

Introduzione

Rolling Stone è una delle riviste musicali e culturali più influenti al mondo, fondata nel 1967 a San Francisco da Jann Wenner e dal critico musicale Ralph J. Gleason. Sin dai suoi inizi, la rivista si è distinta per la sua copertura approfondita della musica, in un primo momento concentrandosi solo sul rock e poi ampliando il proprio sguardo su tutti gli altri generi musicali. Una delle iniziative più celebri della rivista è la pubblicazione periodica di classifiche, tra cui quella de "Le 500 migliori canzoni di tutti i tempi". Queste classifiche non solo riflettono il gusto e le tendenze musicali dell'epoca, ma influenzano anche il modo in cui intere generazioni percepiscono la musica. Ogni edizione di questa classifica è frutto di un processo di selezione che coinvolge critici, musicisti e addetti ai lavori.

Questo progetto si propone di esplorare e confrontare due classifiche de "Le 500 migliori canzoni di tutti i tempi", pubblicate nel 2004 e nel 2021. L'analisi iniziale si concentra su aspetti generali delle due classifiche, come le decadi più rappresentate e gli artisti con il maggior numero di presenze. Questo confronto permette di osservare come le tendenze musicali e i gusti siano cambiati nel tempo.

Successivamente proviamo a rispondere ad una domanda che rappresenta il punto centrale del progetto: "Esiste una correlazione tra i testi delle canzoni e la loro posizione in classifica?". Attraverso l'analisi dei testi delle canzoni e la loro relazione con il ranking, cercheremo di capire se esistono pattern o caratteristiche comuni tra i testi delle canzoni che si trovano ai vertici della classifica rispetto a quelle posizionate più in basso. Questa indagine può essere utile per valutare se coloro che creano questa lista sono condizionati nella valutazione complessiva di una canzone.

1. Importazione dati

In questa sezione, importiamo le librerie necessarie e carichiamo i dataset. Utilizziamo la libreria `pandas` per la manipolazione dei dati. I dataset che stiamo caricando contengono le classifiche delle "500 migliori canzoni di tutti i tempi" pubblicate dalla rivista **Rolling Stone** in due anni diversi: 2004 e 2021. Analizzeremo questi dati per vedere come sono cambiate le classifiche nel tempo.

Questi file li possiamo trovare su kaggle al seguente link:

<https://www.kaggle.com/datasets/bernardinosassoli/rolling-stones-top-500-songs-of-all-time?select=rollingstone.csv>

```
In [1]: import pandas as pd
```

```
In [2]: df_2004 = pd.read_csv('rollingstone_2004.csv')
df_2021 = pd.read_csv('rollingstone_2021.csv')
```

```
In [3]: # ribaltiamo le righe (per avere la prima riga con l'ultimo indice)
df_2004 = df_2004.iloc[::-1]
df_2021 = df_2021.iloc[::-1]

# ribaltiamo la colonna 'ranking' (per avere la prima riga con il primo posto)
```

```
df_2004['Ranking'] = df_2004['Ranking'].iloc[::-1].values
df_2021['Ranking'] = df_2021['Ranking'].iloc[::-1].values

# aggiungiamo 1 a ogni valore di 'ranking' (per avere la classifica partente da 1)
df_2004['Ranking'] += 1
df_2021['Ranking'] += 1

# resettiamo l'indice
df_2004.reset_index(drop=True, inplace=True)
df_2021.reset_index(drop=True, inplace=True)
```

```
In [4]: pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

2. Pulizia dei dati

I file csv che abbiamo scaricato sono stati creati per un altro scopo, motivo per cui sarà necessario eliminare tutte quelle informazioni che non sono necessarie. Le operazioni che andremo ad effettuare saranno: Rimozione colonne superflue, Modifica valori e ordine colonne al fine di migliorare la leggibilità del dataframe.

```
In [5]: df_2004.head(5)
```

```
Out[5]:
```

	Ranking	Artist	Title	Writers	Producer	Year	Spotify id	Popularity
0	1	Bob Dylan	Like a Rolling Stone	Dylan	Tom Wilson	65	3AhXZa8sUQht0UEdBJgpGc	74
1	2	The Rolling Stones	(I Can't Get No) Satisfaction	Mick Jagger, Keith Richards	Andrew Loog Oldham	65	2PzU4IB8Dr6mxV3IHuaG34	79
2	3	John Lennon	Imagine	John Lennon	Lennon, Phil Spector, Yoko Ono	71	7pKfPomDEel4TPT6EOYjn9	78
3	4	Marvin Gaye	What's Going On	Gaye, Renaldo Benson, Al Cleveland	Gaye	71	4RNHbYuRaZesMODlihsUG	63
4	5	Aretha Franklin	Respect	Otis Redding	Jerry Wexler	67	7s25THrKz86DM225dOYwnr	74

```
In [6]: df_2004.drop(columns=['Writers', 'Producer', 'Spotify id', 'Popularity', 'danceability',
df_2021.drop(columns=['Writers', 'Spotify id', 'Popularity', 'danceability', 'energy', 'key'])
```

```
In [7]: df_2004['Year'] = df_2004['Year'].apply(lambda y: '200' + str(y) if str(y) in ['0', '1'] else str(y))
```

```
In [8]: # Cambiamo l'ordine delle colonne
df_2004 = df_2004[['Title', 'Artist', 'Year', 'Ranking']]
df_2021 = df_2021[['Title', 'Artist', 'Year', 'Ranking']]
```

```
In [9]: # Rendiamo le colonne ranking e anno interi
df_2004['Ranking'] = df_2004['Ranking'].astype(int)
```

```
df_2004['Year'] = df_2004['Year'].astype(int)

df_2021['Ranking'] = df_2021['Ranking'].astype(int)
df_2021['Year'] = df_2021['Year'].astype(int)
```

3. Analisi iniziale

In questa prima fase andremo ad esaminare alcuni aspetti dai dataframe, nel dettaglio andremo a controllare:

- Gli artisti più presenti nelle singole classifiche
- Gli artisti presenti in tutte e due le classifiche
- Il numero di canzoni per decennio nelle singole classifiche

```
In [10]: # Artisti piu presenti nelle singole classifiche
artist_2004 = df_2004['Artist'].value_counts()
artist_2021 = df_2021['Artist'].value_counts()

print('10 artisti piu presenti nella classifica del 2004')
print(artist_2004.head(10))

print()

print('10 artisti piu presenti nella classifica del 2021')
print(artist_2021.head(10))
```

10 artisti piu presenti nella classifica del 2004

Artist	
The Beatles	23
The Rolling Stones	14
Bob Dylan	13
Elvis Presley	11
U2	8
The Jimi Hendrix Experience	7
The Beach Boys	7
Sly and the Family Stone	6
Led Zeppelin	6
Chuck Berry	6

Name: count, dtype: int64

10 artisti piu presenti nella classifica del 2021

Artist	
The Beatles	12
Bob Dylan	7
The Rolling Stones	7
David Bowie	6
Bruce Springsteen	5
Joni Mitchell	4
Elton John	4
Aretha Franklin	4
Marvin Gaye	4
Prince	4

Name: count, dtype: int64

Possiamo notare già dal primo confronto come nel corso di 17 anni la situazione sia cambiata. Nel 2004 i primi 4 artisti (The Beatles, Bob Dylan, The Rolling Stones, Elvis Presley) hanno in classifica un totale di 60 canzoni (ovvero il 12%) e ognuno di questi ha più di 10 canzoni in classifica. Nel 2021 solo un artista ha più di 10 canzoni in classifica (The Beatles). Le prime 3 posizioni sono identiche alla classifica precedente tuttavia i numeri sono dimezzati. Per quanto

riguarda le altre posizioni queste cambiano totalmente da una classifica ad un'altra con il caso particolare di Elvis Presley che non rientra neanche nella classifica.

