REPORT

A Problem with Presidents

Purva Puranik

17th January 2023

Table of Contents

1		Introduction	2
2		Summary	2
3		Assumptions and Pre-Processing	2
4		Requirements, and Observations	3
	I	U.S. Presidents Birth and Death Information	3
	II	Top 10 Presidents of the United States from longest-lived to shortest lived	5
	III	Top 10 Presidents of the United States from shortest-lived to longest-lived	6
	IV	Statistical Analysis	7
5		Implementation	9
6		Conclusion	10
7		Future Scope	10

1. Introduction

This report is a summary of the solution to 'A Problem with Presidents; a data to information challenge.' (As posted under https://github.com/senrabc/a_problem_with_presidents.)

We are given data associated with 45 different presidents that the United States of America has had so far. It contains various details about each president: name, birth year, death year, birth location, and death location. Given these details, our challenge is to derive additional information, visualization, and statistical analysis in order to gain more insights.

2. Summary

According to the insights obtained from analyzing and visualization the data of 45 different presidents of the United States of America, we observe that the average age of the presidents so far is 71.9. With the maximum age being 98 for president Jimmy Carter (Alive) and the minimum being 46 for president John F. Kennedy (deceased). This gives that the age range of all the presidents the USA has had so far lies between 46 to 98 with the median age being 71.

3. Assumptions and Pre-Processing –

- While analyzing the data, I made assumptions that the data given is error-free.
- The missing values in CSV pertaining to the fields death date and location of death reflect that the president is alive. Accordingly, the data is pre-processed such that for records with the location of death and death date missing, a value 'Alive' is populated to handle computation and better understand the data.
- A new field 'latest_living_date' is introduced to mark the latest living date for every president. For the deceased president this would be the same as the death date, whereas, for the others, this would be today's date. This is done in order to assist in computing a few other fields.

4. Requirements and Observations

I. U.S. Presidents Birth and Death Information

- Requirement 1: Compute year_of_birth for the data set.

Given the dataset, I derived the field year_of_birth for every president by utilizing the birth date and extracting the year from the value. Figure 1. Below demonstrates the value associated with year_of_birth for all records in the data set.

- Requirement 2: Compute lived_years for the data set.

Based on the difference between the latest living date and birth date, the number of years the president lived is computed. Figure 1. Below demonstrates the value associated with lived_years for all records in the data set.

- Requirement 3: Compute lived_months for the data set.

Based on the difference between the latest living date and birth date, the number of months the president lived is computed. Figure 1. Below demonstrates the value associated with lived_months for all records in the data set.

- Requirement 4: Compute lived_days for the data set.

Based on the difference between the latest living date and birth date, the number of days the president lived is computed. Figure 1. Below demonstrates the value associated with lived_days for all records in the data set.

