Loading and Saving Data with Pandas

The pygmentize command-line tool can be used to display a text file in IPython or Jupyter. It can be installed with `conda install -c conda-forge pygments` or with pip. We can take a look at the CSV with this tool:

```
In [4]:  !pygmentize -l text itunes_data.csv
```

```
Track,Composer,Milliseconds,Bytes,UnitPrice,Genre,Album,Artist
All the Best Cowboys Have Daddy Issues,,2555492,211743651,1.99,TV Shows,"Los
t, Season 1",Lost
Beira Mar,Gilberto Gil,295444,9597994,0.99,Latin,Unplugged,Eric Clapton
Brasil,"Milton Nascimento, Fernando Brant",155428,5252560,0.99,Latin,Milton N
ascimento Ao Vivo,Milton Nascimento
Ben Franklin,,1271938,264168080,1.99,Comedy,"The Office, Season 3",The Office
O Último Romântico (Ao Vivo),,231993,7692697,0.99,Latin,Lulu Santos - RCA 100
Anos De Música - Álbum 02,Lulu Santos
Freewheel Burning,,265952,8713599,0.99,Metal,Living After Midnight,Judas Prie
st
That's The Way,Robert Plant,343431,11248455,0.99,Rock,BBC Sessions [Disc 2]
[Live],Led Zeppelin
Mellowship Slinky In B Major,Anthony Kiedis/Chad Smith/Flea/John Frusciante,2
40091,7971384,0.99,Alternative & Punk,Blood Sugar Sex Magik,Red Hot Chili Pep
pers
"Diga Lá, Coração",,255921,8280636,0.99,Latin,Meus Momentos,Gonzaguinha
Come As You Are,Kurt Cobain,219219,7123357,0.99,Rock,Nevermind,Nirvana
The Spirit Of Radio,Geddy Lee And Alex Lifeson/Geddy Lee And Neil Peart/Rush,
299154,9862012,0.99,Rock,Retrospective I (1974-1980),Rush
Midnight Blue,David Coverdale/Earl Slick,298631,9750990,0.99,Rock,Into The Li
ght,David Coverdale
Se Liga,,410409,13559173,0.99,Hip Hop/Rap,Os Cães Ladram Mas A Caravana Não P
ára,Planet Hemp
Carnival Of Sorts,R.E.M.,233482,7669658,0.99,Alternative & Punk,The Best Of
R.E.M.: The IRS Years,R.E.M.
Paranoid,"A. F. Iommi, W. Ward, T. Butler, J. Osbourne",176352,5729813,0.99,M
etal,Tribute,Ozzy Osbourne
Got That Feeling,Mike Patton,140852,4643227,0.99,Alternative & Punk,Album Of
The Year,Faith No More
What Is And What Should Never Be,"Jimmy Page, Robert Plant",287973,9369385,0.
99,Rock,Led Zeppelin II,Led Zeppelin
The Zephyr Song,"Anthony Kiedis, Flea, John Frusciante, and Chad Smith",23296
0,7690312,0.99,Rock,By The Way,Red Hot Chili Peppers
Run Silent Run Deep,Bruce Dickinson/Steve Harris,275408,4407296,0.99,Metal,No
Prayer For The Dying,Iron Maiden
Realidade Virtual,Bino Farias/Da Gama/Lazão/Toni Garrido,195239,6503533,0.99,
Reggae,Acústico MTV [Live],Cidade Negra
Há Quanto Tempo,,270027,9004470,0.99,Latin,Jota Quest-1995,Jota Quest
Oprah,,6635,224313,0.99,Alternative & Punk,Body Count,Body Count
As Aparências Enganam,,247379,8014346,0.99,Latin,Elis Regina-Minha História,E
lis Regina
Confidence Man,,2615244,223756475,1.99,TV Shows,"Lost, Season 1",Lost
Os Alquimistas Estão Chegando,,367281,12304520,0.99,Latin,Jorge Ben Jor 25 An
os,Jorge Ben
Springsville,J. Carisi,207725,6776219,0.99,Jazz,Miles Ahead,Miles Davis
Wake Up,Alanis Morissette & Glenn Ballard,293485,9703359,0.99,Rock,Jagged Lit
tle Pill,Alanis Morissette
Ave Maria,Franz Schubert,338243,5605648,0.99,Classical,A Soprano Inspired,"Br
itten Sinfonia, Ivor Bolton & Lesley Garrett"
A Banda,,132493,4349539,0.99,Latin,Minha Historia,Chico Buarque
Fight Fire With Fire,Metallica,285753,9420856,0.99,Metal,Ride The Lightning,M
etallica
Sound of a Gun,,260154,4234990,0.99,Alternative,Revelations,Audioslave
RV,,223242,7288162,0.99,Alternative & Punk,Angel Dust,Faith No More
É que Nessa Encarnação Eu Nasci Manga,Lucina/Luli,196519,6568081,0.99,Pop,Axé
```

Bahia 2001,Various Artists
Do The Evolution,Eddie Vedder & Stone Gossard,225462,7377286,0.99,Rock,Live O
n Two Legs [Live],Pearl Jam
Prowler,Steve Harris,255634,4091904,0.99,Metal,A Real Dead One,Iron Maiden
They Can't Take That Away From Me,george gershwin/ira gershwin,161227,524004
3,0.99,Easy Listening,My Way: The Best Of Frank Sinatra [Disc 1],Frank Sinatr
a
The Real Problem,,11650,387360,0.99,Alternative & Punk,Body Count,Body Count
Tempus Fugit,Miles Davis,231784,7548434,0.99,Jazz,The Essential Miles Davis
[Disc 1],Miles Davis
Diwali,,1279904,252726644,1.99,TV Shows,"The Office, Season 3",The Office
Bring'em Back Alive,Audioslave/Chris Cornell,329534,7911634,0.99,Rock,Audiosl
ave,Audioslave
Time Is On My Side,Jerry Ragavoy,179983,5855836,0.99,Rock,"Hot Rocks, 1964-19
71 (Disc 1)",The Rolling Stones
The Ocean,John Bonham/John Paul Jones,271098,8846469,0.99,Rock,Houses Of The
Holy,Led Zeppelin
Minority,Billie Joe Armstrong -Words Green Day -Music,168803,5535061,0.99,Alt
ernative & Punk,International Superhits,Green Day
The Messenger,Daniel Lanois,212062,6975437,0.99,Alternative & Punk,Tangents,T
he Tea Party
Quanta (Live),,357485,11774865,0.99,Latin,Quanta Gente Veio Ver (Live),Gilber
to Gil
Ironic,Alanis Morissette & Glenn Ballard,229825,7598866,0.99,Rock,Jagged Litt
le Pill,Alanis Morissette
Faraó Divindade Do Egito,,228571,7523278,0.99,Latin,Olodum,Olodum
"Nabucco: Chorus, ""Va, Pensiero, Sull'ali Dorate""",Giuseppe Verdi,274504,44
98583,0.99,Classical,Great Opera Choruses,"Chicago Symphony Chorus, Chicago S
ymphony Orchestra & Sir Georg Solti"
Redundant,Billie Joe Armstrong -Words Green Day -Music,198164,6481753,0.99,Al
ternative & Punk,International Superhits,Green Day
Leave Us Leap,,182726,5898810,0.99,Jazz,Up An' Atom,Gene Krupa
Flight Of The Icarus,Adrian Smith/Bruce Dickinson,230269,3686400,0.99,Metal,P
iece Of Mind,Iron Maiden
Adoled (Ocean),,185103,6009946,0.99,Hip Hop/Rap,Os Cães Ladram Mas A Caravana
Não Pára,Planet Hemp
Let's See Action,Pete Townshend,243513,8078418,0.99,Rock,My Generation - The
Very Best Of The Who,The Who
"Toccata and Fugue in D Minor, BWV 565: I. Toccata",Johann Sebastian Bach,153
901,2649938,0.99,Classical,Bach: Toccata & Fugue in D Minor,Ton Koopman
Blind Curve: Vocal Under A Bloodlight / Passing Strangers / Mylo / Perimeter
Walk / Threshold,"Kelly, Mosley, Rothery, Trewaves",569704,18578995,0.99,Roc
k,Misplaced Childhood,Marillion
Dirty Little Thing,"Dave Kushner, Duff, Keith Nelson, Matt Sorum, Scott Weila
nd & Slash",237844,7732982,0.99,Rock,Contraband,Velvet Revolver
Só Tinha De Ser Com Você,Vários,389642,13085596,0.99,Latin,Chill: Brazil (Dis
c 2),Antônio Carlos Jobim
Slug,"Brian Eno, Bono, Adam Clayton, The Edge & Larry Mullen Jnr.",281469,929
5950,0.99,Soundtrack,Original Soundtracks 1,Passengers
Black Mountain Side,Jimmy Page,132702,4440602,0.99,Rock,Led Zeppelin I,Led Ze
ppelin
Pensamento,Bino/Da Gama/Ras Bernardo,192391,6399761,0.99,Reggae,Cidade Negra
- Hits,Cidade Negra
Even Better Than The Real Thing,U2,221361,7279392,0.99,Rock,Achtung Baby,U2
Powerslave,Dickinson,407823,9791106,0.99,Metal,Powerslave,Iron Maiden
Loud Love,Chris Cornell,297456,9660953,0.99,Rock,A-Sides,Soundgarden

2 Minutes To Midnight,Smith/Dickinson,366550,8799380,0.99,Metal,Live After Death,Iron Maiden
Out Of Exile,"Cornell, Commerford, Morello, Wilk",291291,9506571,0.99,Alternative & Punk,Out Of Exile,Audioslave
Spank Thru,Kurt Cobain,190354,6186487,0.99,Rock,From The Muddy Banks Of The Wishkah [Live],Nirvana
London Calling,Joe Strummer/Mick Jones,199706,6569007,0.99,Alternative & Punk,The Singles,The Clash
Break on Through,"Robby Krieger, Ray Manzarek, John Densmore, Jim Morrison",149342,4943144,0.99,Rock,The Doors,The Doors
We Are The Champions,Mercury,182883,5939794,0.99,Rock,News Of The World,Queen
Hard Lovin' Man,"Blackmore, Gillan, Glover, Lord, Paice",431203,13931179,0.99,Rock,Deep Purple In Rock,Deep Purple
Jingo,M.Babatunde Olantunji,592953,19736495,0.99,Rock,Santana - As Years Go By,Santana
200 Years Old,Frank Zappa,272561,8912465,0.99,Rock,Bongo Fury,Frank Zappa & Captain Beefheart
Santa Clara Clareou,,380081,12524725,0.99,Latin,Jorge Ben Jor 25 Anos,Jorge Ben
Juventude Transviada (Ao Vivo),,278622,9183808,0.99,Latin,Cássia Eller - Coleção Sem Limite [Disc 2],Cássia Eller
Rollover D.J.,C. Cester/N. Cester,196702,6406517,0.99,Alternative & Punk,Get Born,JET
Leave,Bill Berry-Peter Buck-Mike Mills-Michael Stipe,437968,14433365,0.99,Rock,New Adventures In Hi-Fi,R.E.M.
God Part II,"Bono/Clayton, Adam/Mullen Jr., Larry/The Edge",195604,6497570,0.99,Rock,Rattle And Hum,U2
Cold Gin,Ace Frehley,262243,8609783,0.99,Rock,Greatest Kiss,Kiss
Sister Morphine,Faithfull/Jagger/Richards,376215,12345289,0.99,Rock,No Security,The Rolling Stones
Nefertiti,Miles Davis,473495,15478450,0.99,Jazz,The Essential Miles Davis [Disc 2],Miles Davis
Take the Box,Luke Smith,199160,3281526,0.99,Pop,Frank,Amy Winehouse
Good Riddance (Time Of Your Life),Billie Joe Armstrong -Words Green Day -Music,153600,5075241,0.99,Alternative & Punk,International Superhits,Green Day
Can't Keep,Eddie Vedder,219428,7215713,0.99,Rock,Riot Act,Pearl Jam
Wall Of Denial,Doyle Bramhall/Stevie Ray Vaughan,336927,11085915,0.99,Blues,In Step,Stevie Ray Vaughan & Double Trouble
If God Will Send His Angels,"Bono, The Edge, Adam Clayton, and Larry Mullen",322533,10563329,0.99,Rock,Pop,U2
Viradouro,Dadinho/Gilbreto Gomes/Gustavo/P.C. Portugal/R. Mocoto,344320,11484362,0.99,Latin,Sambas De Enredo 2001,Various Artists
Surrender,U2,333505,11221406,0.99,Rock,War,U2
Funky Monks,Anthony Kiedis/Chad Smith/Flea/John Frusciante,323395,10708168,0.99,Alternative & Punk,Blood Sugar Sex Magik,Red Hot Chili Peppers
Back in the Village,Dickinson/Smith,320548,7696518,0.99,Metal,Powerslave,Iron Maiden
Freedom For My People,"Mabins, Macie/Magee, Sterling/Robinson, Bobby",38164,1249764,0.99,Rock,Rattle And Hum,U2
Um Jantar Pra Dois,,244009,8021589,0.99,Latin,The Best of Ed Motta,Ed Motta
Hang 'Em High,Alex Van Halen/David Lee Roth/Edward Van Halen/Michael Anthony,210259,6872314,0.99,Rock,Diver Down,Van Halen
Spirit's in the Material World,G M Sumner,181133,5986622,0.99,Rock,The Police Greatest Hits,The Police
Maquinarama,Chico Amaral/Samuel Rosa,245629,8213710,0.99,Rock,Maquinarama,Skank