```
In [11]: artists_2004 = df_2004['Artist']
artists_2021 = df_2021['Artist']
# Artisti nella classifica del 2004
print('Artisti nella classifica del 2004: ' + str(len(artists_2004.unique())))
# Artisti nella classifica del 2021
print('Artisti nella classifica del 2021: ' + str(len(artists_2021.unique())))

common_artists = set(artists_2004).intersection(set(artists_2021))
print("Numero di artisti presenti in entrambe le classifiche: " + str(len(common_artists)))

print("Artisti presenti in entrambe le classifiche:\n")
for artist in common_artists:
    print(artist)
```

Artisti nella classifica del 2004: 255
Artisti nella classifica del 2021: 362
Numero di artisti presenti in entrambe le classifiche: 179
Artisti presenti in entrambe le classifiche:

John Lee Hooker
Black Sabbath
The Kingsmen
The Flamingos
Jerry Lee Lewis
Patsy Cline
Johnny Cash
Kelly Clarkson
The Doors
New York Dolls
Eddie Cochran
Derek and the Dominos
Eric B. and Rakim
Donna Summer
Blue Oyster Cult
The Four Tops
Amy Winehouse
James Taylor
The Isley Brothers
The Band
AC/DC
Otis Redding
Dolly Parton
50 Cent
Prince
Led Zeppelin
Jimmy Cliff
The Eagles
The Shangri-Las
Coldplay
Them
The Cure
The Drifters
Don Henley
Run-DMC
Pavement
Metallica
The Righteous Brothers
Jeff Buckley
The Mamas and the Papas
The Velvet Underground
Roberta Flack
Afrika Bambaataa and Soulsonic Force
Earth Wind and Fire
Bobbie Gentry
Rick James
Joy Division
Billy Joel
Crosby Stills and Nash
The Staple Singers
The Animals
Rod Stewart
Weezer
The Ronettes
The Smiths
The Jimi Hendrix Experience
John Lennon
Pixies
Chic

Gnarls Barkley
Aretha Franklin
Gloria Gaynor
Van Morrison
The Stooges
Kanye West
Dr. Dre
Bonnie Raitt
Muddy Waters
The Beach Boys
The Temptations
Sex Pistols
Prince and the Revolution
Beastie Boys
Chuck Berry
The Kinks
The Strokes
The Everly Brothers
Nirvana
Dusty Springfield
Green Day
The Shirelles
Stevie Wonder
Wilson Pickett
The Allman Brothers Band
U2
Missy Elliott
The Rolling Stones
Ray Charles
Depeche Mode
Tracy Chapman
Grandmaster Flash and the Furious Five
Bo Diddley
The Notorious B.I.G.
Simon and Garfunkel
Procol Harum
Tina Turner
Eminem
Blondie
Smokey Robinson and the Miracles
Public Enemy
Sam Cooke
Glen Campbell
New Order
Radiohead
The Who
Aerosmith
The Zombies
Dionne Warwick
Ike and Tina Turner
Queen
Guns N Roses
Parliament
Neil Young
The Modern Lovers
The Impressions
Michael Jackson
Ramones
Hank Williams
Fleetwood Mac
The Police
Cheap Trick
Alice Cooper
N.W.A

Lou Reed
M.I.A.
Bill Withers
Joni Mitchell
R.E.M.
The Clash
Daft Punk
Al Green
Bruce Springsteen
The Five Satins
The B-52's
The Supremes
Eurythmics
Bob Marley and the Wailers
Fats Domino
Yeah Yeah Yeahs
Iggy Pop
Elvis Presley
Willie Nelson
Elton John
Gladys Knight and the Pips
Bob Dylan
Elvis Costello
Sinead O'Connor
Sly and the Family Stone
Jefferson Airplane
Randy Newman
Martha and the Vandellas
Beyonce
Bee Gees
Beck
Little Richard
2Pac
The Byrds
George Jones
Creedence Clearwater Revival
Television
Roy Orbison
Ben E. King
Tom Petty
Marvin Gaye
Pink Floyd
Mott the Hoople
Lynyrd Skynyrd
The Beatles
Martha Reeves and the Vandellas
Big Star
The Crystals
James Brown
Toots and the Maytals
Buddy Holly
Rihanna
The White Stripes
The Jackson 5
Madonna
David Bowie

Le righe di codice scritte sopra ci mostrano che la classifica del 2021 presenta molti più artisti rispetto a quella del 2004. Inoltre il numero di artisti presenti in entrambe le classifiche è 167 cioè il 45% della classifica del 2021 e addirittura il 64% della classifica del 2004. Questo cambiamento mostra come la musica nel corso degli anni non dipende più da pochi artisti,

tuttavia questa conclusione richiederebbe ulteriori studi, sarebbero necessario raccogliere altre classifiche dei prossimi anni e verificare se tutto ciò continua o se sia una caso limitato.

In [12]: df_2004

Out[12]:

	Title	Artist	Year	Ranking
0	Like a Rolling Stone	Bob Dylan	1965	1
1	(I Can't Get No) Satisfaction	The Rolling Stones	1965	2
2	Imagine	John Lennon	1971	3
3	What's Going On	Marvin Gaye	1971	4
4	Respect	Aretha Franklin	1967	5
5	Good Vibrations	The Beach Boys	1966	6
6	Johnny B. Goode	Chuck Berry	1958	7
7	Hey Jude	The Beatles	1968	8
8	Smells Like Teen Spirit	Nirvana	1991	9
9	What'd I Say	Ray Charles	1959	10
10	My Generation	The Who	1965	11
11	A Change Is Gonna Come	Sam Cooke	1964	12
12	Yesterday	The Beatles	1965	13
13	Blowin' in the Wind	Bob Dylan	1963	14
14	London Calling	The Clash	1980	15
15	I Want to Hold Your Hand	The Beatles	1963	16
16	Purple Haze	The Jimi Hendrix Experience	1967	17
17	Maybellene	Chuck Berry	1955	18
18	Hound Dog	Elvis Presley	1956	19
19	Let It Be	The Beatles	1970	20
20	Born to Run	Bruce Springsteen	1975	21
21	Be My Baby	The Ronettes	1963	22
22	In My Life	The Beatles	1965	23
23	People Get Ready	The Impressions	1965	24
24	God Only Knows	The Beach Boys	1966	25
25	(Sittin' on) the Dock of the Bay	Otis Redding	1968	26
26	Layla	Derek and the Dominos	1970	27
27	A Day in the Life	The Beatles	1967	28
28	Help!	The Beatles	1965	29
29	I Walk the Line	Johnny Cash	1956	30
30	Stairway to Heaven	Led Zeppelin	1971	31
31	Sympathy for the Devil	The Rolling Stones	1968	32
32	River Deep-Mountain High	Ike and Tina Turner	1966	33
33	You've Lost That Lovin' Feelin'	The Righteous Brothers	1964	34
34	Light My Fire	The Doors	1967	35
35	One	U2	1991	36
36	No Woman, No Cry	Bob Marley	1975	37

	Title		Artist	Year	Ranking
37	Gimme Shelter		The Rolling Stones	1969	38
38	That'll Be the Day		Buddy Holly	1957	39
39	Dancing in the Street	Martha Reeves and the Vandellas		1965	40
40	The Weight		The Band	1968	41
41	Waterloo Sunset		The Kinks	1968	42
42	Tutti-Frutti		Little Richard	1955	43
43	Georgia on My Mind		Ray Charles	1960	44
44	Heartbreak Hotel		Elvis Presley	1956	45
45	Heroes		David Bowie	1977	46
46	All Along the Watchtower	The Jimi Hendrix Experience		1968	47
47	Bridge Over Troubled Water	Simon and Garfunkel		1970	48
48	Hotel California		The Eagles	1976	49
49	The Tracks of My Tears	Smokey Robinson and the Miracles		1965	50
50	The Message	Grandmaster Flash and the Furious Five		1982	51
51	When Doves Cry		Prince	1984	52
52	When a Man Loves a Woman		Percy Sledge	1966	53
53	Louie Louie		The Kingsmen	1963	54
54	Long Tall Sally		Little Richard	1956	55
55	Anarchy in the U.K.		Sex Pistols	1977	56
56	A Whiter Shade of Pale		Procol Harum	1967	57
57	Billie Jean		Michael Jackson	1983	58
58	The Times They Are A-Changin'		Bob Dylan	1964	59
59	Let's Stay Together		Al Green	1971	60
60	Whole Lotta Shakin' Goin On		Jerry Lee Lewis	1957	61
61	Bo Diddley		Bo Diddley	1955	62
62	For What It's Worth		Buffalo Springfield	1967	63
