Cleaning Data with multiple Data
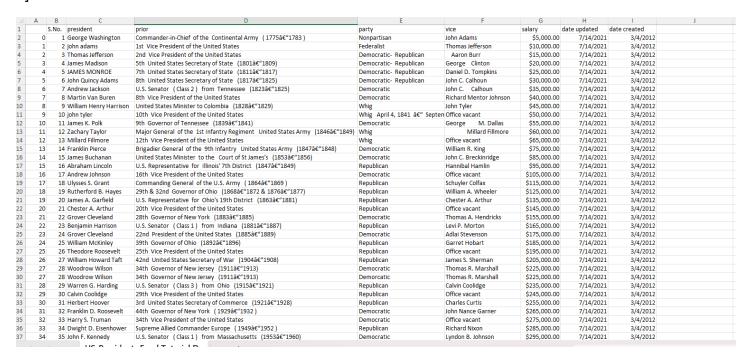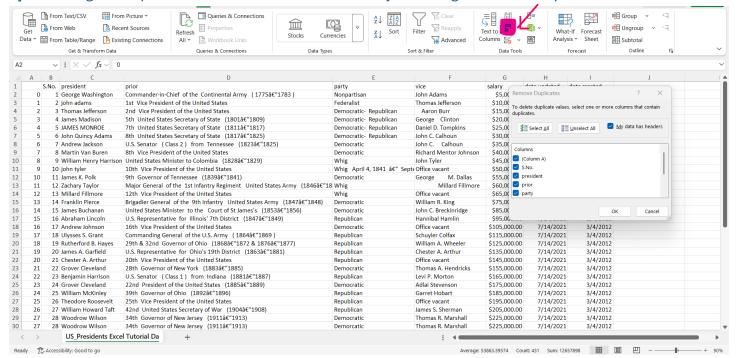
Below is the screenshot of the raw data given to us. From my analysis, I can see that there are

1] Capital letters used in some of the President Name.

2] Unnecessary spaces before the Vice Name.
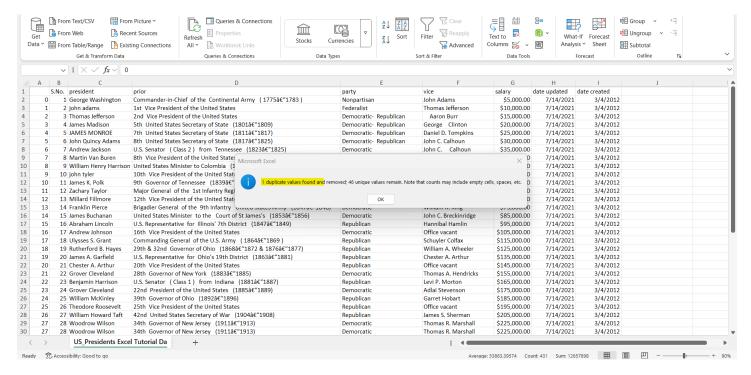
3] Some random values added after the Prior etc

