CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DE DADOS (BIG DATA PROCESSING AND ANALYTICS)

Componente curricular: PARADIGMAS DE LINGUAGEM DE PROGRAMAÇÃO EM CIÊNCIA DE DADOS [TURMA 01D] - 2022/1 - Trilha 1.

Aluno: ROBSON DE FREITAS SAMPAIO.

URL deste notebook: https://github.com/rfsampaio/postgraduate_data_science/blob/main/notebooks/PL_A1.ipynb

---

Dicionário de palavras da biografia em inglês da banda de Rock The Ramones, visando confirmar a Lei de Zipf. Escolhi esse texto por gostar muito da banda e por querer testar a Lei de Zipf na língua inglesa.

Aqui, criamos um dicionário de palavras sobre a biografia em inglês da banda de Rock **The Ramones**.

- A primeira célula obtém o texto da biografia da banda na variável `texto`.
- Na segunda célula cria o dicionário de termos com a estrutura:

      mydict = { 'word1': qty, 'word2': qty, 'word3': qty, ...}

- A terceira célula apresenta um gráfico de distribuição dos termos do seu dicionário para confirmar a lei de Zipf.

Aquisição dos Dados

```python
import urllib.request
import re

texto = []

for line in urllib.request.urlopen('https://rfsampaio.github.io/docs/TheRamones-Biography.txt'):
    texto.append(line.decode('utf-8'))
```

```
for i in range(len(texto)):
    texto[i] = texto[i].lower()
    texto[i] = texto[i].replace('\n','')
    texto[i] = re.sub(r"[^a-z A-Z 0-9]","",texto[i]) # limpeza de caracteres desprezáveis com REGEX

#print(texto)
```

Construção do Dicionário

In [ ]:
```python
mydict = {}                                    # crie um dicionário vazio

for line in texto:

    line = line.lower()                        # converte para lower
    words = line.split()                       # separa cada palavra
    # print(words)

    for word in words:                         # para cada palavra em words
        if word not in mydict.keys():          # se palavra não está no dicionário
            mydict[word] = 1                   # acrescenta a word com o valor 1
        else:                                  # se a entrada já existe
            mydict[word] =  mydict[word] + 1   # apenas soma 1 ao valor já existente

# quantidade de palavras no dicionário
#print(len(mydict.keys()))

# imprime o dicionário
# print(mydict)

# ordenar, de forma decrescente, as palavras do dicionário pelo valor
sorted_mydict = sorted(mydict.items(), key=lambda x: x[1], reverse=True)

print(sorted_mydict)
```