U.S. Presidents Birth and Death Information

PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
George Washington	Feb 22, 1732	Westmoreland Co., Va.	Dec 14, 1799	Mount Vernon, Va.	1732	67	813	24767
John Adams	Oct 30, 1735	Quincy, Mass.	July 4, 1826	Quincy, Mass.	1735	90	1088	33119
Thomas Jefferson	Apr 13, 1743	Albemarle Co., Va.	July 4, 1826	Albemarle Co., Va.	1743	83	998	30397
James Madison	Mar 16, 1751	Port Conway, Va.	June 28, 1836	Orange Co., Va.	1751	85	1023	31150
James Monroe	Apr 28, 1758	Westmoreland Co., Va.	July 4, 1831	New York, New York	1758	73	878	26729
John Quincy Adams	July 11, 1767	Quincy, Mass.	Feb 23, 1848	Washington, D.C.	1767	80	967	29446
Andrew Jackson	Mar 15, 1767	Waxhaws, No/So Carolina	June 8, 1845	Nashville, Tennessee	1767	78	938	28574
Martin Van Buren	Dec 5, 1782	Kinderhook, New York	July 24, 1862	Kinderhook, New York	1782	79	955	29085
William Henry Harrison	Feb 9, 1773	Charles City Co., Va.	Apr 4, 1841	Washington, D.C.	1773	68	817	24890
John Tyler	Mar 29, 1790	Charles City Co., Va.	Jan 18, 1862	Richmond, Va.	1790	71	861	26227
James K. Polk	Nov 2, 1795	Mecklenburg Co., N.C.	June 15, 1849	Nashville, Tennessee	1795	53	643	19583
Zachary Taylor	Nov 24, 1784	Orange County, Va.	July 9, 1850	Washington, D.C	1784	65	787	23967
Millard Fillmore	Jan 7, 1800	Cayuga Co., New York	Mar 8, 1874	Buffalo, New York	1800	74	889	27088
Franklin Pierce	Nov 23, 1804	Hillsborough, N.H.	Oct 8, 1869	Concord, New Hamp.	1804	64	778	23695
James Buchanan	Apr 23, 1791	Cove Gap, Pa.	June 1, 1868	Lancaster, Pa.	1791	77	925	28163
Abraham Lincoln	Feb 12, 1809	LaRue Co., Kentucky	Apr 15, 1865	Washington, D.C.	1809	56	674	20516
Andrew Johnson	Dec 29, 1808	Raleigh, North Carolina	July 31, 1875	Elizabethton, Tenn.	1808	66	799	24320
Ulysses S. Grant	Apr 27, 1822	Point Pleasant, Ohio	July 23, 1885	Wilton, New York	1822	63	758	23098
Rutherford B. Hayes	Oct 4, 1822	Delaware, Ohio	Jan 17, 1893	Fremont, Ohio	1822	70	843	25673
James A. Garfield	Nov 19, 1831	Cuyahoga Co., Ohio	Sep 19, 1881	Elberon, New Jersey	1831	49	598	18202
Chester Arthur	Oct 5, 1829	Fairfield, Vermont	Nov 18, 1886	New York, New York	1829	57	685	20863
Grover Cleveland	Mar 18, 1837	Caldwell, New Jersey	June 24, 1908	Princeton, New Jersey	1837	71	855	26030
Benjamin Harrison	Aug 20, 1833	North Bend, Ohio	Mar 13, 1901	Indianapolis, Indiana	1833	67	810	24676
William McKinley	Jan 29, 1843	Niles, Ohio	Sep 14, 1901	Buffalo, New York	1843	58	703	21412
Theodore Roosevelt	Oct 27, 1858	New York, New York	Jan 6, 1919	Oyster Bay, New York	1858	60	722	21985
William Howard Taft	Sep 15, 1857	Cincinnati, Ohio	Mar 8, 1930	Washington, D.C.	1857	72	869	26471
Woodrow Wilson	Dec 28, 1856	Staunton, Virginia	Feb 3, 1924	Washington, D.C.	1856	67	805	24507
Warren G. Harding	Nov 2, 1865	Morrow County, Ohio	Aug 2, 1923	San Francisco, Cal.	1865	57	692	21091
Calvin Coolidge	July 4, 1872	Plymouth, Vermont	Jan 5, 1933	Northampton, Mass.	1872	60	726	22099
Herbert Hoover	Aug 10, 1874	West Branch, Iowa	Oct 20, 1964	New York, New York	1874	90	1082	32943
Franklin Roosevelt	Jan 30, 1882	Hyde Park, New York	Apr 12, 1945	Warm Springs, Ga.	1882	63	758	23082
Harry S. Truman	May 8, 1884	Lamar, Missouri	Dec 26, 1972	Kansas City, Missouri	1884	88	1063	32373
Dwight Eisenhower	Oct 14, 1890	Denison, Texas	Mar 28, 1969	Washington, D.C.	1890	78	941	28654
John F. Kennedy	May 29, 1917	Brookline, Mass.	Nov 22, 1963	Dallas, Texas	1917	46	557	16978
Lyndon B. Johnson	Aug 27, 1908	Gillespie Co., Texas	Jan 22, 1973	Gillespie Co., Texas	1908	64	772	23524
Richard Nixon	Jan 9, 1913	Yorba Linda, Cal.	Apr 22, 1994	New York, New York	1913	81	975	29688
Gerald Ford	July 14, 1913	Omaha, Nebraska	Dec 26, 2006	Rancho Mirage, Cal.	1913	93	1121	34133
Jimmy Carter	Oct 1, 1924	Plains, Georgia	ALIVE	ALIVE	1924	98	1179	35902
Ronald Reagan	Feb 6, 1911	Tampico, Illinois	June 5, 2004	Los Angeles, Cal.	1911	93	1119	34088
George Bush	June 12, 1924	Milton, Mass.	Nov 30, 2018	Houston, Texas	1924	94	1133	34504
Bill Clinton	Aug 19, 1946	Hope, Arkansas	ALIVE	ALIVE	1946	76	916	27910
George W. Bush	July 6, 1946	New Haven, Conn.	ALIVE	ALIVE	1946	76	918	27954
Barack Obama	Aug 4, 1961	Honolulu, Hawaii	ALIVE	ALIVE	1961	61	737	22446
Donald Trump	June 14, 1946	New York, New York	ALIVE	ALIVE	1946	76	919	27976
Joe Biden	Nov 20, 1942	Scranton, Pa.	ALIVE	ALIVE	1942	80	961	29278

Figure 1. U.S. Presidents Birth and Death Information

II. Top 10 Presidents of the United States from longest-lived to shortest-lived

- Requirement 5: From the given data set, find the top 10 presidents of the United States from longest-lived to shortest-lived age.