	S.No.	president	prior	party	vice	salary	date updated	date created	
1									
2	0	1 George Washington	Commander-in-Chief of the Continental Army (1775â€"1783)	Nonpartisan	John Adams	$5,000.00	7/14/2021	3/4/2012	
3	1	2 john adams	1st Vice President of the United States	Federalist	Thomas Jefferson	$10,000.00	7/14/2021	3/4/2012	
4	2	3 Thomas Jefferson	2nd Vice President of the United States	Democratic- Republican	Aaron Burr	$15,000.00	7/14/2021	3/4/2012	
5	3	4 James Madison	5th United States Secretary of State (1801â€"1809)	Democratic- Republican	George Clinton	$20,000.00	7/14/2021	3/4/2012	
6	4	5 JAMES MONROE	7th United States Secretary of State (1811â€"1817)	Democratic- Republican	Daniel D. Tompkins	$25,000.00	7/14/2021	3/4/2012	
7	5	6 John Quincy Adams	8th United States Secretary of State (1817â€"1825)	Democratic- Republican	John C. Calhoun	$30,000.00	7/14/2021	3/4/2012	
8	6	7 Andrew Jackson	U.S. Senator (Class 2) from Tennessee (1823â€"1825)	Democratic	John C. Calhoun	$35,000.00	7/14/2021	3/4/2012	
9	7	8 Martin Van Buren	8th Vice President of the United States	Democratic	Richard Mentor Johnson	$40,000.00	7/14/2021	3/4/2012	
10	8	9 William Henry Harrison	United States Minister to Colombia (1828â€"1829)	Whig	John Tyler	$45,000.00	7/14/2021	3/4/2012	
11	9	10 john tyler	10th Vice President of the United States	Whig April 4, 1841 â€" Septen	Office vacant	$50,000.00	7/14/2021	3/4/2012	
12	10	11 James K. Polk	9th Governor of Tennessee (1839â€"1841)	Democratic	George M. Dallas	$55,000.00	7/14/2021	3/4/2012	
13	11	12 Zachary Taylor	Major General of the 1st Infantry Regiment United States Army (1846â€"1849)	Whig	Millard Fillmore	$60,000.00	7/14/2021	3/4/2012	
14	12	13 Millard Fillmore	12th Vice President of the United States	Whig	Office vacant	$65,000.00	7/14/2021	3/4/2012	
15	13	14 Franklin Pierce	Brigadier General of the 9th Infantry United States Army (1847â€"1848)	Democratic	William R. King	$75,000.00	7/14/2021	3/4/2012	
16	14	15 James Buchanan	United States Minister to the Court of St James's (1853â€"1856)	Democratic	John C. Breckinridge	$85,000.00	7/14/2021	3/4/2012	
17	15	16 Abraham Lincoln	U.S. Representative for Illinois' 7th District (1847â€"1849)	Republican	Hannibal Hamlin	$95,000.00	7/14/2021	3/4/2012	
18	16	17 Andrew Johnson	16th Vice President of the United States	Democratic	Office vacant	$105,000.00	7/14/2021	3/4/2012	
19	17	18 Ulysses S. Grant	Commanding General of the U.S. Army (1864â€"1869)	Republican	Schuyler Colfax	$115,000.00	7/14/2021	3/4/2012	
20	18	19 Rutherford B. Hayes	29th & 32nd Governor of Ohio (1868â€"1872 & 1876â€"1877)	Republican	William A. Wheeler	$125,000.00	7/14/2021	3/4/2012	
21	19	20 James A. Garfield	U.S. Representative for Ohio's 19th District (1863â€"1881)	Republican	Chester A. Arthur	$135,000.00	7/14/2021	3/4/2012	
22	20	21 Chester A. Arthur	20th Vice President of the United States	Republican	Office vacant	$145,000.00	7/14/2021	3/4/2012	
23	21	22 Grover Cleveland	28th Governor of New York (1883â€"1885)	Democratic	Thomas A. Hendricks	$155,000.00	7/14/2021	3/4/2012	
24	22	23 Benjamin Harrison	U.S. Senator (Class 1) from Indiana (1881â€"1887)	Republican	Levi P. Morton	$165,000.00	7/14/2021	3/4/2012	
25	23	24 Grover Cleveland	22nd President of the United States (1885â€"1889)	Democratic	Adlai Stevenson	$175,000.00	7/14/2021	3/4/2012	
26	24	25 William McKinley	39th Governor of Ohio (1892â€"1896)	Republican	Garret Hobart	$185,000.00	7/14/2021	3/4/2012	
27	25	26 Theodore Roosevelt	25th Vice President of the United States	Republican	Office vacant	$195,000.00	7/14/2021	3/4/2012	
28	26	27 William Howard Taft	42nd United States Secretary of War (1904â€"1908)	Republican	James S. Sherman	$205,000.00	7/14/2021	3/4/2012	
29	27	28 Woodrow Wilson	34th Governor of New Jersey (1911â€"1913)	Democratic	Thomas R. Marshall	$225,000.00	7/14/2021	3/4/2012	
30	27	28 Woodrow Wilson	34th Governor of New Jersey (1911â€"1913)	Demorcatic	Thomas R. Marshall	$225,000.00	7/14/2021	3/4/2012	
31	28	29 Warren G. Harding	U.S. Senator (Class 3) from Ohio (1915â€"1921)	Republican	Calvin Coolidge	$235,000.00	7/14/2021	3/4/2012	
32	29	30 Calvin Coolidge	29th Vice President of the United States	Republican	Office vacant	$245,000.00	7/14/2021	3/4/2012	
33	30	31 Herbert Hoover	3rd United States Secretary of Commerce (1921â€"1928)	Republican	Charles Curtis	$255,000.00	7/14/2021	3/4/2012	
34	31	32 Franklin D. Roosevelt	44th Governor of New York (1929â€"1932)	Democratic	John Nance Garner	$265,000.00	7/14/2021	3/4/2012	
35	32	33 Harry S. Truman	34th Vice President of the United States	Democratic	Office vacant	$275,000.00	7/14/2021	3/4/2012	
36	33	34 Dwight D. Eisenhower	Supreme Allied Commander Europe (1949â€"1952)	Republican	Richard Nixon	$285,000.00	7/14/2021	3/4/2012	
37	34	35 John F. Kennedy	U.S. Senator (Class 1) from Massachusetts (1953â€"1960)	Democratic	Lyndon B. Johnson	$295,000.00	7/14/2021	3/4/2012	