63	She Loves You		The Beatles	1963	64
64	Sunshine of Your Love		Cream	1968	65
65	Redemption Song	Bob Marley and the Wailers		1980	66
66	Jailhouse Rock		Elvis Presley	1957	67
67	Tangled Up in Blue		Bob Dylan	1975	68
68	Crying		Roy Orbison	1961	69
69	Walk On By		Dionne Warwick	1964	70
70	Papa's Got A Brand New Bag		James Brown	1966	71
71	California Girls		The Beach Boys	1965	72
72	Superstition		Stevie Wonder	1972	73
73	Summertime Blues		Eddie Cochran	1958	74

	Title	Artist	Year	Ranking
74	Whole Lotta Love	Led Zeppelin	1969	75
75	Strawberry Fields Forever	The Beatles	1967	76
76	Mystery Train	Elvis Presley	1955	77
77	I Got You (I Feel Good)	James Brown	1965	78
78	Mr. Tambourine Man	The Byrds	1965	79
79	You Really Got Me	The Kinks	1964	80
80	I Heard It Through the Grapevine	Marvin Gaye	1968	81
81	Blueberry Hill	Fats Domino	1956	82
82	Norwegian Wood (This Bird Has Flown)	The Beatles	1965	83
83	Every Breath You Take	The Police	1983	84
84	Crazy	Patsy Cline	1961	85
85	Thunder Road	Bruce Springsteen	1975	86
86	Ring of Fire	Johnny Cash	1963	87
87	My Girl	The Temptations	1965	88
88	California Dreamin'	The Mamas and the Papas	1965	89
89	In the Still of the Night	The Five Satins	1956	90
90	Suspicious Minds	Elvis Presley	1969	91
91	Blitzkrieg Bop	Ramones	1976	92
92	I Still Haven't Found What I'm Looking For	U2	1987	93
93	Good Golly, Miss Molly	Little Richard	1958	94
94	Blue Suede Shoes	Carl Perkins	1956	95
95	Great Balls of Fire	Jerry Lee Lewis	1957	96
96	Roll Over Beethoven	Chuck Berry	1956	97
97	Love and Happiness	Al Green	1972	98
98	Fortunate Son	Creedence Clearwater Revival	1969	99
99	Crazy	Gnarls Barkley	2004	100
100	You Can't Always Get What You Want	The Rolling Stones	1969	101
101	Voodoo Child (Slight Return)	The Jimi Hendrix Experience	1968	102
102	Be-Bop-A-Lula	Gene Vincent	1956	103
103	Hot Stuff	Donna Summer	1979	104
104	Living for the City	Stevie Wonder	1973	105
105	The Boxer	Simon and Garfunkel	1969	106
106	Mr. Tambourine Man	Bob Dylan	1965	107
107	Not Fade Away	Buddy Holly	1957	108
108	Little Red Corvette	Prince	1983	109
109	Brown Eyed Girl	Van Morrison	1967	110
110	I've Been Loving You too Long (to Stop Now)	Otis Redding	1965	111

	Title	Artist	Year	Ranking
111	I'm So Lonesome I Could Cry	Hank Williams	1949	112
112	That's All Right	Elvis Presley	1954	113
113	Up on the Roof	The Drifters	1962	114
114	You Send Me	Sam Cooke	1957	115
115	Honky Tonk Women	The Rolling Stones	1969	116
116	Take Me to the River	Al Green	1974	117
117	Crazy in Love	Beyonce	2003	118
118	Shout (Pt 1 and 2)	The Isley Brothers	1959	119
119	Go Your Own Way	Fleetwood Mac	1977	120
120	I Want You Back	The Jackson 5	1969	121
121	Stand By Me	Ben E. King	1961	122
122	The House of the Rising Sun	The Animals	1964	123
123	It's A Man's Man's Man's World	James Brown	1966	124
124	Jumpin' Jack Flash	The Rolling Stones	1968	125
125	Will You Still Love Me Tomorrow	The Shirelles	1960	126
126	Shake, Rattle and Roll	Big Joe Turner	1954	127
127	Changes	David Bowie	1971	128
128	Rock and Roll Music	Chuck Berry	1957	129
129	Born to Be Wild	Steppenwolf	1968	130
130	Maggie May	Rod Stewart	1971	131
131	With or Without You	U2	1987	132
132	Who Do You Love?	Bo Diddley	1957	133
133	Won't Get Fooled Again	The Who	1971	134
134	In the Midnight Hour	Wilson Pickett	1965	135
135	While My Guitar Gently Weeps	The Beatles	1968	136
136	Your Song	Elton John	1970	137
137	Eleanor Rigby	The Beatles	1966	138
138	Family Affair	Sly and the Family Stone	1971	139
139	I Saw Her Standing There	The Beatles	1964	140
140	Kashmir	Led Zeppelin	1975	141
141	All I Have to Do Is Dream	The Everly Brothers	1958	142
142	Please, Please, Please	James Brown and His Famous Flames	1956	143
143	Purple Rain	Prince and the Revolution	1984	144
144	I Wanna Be Sedated	Ramones	1978	145
145	Everyday People	Sly and the Family Stone	1968	146
146	Rock Lobster	The B-52's	1979	147
147	Me and Bobby McGee	Janis Joplin	1971	148

	Title	Artist	Year	Ranking
148	Lust for Life	Iggy Pop	1977	149
149	Cathy's Clown	The Everly Brothers	1960	150
150	Eight Miles High	The Byrds	1966	151
151	Earth Angel	The Penguins	1954	152
152	Foxy Lady	The Jimi Hendrix Experience	1965	153
153	A Hard Day's Night	The Beatles	1964	154
154	Rave On	Buddy Holly	1958	155
155	Proud Mary	Creedence Clearwater Revival	1969	156
156	The Sounds of Silence	Simon and Garfunkel	1965	157
157	I Only Have Eyes for You	The Flamingos	1959	158
158	Rock Around the Clock	Bill Haley and His Comets	1954	159
159	Moment of Surrender	U2	2009	160
160	I'm Waiting for the Man	The Velvet Underground	1967	161
161	Bring the Noise	Public Enemy	1988	162
162	Folsom Prison Blues	Johnny Cash	1956	163
163	I Can't Stop Loving You	Ray Charles	1962	164
164	Nothing Compares 2 U	Sinead O'Connor	1990	165
165	Bohemian Rhapsody	Queen	1975	166
166	Fast Car	Tracy Chapman	1988	167
167	Let's Get It On	Marvin Gaye	1973	168
168	Papa Was a Rollin' Stone	The Temptations	1972	169
169	Losing My Religion	R.E.M.	1991	170
170	Both Sides Now	Joni Mitchell	1969	171
171	99 Problems	Jay Z	2003	172
172	Dream On	Aerosmith	1973	173
173	Dancing Queen	Abba	1976	174
174	God Save the Queen	Sex Pistols	1977	175
175	Paint it Black	The Rolling Stones	1966	176
176	I Fought the Law	Bobby Fuller Four	1966	177
177	Don't Worry Baby	The Beach Boys	1964	178
178	Free Fallin'	Tom Petty	1989	179
179	September Gurls	Big Star	1974	180
180	Love Will Tear Us Apart	Joy Division	1980	181
181	Hey Ya!	OutKast	2003	182
182	Green Onions	Booker T. and the MG's	1962	183
183	Save the Last Dance for Me	The Drifters	1960	184
184	The Thrill is Gone	B.B. King	1969	185

	Title		Artist	Year	Ranking
185	Please Please Me		The Beatles	1964	186
186	Desolation Row		Bob Dylan	1965	187
187	Who'll Stop the Rain	Creedence Clearwater Revival		1970	188
188	I Never Loved a Man (the Way I Love You)		Aretha Franklin	1967	189
189	Back in Black		AC/DC	1980	190
190	Stayin' Alive		Bee Gees	1977	191
191	Knockin' on Heaven's Door		Bob Dylan	1973	192
192	Free Bird		Lynyrd Skynyrd	1973	193
193	Rehab		Amy Winehouse	2007	194
194	Wichita Lineman		Glen Campbell	1968	195
195	There Goes My Baby		The Drifters	1959	196
196	Peggy Sue		Buddy Holly	1957	197
197	Sweet Child O' Mine		Guns n Roses	1987	198
198	Maybe		The Chantels	1957	199
199	Don't Be Cruel		Elvis Presley	1956	200
200	Hey Joe	The Jimi Hendrix Experience		1966	201
201	Flash Light		Parliament	1977	202
202	Loser		Beck	1993	203
203	Bizarre Love Triangle		New Order	1986	204
204	Come Together		The Beatles	1969	205
205	Positively 4th Street		Bob Dylan	1965	206
206	Try a Little Tenderness		Otis Redding	1966	207
207	Lean on Me		Bill Withers	1972	208
208	Reach Out, I'll Be There		The Four Tops	1966	209
