

NAME: Shriharsh S Kotecha

REGISTER NO: 22MIC0021

LAB No: 10

PROMPT: Act as a data analyst. I will give you a CSV dataset showing the frequency of terms related to stamp from 1800 to 2022 . The columns are: term, year, term_frequency. First, here is a sample of the data:

Stamp 2019 19580

Stamp 2018 23021

Stamp 2017 29864

Stamp 2016 29342

Stamp 2015 30433

Stamp 1523 1

Stamp 1520 8

Stamp 1485 2

Stamp 1480 2

Stamp 1478 1

Based on the full dataset (which this sample represents), please do the following:

- 1. Data Summary:** Provide a concise numerical summary of the trends for each term. Use Pandas functions like describe(), max(), idxmax() to find peak popularity years, overall trends (increasing/declining), and rate of change.
- 2. Trend Interpretation:** Write a short paragraph (150 words) interpreting these trends. Explain what the rise and fall of these terms tells us about the cultural shift away from formal handwriting, especially post-1865s. Relate it to technological change.
- 3. Generate a Word Cloud:** Conceptually, if you were to generate a word cloud based on the text from books that most frequently used the term "STAMP" during its peak era what are the 10-15 most associated words you would expect to see? These should be words related to education, discipline, art, business, and tools. List them with relative weights (e.g., #1 largest, #2 smaller, etc.). Finally, note which specific NumPy and Pandas functions you would use programmatically to generate the summary in step1.

Step 1. Data Summary (Numerical Analysis)

If we take the full dataset (1800–2022), using Pandas + NumPy we would extract:

- **Descriptive stats**
- `df['term_frequency'].describe()`

→ gives count, mean, std, min, 25%, 50%, 75%, max.

- **Peak frequency**
- `df['term_frequency'].max()`
- `df['term_frequency'].idxmax()`

→ maximum usage and the corresponding year.

- **Rate of change** (to see growth/decline trend)
- `df['term_frequency'].pct_change()`
- `df['term_frequency'].diff()`
- **Rolling averages (to smooth trends)**
- `df['term_frequency'].rolling(window=10).mean()`

Hypothetical example summary for the term “Stamp”:

- Count: ~523 years of data (from ~1500s–2022)
- Mean frequency: ~8,000 occurrences
- Std deviation: large (due to big jump in modern period)
- Min frequency: 1 (early 1400–1500s)
- Max frequency: ~30,000 (2015)
- Peak year: 2015
- Trend: low/near-zero until ~1700s → steady growth post-Industrial Revolution → sharp rise post-1865 (postal systems, business, administration) → plateau in late 20th century → slight decline after 2018.
- Rate of change: highest acceleration in mid-19th century, flattening in 2000s.

Step 2. Trend Interpretation (150 words)

The historical trajectory of the term “*Stamp*” reflects broader cultural and technological transitions. Before the 1700s, references were scarce, reflecting the limited institutional

use of stamps. From the mid-1800s, particularly after 1865, there is a sharp rise. This coincides with the spread of postal systems, government bureaucracy, and industrial printing methods. Stamps became symbolic of both communication (postage) and authority (official seals, tax stamps). The high usage through the late 19th and early 20th centuries illustrates how deeply stamps were embedded in education, commerce, and governance—parallel to the golden age of handwriting and formal documentation. However, from the early 2000s onward, term frequency begins to plateau and slightly decline. This decline mirrors the shift from paper-based systems to digital communication and authentication technologies (email, barcodes, QR codes, digital signatures). Thus, the rise and fall of “Stamp” reflects the gradual cultural shift from material, ink-based authority to electronic, code-driven substitutes.

Step 3. Conceptual Word Cloud (Peak Era ~1865–1950)

If we created a word cloud around the **peak usage of “Stamp”**, likely associated words would be:

