

Project Overview:

For my project, I wanted to work with data mining in relationship to the way that people listen to music. Ideally, I would have liked to use Spotify; however, Spotify's API does not allow you to access listening history, which I decided was something that I wanted to work with. Instead, I chose to use lastfm, which is a similar music streaming site that allows you to pull users recent tracks. For this project, I gathered data about users, specifically their country and their listening history and then tried to determine the most popular song in each country based on the results. I wanted to know how many users I have to pull information for in order to get accurate data about the most popular song. For the purposes of this overview, I ran my program with a sample size of 500 total users (with a random amount of users for every country) and 10000 total users.

Implementation:

My program works in 3 discrete parts. The first of which is determining the users to pull listening histories from. Ideally, I would have liked to pull a random list of users, however, lastfm does not have a public list of users, so this was not an option. The method I chose to find users was to pick a very popular user and pull information about their friends, and then recursively pull information from their friends for a random depth. One alternative method would have been to use lastfm's events API to pull data from unrelated users, however, I believe that the lastfm events API is location-based and I wanted to test my program on a collection of users from around the world. As it turns out, friends on lastfm aren't usually related by circles the same way that friends are related on Facebook. So my method seems to work perfectly fine. The data for lastfm is in the form of xml documents containing information tags about location, username, etc. I used the pattern module to pull the xml documents off of their API and then used the ElementTree module to parse the xml data. After pulling thousands of user's username and country off of lastfm, I wrote it to a text file so that I did not have to pull from the internet every time I wanted to test the program.

The next facet of the program was getting information from lastfm about each of the user's listening history. Doing this was pretty straightforward, as it required a simple call to lastfm's API using pattern and then parsing the xml data that was returned. The third facet of this program was using this data to tally how often songs were occurring in each country. To do this, I kept a dictionary that had countries as the key and as the values had a dictionary containing the names of the songs and their tallies. I then read this data into an xml file so that I didn't have to reuse the internet for when I wanted to test the program. I used xml as opposed to using a text file or pickling because it's much more readable than a .dat file used for pickling, it's much easier to work with than text files, and I wanted to gain more experience parsing through them. Lastly, I had to find the most frequently occurring song in each country.

Results:

With a sample size of 500 total users hailing from around the world, I was able to generate this data:

BE -- Quantum Flux by Northlane
FR -- Ball And Chain by XTC
BG -- Russian Winter by Krokus

DK -- Mine munni by Koer
HR -- Zbog tebe za tebe by Ponor
DE -- Echoes Of The Last Paradigm by Medus Lantern
JP -- Stare at Yourself by Hibria
HU -- Supernatural Superserious by REM
BR -- Speechless by Lady Gaga
BS -- The Village by Georges Delerue
FI -- Selvsti pihtynyt by Jenni Vartiainen
BY -- Monte Carlos Mountain by Nik Sand
GR -- Never Had No One Ever by The Smiths
RU -- Linverno Allegro by Antonio Vivaldi
NL -- Dark Doo Wop by MS MR
PT -- No Surprises by Radiohead
RS -- opa hey by Kottarashky
TR -- The Mundane and the Magic by Dark Tranquillity
LT -- Heaven and Hell by CC Catch
TN -- Inception by Eluveitie
SK -- Whispers in the Dark by Mumford Sons
TL -- Wings of Feather And Wax by Killer Be Killed
LV -- Beautiful PrimeMusicru by Alex Gaudino feat Mario
RO -- Night of the Demon by Demon
PL -- Hells Gate by Omen
PM -- When Poets Dreamed Of Angels by David Sylvian
CH -- Rock Superstar by Cypress Hill
CO -- La maladie damour by Matrise des HautsdeSeine
VE -- Is This Happiness by Lana Del Rey
CL -- Living for Love by Madonna
CA -- Everlong by Foo Fighters
IR -- Pagan Poetry by Bjrk
IT -- Runnin Wild by Airbourne
VN -- Honest by Kodaline
CZ -- Human Interest by Hooverphonic
GE -- Smoke Club by Kamandi
AR -- Animal by Neon Trees
AU -- Uptown Funk by Mark Ronson
AT -- Come Undone by Duran Duran
IN -- Birthright by Adharma
LK -- Taphead by Talk Talk
IE -- Silent Picture Show by Unquiet Nights
ID -- One More Time by Daft Punk
ES -- Out of My Mind by Duran Duran
ME -- Forever by HAIM
NO -- My Remedy by Santa Cruz
US -- I Love 64 by mindinabox
UM -- at the end of it all by ADULT
SI -- Ticks Leeches by Tool
KP -- Surrender by Billy Talent
UK -- Fireside by Arctic Monkeys
UA -- Desecration by Nikolo Kotzev

MX -- Vernculo by Future Brown
SE -- White Trash Circus by Mtley Cre

This data is not very accurate, so I decided to see if I could increase the accuracy by running it with 10000 users instead. I generated this list:

Final Frequency Data:

BD -- Thats What You Get by Paramore
BE -- As Daylight Breaks by Serious Black
BG -- Golden Eyes by Buckethead
BA -- Most Of The Time by Turnover
BB -- Towards the Sun From The Home Soundtrack by Rihanna
BN -- The gypsies wife Live by Leonard Cohen
BO -- Ten Feet Tall by Afrojack
JP -- Nothin on You feat Bruno Mars by BoB
JM -- Leggo Dutty Babylon Ways by Chezidek
JO -- Somewhere In Time Believe by Savatage
BR -- Elastic Heart by Sia
BS -- Keep Feeling Fascination by Baxendale
BY -- Eye of the Tiger by Survivor
RU -- Lazars by Blvck Ceiling
RS -- Eklatantno Feat Mimi Mercedez Prod By NSS by Daki BD
LT -- Dream of Flying by Brian Crain
RO -- Ruzica by Goran Bregovi
GT -- Steppin Away by Stereomud
GS -- Reach Out Original Mix by Sander Van Doorn
GR -- Song to Hall Up High by Bathory
BH -- Epitaph by Necrophagist
GE -- Double Fifths by Teebs
GL -- Joan of Arc by Arcade Fire
GI -- Polythene Pam by The Beatles
TN -- Against the Clock by heavytrackerz
HR -- Thinking Out Loud by Ed Sheeran
HU -- Otsegolation by StaticX
HK -- 2 Minutes to Midnight by Iron Maiden
HN -- Down On The Street by The Stooges
HM -- Get Away by Saint Vitus
CR -- Kids With Guns by Gorillaz
PR -- Un Poco de Amor by Shakira
PS -- Sunken Soldiers Ball by Agua de Annique
PW -- All Hands On Deck Remix feat Iggy Azalea by Tinashe
PT -- Move Together by James Bay
PY -- I Put A Spell On You Fifty Shades of Grey From The Fifty Shades of Grey Soundtrack by Annie Lennox
PA -- The Funerary Dirge of a Violinist by Carach Angren
PE -- Throb by Janet Jackson
PK -- Gravity by Papa Roach
PH -- Great Britain by BBC World Service
PL -- Uma Thurman by Fall Out Boy

PM -- When Poets Dreamed Of Angels by David Sylvian
ZM -- Theres a Trollop Up Elmwood Street by The Toy Dolls
EE -- Best of You by Foo Fighters
EG -- Partition by Beyonc
ZA -- Day One by Hans Zimmer
EC -- KILL BILL by Brown Eyed Girls
IT -- Hurricane Venus by Boa
VN -- Hero by Family of the Year
SB -- Untitled by NDE
ES -- Towards the Sun by Rihanna
ME -- Water by Ohgr
MD -- To Find The King by Cerimonial Sacred
MG -- Saturn Valley by
MA -- Aint No Love In The Heart Of The City by Bobby Blue Bland
UZ -- Erotica by Madonna
MO -- Fire by Ohio Players
MN -- Ukrainian Insurgent Army by Drudkh
MH -- Gust of Wind by Pharrell Williams
MK -- Faulty Logic by ManMachine
UM -- at the end of it all by ADULT
MT -- Curtain of Treachery by Thy Serpent
MV -- Aurora Borealis by Astral Projection
MP -- Misty Mountains by Howard Shore
UK -- Love Me Like You Do From Fifty Shades Of Grey by Ellie Goulding
MY -- Uptown Funk ft Bruno Mars by Mark Ronson
MX -- starssss by MRKSTRKF
IL -- Crucify Me feat Lights by Bring Me the Horizon
FR -- La Javanaise by Madeleine Peyroux
FI -- Heliumpallo by PMMP
FO -- Complete Demise by Anata
NI -- One Last Time by Ariana Grande
NL -- PS I Love You by The Beatles
NO -- Pivt kuin unta by Arto Sotavalta
NZ -- Slipshod by Enter Shikari
NP -- Asa branca by Luiz Gonzaga
NU -- Slow Your Ass Down by Black Flag
CH -- Drive Home edit by Steven Wilson
CO -- The Forest Is My Throne by Satyricon
CN -- Lonesome Town by Ricky Nelson
CL -- Alone by Suicidal Tendencies
CC -- Pass This On by The Knife
CA -- Fur Elise by Beethoven Consort Ludwig van Beethoven
CZ -- Moje Mila by PIO Squad
CY -- Einai Poly Noris by
CX -- I Like It Rough by Lady Gaga
UY -- Orphan by Toto
KP -- Go Bang vkcomxclusiveszone by SBTRKT
SY -- On An Island by David Gilmour
KG -- Like That feat Skate by Jack Jack

