and set its value to the filepath. Additionally, due to prior issues with unseen whitespaces, modified the columns with a df.columns.str.strip() method to prevent this from occurring as I called the function. The final part of clean.py was to create the if \_\_name\_\_ == ‘\_\_main\_\_’: condition which would allow me to isolate my functions and run them to ensure they work.
11. My biggest challenge to this data analysis was determine what story do I want to tell. Based off the straightforward data already present in the dataset, I slowly evolved and tightened my focus on showing a general correlation and data trends over time and provide commentary about the visual discoveries. Another challenge was trying to minimize wasted space in my graphs, ensuring that the x and y axis only show relevant data. If my line graph data stays above 1 million, then would modify the y axis limits to fit that scope. Another challenge was trying to show an accurate representation of the data through a ratioed calculation. By narrowing my ratio to 1 to 100,000 and calculating for every calendar year as the numbers fluctuated, I was able to finally get the visual representation I wanted.
12. The most significant key finding the rate of fatal accidents over from 1980 to 2010. For 15 years straight, starting in 1980, there was a significant decline, going from over 70 per 100,000 to a minimum of just under 30 per 100,000 before moderately rising again. It can by hypothesized that this drop can be attributed to an increase and focus of safety programs, enhanced military training, protective equipment, DoD and command policy and its implementation strategy, and cultural shifts.
13. The results of this analysis can be used as historical insight of the U.S. military population and culture, its strength, its challenges, how changes its operations, policy and initiatives have affected it, how changes in leadership, military training, military doctrine, adoption of new technologies, updated equipment, demographics changes and culture trends have had or not had any impact on the statistical data.