  The data is sorted in descending order based on age and the top 10 records are extracted. Figure 2. Below demonstrates this data, reflecting that the longest-living president is president Jimmy Carter (alive). The longest-living president who is no more with us was George Bush aged 94, who died in the year 2018.

Top 10 Presidents of the United States from longest lived to shortest lived

PRESIDENT	BIRTH DATE	DEATH DATE	Age (In Years)
Jimmy Carter	Oct 1, 1924	ALIVE	98
George Bush	June 12, 1924	Nov 30, 2018	94
Gerald Ford	July 14, 1913	Dec 26, 2006	93
Ronald Reagan	Feb 6, 1911	June 5, 2004	93
John Adams	Oct 30, 1735	July 4, 1826	90
Herbert Hoover	Aug 10, 1874	Oct 20, 1964	90
Harry S. Truman	May 8, 1884	Dec 26, 1972	88
James Madison	Mar 16, 1751	June 28, 1836	85
Thomas Jefferson	Apr 13, 1743	July 4, 1826	83
Richard Nixon	Jan 9, 1913	Apr 22, 1994	81

Figure 2. Top 10 Presidents of the United States from longest-lived to shortest-lived

III. Top 10 Presidents of the United States from shortest-lived to longest-lived

- Requirement 6: From the given data set, find the top 10 presidents of the United States from shortest-lived to longest-lived to age.

  The data is sorted in descending order based on age and the top 10 records are extracted. Figure 3. Below demonstrates this data, reflecting that the shorted-living president is president John F. Kennedy (deceased). The shortest-aged president alive is Barack Obama aged 61 as of 17th January 2023.

Top 10 Presidents of the United States from shortest lived to longest lived

PRESIDENT	BIRTH DATE	DEATH DATE	Age (In Years)
John F. Kennedy	May 29, 1917	Nov 22, 1963	46
James A. Garfield	Nov 19, 1831	Sep 19, 1881	49
James K. Polk	Nov 2, 1795	June 15, 1849	53
Abraham Lincoln	Feb 12, 1809	Apr 15, 1865	56
Chester Arthur	Oct 5, 1829	Nov 18, 1886	57
Warren G. Harding	Nov 2, 1865	Aug 2, 1923	57
William McKinley	Jan 29, 1843	Sep 14, 1901	58
Theodore Roosevelt	Oct 27, 1858	Jan 6, 1919	60
Calvin Coolidge	July 4, 1872	Jan 5, 1933	60
Barack Obama	Aug 4, 1961	ALIVE	61

Figure 3. Top 10 Presidents of the United States from shortest-lived to longest-lived

IV. Statistical Analysis

- Requirement 7: Calculate the mean of lived_days

It is observed that the mean age in terms of days of all the presidents is 26427.91. Refer to Figure 4. Below.

- Requirement 8: Calculate the weighted average of lived_days

It is observed that the weighted average age in terms of days of all the presidents is 26427.91. The weighted average will remain the same as we have just one sample data. The error associated with each measurement (lived time) is the same(0). Refer to Figure 4. Below.

- Requirement 9: Calculate the median of lived_days

It is observed that the median in terms of days of all the presidents is 26227.0. Refer to Figure 4. Below.

- Requirement 10: Calculate the mode of lived_days

It is observed that the mode (mode frequent age) in terms of days of all the presidents is 24455 and 27740. In terms of age, this computes to 67 and 76 ages of three presidents each. Refer to Figure 4. Below.

- Requirement 11: Calculate the maximum number from lived_days

It is observed that the maximum number of days a president has lived is 35902. This reflects the longest-living president Jimmy Carter (alive) who is age 98 as of 17th Jan 2023. Refer to Figure 4. Below.

- Requirement 12: Calculate the minimum number from lived_days

It is observed that the minimum number of days a president has lived is 16978. This reflects the shortest-living president John F. Kennedy (deceased) aged 46. Refer to Figure 4. Below.

- Requirement 13: Calculate the standard deviation of lived_days

The standard deviation for the number of days lived is 4591.57.

	Lived Days
Mean	26427.911111
Weighted average	26427.911111
Median	26227.0
Mode	[24455, 27740]
Max	35902
Min	16978
Standard Deviation	4591.574675

Figure 4. Statistics associated with lived days.

- Requirement 14: Plot graph demonstration observations
    - A histogram is plotted based on the living years of all the presidents, instead of using living months, as this helps in getting a better understanding of the data and clearer visualization. Figure 5. The statistical Analysis Graph demonstrates the same.
    - The ages are distributed into various bins between the range of age 40 to 100 to accommodate all ages.
    - The plot also depicts the mean, median, and mode associated with the data.
    - The mean value of the age of the presidents in terms of years is 71.9. The median age in terms of years is 71.0.
    - Whereas the most frequent ages, given using mode are 67 and 76 years.
    - There is just one president between the age of 95 to 100, which is Jimmy Carter.
    - Whereas most president count is 7, pertaining to ages between 65 to 70 and 75 to 80.