Believe,,310778,10317185,0.99,Jazz,Heart of the Night,Spyro Gyra
The Wait,Killing Joke,295418,9688418,0.99,Metal,Garage Inc. (Disc 2),Metallica
Near Wild Heaven,Bill Berry/Michael Stipe/Mike Mills/Peter Buck,199862,6610009,0.99,Alternative & Punk,Out Of Time,R.E.M. Feat. Kate Pearson
Desordem,Charles Gavin/Marcelo Fromer/Sérgio Britto,213289,7067340,0.99,Alternative & Punk,Volume Dois,Titãs
Canário Do Reino,,139337,4552858,0.99,Latin,Serie Sem Limite (Disc 2),Tim Maia
O Pulso,Titãs,199131,6566998,0.99,Alternative & Punk,Acústico,Titãs
Fear Of The Dark,Steve Harris,431333,6906078,0.99,Metal,A Real Live One,Iron Maiden
Every Little Thing She Does is Magic,G M Sumner,261120,8646853,0.99,Rock,The Police Greatest Hits,The Police
O Livro Dos Dias,Renato Russo,257253,8570929,0.99,Latin,A TempestadeTempestade Ou O Livro Dos Dias,Legião Urbana
Arc,Pearl Jam,65593,2099421,0.99,Rock,Riot Act,Pearl Jam
All My Life,Foo Fighters,263653,8665545,0.99,Alternative & Punk,One By One,Foo Fighters
Put You Down,Jerry Cantrell,196231,6420530,0.99,Rock,Facelift,Alice In Chains
Já Foi,Mônica Marianno,245681,8094872,0.99,World,Demorou...,Mônica Marianno
Gatas Extraordinárias,,212506,7095702,0.99,Latin,Cássia Eller - Coleção Sem Limite [Disc 2],Cássia Eller
Carta Ao Tom 74,,162560,5382354,0.99,Bossa Nova,Vinícius De Moraes - Sem Limite,Toquinho & Vinícius
Babyface,U2; Bono,241998,7942573,0.99,Rock,Zooropa,U2
Light My Way,Audioslave/Chris Cornell,303595,7289084,0.99,Rock,Audioslave,Audioslave
Solar,"Milton Nascimento, Fernando Brant",156212,5098288,0.99,Latin,Milton Nascimento Ao Vivo,Milton Nascimento
Something Nice Back Home,,2612779,484711353,1.99,Drama,"LOST, Season 4",Lost
Murder On the Rising Star,,2935894,551759986,1.99,Sci Fi & Fantasy,"Battlestar Galactica (Classic), Season 1",Battlestar Galactica (Classic)
Think,Lowman Pauling,166739,5513208,0.99,R&B/Soul,Sex Machine,James Brown
Ando Meio Desligado,Caetano Veloso,195813,6547648,0.99,Latin,Chill: Brazil (Disc 1),Marcos Valle
Trac Trac,Fito Paez/Herbert Vianna,231653,7638256,0.99,Latin,Arquivo II,Os Paralamas Do Sucesso
Whiplash,"James Hetfield, Lars Ulrich",249208,8102839,0.99,Metal,Kill 'Em All,Metallica
Brave New World,Bruce Dickinson/David Murray/Steve Harris,378984,15161472,0.99,Rock,Brave New World,Iron Maiden
A Statistic,,6373,211997,0.99,Alternative & Punk,Body Count,Body Count
Manifest Destiny,Toby Smith,382197,12676962,0.99,Electronica/Dance,The Return Of The Space Cowboy,Jamiroquai
Falamansa Song,,237165,7921313,0.99,Latin,Deixa Entrar,Falamansa
Song For Lorraine,Jay Beckenstein,240091,8101723,0.99,Jazz,Morning Dance,Spyro Gyra
Diamante De Mendigo,,206053,6775101,0.99,Rock,Raul Seixas,Raul Seixas
Corcovado (Quiet Nights Of Quiet Stars),,205662,6687994,0.99,Jazz,Warner 25 Anos,Antônio Carlos Jobim
KKK Bitch,,173008,5709631,0.99,Alternative & Punk,Body Count,Body Count
União Da Ilha,"Dito/Djalma Falcao/Ilha, Almir Da/Márcio André",330945,11100945,0.99,Latin,Sambas De Enredo 2001,Various Artists
Your Mirror,Mick Hucknall,240666,7893821,0.99,Rock,Greatest Hits,Lenny Kravitz

Maybe I'm A Leo,Ian Gillan/Ian Paice/Jon Lord/Ritchie Blckmore/Roger Glover,2
90455,9502646,0.99,Rock,Machine Head,Deep Purple
Not The Doctor,Alanis Morissette & Glenn Ballard,227631,7604601,0.99,Rock,Jag
ged Little Pill,Alanis Morissette
Suck My Kiss,Anthony Kiedis/Chad Smith/Flea/John Frusciante,217234,7129137,0.
99,Alternative & Punk,Blood Sugar Sex Magik,Red Hot Chili Peppers
The Unforgettable Fire,U2,295183,9684664,0.99,Rock,The Best Of 1980-1990,U2
You Learn,Alanis Morissette & Glenn Ballard,239699,7824837,0.99,Rock,Jagged L
ittle Pill,Alanis Morissette
Losing My Religion,Bill Berry/Michael Stipe/Mike Mills/Peter Buck,269035,8885
672,0.99,Alternative & Punk,Out Of Time,R.E.M. Feat. Kate Pearson
"Concert pour 4 Parties de V**les, H. 545: I. Prelude",Marc-Antoine Charpenti
er,110266,1973559,0.99,Classical,"Charpentier: Divertissements, Airs & Concer
ts",Les Arts Florissants & William Christie
Flip The Switch,Jagger/Richards,252421,8336591,0.99,Rock,No Security,The Roll
ing Stones
Enquanto O Mundo Explode,Chico Science,88764,2968650,0.99,Latin,Afrociberdeli
a,Chico Science & Nação Zumbi
So Central Rain,R.E.M.,194768,6414550,0.99,Alternative & Punk,The Best Of R.
E.M.: The IRS Years,R.E.M.
Riviera Paradise,Stevie Ray Vaughan,528692,17232776,0.99,Blues,In Step,Stevie
Ray Vaughan & Double Trouble
Caffeine,,267937,8747367,0.99,Alternative & Punk,Angel Dust,Faith No More
Palco (Live),,238315,8026622,0.99,Latin,Quanta Gente Veio Ver (Live),Gilberto
Gil
Walk On,"Adam Clayton, Bono, Larry Mullen, The Edge",296280,9800861,0.99,Roc
k,All That You Can't Leave Behind,U2
Layla,Clapton/Gordon,430733,14115792,0.99,Blues,The Cream Of Clapton,Eric Cla
pton
End Of The Night,"Robby Krieger, Ray Manzarek, John Densmore, Jim Morrison",1
72695,5589732,0.99,Rock,The Doors,The Doors
Revolution,,256026,8371254,0.99,Rock,"Pure Cult: The Best Of The Cult (For Ro
ckers, Ravers, Lovers & Sinners) [UK]",The Cult
Sweet Child O' Mine,,356424,5879347,0.99,Rock,Appetite for Destruction,Guns
N' Roses
Heaven Can Wait,,440555,10577743,0.99,Rock,Live At Donington 1992 (Disc 2),Ir
on Maiden
Love Comes,"Darius ""Take One"" Minwalla/Jon Auer/Ken Stringfellow/Matt Harri
s",199923,3240609,0.99,Rock,Every Kind of Light,The Posies
The 23rd Psalm,,2610416,487401604,1.99,TV Shows,"Lost, Season 2",Lost
Dude (Looks Like A Lady),"Steven Tyler, Joe Perry, Desmond Child",264855,8679
940,0.99,Rock,Big Ones,Aerosmith
Minas,"Milton Nascimento, Caetano Veloso",152293,4921056,0.99,Latin,Minas,Mil
ton Nascimento
Pride (In The Name Of Love),U2,230243,7549085,0.99,Rock,The Best Of 1980-199
0,U2
Put The Finger On You,"Angus Young, Malcolm Young, Brian Johnson",205662,6713
451,0.99,Rock,For Those About To Rock We Salute You,AC/DC
Get On The Snake,Chris Cornell/Kim Thayil,225123,7313744,0.99,Rock,A-Sides,So
undgarden
A Festa Do Santo Reis,,159791,5204995,0.99,Latin,Serie Sem Limite (Disc 2),Ti
m Maia
Wanted Dread And Alive,,260310,8670933,0.99,Reggae,Greatest Hits,Lenny Kravit
z
Tanto Tempo,Vários,170292,5572240,0.99,Latin,Chill: Brazil (Disc 2),Antônio C
arlos Jobim

Infinite Dreams,Steve Harris,369005,8858669,0.99,Metal,Seventh Son of a Seventh Son,Iron Maiden
Feirinha da Pavuna/Luz do Repente/Bagaço da Laranja,"Arlindo Cruz/Franco/Marquinhos PQD/Negro, Jovelina Pérolo/Zeca Pagodinho",107206,3593684,0.99,Latin,Ao Vivo [IMPORT],Zeca Pagodinho
Motéis,,213498,7041077,0.99,Latin,Roda De Funk,Funk Como Le Gusta
Corduroy,Pearl Jam & Eddie Vedder,305293,9991106,0.99,Rock,Live On Two Legs [Live],Pearl Jam
Why Go,Jeff Ament,200254,6539287,0.99,Rock,Ten,Pearl Jam
"Hail, Hail",Stone Gossard & Eddie Vedder & Jeff Ament & Mike McCready,223764,7364206,0.99,Rock,Live On Two Legs [Live],Pearl Jam
Flight Of The Rat,"Blackmore, Gillan, Glover, Lord, Paice",478302,15563967,0.99,Rock,Deep Purple In Rock,Deep Purple
Groovus Interruptus,Jim Beard,319373,10602166,0.99,Jazz,Outbreak,Dennis Chambers
Wrathchild,Steve Harris,185808,7434368,0.99,Metal,Rock In Rio [CD1],Iron Maiden
Nene 2001,,284969,9694508,0.99,Soundtrack,Carnaval 2001,Various Artists
Nightrain,,268537,4457283,0.99,Rock,Appetite for Destruction,Guns N' Roses
Elderly Woman Behind The Counter In A Small Town,Dave Abbruzzese & Jeff Ament & Stone Gossard & Mike McCready & Eddie Vedder,229328,7509304,0.99,Rock,Live On Two Legs [Live],Pearl Jam
The Outlaw Torn,"James Hetfield, Lars Ulrich",588721,19286261,0.99,Metal,Load,Metallica
Battery,J.Hetfield/L.Ulrich,312424,10229577,0.99,Metal,Master Of Puppets,Metallica
Lords Of The Backstage,"Kelly, Mosley, Rothery, Trewaves",112875,3741319,0.99,Rock,Misplaced Childhood,Marillion
The River (Remix),The Tea Party,343170,11193268,0.99,Alternative & Punk,Tangents,The Tea Party
Bowels Of The Devil,,223216,7324125,0.99,Alternative & Punk,Body Count,Body Count
Don't Go Back To Rockville,R.E.M.,272352,9010715,0.99,Alternative & Punk,The Best Of R.E.M.: The IRS Years,R.E.M.
Macô,Chico Science,249600,8253934,0.99,Latin,Afrociberdelia,Chico Science & Nação Zumbi
Love Is The Colour,R. Carless,251585,8419165,0.99,Jazz,Blue Moods,Incognito
Without You,Van Halen,390295,12619558,0.99,Rock,Van Halen III,Van Halen
Phantom Lord,"James Hetfield, Lars Ulrich, Dave Mustaine",302053,9817143,0.99,Metal,Kill 'Em All,Metallica
Don't Take Your Love From Me,,282331,9244238,0.99,Jazz,Up An' Atom,Gene Krupa
On The Run,"Gilmour, Waters",214595,7206300,0.99,Rock,Dark Side Of The Moon,Pink Floyd
Set It Off,Audioslave/Chris Cornell,263262,6321091,0.99,Rock,Audioslave,Audioslave
Get Me Outta Here,C. Cester/N. Cester,176274,5729098,0.99,Alternative & Punk,Get Born,JET
Surprise! You're Dead!,Faith No More,147226,4823036,0.99,Alternative & Punk,The Real Thing,Faith No More
Linha Do Horizonte,Vários,279484,9275929,0.99,Latin,Chill: Brazil (Disc 2),Antônio Carlos Jobim
Cigano,,304692,10037362,0.99,Latin,Djavan Ao Vivo - Vol. 02,Djavan
She Wears Black,G Harvey/R Hope-Taylor,528666,17617944,0.99,Jazz,Blue Moods,Incognito
Losfer Words,Steve Harris,252891,6074756,0.99,Metal,Powerslave,Iron Maiden
Blood Brothers,Steve Harris,434442,17379456,0.99,Rock,Brave New World,Iron Ma