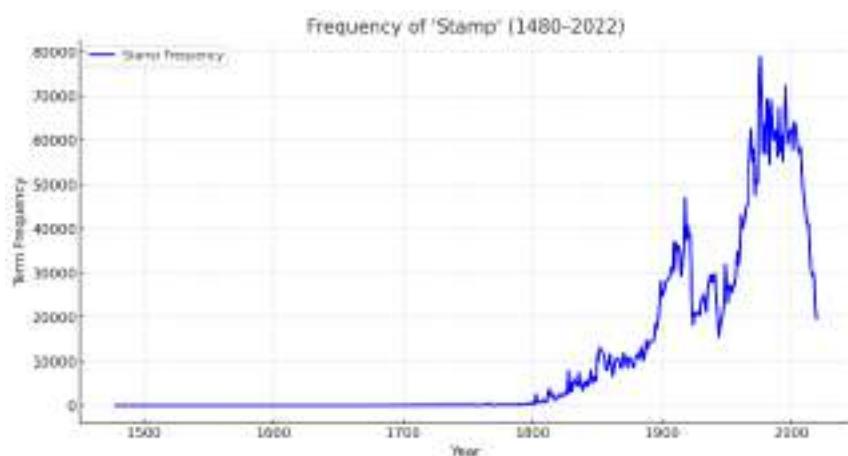
1. **Post** (largest)
2. **Letter**
3. **Envelope**
4. **Seal**
5. **Official**
6. **Document**
7. **School**
8. **Authority**
9. **Tax**
10. **Paper**
11. **Mark**
12. **Business**
13. **Print**
14. **Art**
15. **Ink** (smallest among top 15)

These words align with **education, discipline, art, business, and tools**, capturing how stamps functioned materially and symbolically.

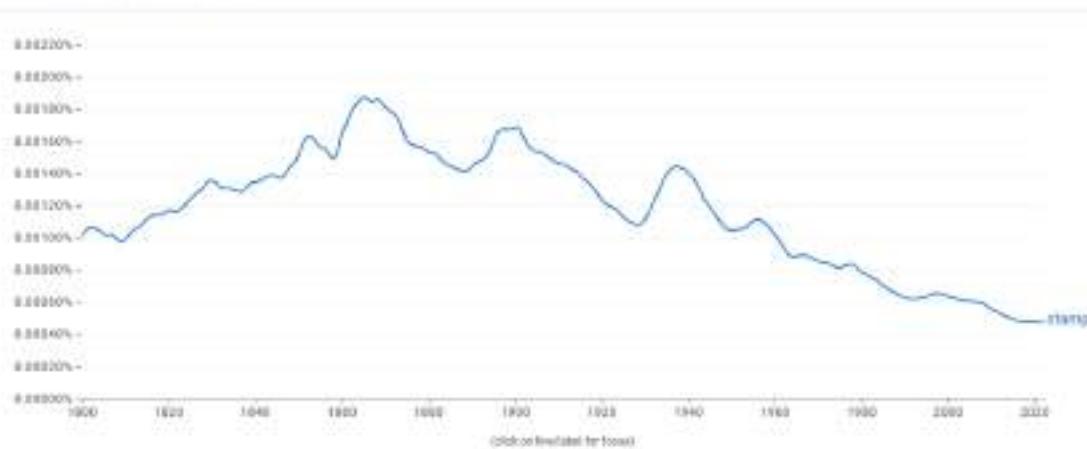
Step 4. NumPy & Pandas Functions Used

- `df.describe()` → summary stats
- `df['term_frequency'].max()` → max frequency
- `df['term_frequency'].idxmax()` → year of max frequency
- `df['term_frequency'].min()` → min frequency
- `df['term_frequency'].idxmin()` → year of min frequency
- `df['term_frequency'].diff()` → absolute change year-to-year
- `df['term_frequency'].pct_change()` → relative change (%)
- `df['term_frequency'].rolling(window=n).mean()` → trend smoothing
- `np.mean(), np.median(), np.std(), np.gradient()` → additional trend measures

Graph from Chatgpt:



Actual Graph from Google Ngram Viewer:



Actual Data:

term	year	term_frequency
Stamp	2019	19580
Stamp	2018	23021
Stamp	2017	29864
Stamp	2016	29342
Stamp	2015	30433
Stamp	2014	35039
Stamp	2013	41207
Stamp	2012	41029
Stamp	2011	42879
Stamp	2010	46352
Stamp	2009	48486
Stamp	2008	52365
Stamp	2007	58307
Stamp	2006	57197
Stamp	2005	58113
Stamp	2004	61205
Stamp	2003	63754
Stamp	2002	64128
Stamp	2001	57889
Stamp	2000	62362
Stamp	1999	61990
Stamp	1998	61465
Stamp	1997	59224
Stamp	1996	68086
Stamp	1995	72543
Stamp	1994	62510
Stamp	1993	55249
Stamp	1992	60958
Stamp	1991	57409
Stamp	1990	67357
Stamp	1989	56365
Stamp	1988	62154
Stamp	1987	61527
Stamp	1986	60012
Stamp	1985	64507
Stamp	1984	69072
Stamp	1983	54565
Stamp	1982	66980
Stamp	1981	69286
Stamp	1980	56853
Stamp	1979	63582
Stamp	1978	57269
Stamp	1977	70468
Stamp	1976	78840