[('the', 217), ('of', 84), ('a', 81), ('and', 78), ('in', 59), ('to', 45), ('ramones', 34), ('their', 33), ('on', 30), ('was', 27), ('as', 27), ('with', 25), ('by', 23), ('for', 23), ('at', 22), ('band', 21), ('dee', 20), ('ramone', 18), ('rock', 17), ('were', 16), ('they', 16), ('new', 14), ('after', 13), ('it', 13), ('an', 12), ('his', 12), ('album', 12), ('he', 11), ('would', 10), ('bands', 10), ('had', 10), ('pop', 9), ('into', 8), ('who', 8), ('music', 8), ('first', 8), ('its', 7), ('group', 7), ('but', 7), ('that', 7), ('johnny', 7), ('high', 7), ('lp', 6), ('roll', 6), ('songs', 6), ('about', 6), ('most', 6), ('career', 6), ('joey', 6), ('also', 6), ('up', 6), ('school', 6), ('former', 6), ('early', 6), ('two', 6), ('sound', 6), ('no', 6), ('more', 5), ('n', 5), ('this', 5), ('punk', 5), ('when', 5), ('some', 5), ('you', 5), ('all', 5), ('from', 5), ('took', 5), ('guitar', 5), ('tommy', 5), ('1974', 5), ('which', 5), ('began', 5), ('not', 5), ('one', 5), ('i', 5), ('albums', 5), ('than', 4), ('years', 4), ('is', 4), ('debut', 4), ('released', 4), ('cover', 4), ('soon', 4), ('there', 4), ('like', 4), ('record', 4), ('brothers', 4), ('we', 4), ('york', 4), ('born', 4), ('erdelyi', 4), ('played', 4), ('studio', 4), ('colvin', 4), ('heroin', 4), ('performance', 4), ('richie', 4), ('months', 4), ('60s', 4), ('be', 4), ('manager', 4), ('records', 4), ('dont', 4), ('concert', 4), ('contained', 4), ('marky', 4), ('out', 4), ('my', 4), ('30', 3), ('1976', 3), ('during', 3), ('era', 3), ('under', 3), ('four', 3), ('members', 3), ('original', 3), ('club', 3), ('yorks', 3), ('bowery', 3), ('local', 3), ('groups', 3), ('kristal', 3), ('television', 3), ('without', 3), ('lead', 3), ('our', 3), ('people', 3), ('hyman', 3), ('cummings', 3), ('america', 3), ('where', 3), ('garage', 3), ('work', 3), ('street', 3), ('love', 3), ('glitter', 3), ('bassist', 3), ('whose', 3), ('same', 3), ('us', 3), ('place', 3), ('release', 3), ('stein', 3), ('radio', 3), ('producer', 3), ('recorded', 3), ('other', 3), ('wanna', 3), ('around', 3), ('now', 3), ('made', 3), ('hit', 3), ('included', 3), ('such', 3), ('film', 3), ('song', 3), ('replaced', 3), ('metal', 3), ('attempt', 3), ('brain', 3), ('im', 3), ('moves', 3), ('come', 3), ('produced', 3), ('classic', 3), ('next', 3), ('time', 3), ('drummer', 3), ('1989', 3), ('april', 2), ('world', 2), ('amid', 2), ('disco', 2), ('collection', 2), ('loud', 2), ('short', 2), ('fast', 2), ('almost', 2), ('announced', 2), ('bad', 2), ('uniform', 2), ('jeans', 2), ('tshirts', 2), ('party', 2), ('day', 2), ('own', 2), ('material', 2), ('hilly', 2), ('back', 2), ('patti', 2), ('smith', 2), ('blondie', 2), ('couple', 2), ('kind', 2), ('shared', 2), ('forest', 2), ('hills', 2), ('1951', 2), ('drumming', 2), ('beatles', 2), ('met', 2), ('tom', 