

Sparkify Project Workspace

This workspace contains a tiny subset (128MB) of the full dataset available (12GB). Feel free to use this workspace to build your project, or to explore a smaller subset with Spark before deploying your cluster on the cloud. Instructions for setting up your Spark cluster is included in the last lesson of the Extracurricular Spark Course content.

You can follow the steps below to guide your data analysis and model building portion of this project.

```
In [1]: # import libraries
import pyspark
from pyspark.sql import SparkSession

from pyspark.sql.functions import isnan, when, count, col, isnull, avg
from pyspark.sql.functions import udf
from pyspark.sql.types import StringType
from pyspark.sql.types import IntegerType

from pyspark.ml import Pipeline
from pyspark.ml.feature import StringIndexer, OneHotEncoder, VectorAssembler

from pyspark.ml.regression import LinearRegression
from pyspark.ml.classification import LogisticRegression

import datetime

import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from seaborn import barplot, countplot

from pyspark.sql.functions import percent_rank
from pyspark.sql import Window

from pyspark.ml.evaluation import BinaryClassificationEvaluator
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
from pyspark.ml.tuning import CrossValidator, ParamGridBuilder
from pyspark.mllib.evaluation import MulticlassMetrics

from pyspark.ml.classification import GBTClassifier
from pyspark.ml.classification import LinearSVC
from pyspark.ml.classification import OneVsRest
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.regression import LinearRegression
from pyspark.ml.classification import LogisticRegression
```

```
In [2]: # create a Spark session
spark = SparkSession \
    .builder \
    .appName("ZeizerChurnProject") \
    .getOrCreate()
```

Load and Clean Dataset

In this workspace, the mini-dataset file is `mini_sparkify_event_data.json`. Load and clean the dataset, checking for invalid or missing data - for example, records without `userids` or `sessionids`.

Loading the data

```
In [3]: path = "mini_sparkify_event_data.json"
data_churn = spark.read.json(path)
```

Cleaning the data

```
In [4]: data_churn.printSchema()
```

```
root
|-- artist: string (nullable = true)
|-- auth: string (nullable = true)
|-- firstName: string (nullable = true)
|-- gender: string (nullable = true)
|-- itemInSession: long (nullable = true)
|-- lastName: string (nullable = true)
|-- length: double (nullable = true)
|-- level: string (nullable = true)
|-- location: string (nullable = true)
|-- method: string (nullable = true)
|-- page: string (nullable = true)
|-- registration: long (nullable = true)
|-- sessionId: long (nullable = true)
|-- song: string (nullable = true)
|-- status: long (nullable = true)
|-- ts: long (nullable = true)
|-- userAgent: string (nullable = true)
|-- userId: string (nullable = true)
```

```
In [5]: data_churn.take(3)
```

```
Out[5]: [Row(artist='Martha Tilston', auth='Logged In', firstName='Colin', gender='M', itemInSession=50, lastName='Freeman', length=277.89016, level='paid', location='Bakersfield, CA', method='PUT', page='NextSong', registration=1538173362000, sessionId=29, song='Rockpools', status=200, ts=1538352117000, userAgent='Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101 Firefox/31.0', userId='30'),
Row(artist='Five Iron Frenzy', auth='Logged In', firstName='Micah', gender='M', itemInSession=79, lastName='Long', length=236.09424, level='free', location='Boston-Cambridge-Newton, MA-NH', method='PUT', page='NextSong', registration=1538331630000, sessionId=8, song='Canada', status=200, ts=1538352180000, userAgent='\"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/37.0.2062.103 Safari/537.36\"', userId='9'),
Row(artist='Adam Lambert', auth='Logged In', firstName='Colin', gender='M', itemInSession=51, lastName='Freeman', length=282.8273, level='paid', location='Bakersfield, CA', method='PUT', page='NextSong', registration=1538173362000, sessionId=29, song='Time For Miracles', status=200, ts=1538352394000, userAgent='Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101 Firefox/31.0', userId='30')]
```

In [6]: `data_churn.describe().show()`

```
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+
|summary|          artist|      auth|firstName|gender|      itemInSe
ssion|lastName|      length|  level|      location|method|  p
age|      registration|      sessionId|      song|
status|      ts|      userAgent|      userId|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+
|  count|          228108|      286500|      278154|278154|          2
86500|  278154|          228108|286500|          278154|286500| 286
500|          278154|          286500|          228108|
286500|          286500|          278154|          286500|
|  mean| 551.0852017937219|      null|      null|  null|114.414212914
48516|      null|249.1171819778458|      null|      null|  null|  n
ull|1.535358834084427...|1041.526554973822|      Infinity|210.0
5459685863875|1.540956889810483...|      null|59682.0227859
3872|
| stddev|1217.7693079161374|      null|      null|  null|129.767262011
40994|      null|99.23517921058361|      null|      null|  null|  n
ull| 3.291321616327586E9|726.7762634630741|      NaN| 31.5
0507848842214|1.5075439608226302E9|      null|109091.949999
1047|
|  min|          !!!| Cancelled| Adelaida|      F|
0|  Adams|          0.78322|  free|      Albany, OR|  GET|  About|
1521380675000|          1|⓪ÃÃg ÃÃtti Gr...|          200|
1538352117000|"Mozilla/5.0 (Mac...|
|  max| ÃÃlafur Arnalds|Logged Out|  Zyonna|      M|          1
321|  Wright|          3024.66567|  paid|Winston-Salem, NC|  PUT|Upgrad
e|          1543247354000|          2474|ÃÃau hafa slopp...|
404|          1543799476000|Mozilla/5.0 (comp...|          99|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+
```

Number of lines and Columns

In [7]: `print(data_churn.count(), len(data_churn.columns))`

286500 18

tem o DropDuplicates

Searching for NaN's

from: <https://stackoverflow.com/questions/44627386/how-to-find-count-of-null-and-nan-values-for-each-column-in-a-pyspark-dataframe>
 (https://stackoverflow.com/questions/44627386/how-to-find-count-of-null-and-nan-values-for-each-column-in-a-pyspark-dataframe)

In [8]: `data_churn.select([count(when(isnan(c), c)).alias(c) for c in data_churn.columns]).show()`

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|artist|auth|firstName|gender|itemInSession|lastName|length|level|location|method|page|registration|sessionId|song|status|ts|userAgent|userId|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
```

In [9]: `data_churn.select([count(when(isnull(c), c)).alias(c) for c in data_churn.columns]).show()`

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|artist|auth|firstName|gender|itemInSession|lastName|length|level|location|method|page|registration|sessionId|song|status|ts|userAgent|userId|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
| 58392|      0|      8346|      8346|      0|      8346| 58392|      0|      8346|      0|      0|      8346|      0|      0|      8346|      0|      0|      8346|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
```

```
In [10]: data_churn.select(['artist', 'length', 'song']).collect()
```

```

Out[10]: [Row(artist='Martha Tilston', length=277.89016, song='Rockpools'),
          Row(artist='Five Iron Frenzy', length=236.09424, song='Canada'),
          Row(artist='Adam Lambert', length=282.8273, song='Time For Miracle
          s'),
          Row(artist='Enigma', length=262.71302, song='Knocking On Forbidden D
          oors'),
          Row(artist='Daft Punk', length=223.60771, song='Harder Better Faster
          Stronger'),
          Row(artist='The All-American Rejects', length=208.29995, song="Don't
          Leave Me"),
          Row(artist='The Velvet Underground / Nico', length=260.46649, song
          ='Run Run Run'),
          Row(artist='Starflyer 59', length=185.44281, song='Passengers (Old A
          lbum Version)'),
          Row(artist=None, length=None, song=None),
          Row(artist='Frumpies', length=134.47791, song='Fuck Kitty'),
          Row(artist='Britt Nicole', length=229.8771, song='Walk On The Wate
          r'),
          Row(artist=None, length=None, song=None),
          Row(artist='Edward Sharpe & The Magnetic Zeros', length=223.58159, s
          ong='Jade'),
          Row(artist='Tesla', length=201.06404, song="Gettin' Better"),
          Row(artist=None, length=None, song=None),
          Row(artist='Stan Mosley', length=246.69995, song='So-Called Friend
          s'),
          Row(artist='Florence + The Machine', length=168.64608, song="You've
          Got The Love"),
          Row(artist='Tokyo Police Club', length=166.1122, song='Citizens Of T
          omorrow'),
          Row(artist='Orishas', length=222.22322, song='Represent'),
          Row(artist='Ratatat', length=229.77261, song='Swisha'),
          Row(artist='Manolo Garcia', length=283.74159, song='Carbon Y Ramas S
          ecas'),
          Row(artist='Downhere', length=223.92118, song='Here I Am'),
          Row(artist='Modjo', length=250.93179, song='What I Mean'),
          Row(artist='MÃ\x83Ãtley CrÃ\x83Ãe', length=231.26159, song='Sticky
          Sweet'),
          Row(artist='David Bowie', length=174.41914, song='Sorrow (1997 Digit
          al Remaster)'),
          Row(artist='Skillet', length=233.32526, song='Rebirthing (Album Vers
          ion)'),
          Row(artist='Edwyn Collins', length=216.842, song="You'll Never Know
          (My Love) (Bovellian 07 Mix)"),
          Row(artist='Telepopmusik', length=241.60608, song='Smile'),
          Row(artist='Kings Of Leon', length=307.46077, song='I Want You'),
          Row(artist='Florence + The Machine', length=219.66322, song='Dog Day
          s Are Over (Radio Edit)'),
          Row(artist='Coldplay', length=343.32689, song='For You'),
          Row(artist='Nine Inch Nails', length=373.2371, song='Closer'),
          Row(artist='Nick Drake', length=222.92853, song='Tomorrow Is A Long
          Time'),
          Row(artist='Campo', length=429.50485, song='Mi CorazÃ\x83Ãn'),
          Row(artist='Kix', length=246.25587, song='Girl Money'),
          Row(artist=None, length=None, song=None),
          Row(artist='Lambert_ Hendricks & Ross', length=139.04934, song='Hall
          oween Spooks'),
          Row(artist='Edgar Oceransky', length=309.62893, song='Demasiado Buen

```



```

o'),
Row(artist='Kanye West', length=311.84934, song='Stronger'),
Row(artist='Eminem', length=269.03465, song='My Name Is'),
Row(artist='Charttraxx Karaoke', length=225.17506, song='Fireflies'),
Row(artist='The Acacia Strain', length=178.15465, song='Whoa! Shut It Down'),
Row(artist='Hate Eternal', length=207.67302, song='Spiritual Holocaust'),
Row(artist='Vampire Weekend', length=243.04281, song='The Kids Don't Stand A Chance (Album)'),
Row(artist=None, length=None, song=None),
Row(artist='Redman / Ready Roc', length=217.91302, song='Dis Iz Brick City'),
Row(artist='Scars On Broadway', length=196.12689, song='World Long Gone'),
Row(artist=None, length=None, song=None),
Row(artist='Bram Vermeulen', length=251.42812, song='Mamma'),
Row(artist='The All-American Rejects', length=238.86322, song='Move Along'),
Row(artist='Lily Allen', length=185.83465, song='22'),
Row(artist=None, length=None, song=None),
Row(artist='The Last Goodnight', length=189.85751, song='Pictures Of You'),
Row(artist=None, length=None, song=None),
Row(artist='Death Cab for Cutie', length=290.29832, song='I Will Possess Your Heart (Radio Edit)'),
Row(artist='Kings Of Leon', length=204.2771, song='Manhattan'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Tub Ring', length=233.69098, song='Invalid'),
Row(artist='Björk', length=348.57751, song='Undo'),
Row(artist='Half Pint', length=218.5922, song='Don't Close The Door'),
Row(artist='Otis Redding', length=151.61424, song='These Arms Of Mine'),
Row(artist='Radiohead', length=141.322, song='I Will (Los Angeles Version)'),
Row(artist='Boston', length=357.72036, song='Don't Look Back'),
Row(artist='The Prodigy', length=260.07465, song='The Big Gundown'),
Row(artist='Future Rock', length=239.90812, song='Gears'),
Row(artist='Evan Rachel Wood', length=184.92036, song='Black Bird'),
Row(artist=None, length=None, song=None),
Row(artist='Phoenix', length=184.73751, song='Long Distance Call'),
Row(artist='Antonio Vivaldi', length=302.8371, song='L'automne : Presto _Tempo impetuoso d'estate'),
Row(artist='Dwight Yoakam', length=239.3073, song='You're The One'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Ultra Vivid Scene', length=196.93669, song='Nausea'),
Row(artist='Breakage', length=42.03057, song='If'),

```

```

Row(artist=None, length=None, song=None),
Row(artist='Charles Mingus', length=279.84934, song='Eat That Chicke
n (LP Version)'),
Row(artist='Parachute', length=221.67465, song='Ghost'),
Row(artist='Nevermore', length=320.57424, song='Matricide'),
Row(artist='Sea Wolf', length=241.76281, song='The Garden You Plante
d'),
Row(artist=None, length=None, song=None),
Row(artist='Electric Light Orchestra', length=274.88608, song='Roll
Over Beethoven (Single Version) (2003 Digital Remaster)'),
Row(artist=None, length=None, song=None),
Row(artist='Jason Mraz', length=218.40934, song='Geek In The Pink [P
hil Tan Remix]'),
Row(artist='Erin McKeown', length=242.59873, song='You Were Right Ab
out Everything'),
Row(artist=None, length=None, song=None),
Row(artist='Amy Winehouse', length=229.85098, song='Valerie'),
Row(artist='Mastodon', length=215.97995, song='The Wolf Is Loose (Al
bum Version)'),
Row(artist='Beyonce & Shakira', length=201.50812, song='Beautiful Li
ar'),
Row(artist='Taylor Swift', length=192.73098, song='Crazier'),
Row(artist='Traveling Wilburys', length=196.51873, song='Margarit
a'),
Row(artist='Dead Can Dance', length=203.91138, song='The Promised Wo
mb (Remastered)'),
Row(artist='Cristina Pato', length=337.8673, song='MiÃ\x83Ã±a Nai LÃ
\x83Ã±a'),
Row(artist='The Killers', length=246.80444, song='Read My Mind'),
Row(artist='Jacky Terrasson', length=342.7522, song='Le Jardin d'Hiv
er'),
Row(artist=None, length=None, song=None),
Row(artist='En Vogue', length=250.46159, song='Free Your Mind'),
Row(artist='Foo Fighters', length=235.28444, song='Learn To Fly'),
Row(artist='Galaxie 500', length=276.32281, song='Melt Away'),
Row(artist=None, length=None, song=None),
Row(artist='Faith Evans', length=324.28363, song='Soon As I Get Home
(Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='Linkin Park', length=268.61669, song='New Divide (Album
Version)'),
Row(artist='Owl City', length=230.81751, song='Umbrella Beach'),
Row(artist=None, length=None, song=None),
Row(artist='Paramore', length=198.76526, song='Franklin (Album Versi
on)'),
Row(artist=None, length=None, song=None),
Row(artist='Dwight Yoakam', length=239.3073, song='You're The One'),
Row(artist='Moose', length=384.23465, song='This River Will Never Ru
n Dry'),
Row(artist='Righeira', length=218.53995, song='Vamos a la Playa'),
Row(artist='Klaus Schulze', length=1242.46159, song='North Of the Yu
kon'),
Row(artist='Miranda!', length=229.82485, song='Perfecta'),
Row(artist=None, length=None, song=None),
Row(artist='The Devil Wears Prada', length=69.69424, song='The Ascen
t (LP Version)'),
Row(artist='Justin Bieber', length=245.73342, song='Down To Earth'),