209	Bye Bye Love		The Everly Brothers	1957	210
210	Gloria		Them	1965	211
211	In My Room		The Beach Boys	1963	212
212	96 Tears		and the Mysterians	1966	213
213	Caroline, No		The Beach Boys	1966	214
214	1999		Prince	1982	215
215	Rockin' in the Free World		Neil Young	1989	216
216	Your Cheatin' Heart		Hank Williams	1953	217
217	Do You Believe in Magic		The Lovin Spoonful	1965	218
218	Jolene		Dolly Parton	1974	219
219	Boom Boom		John Lee Hooker	1962	220
220	Spoonful		Howlin Wolf	1960	221
221	Walk Away Renee		The Left Banke	1966	222

	Title	Artist	Year	Ranking
222	Walk on the Wild Side	Lou Reed	1972	223
223	Oh, Pretty Woman	Roy Orbison	1964	224
224	Dance to the Music	Sly and the Family Stone	1968	225
225	Hoochie Coochie Man	Muddy Waters	1954	226
226	Fire and Rain	James Taylor	1970	227
227	Should I Stay or Should I Go	The Clash	1982	228
228	Good Times	Chic	1979	229
229	Mannish Boy	Muddy Waters	1955	230
230	Moondance	Van Morrison	1970	231
231	Just Like a Woman	Bob Dylan	1966	232
232	Sexual Healing	Marvin Gaye	1982	233
233	Only the Lonely	Roy Orbison	1960	234
234	We Gotta Get Out of This Place	The Animals	1965	235
235	Paper Planes	M.I.A.	2007	236
236	I'll Feel a Whole Lot Better	The Byrds	1965	237
237	Everyday	Buddy Holly	1957	238
238	I've Got a Woman	Ray Charles	1954	239
239	Planet Rock	Afrika Bambaataa and Soulsonic Force	1982	240
240	I Fall to Pieces	Patsy Cline	1961	241
241	Son of a Preacher Man	Dusty Springfield	1968	242
242	The Wanderer	Dion	1961	243
243	Stand!	Sly and the Family Stone	1969	244
244	Rocket Man	Elton John	1972	245
245	Love Shack	The B-52's	1989	246
246	Gimme Some Lovin'	The Spencer Davis Group	1966	247
247	(Your Love Keeps Lifting Me) Higher and Higher	Jackie Wilson	1967	248
248	The Night They Drove Old Dixie Down	The Band	1969	249
249	Hot Fun in the Summertime	Sly and the Family Stone	1969	250
250	Rapper's Delight	Sugarhill Gang	1979	251
251	Chain of Fools	Aretha Franklin	1967	252
252	Paranoid	Black Sabbath	1970	253
253	Money Honey	The Drifters	1953	254
254	Mack the Knife	Bobby Darin	1959	255
255	All the Young Dudes	Mott the Hoople	1972	256
256	Paranoid Android	Radiohead	1997	257
257	Highway to Hell	AC/DC	1979	258

	Title	Artist	Year	Ranking
258	Heart of Glass	Blondie	1978	259
259	Mississippi	Bob Dylan	2001	260
260	Wild Thing	The Troggs	1966	261
261	I Can See For Miles	The Who	1967	262
262	Oh, What A Night	The Dells	1969	263
263	Hallelujah	Jeff Buckley	1994	264
264	Higher Ground	Stevie Wonder	1973	265
265	Ooo Baby Baby	Smokey Robinson and The Miracles	1965	266
266	He's A Rebel	The Crystals	1962	267
267	Sail Away	Randy Newman	1972	268
268	Walking in the Rain	The Ronettes	1964	269
269	Tighten Up	Archie Bell and The Drells	1968	270
270	Personality Crisis	New York Dolls	1973	271
271	Sunday Bloody Sunday	U2	1983	272
272	Jesus Walks	Kanye West	2004	273
273	Roadrunner	The Modern Lovers	1976	274
274	He Stopped Loving Her Today	George Jones	1980	275
275	Sloop John B	The Beach Boys	1966	276
276	Sweet Little Sixteen	Chuck Berry	1958	277
277	Something	The Beatles	1969	278
278	Somebody to Love	Jefferson Airplane	1967	279
279	Born in the U.S.A.	Bruce Springsteen	1984	280
280	I'll Take You There	The Staple Singers	1972	281
281	Ziggy Stardust	David Bowie	1972	282
282	Pictures of You	The Cure	1989	283
283	Chapel of Love	The Dixie Cups	1964	284
284	Ain't No Sunshine	Bill Withers	1971	285
285	Seven Nation Army	The White Stripes	2003	286
286	You Are the Sunshine of My Life	Stevie Wonder	1972	287
287	Help Me	Joni Mitchell	1974	288
288	Call Me	Blondie	1980	289
289	(What's So Funny bout) Peace, Love and Underst...	Elvis Costello	1979	290
290	Smokestack Lightning	Howlin Wolf	1956	291
291	Summer Babe (Winter Version)	Pavement	1992	292
292	Walk This Way	Run-DMC	1986	293
293	Money (That's What I Want)	Barrett Strong	1960	294
294	Can't Buy Me Love	The Beatles	1964	295

	Title		Artist	Year	Ranking
295	Marshall Mathers		Eminem	2000	296
296	She's Not There		The Zombies	1964	297
297	Train in Vain		The Clash	1979	298
298	Tired of Being Alone		Al Green	1971	299
299	Black Dog		Led Zeppelin	1971	300
300	Street Fighting Man		The Rolling Stones	1968	301
301	Get Up, Stand Up	Bob Marley and the Wailers		1975	302
302	Heart of Gold		Neil Young	1972	303
303	Sign 'O' The Times		Prince	1987	304
304	One Way or Another		Blondie	1978	305
305	Like A Prayer		Madonna	1989	306
306	One More Time		Daft Punk	2000	307
307	Da Ya Think I'm Sexy?		Rod Stewart	1978	308
308	Blue Eyes Crying in the Rain		Willie Nelson	1975	309
309	Ruby Tuesday		The Rolling Stones	1967	310
310	With a Little Help From My Friends		The Beatles	1967	311
311	Say It Loud - I'm Black and I'm Proud		James Brown	1968	312
312	That's Entertainment		The Jam	1980	313
313	Why Do Fools Fall In Love	Frankie Lymon and The Teenagers		1956	314
314	Lonely Teardrops		Jackie Wilson	1958	315
315	What's Love Got to Do With It		Tina Turner	1984	316
316	Iron Man		Black Sabbath	1971	317
317	Wake Up Little Susie		The Everly Brothers	1957	318
318	In Dreams		Roy Orbison	1963	319
319	I Put a Spell on You	Screamin Jay Hawkins		1956	320
320	Comfortably Numb		Pink Floyd	1979	321
321	Don't Let Me Be Misunderstood		The Animals	1965	322
322	Alison		Elvis Costello	1977	323
323	Wish You Were Here		Pink Floyd	1975	324
324	Many Rivers to Cross		Jimmy Cliff	1969	325
325	School's Out		Alice Cooper	1972	326
326	Take Me Out		Franz Ferdinand	2004	327
327	Heartbreaker		Led Zeppelin	1969	328
328	Cortez the Killer		Neil Young	1975	329
329	Fight the Power		Public Enemy	1989	330
330	Dancing Barefoot		Patti Smith Group	1979	331
331	Baby Love		The Supremes	1964	332

	Title	Artist	Year	Ranking
332	Good Lovin'	The Young Rascals	1966	333
333	Get Up (I Feel Like Being a) Sex Machine	James Brown	1970	334
334	For Your Precious Love	Jerry Butler and The Impressions	1958	335
335	The End	The Doors	1967	336
336	That's the Way of the World	Earth Wind and Fire	1975	337
337	We Will Rock You	Queen	1977	338
338	I Can't Make You Love Me	Bonnie Raitt	1991	339
339	Subterranean Homesick Blues	Bob Dylan	1965	340
340	Spirit in the Sky	Norman Greenbaum	1970	341
341	Sweet Jane	The Velvet Underground	1970	342
342	Wild Horses	The Rolling Stones	1971	343
343	Beat It	Michael Jackson	1982	344
344	Beautiful Day	U2	2000	345
345	Walk This Way	Aerosmith	1976	346
346	Maybe I'm Amazed	Paul McCartney	1970	347
347	You Keep Me Hanging On	The Supremes	1966	348
348	Baba O'Riley	The Who	1971	349
349	The Harder They Come	Jimmy Cliff	1975	350
350	Runaround Sue	Dion	1961	351
351	Jim Dandy	Lavern Baker	1956	352
352	Piece of My Heart	Big Brother and the Holding Company	1968	353
353	La Bamba	Ritchie Valens	1958	354
354	California Love	2Pac	1996	355