2), ('1952', 2), ('pair', 2), ('called', 2), ('recording', 2), ('engineer', 2), ('sept', 2), ('then', 2), ('manhattan', 2), ('together', 2), ('center', 2), ('these', 2), ('outsider', 2), ('stooges', 2), ('mc5', 2), ('those', 2), ('bass', 2), ('monte', 2), ('melnick', 2), ('later', 2), ('rehearsal', 2), ('while', 2), ('shows', 2), ('started', 2), ('rehearsals', 2), ('days', 2), ('him', 2), ('following', 2), ('themselves', 2), ('paul', 2), ('tour', 2), ('image', 2), ('stage', 2), ('none', 2), ('flop', 2), ('friends', 2), ('down', 2), ('didnt', 2), ('want', 2), ('cbgbs', 2), ('richard', 2), ('performed', 2), ('august', 2), ('frequently', 2), ('became', 2), ('act', 2), ('talking', 2), ('heads', 2), ('danny', 2), ('fields', 2), ('went', 2), ('linda', 2), ('sire', 2), ('hall', 2), ('today', 2), ('glue', 2), ('dees', 2), ('experience', 2), ('three', 2), ('july', 2), ('roundhouse', 2), ('london', 2), ('show', 2), ('remain', 2), ('returned', 2), ('set', 2), ('lineup', 2), ('leave', 2), ('home', 2), ('russia', 2), ('touches', 2), ('horror', 2), ('gimme', 2), ('probably', 2), ('top', 2), ('1978', 2), ('duties', 2), ('road', 2), ('mental', 2), ('shes', 2), ('movie', 2), ('arkush', 2), ('originally', 2), ('dvd', 2), ('seeking', 2), ('spector', 2), ('end', 2), ('century', 2), ('version', 2), ('fan', 2), ('through', 2), ('public', 2), ('did', 2), ('gary', 2), ('kurfirst', 2), ('important', 2), ('heavy', 2), ('title', 2), ('still', 2), ('charts', 2), ('player', 2), ('1987', 2), ('king', 2), ('year', 2), ('rap', 2), ('know', 2), ('appearance', 2), ('could', 2), ('last', 2), ('fans', 2), ('hollywood', 2), ('cancer', 2), ('15', 2), ('biography', 1), ('fact', 1), ('hard', 1), ('recapture', 1), ('impact', 1), ('23', 1), ('caught', 1), ('surprise', 1), ('issued', 1), ('wretched', 1), ('excess', 1), ('virtuosic', 1), ('indulgences', 1), ('progressive', 1), ('partyhearty', 1), ('exhortations', 1), ('something', 1), ('sun', 1), ('29minute', 1), ('14track', 1), ('funny', 1), ('invariably', 1), ('chanted', 1), ('onetwothreefour', 1), ('lyrics', 1), ('gluesniffing', 1), ('nazism', 1), ('violence', 1), ('stupid', 1), ('girlfriends', 1), ('so', 1), ('dumb', 1), ('smart', 1), ('stared', 1), ('impassively', 1), ('camera', 1), ('lens', 1), ('clad', 1), ('leather', 1), ('jackets', 1), ('if', 1), ('inviting', 1), ('listener', 1), ('block', 1), ('or', 1), ('brawl', 1), ('pretty', 1), ('name', 1), ('stuff', 1), ('rigorously', 1), ('stylized', 1), ('fortunate', 1), ('enough', 1), ('begin', 1), ('neophyte', 1), ('owner', 1), ('willing', 1), ('take', 1), ('chances', 1), ('unproven', 1), ('playing', 1), ('proprietor', 1), ('cbgb', 1), ('315', 1), ('told', 1), ('nobody', 1), ('going', 1), ('guys', 1), ('ill', 1), ('have', 1), ('contemporaries', 1), ('kickstarted', 1), ('revolt', 1), ('prophets', 1),