iden
"War of the Gods, Pt. 2",,2923381,487899692,1.99,Sci Fi & Fantasy,"Battlestar
Galactica (Classic), Season 1",Battlestar Galactica (Classic)
Let There Be Rock,AC/DC,366654,12021261,0.99,Rock,Let There Be Rock,AC/DC
The Negotiation,,1767851,371663719,1.99,Comedy,"The Office, Season 3",The Off
ice
Vida Boa,Fausto Nilo - Armandinho,281730,9411272,0.99,Latin,Prenda Minha,Caet
ano Veloso
Fast And Loose,Clarke/Kilmister/Taylor,203337,6643350,0.99,Metal,Ace Of Spade
s,Motörhead
They're Red Hot,Robert Johnson,71941,2382220,0.99,Alternative & Punk,Blood Su
gar Sex Magik,Red Hot Chili Peppers
Dr. Heckyll & Mr. Jive,,278465,9110403,0.99,Rock,The Best Of Men At Work,Men
At Work
Territorial Pissings,Kurt Cobain,143281,4613880,0.99,Rock,Nevermind,Nirvana
The Pan Piper,Miles Davis,233769,7593713,0.99,Jazz,The Essential Miles Davis
[Disc 1],Miles Davis
Ali,Nando Reis/Samuel Rosa,306390,10110351,0.99,Rock,Maquinarama,Skank
Sangue De Bairro,Chico Science,132231,4415557,0.99,Latin,Afrociberdelia,Chico
Science & Nação Zumbi
Beijo do Olhar,Mônica Marianno,252682,8369029,0.99,World,Demorou...,Mônica Ma
rianno
Pretinha,,265273,8914579,0.99,Jazz,Quanta Gente Veio ver--Bônus De Carnaval,G
ilberto Gil
Fortuneteller,"Blackmore, Glover, Turner, Lord, Paice",349335,11369671,0.99,R
ock,Slaves And Masters,Deep Purple
Surrender,,287634,9422926,0.99,Jazz,Heart of the Night,Spyro Gyra
Killers,Paul Di'Anno/Steve Harris,309995,10009697,0.99,Rock,The Beast Live,Pa
ul D'Ianno
Be Quick Or Be Dead,Bruce Dickinson/Janick Gers,196911,3151872,0.99,Metal,A R
eal Live One,Iron Maiden
Pau-De-Arara,"Guio De Morais E Seus ""Parentes""/Luiz Gonzaga",191660,634064
9,0.99,Soundtrack,As Canções de Eu Tu Eles,Gilberto Gil
On Mercury,"Anthony Kiedis, Flea, John Frusciante, and Chad Smith",208509,683
4762,0.99,Rock,By The Way,Red Hot Chili Peppers
On A Plain,Kurt Cobain,196440,6390635,0.99,Rock,Nevermind,Nirvana
"Lost (Pilot, Part 2)",,2436583,204995876,1.99,TV Shows,"Lost, Season 1",Lost
O Bicho Tá Pregando,,171964,5683369,0.99,Hip Hop/Rap,Os Cães Ladram Mas A Car
avana Não Pára,Planet Hemp
Lookin' For A Reason,J.C. Fogerty,209789,6933135,0.99,Rock,"Chronicle, Vol.
2",Creedence Clearwater Revival
Be Yourself,"Cornell, Commerford, Morello, Wilk",279484,9106160,0.99,Alternat
ive & Punk,Out Of Exile,Audioslave
Não Sei O Que Eu Quero Da Vida,,151849,5024963,0.99,Latin,Cássia Eller - Cole
ção Sem Limite [Disc 2],Cássia Eller
Shamrocks And Shenanigans (Boom Shalock Lock Boom/Butch Vig Mix),E. Schrody/
L. Dimant,237035,9483705,0.99,Hip Hop/Rap,House of Pain,House Of Pain
Talk About The Passion,R.E.M.,203206,6725435,0.99,Alternative & Punk,The Best
Of R.E.M.: The IRS Years,R.E.M.
Rasul,Jeremy Wall,238315,7854737,0.99,Jazz,Morning Dance,Spyro Gyra
Tim Tim Por Tim Tim,Vários,213237,7143328,0.99,Latin,Chill: Brazil (Disc 2),A
ntônio Carlos Jobim
Pini Di Roma (Pinien Von Rom) \ I Pini Della Via Appia,,286741,4718950,0.99,C
lassical,Respighi:Pines of Rome,Eugene Ormandy
Die Die My Darling,Danzig,149315,4867667,0.99,Metal,Garage Inc. (Disc 1),Meta
llica

Enter Sandman,Apocalyptica,221701,7286305,0.99,Metal,Plays Metallica By Four Cellos,Apocalyptica
Harvester Of Sorrow,James Hetfield and Lars Ulrich,345547,11377339,0.99,Metal,...And Justice For All,Metallica
Child In Time (Son Of Aleric - Instrumental),"Richie Blackmore, Ian Gillian, Roger Glover, Jon Lord, Ian Paice",602880,19712753,0.99,Rock,Knocking at Your Back Door: The Best Of Deep Purple in the 80's,Deep Purple
I'm A Greedy Man Pt.1,Charles Bobbitt/James Brown,217730,7251211,0.99,R&B/Soul,Sex Machine,James Brown
The One I Love,R.E.M.,197355,6495125,0.99,Alternative & Punk,The Best Of R.E.M.: The IRS Years,R.E.M.
Onibusfobia,,315977,10474904,0.99,Latin,Jota Quest-1995,Jota Quest
O Estrangeiro,,374700,12472890,0.99,Latin,Minha Historia,Chico Buarque
Voce Nao Entende Nada - Cotidiano,,421982,13885612,0.99,Latin,Minha Historia,Chico Buarque
Planet Home,Jay Kay/Toby Smith,284447,9566237,0.99,R&B/Soul,Synkronized,Jamiroquai
Give Peace a Chance,,274644,4448025,0.99,Pop,Instant Karma: The Amnesty International Campaign to Save Darfur,U2
Flor De Lis,Djavan,236355,7801108,0.99,Latin,Djavan Ao Vivo - Vol. 1,Djavan
Esperando Por Mim,Renato Russo,261668,8844133,0.99,Latin,A TempestadeTempestade Ou O Livro Dos Dias,Legião Urbana
Iron Maiden,Steve Harris,351869,14076032,0.99,Rock,Rock In Rio [CD2],Iron Maiden
Dead Horse,,257600,4394014,0.99,Rock,Use Your Illusion I,Guns N' Roses
Sheer Heart Attack,Taylor,207386,6642685,0.99,Rock,News Of The World,Queen
Light My Fire,"Robby Krieger, Ray Manzarek, John Densmore, Jim Morrison",428329,13963351,0.99,Rock,The Doors,The Doors
Chão de Giz (Elba Ramalho),,274834,9016916,0.99,Latin,Vozes do MPB,Various Artists
Waiting On A Sign,The Tea Party,261903,8558590,0.99,Alternative & Punk,Tangents,The Tea Party
Wildest Dreams,Adrian Smith/Steve Harris,232777,9312384,0.99,Heavy Metal,Dance Of Death,Iron Maiden
Murders In The Rue Morgue,Steve Harris,258377,6205786,0.99,Heavy Metal,Killers,Iron Maiden
Pot-Pourri N.º 4,,392437,13125975,0.99,Bossa Nova,Vinícius De Moraes - Sem Limite,Toquinho & Vinícius
The Witch,,258768,8725403,0.99,Rock,"Pure Cult: The Best Of The Cult (For Rockers, Ravers, Lovers & Sinners) [UK]",The Cult
Sobremesa,Chico Science,240091,7960868,0.99,Latin,Afrociberdelia,Chico Science & Nação Zumbi
Outbreak,Jim Beard & Jon Herington,659226,21685807,0.99,Jazz,Outbreak,Dennis Chambers
Construção / Deus Lhe Pague,,383059,12675305,0.99,Latin,Minha Historia,Chico Buarque
My Funny Valentine (Live),Miles Davis,907520,29416781,0.99,Jazz,The Essential Miles Davis [Disc 2],Miles Davis
Angel,Carolyn Franklin and Sonny Saunders,240561,7880256,0.99,Rock,Greatest Hits,Lenny Kravitz
Heaven Can Wait,Steve Harris,441417,10596431,0.99,Metal,Somewhere in Time,Iron Maiden
Company Man,,2601101,503786316,1.99,Drama,"Heroes, Season 1",Heroes
Perfect,Billy Corgan,203023,6734636,0.99,Alternative & Punk,Rotten Apples: Greatest Hits,Smashing Pumpkins
Sleep On The Sidewalk,May,187428,6099840,0.99,Rock,News Of The World,Queen

Domingo,Sérgio Britto/Toni Bellotto,208613,6883180,0.99,Alternative & Punk,Volume Dois,Titãs
Seu Balancê,Paulinho Rezende/Toninho Geraes,219454,7311219,0.99,Latin,Ao Vivo [IMPORT],Zeca Pagodinho
Alive,Stone Gossard,341080,11176623,0.99,Rock,Ten,Pearl Jam
Get Up (I Feel Like Being A) Sex Machine,Bobby Byrd/James Brown/Ron Lenhoff,316551,10498031,0.99,R&B/Soul,Sex Machine,James Brown
Blood Brothers,Steve Harris,435513,17422464,0.99,Metal,Rock In Rio [CD1],Iron Maiden
Runnin' With The Devil,"Edward Van Halen, Alex Van Halen, David Lee Roth, Michael Anthony",215902,7061901,0.99,Rock,"The Best Of Van Halen, Vol. I",Van Halen
Milk It,Kurt Cobain,225724,7406945,0.99,Rock,From The Muddy Banks Of The Wishkah [Live],Nirvana
Gone,Eddie Vedder,249547,8158204,0.99,Alternative & Punk,Pearl Jam,Pearl Jam
When You Gonna Learn (Digeridoo),"Jay Kay/Kay, Jay",230635,7655482,0.99,Rock,Emergency On Planet Earth,Jamiroquai
Angela,,169508,5574957,0.99,Jazz,Warner 25 Anos,Antônio Carlos Jobim
Locomotive,Slash/W. Axl Rose,522396,17236842,0.99,Metal,Use Your Illusion II,Guns N' Roses
The Angel And The Gambler,Steve Harris,592744,23711872,0.99,Rock,Virtual XI,Iron Maiden
Bye Bye Blackbird,Miles Davis,476003,15549224,0.99,Jazz,The Essential Miles Davis [Disc 1],Miles Davis
Angel Of Harlem,"Bono/Clayton, Adam/Mullen Jr., Larry/The Edge",229276,7498022,0.99,Rock,Rattle And Hum,U2
Hello Mary Lou,J.C. Fogerty,132832,4476563,0.99,Rock,"Chronicle, Vol. 2",Creedence Clearwater Revival
Bushleager,Stone Gossard,237479,7849757,0.99,Rock,Riot Act,Pearl Jam
The Worst,Jagger/Richards,144613,4750094,0.99,Rock,Voodoo Lounge,The Rolling Stones
Saint Of Me,Jagger/Richards,325694,10725160,0.99,Rock,No Security,The Rolling Stones
Dirty Water Dog,Van Halen,327392,10709202,0.99,Rock,Van Halen III,Van Halen
Times of Trouble,,342539,5795951,0.99,Alternative,Temple of the Dog,Temple of the Dog
Jerusalem,Roy Z,402390,13194463,0.99,Metal,Chemical Wedding,Bruce Dickinson
Pretty Noose,Chris Cornell,253570,8317931,0.99,Rock,A-Sides,Soundgarden
Bê-a-Bá,Rodolfo,249051,8130636,0.99,Alternative & Punk,Cesta Básica,Raimundos
Esquinas,,280999,9096726,0.99,Latin,Djavan Ao Vivo - Vol. 02,Djavan
Wasting Love,,336953,8091301,0.99,Rock,Live At Donington 1992 (Disc 1),Iron Maiden
Hot Dog,Jimmy Page & Robert Plant,197198,6536167,0.99,Rock,In Through The Out Door,Led Zeppelin
As Rosas Não Falam (Beth Carvalho),,116767,3836641,0.99,Latin,Vozes do MPB,Various Artists
Boogie Blues,,204199,6603153,0.99,Jazz,Up An' Atom,Gene Krupa
Numbers,,2609772,214709143,1.99,TV Shows,"Lost, Season 1",Lost
Beach Games,,1676134,333671149,1.99,Comedy,"The Office, Season 3",The Office
Maracatu Atômico,Chico Science,284264,9670057,0.99,Latin,Afrociberdelia,Chico Science & Nação Zumbi
Eruption,"Edward Van Halen, Alex Van Halen, Michael Anthony and David Lee Roth",102556,3286026,0.99,Rock,Van Halen,Van Halen
Tempo Perdido,Renato Russo,302158,9963914,0.99,Latin,Mais Do Mesmo,Legião Urbana
You Got No Right,"Dave Kushner, Duff, Matt Sorum, Scott Weiland & Slash",3354

```
12,10991094,0.99,Rock,Contraband,Velvet Revolver
Wasting Love,Bruce Dickinson/Janick Gers,350981,14041216,0.99,Rock,Fear Of Th
e Dark,Iron Maiden
Somebody Stole My Guitar,"Ian Gillan, Roger Glover, Jon Lord, Steve Morse, Ia
n Paice",249443,8180421,0.99,Rock,Purpendicular,Deep Purple
'Round Midnight,Miles Davis,357459,11590284,0.99,Jazz,The Essential Miles Dav
is [Disc 1],Miles Davis
Aquilo,,246073,8167819,0.99,Latin,Lulu Santos - RCA 100 Anos De Música - Álbu
m 02,Lulu Santos
Flash,"May, Brian",168489,5464986,0.99,Rock,Greatest Hits I,Queen
Smooth,M. Itaal Shur & Rob Thomas,298161,9867455,0.99,Rock,Supernatural,Santa
na
Childhoods End?,"Kelly, Mosley, Rothery, Trewaves",272796,9015366,0.99,Rock,M
isplaced Childhood,Marillion
Dream Of Mirrors,Janick Gers/Steve Harris,561162,22448256,0.99,Rock,Brave New
World,Iron Maiden
Royal Orleans,John Bonham/John Paul Jones,179591,5930027,0.99,Rock,Presence,L
ed Zeppelin
The Wicker Man,Adrian Smith/Bruce Dickinson/Steve Harris,281782,11272320,0.9
9,Metal,Rock In Rio [CD1],Iron Maiden
It's Too Funky In Here,Brad Shapiro/George Jackson/Robert Miller/Walter Shaw,
239072,7973979,0.99,R&B/Soul,Sex Machine,James Brown
Night Flight,John Paul Jones/Robert Plant,217547,7160647,0.99,Rock,Physical G
raffiti [Disc 2],Led Zeppelin
Won't Get Fooled Again (Full Length Version),John Entwistle/Pete Townshend,51
3750,16855521,0.99,Rock,My Generation - The Very Best Of The Who,The Who
She Loves Me Not,Bill Gould/Mike Bordin/Mike Patton,209867,6887544,0.99,Alter
native & Punk,Album Of The Year,Faith No More
Beth,"Peter Criss, Stan Penridge, Bob Ezrin",170187,5577807,0.99,Rock,Unplugg
ed [Live],Kiss
Exploder,Audioslave/Chris Cornell,206053,4948095,0.99,Rock,Audioslave,Audiosl
ave
Os Cegos Do Castelo,Titãs,296829,9868187,0.99,Alternative & Punk,Acústico,Tit
ãs
The Star Spangled Banner,"Hendrix, Jimi",43232,1385810,0.99,Rock,Rattle And H
um,U2
```