355	Candle in the Wind	Elton John	1973	356
356	That Lady (Pts 1 and 2)	The Isley Brothers	1973	357
357	Spanish Harlem	Ben E. King	1960	358
358	The Loco-Motion	Little Eva	1962	359
359	The Great Pretender	The Platters	1955	360
360	All Shook Up	Elvis Presley	1957	361
361	Tears in Heaven	Eric Clapton	1992	362
362	Watching the Detectives	Elvis Costello	1977	363
363	Bad Moon Rising	Creedence Clearwater Revival	1969	364
364	Sweet Dreams (Are Made of This)	Eurythmics	1983	365
365	Little Wing	The Jimi Hendrix Experience	1968	366
366	Nowhere to Run	Martha and the Vandellas	1965	367
367	Got My Mojo Working	Muddy Waters	1957	368
368	Killing Me Softly With His Song	Roberta Flack	1973	369

	Title	Artist	Year	Ranking
369	All You Need is Love	The Beatles	1967	370
370	Complete Control	The Clash	1979	371
371	The Letter	The Box Tops	1967	372
372	Highway 61 Revisited	Bob Dylan	1965	373
373	Unchained Melody	The Righteous Brothers	1965	374
374	How Deep Is Your Love	Bee Gees	1977	375
375	White Room	Cream	1968	376
376	Personal Jesus	Depeche Mode	1989	377
377	I'm a Man	Bo Diddley	1955	378
378	The Wind Cries Mary	The Jimi Hendrix Experience	1967	379
379	I Can't Explain	The Who	1965	380
380	Marquee Moon	Television	1977	381
381	Wonderful World	Sam Cooke	1960	382
382	Brown Eyed Handsome Man	Chuck Berry	1956	383
383	Another Brick in the Wall Part 2	Pink Floyd	1979	384
384	Fake Plastic Trees	Radiohead	1995	385
385	Maps	Yeah Yeah Yeahs	2004	386
386	Hit the Road Jack	Ray Charles	1961	387
387	Pride (In the Name of Love)	U2	1984	388
388	Radio Free Europe	R.E.M.	1983	389
389	Goodbye Yellow Brick Road	Elton John	1973	390
390	Tell It Like It Is	Aaron Neville	1966	391
391	Bitter Sweet Symphony	The Verve	1997	392
392	Whipping Post	The Allman Brothers Band	1969	393
393	Ticket to Ride	The Beatles	1965	394
394	Ohio	Crosby Stills Nash and Young	1970	395
395	I Know You Got Soul	Eric B. and Rakim	1987	396
396	Tiny Dancer	Elton John	1971	397
397	Roxanne	The Police	1979	398
398	Just My Imagination	The Temptations	1971	399
399	Baby I Need Your Loving	The Four Tops	1964	400
400	Summer in the City	The Lovin Spoonful	1966	401
401	O-o-h Child	The Five Steps	1970	402
402	Can't Help Falling in Love	Elvis Presley	1961	403
403	Remember (Walkin' in the Sand)	The Shangri Las	1964	404
404	(Don't Fear) the Reaper	Blue Oyster Cult	1976	405
405	Thirteen	Big Star	1972	406

	Title	Artist	Year	Ranking
406	Sweet Home Alabama	Lynyrd Skynyrd	1974	407
407	Enter Sandman	Metallica	1991	408
408	Tonight's the Night	The Shirelles	1960	409
409	Thank You (Falettinme Be Mice Elf Agin)	Sly and the Family Stone	1970	410
410	C'mon Everybody	Eddie Cochran	1958	411
411	Umbrella	Rihanna	2007	412
412	Visions of Johanna	Bob Dylan	1966	413
413	We've Only Just Begun	Carpenters	1970	414
414	In Bloom	Nirvana	1991	415
415	Sweet Emotion	Aerosmith	1975	416
416	Monkey Gone to Heaven	Pixies	1989	417
417	I Feel Love	Donna Summer	1977	418
418	Ode to Billie Joe	Bobbie Gentry	1967	419
419	The Girl Can't Help It	Little Richard	1957	420
420	Youngblood	The Coasters	1957	421
421	I Can't Help Myself (Sugar Pie Honey Bunch)	The Four Tops	1965	422
422	The Boys of Summer	Don Henley	1984	423
423	Juicy	The Notorious B.I.G.	1994	424
424	Fuck tha Police	N.W.A	1989	425
425	Suite: Judy Blue Eyes	Crosby Stills and Nash	1969	426
426	Nuthin' But a 'G' Thang	Dr. Dre	1993	427
427	It's Your Thing	The Isley Brothers	1969	428
428	Piano Man	Billy Joel	1973	429
429	Blue Suede Shoes	Elvis Presley	1956	430
430	William, It Was Really Nothing	The Smiths	1984	431
431	American Idiot	Green Day	2004	432
432	Tumbling Dice	The Rolling Stones	1972	433
433	Smoke on the Water	Deep Purple	1973	434
434	New Year's Day	U2	1983	435
435	Everybody Needs Somebody to Love	Solomon Burke	1964	436
436	(White Man) In Hammersmith Palais	The Clash	1979	437
437	Ain't That a Shame	Fats Domino	1955	438
438	Midnight Train to Georgia	Gladys Knight and the Pips	1973	439
439	Ramble On	Led Zeppelin	1969	440
440	Mustang Sally	Wilson Pickett	1966	441
441	Alone Again Or	Love	1968	442
442	Beast of Burden	The Rolling Stones	1978	443

	Title	Artist	Year	Ranking
443	Love Me Tender	Elvis Presley	1956	444
444	I Wanna Be Your Dog	The Stooges	1969	445
445	n Pepa, 'Push It	Salt	1987	446
446	Pink Houses	John Mellencamp	1983	447
447	In Da Club	50 Cent	2002	448
448	Come Go With Me	The Del-Vikings	1957	449
449	I Shot the Sheriff	Bob Marley and the Wailers	1973	450
450	I Got You Babe	Sonny and Cher	1965	451
451	Come As You Are	Nirvana	1991	452
452	Pressure Drop	Toots and the Maytals	1973	453
453	Leader of the Pack	The Shangri-Las	1964	454
454	Heroin	The Velvet Underground	1967	455
455	Penny Lane	The Beatles	1967	456
456	The Twist	Chubby Checker	1960	457
457	Cupid	Sam Cooke	1961	458
458	Paradise City	Guns N Roses	1987	459
459	My Sweet Lord	George Harrison	1970	460
460	Sheena Is a Punk Rocker	Ramones	1977	461
461	All Apologies	Nirvana	1993	462
462	Soul Man	Sam and Dave	1967	463
463	Kiss	Prince and the Revolution	1986	464
464	Rollin' Stone	Muddy Waters	1948	465
465	Get Ur Freak On	Missy Elliott	2001	466
466	Big Pimpin'	Jay Z	1999	467
467	Respect Yourself	The Staple Singers	1971	468
468	Rain	The Beatles	1966	469
469	Standing in the Shadows of Love	The Four Tops	1966	470
470	Surrender	Cheap Trick	1978	471
471	Runaway	Del Shannon	1961	472
472	Welcome to the Jungle	Guns N Roses	1987	473
473	Into the Mystic	Van Morrison	1970	474
474	Where Did Our Love Go	The Supremes	1964	475
475	Do Right Woman-Do Right Man	Aretha Franklin	1967	476
476	How Soon Is Now?	The Smiths	1985	477
477	Last Nite	The Strokes	2001	478
478	I Want to Know What Love Is	Foreigner	1984	479
479	Sabotage	Beastie Boys	1994	480

	Title	Artist	Year	Ranking
480	Super Freak	Rick James	1981	481
481	Since U Been Gone	Kelly Clarkson	2004	482
482	White Rabbit	Jefferson Airplane	1967	483
483	Cry Me a River	Justin Timberlake	2002	484
484	Lady Marmalade	LaBelle	1975	485
485	Young Americans	David Bowie	1975	486
486	I'm Eighteen	Alice Cooper	1971	487
487	Just Like Heaven	The Cure	1987	488
488	Under the Boardwalk	The Drifters	1964	489
489	Clocks	Coldplay	2002	490
490	I Love Rock 'N Roll	Joan Jett and the Blackhearts	1982	491
491	I Will Survive	Gloria Gaynor	1978	492
492	Time to Pretend	MGMT	2008	493
493	Ignition (Remix)	R. Kelly	2002	494
494	Brown Sugar	The Rolling Stones	1971	495
495	Running on Empty	Jackson Browne	1977	496
496	The Rising	Bruce Springsteen	2002	497
497	Miss You	The Rolling Stones	1978	498
498	Buddy Holly	Weezer	1994	499
499	Shop Around	Smokey Robinson and the Miracles	1960	500

In [13]: *# Contiamo quante canzoni per decennio sono presenti nelle due classifiche*