('honor', 1), ('sales', 1), ('divisive', 1), ('controversial', 1), ('presence', 1), ('late', 1), ('70s', 1), ('influence', 1), ('countless', 1), ('others', 1), ('artistic', 1), ('wake', 1), ('singer', 1), ('encapsulated', 1), ('roots', 1), ('appeal', 1), ('succinct', 1), ('sentences', 1), ('loners', 1), ('outcasts', 1), ('thats', 1), ('audience', 1), ('dark', 1), ('black', 1), ('sense', 1), ('humor', 1), ('only', 1), ('few', 1), ('understood', 1), ('grew', 1), ('neighborhood', 1), ('borough', 1), ('queens', 1), ('jeffrey', 1), ('may', 1), ('19', 1), ('tall', 1), ('gangling', 1), ('myopic', 1), ('suffering', 1), ('obsessivecompulsive', 1), ('disorder', 1), ('teenager', 1), ('john', 1), ('oct', 1), ('8', 1), ('reached', 1), ('1954', 1), ('aspiring', 1), ('baseball', 1), ('star', 1), ('bought', 1), ('jan', 1), ('29', 1), ('budapest', 1), ('hungary', 1), ('tangerine', 1), ('puppets', 1), ('noted', 1), ('plant', 1), ('douglas', 1), ('18', 1), ('kid', 1), ('developed', 1), ('habit', 1), ('age', 1), ('14', 1), ('ran', 1), ('buddy', 1), ('construction', 1), ('worker', 1), ('purchased', 1), ('guitars', 1), ('mannys', 1), ('misfits', 1), ('heroes', 1), ('dolls', 1), ('dictators', 1), ('gravitated', 1), ('toward', 1), ('flamboyance', 1), ('butch', 1), ('partner', 1), ('space', 1), ('fronted', 1), ('outfit', 1), ('sniper', 1), ('attend', 1), ('crony', 1), ('rhythm', 1), ('vocals', 1), ('drums', 1), ('stern', 1), ('left', 1), ('instrument', 1), ('hymans', 1), ('drum', 1), ('skills', 1), ('deemed', 1), ('inadequate', 1), ('tabbed', 1), ('vocalist', 1), ('colvins', 1), ('singing', 1), ('failed', 1), ('fruitless', 1), ('auditions', 1), ('finally', 1), ('reluctantly', 1), ('permanent', 1), ('position', 1), ('behind', 1), ('kit', 1), ('showed', 1), ('march', 1), ('named', 1), ('hattip', 1), ('pseudonym', 1), ('mccartney', 1), ('use', 1), ('checked', 1), ('hotels', 1), ('decided', 1), ('craft', 1), ('dressed', 1), ('leathers', 1), ('garb', 1), ('wore', 1), ('wrote', 1), ('credited', 1), ('collectively', 1), ('drew', 1), ('inspiration', 1), ('girlgroup', 1), ('bubblegum', 1), ('detritus', 1), ('culture', 1), ('skilled', 1), ('musician', 1), ('attacked', 1), ('speed', 1), ('volume', 1), ('wecandothis', 1), ('inventiveness', 1), ('gig', 1), ('recalled', 1), ('invited', 1), ('see', 1), ('hated', 1), ('anymore', 1), ('needed', 1), ('find', 1), ('primed', 1), ('listen', 1), ('prejudice', 1), ('oddball', 1), ('described', 1), ('press', 1), ('unlike', 1), ('drill', 1), ('rear', 1), ('molar', 1), ('opened', 1), ('december', 1), ('1973', 1), ('cruddy', 1), ('welcome', 1), ('spring', 1), ('televisions', 1), ('guitarists', 1), ('verlaine', 1), ('lloyd', 1), ('talked', 1), ('letting', 1), ('them', 1), ('play', 1), ('sunday', 1), ('residency', 1), ('1617', 1), ('sets', 1), ('often', 