**Analysis Report of U.S. Presidents' Lived Years and Days.**

**Author**: Your Name **Date**: Current Date

Table 1: Top 10 Longest-Lived Presidents

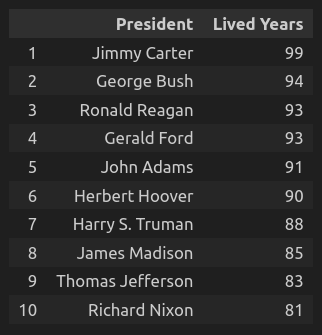


Table 2: Top 10 Shortest-Lived Presidents

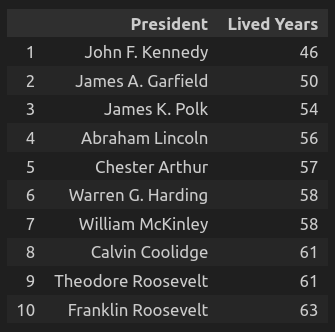
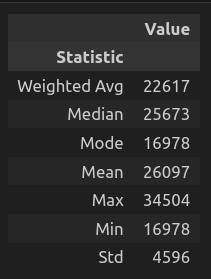
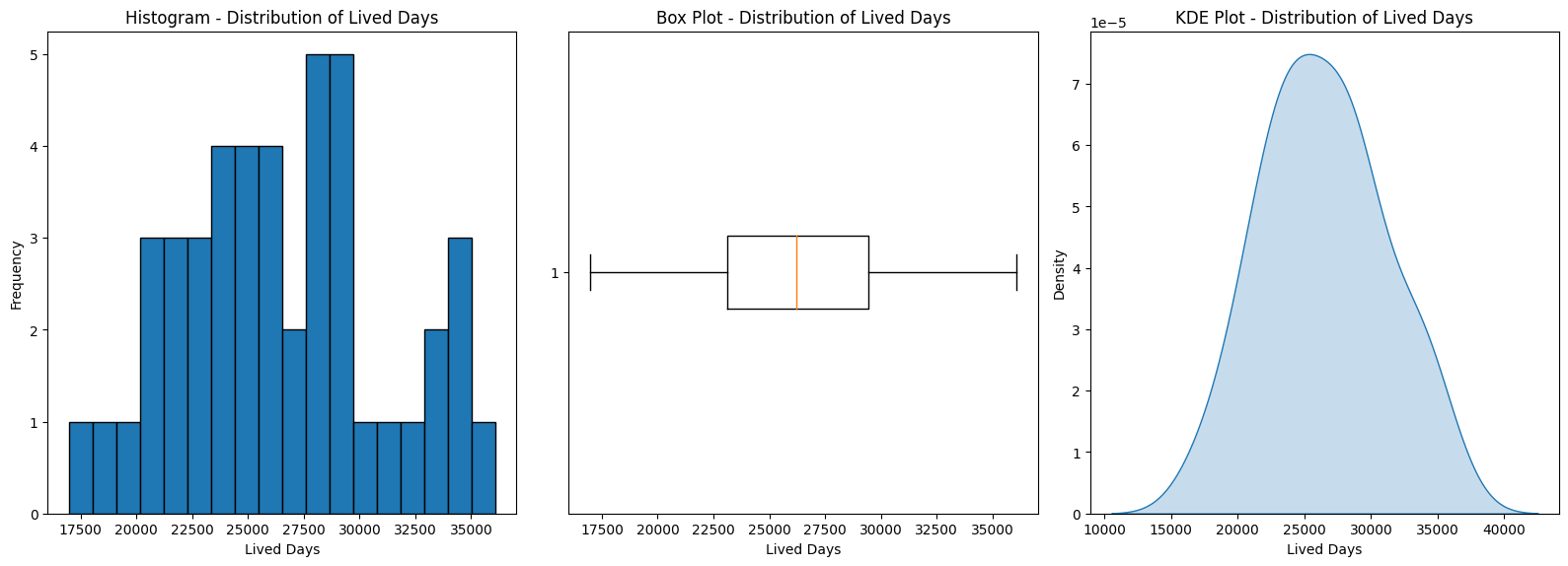