```

```

Row(artist='Trivium', length=293.66812, song='Requiem'),
Row(artist=None, length=None, song=None),
Row(artist='The Cardigans', length=198.21669, song='Lovefool'),
Row(artist='OutKast', length=234.29179, song='The Way You Move'),
Row(artist='Collective Soul', length=273.47546, song='Run (LP Versio
n)'),
Row(artist='Foolish Things', length=266.86649, song='Who Can Compar
e'),
Row(artist=None, length=None, song=None),
Row(artist='Skunk Anansie', length=194.58567, song='Pickin On Me'),
Row(artist='Underworld', length=788.92363, song='mmm skyscraper...i
love you'),
Row(artist=None, length=None, song=None),
Row(artist='Dario Marianelli', length=146.99057, song='Cello Lesso
n'),
Row(artist='The Black Keys', length=1400.2673, song='240 Years Befor
e Your Time'),
Row(artist='Sick Puppies', length=198.94812, song='War'),
Row(artist='David Sanborn', length=217.73016, song='Little Flower'),
Row(artist='Alice In Chains', length=153.3122, song='We Die Young'),
Row(artist='Michael Franks', length=346.3571, song="When Sly Calls
[Don't Touch That Phone] (LP Version)"),
Row(artist='The Airborne Toxic Event', length=216.21506, song='This
Losing'),
Row(artist='Letzte Instanz', length=222.58893, song='Bittere Nach
t'),
Row(artist=None, length=None, song=None),
Row(artist='Atreyu', length=213.52444, song='Your Private War (LP Ve
rsion)'),
Row(artist='Florence + The Machine', length=258.55955, song='Cosmic
Love'),
Row(artist='Daft Punk', length=213.28934, song='Harder Better Faster
Stronger (Alive Radio Edit 2007)'),
Row(artist='Sick Puppies', length=200.75057, song='Too Many Words'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Creed', length=295.3922, song='My Sacrifice'),
Row(artist='Shakira Featuring Wyclef Jean', length=217.36444, song
="Hips Don't Lie (featuring Wyclef Jean)"),
Row(artist='The Dingees', length=221.23057, song='Summer'),
Row(artist='Taylor Swift', length=245.36771, song='The Way I Loved Y
ou'),
Row(artist=None, length=None, song=None),
Row(artist='Rick Astley', length=261.69424, song='Hold Me In Your Ar
ms'),
Row(artist='Shafiq Husayn', length=186.14812, song='No Moor'),
Row(artist='The Clientele', length=264.95955, song='House On Fire'),
Row(artist='Nelly / Jaheim', length=336.56118, song='My Place'),
Row(artist='A&E', length=239.69914, song='Shima Guni'),
Row(artist=None, length=None, song=None),
Row(artist='We The Kings', length=206.52363, song='Secret Valentin
e'),
Row(artist='Barry Tuckwell/Academy of St Martin-in-the-Fields/Sir Ne
ville Marriner', length=277.15873, song='Horn Concerto No. 4 in E fla
t K495: II. Romance (Andante cantabile)'),

```

```

Row(artist='Belanova', length=185.88689, song='Rosa Pastel'),
Row(artist='JoÃ\x83Ã£o Gilberto', length=258.37669, song='Ave Maria
No Morro'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Ginuwine', length=243.30404, song='In Those Jeans'),
Row(artist='Cherie', length=202.13506, song="I'm Ready (Album Versio
n)"),
Row(artist=None, length=None, song=None),
Row(artist='Griffin House', length=195.7873, song='The Guy That Says
Goodbye To You Is Out Of His Mind'),
Row(artist='Stephen Marley / Damian Marley', length=220.52526, song
='The Traffic Jam'),
Row(artist='Ashley Tisdale/Lucas Grabeel', length=277.05424, song='I
Want It All'),
Row(artist='Colony 5', length=286.71955, song='Science'),
Row(artist='Starfield', length=265.7171, song='No Other Saviour'),
Row(artist='Queen', length=139.17995, song='Who Wants To Live Foreve
r (With Commentary)'),
Row(artist=None, length=None, song=None),
Row(artist='BjÃ\x83Ã¶rk', length=348.57751, song='Undo'),
Row(artist=None, length=None, song=None),
Row(artist='Stevie Ray Vaughan And Double Trouble', length=256.1562
8, song='Leave My Girl Alone'),
Row(artist='Black Eyed Peas', length=289.12281, song='I Gotta Feelin
g'),
Row(artist='JosÃ\x83Ã© GonzÃ\x83Ã¡lez', length=168.98567, song='Hint
s'),
Row(artist=None, length=None, song=None),
Row(artist='Dario Marianelli', length=282.95791, song='With My Own E
yes'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Coldplay', length=268.38159, song='Yellow'),
Row(artist='DEATH', length=251.37587, song='Together as one'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Coldplay', length=307.51302, song='Clocks'),
Row(artist='Firewind', length=222.22322, song='Before The Storm'),
Row(artist='The Kooks', length=203.96363, song='Naive'),
Row(artist=None, length=None, song=None),
Row(artist='Kurd Maverick', length=398.18404, song='Blue Monday'),
Row(artist='JET', length=213.81179, song='Are You Gonna Be My Girl
(Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='J Mascis', length=222.45832, song='Waistin'),
Row(artist='James Blunt', length=244.13995, song='High (Album Versio
n)'),
Row(artist=None, length=None, song=None),
Row(artist='Eagles', length=372.29669, song='Hotel California'),
Row(artist='Eric Church', length=164.362, song='Hell On The Heart'),
Row(artist='Los Rebutitos', length=226.61179, song='Un Bonito Fina

```

```

l'),
Row(artist='Shooter Jennings', length=249.36444, song='Sweet Savanna
h'),
Row(artist='Nirvana', length=234.97098, song='Dive'),
Row(artist=None, length=None, song=None),
Row(artist='Harmonia', length=655.77751, song='Sehr kosmisch'),
Row(artist='Hootie And The Blowfish', length=217.67791, song='Runnin
g From An Angel (LP Version)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Magnolia Electric Co.', length=258.55955, song='Hard To
Love A Man'),
Row(artist=None, length=None, song=None),
Row(artist='Silverstein', length=197.3024, song='Already Dead (Album
Version)'),
Row(artist='Carpenters', length=218.85342, song='Jambalaya (On The B
ayou)'),
Row(artist=None, length=None, song=None),
Row(artist='Björk', length=259.13424, song='Alarm Call'),
Row(artist='Devo', length=259.13424, song='The Super Thing'),
Row(artist='Phoenix', length=167.83628, song='Lasso'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='DJ Jazzy Jeff', length=295.02649, song='Supa Jean feat.
Jean Grae (Instrumental)'),
Row(artist=None, length=None, song=None),
Row(artist='Travis', length=295.23546, song='Luv'),
Row(artist=None, length=None, song=None),
Row(artist='Justin Timberlake', length=288.93995, song='Cry Me A Riv
er'),
Row(artist='Enrique Iglesias', length=208.5873, song='Escape'),
Row(artist='Donna Fargo', length=158.64118, song='Funny Face'),
Row(artist='Seine', length=173.66159, song='So far so long'),
Row(artist='Sheena Easton', length=199.05261, song='Morning Train (N
ine to Five)'),
Row(artist='Bachman-Turner Overdrive', length=238.49751, song='Roll
On Down The Highway'),
Row(artist=None, length=None, song=None),
Row(artist='Keith Sweat', length=312.99873, song='In Your Eyes (LP V
ersion)'),
Row(artist='Metronomy', length=215.35302, song='Radio Ladio'),
Row(artist='Onar', length=306.6771, song='Xehasmeni Melodia'),
Row(artist='Eminem', length=181.34159, song='Insane'),
Row(artist='The Dodos', length=282.53995, song='Fools'),
Row(artist='Emery', length=256.10404, song='Holding Out For A Her
o'),
Row(artist='Mando Diao', length=211.59138, song='Never Seen The Ligh
t Of Day (Radio Edit)'),
Row(artist='Slowdive', length=406.07302, song='J's Heaven'),
Row(artist='The All-American Rejects', length=268.85179, song='Back
To Me'),
Row(artist='Alliance Ethnik', length=252.21179, song='Repräsentation
te'),

```

```

Row(artist='Beirut', length=117.52444, song='La Banlieue'),
Row(artist='Third Day', length=243.43465, song='All The Heavens'),
Row(artist='Evanescence', length=263.78404, song='My Immortal (Album
Version)'),
Row(artist='Cat Stevens', length=183.19628, song='The First Cut Is T
he Deepest'),
Row(artist='Duran Duran', length=231.31383, song='All Along The Wate
r'),
Row(artist=None, length=None, song=None),
Row(artist='Brett Dennen', length=271.51628, song='She's Mine'),
Row(artist='Simon Harris', length=195.83955, song='Sample Track 2'),
Row(artist='The Decemberists', length=252.81261, song='The Bachelor
and the Bride'),
Row(artist='Tenacious D', length=165.95546, song='The Metal'),
Row(artist='Kerry Livgren', length=248.58077, song='Colonnade Garden
s'),
Row(artist='Metallica', length=515.21261, song='Master Of Puppets'),
Row(artist='STEVE CAMP', length=201.82159, song='Lazy Jane'),
Row(artist='Relient K', length=210.33751, song='Bite My Tongue'),
Row(artist='DJ Koze', length=208.74404, song='Barock Am Ring'),
Row(artist='The Brian Setzer Orchestra', length=182.282, song='Ameri
cano'),
Row(artist='Lifehouse', length=249.18159, song='By Your Side'),
Row(artist='Giuliano Palma & The Bluebeaters', length=204.95628, son
g='Se Ne Dicon Di Parole'),
Row(artist='Frumpies', length=134.47791, song='Fuck Kitty'),
Row(artist='Usher Featuring Lil' Jon & Ludacris', length=250.38322,
song='Yeah!'),
Row(artist=None, length=None, song=None),
Row(artist='Povia', length=172.66893, song='Ma Tu Sei Scemo'),
Row(artist=None, length=None, song=None),
Row(artist='Maxwell', length=255.242, song='Pretty Wings'),
Row(artist='Four Tet', length=447.79057, song='Wing Body Wing'),
Row(artist='Harmonia', length=655.77751, song='Sehr kosmisch'),
Row(artist='Billy Bragg', length=227.52608, song='Sulk'),
Row(artist='The Reverend Peyton's Big Damn Band', length=179.46077,
song='My Old Man Boogie'),
Row(artist='Jack Johnson', length=164.98893, song='The Sharing Son
g'),
Row(artist='Flying Lotus', length=134.3473, song='SexSlaveShip'),
Row(artist='Kate Nash', length=265.89995, song='Birds'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Bob Newhart', length=367.98649, song='Introducing Tobacc
o To Civilization (LP Version)'),
Row(artist='OneRepublic', length=224.67873, song='Secrets'),
Row(artist='Goldfrapp', length=251.14077, song='Train'),
Row(artist='Jules Shear', length=177.50159, song='The Once Lost Retu
rns'),
Row(artist='John Powell', length=148.61016, song='Dragon's Den'),
Row(artist=None, length=None, song=None),
Row(artist='The All-American Rejects', length=260.77995, song='Your
Star'),
Row(artist='Muse', length=209.34485, song='Supermassive Black Hole
(Album Version)'),
Row(artist=None, length=None, song=None),

