A Problem with Presidents: Converting Data to Information

The objective of this project is to demonstrate the process of converting data into information using Python and mathematics. I analyzed the data related to the lifespan of US Presidents and perform basic exploratory data analysis with the dataset given and broken down into the three figures. Each of these figures shows what I have found in terms of solving this problem.

Figure 1 shows that the top 10 presidents from longest lived to shortest lived and the top 10 presidents from shortest lived to longest lived. I’ve found that President George Bush lived the longest, with a lifespan of 94 years and 84 days. While President John F. Kennedy had the shortest lifespan, with 46 years and 183 days. Both presidents are statistically recorded as Maximum Value and Minimum Value respectively in the next figure.

Figure 2 shows the statistical table of the average lifespan of US Presidents. The Mean is calculated from adding all the lived\_days variable and divide it by the total number of presidents. The Weighted Average is similar to the mean but add significance for every year the Presidents have lived. The Standard Deviation for this particular dataset is 4596 (around 12 years). It is also important to note that there are living presidents (such as Jimmy Carter) are not part of this data.

Figure 3 shows a histogram of how many frequently range of values for every president have lived. It revealed that the distribution of the data peaked in both 21000 and 25000 days where presidents lived in those two values are the most frequent. The right end of the Histogram is rising in frequency as an average lifespan of healthy Presidents that last longer while the left end of the Histogram listing Presidents who are mostly assassinated.

Concluding this report using Python to parse the given dataset of US presidents, my analysis of the US Presidents' lifespan data revealed insights that can be used to gain a better understanding of this historical aspect of the US Presidency. The use of descriptive statistics and visualization techniques allowed us to communicate the findings in a way that is easily understandable to a broad audience.

# References

[1] Woolf, M. (n.d.). Statistics and Probability Background. LibreTexts. <https://eng.libretexts.org/Bookshelves/Industrial_and_Systems_Engineering/Book%3A_Chemical_Process_Dynamics_and_Controls_(Woolf)/13%3A_Statistics_and_Probability_Background/13.01%3A_Basic_statistics-_mean%2C_median%2C_average%2C_standard_deviation%2C_z-scores%2C_and_p-value>.

Graphical user interface

Description automatically generated with medium confidence

Figure 1 – The Top 10 Lists

Text

Description automatically generated

Figure 2 – The Statistics

Chart, histogram

Description automatically generated

Figure 3 – The Histogram