In [5]:
```python
import pandas as pd
```

In [ ]:
```python
csv_df = pd.read_csv("itunes_data.csv")
csv_df.head()
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	
0	All the Best Cowboys Have Daddy Issues	NaN	2555492	211743651	1.99	TV Shows	Lost, Season 1	
1	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged	C
2	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo	Nasci
3	Ben Franklin	NaN	1271938	264168080	1.99	Comedy	The Office, Season 3	The
4	O Último Romântico (Ao Vivo)	NaN	231993	7692697	0.99	Latin	Lulu Santos - RCA 100 Anos De Música - Álbum 02	Lulu

```python
excel_df = pd.read_excel("itunes_data.xlsx", engine="openpyxl")
excel_df.head()
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	Art
0	My Time After Awhile	Robert Geddins/Ron Badger/Sheldon Feinberg	182491	6022698	0.99	Blues	The Best Of Buddy Guy - The Millenium Collection	Bud G
1	Be Quick Or Be Dead	Bruce Dickinson/Janick Gers	204512	8181888	0.99	Rock	Fear Of The Dark	Ir Maid
2	Água E Fogo	Chico Amaral/Edgard Scandurra/Samuel Rosa	278987	9272272	0.99	Rock	Maquinarama	Ska
3	Ozone Baby	Jimmy Page, Robert Plant	215954	7079588	0.99	Rock	Coda	L Zeppe
4	Bop Boogie	NaN	189596	6093124	0.99	Jazz	Up An' Atom	Ge Kru

```python
from sqlalchemy import create_engine

engine = create_engine("sqlite:///chinook.db")
```

```python
query = """SELECT tracks.name as Track, tracks.composer, tracks.milliseconds
tracks.bytes, tracks.unitprice,
genres.name as Genre,
albums.title as Album,
artists.name as Artist
FROM tracks
JOIN genres ON tracks.genreid = genres.genreid
JOIN albums ON tracks.albumid = albums.albumid
JOIN artists ON albums.artistid = artists.artistid;
"""
```

```python
with engine.connect() as connection:
    sql_df = pd.read_sql_query(query, connection)
```

```python
sql_df.head(2).T
```

Out[12]:

	0	1
Track	For Those About To Rock (We Salute You)	Put The Finger On You
Composer	Angus Young, Malcolm Young, Brian Johnson	Angus Young, Malcolm Young, Brian Johnson
Milliseconds	343719	205662
Bytes	11170334	6713451
UnitPrice	0.99	0.99
Genre	Rock	Rock
Album	For Those About To Rock We Salute You	For Those About To Rock We Salute You
Artist	AC/DC	AC/DC

```python
# create dataframe from lists
df = pd.DataFrame(data={"seconds": [1, 2, 3, 4], "intensity": [12, 11, 12, 1
df.head()
```

Out[ ]:

	seconds	intensity
0	1	12
1	2	11
2	3	12
3	4	14

Understanding the DataFrame Structure and Combining/Concatenating Multiple DataFrames

```
In [14]: sql_df.index
```

Out[14]: RangeIndex(start=0, stop=3503, step=1)

```
In [15]: sql_df.columns
```

Out[15]: Index(['Track', 'Composer', 'Milliseconds', 'Bytes', 'UnitPrice', 'Genre',
               'Album', 'Artist'],
              dtype='object')

```
In [16]: type(sql_df)
```

Out[16]: pandas.core.frame.DataFrame

```
In [17]: itunes_df = pd.concat([csv_df, excel_df, sql_df], axis=0)
         itunes_df.head()
```

Out[17]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	
0	All the Best Cowboys Have Daddy Issues	NaN	2555492	211743651	1.99	TV Shows	Lost, Season 1	
1	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged	C
2	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo	Nasci
3	Ben Franklin	NaN	1271938	264168080	1.99	Comedy	The Office, Season 3	The
4	O Último Romântico (Ao Vivo)	NaN	231993	7692697	0.99	Latin	Lulu Santos - RCA 100 Anos De Música - Álbum 02	Lulu

Exploratory Data Analysis (EDA) and Basic Data Cleaning with pandas

```
In [18]: itunes_df.tail()
```

Out[18]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album
3498	Pini Di Roma (Pinien Von Rom) \ I Pini Della V...	None	286741	4718950	0.99	Classical	Respighi:Pine of Rom...
3499	String Quartet No. 12 in C Minor, D. 703 "Quar...	Franz Schubert	139200	2283131	0.99	Classical	Schubert: The Late String Quartets & String Qu...
3500	L'orfeo, Act 3, Sinfonia (Orchestra)	Claudio Monteverdi	66639	1189062	0.99	Classical	Monteverdi L'Orfeo
3501	Quintet for Horn, Violin, 2 Violas, and Cello ...	Wolfgang Amadeus Mozart	221331	3665114	0.99	Classical	Mozart Chamber Music
3502	Koyaanisqatsi	Philip Glass	206005	3305164	0.99	Soundtrack	Koyaanisqatsi (Soundtrack from the Motion Pict...

In [19]:
```python
print(itunes_df.iloc[0])
print(itunes_df.iloc[-1])
```

```
Track           All the Best Cowboys Have Daddy Issues
Composer                                           NaN
Milliseconds                                   2555492
Bytes                                        211743651
UnitPrice                                         1.99
Genre                                         TV Shows
Album                                    Lost, Season 1
Artist                                            Lost
Name: 0, dtype: object
Track                                    Koyaanisqatsi
Composer                                  Philip Glass
Milliseconds                                    206005
Bytes                                          3305164
UnitPrice                                         0.99
Genre                                       Soundtrack
Album           Koyaanisqatsi (Soundtrack from the Motion Pict...
Artist                            Philip Glass Ensemble
Name: 3502, dtype: object
```

In [20]:
```python
itunes_df.iloc[0, 0]
```

Out[20]: 'All the Best Cowboys Have Daddy Issues'

```
In [21]:  itunes_df.iloc[-1, -1]
```

```
Out[21]:  'Philip Glass Ensemble'
```

```
In [22]:  itunes_df.loc[3502]
```

```
Out[22]:  Track                                      Koyaanisqatsi
          Composer                                    Philip Glass
          Milliseconds                                      206005
          Bytes                                            3305164
          UnitPrice                                           0.99
          Genre                                         Soundtrack
          Album           Koyaanisqatsi (Soundtrack from the Motion Pict...
          Artist                              Philip Glass Ensemble
          Name: 3502, dtype: object
```

```
In [ ]:   test_df = itunes_df.copy()
          test_df = pd.concat(
              [pd.DataFrame([itunes_df.loc[3502]], columns=test_df.columns), test_df],
              ignore_index=True,
          )
          test_df.loc[3502]
```

```
Out[ ]:   Track           Hallelujah Here She Comes
          Composer                               U2
          Milliseconds                       242364
          Bytes                             8027028
          UnitPrice                            0.99
          Genre                                Rock
          Album                  B-Sides 1980-1990
          Artist                                 U2
          Name: 3502, dtype: object
```

```
In [26]:  test_df.reset_index(inplace=True, drop=True)
          test_df.head()
```

Out[26]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album
0	Koyaanisqatsi	Philip Glass	206005	3305164	0.99	Soundtrack	Koyaanisqatsi (Soundtrack from the Motion Pict...
1	All the Best Cowboys Have Daddy Issues	NaN	2555492	211743651	1.99	TV Shows	Lost, Season 1
2	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged
3	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo
4	Ben Franklin	NaN	1271938	264168080	1.99	Comedy	The Office, Season 3

In [27]: `test_df.index`

Out[27]: `RangeIndex(start=0, stop=4022, step=1)`

In [28]: `itunes_df.shape`

Out[28]: `(4021, 8)`

In [29]: `itunes_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 4021 entries, 0 to 3502
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Track         4021 non-null   object
 1   Composer      2908 non-null   object
 2   Milliseconds  4021 non-null   int64
 3   Bytes         4021 non-null   int64
 4   UnitPrice     4021 non-null   float64
 5   Genre         4021 non-null   object
 6   Album         4021 non-null   object
 7   Artist        4021 non-null   object
dtypes: float64(1), int64(2), object(5)
memory usage: 411.8+ KB
```

In [30]: `itunes_df.isna().sum()`

Out[30]: Track   0
          Composer   1113
          Milliseconds   0
          Bytes   0
          UnitPrice   0
          Genre   0
          Album   0
          Artist   0
          dtype: int64

In [31]: `type(itunes_df.isna())`

Out[31]: `pandas.core.frame.DataFrame`

In [32]: `itunes_df.describe()`

Out[32]:

	Milliseconds	Bytes	UnitPrice
count	4.021000e+03	4.021000e+03	4021.000000
mean	3.927276e+05	3.311048e+07	1.050184
std	5.337745e+05	1.042268e+08	0.237857
min	1.071000e+03	3.874700e+04	0.990000
25%	2.069680e+05	6.372433e+06	0.990000
50%	2.554770e+05	8.102839e+06	0.990000
75%	3.217240e+05	1.025143e+07	0.990000
max	5.286953e+06	1.059546e+09	1.990000

In [ ]: `itunes_df["Genre"].mode()`

Out[ ]: 0   Rock
         Name: Genre, dtype: object

In [ ]: `itunes_df["Genre"].value_counts()`

Out[ ]: Genre
         Rock   1498
         Latin   656
         Metal   420
         Alternative & Punk   393
         Jazz   160
         TV Shows   105
         Blues   92
         Classical   85
         Drama   71
         R&B/Soul   69
         Reggae   64
         Pop   51
         Soundtrack   50
         Alternative   45
         Hip Hop/Rap   40
         Electronica/Dance   35
         World   32
         Heavy Metal   31
         Sci Fi & Fantasy   31
         Easy Listening   28
         Comedy   20
         Bossa Nova   17
         Science Fiction   15
         Rock And Roll   12
         Opera   1
         Name: count, dtype: int64

In [ ]:
```python
itunes_df["Genre"].value_counts()[:5]
```

Out[ ]: Genre
         Rock   1498
         Latin   656
         Metal   420
         Alternative & Punk   393
         Jazz   160
         Name: count, dtype: int64

In [ ]:
```python
itunes_df["Artist"].unique().shape
```

Out[ ]: (204,)

In [39]:
```python
import numpy as np

itunes_df.select_dtypes(include=np.number).corr()
```
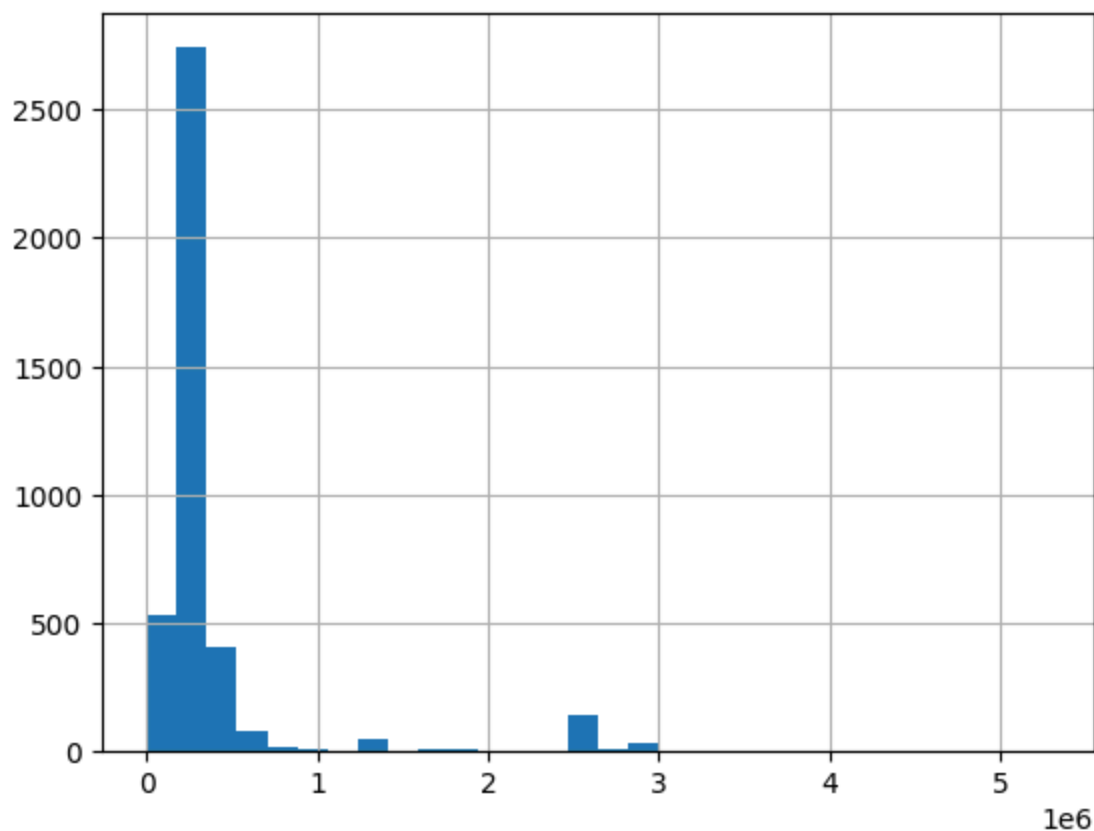
Out[39]:

|	Milliseconds	Bytes	UnitPrice
Milliseconds	1.000000	0.957791	0.934829
Bytes	0.957791	1.000000	0.938734
UnitPrice	0.934829	0.938734	1.000000

Plotting with DataFrames