To begin with we need to format and standardize the data.

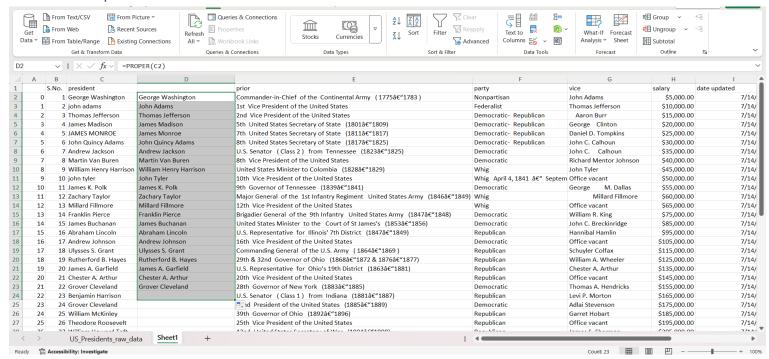## 1] Removing the duplicates values from the Raw Data by following the below steps

When clicked in Ok to remove the duplicates, it found 1 duplicate record and deleted the records as shown below
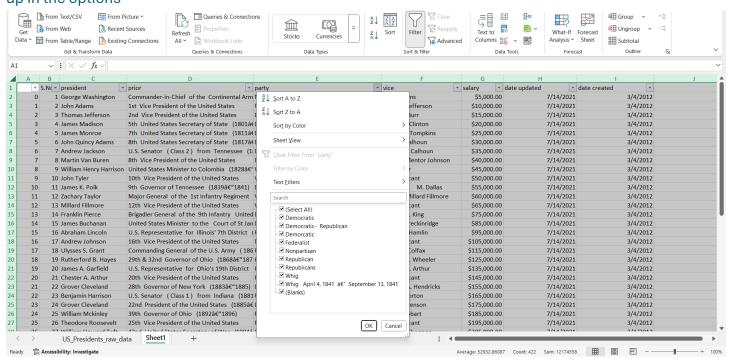

Created a new column to format the president Name Column such that if any mistake occurs I do not damage the original data. Next, Used the =PROPER(Value) formula to set the Name in proper case as few names were in the Capital letters. This formula helps to standardize the format. If you observe the C6 cell we notice that all in Capital letter.