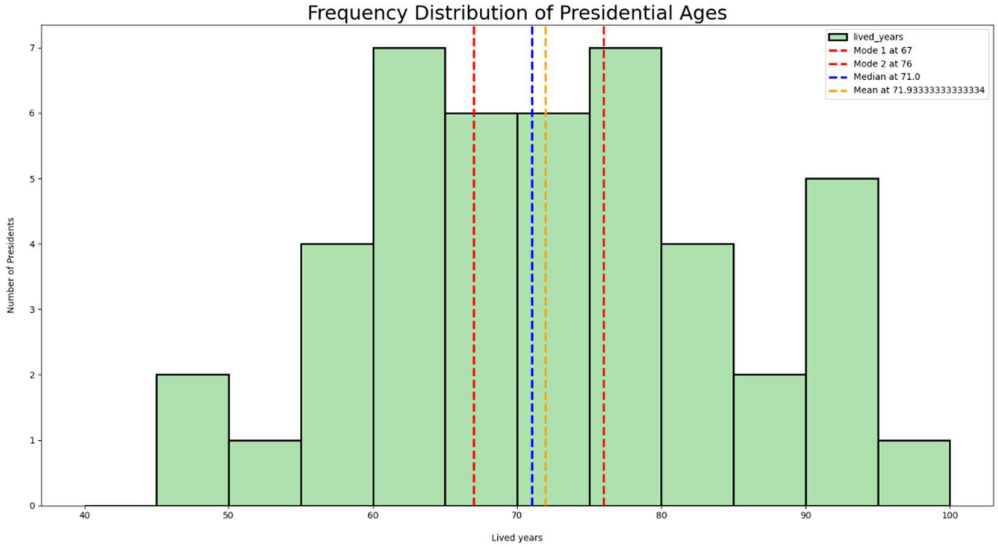

Figure 5. Frequency Distribution of Presidential Ages

5. Implementation

This challenge is solved using python. Below are the files included –

A. PresidentProblem.py
- This is the main file which contains a class PresidentData consisting of a member variable df that is dataframe to hold the data imported from CSV.
- The class consists of methods given below which process and work upon the dataframe loaded.
    - preprocessPresidentData()
        - This function helps to load the data from CSV and perform the necessary pre-processing. This solves the requirement 1, 2, 3, and 4
    - topTenLongestToShortestLivedPresident()
        - This function helps in obtaining the top 10 presidents from longest to shortest lived, providing the solution to requirement 5
    - topTenShortestToLongestLivedPresident()
        - This function helps in obtaining the top 10 presidents from shortest to longest lived, providing the solution to requirement 6
    - calculateStats()
        - This function helps in obtaining the mean, weighted average, median, mode, max, min, and standard deviation of lived_days associated with requirements 7, 8, 9, 10, 11, 12, and 13
    - plotGraph()
        - This function helps in plotting histogram for Frequency Distribution of Presidential Ages which also highlights the mean, median and mode of data based on age in years.

B. Requirements.txt

This file mentions the third-party packages/libraries required to be downloaded in order to perform the necessary computations to solve this challenge. The packages/libraries used are:-
- Numpy 1.23.2 [Majorly used to perform necessary numerical computation]
- Pandas 1.4.4 [Majorly used to deal with data frames]
- dataframe-image 0.1.2 [Used to export the data frame images]
- matplotlib 3.5.3 [Used to plot graphs]

C. PresidentData.csv

Data for 45 different presidents of the United States of America are included in the input CSV file. Each president's name, birth year, death year, place of birth, and place of death are among the information provided.

6. Conclusion

- On examining the data, we get insights into the life expectancy of the presidents. We note that the presidents' average age to date is 71.9.
- President Jimmy Carter (Alive) is the oldest living president at 98, while President John F. Kennedy is the youngest at 46. (deceased).
- The median age of all the presidents the United States has had so far is 71, with a range of 46 to 98.
- The median age closely reflects the fact that most presidents have lived over the age of 60.
- The lived days values will display better details and better illustrate the distribution.
- While there is a clear trend toward normal distribution in the frequency distribution of presidential ages, there are also progressively more presidents who live lengthy lives.

7. Future Scope

- Segregating the data of alive and deceased presidents can help us in getting better insights.
- The fact that few of the presidents are no more due to untimely death associated with unnatural conditions, makes the data skewed and hampers the computation of various parameters.
- Such as the assassination of the shortest-lived president affects the age range of the presidential data.
- Thus, a more detailed analysis of other external factors, such as the reason for death, can help in obtaining more insights.
- A few other observations that could be done are based on the birth year and the number of years the presidents have lived, to help find information regarding if with increasing years, the life expectancy has increased.