```
In [40]:  import matplotlib.pyplot as plt
```

```
In [ ]:  itunes_df["Milliseconds"].hist(bins=30)
         plt.show()
```



```
In [ ]:  # this cell is for saving the image; more on this is chapter 5
         f = plt.figure(
             figsize=(5.5, 5.5)
         )   # this changes the size of the image -- more on this is chapter 5
         f.patch.set_facecolor("w")   # sets background color behind axis labels
         itunes_df["Milliseconds"].hist(bins=30)
         plt.tight_layout()   # auto-adjust margins
```
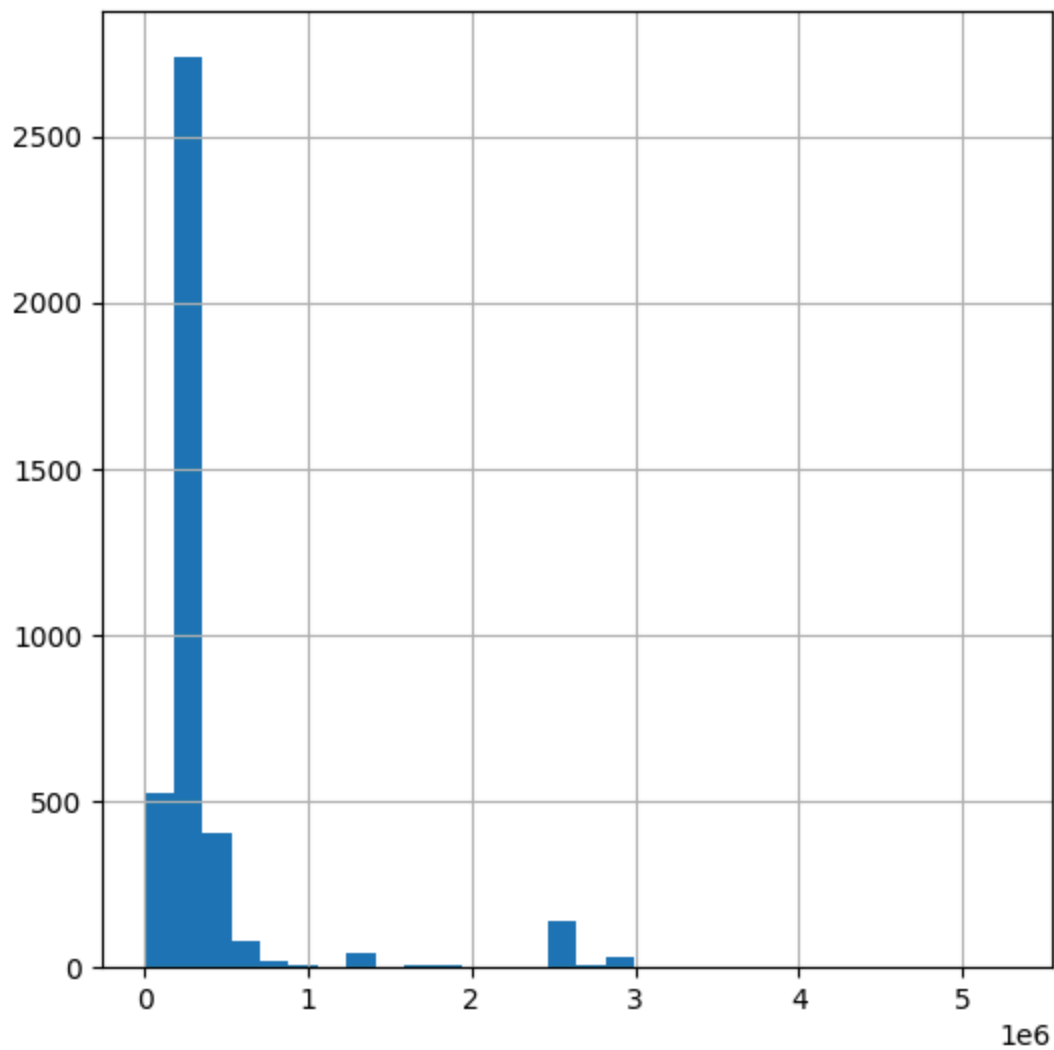
```
In [ ]:  # figsize increases the size of the image -- more on this is chapter 5
         itunes_df.plot.scatter(x="Milliseconds", y="Bytes", figsize=(8, 8))
         plt.show()
```
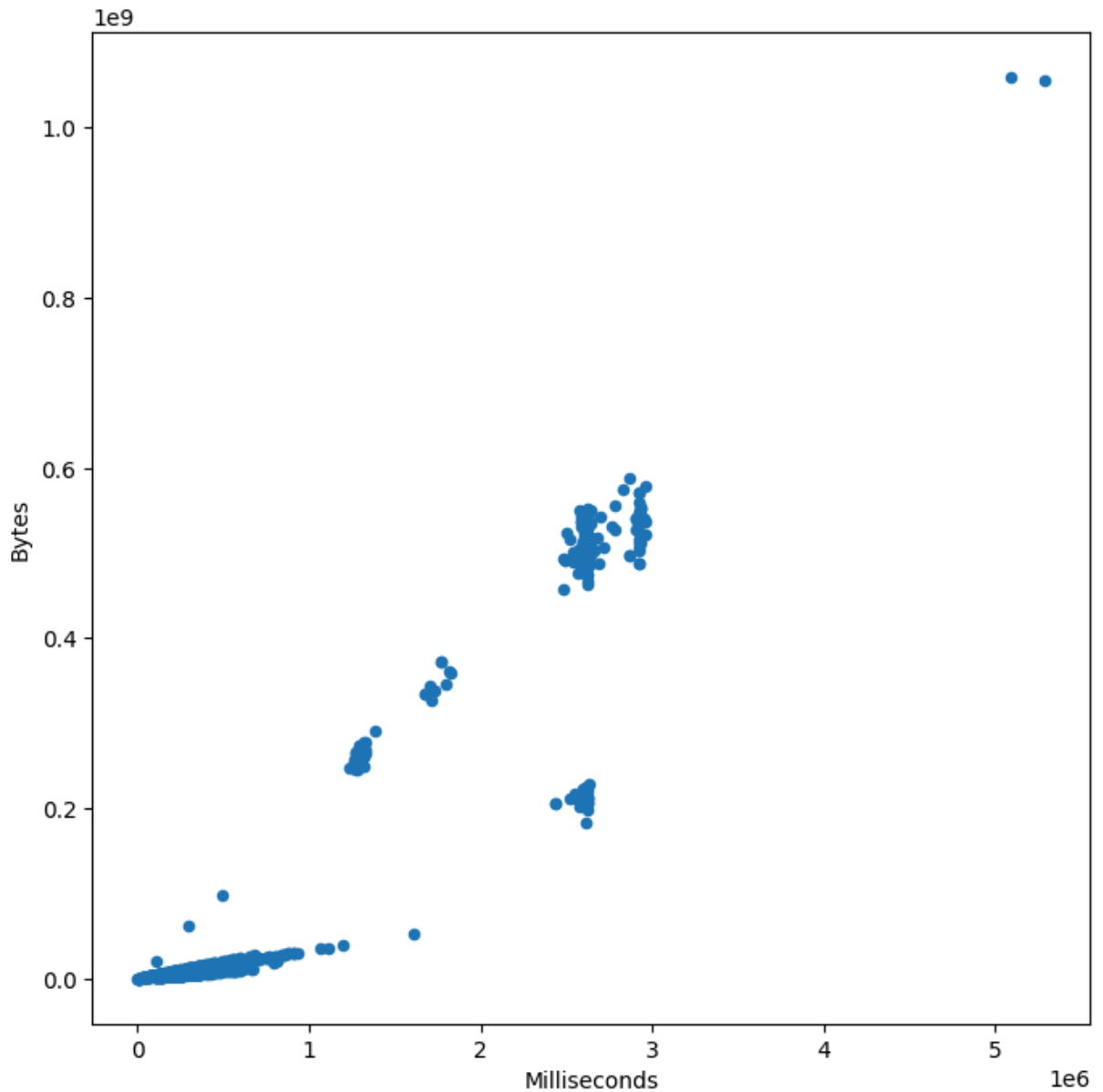
```
In [ ]:  # saving the image -- more on this in chapter 5
         f = plt.figure()
         itunes_df.plot.scatter(x="Milliseconds", y="Bytes", figsize=(5.5, 5.5))
         f.patch.set_facecolor("w")  # sets background color behind axis labels
         plt.tight_layout()  # auto-adjust margins
```

<Figure size 640x480 with 0 Axes>
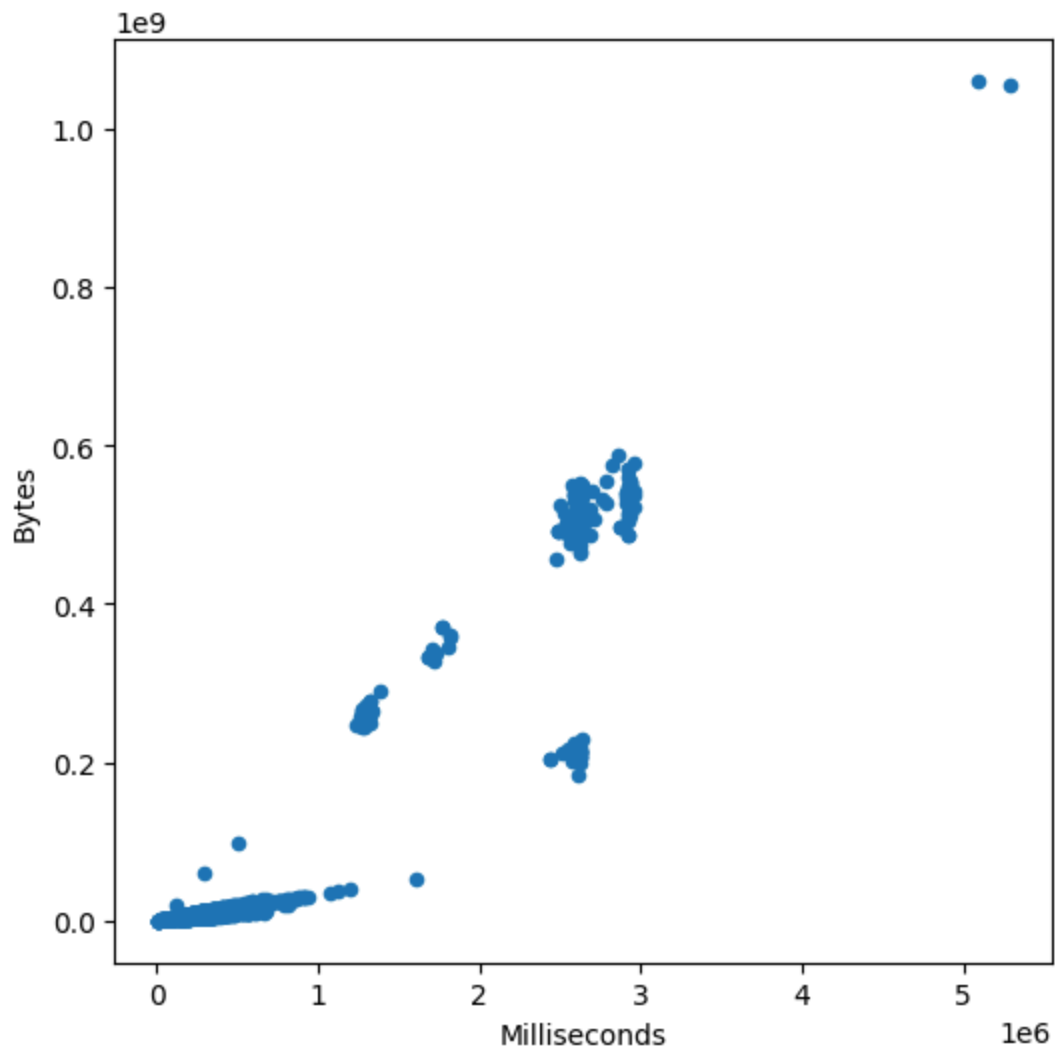
```
In [ ]:  itunes_df["Genre"].value_counts().plot.bar()
         plt.show()
```