```

```

Row(artist='Sheena Easton', length=239.62077, song='Strut (1993 Digital Remaster)'),
Row(artist=None, length=None, song=None),
Row(artist='Puddle Of Mudd', length=210.80771, song='Psycho'),
Row(artist='Memory Cassette', length=255.99955, song='Asleep At A Party'),
Row(artist=None, length=None, song=None),
Row(artist='Ricky Martin', length=279.48363, song='Tal Vez'),
Row(artist='Vilma Palma e Vampiros', length=289.82812, song='Me vuelvo loco por vos'),
Row(artist='Fate', length=244.1922, song='Limbo A Go Go'),
Row(artist='Faithless', length=218.04363, song='Drifting Away'),
Row(artist=None, length=None, song=None),
Row(artist='Base Ball Bear', length=255.60771, song='Sayonara-Nostalgia'),
Row(artist=None, length=None, song=None),
Row(artist='Kid Cudi / MGMT / Ratatat', length=295.67955, song='Pursuit Of Happiness (nightmare)'),
Row(artist='Weird Al" Yankovic', length=303.5424, song='Angry White Boy Polka'),
Row(artist='Lupe Fiasco', length=182.69995, song='Put You On Game (Explicit Album Version)'),
Row(artist='Pixies', length=166.68689, song='Debaser (live)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='King Tubby', length=142.21016, song='Dub On My Mind Dub'),
Row(artist='Erykah Badu', length=225.90649, song='Umm Hmm'),
Row(artist='"O"Rosko Raricim", length=90.56608, song='Terre Promise'),
Row(artist='Gorillaz', length=192.91383, song='Tomorrow Comes Today'),
Row(artist='Bonnie Tyler', length=418.63791, song='Total Eclipse Of The Heart'),
Row(artist=None, length=None, song=None),
Row(artist='Iron And Wine', length=272.43057, song='Pagan Angel and a Borrowed Car (Album)'),
Row(artist=None, length=None, song=None),
Row(artist='Rusted Root', length=260.80608, song='Hands Are Law'),
Row(artist='Joni Mitchell', length=183.64036, song='Free Man In Paris (LP Version)'),
Row(artist='Four Tet', length=209.05751, song='Hilarious Movie Of The 90s'),
Row(artist='Tiziano Ferro', length=227.05587, song='Il Regalo Più Grande'),
Row(artist='Eminem', length=320.83546, song='Hailie's Song'),
Row(artist='Cut Copy', length=272.95302, song='Time Stands Still'),
Row(artist=None, length=None, song=None),
Row(artist='Buckcherry', length=252.70812, song='Sunshine (Album Version)'),
Row(artist='Sound 5', length=350.32771, song='The Hacienda Must Be (Re:) Built'),
Row(artist=None, length=None, song=None),
Row(artist='The Kooks', length=169.45587, song='Love It All'),
Row(artist='The Car Is On Fire', length=218.69669, song='Manuel'),
Row(artist='K'naan / Chubb Rock", length=189.64853, song='ABC's'),
Row(artist='Delirious', length=457.19465, song='Dance Flow'),

```

```

Row(artist='Plasmatics', length=256.7571, song='The Damned'),
Row(artist='Escape The Fate', length=210.57261, song='We Won't Back
Down'),
Row(artist='Kevin Blechdom', length=222.56281, song='Running Away'),
Row(artist='B.o.B', length=269.63546, song='Nothin' On You [feat. Br
uno Mars] (Album Version)'),
Row(artist='Alliance Ethnik', length=195.94404, song='SinceritÃ\x83Â
© Et Jalousie'),
Row(artist='Aly', length=532.84526, song='Rosaires'),
Row(artist='Benny Benassi Presents The Biz', length=285.54404, song
='Satisfaction'),
Row(artist='Jimmy Needham', length=212.27057, song='Forgiven And Lov
ed'),
Row(artist='Osvaldo Pugliese', length=173.24363, song='Canaro En Par
is'),
Row(artist='Tears For Fears', length=361.29914, song='Change'),
Row(artist='Ray LaMontagne', length=344.86812, song='Forever My Frie
nd'),
Row(artist=None, length=None, song=None),
Row(artist='Doc Watson', length=134.03383, song='I'm Going Back To T
he Old Home'),
Row(artist='Marc Anthony', length=289.4624, song='Valio La Pena'),
Row(artist='Flyleaf', length=194.92526, song='Set Apart This Drea
m'),
Row(artist='Dominique A', length=153.20771, song='Le Courage Des Ois
eaux'),
Row(artist=None, length=None, song=None),
Row(artist='B5', length=189.36118, song='Heartbreak'),
Row(artist='The Futureheads', length=120.81587, song='Robot'),
Row(artist=None, length=None, song=None),
Row(artist='Jordan Francis/Roshon Bernard Fegan', length=155.01016,
song='Hasta La Vista'),
Row(artist=None, length=None, song=None),
Row(artist='The Smiths', length=138.44853, song='Frankly_ Mr. Shankl
y'),
Row(artist='Daft Punk', length=297.89995, song='Digital Love'),
Row(artist='The Killers', length=229.0673, song='Bones'),
Row(artist='JoÃ\x83Âfo Gilberto', length=249.25995, song='Bolinha De
Papel'),
Row(artist=None, length=None, song=None),
Row(artist='Days Of The New', length=353.27955, song='Provider'),
Row(artist='Plies', length=278.22975, song='Somebody (Loves You) (Ex
plicit Album Version)'),
Row(artist='Story Of The Year', length=235.78077, song='Until The Da
y I Die (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Flamin' Groovies', length=299.65016, song='Slow Death'),
Row(artist='Kode9', length=369.47546, song='2 Far Gone'),
Row(artist='The Prodigy', length=225.09669, song='Firestarter'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Yeasayer', length=298.91873, song='Red Cave'),
Row(artist=None, length=None, song=None),
Row(artist='Dwight Yoakam', length=239.3073, song='You're The One'),

```



```

Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Eminem', length=284.86485, song='The Real Slim Shady'),
Row(artist=None, length=None, song=None),
Row(artist='Damian Marley / Nas', length=317.30893, song='Road To Zi
on'),
Row(artist=None, length=None, song=None),
Row(artist='Total Science', length=425.89995, song='Flip Flop'),
Row(artist=None, length=None, song=None),
Row(artist='Nickelback', length=262.81751, song="I'd Come For You (A
lbum Version)"),
Row(artist='Faith Hill', length=222.30159, song="There You'll Be (Al
bum Version)"),
Row(artist='Kronos Quartet', length=191.32036, song='Tabu (Taboo)'),
Row(artist='Venetian Snares', length=71.70567, song='Fuck Off'),
Row(artist='Evanescence', length=220.15955, song='Hello'),
Row(artist=None, length=None, song=None),
Row(artist='Fall Out Boy', length=191.37261, song='Grand Theft Autum
n/Where Is Your Boy (Album Version)'),
Row(artist='The Red Crayola', length=153.80853, song='Transparent Ra
diation - Original'),
Row(artist="Lil' Zane", length=225.27955, song="What's Up"),
Row(artist='Tune Up! vs. Italobrothers', length=319.73832, song='Col
ours Of The Rainbow'),
Row(artist='Ludovico Einaudi & BallakÃ© Sissoko', length=348.29
016, song='Dessert Dans Le Desert'),
Row(artist='Radiohead', length=247.06567, song='Black Star'),
Row(artist='Enya', length=104.56771, song='Deireadh An Tuath'),
Row(artist='SOJA', length=271.67302, song="Decide You're Gone"),
Row(artist='Grinderman', length=225.64526, song="(I Don't Need You T
o) Set Me Free (Edit)"),
Row(artist='Binary Star', length=54.22975, song='The KGB (Intro)'),
Row(artist='Erasure', length=246.46485, song="Fingers And Thumbs (Co
ld Summer's Day) (Sound Factory Remix Radop Edit)"),
Row(artist='Silverio', length=236.66893, song='XXX'),
Row(artist='The White Stripes', length=232.61995, song='Seven Nation
Army'),
Row(artist='Yellowcard', length=218.74893, song='Breathing'),
Row(artist='Usher featuring will.i.am', length=395.72853, song='OM
G'),
Row(artist='Eddie Floyd', length=198.19057, song='Guess Who'),
Row(artist='Fudge Fingas', length=405.96853, song='Me & U'),
Row(artist=None, length=None, song=None),
Row(artist='Madness', length=168.59383, song='Baggy Trousers'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='The Get Up Kids', length=260.07465, song="I'll Catch Yo
u"),
Row(artist='No Te Va Gustar', length=274.12853, song='Te Voy a Lleva
r'),
Row(artist=None, length=None, song=None),
Row(artist='HYPOCRISY', length=200.98567, song='Orgy In Blood'),
Row(artist='Shadows Fall', length=303.62077, song='December (origina
lly by Only Living Witness)'),
Row(artist='The Gathering', length=380.23791, song='Rescue Me'),
Row(artist='The Solsonics', length=263.8624, song='Keep The Rhythm S
trong'),

```

```

Row(artist=None, length=None, song=None),
Row(artist='Bon Jovi', length=224.522, song="It's My Life"),
Row(artist='The Police', length=223.84281, song='Invisible Sun'),
Row(artist='the bird and the bee', length=226.21995, song='Polite Dance Song'),
Row(artist='Junip', length=250.87955, song='The Ghost Of Tom Joad'),
Row(artist='The Black Ghosts', length=225.77587, song='Repetition Kills You'),
Row(artist='The Libertines', length=182.38649, song='Music When The Lights Go Out'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Blood Red Shoes', length=222.24934, song='You Bring Me Down'),
Row(artist='Aerosmith', length=319.73832, song='Crazy'),
Row(artist='Afroman', length=202.37016, song='Because I Got High'),
Row(artist='Crotchduster', length=871.31383, song='Crotchopus'),
Row(artist='CKY', length=197.04118, song='Sniped'),
Row(artist='From Autumn To Ashes', length=258.42893, song='Eulogy For An Angel'),
Row(artist='Black Eyed Peas', length=222.6673, song='Third Eye'),
Row(artist='Train', length=278.04689, song='Ramble On'),
Row(artist='Tokio Hotel', length=248.73751, song='Spring nicht'),
Row(artist='Ronan Hardiman', length=171.36281, song='Victory'),
Row(artist='The Killers', length=360.75057, song='Human'),
Row(artist='Esperanza Spalding', length=419.47383, song='Two Bad'),
Row(artist='Easy Star All-Stars', length=246.9873, song='Brain Damage (featuring Dr. Israel)'),
Row(artist=None, length=None, song=None),
Row(artist='Faithless Featuring Dido', length=205.76608, song='One Step Too Far'),
Row(artist='Zero 7', length=237.60934, song='Swing'),
Row(artist=None, length=None, song=None),
Row(artist='3 Doors Down', length=233.74322, song='Kryptonite'),
Row(artist=None, length=None, song=None),
Row(artist='Bruce Springsteen', length=288.62649, song='Thunder Road'),
Row(artist='Mariah Carey / Jadakiss / Styles P', length=267.67628, song='We Belong Together'),
Row(artist='Gene Chandler', length=153.33832, song='Duke Of Earl'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Papoose', length=154.22649, song='We Got Gunz'),
Row(artist='Eminem / Dr. Dre / 50 Cent', length=297.482, song='Crack A Bottle'),
Row(artist='Left Lane Cruiser', length=246.56934, song='KFD'),
Row(artist='Foo Fighters', length=349.88363, song='Exhausted'),
Row(artist='Yeah Yeah Yeahs', length=220.96934, song='Heads Will Roll'),
Row(artist='Jami Smith', length=386.79465, song='Water Falls'),
Row(artist='The xx', length=188.55138, song='Basic Space'),
Row(artist='Metallica', length=450.19383, song='Ride The Lightning'),
Row(artist=None, length=None, song=None),
Row(artist='Bullet For My Valentine', length=236.32934, song='Cries

```

```

In Vain'),
Row(artist='Elliott Smith', length=226.97751, song='Pictures Of Me'),
Row(artist='The Mars Volta', length=257.07057, song='Concertina'),
Row(artist='Jimi Hendrix', length=372.87138, song='Electric Church Re-
ed House'),
Row(artist='Neneh Cherry', length=232.202, song='Manchild'),
Row(artist='Johnny Gill', length=276.58404, song='Fairweather Friend'),
Row(artist='Ryan Bingham', length=123.11465, song='Boracho Station'),
Row(artist='Michael Bublé', length=190.40608, song='Crazy Little Thing Called Love (Album Version)'),
Row(artist='Fugees', length=290.2722, song='Killing Me Softly With This Song'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='H-Blockx', length=230.08608, song='Countdown To Insanity'),
Row(artist='Jason Mraz & Colbie Caillat', length=189.6224, song='Lucky (Album Version)'),
Row(artist='Justin Bieber', length=220.89098, song='Somebody To Love'),
Row(artist='Blackfoot', length=632.94649, song='Highway Song'),
Row(artist=None, length=None, song=None),
Row(artist='Cupid', length=390.13832, song='Cupid Shuffle [The Scumfrog Club Mix]'),
Row(artist=None, length=None, song=None),
Row(artist='Red Astaire', length=264.09751, song='S.B.P.'),
Row(artist='Mikromusic', length=181.99465, song='Dobrze Jest'),
Row(artist='Cam\'Ron / Juelz Santana', length=208.3522, song='Oh Boy'),
Row(artist=None, length=None, song=None),
Row(artist='The Smiths', length=236.48608, song='Unloveable'),
Row(artist='Justin Bieber', length=220.89098, song='Somebody To Love'),
Row(artist='Nana Caymmi - Part. Esp.: Emilio Santiago-Artista Gentilmente Cedido Pela Sigla', length=227.5522, song='Doralinda/Cafe Com Pao(Jodel)/Dora'),
Row(artist='Beirut', length=257.48853, song='Postcards From Italy'),
Row(artist='Avril Lavigne', length=240.40444, song='When You\'re Gone'),
Row(artist='Duran Duran', length=321.88036, song='The Seventh Stranger'),
Row(artist='Benny Spellman', length=146.65098, song='Lipstick Traces (On A Cigarette) (2002 Digital Remaster)'),
Row(artist='AFI', length=266.84036, song='Endlessly_ She Said'),
Row(artist='Mavado', length=200.9073, song='Touch The Road'),
Row(artist='Neikka RPM', length=231.60118, song='Here\'s Your Revolution (Red Venom)'),
Row(artist='Coal Chamber', length=215.14404, song='Sway (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='Tech N9ne', length=115.82649, song='Enter Everready'),
Row(artist='The Ordinary Boys', length=162.87302, song='Boys Will Be Boys'),
Row(artist='Interpol', length=203.12771, song='The Scale'),