```
decade_2004 = {}
for year in df_2004['Year']:
    decade = str(year)[0:3] + '0'
    if decade in decade_2004:
        decade_2004[decade] += 1
    else:
        decade_2004[decade] = 1

decade_2021 = {}
for year in df_2021['Year']:
    decade = str(year)[0:3] + '0'
    if decade in decade_2021:
        decade_2021[decade] += 1
    else:
        decade_2021[decade] = 1

# Ordiniamo i dizionari in base all'ordine temporale
decade_2004 = dict(sorted(decade_2004.items()))
decade_2021 = dict(sorted(decade_2021.items()))
```

In [14]: decade_2004

```
Out[14]: {'1940': 2,
          '1950': 68,
          '1960': 196,
          '1970': 131,
          '1980': 55,
          '1990': 22,
          '2000': 26}
```

```
In [15]: decade_2021
```

```
Out[15]: {'1930': 2,
          '1940': 1,
          '1950': 24,
          '1960': 108,
          '1970': 144,
          '1980': 80,
          '1990': 70,
          '2000': 36,
          '2010': 32,
          '2020': 3}
```

Quest'ultimo confronto conferma il trend già visto con i punti precedenti, ovvero andando avanti nel tempo la classifica risulta meno incentrata su pochi cantanti o anni, infatti se nella classifica del 2004 circa la metà delle canzoni sono concentrate negli anni 70 e 80, nella classifica del 2021, questi restano gli anni con più canzoni ma, la distanza con gli altri anni si riduce, in seguito ad un netto aumento delle canzoni degli anni successivi. Ricordiamo sempre che questa classifica è redatta da Rolling Stones, quindi l'analisi effettuata è utile per capire l'opinione di una testata giornalistica che rappresenta una figura importante nello scenario musicale, tuttavia non è comunque possibile dire se le canzoni scelte sono oggettivamente le migliori 500.

4. Scraping testi

Per la seconda parte della nostra analisi sarà necessario lavorare con i testi delle canzoni. Per ottenere questi testi dovremmo fare web scraping, è quindi necessario l'utilizzo di alcune librerie fondamentali come `requests` e `BeautifulSoups`. Request è una libreria che permette di ottenere pagine HTTP. BeautifulSoup riesce ad estrarre informazioni da pagine web. Infine abbiamo re, una libreria python che introduce le espressioni regolari, utili per pulire i testi.

Request: <https://requests.readthedocs.io/en/latest/>

BeautifulSoup: <https://www.crummy.com/software/BeautifulSoup/>

```
In [16]: import requests
         from bs4 import BeautifulSoup as soup
         import re
```

```
In [17]: def format_url(title, artist):
         # Rimuoviamo i caratteri speciali
         title = re.sub(r'[\(\)\[\]]', '', title)
         title = re.sub(r'-' , ' ', title)
         title = re.sub(r'+' , ' ', title)
         title = re.sub(r' ' , '-' , title)
         title = re.sub(r'"', '', title)
         title = re.sub(r',', '', title)
```

```

title = re.sub(r':', '', title)
title = re.sub(r'\.', '', title)
title = re.sub(r'\?', '', title)
title = re.sub(r'!', '', title)
title = re.sub(r'/', '-', title)

artist = re.sub(r'[\(\)\[\]]', '', artist)
artist = re.sub(r' - ', ' ', artist)
artist = re.sub(r'\+', ' ', artist)
artist = re.sub(r' + ', ' ', artist)
artist = re.sub(r' ', '-', artist)
artist = re.sub(r'"', '', artist)
artist = re.sub(r',', ' ', artist)
artist = re.sub(r'\.', ' ', artist)
artist = re.sub(r'/', '-', artist)

# elimina il testo "feat." e la parte successiva
artist = re.sub(r'feat.*', '', artist)

return title, artist

```

```

In [18]: def add_url(df):
# Copia del dataframe
    urls = []
    titles = df['Title']
    artists = df['Artist']

    for i in range(500):
        title = titles[i]
        artist = artists[i]

        title, artist = format_url(title, artist)

        url = 'https://www.genius.com/' + artist + '-' + title + '-lyrics'
        urls.append(url)

    df['URL'] = urls
    return df

```

Le funzioni presenti sopra lavorano insieme e hanno lo scopo di trovare gli URL relativi ad ogni canzone. La funzione `format_url` si occupa di riscrivere le informazioni relative all'artista e al titolo in un formato adeguato per un URL. La funzione `add_url` crea una nuova colonna con gli url relativi alle singole canzoni e per creare questi sfrutta la funzione precedente e poi genera l'url corretto.

```

In [19]: df_2004 = add_url(df_2004)
df_2021 = add_url(df_2021)

```

```

In [20]: def get_clear_text(text_raw):
    full_text = ""

    for t in text_raw:
        full_text += " " + t.text.strip()

    full_text = re.sub(r"[\.\*\?]", "", full_text) #rimuove il testo tra le parentesi
    full_text = re.sub(r"\([^\)]*\)", "", full_text) #rimuove il testo tra le parentesi
    full_text = re.sub(r"(\w)([A-Z])", r"\1 \2", full_text) #trova una lettera minuscola
    full_text = full_text.replace("'", " ").replace('"', " ")
    full_text = full_text.replace("!", "! ").replace("?", "? ") #aggiunge uno spazio
    full_text = full_text.lower() #trasforma tutto in minuscolo

```



```
return full_text
```

```
In [21]: def extract_lyrics(url):
page = requests.get(url)
page_soup = soup(page.content, 'html.parser')
text_raw = page_soup.find_all("div", {"class": "Lyrics__Container-sc-1ynbvzw-1"})
text = get_clear_text(text_raw)
if text == "":
    text = "ERRORE"
return text
```

Le due funzioni presenti sopra lavorano con i testi delle canzoni. `get_clear_text` si occupa di eliminare tutte le informazioni superflue dal testo (il sito infatti presenta parti di testo inutili, ad esempio "Strofa 1", "Rit.", queste non sono utili per il nostro obiettivo). La funzione `extract_lyrics` sfrutta l'url di una canzone per effettuare la richiesta web, trovare la pagina, prendere solo il testo, pulirlo tramite la funzione precedente e tornarlo in output.