```
In [ ]:  # saving the image -- more on this in chapter 5
         f = plt.figure(figsize=(5.5, 5.5))
         itunes_df["Genre"].value_counts().plot.bar()
         f.patch.set_facecolor("w")   # sets background color behind axis labels
         plt.tight_layout()   # auto-adjust margins
```
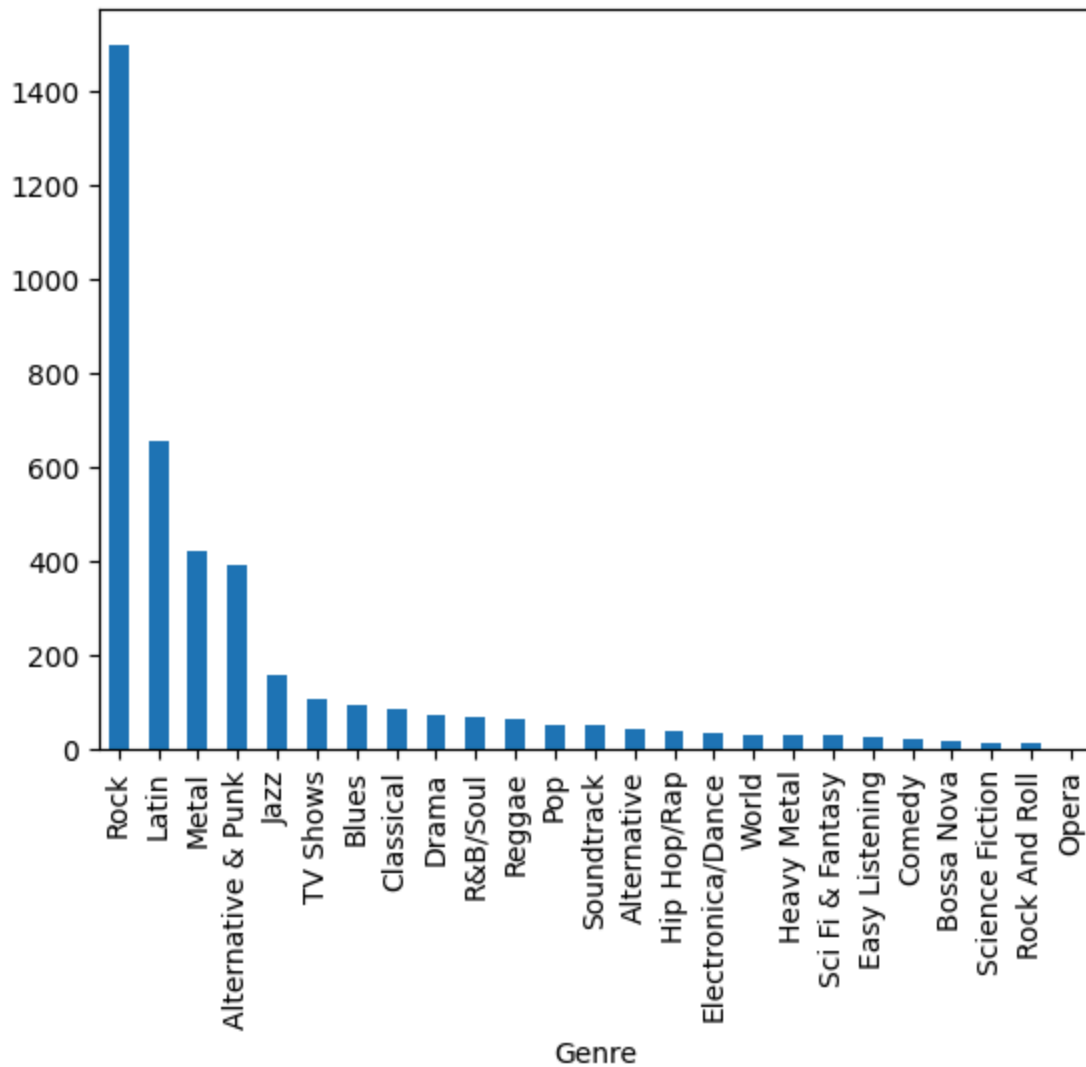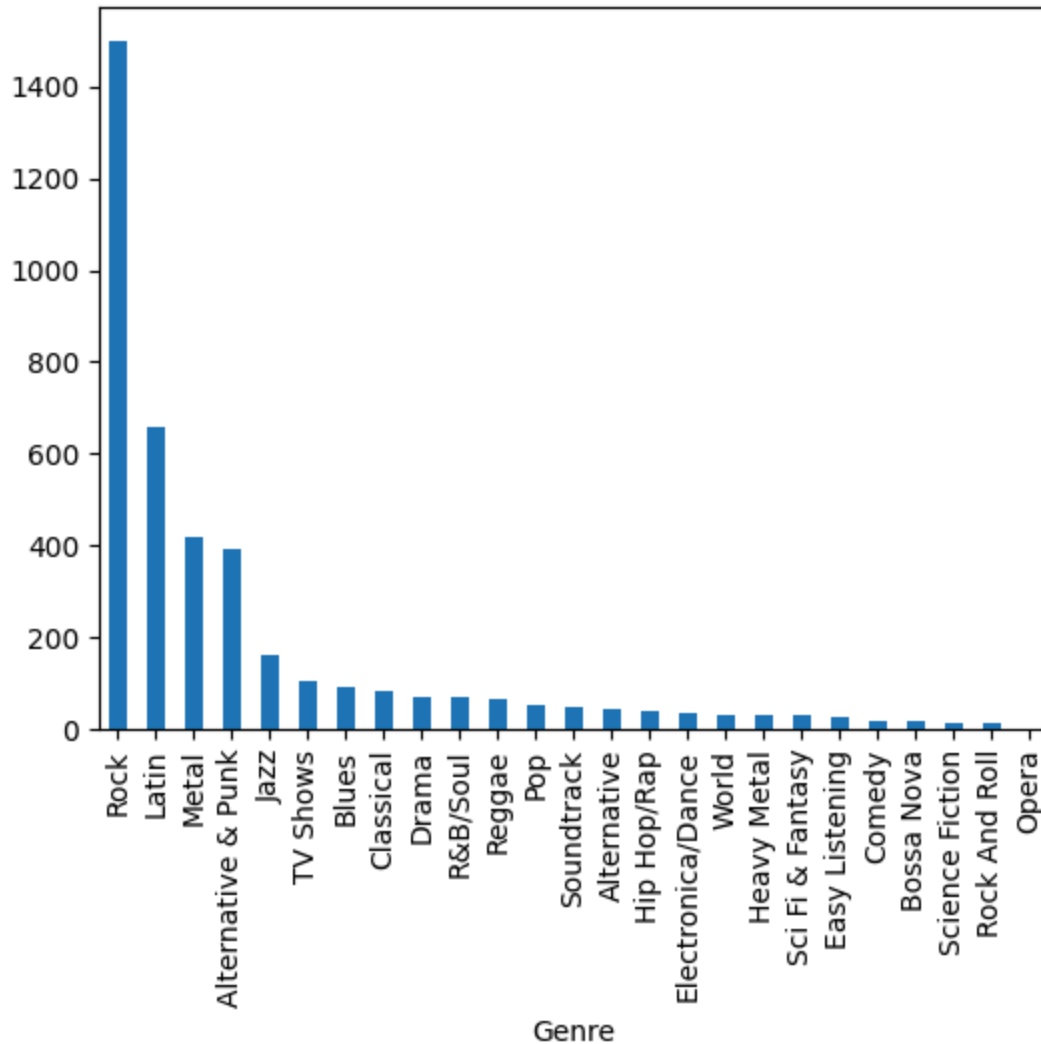
Cleaning Data, Filtering DataFrames

```
In [ ]:  itunes_df[itunes_df["Milliseconds"] > 4e6]
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	A
2833	Occupation / Precipice	None	5286953	1054423946	1.99	TV Shows	Battlestar Galactica, Season 3	Battle Gala
2898	Through a Looking Glass	None	5088838	1059546140	1.99	Drama	Lost, Season 3	

```
In [ ]:  print(itunes_df[itunes_df["Milliseconds"] > 4e6])
```

```
                      Track Composer  Milliseconds       Bytes  UnitPrice
\
2833   Occupation / Precipice       None       5286953  1054423946       1.99
2898   Through a Looking Glass      None       5088838  1059546140       1.99

                Genre                        Album                  Artist
2833   TV Shows  Battlestar Galactica, Season 3  Battlestar Galactica
2898      Drama                   Lost, Season 3                    Lost
```

In [ ]: `print(itunes_df[itunes_df["Milliseconds"] > 4e6][["Genre", "Artist"]])`

```
                Genre                  Artist
2833   TV Shows  Battlestar Galactica
2898      Drama                   Lost
```

In [ ]: `itunes_df["Milliseconds"] > 4e6`

Out[ ]:
```
0       False
1       False
2       False
3       False
4       False
         ...
3498    False
3499    False
3500    False
3501    False
3502    False
Name: Milliseconds, Length: 4021, dtype: bool
```

In [ ]: `itunes_df[itunes_df["Milliseconds"] > 2e6]["Genre"].value_counts()`

Out[ ]:
```
Genre
Drama              69
TV Shows           67
Sci Fi & Fantasy   31
Science Fiction    15
Comedy              2
Name: count, dtype: int64
```

In [ ]: `itunes_df[(itunes_df["Milliseconds"] > 2e6) & (itunes_df["Bytes"] < 0.4e9)][`
        `    "Genre"`
        `].value_counts()`

Out[ ]:
```
Genre
TV Shows    32
Name: count, dtype: int64
```

In [ ]: `itunes_df[itunes_df["Genre"] != "TV Shows"]["Genre"].value_counts()`

Out[ ]:   Genre
           Rock   1498
           Latin   656
           Metal   420
           Alternative & Punk   393
           Jazz   160
           Blues   92
           Classical   85
           Drama   71
           R&B/Soul   69
           Reggae   64
           Pop   51
           Soundtrack   50
           Alternative   45
           Hip Hop/Rap   40
           Electronica/Dance   35
           World   32
           Heavy Metal   31
           Sci Fi & Fantasy   31
           Easy Listening   28
           Comedy   20
           Bossa Nova   17
           Science Fiction   15
           Rock And Roll   12
           Opera   1
           Name: count, dtype: int64

In [ ]:
```python
itunes_df[~(itunes_df["Genre"] == "TV Shows")]["Genre"].value_counts()
```

Out[ ]:   Genre
           Rock   1498
           Latin   656
           Metal   420
           Alternative & Punk   393
           Jazz   160
           Blues   92
           Classical   85
           Drama   71
           R&B/Soul   69
           Reggae   64
           Pop   51
           Soundtrack   50
           Alternative   45
           Hip Hop/Rap   40
           Electronica/Dance   35
           World   32
           Heavy Metal   31
           Sci Fi & Fantasy   31
           Easy Listening   28
           Comedy   20
           Bossa Nova   17
           Science Fiction   15
           Rock And Roll   12
           Opera   1
           Name: count, dtype: int64

```
In [ ]: itunes_df[itunes_df["Genre"].str.contains("TV")]["Genre"].value_counts()
```

```
Out[ ]: Genre
        TV Shows      105
        Name: count, dtype: int64
```

```
In [ ]: itunes_df_copy = itunes_df.copy()  # make a backup to test dropping a column
        itunes_df_copy.drop("Composer", axis=1, inplace=True)
        itunes_df_copy.columns
```

```
Out[ ]: Index(['Track', 'Milliseconds', 'Bytes', 'UnitPrice', 'Genre', 'Album',
               'Artist'],
              dtype='object')
```

```
In [ ]: only_music = itunes_df[
            ~itunes_df["Genre"].isin(
                ["Drama", "TV Shows", "Sci Fi & Fantasy", "Science Fiction", "Comedy"
            )
        ]
```

Missing values

```
In [ ]: itunes_df[itunes_df["Composer"].isna()].sample(5, random_state=42).head()
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	
2947	Three Minutes	None	2763666	531556853	1.99	TV Shows	S
98	Romance Ideal	NaN	250070	8260477	0.99	Latin	Arq Par S
3327	Todo o Carnaval tem seu Fim	None	237426	9499371	0.99	Electronica/Dance	B Van
2457	We've Got To Get Together/Jingo	None	1070027	34618222	0.99	Rock	S
587	Vamos Dançar	None	226194	7617432	0.99	Latin	The E

```
In [59]: itunes_df_copy = itunes_df.copy()  # make a backup for testing dropping a co
         # print shapes to check that rows were dropped
         print(itunes_df_copy.shape)
         itunes_df_copy.dropna(inplace=True)
         print(itunes_df_copy.shape)
```

```
(4021, 8)
(2908, 8)
```

In [ ]:
```python
itunes_df_copy = itunes_df.copy()  # make a backup for testing dropping a co
itunes_df_copy.loc[itunes_df["Composer"].isna(), "Composer"] = "Unknown"
itunes_df_copy.head()
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	
0	All the Best Cowboys Have Daddy Issues	Unknown	2555492	211743651	1.99	TV Shows	Lost, Season 1	
1	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged	C
2	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo	Nasci
3	Ben Franklin	Unknown	1271938	264168080	1.99	Comedy	The Office, Season 3	The
4	O Último Romântico (Ao Vivo)	Unknown	231993	7692697	0.99	Latin	Lulu Santos - RCA 100 Anos De Música - Álbum 02	Lulu