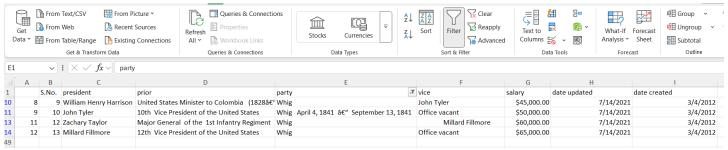## This is the output for this formula

Observing the Party column and used the filter option to check the filters there a few options that popped up in the options
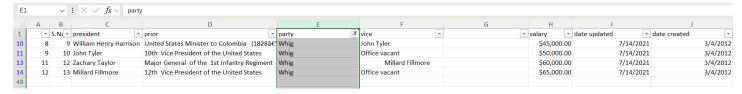

there are two republican and republicans options, when I selected only those two options, I found that it could be the spelling mistake as I could just see one record with a extra "s" in it.

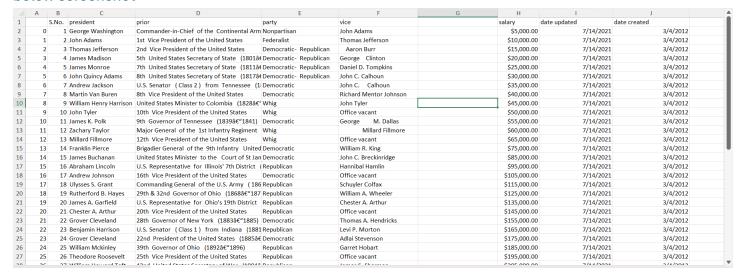|	A	B	C	D	E	F	G	H	I
1		S.N	president	prior	party	vice	salary	date updated	date created
17	15	16	Abraham Lincoln	U.S. Representative for Illinois' 7th District (	Republican	Hannibal Hamlin	$95,000.00	7/14/2021	3/4/2012
19	17	18	Ulysses S. Grant	Commanding General of the U.S. Army (186	Republican	Schuyler Colfax	$115,000.00	7/14/2021	3/4/2012
20	18	19	Rutherford B. Hayes	29th & 32nd Governor of Ohio (1868â€"187	Republican	William A. Wheeler	$125,000.00	7/14/2021	3/4/2012
21	19	20	James A. Garfield	U.S. Representative for Ohio's 19th District	Republican	Chester A. Arthur	$135,000.00	7/14/2021	3/4/2012
22	20	21	Chester A. Arthur	20th Vice President of the United States	Republican	Office vacant	$145,000.00	7/14/2021	3/4/2012
24	22	23	Benjamin Harrison	U.S. Senator (Class 1) from Indiana (1881	Republican	Levi P. Morton	$165,000.00	7/14/2021	3/4/2012
26	24	25	William Mckinley	39th Governor of Ohio (1892â€"1896)	Republican	Garret Hobart	$185,000.00	7/14/2021	3/4/2012
27	25	26	Theodore Roosevelt	25th Vice President of the United States	Republican	Office vacant	$195,000.00	7/14/2021	3/4/2012
28	26	27	William Howard Taft	42nd United States Secretary of War (1904â	Republican	James S. Sherman	$205,000.00	7/14/2021	3/4/2012
31	28	29	Warren G. Harding	U.S. Senator (Class 3) from Ohio (1915â€	Republican	Calvin Coolidge	$235,000.00	7/14/2021	3/4/2012
32	29	30	Calvin Coolidge	29th Vice President of the United States	Republican	Office vacant	$245,000.00	7/14/2021	3/4/2012
33	30	31	Herbert Hoover	3rd United States Secretary of Commerce (1	Republican	Charles Curtis	$255,000.00	7/14/2021	3/4/2012
36	33	34	Dwight D. Eisenhower	Supreme Allied Commander Europe (1949â€	Republican	Richard Nixon	$285,000.00	7/14/2021	3/4/2012
39	36	37	Richard Nixon	36th Vice President of the United States (19!	Republican	Spiro Agnew	$315,000.00	7/14/2021	3/4/2012
40	37	38	Gerald Ford	40th Vice President of the United States	Republican	Office vacant	$325,000.00	7/14/2021	3/4/2012
42	39	40	Ronald Reagan	33rd Governor of California (1967â€"1975	Republican	George H. W. Bush	$345,000.00	7/14/2021	3/4/2012
43	40	41	George H. W. Bush	43rd Vice President of the United States	Republican	Dan Quayle	$355,000.00	7/14/2021	3/4/2012
45	42	43	George W. Bush	46th Governor of Texas (1995â€"2000)	Republican	Dick Cheney	$375,000.00	7/14/2021	3/4/2012
47	44	45	Donald Trump	Chairman of The Trump Organization (1971	Republicans	Mike Pence	$405,000.00	Wednesday, July 14, 2021	Saturday, February 1, 2020
49									