```
In [22]: df_2004['Lyrics'] = df_2004['URL'].apply(extract_lyrics)
df_2021['Lyrics'] = df_2021['URL'].apply(extract_lyrics)

df_2004.to_csv('rollingstone_2004(lyrics).csv', index=False)
df_2021.to_csv('rollingstone_2021(lyrics).csv', index=False)
# Questo codice richiede molto tempo per eseguire, motivo per cui portiamo i dataframe
```

Per effettuare lo scraping dei testi è stato necessario non solo pulire i testi raccolti tramite web, rimuovendo tutte le informazioni inutili inserite per migliorare l'estetica del sito web, ma anche lavorare sugli artisti e i titoli delle canzoni, infatti molti di questi presentavano degli errori, dovuti probabilmente allo scraping effettuato da un sito diverso rispetto a quello che abbiamo utilizzato per trovare i testi, quindi è stato necessario cambiare, sia manualmente sia tramite codice, alcune informazioni presenti nei dataframe di partenza, per questo motivo abbiamo creato i due nuovi dataframe e li abbiamo esportati come file csv, in modo tale che chiunque voglia lavorare con questi possa farlo tranquillamente, risparmiando tempo.

Alcuni testi sono assenti e al loro posto è presente una keyword "ERRORE", questa è stata utilizzata in un primo momento per capire quali testi non si riuscivano ad ottenere tramite scraping, tuttavia una volta completato possiamo vedere questa keyword come un'indicazione che le canzoni contrassegnate non hanno un testo, cioè sono solo base musicale. Le canzoni in questione sono un numero molto basso che non va ad intralciare il nostro lavoro.

5. Natural Language Processing (NLP)

Una volta raccolte le informazioni e i testi relativi ad ogni canzone, eseguiamo alcuni passaggi della pipeline NLP. Per farlo sfruttiamo `nltk`, una libreria python utilizzata per lavorare con il linguaggio naturale.

nltk: <https://www.nltk.org/>

```
In [23]: df_2004 = pd.read_csv('rollingstone_2004(lyrics).csv')
df_2021 = pd.read_csv('rollingstone_2021(lyrics).csv')
```

```
In [24]: import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import string

nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

stop_words = set(stopwords.words('english'))
lemmatizer = WordNetLemmatizer()
```

```
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\alesc\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\alesc\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]   C:\Users\alesc\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
In [25]: all_words = {}

def get_most_common_words(df):
    for i in range(500):
        lyrics = df['Lyrics'][i]
        words = word_tokenize(lyrics)
        words = [w for w in words if not w in stop_words]
        words = [w for w in words if not w in string.punctuation]
        words = [lemmatizer.lemmatize(w) for w in words]
        for word in words:
            if word in all_words:
                all_words[word] += 1
            else:
                all_words[word] = 1
```

La funzione `get_most_common_words` esegue alcuni passaggi della pipeline NLP, ovvero tokenizzazione, rimozione delle stop words e lemmatizzazione.

La tokenizzazione è il processo di suddivisione di un testo in unità più piccole chiamate "token". La tokenizzazione è il primo passo in molte applicazioni di NLP.

Le "stop words" sono parole comuni che spesso vengono rimosse dai testi durante l'elaborazione del linguaggio naturale. Queste parole sono considerate non essenziali per il significato complessivo del testo e includono termini come articoli, preposizioni, pronomi e congiunzioni.

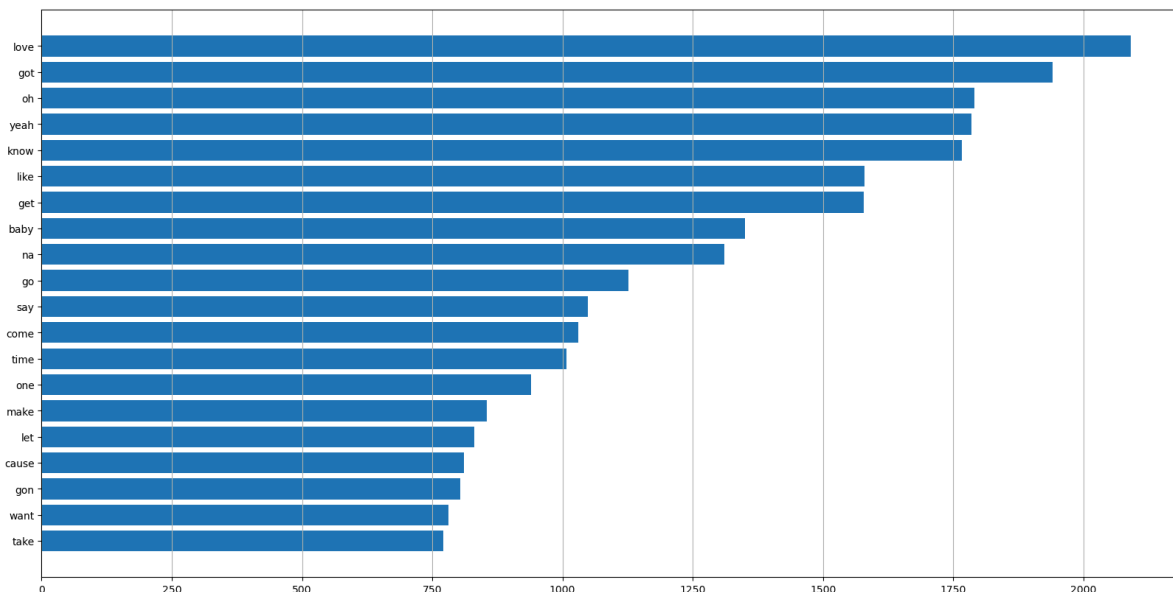
La lemmatizzazione è il processo di riduzione delle forme flesse di una parola alla sua forma base o lemma. La forma base è il modo in cui una parola appare nel dizionario. La lemmatizzazione consente di trattare le forme diverse di una parola come un'unica entità.

```
In [26]: get_most_common_words(df_2004)
get_most_common_words(df_2021)
```

```
In [36]: import matplotlib.pyplot as plt

sorted_words = dict(sorted(all_words.items(), key=lambda item: item[1], reverse=True))
```

```
plt.figure(figsize=(20, 10))
plt.barh(list(sorted_words.keys())[:20], list(sorted_words.values())[:20])
plt.gca().invert_yaxis()
plt.grid(axis='x')
plt.show()
```



Eseguiti i passaggi NLP, andiamo a plottare le 20 parole piu utilizzate(keywords), queste ci possono dare delle informazioni in generale sui testi delle canzoni.

6. Sentiment Analysis

La Sentiment Analysis è una tecnica di elaborazione del linguaggio naturale (NLP) utilizzata per determinare il tono emotivo dietro un testo.

Eseguiamo adesso il punto centrale del progetto, ovvero vogliamo verificare se esiste una correlazione tra la posizione nella classifica e il sentimento della canzone, quindi vogliamo capire per esempio se una canzone con un sentimento positivo tende ad occupare posizioni piu alte rispetto ad una canzone triste o viceversa.

```
In [37]: from nltk.sentiment.vader import SentimentIntensityAnalyzer
#nltk.download('vader_lexicon')

sentiment_analyzer = SentimentIntensityAnalyzer()
```

Per effettuare la sentiment analysis utilizziamo **VADER** (Valence Aware Dictionary and sEntiment Reasoner), uno strumento per analizzare testi informali, come in questo caso i testi delle canzoni. VADER è un modello basato su regole che utilizza un dizionario di termini già etichettati con un punteggio di valenza emotiva (positivo, negativo o neutro), inoltre è in grado di gestire con precisione espressioni come intensificatori, negazioni e persino emoticon o punteggiatura.

```
In [38]: def get_sentiment(text):
    if text == "ERRORE":
        return 0, 0, 0, 0
    sentiment = sentiment_analyzer.polarity_scores(text)
    return sentiment['neg'], sentiment['neu'], sentiment['pos'], sentiment['compound']
```

```
df_2004['Negative'], df_2004['Neutral'], df_2004['Positive'], df_2004['Compound'] = ...
df_2021['Negative'], df_2021['Neutral'], df_2021['Positive'], df_2021['Compound'] = ...
```

La funzione `get_sentiment` ha lo scopo di calcolare i punteggi di sentiment di un dato testo utilizzando il metodo di sentiment analysis di VADER. Bisogna far attenzione ai testi che contengono la parola "ERRORE", questi infatti vengono impostati ad un valore standard (0,0,0,0) e non verranno conteggiati successivamente.