In [ ]:
```python
itunes_df_copy = itunes_df.copy()  # make a backup for testing dropping a co
itunes_df_copy["Composer"].fillna("Unknown", inplace=True)
itunes_df_copy.head()
```

```
/tmp/ipykernel_1089788/888296175.py:2: FutureWarning: A value is trying to be
set on a copy of a DataFrame or Series through chained assignment using an in
place method.
The behavior will change in pandas 3.0. This inplace method will never work b
ecause the intermediate object on which we are setting values always behaves
as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.
method({col: value}, inplace=True)' or df[col] = df[col].method(value) instea
d, to perform the operation inplace on the original object.


  itunes_df_copy['Composer'].fillna('Unknown', inplace=True)
```

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	
0	All the Best Cowboys Have Daddy Issues	Unknown	2555492	211743651	1.99	TV Shows	Lost, Season 1	
1	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged	C
2	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo	Nasci
3	Ben Franklin	Unknown	1271938	264168080	1.99	Comedy	The Office, Season 3	The
4	O Último Romântico (Ao Vivo)	Unknown	231993	7692697	0.99	Latin	Lulu Santos - RCA 100 Anos De Música - Álbum 02	Lulu

In [ ]:
```python
itunes_df["UnitPrice"].value_counts().iloc[0] / itunes_df.shape[0]
```

Out[ ]: np.float64(0.9398159661775678)

In [ ]:
```python
itunes_df_copy = itunes_df.copy()  # make a backup for testing dropping a co
itunes_df_copy["UnitPrice"].fillna(itunes_df_copy["UnitPrice"].mode(), inpla
itunes_df_copy.head()
```

/tmp/ipykernel_1089788/2348392847.py:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

 itunes_df_copy['UnitPrice'].fillna(itunes_df_copy['UnitPrice'].mode(), inplace=True)

Out[ ]:

	Track	Composer	Milliseconds	Bytes	UnitPrice	Genre	Album	
0	All the Best Cowboys Have Daddy Issues	NaN	2555492	211743651	1.99	TV Shows	Lost, Season 1	
1	Beira Mar	Gilberto Gil	295444	9597994	0.99	Latin	Unplugged	C
2	Brasil	Milton Nascimento, Fernando Brant	155428	5252560	0.99	Latin	Milton Nascimento Ao Vivo	Nasci
3	Ben Franklin	NaN	1271938	264168080	1.99	Comedy	The Office, Season 3	The
4	O Último Romântico (Ao Vivo)	NaN	231993	7692697	0.99	Latin	Lulu Santos - RCA 100 Anos De Música - Álbum 02	Lulu

In [ ]:
```python
# create distributions for plotting
from scipy.stats import skewnorm

df = pd.DataFrame(
    {"normal": skewnorm.rvs(0, size=10000), "skewed": skewnorm.rvs(10, size=
)

df.plot.hist(bins=30, alpha=0.5, figsize=(5.5, 5.5))
plt.vlines(df["skewed"].mean(), ymin=0, ymax=2000, label="mean", color="r")
plt.vlines(df["skewed"].median(), ymin=0, ymax=2000, label="median", color="
f.patch.set_facecolor("w")  # sets background color behind axis labels
plt.tight_layout()  # auto-adjust margins
```
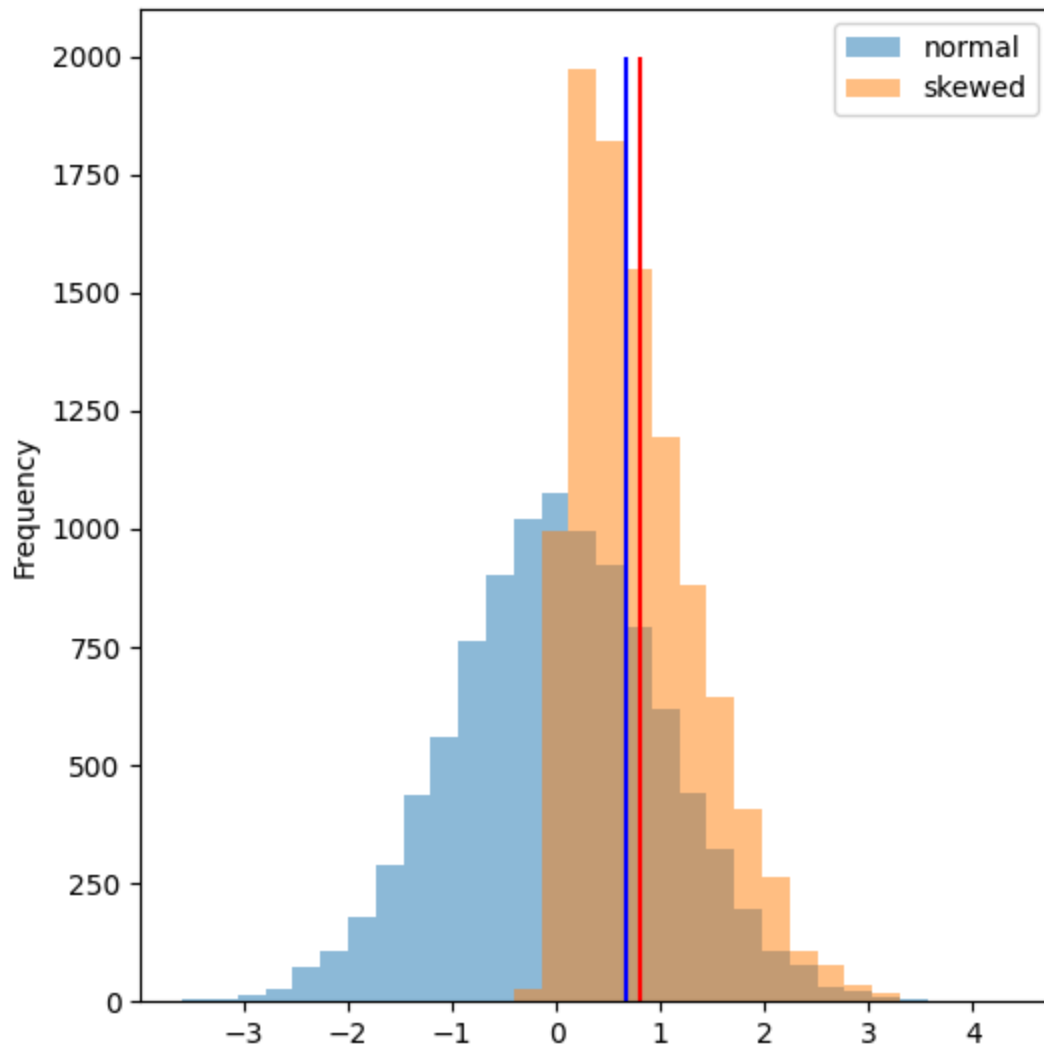
KNN imputation

```
In [ ]:  import numpy as np

         itunes_df_copy = (
             itunes_df.copy()
         )  # create copy of the dataframe so we don't alter the original
         itunes_df_copy.loc[0, "Bytes"] = np.nan
```

```
In [ ]:  from sklearn.impute import KNNImputer

         imputer = KNNImputer()
         imputed = imputer.fit_transform(itunes_df_copy[["Milliseconds", "Bytes", "Un
```

```
In [ ]:  # there are multiple values with the index value of 0, which is why we get m
         itunes_df.loc[0, "Bytes"]
```

```
Out[ ]:  0     211743651
         0       6022698
         0      11170334
         Name: Bytes, dtype: int64
```

```
In [ ]:  itunes_df_copy["Bytes"] = imputed[:, 1]
```

```
In [ ]:  itunes_df_copy.loc[0, "Bytes"]
```

```
Out[ ]:  0      381737603.6
         0        5454572.2
         0       11232851.2
         Name: Bytes, dtype: float64
```

```
In [ ]:  itunes_df["Bytes"].mean()
```

```
Out[ ]:  np.float64(33110484.730912708)
```

```
In [73]:  # adapted from this SO answer: https://stackoverflow.com/a/46740476/4549682
          def remove_outliers(df, column):
              q1 = df[column].quantile(0.25)
              q3 = df[column].quantile(0.75)
              iqr = q3 - q1
              upper_boundary = q3 + 1.5 * iqr
              lower_boundary = q1 - 1.5 * iqr
              new_df = df.loc[(df[column] > lower_boundary) & (df[column] < upper_boun
              return new_df
```

```
In [ ]:  itunes_df_clean = remove_outliers(itunes_df, "Milliseconds")
```

```
In [75]:  itunes_df_clean.shape
```

```
Out[75]:  (3614, 8)
```

```
In [76]:  itunes_df.shape
```

```
Out[76]:  (4021, 8)
```

Duplicate values

```
In [77]:  itunes_df.duplicated().sum()
```

```
Out[77]:  np.int64(518)
```

```
In [78]:  itunes_df.drop_duplicates(inplace=True)
```

Data Transformations

```
In [ ]:  itunes_df["Seconds"] = itunes_df["Milliseconds"] / 1000
```

```
In [ ]:  itunes_df["len_byte_ratio"] = itunes_df["Milliseconds"] / itunes_df["Bytes"]
```

Apply and Map

```
In [ ]: genre_dict = {"metal": "Metal", "met": "Metal"}
        itunes_df["Genre"].replace(genre_dict)
```

```
Out[ ]: 0          TV Shows
        1            Latin
        2            Latin
        3           Comedy
        4            Latin
                   ...
        3496     Classical
        3499     Classical
        3500     Classical
        3501     Classical
        3502     Soundtrack
        Name: Genre, Length: 3503, dtype: object
```

```
In [ ]: itunes_df["Genre"].apply(lambda x: x.lower())
```

```
Out[ ]: 0          tv shows
        1            latin
        2            latin
        3           comedy
        4            latin
                   ...
        3496     classical
        3499     classical
        3500     classical
        3501     classical
        3502     soundtrack
        Name: Genre, Length: 3503, dtype: object
```

```
In [ ]: # the above is the same as this
        def lowercase(x):
            return x.lower()


        itunes_df["Genre"].apply(lowercase)
```

```
Out[ ]: 0          tv shows
        1            latin
        2            latin
        3           comedy
        4            latin
                   ...
        3496     classical
        3499     classical
        3500     classical
        3501     classical
        3502     soundtrack
        Name: Genre, Length: 3503, dtype: object
```

```
In [ ]: # but using built-in functions is almost always faster
        itunes_df["Genre"].str.lower()
```

```
Out[ ]:  0          tv shows
         1             latin
         2             latin
         3            comedy
         4             latin
                       ...
         3496      classical
         3499      classical
         3500      classical
         3501      classical
         3502     soundtrack
         Name: Genre, Length: 3503, dtype: object
```

Here would be an exception to the rule for using apply -- we want to get a hybrid score of the polarity and subjectivty of text:

```python
In [ ]:  # this is a common sentiment analysis library; polarity is positive/negative
         # subjectivety is subjective/objective rating.
         from textblob import TextBlob

         test = TextBlob("Textblob is amazingly simple to use. What great fun!")
         test.sentiment
```

```
Out[ ]:  Sentiment(polarity=0.39166666666666666, subjectivity=0.4357142857142857)
```

```python
In [87]:  test.sentiment.polarity
```

```
Out[87]:  0.39166666666666666
```

```python
In [ ]:  # it would be better than apply to use a list comprehension to get sentiment
         itunes_df["Track_sentiment"] = [
             TextBlob(x).sentiment.polarity for x in itunes_df["Track"]
         ]
```

```python
In [ ]:  # but, if we wanted to mix polarity and subjectivity into one column, it wou
         def pol_sub_mix(x):
             tb = TextBlob(x)
             return tb.polarity * tb.subjectivity


         itunes_df["Track_pol_sub_mix"] = itunes_df["Track"].apply(pol_sub_mix)
```

```python
In [ ]:  # delete these columns
         itunes_df.drop(["Track_pol_sub_mix", "Track_sentiment"], inplace=True, axis=
```

```python
In [91]:  # currently doesn't work with python 3.9
          # import swifter
          # itunes_df['Genre'].swifter.apply(lambda x: x.lower())
```

```python
In [96]:  itunes_df.select_dtypes(include=np.number).columns
```

```
Out[96]:  Index(['Milliseconds', 'Bytes', 'UnitPrice', 'Seconds', 'len_byte_ratio'],
          dtype='object')
```

Group By

```
In [ ]:   itunes_df[
              ["Milliseconds", "Bytes", "UnitPrice", "Seconds", "len_byte_ratio", "Gen
          ].groupby("Genre").mean()["Seconds"].sort_values().head()
```

```
Out[ ]:   Genre
          Rock And Roll      134.643500
          Opera              174.813000
          Hip Hop/Rap        178.176286
          Easy Listening     189.164208
          Bossa Nova         219.590000
          Name: Seconds, dtype: float64
```

Saving data

```
In [ ]:   itunes_df.to_csv("cleaned_itunes_data.csv", index=False)
```

Bitcoin data analysis

```
In [ ]:   btc_df = pd.read_csv("bitcoin_price.csv")
          btc_df.head()
```