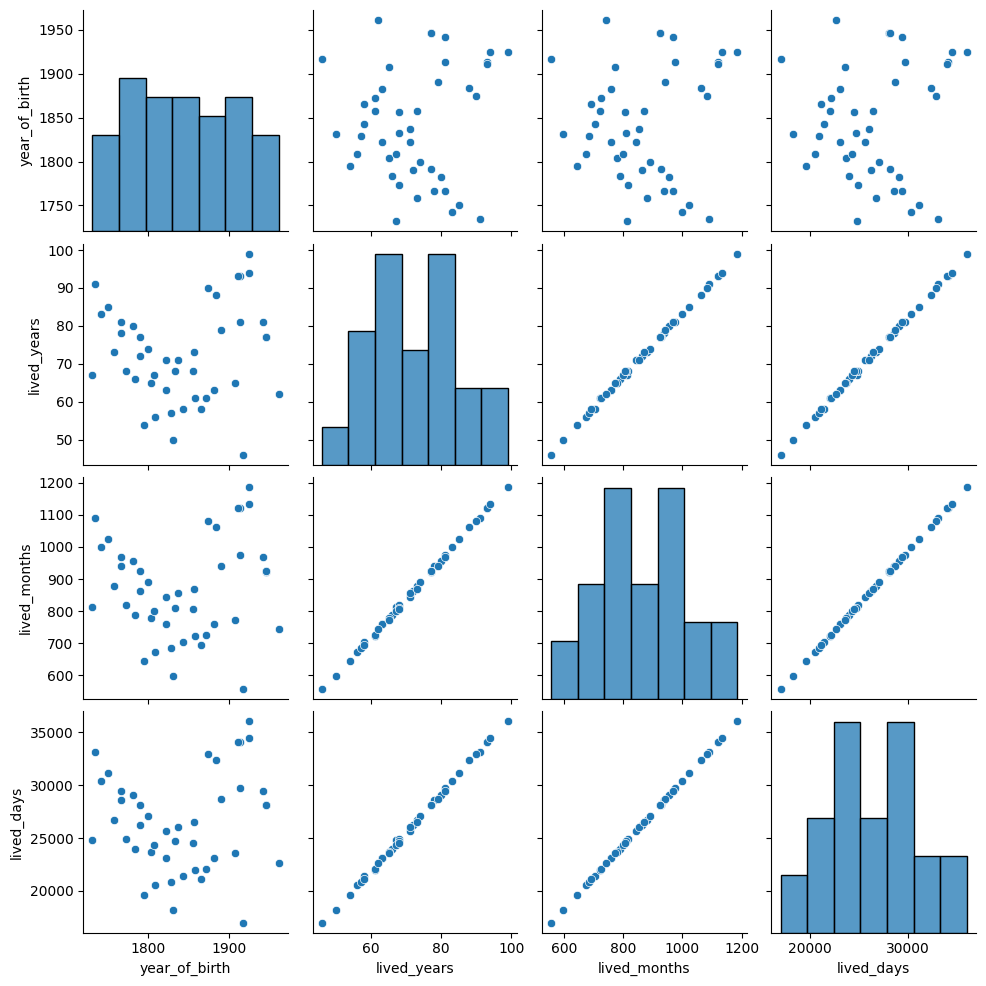
Table 3: Descriptive Statistics of Lived Days



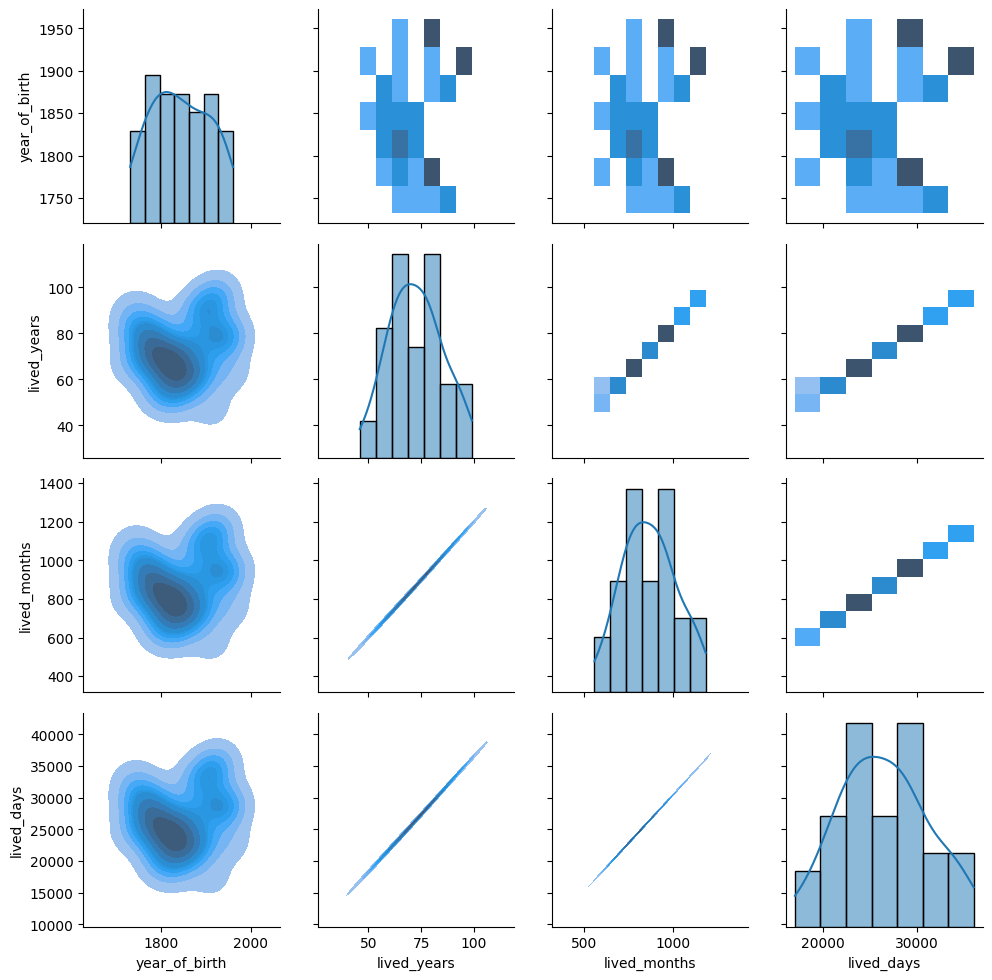
Plot: Distribution of Lived Days



Pair Plot:



Pair Grid:



Commentary:

* During the data preprocessing phase, the last row of the dataset was removed. This step was necessary as the row contained a reference link that was not relevant to the analysis and could potentially interfere with the accuracy of the results.
* It was discovered that the DEATH DATE column had missing values (NaN) for certain presidents, including Jimmy Carter, Bill Clinton, George W. Bush, Barack Obama, Donald Trump, and Joe Biden. This was expected since these presidents are still alive at the time of the analysis. This distinction was taken into account to ensure the integrity of subsequent calculations and analysis.
* In order to standardize the data and facilitate accurate analysis, the BIRTH DATE and DEATH DATE columns were converted to the pandas' DateTime format. These columns contained dates in two different formats: one format with the month represented in short form (e.g., Jan) and the other with the month represented in full form (e.g., January). By converting both formats to a uniform structure of (YYYY-MM-DD), the data became consistent and compatible with the chosen analytical methods.
* To handle the missing values in the DEATH DATE column for the living presidents mentioned earlier, the index of rows with NaN values was identified. As these presidents are still alive, their DEATH DATE values were updated with the current date. This step ensured that the analysis captured the correct statistical values and provided accurate plot representations.
* As part of the analysis process, the rows corresponding to the living presidents were excluded from further calculations and analysis. This step allowed for a focused examination of the historical presidents' lived years and days, providing valuable insights into their lifespans and enabling meaningful comparisons and visualizations.

Conclusion:

**Stats**:

Weighted Average: The weighted average lived days for US Presidents is approximately 22,617. This value takes into account the lifespan distribution of each president, giving more weight to those with longer-lived days. It provides a representative estimate of the average duration of life among the presidents.

Median: The median lived days value is 25,673. This value represents the midpoint of the sorted list of lived days. It indicates that half of the US Presidents in the dataset had lived for fewer than 25,673 days, while the other half had lived for more than 25,673 days. It serves as a robust measure of the central tendency of the lived days distribution.

Mode: The mode lived days value is 16,978. This value represents the most frequently occurring lived days in the dataset. It suggests that there is a significant cluster of US Presidents who lived for approximately 16,978 days, indicating a relatively common duration among a subset of presidents.

Mean: The mean lived days value is approximately 26,097. This value represents the arithmetic average of the lived days for US Presidents. It provides an overall measure of the central tendency of the distribution. It should be noted that the mean can be influenced by outliers in the data, which may skew the result.

Maximum: The maximum lived days value is 34,504. This value represents the longest lifespan among the US Presidents in the dataset. It indicates that one president had a significantly longer duration of life compared to others, serving as an upper boundary for lived days.

Minimum: The minimum lived days value is 16,978. This value represents the shortest lifespan among the US Presidents in the dataset. It indicates that one president had a relatively shorter duration of life compared to others, serving as a lower boundary for lived days.

Standard Deviation: The standard deviation of approximately 4,596 provides a measure of the spread or variability of the lived days distribution. It quantifies the average amount of deviation or dispersion of the lived days from the mean. A higher standard deviation suggests a wider range of variations in the lived days among US Presidents.

These statistics offer valuable insights into the distribution, central tendency, and variability of the lived days among US Presidents. They provide a quantitative understanding of the average, typical, and extreme values, allowing for meaningful comparisons and analysis of the lifespan of US Presidents.

**Plots:**

* Histogram:
  + The Histogram plot provides a visual representation of the distribution of lived days among the US Presidents.
  + It allows us to observe the frequency or count of presidents within different ranges of lived days.
  + Insights can be gained regarding the shape of the distribution
* Box Plot:
  + The Box Plot provides a summary of the distribution of lived days, including measures such as median, quartiles, and potential outliers.
  + It allows us to identify the central tendency and spread of lived days.
  + In Box Plot we can identify any extreme values or variations in the lived days among the US Presidents.
* KDE Plot:
  + The KDE Plot (Kernel Density Estimation) has provided an estimation of the probability density function of the lived days distribution.
  + It has provided insights into the smoothness and shape of the distribution.
  + The KDE Plot has highlighted areas of high and low density, allowing us to observe patterns and peaks in the lived days distribution.
* Pairplot:
  + The Pairplot has displayed pairwise relationships between the numerical variables in your dataset.
  + We have observed correlations, trends, or patterns between lived years, lived months, and lived days.
  + Insights have been gained regarding the relationships between these variables and potential insights into the relationship between age and lifespan.
* PairGrid:
  + The PairGrid is a grid of subplots that has helped us see detailed comparisons and analysis between variables.
  + We can also observe relationships between multiple variables simultaneously.
  + By examining scatter plots and distributions across different combinations of variables, we observed relationships, clusters, or outliers within the dataset.