```
In [39]: def plot_sentiment_vs_ranking(df):
fig, axs = plt.subplots(2, 2, figsize=(25, 10))

axs[0, 0].scatter(df['Ranking'], df['Compound'], color='blue')
axs[0, 0].set_title('Ranking vs Compound Sentiment Score')
axs[0, 0].set_xlabel('Ranking')
axs[0, 0].set_ylabel('Compound Sentiment Score')

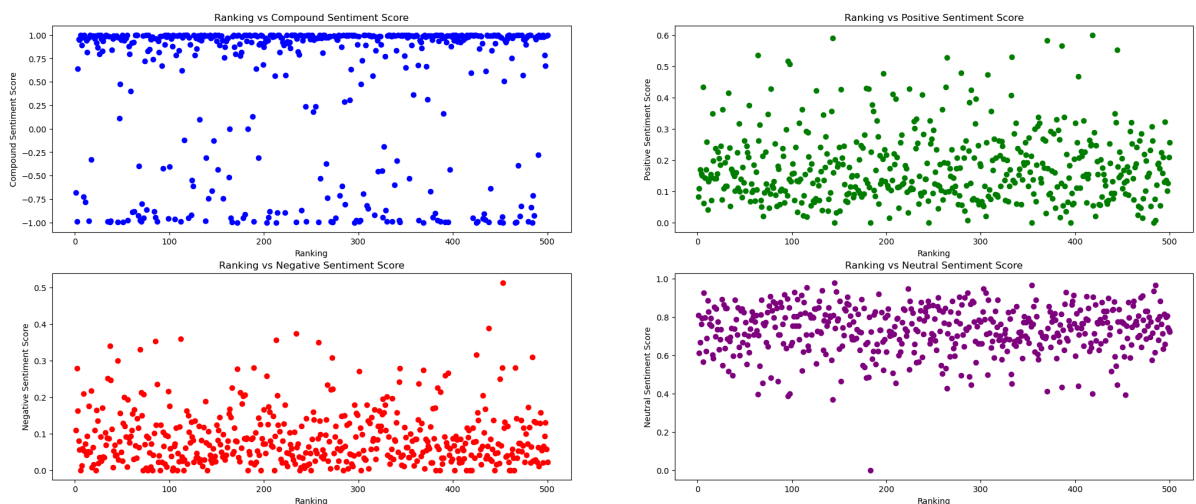
axs[0, 1].scatter(df['Ranking'], df['Positive'], color='green')
axs[0, 1].set_title('Ranking vs Positive Sentiment Score')
axs[0, 1].set_xlabel('Ranking')
axs[0, 1].set_ylabel('Positive Sentiment Score')

axs[1, 0].scatter(df['Ranking'], df['Negative'], color='red')
axs[1, 0].set_title('Ranking vs Negative Sentiment Score')
axs[1, 0].set_xlabel('Ranking')
axs[1, 0].set_ylabel('Negative Sentiment Score')

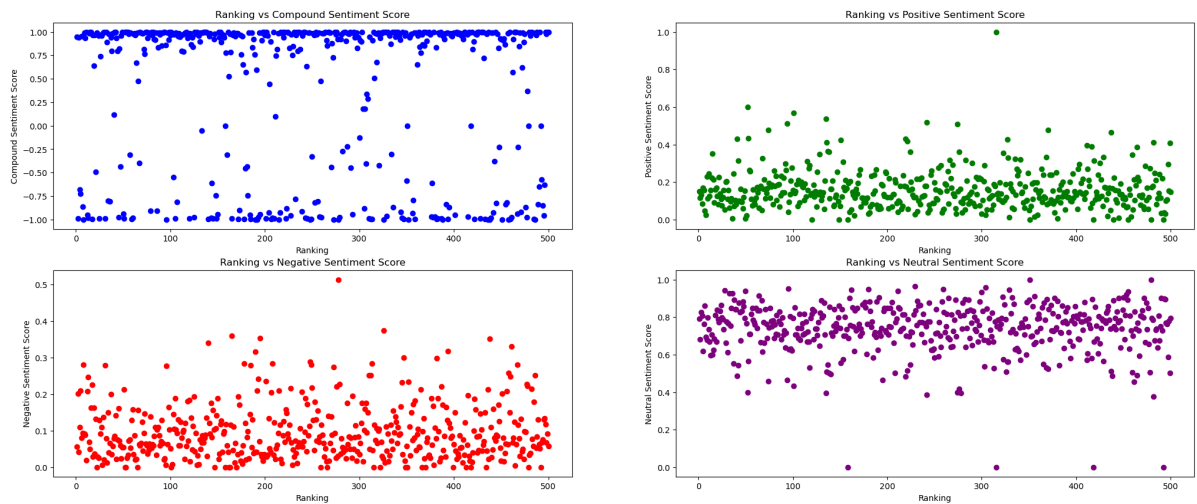
axs[1, 1].scatter(df['Ranking'], df['Neutral'], color='purple')
axs[1, 1].set_title('Ranking vs Neutral Sentiment Score')
axs[1, 1].set_xlabel('Ranking')
axs[1, 1].set_ylabel('Neutral Sentiment Score')

plt.show()
```

```
In [40]: plot_sentiment_vs_ranking(df_2004)
```



```
In [41]: plot_sentiment_vs_ranking(df_2021)
```



Analizziamo la situazione che ci viene presentata nei seguenti grafici. Possiamo vedere come in tutte e due le classifiche la maggior parte delle canzoni hanno un sentimento "Compound" vicino al valore 1, questo vuol dire che quei testi esprimono un sentimento positivo. Nonostante ciò ci sono comunque un numero non indifferente di canzoni con uno score che tende a -1, tuttavia sono sparse equamente per tutta l'asse x, questo vuol dire che non esiste una concentrazione basata sulla loro posizione in classifica. Per quanto riguarda gli altri grafici possiamo notare anche qui come questi siano abbastanza simili tra le due classifiche, anche se "Neutral" e "Positive" risultano più concentrate nel 2021 motivo per cui ci aspettiamo un cambiamento anche in 'Negative'.

```
In [42]: def get_sentiment_mean(df):
          df = df[df['Lyrics'] != "ERRORE"]
          return df[['Negative', 'Neutral', 'Positive', 'Compound']].mean()
```

La funzione `get_sentiment_mean` torna la media dei sentiment

```
In [43]: get_sentiment_mean(df_2004)
```

```
Out[43]: Negative    0.089253
          Neutral    0.732818
          Positive    0.177896
          Compound    0.494105
          dtype: float64
```

```
In [44]: get_sentiment_mean(df_2021)
```

```
Out[44]: Negative    0.095588
          Neutral    0.743372
          Positive    0.161016
          Compound    0.421427
          dtype: float64
```

I risultati mostrano ciò che avevamo visto precedentemente con i grafici, ovvero che nel 2021 il sentimento 'Positive' e 'Neutral' cambia rispetto al 2004, e ovviamente cambia anche 'Negative' (tuttavia questo non si riusciva a notare analizzando i risultati solo visivamente).

Conclusione

L'analisi condotta sui testi delle canzoni presenti nelle classifiche delle 500 migliori canzoni di tutti i tempi, pubblicate da Rolling Stone nel 2004 e nel 2021, ha evidenziato un aspetto

interessante: non esiste una correlazione significativa tra il sentiment dei testi e la posizione delle canzoni nel ranking. I risultati mostrano che, indipendentemente dal posizionamento in classifica, le canzoni tendono a presentare sentiment simili. Questa scoperta suggerisce che, almeno per quanto riguarda i sentiment espressi nei testi, altri fattori giocano un ruolo più determinante nella valutazione e classificazione delle canzoni. Questi potrebbero includere aspetti musicali, l'influenza culturale dell'artista, la popolarità, o persino il contesto storico in cui la canzone è stata rilasciata. Dunque il valore di una canzone, come percepito dai critici e dal pubblico, sembra essere più complesso e multidimensionale rispetto al semplice tono emotivo dei suoi testi.

Questo studio tuttavia potrebbe essere approfondito aggiungendo altre classifiche o tenendo conto di altri fattori come la musica o il ritmo di una canzone, infatti non analizzando questi aspetti abbiamo dovuto escludere alcune canzoni che non presentavano testo e altre canzoni che hanno contribuito in realtà presentano poche parole oppure ripetizioni di frasi(un esempio è Acknowledgement di John Coltrane).