```

```

Row(artist='Timbaland / Katy Perry', length=238.70649, song='If We E
ver Meet Again'),
Row(artist='The Black Keys', length=201.69098, song='Stack Shot Bill
y'),
Row(artist='The Decemberists', length=272.40444, song='Days Of Elain
e'),
Row(artist='Freeform Five', length=421.92934, song='No More Conversa
tions'),
Row(artist='Epica', length=348.3424, song='Unleashed'),
Row(artist='Ron Carter', length=497.13587, song="I CAN'T GET STARTE
D"),
Row(artist='The Black Keys', length=156.18567, song='Heavy Soul '),
Row(artist='Madonna', length=260.75383, song='Beautiful Stranger'),
Row(artist='Mutemath', length=283.71546, song='You Are Mine (Album V
ersion)'),
Row(artist='BeyoncÃ\x83Â©', length=224.41751, song='Trust In Me'),
Row(artist=None, length=None, song=None),
Row(artist='Faith No More', length=223.81669, song='Cuckoo For Cac
a'),
Row(artist=None, length=None, song=None),
Row(artist='Weird Al Yankovic', length=219.89832, song='Bohemian Pol
ka'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Flora Purim', length=399.69914, song='O Sonho (Moon Drea
ms)'),
Row(artist=None, length=None, song=None),
Row(artist='Confederate Railroad', length=199.78404, song='Queen Of
Memphis (Dance Mix)'),
Row(artist='Kanye West', length=237.42649, song='Diamonds From Sierr
a Leone'),
Row(artist='The Pussycat Dolls / Snoop Dogg', length=210.31138, song
='Bottle Pop'),
Row(artist='Crystal Castles', length=213.05424, song='Not In Love'),
Row(artist='Kid Cudi', length=227.34322, song='Up Up & Away'),
Row(artist='Il Divo', length=280.68526, song='Regresa A Mi (Unbreak
My Heart)'),
Row(artist=None, length=None, song=None),
Row(artist='John Zorn', length=268.5122, song='Chinatown'),
Row(artist=None, length=None, song=None),
Row(artist='Zoegirl', length=283.66322, song='No You'),
Row(artist='Five Iron Frenzy', length=783.25506, song='Every New Da
y'),
Row(artist='Cage The Elephant', length=241.162, song='In One Ear'),
Row(artist=None, length=None, song=None),
Row(artist='The Lonely Island / Natalie Portman / Chris Parnell', le
ngth=146.83383, song="Natalie's Rap"),
Row(artist=None, length=None, song=None),
Row(artist='BjÃ\x83Â¶rk', length=348.57751, song='Undo'),
Row(artist=None, length=None, song=None),
Row(artist='Performed by Jennifer Hudson;BeyoncÃ\x83Â© Knowles;Anik
a Noni Rose;Dreamgirls (Motion Picture Soundtrack)', length=199.3399
5, song='Dreamgirls'),
Row(artist='Jonas Brothers', length=150.96118, song='Goodnight And G
oodbye'),
Row(artist='Eddie Vedder', length=199.23546, song='End Of The Roa

```

```

d'),
  Row(artist='Getfar', length=423.75791, song='Shining Star'),
  Row(artist='Natiruts', length=210.88608, song='Jamaica Roots II(Agora E Sempre)'),
  Row(artist='The Doors', length=182.49098, song='Spanish Caravan [New Stereo Mix] (Advanced Resolution)'),
  Row(artist='TECH N9NE feat. Kutt Calhoun & Big Krizz Kaliko', length=330.68363, song='Sex Out South'),
  Row(artist='Frumpies', length=134.47791, song='Fuck Kitty'),
  Row(artist='Violent Femmes', length=284.3424, song='Add It Up'),
  Row(artist='Metallica', length=515.21261, song='Master Of Puppets'),
  Row(artist='Taylor Swift / Colbie Caillat', length=264.95955, song='Breathe'),
  Row(artist='Mika Miko', length=74.68363, song="Don't Shake It Off"),
  Row(artist='Linkin Park', length=226.95138, song='High Voltage'),
  Row(artist='Air France', length=194.0371, song='Karibien'),
  Row(artist='Insane Clown Posse', length=262.89587, song='Piggy Pie (Old School)'),
  Row(artist=None, length=None, song=None),
  Row(artist='Kings Of Leon', length=201.79546, song='Revelry'),
  Row(artist='Lady GaGa', length=274.18077, song='Alejandro'),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist='Default', length=268.66893, song='Wasting My Time'),
  Row(artist='Nickelback', length=227.18649, song='Never Gonna Be Alone (Album Version)'),
  Row(artist='Joshua Radin', length=204.43383, song='Sky'),
  Row(artist='Cabaret Voltaire', length=236.38159, song='Sex In Secret'),
  Row(artist='Jimmy Needham', length=156.18567, song='Firefly'),
  Row(artist='JosÃ© Cid', length=287.4771, song='Nossa Senhora Do Tejo'),
  Row(artist='Kings Of Leon', length=231.26159, song='Use Somebody'),
  Row(artist=None, length=None, song=None),
  Row(artist='The Radio Dept.', length=245.9424, song='A Token Of Gratitude'),
  Row(artist='Randy Crawford', length=270.75873, song='Almaz'),
  Row(artist='Daniel Bedingfield', length=235.17995, song='Never Gonna Leave Your Side'),
  Row(artist='Jimmy Eat World', length=434.28526, song='23 (Album Version)'),
  Row(artist='Daddy Yankee', length=172.90404, song='Descontrol'),
  Row(artist=None, length=None, song=None),
  Row(artist='Simply Red', length=303.96036, song='Thrill Me'),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist=None, length=None, song=None),
  Row(artist='Zinc / Slarta John', length=191.16363, song='Flim'),
  Row(artist='Pavement', length=99.16036, song='Mercy:The Laundromat'),
  Row(artist=None, length=None, song=None),
  Row(artist='Ram Jam', length=236.7473, song='Black Betty'),
  Row(artist='Corinne Bailey Rae', length=234.05669, song='Daydreamin

```

```

g'),
Row(artist='Downhere', length=240.45669, song='My Last Amen'),
Row(artist='The Killers', length=304.84853, song='Believe Me Natali
e'),
Row(artist='Caribou', length=282.33098, song='Kaili'),
Row(artist='The Divine Comedy', length=203.4673, song='Songs of Lov
e'),
Row(artist='Lifhouse', length=246.93506, song='Unknown'),
Row(artist='CÃx83Ãline Dion & Josh Groban', length=277.99465, song
='The Prayer'),
Row(artist='Love Of Lesbian', length=269.29587, song='Universos Infi
nitos'),
Row(artist='Justin Timberlake', length=288.93995, song='Cry Me A Riv
er'),
Row(artist='Tub Ring', length=233.69098, song='Invalid'),
Row(artist='John Mayer', length=269.71383, song='Heartbreak Warfar
e'),
Row(artist='Train', length=205.45261, song='Marry Me'),
Row(artist='Mae', length=444.70812, song='Communication'),
Row(artist='Nicholas Gunn', length=247.66649, song='Canyon Nights'),
Row(artist='TV On The Radio', length=176.1171, song='Dancing Choos
e'),
Row(artist='City And Colour', length=154.40934, song='Faithless'),
Row(artist='Tub Ring', length=233.69098, song='Invalid'),
Row(artist='Yeastayer', length=323.44771, song='2080'),
Row(artist='Justin Hinds & The Dominoes', length=165.01506, song='Sa
ve A Bread'),
Row(artist='Jack Johnson', length=373.2371, song='Angel / Better Tog
ether'),
Row(artist='Fanfarlo', length=270.78485, song="I'm a Pilot"),
Row(artist='First Aid Kit', length=149.57669, song="You're Not Comin
g Home Tonight"),
Row(artist='Florence + The Machine', length=258.55955, song='Cosmic
Love'),
Row(artist=None, length=None, song=None),
Row(artist='Slim Dusty', length=198.922, song='Long Black Road'),
Row(artist='Fluke', length=391.00036, song='Cut'),
Row(artist='Muse', length=258.06322, song='Map Of The Problematiqu
e'),
Row(artist='The Pussycat Dolls', length=219.50649, song="I Don't Nee
d A Man"),
Row(artist='HÃx83Ãroes del Silencio', length=198.89587, song='El M
ar No Cesa (Live)'),
Row(artist=None, length=None, song=None),
Row(artist='Weezer', length=311.87546, song="Dreamin'"),
Row(artist='Metallica', length=415.16363, song='Fade To Black'),
Row(artist='Dangerdoom', length=209.47546, song='Space Hos'),
Row(artist=None, length=None, song=None),
Row(artist='Surfer Blood', length=236.69506, song='Take It Easy'),
Row(artist='Samy Deluxe', length=238.36689, song='Intro'),
Row(artist='JET', length=243.59138, song='Cold Hard Bitch (Album Ver
sion)'),
Row(artist='Josh Groban (Featuring Lili Haydn)', length=299.12771, s
ong="Jesu_ Joy Of Man's Desiring (Featuring Lili Haydn) (Album Versio
n)"),
Row(artist='Television', length=312.24118, song='Carried Away (Rema
stered LP Version)'),

```

```

Row(artist='Alexia', length=174.47138, song='Gimme Love'),
Row(artist='Pearl Jam', length=276.40118, song='Nothingman (Live)'),
Row(artist='The Misfits', length=143.35955, song='TV Casualty'),
Row(artist='Efecto Mariposa', length=254.37995, song='Por querert
e'),
Row(artist=None, length=None, song=None),
Row(artist="O'Rosko Raricim", length=90.56608, song='Terre Promis
e'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Kid Cudi / MGMT / Ratatat', length=295.67955, song='Purs
uit Of Happiness (nightmare)'),
Row(artist='Adam & The Ants', length=195.65669, song='Prince Charmin
g'),
Row(artist='Selena Gomez & The Scene', length=241.91955, song='Natur
ally'),
Row(artist='Frumpies', length=134.47791, song='Fuck Kitty'),
Row(artist='Linkin Park', length=197.51138, song='Figure.09 (Album V
ersion)'),
Row(artist=None, length=None, song=None),
Row(artist='Linkin Park', length=494.99383, song='Bleed It Out [Live
At Milton Keynes]'),
Row(artist='Kaiser Chiefs', length=222.82404, song='Half The Trut
h'),
Row(artist='Coldplay', length=311.27465, song='The Scientist'),
Row(artist='Jel', length=243.59138, song='Thrashin'),
Row(artist='Rob Zombie', length=251.16689, song='Iron Head'),
Row(artist='Kelly Clarkson', length=156.65587, song='(You Make Me Fe
el Like A) Natural Woman'),
Row(artist='Thursday', length=214.43873, song='Voices On A String (A
lbum Version)'),
Row(artist='Darude', length=516.28363, song='Feel The Beat'),
Row(artist='Third Eye Blind', length=272.87465, song='Jumper (2006 R
emastered LP Version)'),
Row(artist='Benny Benassi Presents The Biz', length=285.54404, song
='Satisfaction'),
Row(artist=None, length=None, song=None),
Row(artist='Travie McCoy', length=211.12118, song='Billionaire [fea
t. Bruno Mars] (Explicit Album Version)'),
Row(artist='Mario Rosenstock', length=12.64281, song='Keithipedia
5'),
Row(artist='Harmonia', length=655.77751, song='Sehr kosmisch'),
Row(artist='Wax Tailor', length=177.68444, song='The Tune'),
Row(artist='Alexi Murdoch', length=288.93995, song='Dream About Flyi
ng'),
Row(artist='Coldplay', length=307.51302, song='Clocks'),
Row(artist='Kieran', length=219.84608, song='Best Friends'),
Row(artist='Bloc Party', length=235.65016, song='Luno'),
Row(artist='Alicia Keys featuring Tony! Toni! TonÃ\x83Ã©! and Jermai
ne Paul', length=284.15955, song='Diary'),
Row(artist='The Killers', length=254.37995, song='Tidal Wave'),
Row(artist='Philmont', length=217.20771, song='Another Name'),
Row(artist='Helen Reddy', length=211.35628, song='Angie Baby'),
Row(artist='Keyshia Cole', length=244.00934, song='Work It Out'),
Row(artist='Charttraxx Karaoke', length=225.17506, song='Fireflie
s'),

```

```

Row(artist=None, length=None, song=None),
Row(artist="Soulja Boy Tell'em", length=225.12281, song='Snap And Roll'),
Row(artist='BlÃ\x83Ãf', length=202.84036, song='Oktober'),
Row(artist='Cosmo Vitelli', length=206.05342, song='Robot Soul (Radio Edit)'),
Row(artist='The Crests', length=182.88281, song='16 Candles'),
Row(artist='Blink-182', length=280.00608, song='All Of This'),
Row(artist='Joan Baez', length=465.24036, song='The Greenwood Side'),
Row(artist=None, length=None, song=None),
Row(artist='Jay-Z / M.O.P.', length=267.51955, song="U Don't Know Remix"),
Row(artist='Eminem', length=251.55873, song='Without Me'),
Row(artist='Midge Ure', length=277.9424, song='Remembrance Day'),
Row(artist='The Black Keys', length=183.27465, song='Keep Your Hands Off Her'),
Row(artist='Derek Webb', length=166.19057, song='Please_ Before I Go'),
Row(artist='I Was A Cub Scout', length=216.99873, song='Close To Me'),
Row(artist=None, length=None, song=None),
Row(artist='Justin Bieber', length=232.69832, song='That Should Be Me'),
Row(artist='Justin Bieber', length=220.89098, song='Somebody To Love'),
Row(artist='The Avett Brothers', length=300.85179, song='I And Love And You'),
Row(artist='Cartola', length=208.92689, song='Sala De RecepÃ\x83Ãf\x83Ãfo'),
Row(artist='Sam Cooke', length=122.04363, song="Ain't Misbehavin"),
Row(artist=None, length=None, song=None),
Row(artist='Richard Grey feat.Sir.Charles', length=411.42812, song="Don't Take My Love Away (Al Velilla)"),
Row(artist='matchbox twenty', length=302.8371, song='Hand Me Down (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Passion Pit', length=372.76689, song="I've Got Your Number"),
Row(artist='System of a Down', length=168.35873, song='Lonely Day'),
Row(artist='Julian Lennon', length=247.30077, song='Saltwater (LP Version)'),
Row(artist='Rihanna', length=267.49342, song="Don't Stop The Music"),
Row(artist='Max Richter', length=78.47138, song="Vladimir's Blues"),
Row(artist='Hannah Montana', length=223.79057, song="You'll Always Find Your Way Back Home"),
Row(artist='Athlete', length=290.58567, song='Chances'),
Row(artist='All Time Low', length=227.49995, song='Lost In Stereo'),
Row(artist='Ne-Yo / LL Cool J', length=264.01914, song='So Sick'),
Row(artist='Vitalic', length=225.51465, song='Repair Machines'),
Row(artist='Bryan Ferry', length=240.84853, song='Amazing Grace (1999 Digital Remaster)'),
Row(artist=None, length=None, song=None),