Since there was only One record, I went ahead to correct the single record. Doing the same procedure with the "WHIG" party, seems like there are two types of whig one with a random date and another only whig. It looks like an error, hence we would get rid of that random data after "Whig"
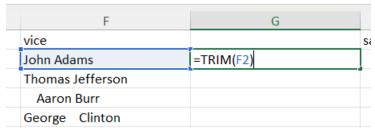

	A	B	C	D	E	F	G	H	I
1		S.No.	president	prior	party	vice	salary	date updated	date created
10	8	9	William Henry Harrison	United States Minister to Colombia (1828â€"	Whig		John Tyler	$45,000.00	7/14/2021	3/4/2012
11	9	10	John Tyler	10th Vice President of the United States	Whig April 4, 1841 â€" September 13, 1841	Office vacant	$50,000.00	7/14/2021	3/4/2012
13	11	12	Zachary Taylor	Major General of the 1st Infantry Regiment	Whig	Millard Fillmore	$60,000.00	7/14/2021	3/4/2012
14	12	13	Millard Fillmore	12th Vice President of the United States	Whig	Office vacant	$65,000.00	7/14/2021	3/4/2012
49									

Below is the corrected version of the column

	A	B	C	D	E	F	G	H	I	J
1		S.N	president	prior	party	vice		salary	date updated	date created
10	8	9	William Henry Harrison	United States Minister to Colombia (1828â€"	Whig	John Tyler		$45,000.00	7/14/2021	3/4/2012
11	9	10	John Tyler	10th Vice President of the United States	Whig	Office vacant		$50,000.00	7/14/2021	3/4/2012
13	11	12	Zachary Taylor	Major General of the 1st Infantry Regiment	Whig	Millard Fillmore		$60,000.00	7/14/2021	3/4/2012
14	12	13	Millard Fillmore	12th Vice President of the United States	Whig	Office vacant		$65,000.00	7/14/2021	3/4/2012
48										