1), ('lasted', 1), ('20', 1), ('minutes', 1), ('devolve', 1), ('onstage', 1), ('arguments', 1), ('between', 1), ('tight', 1), ('hot', 1), ('style', 1), ('apparent', 1), ('precedent', 1), ('regulars', 1), ('linchpin', 1), ('among', 1), ('key', 1), ('performers', 1), ('including', 1), ('stoked', 1), ('underground', 1), ('197475', 1), ('1975', 1), ('acquired', 1), ('simpatico', 1), ('formerly', 1), ('familiar', 1), ('andy', 1), ('warhols', 1), ('circle', 1), ('velvet', 1), ('undergrounds', 1), ('heyday', 1), ('staffer', 1), ('elektra', 1), ('man', 1), ('signed', 1), ('both', 1), ('manage', 1), ('wife', 1), ('president', 1), ('seymour', 1), ('joined', 1), ('comanager', 1), ('she', 1), ('advised', 1), ('her', 1), ('husband', 1), ('sign', 1), ('innovative', 1), ('february', 1), ('manhattans', 1), ('city', 1), ('craig', 1), ('leon', 1), ('cutting', 1), ('17', 1), ('6400', 1), ('arrived', 1), ('navelgazing', 1), ('singersongwriters', 1), ('selfindulgent', 1), ('progrockers', 1), ('emptyheaded', 1), ('dollies', 1), ('challenge', 1), ('instantly', 1), ('divided', 1), ('critics', 1), ('camps', 1), ('hailed', 1), ('genius', 1), ('purest', 1), ('garbage', 1), ('blitzkrieg', 1), ('bop', 1), ('beat', 1), ('brat', 1), ('walk', 1), ('speedy', 1), ('slap', 1), ('face', 1), ('nazi', 1), ('imagery', 1), ('your', 1), ('tomorrow', 1), ('testorshuffing', 1), ('sniff', 1), ('violent', 1), ('saga', 1), ('male', 1), ('prostitution', 1), ('53rd', 1), ('3rd', 1), ('chicken', 1), ('hawk', 1), ('trade', 1), ('queasy', 1), ('thing', 1), ('certain', 1), ('nothing', 1), ('else', 1), ('4', 1), ('bicentennial', 1), ('british', 1), ('invasion', 1), ('reverse', 1), ('headlining', 1), ('soldout', 1), ('outside', 1), ('word', 1), ('copies', 1), ('english', 1), ('shores', 1), ('likeminded', 1), ('young', 1), ('uk', 1), ('already', 1), ('absorbing', 1), ('lessons', 1), ('attended', 1), ('royalty', 1), ('sex', 1), ('pistols', 1), ('clash', 1), ('damned', 1), ('generation', 1), ('x', 1), ('pretenders', 1), ('england', 1), ('bastion', 1), ('support', 1), ('throughout', 1), ('eve', 1), ('76', 1), ('rainbow', 1), ('twolp', 1), ('alive', 1), ('featuring', 1), ('rocket', 1), ('followed', 1), ('quick', 1), ('succession', 1), ('separated', 1), ('just', 1), ('nine', 1), ('1977', 1), ('further', 1), ('refined', 1), ('buzzsaw', 1), ('assault', 1), ('accenting', 1), ('strongly', 1), ('bubbled', 1), ('surface', 1), ('staples', 1), ('pinhead', 1), ('inspired', 1), ('tod', 1), ('brownings', 1), ('freaks', 1), ('babysitter', 1), ('shock', 1), ('treatment', 1), ('huffers', 1), ('anthem', 1), ('carbona', 1), ('latter', 1), ('repressings', 1), ('bouncy', 1), ('sheena', 1), ('rocker', 1), ('found', 1), ('slot', 1), ('accessible', 1), ('crack', 1), ('50', 1), ('peaking', 1), ('48', 1), ('cretin', 1), ('hop', 1), ('rockawa