```



```

Row(artist='Breeder', length=677.61587, song='Twilo Thunder'),
Row(artist='Interpol', length=302.8371, song='PDA (E.P Version)'),
Row(artist='Colbie Caillat', length=224.70485, song='The Little Thin
gs'),
Row(artist='Various Artists - Delicious Vinyl', length=264.25424, so
ng='Bust A Move'),
Row(artist='Colbie Caillat', length=196.17914, song='Bubbly'),
Row(artist='Tavares', length=394.00444, song='Heaven Must Be Missing
An Angel'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Fever Ray', length=225.09669, song='Dry And Dusty'),
Row(artist='The Verve', length=399.51628, song='Come On'),
Row(artist='Cabas', length=215.58812, song='Llega La Noche'),
Row(artist='Scritti Politti', length=286.9024, song='Wood Beez'),
Row(artist='Jonathan Butler', length=195.89179, song='Sugar Candy Ki
sses'),
Row(artist=None, length=None, song=None),
Row(artist='The Tony Rich Project', length=220.62975, song='No Time
Soon'),
Row(artist='Revive', length=234.89261, song='Blink'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='504 Boyz', length=214.12526, song='Wobble Wobble'),
Row(artist='Covenant', length=300.19873, song='XRDS'),
Row(artist='Ezra Furman And The Harpoons', length=233.16853, song='T
he Worm In The Apple'),
Row(artist='Jhonny Rivera', length=211.09506, song='Es mi madre'),
Row(artist='J. Karjalainen & Mustat Lasit', length=336.74404, song
='Sinisten tÃ¤htien alla'),
Row(artist='Alliance Ethnik', length=195.94404, song='SinceritÃ¤
© Et Jalousie'),
Row(artist='Adoniran Barbosa', length=181.68118, song='Trem Das Onz
e'),
Row(artist='Slick Rick', length=275.64363, song='Hey Young World'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='DESTRUCTION', length=274.36363, song='Release From Agon
y'),
Row(artist='Beatsteaks', length=169.24689, song='48/49'),
Row(artist='Amy Winehouse', length=240.43057, song='Back To Black'),
Row(artist='Idlewild', length=306.6771, song='Post-Electric'),
Row(artist='Kate Nash', length=224.7571, song='Foundations'),
Row(artist='Rachid Taha', length=289.74975, song='Kelma'),
Row(artist='Future Breeze', length=371.90485, song='Keep The Fire Bu
rnin'),
Row(artist='The Buggles', length=198.08608, song='Video Killed The R
adio Star'),
Row(artist=None, length=None, song=None),
Row(artist='Simon Webbe', length=240.48281, song='Sunshine (Love Lik
e That)'),
Row(artist=None, length=None, song=None),
Row(artist='The Rural Alberta Advantage', length=168.46322, song='Dr
ain the Blood'),
Row(artist=None, length=None, song=None),

```

```

Row(artist='Dwight Yoakam', length=239.3073, song='You're The One'),
Row(artist='Klaus Doldinger's Passport', length=460.43383, song='Fairy Tale'),
Row(artist='Buckcherry', length=208.66567, song='Next 2 You (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='David Crowder*Band', length=241.89342, song='Only You (Illuminate Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Florence + The Machine', length=219.66322, song='Dog Days Are Over (Radio Edit)'),
Row(artist=None, length=None, song=None),
Row(artist='Monolithe', length=3024.66567, song='Monolithe II'),
Row(artist='The Wombats', length=210.54649, song='Moving To New York (album version)'),
Row(artist='Justin Bieber', length=229.0673, song='One Less Lonely Girl'),
Row(artist='First Aid Kit', length=333.03465, song='Pervigilo'),
Row(artist='Future Loop Foundation', length=327.05261, song='What's Your Name'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='B.o.B', length=269.63546, song='Nothin' On You [feat. Bruno Mars] (Album Version)'),
Row(artist='Crystal', length=251.97669, song='VÃ\x83Ã\x81NDORMADÃ\x83Ã\x81R'),
Row(artist='Mission Of Burma', length=112.16934, song='Go Fun_ Burn Man'),
Row(artist='The Bats', length=249.62567, song='The Looming Past'),
Row(artist=None, length=None, song=None),
Row(artist='Lil Wayne', length=254.27546, song='Ain't That A Bitc h'),
Row(artist='Aesthetic Perfection', length=297.24689, song='The Great Depression'),
Row(artist=None, length=None, song=None),
Row(artist='Marilyn Manson', length=201.40363, song='This Is Halloween'),
Row(artist=None, length=None, song=None),
Row(artist='Live', length=205.37424, song='Selling The Drama'),
Row(artist='Rank 1', length=460.30322, song='L.E.D. There Be Light'),
Row(artist='Depeche Mode', length=314.67057, song='Come Back'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='36 Crazyfists', length=249.52118, song='Waiting On A War'),
Row(artist='Lisa Hannigan', length=245.39383, song='Teeth'),
Row(artist='Bell Orchestre', length=137.82159, song='Stripes'),
Row(artist='Kings Of Leon', length=231.26159, song='Use Somebody'),
Row(artist='Bart Peeters', length=217.52118, song='Er kan altijd nog een tandje bij'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='BjÃ\x83Ã¶rk', length=348.57751, song='Undo'),

```

```

Row(artist='Florence + The Machine', length=258.55955, song='Cosmic
Love'),
Row(artist='Green Day', length=192.9922, song='Brain Stew (Album Ver
sion)'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Vanessa Paradis', length=307.04281, song="L'Amour En So
i"),
Row(artist='UB40', length=196.41424, song='Guilty'),
Row(artist='Florence + The Machine', length=185.3122, song="I'm Not
Calling You A Liar"),
Row(artist=None, length=None, song=None),
Row(artist='Animal Collective', length=294.79138, song='Winters Lov
e'),
Row(artist='Diana Ross', length=248.5024, song='If We Hold On Togeth
er'),
Row(artist='Sugarland', length=247.77098, song='Just Might (Make Me
Believe)'),
Row(artist=None, length=None, song=None),
Row(artist='Eminem', length=250.8273, song='Mockingbird'),
Row(artist='Portishead', length=255.99955, song='Seven Months'),
Row(artist=None, length=None, song=None),
Row(artist='Cradle Of Filth', length=371.66975, song='Better To Reig
n In Hell'),
Row(artist=None, length=None, song=None),
Row(artist='Belle and Sebastian', length=215.37914, song='Expectatio
ns'),
Row(artist=None, length=None, song=None),
Row(artist='Angels and Airwaves', length=305.37098, song='Call To Ar
ms'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='Juan Carlos Baglietto', length=167.49669, song='Versos A
Susana'),
Row(artist='Booka Shade', length=256.9922, song='Karma Car'),
Row(artist='Los Tipitos', length=176.53506, song='Solo Figuras'),
Row(artist='Aqua', length=203.78077, song='My Oh My'),
Row(artist='Kurtis Blow', length=353.88036, song='The Breaks'),
Row(artist=None, length=None, song=None),
Row(artist='Daze', length=204.72118, song='Superhero'),
Row(artist=None, length=None, song=None),
Row(artist='Enya', length=212.29669, song='Boadicea'),
Row(artist='Ed Harcourt', length=210.1024, song='The Last Cigarette'),
Row(artist='Jason Mraz & Colbie Caillat', length=189.6224, song='Luc
ky (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='Amos Lee', length=119.64036, song='Bottom Of the Barre
l'),
Row(artist='Coldplay', length=294.1122, song='Fix You'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='David Crowder*Band', length=322.89914, song='foreverande
ver_ etc...'),
Row(artist='Steve Forbert', length=212.89751, song="Romeo's Tune"),

```

```
Row(artist='Spook And The Guay', length=187.24526, song='Etre Et Avo  
ir')),  
Row(artist='Dot Allison', length=471.97995, song='Substance (Felix D  
a Housecat Remix)'),  
Row(artist=None, length=None, song=None),  
Row(artist='Celtic Woman', length=156.62975, song='Over The Rainbo  
w'),  
Row(artist='Future Rock', length=239.90812, song='Gears'),  
Row(artist='Gentleman', length=199.81016, song='Tranquility (Acousti  
c)'),  
Row(artist='Santogold', length=213.18485, song='Creator [vs. Switch  
and FreQ Nasty]'),  
Row(artist='Zac Brown Band', length=167.99302, song='Mary (Album)'),  
Row(artist=None, length=None, song=None),  
Row(artist='Lily Allen', length=207.77751, song='Chinese'),  
Row(artist=None, length=None, song=None),  
Row(artist='Phil Collins', length=144.32608, song='Transformation (P  
hil Version)'),  
Row(artist=None, length=None, song=None),  
Row(artist='DJ Dizzy', length=221.1522, song='Sexy Bitch'),  
Row(artist='Totta NÃslund', length=236.9824, song='Hemma FÃ  
3Ãre MÃrkret'),  
Row(artist='First Aid Kit', length=149.57669, song='You're Not Comin  
g Home Tonight'),  
Row(artist='Skillet', length=233.32526, song='Rebirthing (Album Vers  
ion)'),  
Row(artist='Griffin House', length=195.7873, song='The Guy That Says  
Goodbye To You Is Out Of His Mind'),  
Row(artist='Kany Garcia', length=235.80689, song='Hoy Ya Me Voy'),  
Row(artist='Pedro Guerra', length=224.13016, song='NiÃos'),  
Row(artist='Boubacar TraorÃ', length=368.8224, song='Soundiat  
a'),  
Row(artist=None, length=None, song=None),  
Row(artist='Los Amigos Invisibles', length=272.56118, song='Ponerte  
En Cuatro'),  
Row(artist='The Prodigy', length=260.07465, song='The Big Gundown'),  
Row(artist=None, length=None, song=None),  
Row(artist='Holger Czukay', length=792.52853, song='0h Lord_ give us  
more money'),  
Row(artist='Alice In Chains', length=149.49832, song='Them Bones'),  
Row(artist='Eagles Of Death Metal', length=158.32771, song='English  
Girl'),  
Row(artist='Bon Jovi', length=224.522, song='It's My Life'),  
Row(artist='SONATA ARCTICA', length=243.35628, song='Respect the Wil  
derness'),  
Row(artist='SOJA', length=370.52036, song='You Dont Know Me'),  
Row(artist='Flyleaf', length=166.86975, song='The Kind'),  
Row(artist=None, length=None, song=None),  
Row(artist='Aerosmith', length=249.57342, song='F.I.N.E.'),  
Row(artist='Donna Lewis', length=240.95302, song='I Love You Always  
Forever ( LP Version )'),  
Row(artist='Bran Van 3000', length=235.28444, song='Everywhere'),  
Row(artist=None, length=None, song=None),  
Row(artist=None, length=None, song=None),  
Row(artist='A Fine Frenzy', length=215.66649, song='Come On_ Come Ou  
t'),  
Row(artist=None, length=None, song=None),
```

```

Row(artist='Phish', length=642.42893, song='Demand'),
Row(artist='Ocote Soul Sounds & Adrian Quesada', length=306.07628, s
ong='La Reja'),
Row(artist='The Smiths', length=357.58975, song='Reel Around The Fou
ntain'),
Row(artist='Phil Collins', length=258.48118, song="You'll Be In My H
eart"),
Row(artist='LCD Soundsystem', length=327.96689, song='North American
Scum'),
Row(artist='Ut', length=313.86077, song='I.D.'),
Row(artist='Erin McKeown', length=338.46812, song='Fast As I Can'),
Row(artist='Kathy Mattea', length=302.96771, song="You'll Never Leav
e Harlan Alive"),
Row(artist='Booka Shade', length=324.44036, song='wasting time'),
Row(artist='Roberto Cardinali', length=226.08934, song='Maybe Toda
y'),
Row(artist=None, length=None, song=None),
Row(artist='Ministry', length=283.66322, song='Work For Love'),
Row(artist='Balkan Beat Box', length=352.60036, song='Bulgarian Chi
cks'),
Row(artist='Postal Service', length=261.25016, song='Sleeping In (Al
bum)'),
Row(artist='Josh Groban', length=249.46893, song='Let Me Fall (From
Cirque du Soleil) (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='Elbow', length=221.85751, song='Grounds For Divorce'),
Row(artist=None, length=None, song=None),
Row(artist='The Police', length=287.81669, song='So Lonely'),
Row(artist='Sidewalk Prophets', length=260.62322, song='You Love Me
Anyway (Album)'),
Row(artist='Cut Copy', length=299.44118, song='Lights & Music'),
Row(artist=None, length=None, song=None),
Row(artist='Borodin String Quartet', length=694.282, song='Mahler: Q
uartet for Piano and Strings in A minor: Allegro'),
Row(artist='Radiohead', length=249.93914, song='Everything In Its Ri
ght Place'),
Row(artist='Eminem', length=250.8273, song='Mockingbird'),
Row(artist='Michael BublÃ\x83Â©', length=225.90649, song='Home (Albu
m Version)'),
Row(artist='Iron And Wine', length=229.90322, song='Dearest Forsaken
(Album)'),
Row(artist=None, length=None, song=None),
Row(artist='Jack Johnson', length=193.88036, song='Posters'),
Row(artist=None, length=None, song=None),
Row(artist='Afromental', length=196.20526, song='Bootycall'),
Row(artist='Los Bunkers', length=351.63383, song='Ahora Que No EstÃ
\x83Â¡s'),
Row(artist=None, length=None, song=None),
Row(artist='M.A. Numminen', length=166.55628, song='TULENLIEKKI'),
Row(artist='Local H', length=223.00689, song='Bound For The Floor'),
Row(artist=None, length=None, song=None),
Row(artist='Il Divo', length=266.70975, song='Sei Parte Ormai Di M
e'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),