Next lets observe the Vice column, I can observe that there is extra space in that column as shown in the below screenshot
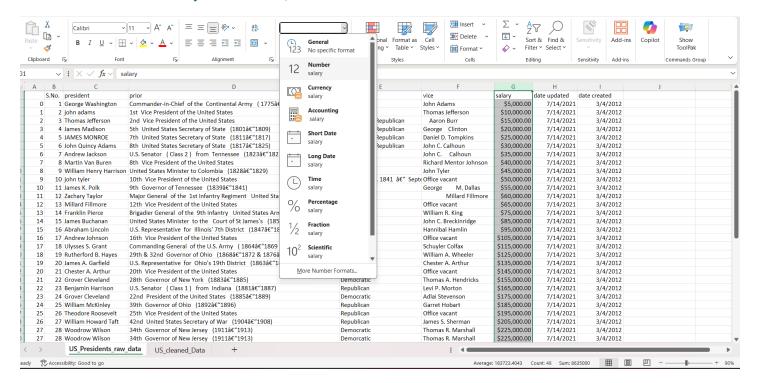
	A	B	C	D	E	F	G	H	I	J
1		S.No.	president	prior	party	vice		salary	date updated	date created
2	0	1 George Washington	Commander-in-Chief of the Continental Arm	Nonpartisan	John Adams		$5,000.00	7/14/2021	3/4/2012
3	1	2 John Adams	1st Vice President of the United States	Federalist	Thomas Jefferson		$10,000.00	7/14/2021	3/4/2012
4	2	3 Thomas Jefferson	2nd Vice President of the United States	Democratic- Republican	Aaron Burr		$15,000.00	7/14/2021	3/4/2012
5	3	4 James Madison	5th United States Secretary of State (1801â€ Democratic- Republican	George Clinton		$20,000.00	7/14/2021	3/4/2012
6	4	5 James Monroe	7th United States Secretary of State (1811â€ Democratic- Republican	Daniel D. Tompkins		$25,000.00	7/14/2021	3/4/2012
7	5	6 John Quincy Adams	8th United States Secretary of State (1817â€ Democratic- Republican	John C. Calhoun		$30,000.00	7/14/2021	3/4/2012
8	6	7 Andrew Jackson	U.S. Senator (Class 2) from Tennessee (1: Democratic	John C. Calhoun		$35,000.00	7/14/2021	3/4/2012
9	7	8 Martin Van Buren	8th Vice President of the United States	Democratic	Richard Mentor Johnson		$40,000.00	7/14/2021	3/4/2012
10	8	9 William Henry Harrison	United States Minister to Colombia (1828â€ Whig	John Tyler		$45,000.00	7/14/2021	3/4/2012
11	9	10 John Tyler	10th Vice President of the United States	Whig	Office vacant		$50,000.00	7/14/2021	3/4/2012
12	10	11 James K. Polk	9th Governor of Tennessee (1839â€"1841) Democratic	George M. Dallas		$55,000.00	7/14/2021	3/4/2012
13	11	12 Zachary Taylor	Major General of the 1st Infantry Regiment Whig	Millard Fillmore		$60,000.00	7/14/2021	3/4/2012
14	12	13 Millard Fillmore	12th Vice President of the United States	Whig	Office vacant		$65,000.00	7/14/2021	3/4/2012
15	13	14 Franklin Pierce	Brigadier General of the 9th Infantry United Democratic	William R. King		$75,000.00	7/14/2021	3/4/2012
16	14	15 James Buchanan	United States Minister to the Court of St Jan Democratic	John C. Breckinridge		$85,000.00	7/14/2021	3/4/2012
17	15	16 Abraham Lincoln	U.S. Representative for Illinois' 7th District Republican	Hannibal Hamlin		$95,000.00	7/14/2021	3/4/2012
18	16	17 Andrew Johnson	16th Vice President of the United States	Democratic	Office vacant		$105,000.00	7/14/2021	3/4/2012
19	17	18 Ulysses S. Grant	Commanding General of the U.S. Army (186 Republican	Schuyler Colfax		$115,000.00	7/14/2021	3/4/2012
20	18	19 Rutherford B. Hayes	29th & 32nd Governor of Ohio (1868â€"187 Republican	William A. Wheeler		$125,000.00	7/14/2021	3/4/2012
21	19	20 James A. Garfield	U.S. Representative for Ohio's 19th District Republican	Chester A. Arthur		$135,000.00	7/14/2021	3/4/2012
22	20	21 Chester A. Arthur	20th Vice President of the United States	Republican	Office vacant		$145,000.00	7/14/2021	3/4/2012
23	21	22 Grover Cleveland	28th Governor of New York (1883â€"1885) Democratic	Thomas A. Hendricks		$155,000.00	7/14/2021	3/4/2012
24	22	23 Benjamin Harrison	U.S. Senator (Class 1) from Indiana (1881 Republican	Levi P. Morton		$165,000.00	7/14/2021	3/4/2012
25	23	24 Grover Cleveland	22nd President of the United States (1885â€ Democratic	Adlai Stevenson		$175,000.00	7/14/2021	3/4/2012
26	24	25 William Mckinley	39th Governor of Ohio (1892â€"1896) Republican	Garret Hobart		$185,000.00	7/14/2021	3/4/2012
27	25	26 Theodore Roosevelt	25th Vice President of the United States	Republican	Office vacant		$195,000.00	7/14/2021	3/4/2012

To fix this issue, I am going to use the TRIM formula that is =TRIM (value). This helps to eliminate the extra space. Lets see the Output below

F	G	
vice		s
John Adams	=TRIM(F2)	
Thomas Jefferson		
Aaron Burr		
George Clinton		

Output:

F	G
vice	
John Adams	John Adams
Thomas Jefferson	Thomas Jefferson
Aaron Burr	Aaron Burr
George Clinton	George Clinton
Daniel D. Tompkins	Daniel D. Tompkins
John C. Calhoun	John C. Calhoun
John C. Calhoun	John C. Calhoun
Richard Mentor Johnson	Richard Mentor Johnson
John Tyler	John Tyler
Office vacant	Office vacant
George M. Dallas	George M. Dallas
Millard Fillmore	Millard Fillmore

Next, while observing the Salary Column we observe that there is a dollar sign which makes it had to put it in the database and may cause issue,

let's just make it simple for us and not get into issues, I am converting the TEXT to number as shown below and we observed that we removed the $ symbol and also the Cents are removed.

Now, Lets see the Date field some are which are not in same format as shown in the figure, Hence we use the Short Date option to fix this issue

Output: All have the same format i.e short date

Now, deleting the columns which are not required for us to analyze.

1] We have two serial no. column which is not required, I have deleted the column.

2] The prior Column as so much data which looks like we do not require those details and hence we are deleting that data as well, as shown in the below.

When we compare the raw data we got, the below is the much cleaner version of the data and looks good for analysis and to do the visualization.

	A	B	C	D	E	F	G
1	S.No.	president-Fixed	party	Vice	salary	date updated	date created
2	1	George Washington	Nonpartisan	John Adams	5000	7/14/2021	3/4/2012
3	2	John Adams	Federalist	Thomas Jefferson	10000	7/14/2021	3/4/2012
4	3	Thomas Jefferson	Democratic- Republican	Aaron Burr	15000	7/14/2021	3/4/2012
5	4	James Madison	Democratic- Republican	George Clinton	20000	7/14/2021	3/4/2012
6	5	James Monroe	Democratic- Republican	Daniel D. Tompkins	25000	7/14/2021	3/4/2012
7	6	John Quincy Adams	Democratic- Republican	John C. Calhoun	30000	7/14/2021	3/4/2012
8	7	Andrew Jackson	Democratic	John C. Calhoun	35000	7/14/2021	3/4/2012
9	8	Martin Van Buren	Democratic	Richard Mentor Johnson	40000	7/14/2021	3/4/2012
10	9	William Henry Harrison	Whig	John Tyler	45000	7/14/2021	3/4/2012
11	10	John Tyler	Whig	Office vacant	50000	7/14/2021	3/4/2012
12	11	James K. Polk	Democratic	George M. Dallas	55000	7/14/2021	3/4/2012
13	12	Zachary Taylor	Whig	Millard Fillmore	60000	7/14/2021	3/4/2012
14	13	Millard Fillmore	Whig	Office vacant	65000	7/14/2021	3/4/2012
15	14	Franklin Pierce	Democratic	William R. King	75000	7/14/2021	3/4/2012
16	15	James Buchanan	Democratic	John C. Breckinridge	85000	7/14/2021	3/4/2012
17	16	Abraham Lincoln	Republican	Hannibal Hamlin	95000	7/14/2021	3/4/2012
18	17	Andrew Johnson	Democratic	Office vacant	105000	7/14/2021	3/4/2012
19	18	Ulysses S. Grant	Republican	Schuyler Colfax	115000	7/14/2021	3/4/2012
20	19	Rutherford B. Hayes	Republican	William A. Wheeler	125000	7/14/2021	3/4/2012
21	20	James A. Garfield	Republican	Chester A. Arthur	135000	7/14/2021	3/4/2012
22	21	Chester A. Arthur	Republican	Office vacant	145000	7/14/2021	3/4/2012
23	22	Grover Cleveland	Democratic	Thomas A. Hendricks	155000	7/14/2021	3/4/2012
24	23	Benjamin Harrison	Republican	Levi P. Morton	165000	7/14/2021	3/4/2012
25	24	Grover Cleveland	Democratic	Adlai Stevenson	175000	7/14/2021	3/4/2012
26	25	William Mckinley	Republican	Garret Hobart	185000	7/14/2021	3/4/2012
27	26	Theodore Roosevelt	Republican	Office vacant	195000	7/14/2021	3/4/2012