KE -- Aman bre deryalar by Arif entrk
KI -- War of the Thrones by Blind Guardian
SV -- Postmortem by Slayer
ST -- Passage by Lowercase Noises
SK -- Sun King by The Beatles
SJ -- Golden Earrings by Gandalf
SI -- Alive Live On 889 WERSBoston by Misery Index
SH -- Its My Life by No Doubt
SL -- Give Me Some Free by Materia
KZ -- 1 2005 astvatsurovcom by
SA -- Chump by Green Day
SG -- Mistletoe by Justin Bieber
SE -- Why Dont We Go There by One Direction
DO -- Say by John Mayer
DJ -- This Wifi Sucks by Nevada Rose
DK -- Royal Blue by Cold War Kids
DE -- No pain No game by
DZ -- Old Mornings Dawn by Summoning
US -- Side FX Include Mikeys Middle Finger TV 4 by Ministry
LB -- Wucan by Black Mountain
LC -- Parisienne Walkways by Gary Moore
TT -- Illuminator by William Control
TR -- We Found Love feat Calvin Harris by Rihanna
LK -- Saree Ke Fall Sa by Nakash Aziz Antara Mitra
LI -- Perfect Now by Sarah Blasko
LV -- Requiem in D Minor III I Dies irae by Wolfgang Amadeus Mozart
TL -- Wings of Feather And Wax by Killer Be Killed
LU -- The Upgrade Part II by Slaves
TH -- No Way by The Naked and Famous
TF -- Insight by Joy Division
VA -- Vile Vicious Vision by Acid Drinkers
VE -- Dame tu Cosita by El Chombo
IQ -- Far From Any Road by The Handsome Family
IS -- Black Woods by Mono
IR -- Soulmate by Natasha Bedingfield
AM -- Demon You Lily Anne by Lake of Tears
AL -- Work Bitch by Britney Spears
AO -- Love Is Dead by Tokio Hotel
AN -- Madonna by Drake
AQ -- mpodobeala timpului by Negur Bunget
AR -- Problem by Ariana Grande
AU -- Episode 11 Showdown in the Past by Dragon Ball Z Saban dub
AT -- Nostalgia by 4Lyn
IO -- Soothe My Soul by Depeche Mode
IN -- Take Me to Church I Put a Spell on You Medley 57th GRAMMYs by Hozier Annie Lennox
AZ -- Bitter Sweet Faith by BitterSweet
IE -- Space Oddity by David Bowie
ID -- Dizzy by Goodnight Electric
UA -- In by May Of Sorrow

This was the size of the user pool for each country:

BD -- 12
BE -- 63
BG -- 65
BA -- 13
BB -- 2
BN -- 1
BO -- 3
JP -- 73
JM -- 2
JO -- 2
BR -- 2769
BS -- 4
BY -- 78
RU -- 519
RS -- 54
TL -- 1
PA -- 1
RO -- 46
GT -- 4
GS -- 1
GR -- 31
BH -- 1
GE -- 5
GL -- 1
GI -- 2
HR -- 116
HU -- 35
HK -- 5
HN -- 2
HM -- 1
PR -- 4
PS -- 1
PW -- 1
PT -- 80
PY -- 10
LV -- 24
PE -- 20
PK -- 1
PH -- 10
PL -- 884
PM -- 4
ZM -- 1
EE -- 15
EG -- 17
ZA -- 15
EC -- 6
AL -- 2

AO -- 3
KZ -- 11
ES -- 109
ME -- 7
MD -- 4
MG -- 2
UY -- 11
UZ -- 2
MO -- 1
MN -- 2
MH -- 1
MK -- 12
UM -- 2
MT -- 1
MV -- 1
MP -- 1
UK -- 618
MY -- 5
MX -- 173
AT -- 40
FR -- 126
MA -- 2
SH -- 3
FI -- 237
FO -- 1
NI -- 1
NL -- 187
NO -- 46
NZ -- 11
NP -- 2
NU -- 1
CH -- 36
CO -- 51
CN -- 10
CL -- 76
CC -- 3
CA -- 163
CZ -- 73
CY -- 1
CX -- 3
CR -- 8
SY -- 3
KG -- 1
KE -- 1
KI -- 1
SV -- 2
ST -- 2
SK -- 44
SJ -- 1

SI -- 8
KP -- 6
SL -- 1
SB -- 1
SA -- 3
SG -- 9
SE -- 166
DO -- 1
DJ -- 1
DK -- 17
DE -- 454
DZ -- 3
US -- 1310
LB -- 3
LC -- 1
TT -- 2
TR -- 98
LK -- 3
LI -- 3
TN -- 7
LT -- 15
LU -- 2
TH -- 3
TF -- 1
VA -- 8
VE -- 22
IQ -- 1
IS -- 6
IR -- 20
AM -- 2
IT -- 170
VN -- 16
AN -- 1
AQ -- 11
AR -- 67
AU -- 130
IL -- 27
IO -- 1
IN -- 57
AZ -- 2
IE -- 37
ID -- 69
UA -- 165

This is obviously not the most accurate data as well. This is because in the grand scheme of things, 10000 users is not significant. Likewise, Lastfm is probably not representative of music listeners as a whole. Spotify data would have been so much more accurate. Additionally, there were other cooler projects I would have liked to do, however, given time constraints, I was limited.

Reflection:

I got a lot of practice storing and parsing xml data, which is really useful for future projects. However, I didn't have a plan for this project going into it, and I don't feel like this was really my best project idea, it sort of developed out of the available APIs that I was playing around with as I was trying to come up with an idea. Furthermore, I think that doing this project in Java would have been considerably more efficient. One reason is that Java as a whole runs a lot faster than Python. Another reason is that Java allows multithreading, which would have allowed me to do all three facets of my project simultaneously as I was pulling data from lastfm.