```

```

Row(artist='Kevin Michael feat. Wyclef Jean', length=224.9922, song="It Don't Make Any Difference To Me (1 Love) Feat. Wyclef (Radio Edit)"),
Row(artist=None, length=None, song=None),
Row(artist='Aloe Blacc', length=225.2273, song='Take Me Back'),
Row(artist=None, length=None, song=None),
Row(artist='The George Shearing Quintet with Nancy Wilson', length=137.06404, song='Ghost Of Yesterday (2001 Digital Remaster)'),
Row(artist=None, length=None, song=None),
Row(artist='Lonnie Gordon', length=181.21098, song='Catch You Baby (Steve Pitron & Max Sanna Radio Edit)'),
Row(artist='Zero 7', length=263.33995, song='Pop Art Blue'),
Row(artist=None, length=None, song=None),
Row(artist='Franz Ferdinand', length=237.89669, song='The Fallen (Ruined By Justice)'),
Row(artist=None, length=None, song=None),
Row(artist='Copywrite', length=224.80934, song='Jeah'),
Row(artist='Café Tacvba', length=135.33995, song='Las flores'),
Row(artist='Jesse Cook', length=215.30077, song='Falling From Grace'),
Row(artist=None, length=None, song=None),
Row(artist='Soundgarden', length=345.83465, song='Blow Up The Outside World'),
Row(artist='12 Stones', length=225.98485, song='The Way I Feel (Not Our Master)'),
Row(artist=None, length=None, song=None),
Row(artist='Daft Punk', length=472.05832, song='Phoenix (Basement Jaxx Remix)'),
Row(artist='Brooklyn Bounce', length=359.13098, song='Get Ready To Bounce Recall 08'),
Row(artist='Noisy Stylus', length=206.57587, song='Stand Back'),
Row(artist='Dead Kennedys', length=208.1171, song='California Å\x83Å\x9cber Alles'),
Row(artist='My Chemical Romance', length=231.10485, song="Honey_ This Mirror Isn't Big Enough For The Two Of Us"),
Row(artist='Peaches', length=215.17016, song='Shake Yer Dix'),
Row(artist='Ashley MacIsaac', length=222.14485, song='Captain America'),
Row(artist='Tony Lucca/Jay Nash/Matt Duke', length=306.46812, song='Pretty Things'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='10cc', length=267.98975, song='Dreadlock Holiday'),
Row(artist=None, length=None, song=None),
Row(artist='Theuns Jordaan', length=284.02893, song='Beautiful In Beaufort Wes'),
Row(artist='Suzi Quatro', length=197.14567, song="Keep A Knockin'"),
Row(artist='Righteous Brothers', length=200.93342, song='Hung On You'),
Row(artist='En Vogue', length=289.01832, song="Don't Let Go (Love)"),
Row(artist='Kate Ryan', length=220.83873, song='DÃ\x83Ã©s enchantÃ\x83Ã©e'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='The Redwalls', length=213.81179, song='Universal Blue

```

```

s'),
Row(artist='Muse', length=146.25914, song='Overdue'),
Row(artist=None, length=None, song=None),
Row(artist=None, length=None, song=None),
Row(artist='the bird and the bee', length=181.02812, song='Heard It
On The Radio'),
Row(artist='Pandora's Box', length=164.25751, song='The Want Ad'),
Row(artist='50 Cent / Lloyd Banks / Eminem', length=248.94649, song
="Don't Push Me"),
Row(artist='Super Furry Animals', length=213.99465, song='Hello Suns
hine'),
Row(artist='Incubus', length=263.60118, song='Are You In?'),
Row(artist='Caribou (formerly Dan Snaith's Manitoba)', length=257.93
261, song='Bijoux'),
Row(artist='The Presidents of the United States of America', length=
224.70485, song='Naked and Famous'),
Row(artist='BBE', length=251.61098, song='Seven Days And One Week (R
adio Edit)'),
Row(artist='Serj Tankian', length=263.60118, song='Praise The Lord A
nd Pass The Ammunition (Album Version)'),
Row(artist=None, length=None, song=None),
Row(artist='Willy Mason', length=146.02404, song='Still A Fly'),
Row(artist='Bananarama', length=219.0624, song='Venus'),
Row(artist='Color Me Badd', length=256.93995, song='The Earth_ The S
un_ The Rain (LP Version)'),
Row(artist='Anthrax', length=200.07138, song='Refuse To Be Denied'),
Row(artist='DENNIS EDWARDS FEAT. THE WEATHER GIRLS', length=240.1432
2, song="Don't Look Any Further - 13 Studio Extended Version"),
Row(artist='Toasters', length=279.09179, song='Sitting on Top of the
World'),
Row(artist='Iron And Wine', length=153.05098, song='Naked As We Ca
n'),
Row(artist='Evanescence', length=236.12036, song='Bring Me To Lif
e'),
Row(artist='The Thermals', length=141.47873, song='No Culture Icons
(Album)'),
Row(artist='Gramophonedzie', length=401.29261, song="Why Don't You
(Original Mix)"),
Row(artist='Violeta Parra', length=190.77179, song='La Inhumana'),
Row(artist='Lucas Masciano', length=520.77669, song='Besos'),
Row(artist='The Police', length=289.85424, song='So Lonely'),
Row(artist='Luis Fonsi', length=250.53995, song='Vivo Muriendo'),
Row(artist='Axwell_ Ingrosso_ Angello_ Laidback Luke feat. Deborah C
ox', length=161.72363, song='Leave The World Behind'),
Row(artist=None, length=None, song=None),
Row(artist='Juan Diego FlÃ³rez / Daniel Binelli / Fort Worth Sy
mphony Orchestra / Miguel Harth-Bedoya', length=278.09914, song='El D
Ã¡xada Que Me Quieras'),
Row(artist='Eminem', length=205.81832, song='FACK'),
Row(artist='The Ordinary Boys', length=162.87302, song='Boys Will Be
Boys'),
Row(artist='Five Iron Frenzy', length=236.09424, song='Canada'),
Row(artist='Joan Baez', length=233.9522, song='John Riley'),
Row(artist='Whiskeytown', length=168.12363, song='My Hometown'),
Row(artist='Carpenters', length=238.39302, song='Yesterday Once Mor
e'),
Row(artist='Yvonne Elliman', length=176.09098, song="If I Can't Have

```

```

You"),
  Row(artist='Hayko Cepkin', length=278.41261, song='Gelin Olmus'),
  Row(artist='Noisettes', length=210.88608, song='When You Were Young'),
  Row(artist='Hinder', length=262.05995, song='Lips Of An Angel'),
  Row(artist='Blink-182', length=246.12526, song="Adam's Song"),
  Row(artist='Busta Rhymes', length=220.78649, song='Everything Remains Raw (Explicit Version)'),
  Row(artist='Foolish Things', length=266.86649, song='Who Can Compare'),
  Row(artist=None, length=None, song=None),
  Row(artist='Morgan Page', length=235.54567, song='Strange Condition'),
  Row(artist='Kanye West', length=238.73261, song='Bad News'),
  Row(artist='DeGarmo & Key', length=166.79138, song='When The Son Begins To Reign (Degarmo And Key Album Version)'),
  Row(artist=None, length=None, song=None),
  Row(artist='The Dears', length=289.95873, song='Lost In The Plot'),
  Row(artist='Shakira ft. Wyclef Jean', length=215.90159, song="Hips Don't Lie"),
  Row(artist="Slash's Snakepit", length=294.63465, song='Lower'),
  Row(artist='Miley Cyrus', length=166.50404, song='Wake Up America'),
  Row(artist=None, length=None, song=None),
  Row(artist='The Black Keys', length=211.01669, song='Tighten Up'),
  Row(artist=None, length=None, song=None),
  Row(artist='Swans', length=365.24363, song='Love Will Save You'),
  Row(artist='Shriekback', length=290.84689, song='Underwaterboys'),
  Row(artist='Portishead', length=259.99628, song='It Could Be Sweet'),
  Row(artist=None, length=None, song=None),
  ...]

```

```
In [11]: avg_len=data_churn.agg(avg('length')).take(1)[0][0]
```

```
In [12]: to_replace='missing_val'
```

```
In [13]: data_churn=data_churn.fillna(to_replace,subset=['artist','song'])
```

```
In [14]: data_churn=data_churn.fillna(avg_len,subset=['length'])
```

```
In [ ]:
```



```
In [15]: data_churn.select([count(when(isnull(c), c)).alias(c) for c in data_churn.columns]).show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
|artist|auth|firstName|gender|itemInSession|lastName|length|level|location|method|page|registration|sessionId|song|status|ts|userAgent|userId|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
|      0|      0|      8346|      8346|      0|      8346|      0|      0|      8346|
8346|      0|      0|      8346|      0|      0|      0|      0|      8346|
0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
```

Let us drop the remainder rows with NaN's

```
In [16]: to_replace2='missing_val2'
```

```
In [17]: data_churn=data_churn.fillna(to_replace2,subset=['firstName','gender',
,'lastName','location','userAgent'])
```

```
In [18]: data_churn.select([count(when(isnull(c), c)).alias(c) for c in data_churn.columns]).show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
|artist|auth|firstName|gender|itemInSession|lastName|length|level|location|method|page|registration|sessionId|song|status|ts|userAgent|userId|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|
0|      0|      0|      8346|      0|      0|      0|      0|      0|      0|
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
```

In [19]: data_churn.columns

```
Out[19]: ['artist',
          'auth',
          'firstName',
          'gender',
          'itemInSession',
          'lastName',
          'length',
          'level',
          'location',
          'method',
          'page',
          'registration',
          'sessionId',
          'song',
          'status',
          'ts',
          'userAgent',
          'userId']
```

```
In [20]: #data_churn.select('registration').collect()
avg_reg=int(data_churn.agg(avg('registration')).take(1)[0][0])
print(avg_reg)

1535358834084
```

In [21]: data_churn=data_churn.fillna(avg_reg,subset=['registration'])

In [22]: data_churn.select([count(when(isnull(c), c)).alias(c) for c in data_churn.columns]).show()

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|artist|auth|firstName|gender|itemInSession|lastName|length|level|location|method|page|registration|sessionId|song|status|ts|userAgent|userId|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
```

In []:

Exploratory Data Analysis

When you're working with the full dataset, perform EDA by loading a small subset of the data and doing basic manipulations within Spark. In this workspace, you are already provided a small subset of data you can explore.

Define Churn

Once you've done some preliminary analysis, create a column `Churn` to use as the label for your model. I suggest using the `Cancellation Confirmation` events to define your churn, which happen for both paid and free users. As a bonus task, you can also look into the `Downgrade` events.

Explore Data

Once you've defined churn, perform some exploratory data analysis to observe the behavior for users who stayed vs users who churned. You can start by exploring aggregates on these two groups of users, observing how much of a specific action they experienced per a certain time unit or number of songs played.

```
In [23]: flag_churn_event = udf(lambda x: 1 if (x == "Submit Downgrade" or x ==
'Downgrade' or x == 'Cancellation Confirmation' or x == 'Cancel') else 0,
IntegerType())
```

Creating the churn variable

```
In [24]: data_churn = data_churn.withColumn("churn", flag_churn_event("page"))
```

```
In [25]: data_churn.take(2)
```

```
Out[25]: [Row(artist='Martha Tilston', auth='Logged In', firstName='Colin', ge
nder='M', itemInSession=50, lastName='Freeman', length=277.89016, lev
el='paid', location='Bakersfield, CA', method='PUT', page='NextSong',
registration=1538173362000, sessionId=29, song='Rockpools', status=20
0, ts=1538352117000, userAgent='Mozilla/5.0 (Windows NT 6.1; WOW64; r
v:31.0) Gecko/20100101 Firefox/31.0', userId='30', churn=0),
Row(artist='Five Iron Frenzy', auth='Logged In', firstName='Micah',
gender='M', itemInSession=79, lastName='Long', length=236.09424, leve
l='free', location='Boston-Cambridge-Newton, MA-NH', method='PUT', pa
ge='NextSong', registration=1538331630000, sessionId=8, song='Canad
a', status=200, ts=1538352180000, userAgent='\"Mozilla/5.0 (Windows NT
6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/37.0.2062.1
03 Safari/537.36\"', userId='9', churn=0)]
```

```
In [26]: data_churn.select('churn').distinct().collect()
```

```
Out[26]: [Row(churn=1), Row(churn=0)]
```

Creating the hour variable

```
In [27]: get_hour = udf(lambda x: datetime.datetime.fromtimestamp(x / 3600.0).hour)
```

```
In [28]: data_churn = data_churn.withColumn("hour", get_hour(data_churn.ts))
```

```
In [29]: data_churn.show(2)
```

```
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
--+-+-----+
|          artist|          auth|firstName|gender|itemInSession|lastName|
length|level|          location|method|    page| registration|sessi
onId|    song|status|          ts|          userAgent|userId|churn
|hour|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
--+-+-----+
| Martha Tilston|Logged In|    Colin|    M|          50| Freeman|2
77.89016| paid|    Bakersfield, CA|  PUT|NextSong|1538173362000|
29|Rockpools|    200|1538352117000|Mozilla/5.0 (Wind...|    30|    0|
20|
|Five Iron Frenzy|Logged In|    Micah|    M|          79|    Long|2
36.09424| free|Boston-Cambridge-...|  PUT|NextSong|1538331630000|
8|    Canada|    200|1538352180000|"Mozilla/5.0 (Win...|    9|    0|
20|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
--+-+-----+
only showing top 2 rows
```

```
In [30]: data_churn.select(['page']).distinct().collect()
```

```
Out[30]: [Row(page='Cancel'),
Row(page='Submit Downgrade'),
Row(page='Thumbs Down'),
Row(page='Home'),
Row(page='Downgrade'),
Row(page='Roll Advert'),
Row(page='Logout'),
Row(page='Save Settings'),
Row(page='Cancellation Confirmation'),
Row(page='About'),
Row(page='Submit Registration'),
Row(page='Settings'),
Row(page='Login'),
Row(page='Register'),
Row(page='Add to Playlist'),
Row(page='Add Friend'),
Row(page='NextSong'),
Row(page='Thumbs Up'),
Row(page='Help'),
Row(page='Upgrade'),
Row(page='Error'),
Row(page='Submit Upgrade')]
```