y', 1), ('beach', 1), ('teenage', 1), ('lobotomy', 1), ('complementary', 1), ('trashmens', 1), ('epic', 1), ('surfin', 1), ('bird', 1), ('coproduced', 1), ('second', 1), ('third', 1), ('tony', 1), ('bongiovi', 1), ('wearied', 1), ('grind', 1), ('gave', 1), ('replacement', 1), ('marc', 1), ('bell', 1), ('member', 1), ('trio', 1), ('dust', 1), ('recently', 1), ('drummed', 1), ('hells', 1), ('attitudinal', 1), ('voidoids', 1), ('thus', 1), ('tours', 1), ('duty', 1), ('ruin', 1), ('strong', 1), ('diversify', 1), ('hopes', 1), ('obtaining', 1), ('airtime', 1), ('favor', 1), ('getting', 1), ('40', 1), ('containing', 1), ('illness', 1), ('sedated', 1), ('go', 1), ('nihilism', 1), ('against', 1), ('along', 1), ('sharp', 1), ('sublime', 1), ('questioningly', 1), ('searchers', 1), ('needles', 1), ('pins', 1), ('gloriously', 1), ('hummable', 1), ('close', 1), ('shortly', 1), ('residence', 1), ('los', 1), ('angeles', 1), ('filming', 1), ('b', 1), ('picture', 1), ('czar', 1), ('roger', 1), ('corman', 1), ('directed', 1), ('fillmore', 1), ('east', 1), ('lighting', 1), ('director', 1), ('allan', 1), ('completed', 1), ('joe', 1), ('dante', 1), ('suffered', 1), ('heart', 1), ('attack', 1), ('midproduction', 1), ('earmarked', 1), ('vehicle', 1), ('cheap', 1), ('trick', 1), ('goofy', 1), ('loving', 1), ('spoof', 1), ('americaninternationals', 1), ('teenbeach', 1), ('pictures', 1), ('surrounded', 1), ('youthful', 1), ('cast', 1), ('anarchic', 1), ('musical', 1), ('climaxed', 1), ('demolition', 1), ('titular', 1), ('driveins', 1), ('1979', 1), ('nevertheless', 1), ('enhanced', 1), ('becoming', 1), ('cult', 1), ('vhs', 1), ('laser', 1), ('disc', 1), ('cable', 1), ('well', 1), ('midnight', 1), ('hopefully', 1), ('commercial', 1), ('shift', 1), ('direction', 1), ('hired', 1), ('unlikely', 1), ('phil', 1), ('reclusive', 1), ('volatile', 1), ('hitmaker', 1), ('been', 1), ('pursuing', 1), ('making', 1), ('1980', 1), ('la', 1), ('fractious', 1), ('affair', 1), ('walked', 1), ('session', 1), ('manic', 1), ('perfectionist', 1), ('opening', 1), ('chord', 1), ('nearly', 1), ('100', 1), ('times', 1), ('allegedly', 1), ('brandished', 1), ('pistol', 1), ('spectorized', 1), ('memorable', 1), ('tracks', 1), ('chinese', 1), ('scoring', 1), ('cowritten', 1), ('thunders', 1), ('heartbreakers', 1), ('do', 1), ('remember', 1), ('affected', 1), ('says', 1), ('aint', 1), ('havana', 1), ('highestcharting', 1), ('reaching', 1), ('44', 1), ('unpleasant', 1), ('dawn', 1), ('80s', 1), ('internal', 1), ('stress', 1), ('develop', 1), ('within', 1), ('date', 1), ('marry', 1), ('joeys', 1), ('girlfriend', 1), ('remainder', 1), ('mates', 1), ('avoid', 1), ('direct', 1), ('communication', 1), ('each', 1), ('using', 1), ('intermediaries', 1), ('remained', 1), ('popular', 1), ('touring', 1), ('attraction', 