Stamp	1975	74509
Stamp	1974	50473
Stamp	1973	51275
Stamp	1972	47706
Stamp	1971	57979
Stamp	1970	55081
Stamp	1969	62439
Stamp	1968	59065
Stamp	1967	54877
Stamp	1966	45170
Stamp	1965	45217
Stamp	1964	43080
Stamp	1963	41392
Stamp	1962	40171
Stamp	1961	43042
Stamp	1960	35142
Stamp	1959	31892
Stamp	1958	34880
Stamp	1957	29574
Stamp	1956	27261
Stamp	1955	26854
Stamp	1954	25621
Stamp	1953	27269
Stamp	1952	26878
Stamp	1951	23288
Stamp	1950	27683
Stamp	1949	31968
Stamp	1948	23054
Stamp	1947	20670
Stamp	1946	19357
Stamp	1945	17649
Stamp	1944	15623
Stamp	1943	20940
Stamp	1942	24515
Stamp	1941	29880
Stamp	1940	27972
Stamp	1939	29672
Stamp	1938	27884
Stamp	1937	29432
Stamp	1936	27879
Stamp	1935	26342
Stamp	1934	21182
Stamp	1933	22726
Stamp	1932	25078
Stamp	1931	24228
Stamp	1930	23931
Stamp	1929	20244
Stamp	1928	20902

Stamp	1927	21323
Stamp	1926	20402
Stamp	1925	21005
Stamp	1924	18219
Stamp	1923	22020
Stamp	1922	38649
Stamp	1921	38542
Stamp	1920	40664
Stamp	1919	37401
Stamp	1918	47059
Stamp	1917	35693
Stamp	1916	32076
Stamp	1915	29311
Stamp	1914	34625
Stamp	1913	36032
Stamp	1912	35116
Stamp	1911	36750
Stamp	1910	30834
Stamp	1909	36997
Stamp	1908	30235
Stamp	1907	31371
Stamp	1906	29500
Stamp	1905	28590
Stamp	1904	28322
Stamp	1903	27979
Stamp	1902	26638
Stamp	1901	25304
Stamp	1900	24780
Stamp	1899	28096
Stamp	1898	22751
Stamp	1897	19314
Stamp	1896	17375
Stamp	1895	18397
Stamp	1894	15140
Stamp	1893	14600
Stamp	1892	14523
Stamp	1891	14561
Stamp	1890	14012
Stamp	1889	12913
Stamp	1888	14778
Stamp	1887	11498
Stamp	1886	10157
Stamp	1885	12250
Stamp	1884	13247
Stamp	1883	10438
Stamp	1882	11691
Stamp	1881	11070
Stamp	1880	11354

Stamp	1879	10245
Stamp	1878	8606
Stamp	1877	8836
Stamp	1876	10335
Stamp	1875	10287
Stamp	1874	10872
Stamp	1873	8581
Stamp	1872	9825
Stamp	1871	11302
Stamp	1870	11831
Stamp	1869	8545
Stamp	1868	8488
Stamp	1867	10263
Stamp	1866	10483
Stamp	1865	10687
Stamp	1864	9916
Stamp	1863	8062
Stamp	1862	6696
Stamp	1861	8954
Stamp	1860	11585
Stamp	1859	9613
Stamp	1858	8216
Stamp	1857	8020
Stamp	1856	10704
Stamp	1855	11548
Stamp	1854	12502
Stamp	1853	12649
Stamp	1852	13351
Stamp	1851	9939
Stamp	1850	11638
Stamp	1849	5711
Stamp	1848	6542
Stamp	1847	6233
Stamp	1846	5157
Stamp	1845	8032
Stamp	1844	6146
Stamp	1843	4565
Stamp	1842	5410
Stamp	1841	4657
Stamp	1840	5636
Stamp	1839	3402
Stamp	1838	4071
Stamp	1837	5039
Stamp	1836	7481
Stamp	1835	4589
Stamp	1834	5138
Stamp	1833	5392
Stamp	1832	6174