```
In [31]: data_churn.columns
```

```
Out[31]: ['artist',
'auth',
'firstName',
'gender',
'itemInSession',
'lastName',
'length',
'level',
'location',
'method',
'page',
'registration',
'sessionId',
'song',
'status',
'ts',
'userAgent',
'userId',
'churn',
'hour']
```

```
In [32]: data_churn.select(['userId', 'page']).distinct().collect()
```

```
Out[32]: [Row(userId='88', page='Home'),
Row(userId='88', page='NextSong'),
Row(userId='13', page='Add to Playlist'),
Row(userId='87', page='Home'),
Row(userId='131', page='Upgrade'),
Row(userId='59', page='Add Friend'),
Row(userId='40', page='Help'),
Row(userId='92', page='Settings'),
Row(userId='12', page='Thumbs Up'),
Row(userId='50', page='Error'),
Row(userId='81', page='Upgrade'),
Row(userId='81', page='Submit Upgrade'),
Row(userId='53', page='Submit Upgrade'),
Row(userId='99', page='Downgrade'),
Row(userId='200017', page='NextSong'),
Row(userId='300015', page='Add Friend'),
Row(userId='300007', page='Roll Advert'),
Row(userId='300005', page='Thumbs Up'),
Row(userId='300014', page='Downgrade'),
Row(userId='300019', page='Upgrade'),
Row(userId='300002', page='Help'),
Row(userId='300019', page='About'),
Row(userId='131', page='Help'),
Row(userId='88', page='Thumbs Up'),
Row(userId='122', page='Home'),
Row(userId='93', page='Upgrade'),
Row(userId='4', page='Add Friend'),
Row(userId='83', page='Error'),
Row(userId='139', page='Downgrade'),
Row(userId='155', page='Roll Advert'),
Row(userId='100023', page='Downgrade'),
Row(userId='100018', page='About'),
Row(userId='200019', page='Submit Upgrade'),
Row(userId='200019', page='Thumbs Up'),
Row(userId='200020', page='Settings'),
Row(userId='200011', page='Add to Playlist'),
Row(userId='200001', page='Save Settings'),
Row(userId='300021', page='Downgrade'),
Row(userId='40', page='Thumbs Down'),
Row(userId='132', page='NextSong'),
Row(userId='121', page='Logout'),
Row(userId='70', page='Upgrade'),
Row(userId='49', page='Thumbs Down'),
Row(userId='42', page='Save Settings'),
Row(userId='25', page='Settings'),
Row(userId='57', page='Thumbs Down'),
Row(userId='90', page='Home'),
Row(userId='3', page='Roll Advert'),
Row(userId='72', page='Thumbs Down'),
Row(userId='200019', page='Roll Advert'),
Row(userId='65', page='Thumbs Up'),
Row(userId='136', page='Roll Advert'),
Row(userId='35', page='Settings'),
Row(userId='111', page='Logout'),
Row(userId='141', page='Downgrade'),
Row(userId='66', page='Add to Playlist'),
Row(userId='104', page='NextSong'),
```

```
Row(userId='142', page='Save Settings'),
Row(userId='2', page='Logout'),
Row(userId='12', page='Thumbs Down'),
Row(userId='77', page='Add to Playlist'),
Row(userId='125', page='Cancel'),
Row(userId='147', page='Upgrade'),
Row(userId='100013', page='Roll Advert'),
Row(userId='100018', page='NextSong'),
Row(userId='100023', page='Roll Advert'),
Row(userId='200021', page='Save Settings'),
Row(userId='300001', page='Thumbs Up'),
Row(userId='300020', page='Settings'),
Row(userId='95', page='Logout'),
Row(userId='70', page='NextSong'),
Row(userId='100', page='Add to Playlist'),
Row(userId='105', page='Submit Upgrade'),
Row(userId='96', page='Thumbs Down'),
Row(userId='66', page='NextSong'),
Row(userId='127', page='Logout'),
Row(userId='104', page='Logout'),
Row(userId='145', page='Add Friend'),
Row(userId='10', page='Logout'),
Row(userId='51', page='Error'),
Row(userId='56', page='Save Settings'),
Row(userId='103', page='Submit Downgrade'),
Row(userId='126', page='Downgrade'),
Row(userId='21', page='Help'),
Row(userId='88', page='Settings'),
Row(userId='152', page='Error'),
Row(userId='147', page='Downgrade'),
Row(userId='139', page='Logout'),
Row(userId='84', page='Help'),
Row(userId='100009', page='Upgrade'),
Row(userId='100007', page='Save Settings'),
Row(userId='200022', page='Thumbs Down'),
Row(userId='200008', page='Add to Playlist'),
Row(userId='200017', page='Help'),
Row(userId='300008', page='NextSong'),
Row(userId='300016', page='Add Friend'),
Row(userId='54', page='Downgrade'),
Row(userId='102', page='Thumbs Up'),
Row(userId='29', page='Downgrade'),
Row(userId='8', page='Roll Advert'),
Row(userId='138', page='Help'),
Row(userId='23', page='NextSong'),
Row(userId='153', page='Add to Playlist'),
Row(userId='29', page='Cancel'),
Row(userId='100025', page='Cancellation Confirmation'),
Row(userId='200019', page='Home'),
Row(userId='200018', page='Add Friend'),
Row(userId='300016', page='NextSong'),
Row(userId='300006', page='About'),
Row(userId='39', page='Roll Advert'),
Row(userId='132', page='Home'),
Row(userId='106', page='Add Friend'),
Row(userId='58', page='Upgrade'),
Row(userId='66', page='Home'),
```



```
Row(userId='117', page='Settings'),
Row(userId='129', page='NextSong'),
Row(userId='120', page='Thumbs Down'),
Row(userId='43', page='NextSong'),
Row(userId='43', page='Thumbs Down'),
Row(userId='70', page='About'),
Row(userId='71', page='About'),
Row(userId='94', page='Add Friend'),
Row(userId='128', page='Save Settings'),
Row(userId='300023', page='Downgrade'),
Row(userId='300025', page='Upgrade'),
Row(userId='93', page='Thumbs Up'),
Row(userId='132', page='Help'),
Row(userId='52', page='Upgrade'),
Row(userId='28', page='Thumbs Down'),
Row(userId='37', page='Roll Advert'),
Row(userId='114', page='Roll Advert'),
Row(userId='75', page='About'),
Row(userId='152', page='Roll Advert'),
Row(userId='152', page='Thumbs Up'),
Row(userId='52', page='Downgrade'),
Row(userId='100014', page='Roll Advert'),
Row(userId='200014', page='Thumbs Down'),
Row(userId='300013', page='Add Friend'),
Row(userId='118', page='Add to Playlist'),
Row(userId='145', page='Thumbs Up'),
Row(userId='2', page='Add Friend'),
Row(userId='109', page='Save Settings'),
Row(userId='148', page='Logout'),
Row(userId='89', page='Add Friend'),
Row(userId='152', page='Thumbs Down'),
Row(userId='72', page='Error'),
Row(userId='152', page='About'),
Row(userId='100004', page='NextSong'),
Row(userId='100012', page='Upgrade'),
Row(userId='200008', page='NextSong'),
Row(userId='200022', page='Settings'),
Row(userId='200009', page='Add to Playlist'),
Row(userId='200008', page='Downgrade'),
Row(userId='200001', page='Upgrade'),
Row(userId='200007', page='Help'),
Row(userId='300010', page='Help'),
Row(userId='300006', page='Thumbs Down'),
Row(userId='9', page='NextSong'),
Row(userId='38', page='Home'),
Row(userId='24', page='NextSong'),
Row(userId='150', page='NextSong'),
Row(userId='8', page='Thumbs Up'),
Row(userId='11', page='Downgrade'),
Row(userId='101', page='Cancel'),
Row(userId='24', page='Downgrade'),
Row(userId='62', page='Roll Advert'),
Row(userId='153', page='Thumbs Up'),
Row(userId='22', page='Add Friend'),
Row(userId='98', page='About'),
Row(userId='127', page='Error'),
Row(userId='100022', page='NextSong'),
```

```
Row(userId='200010', page='Logout'),
Row(userId='200015', page='Error'),
Row(userId='300009', page='NextSong'),
Row(userId='300022', page='Add to Playlist'),
Row(userId='118', page='Thumbs Down'),
Row(userId='32', page='Logout'),
Row(userId='85', page='Settings'),
Row(userId='45', page='Thumbs Down'),
Row(userId='70', page='Downgrade'),
Row(userId='116', page='Home'),
Row(userId='79', page='Save Settings'),
Row(userId='136', page='About'),
Row(userId='22', page='Roll Advert'),
Row(userId='100008', page='Thumbs Down'),
Row(userId='200003', page='Add to Playlist'),
Row(userId='200009', page='Home'),
Row(userId='200017', page='Error'),
Row(userId='300023', page='Thumbs Down'),
Row(userId='300023', page='Help'),
Row(userId='101', page='Help'),
Row(userId='6', page='NextSong'),
Row(userId='73', page='Roll Advert'),
Row(userId='18', page='About'),
Row(userId='106', page='Roll Advert'),
Row(userId='16', page='NextSong'),
Row(userId='129', page='Thumbs Up'),
Row(userId='18', page='Cancellation Confirmation'),
Row(userId='140', page='Submit Downgrade'),
Row(userId='71', page='Home'),
Row(userId='88', page='Upgrade'),
Row(userId='33', page='About'),
Row(userId='25', page='Roll Advert'),
Row(userId='100018', page='Save Settings'),
Row(userId='100014', page='NextSong'),
Row(userId='200020', page='Add Friend'),
Row(userId='300011', page='Home'),
Row(userId='300004', page='Settings'),
Row(userId='115', page='Thumbs Up'),
Row(userId='61', page='Upgrade'),
Row(userId='105', page='Roll Advert'),
Row(userId='106', page='NextSong'),
Row(userId='99', page='NextSong'),
Row(userId='112', page='Logout'),
Row(userId='72', page='Thumbs Up'),
Row(userId='56', page='Submit Upgrade'),
Row(userId='200002', page='Settings'),
Row(userId='200015', page='Add Friend'),
Row(userId='300021', page='Thumbs Up'),
Row(userId='300007', page='Settings'),
Row(userId='300019', page='NextSong'),
Row(userId='142', page='Help'),
Row(userId='61', page='Home'),
Row(userId='75', page='NextSong'),
Row(userId='37', page='Settings'),
Row(userId='17', page='Add to Playlist'),
Row(userId='89', page='Thumbs Up'),
Row(userId='122', page='Cancellation Confirmation'),
```

```
Row(userId='75', page='Logout'),
Row(userId='75', page='Settings'),
Row(userId='152', page='Submit Upgrade'),
Row(userId='100018', page='Add to Playlist'),
Row(userId='100018', page='Downgrade'),
Row(userId='100021', page='Thumbs Up'),
Row(userId='200021', page='Submit Upgrade'),
Row(userId='200020', page='Submit Upgrade'),
Row(userId='200005', page='Thumbs Up'),
Row(userId='200020', page='Cancel'),
Row(userId='300004', page='Thumbs Up'),
Row(userId='101', page='NextSong'),
Row(userId='95', page='Settings'),
Row(userId='118', page='Settings'),
Row(userId='73', page='Help'),
Row(userId='49', page='NextSong'),
Row(userId='4', page='Thumbs Up'),
Row(userId='155', page='Thumbs Down'),
Row(userId='151', page='Help'),
Row(userId='100019', page='Logout'),
Row(userId='200022', page='Thumbs Up'),
Row(userId='200023', page='Logout'),
Row(userId='300009', page='Add Friend'),
Row(userId='300001', page='Cancel'),
Row(userId='25', page='Home'),
Row(userId='92', page='Help'),
Row(userId='69', page='Help'),
Row(userId='20', page='Logout'),
Row(userId='8', page='Upgrade'),
Row(userId='56', page='Add Friend'),
Row(userId='133', page='Add Friend'),
Row(userId='49', page='Save Settings'),
Row(userId='96', page='Save Settings'),
Row(userId='96', page='Submit Upgrade'),
Row(userId='100015', page='Thumbs Down'),
Row(userId='100004', page='Roll Advert'),
Row(userId='100012', page='Logout'),
Row(userId='100019', page='Add to Playlist'),
Row(userId='100014', page='Home'),
Row(userId='200020', page='Add to Playlist'),
Row(userId='200021', page='Thumbs Down'),
Row(userId='200009', page='Roll Advert'),
Row(userId='200008', page='Help'),
Row(userId='58', page='Settings'),
Row(userId='73', page='Home'),
Row(userId='59', page='Settings'),
Row(userId='104', page='Roll Advert'),
Row(userId='36', page='Add Friend'),
Row(userId='137', page='Upgrade'),
Row(userId='49', page='Submit Upgrade'),
Row(userId='100024', page='Roll Advert'),
Row(userId='100008', page='Downgrade'),
Row(userId='200006', page='Help'),
Row(userId='200003', page='NextSong'),
Row(userId='200005', page='Submit Upgrade'),
Row(userId='300013', page='Thumbs Up'),
Row(userId='300009', page='Settings'),
```