1), ('solid', 1), ('base', 1), ('attempts', 1), ('break', 1), ('proved', 1), ('futile', 1), ('10ccs', 1), ('graham', 1), ('gouldman', 1), ('pleasant', 1), ('dreams', 1), ('1981', 1), ('squeeze', 1), ('underrated', 1), ('shouldve', 1), ('beens', 1), ('kkk', 1), ('baby', 1), ('away', 1), ('sensation', 1), ('711', 1), ('veterans', 1), ('ritchie', 1), ('cordell', 1), ('glen', 1), ('kolotkin', 1), ('subterranean', 1), ('jungle', 1), ('1983', 1), ('nice', 1), ('gutsy', 1), ('wicked', 1), ('originals', 1), ('psycho', 1), ('therapy', 1), ('bomb', 1), ('girl', 1), ('mixed', 1), ('pitch', 1), ('perfect', 1), ('chambers', 1), ('has', 1), ('explosions', 1), ('little', 1), ('bit', 1), ('o', 1), ('soul', 1), ('once', 1), ('again', 1), ('understand', 1), ('brand', 1), ('trashed', 1), ('context', 1), ('modern', 1), ('demands', 1), ('parted', 1), ('company', 1), ('managers', 1), ('linked', 1), ('stability', 1), ('1984', 1), ('aggressive', 1), ('too', 1), ('tough', 1), ('die', 1), ('brought', 1), ('erdelyiramone', 1), ('coproduce', 1), ('longtime', 1), ('ed', 1), ('stasium', 1), ('fired', 1), ('advancing', 1), ('alcoholism', 1), ('creating', 1), ('problems', 1), ('ongoing', 1), ('addiction', 1), ('journeyman', 1), ('reinhardt', 1), ('provided', 1), ('somewhat', 1), ('renaissance', 1), ('updated', 1), ('adding', 1), ('hardcore', 1), ('mix', 1), ('resulting', 1), ('many', 1), ('future', 1), ('irresistible', 1), ('magic', 1), ('howing', 1), ('moon', 1), ('chasing', 1), ('night', 1), ('daytime', 1), ('dilemma', 1), ('offensive', 1), ('thrash', 1), ('warthog', 1), ('endless', 1), ('vacation', 1), ('gutter', 1), ('mamas', 1), ('boys', 1), ('track', 1), ('proving', 1), ('creative', 1), ('goods', 1), ('grope', 1), ('even', 1), ('agree', 1), ('animal', 1), ('boy', 1), ('1986', 1), ('best', 1), ('hanging', 1), ('upside', 1), ('bonzo', 1), ('goes', 1), ('bitburg', 1), ('antireagan', 1), ('screed', 1), ('much', 1), ('johnnys', 1), ('chagrin', 1), ('staunch', 1), ('right', 1), ('winger', 1), ('halfway', 1), ('sanity', 1), ('fate', 1), ('immediate', 1), ('predecessors', 1), ('chart', 1), ('lives', 1), ('hits', 1), ('jumped', 1), ('ship', 1), ('brief', 1), ('unsuccessful', 1), ('flirtation', 1), ('blondies', 1), ('clem', 1), ('burke', 1), ('nowsober', 1), ('drain', 1), ('notable', 1), ('inclusion', 1), ('pet', 1), ('sematary', 1), ('theme', 1), ('liketitled', 1), ('based', 1), ('novel', 1), ('stephen', 1), ('principal', 1), ('writer', 1), ('quit', 1), ('nom', 1), ('de', 1), ('disque', 1), ('how', 1), ('negro', 1), ('continue', 1), ('write', 1), ('until', 1), ('dissolution', 1), ('assumed', 1), ('christopher', 1), ('joseph', 1), ('ward', 1), ('cj', 1), ('sort', 1), ('awol', 1), ('status', 1), ('marine', 1), ('corps', 1), ('before', 1), ('formally', 1), ('joining', 1), ('jerry', 1), ('lewis',

