

A Problem with Presidents

Mustafa Gangardiwala

June 2, 2023

1. Introduction

The given dataset with the problem statements contains information about each president starting from first president George Washington to current president Joe Biden. The information is birth date, birthplace, death date and location of death.

2. Implementation

a. Technologies

- i. I have used Python for the problem and performing to task asked in the problem statement.
- ii. The data analysis part is done in a Jupyter Notebook.

b. Libraires

- i. Pandas
- ii. NumPy
- iii. Datetime
- iv. Matplotlib
- v. Seaborn

c. Assumptions

- i. The number of years, months, and days a president has lived is calculated from their birthday to present date which is June 2, 2023.
- ii. While calculating weighted average I have used president's lived years as weights

d. Running Instruction

- i. Requirement.txt contains all the libraries that are used in the analysis to and can be ran using "*pip install -r requirements.txt*"
- ii. The analysis is in the "Solution.ipynb" file which is Jupyter Notebook.

3. Summary

The dataset contains 47 rows, with 5 columns out which last row is a reference which I removed. The longest-lived president is Jimmy Carter aged 99 years old currently. The shortest-lived president is John F. Kennedy who died at the age of 46 years old. The average life span of a US president is 72.467 years which comes out to be 26446 days and median is 72 years which is 26227 days.

4. Observations

a. Top 10 Presidents from longest lived to shortest lived

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
37	Jimmy Carter	1924-10-01	Plains, Georgia	NaT	NaN	1924	99	1184	36038
39	George Bush	1924-06-12	Milton, Mass.	2018-11-30	Houston, Texas	1924	94	1133	34504
36	Gerald Ford	1913-07-14	Omaha, Nebraska	2006-12-26	Rancho Mirage, Cal.	1913	93	1121	34133
38	Ronald Reagan	1911-02-06	Tampico, Illinois	2004-06-05	Los Angeles, Cal.	1911	93	1120	34088
1	John Adams	1735-10-30	Quincy, Mass.	1826-07-04	Quincy, Mass.	1735	91	1089	33119
29	Herbert Hoover	1874-08-10	West Branch, Iowa	1964-10-20	New York, New York	1874	90	1082	32943
31	Harry S. Truman	1884-05-08	Lamar, Missouri	1972-12-26	Kansas City, Missouri	1884	88	1063	32373
3	James Madison	1751-03-16	Port Conway, Va.	1836-06-28	Orange Co., Va.	1751	85	1023	31150
2	Thomas Jefferson	1743-04-13	Albemarle Co., Va.	1826-07-04	Albemarle Co., Va.	1743	83	999	30397
35	Richard Nixon	1913-01-09	Yorba Linda, Cal.	1994-04-22	New York, New York	1913	81	975	29688

Figure 1 Top 10 Longest lived presidents

b. Top 10 Presidents from shortest lived to longest lived

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
33	John F. Kennedy	1917-05-29	Brookline, Mass.	1963-11-22	Dallas, Texas	1917	46	558	16978
19	James A. Garfield	1831-11-19	Cuyahoga Co., Ohio	1881-09-19	Elberon, New Jersey	1831	50	598	18202
10	James K. Polk	1795-11-02	Mecklenburg Co., N.C.	1849-06-15	Nashville, Tennessee	1795	54	643	19583
15	Abraham Lincoln	1809-02-12	LaRue Co., Kentucky	1865-04-15	Washington, D.C.	1809	56	674	20516
20	Chester Arthur	1829-10-05	Fairfield, Vermont	1886-11-18	New York, New York	1829	57	685	20863
27	Warren G. Harding	1865-11-02	Morrow County, Ohio	1923-08-02	San Francisco, Cal.	1865	58	693	21091
23	William McKinley	1843-01-29	Niles, Ohio	1901-09-14	Buffalo, New York	1843	58	704	21412
24	Theodore Roosevelt	1858-10-27	New York, New York	1919-01-06	Oyster Bay, New York	1858	61	723	21985
28	Calvin Coolidge	1872-07-04	Plymouth, Vermont	1933-01-05	Northampton, Mass.	1872	61	726	22099
42	Barack Obama	1961-08-04	Honolulu, Hawaii	NaT	NaN	1961	62	742	22582

Figure 2 Top 10 Shortest lived presidents

c. Statistical Values

Statistic	Value
Mean	26446.044444
Weighted Average	27223.485434
Median	26227.0
Mode	[16978, 18202, 19583, 20516, 20863, 21091, 214...
Max	36038
Min	16978
Standard Deviation	4600.48696

Figure 3 Statistical Data about "lived_days"

	year_of_birth	lived_years	lived_months	lived_days
count	45.000000	45.000000	45.000000	45.000000
mean	1844.333333	72.466667	868.977778	26446.044444
std	65.584782	12.531052	151.055602	4600.486960
min	1732.000000	46.000000	558.000000	16978.000000
25%	1791.000000	63.000000	759.000000	23098.000000
50%	1837.000000	72.000000	862.000000	26227.000000
75%	1908.000000	81.000000	967.000000	29414.000000
max	1961.000000	99.000000	1184.000000	36038.000000

Figure 4 Statistical Data about newly added columns

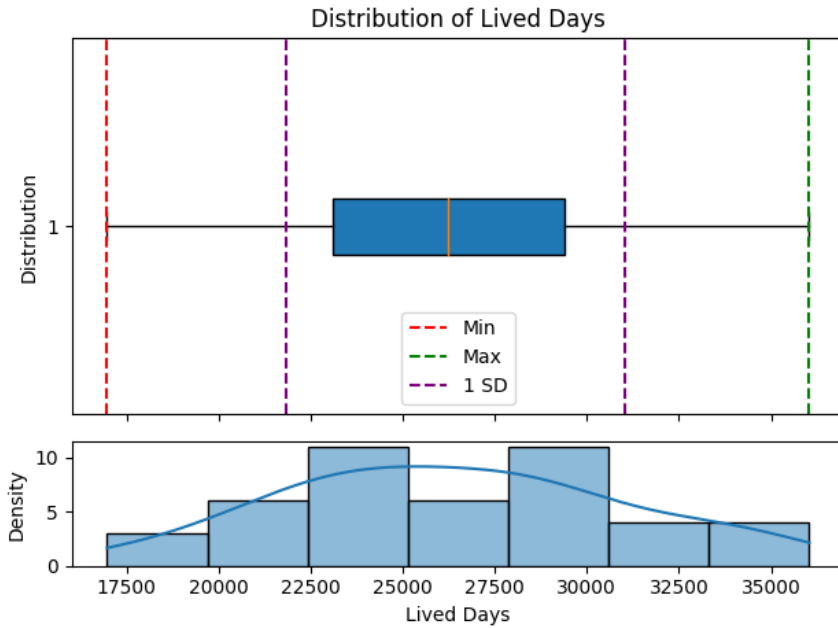


Figure 5 Box Plot with histogram showing the desnsity of presidents lived days

I decided to go with a box plot because it the way to show distribution of data.

5. Conclusion

Based on the above analysis it can be concluded that the average life-span of a US president is 72.467 years with the range being between 46 years to 99 years. There is a need for more data for further analysis.

6. Future Scope

A heat maps of US states in which presidents were born and died can created for further analysis. Apart, from that we can also include the data about the causes of deaths of US presidents. Another, thing I want to do in the analysis is to see the average life-span of a president after leaving office.