README for Python Data Science Midterm Project

Overview

This project analyzes a dataset obtained from www.kaggle.com, which contains a historical

numerical record of all U.S. military deaths from 1980 to 2010. The objective of the analysis

was to develop visual charts based on the data points and provide insights into trends and patterns observed in the dataset.

Project Details

Dataset: The dataset contains information on U.S. military fatalities, including columns

such as Calendar Year, Total Military FTE, Total Deaths, Accident, Hostile Action, Homicide,

Illness, Self-Inflicted, and Terrorist Attack.

Questionnaire and Analysis

What was the main objective of your analysis?

The main objective of the analysis was to develop visual charts based on the data points and provide

a generalized observation of trends in U.S. military fatalities. The analysis aimed to explore trends

related to specific conflicts, economic activity, military force restructuring, program implementation,

medical treatment, and other factors.

What visualizations did you create, and what did you observe from them?

a. Line plot showing year-by-year changes in total military strength.

b. Three doughnut charts illustrating force distribution in 1980, 1990, and 2010.

c. Doughnut chart representing the categorical breakdown of U.S. military fatalities.

d. Line charts representing trends in different categories of fatalities per 100,000 soldiers.

e. Bar charts showing the average annual decrease in fatalities per 100,000 soldiers over 10-year periods.

How did you use data visualization techniques in your analysis?

Various data visualization techniques were employed:

Line plots to show year-by-year changes.

Doughnut charts to illustrate force distribution and categorical breakdown.

Line charts to represent trends over time.

Bar charts to show changes over time periods.

Visualization choices were made to ensure accurate representation of

the data and to highlight trends effectively.

How did you handle data formatting or cleaning?

The dataset was already well-structured, with no missing values.

Column names were straightforward but had leading or trailing whitespaces,

which were removed using the .strip() method.

What challenges did you face during the analysis?

Determining the story to tell based on the available data.

Minimizing wasted space in the graphs and ensuring relevant data on the axes.

Calculating accurate representations of data through ratioed calculations.

What was the most significant finding of your analysis?

The rate of fatal accidents showed a significant decline from 1980 to 2010, dropping from over 70 per 100,000 to

a minimum of just under 30 per 100,000. This could be attributed to safety programs, enhanced military training,

protective equipment, policy implementation, and cultural shifts.

How can the results of your analysis be used?

The results provide historical insights into the U.S. military population, culture,

and the impact of various factors on the statistical data. They can be used to understand

changes in military operations, policy initiatives, training, technology adoption, and demographic shifts over time.

Running the Code

To run the code for this analysis:

Make sure you have Python and the required dependencies installed.

Download the dataset from www.kaggle.com and place it in the same directory as the code.

Run the Python script to execute the analysis and generate the visualizations.

Additional Notes

The code provided is for reference purposes only and may require adjustments based on your specific environment and dataset.

Feel free to modify the code or adapt it to suit your requirements.

Make sure to provide appropriate credits and references for the dataset used in your analysis.

Enjoy analyzing the U.S. military fatalities dataset!