1), ('labor', 1), ('muscular', 1), ('dystrophy', 1), ('telethon', 1), ('september', 1), ('16', 1), ('shifted', 1), ('radioactiv
e', 1), ('imprint', 1), ('operated', 1), ('stock', 1), ('influences', 1), ('soared', 1), ('90s', 1), ('seattle', 1), ('grunge',
1), ('explosion', 1), ('off', 1), ('almighty', 1), ('motorheadwho', 1), ('claimed', 1), ('metals', 1), ('tribute', 1), ('1991',
1), ('1916', 1), ('titled', 1), ('mondo', 1), ('bizarro', 1), ('1992', 1), ('consistent', 1), ('good', 1), ('acid', 1), ('eater
s', 1), ('1994', 1), ('covers', 1), ('psychedelic', 1), ('continued', 1), ('languish', 1), ('near', 1), ('bottom', 1), ('sell',
1), ('halls', 1), ('europe', 1), ('fill', 1), ('soccer', 1), ('stadiums', 1), ('south', 1), ('received', 1), ('large', 1), ('sca
le', 1), ('homeland', 1), ('plethora', 1), ('younger', 1), ('rockers', 1), ('owed', 1), ('everything', 1), ('taken', 1), ('immin
ent', 1), ('exit', 1), ('1995', 1), ('adios', 1), ('amigos', 1), ('bade', 1), ('farewell', 1), ('6', 1), ('1996', 1), ('palace',
1), ('vine', 1), ('calif', 1), ('1997', 1), ('cd', 1), ('outta', 1), ('here', 1), ('guests', 1), ('lemmy', 1), ('motorhead', 1),
('eddie', 1), ('vedder', 1), ('pearl', 1), ('jam', 1), ('chris', 1), ('cornell', 1), ('soundgarden', 1), ('unrecognizable', 1),
('shaved', 1), ('head', 1), ('painfully', 1), ('thin', 1), ('torso', 1), ('except', 1), ('1999', 1), ('signing', 1), ('virgin',
1), ('megastore', 1), ('appear', 1), ('underwent', 1), ('protracted', 1), ('battle', 1), ('ended', 1), ('death', 1), ('2001',
1), ('solo', 1), ('worry', 1), ('me', 1), ('posthumously', 1), ('inducted', 1), ('fame', 1), ('ceremony', 1), ('june', 1), ('5',
1), ('2002', 1), ('died', 1), ('overdose', 1), ('succumbed', 1), ('prostate', 1), ('2004', 1), ('documentary', 1), ('enter', 1),
('theaters', 1), ('country', 1), ('reshaped', 1), ('landscape', 1), ('will', 1), ('forever', 1), ('hold', 1), ('hearts', 1), ('f
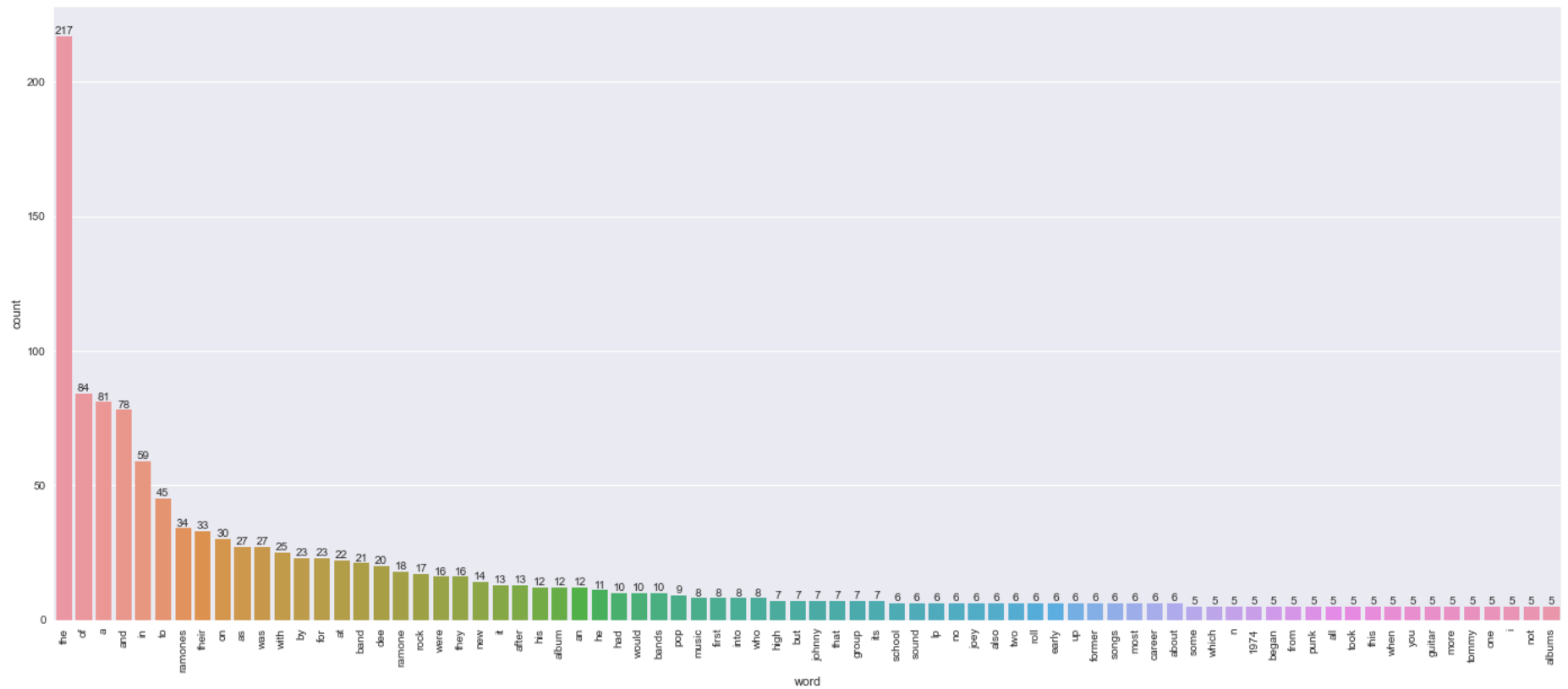inest', 1)]

Exibição dos Resultados

```python
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt

df = pd.DataFrame(mydict.items(),  columns=['word', 'count']).sort_values('count',ascending=False)
df = df[df['count'] > 4] # somente termos com mais de 4 ocorrências
# df = df.iloc[ np.int(len(df)/2) - 10 : np.int(len(df)/2) + 10 ] # para livros ou textos com muitos termos limita a um número mi

plt.figure(figsize=(24,10))
mpl.style.use(['seaborn'])

# incluindo labels nas barras
ax = sns.barplot(x=df.word,y=df['count'],data=df)
for i in ax.containers:
    ax.bar_label(i,)

plt.xticks(rotation=90)
plt.show()
```

O gráfico de distribuição de frequência acima confirma a Lei de Zipf.