Out[ ]:

	symbol	time	open	close	high	low	volume
0	btcusd	1364688000000	92.500000	93.033000	93.74999	91.00000	3083.079791
1	btcusd	1364774400000	93.250000	103.999000	105.90000	92.49999	5224.401313
2	btcusd	1364860800000	104.000000	118.229354	118.38670	99.00000	8376.527478
3	btcusd	1364947200000	117.958261	134.700000	146.88000	101.51088	12996.245072
4	btcusd	1365033600000	134.716560	132.899000	143.00000	119.00000	6981.668305

```
In [ ]:   btc_df["symbol"].unique()
```

```
Out[ ]:   array(['btcusd'], dtype=object)
```

```
In [ ]:   btc_df.drop("symbol", axis=1, inplace=True)
```

```
In [ ]:   btc_df["time"] = pd.to_datetime(btc_df["time"], unit="ms")
```

```
In [ ]:   btc_df["time"].dtype
```

```
Out[ ]:   dtype('<M8[ns]')
```

```
In [104…   btc_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2793 entries, 0 to 2792
Data columns (total 6 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   time    2793 non-null   datetime64[ns]
 1   open    2793 non-null   float64
 2   close   2793 non-null   float64
 3   high    2793 non-null   float64
 4   low     2793 non-null   float64
 5   volume  2793 non-null   float64
dtypes: datetime64[ns](1), float64(5)
memory usage: 131.1 KB
```

In [ ]: `btc_df.set_index("time", inplace=True)`

In [106… `btc_df.head()`

Out[106…

| time	open	close	high	low	volume
2013-03-31	92.500000	93.033000	93.74999	91.00000	3083.079791
2013-04-01	93.250000	103.999000	105.90000	92.49999	5224.401313
2013-04-02	104.000000	118.229354	118.38670	99.00000	8376.527478
2013-04-03	117.958261	134.700000	146.88000	101.51088	12996.245072
2013-04-04	134.716560	132.899000	143.00000	119.00000	6981.668305

In [ ]: `btc_df[["close"]].plot(logy=True)`

Out[ ]: `<Axes: xlabel='time'>`

```
In [ ]: f = plt.figure(figsize=(5.5, 5.5))
        btc_df.iloc[-3000:][["close"]].plot(logy=True, figsize=(5.5, 5.5))
        f.patch.set_facecolor("w")  # sets background color behind axis labels
        plt.tight_layout()  # auto-adjust margins
```

<Figure size 550x550 with 0 Axes>
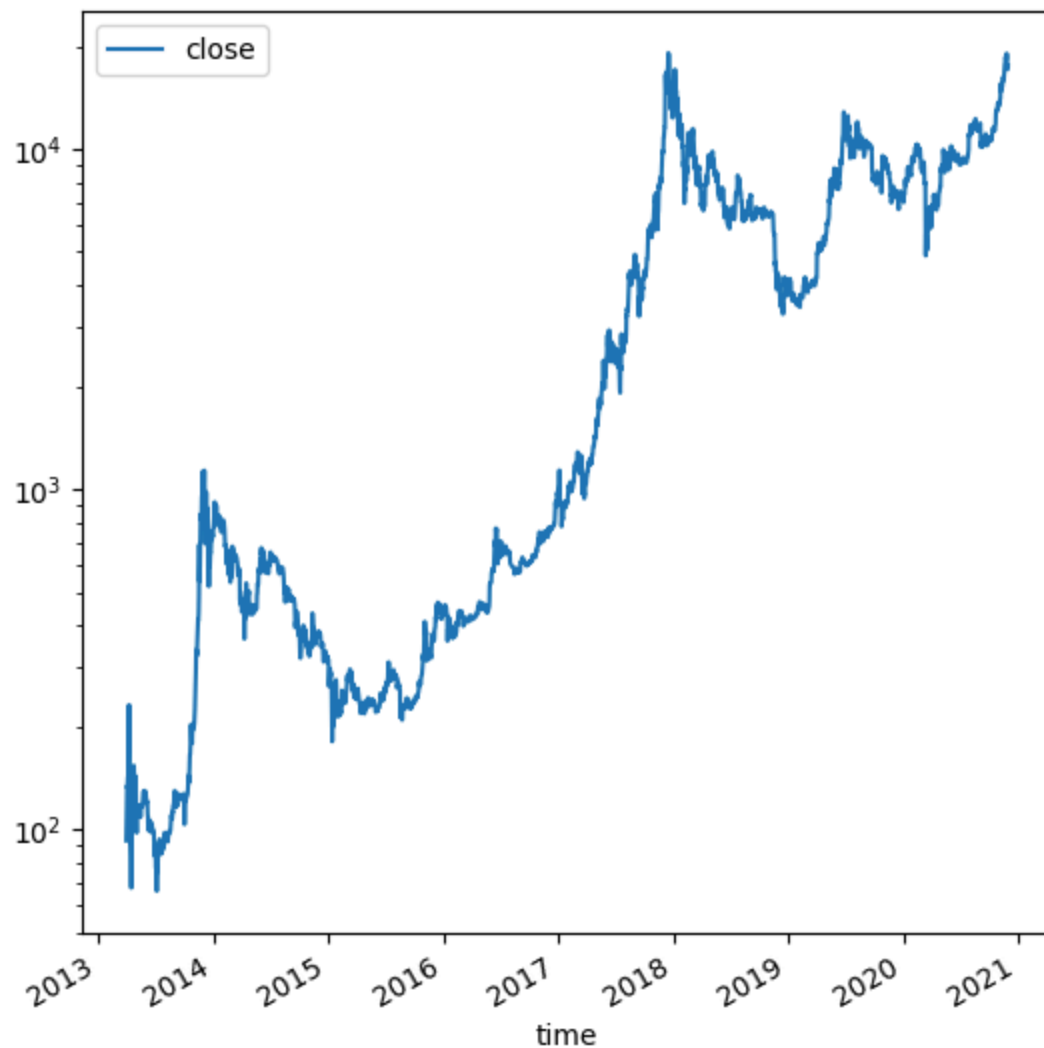
We could load a datetime as a pandas timestamp like so. Unfortunately, this trick doesn't work with seconds since the epoch, but does work with other datetime formats. For seconds since the epoch, we can provide a conversion function to the argument `date_parser` in `read_csv`, or convert it with `to_datetime` after loading.

```
In [ ]:   btc_df2 = pd.read_csv(
              "bitcoin_price.csv",
              index_col="time",
              parse_dates=["time"],
              infer_datetime_format=True,
          )
          btc_df2.head()
```

```
/tmp/ipykernel_1089788/542906089.py:1: FutureWarning: The argument 'infer_dat
etime_format' is deprecated and will be removed in a future version. A strict
version of it is now the default, see https://pandas.pydata.org/pdeps/0004-co
nsistent-to-datetime-parsing.html. You can safely remove this argument.
  btc_df2 = pd.read_csv('bitcoin_price.csv', index_col='time', parse_dates=
['time'], infer_datetime_format=True)
/tmp/ipykernel_1089788/542906089.py:1: UserWarning: Could not infer format, s
o each element will be parsed individually, falling back to `dateutil`. To en
sure parsing is consistent and as-expected, please specify a format.
  btc_df2 = pd.read_csv('bitcoin_price.csv', index_col='time', parse_dates=
['time'], infer_datetime_format=True)
```

Out[ ]:

time	symbol	open	close	high	low	volume
1364688000000	btcusd	92.500000	93.033000	93.74999	91.00000	3083.079791
1364774400000	btcusd	93.250000	103.999000	105.90000	92.49999	5224.401313
1364860800000	btcusd	104.000000	118.229354	118.38670	99.00000	8376.527478
1364947200000	btcusd	117.958261	134.700000	146.88000	101.51088	12996.245072
1365033600000	btcusd	134.716560	132.899000	143.00000	119.00000	6981.668305

In [ ]:
```python
date_parser = lambda x: pd.to_datetime(x, unit="ms")
btc_df2 = pd.read_csv(
    "bitcoin_price.csv", index_col="time", parse_dates=["time"], date_parser
)
btc_df2.head()
```

```
/tmp/ipykernel_1089788/4261904790.py:2: FutureWarning: The argument 'date_par
ser' is deprecated and will be removed in a future version. Please use 'date_
format' instead, or read your data in as 'object' dtype and then call 'to_dat
etime'.
  btc_df2 = pd.read_csv('bitcoin_price.csv', index_col='time', parse_dates=
['time'], date_parser=date_parser)
/tmp/ipykernel_1089788/4261904790.py:1: FutureWarning: The behavior of 'to_da
tetime' with 'unit' when parsing strings is deprecated. In a future version,
strings will be parsed as datetime strings, matching the behavior without a
'unit'. To retain the old behavior, explicitly cast ints or floats to numeric
type before calling to_datetime.
  date_parser = lambda x: pd.to_datetime(x, unit='ms')
```

Out[ ]:

time	symbol	open	close	high	low	volume
2013-03-31	btcusd	92.500000	93.033000	93.74999	91.00000	3083.079791
2013-04-01	btcusd	93.250000	103.999000	105.90000	92.49999	5224.401313
2013-04-02	btcusd	104.000000	118.229354	118.38670	99.00000	8376.527478
2013-04-03	btcusd	117.958261	134.700000	146.88000	101.51088	12996.245072
2013-04-04	btcusd	134.716560	132.899000	143.00000	119.00000	6981.668305

```
In [ ]:  btc_df.loc["1-1-2019":"12-31-2019"]
```

Out[ ]:

time	open	close	high	low	volume
2019-01-01	3832.628624	3963.1	3987.000000	3788.100000	13848.346752
2019-01-02	3963.000000	4048.8	4083.470676	3917.900000	19437.727989
2019-01-03	4049.000000	3924.3	4056.644331	3875.100000	10748.083712
2019-01-04	3923.200000	3954.9	3986.000000	3850.100000	13228.660465
2019-01-05	3954.900000	3911.9	4014.900000	3901.000000	14873.962225
...	...	...	...	...	...
2019-12-27	7231.600000	7287.0	7300.000000	7113.300000	3996.647227
2019-12-28	7287.000000	7334.4	7390.000000	7279.200000	2731.060711
2019-12-29	7334.400000	7413.6	7551.600000	7311.600000	2494.487206
2019-12-30	7416.100000	7260.9	7421.700000	7201.290265	4088.958159
2019-12-31	7260.900000	7208.3	7333.000000	7161.300000	2992.686133

365 rows × 5 columns

```
In [ ]:  btc_df.loc["2019"]
```

Out[ ]:

time	open	close	high	low	volume
2019-01-01	3832.628624	3963.1	3987.000000	3788.100000	13848.346752
2019-01-02	3963.000000	4048.8	4083.470676	3917.900000	19437.727989
2019-01-03	4049.000000	3924.3	4056.644331	3875.100000	10748.083712
2019-01-04	3923.200000	3954.9	3986.000000	3850.100000	13228.660465
2019-01-05	3954.900000	3911.9	4014.900000	3901.000000	14873.962225
...	...	...	...	...	...
2019-12-27	7231.600000	7287.0	7300.000000	7113.300000	3996.647227
2019-12-28	7287.000000	7334.4	7390.000000	7279.200000	2731.060711
2019-12-29	7334.400000	7413.6	7551.600000	7311.600000	2494.487206
2019-12-30	7416.100000	7260.9	7421.700000	7201.290265	4088.958159
2019-12-31	7260.900000	7208.3	7333.000000	7161.300000	2992.686133

365 rows × 5 columns

NumPy

```
In [ ]:    close_array = btc_df["close"].values
```

```
In [115…   close_array
```

```
Out[115…   array([   93.033     ,    103.999      ,    118.22935407, ...,
                  17211.69580098, 17171.          , 17686.840768  ], shape=(2793,))
```

```
In [116…   close_array.shape
```

```
Out[116…   (2793,)
```

```
In [117…   close_array.dtype
```

```
Out[117…   dtype('float64')
```

```
In [ ]:    import numpy as np

           close_list = btc_df["close"].to_list()
           close_array = np.array(close_list)
           close_array
```

```
Out[ ]:    array([   93.033     ,    103.999      ,    118.22935407, ...,
                  17211.69580098, 17171.          , 17686.840768  ], shape=(2793,))
```

```
In [119…   %timeit kd_close = close_array / 1000
```

```
           3.68 µs ± 432 ns per loop (mean ± std. dev. of 7 runs, 100,000 loops each)
```

```
In [120…   %timeit kd_close_list = [c / 1000 for c in close_list]
```

```
           182 µs ± 37.1 µs per loop (mean ± std. dev. of 7 runs, 1,000 loops each)
```

```
In [ ]:    volume_array = btc_df["volume"].values
           close_array * volume_array
```

```
Out[ ]:    array([2.86828162e+05, 5.43332512e+05, 9.90351433e+05, ...,
                  5.91110933e+08, 1.92622769e+08, 6.25398332e+07], shape=(2793,))
```

```
In [ ]:    btc_df["market_cap"] = btc_df["close"] * btc_df["volume"]
           btc_df["market_cap"]
```

```
Out[ ]:  time
         2013-03-31      2.868282e+05
         2013-04-01      5.433325e+05
         2013-04-02      9.903514e+05
         2013-04-03      1.750594e+06
         2013-04-04      9.278567e+05
                             ...
         2020-11-24      2.857665e+08
         2020-11-25      2.523951e+08
         2020-11-26      5.911109e+08
         2020-11-27      1.926228e+08
         2020-11-28      6.253983e+07
         Name: market_cap, Length: 2793, dtype: float64
```

```
In [ ]:  np.log(btc_df["close"])
```

```
Out[ ]:  time
         2013-03-31      4.532954
         2013-04-01      4.644381
         2013-04-02      4.772626
         2013-04-03      4.903050
         2013-04-04      4.889589
                           ...
         2020-11-24      9.859065
         2020-11-25      9.837722
         2020-11-26      9.753344
         2020-11-27      9.750977
         2020-11-28      9.780576
         Name: close, Length: 2793, dtype: float64
```