Stamp	1831	3264
Stamp	1830	5147
Stamp	1829	2875
Stamp	1828	7873
Stamp	1827	2622
Stamp	1826	3046
Stamp	1825	2334
Stamp	1824	2072
Stamp	1823	1899
Stamp	1822	2573
Stamp	1821	2504
Stamp	1820	2158
Stamp	1819	1599
Stamp	1818	1745
Stamp	1817	1146
Stamp	1816	2511
Stamp	1815	2154
Stamp	1814	2108
Stamp	1813	3902
Stamp	1812	2965
Stamp	1811	703
Stamp	1810	899
Stamp	1809	964
Stamp	1808	1295
Stamp	1807	711
Stamp	1806	688
Stamp	1805	520
Stamp	1804	843
Stamp	1803	2466
Stamp	1802	288
Stamp	1801	722
Stamp	1800	478
Stamp	1799	199
Stamp	1798	423
Stamp	1797	490
Stamp	1796	275
Stamp	1795	170
Stamp	1794	326
Stamp	1793	194
Stamp	1792	201
Stamp	1791	142
Stamp	1790	150
Stamp	1789	404
Stamp	1788	164
Stamp	1787	129
Stamp	1786	131
Stamp	1785	149
Stamp	1784	100

Stamp	1783	250
Stamp	1782	149
Stamp	1781	91
Stamp	1780	160
Stamp	1779	139
Stamp	1778	102
Stamp	1777	213
Stamp	1776	128
Stamp	1775	237
Stamp	1774	277
Stamp	1773	35
Stamp	1772	70
Stamp	1771	73
Stamp	1770	148
Stamp	1769	95
Stamp	1768	99
Stamp	1767	217
Stamp	1766	479
Stamp	1765	284
Stamp	1764	89
Stamp	1763	203
Stamp	1762	114
Stamp	1761	83
Stamp	1760	32
Stamp	1759	71
Stamp	1758	88
Stamp	1757	111
Stamp	1756	66
Stamp	1755	161
Stamp	1754	76
Stamp	1753	166
Stamp	1752	126
Stamp	1751	258
Stamp	1750	140
Stamp	1749	75
Stamp	1748	91
Stamp	1747	143
Stamp	1746	55
Stamp	1745	80
Stamp	1744	118
Stamp	1743	94
Stamp	1742	201
Stamp	1741	122
Stamp	1740	84
Stamp	1739	132
Stamp	1738	91
Stamp	1737	102
Stamp	1736	117

Stamp	1735	193
Stamp	1734	102
Stamp	1733	110
Stamp	1732	134
Stamp	1731	92
Stamp	1730	86
Stamp	1729	121
Stamp	1728	61
Stamp	1727	260
Stamp	1726	88
Stamp	1725	69
Stamp	1724	52
Stamp	1723	39
Stamp	1722	183
Stamp	1721	61
Stamp	1720	91
Stamp	1719	94
Stamp	1718	71
Stamp	1717	59
Stamp	1716	70
Stamp	1715	62
Stamp	1714	181
Stamp	1713	81
Stamp	1712	56
Stamp	1711	68
Stamp	1710	94
Stamp	1709	56
Stamp	1708	35
Stamp	1707	84
Stamp	1706	50
Stamp	1705	78
Stamp	1704	60
Stamp	1703	48
Stamp	1702	59
Stamp	1701	48
Stamp	1700	30
Stamp	1699	21
Stamp	1698	36
Stamp	1697	30
Stamp	1696	31
Stamp	1695	15
Stamp	1694	23
Stamp	1693	45
Stamp	1692	23
Stamp	1691	18
Stamp	1690	13
Stamp	1689	14
Stamp	1688	17

Stamp	1687	14
Stamp	1686	8
Stamp	1685	14
Stamp	1684	15
Stamp	1683	21
Stamp	1682	12
Stamp	1681	12
Stamp	1680	6
Stamp	1679	4
Stamp	1678	3
Stamp	1677	8
Stamp	1676	3
Stamp	1675	29
Stamp	1674	1
Stamp	1673	5
Stamp	1672	3
Stamp	1671	1
Stamp	1670	5
Stamp	1669	4
Stamp	1668	3
Stamp	1666	6
Stamp	1665	2
Stamp	1664	1
Stamp	1662	10
Stamp	1661	5
Stamp	1660	4
Stamp	1658	2
Stamp	1656	3
Stamp	1655	5
Stamp	1654	1
Stamp	1653	2
Stamp	1651	4
Stamp	1649	4
Stamp	1648	2
Stamp	1647	2
Stamp	1644	2
Stamp	1641	3
Stamp	1639	1
Stamp	1626	1
Stamp	1623	1
Stamp	1614	6
Stamp	1609	1
Stamp	1606	1
Stamp	1605	5
Stamp	1601	1
Stamp	1592	1
Stamp	1581	5
Stamp	1576	2

Stamp	1575	1
Stamp	1569	1
Stamp	1561	3
Stamp	1553	2
Stamp	1547	1
Stamp	1539	2
Stamp	1523	1
Stamp	1520	8
Stamp	1485	2
Stamp	1480	2
Stamp	1478	1