```
Row(userId='124', page='Home'),
Row(userId='117', page='Upgrade'),
Row(userId='16', page='Thumbs Down'),
Row(userId='74', page='About'),
Row(userId='95', page='Error'),
Row(userId='83', page='Downgrade'),
Row(userId='47', page='Settings'),
Row(userId='100001', page='Thumbs Up'),
Row(userId='100006', page='NextSong'),
Row(userId='300001', page='NextSong'),
Row(userId='300021', page='Thumbs Down'),
Row(userId='300024', page='Add Friend'),
Row(userId='28', page='Help'),
Row(userId='93', page='Roll Advert'),
Row(userId='113', page='NextSong'),
Row(userId='17', page='Roll Advert'),
Row(userId='91', page='Thumbs Up'),
Row(userId='117', page='Logout'),
Row(userId='17', page='Settings'),
Row(userId='17', page='Downgrade'),
Row(userId='74', page='Submit Upgrade'),
Row(userId='21', page='Thumbs Down'),
Row(userId='124', page='Save Settings'),
Row(userId='46', page='Upgrade'),
Row(userId='28', page='Downgrade'),
Row(userId='63', page='NextSong'),
Row(userId='100009', page='Add Friend'),
Row(userId='100016', page='Help'),
Row(userId='300004', page='Logout'),
Row(userId='65', page='Roll Advert'),
Row(userId='132', page='Add Friend'),
Row(userId='117', page='NextSong'),
Row(userId='135', page='NextSong'),
Row(userId='26', page='Thumbs Up'),
Row(userId='154', page='Logout'),
Row(userId='139', page='Thumbs Down'),
Row(userId='62', page='About'),
Row(userId='200024', page='Add Friend'),
Row(userId='200011', page='Error'),
Row(userId='54', page='NextSong'),
Row(userId='78', page='Add to Playlist'),
Row(userId='27', page='Upgrade'),
Row(userId='73', page='Thumbs Up'),
Row(userId='12', page='Upgrade'),
Row(userId='11', page='Thumbs Up'),
Row(userId='119', page='Add to Playlist'),
Row(userId='97', page='Settings'),
Row(userId='7', page='Settings'),
Row(userId='145', page='About'),
Row(userId='123', page='Help'),
Row(userId='100023', page='Cancel'),
Row(userId='200009', page='Thumbs Down'),
Row(userId='300021', page='NextSong'),
Row(userId='300015', page='Submit Downgrade'),
Row(userId='300004', page='Add Friend'),
Row(userId='65', page='Settings'),
Row(userId='102', page='NextSong'),
```

```
Row(userId='44', page='Logout'),
Row(userId='10', page='NextSong'),
Row(userId='8', page='Home'),
Row(userId='81', page='Thumbs Down'),
Row(userId='149', page='Logout'),
Row(userId='92', page='Error'),
Row(userId='123', page='Thumbs Down'),
Row(userId='154', page='Add Friend'),
Row(userId='100005', page='NextSong'),
Row(userId='100019', page='Thumbs Down'),
Row(userId='100016', page='Add Friend'),
Row(userId='300017', page='Add to Playlist'),
Row(userId='30', page='Roll Advert'),
Row(userId='113', page='Home'),
Row(userId='132', page='Submit Upgrade'),
Row(userId='105', page='Thumbs Up'),
Row(userId='73', page='Add to Playlist'),
Row(userId='114', page='NextSong'),
Row(userId='50', page='Home'),
Row(userId='75', page='Add Friend'),
Row(userId='42', page='Add to Playlist'),
Row(userId='60', page='Home'),
Row(userId='97', page='Add to Playlist'),
Row(userId='122', page='Thumbs Up'),
Row(userId='73', page='Cancel'),
Row(userId='25', page='Save Settings'),
Row(userId='26', page='Upgrade'),
Row(userId='100006', page='Home'),
Row(userId='200002', page='Help'),
Row(userId='9', page='Thumbs Up'),
Row(userId='74', page='Thumbs Up'),
Row(userId='', page='Home'),
Row(userId='146', page='Thumbs Up'),
Row(userId='12', page='Add Friend'),
Row(userId='12', page='Logout'),
Row(userId='54', page='Save Settings'),
Row(userId='66', page='Settings'),
Row(userId='85', page='Help'),
Row(userId='14', page='Downgrade'),
Row(userId='126', page='Settings'),
Row(userId='127', page='Save Settings'),
Row(userId='100007', page='NextSong'),
Row(userId='200015', page='Roll Advert'),
Row(userId='200023', page='Roll Advert'),
Row(userId='200025', page='About'),
Row(userId='300015', page='Error'),
Row(userId='300019', page='Add to Playlist'),
Row(userId='25', page='Add Friend'),
Row(userId='126', page='Add to Playlist'),
Row(userId='94', page='Add to Playlist'),
Row(userId='105', page='NextSong'),
Row(userId='73', page='Thumbs Down'),
Row(userId='136', page='Logout'),
Row(userId='37', page='About'),
Row(userId='78', page='Settings'),
Row(userId='42', page='Downgrade'),
Row(userId='19', page='Add Friend'),
```

```
Row(userId='50', page='Save Settings'),
Row(userId='26', page='Home'),
Row(userId='41', page='Thumbs Down'),
Row(userId='64', page='Logout'),
Row(userId='108', page='Settings'),
Row(userId='79', page='Logout'),
Row(userId='83', page='Add to Playlist'),
Row(userId='39', page='Submit Downgrade'),
Row(userId='100024', page='Thumbs Up'),
Row(userId='100006', page='Cancel'),
Row(userId='200018', page='Upgrade'),
Row(userId='78', page='Help'),
Row(userId='95', page='Thumbs Up'),
Row(userId='142', page='Upgrade'),
Row(userId='55', page='Roll Advert'),
Row(userId='49', page='Thumbs Up'),
Row(userId='11', page='Settings'),
Row(userId='19', page='Thumbs Down'),
Row(userId='140', page='About'),
Row(userId='103', page='Logout'),
Row(userId='132', page='Error'),
Row(userId='89', page='Save Settings'),
Row(userId='7', page='Add to Playlist'),
Row(userId='91', page='Submit Upgrade'),
Row(userId='126', page='Error'),
Row(userId='111', page='Error'),
Row(userId='72', page='Add Friend'),
Row(userId='53', page='Cancellation Confirmation'),
Row(userId='200023', page='Help'),
Row(userId='200022', page='Add Friend'),
Row(userId='300005', page='Add to Playlist'),
Row(userId='92', page='Roll Advert'),
Row(userId='81', page='Save Settings'),
Row(userId='77', page='Downgrade'),
Row(userId='18', page='Logout'),
Row(userId='136', page='Add to Playlist'),
Row(userId='6', page='Upgrade'),
Row(userId='30', page='Settings'),
Row(userId='109', page='Downgrade'),
Row(userId='25', page='Logout'),
Row(userId='56', page='Home'),
Row(userId='139', page='Home'),
Row(userId='100015', page='Roll Advert'),
Row(userId='200022', page='NextSong'),
Row(userId='300008', page='Thumbs Up'),
Row(userId='300015', page='Help'),
Row(userId='92', page='NextSong'),
Row(userId='92', page='Thumbs Up'),
Row(userId='42', page='Home'),
Row(userId='136', page='Help'),
Row(userId='75', page='Downgrade'),
Row(userId='98', page='Help'),
Row(userId='65', page='Submit Upgrade'),
Row(userId='15', page='Save Settings'),
Row(userId='136', page='Settings'),
Row(userId='149', page='Help'),
Row(userId='90', page='NextSong'),
```

```
Row(userId='104', page='Thumbs Down'),
Row(userId='9', page='Submit Downgrade'),
Row(userId='116', page='Roll Advert'),
Row(userId='70', page='Error'),
Row(userId='100011', page='Thumbs Down'),
Row(userId='200012', page='Roll Advert'),
Row(userId='300009', page='Logout'),
Row(userId='25', page='NextSong'),
Row(userId='131', page='Add Friend'),
Row(userId='81', page='NextSong'),
Row(userId='100', page='Settings'),
Row(userId='100', page='About'),
Row(userId='115', page='Add Friend'),
Row(userId='85', page='About'),
Row(userId='104', page='Home'),
Row(userId='148', page='Thumbs Up'),
Row(userId='118', page='Logout'),
Row(userId='80', page='Add Friend'),
Row(userId='7', page='Error'),
Row(userId='75', page='Roll Advert'),
Row(userId='57', page='Settings'),
Row(userId='200010', page='Help'),
Row(userId='200016', page='Thumbs Down'),
Row(userId='40', page='NextSong'),
Row(userId='61', page='NextSong'),
Row(userId='145', page='Downgrade'),
Row(userId='70', page='Cancel'),
Row(userId='24', page='Thumbs Down'),
Row(userId='100005', page='Add to Playlist'),
Row(userId='100014', page='Cancel'),
Row(userId='200006', page='NextSong'),
Row(userId='200023', page='About'),
Row(userId='200001', page='Thumbs Up'),
Row(userId='200011', page='About'),
Row(userId='300017', page='Thumbs Down'),
Row(userId='300012', page='Settings'),
Row(userId='137', page='Logout'),
Row(userId='124', page='Settings'),
Row(userId='32', page='NextSong'),
Row(userId='39', page='Help'),
Row(userId='45', page='Thumbs Up'),
Row(userId='56', page='Upgrade'),
Row(userId='29', page='About'),
Row(userId='35', page='About'),
Row(userId='57', page='NextSong'),
Row(userId='45', page='Submit Upgrade'),
Row(userId='62', page='Error'),
Row(userId='132', page='Roll Advert'),
Row(userId='108', page='Upgrade'),
Row(userId='148', page='Thumbs Down'),
Row(userId='153', page='About'),
Row(userId='100012', page='NextSong'),
Row(userId='100014', page='Thumbs Up'),
Row(userId='300008', page='Downgrade'),
Row(userId='300018', page='About'),
Row(userId='54', page='Thumbs Up'),
Row(userId='33', page='NextSong'),
```

```
Row(userId='137', page='Settings'),
Row(userId='129', page='Add to Playlist'),
Row(userId='18', page='Cancel'),
Row(userId='45', page='Upgrade'),
Row(userId='10', page='Home'),
Row(userId='89', page='Submit Upgrade'),
Row(userId='97', page='Save Settings'),
Row(userId='81', page='Error'),
Row(userId='100011', page='Home'),
Row(userId='100013', page='Submit Upgrade'),
Row(userId='100003', page='Help'),
Row(userId='100025', page='Submit Downgrade'),
Row(userId='100004', page='Thumbs Down'),
Row(userId='200023', page='Thumbs Down'),
Row(userId='300001', page='Settings'),
Row(userId='300012', page='Thumbs Down'),
Row(userId='95', page='NextSong'),
Row(userId='46', page='Add Friend'),
Row(userId='16', page='Roll Advert'),
Row(userId='74', page='Logout'),
Row(userId='36', page='About'),
Row(userId='21', page='Save Settings'),
Row(userId='153', page='Upgrade'),
Row(userId='100009', page='Thumbs Down'),
Row(userId='100010', page='Add Friend'),
Row(userId='100015', page='Downgrade'),
Row(userId='74', page='NextSong'),
Row(userId='93', page='Help'),
Row(userId='39', page='Upgrade'),
Row(userId='127', page='Roll Advert'),
Row(userId='67', page='NextSong'),
Row(userId='6', page='Roll Advert'),
Row(userId='19', page='Thumbs Up'),
Row(userId='60', page='Upgrade'),
Row(userId='100009', page='Cancellation Confirmation'),
Row(userId='200014', page='Settings'),
Row(userId='300008', page='Roll Advert'),
Row(userId='53', page='NextSong'),
Row(userId='61', page='Help'),
Row(userId='18', page='Home'),
Row(userId='54', page='About'),
Row(userId='119', page='Roll Advert'),
Row(userId='6', page='Error'),
Row(userId='80', page='Upgrade'),
Row(userId='45', page='Downgrade'),
Row(userId='62', page='Logout'),
Row(userId='44', page='Settings'),
Row(userId='200011', page='Logout'),
Row(userId='200021', page='Help'),
Row(userId='300023', page='Add to Playlist'),
Row(userId='300025', page='NextSong'),
Row(userId='300016', page='Save Settings'),
Row(userId='300004', page='Submit Downgrade'),
Row(userId='45', page='Logout'),
Row(userId='27', page='Home'),
Row(userId='27', page='Logout'),
Row(userId='115', page='NextSong'),
```



```
Row(userId='127', page='Add to Playlist'),
Row(userId='44', page='Help'),
Row(userId='117', page='Add to Playlist'),
Row(userId='150', page='Settings'),
Row(userId='8', page='Add to Playlist'),
Row(userId='121', page='Cancellation Confirmation'),
Row(userId='25', page='About'),
Row(userId='137', page='Submit Upgrade'),
Row(userId='200003', page='Help'),
Row(userId='300022', page='Logout'),
Row(userId='74', page='Add to Playlist'),
Row(userId='39', page='NextSong'),
Row(userId='36', page='Upgrade'),
Row(userId='69', page='Add to Playlist'),
Row(userId='7', page='Home'),
Row(userId='42', page='Settings'),
Row(userId='66', page='Thumbs Up'),
Row(userId='126', page='Thumbs Down'),
Row(userId='148', page='Add Friend'),
Row(userId='65', page='Help'),
Row(userId='21', page='Logout'),
Row(userId='62', page='Home'),
Row(userId='41', page='Settings'),
Row(userId='3', page='Add to Playlist'),
Row(userId='86', page='Submit Upgrade'),
Row(userId='100005', page='Home'),
Row(userId='100019', page='Cancellation Confirmation'),
Row(userId='200005', page='Add Friend'),
Row(userId='300025', page='Save Settings'),
Row(userId='101', page='Thumbs Down'),
Row(userId='42', page='Thumbs Up'),
Row(userId='102', page='Roll Advert'),
Row(userId='62', page='NextSong'),
Row(userId='148', page='Upgrade'),
Row(userId='10', page='About'),
Row(userId='145', page='Thumbs Down'),
Row(userId='69', page='About'),
Row(userId='14', page='Add Friend'),
Row(userId='200019', page='Thumbs Down'),
Row(userId='300024', page='NextSong'),
Row(userId='112', page='Thumbs Up'),
Row(userId='92', page='Submit Downgrade'),
Row(userId='38', page='About'),
Row(userId='300011', page='Thumbs Down'),
Row(userId='300025', page='Home'),
Row(userId='300017', page='Settings'),
Row(userId='140', page='Error'),
Row(userId='40', page='Home'),
Row(userId='114', page='Thumbs Up'),
Row(userId='37', page='Home'),
Row(userId='145', page='Add to Playlist'),
Row(userId='8', page='Thumbs Down'),
Row(userId='94', page='Upgrade'),
Row(userId='132', page='Save Settings'),
Row(userId='53', page='Cancel'),
Row(userId='76', page='About'),
Row(userId='100009', page='Submit Downgrade'),
```

```
Row(userId='100016', page='Submit Downgrade'),
Row(userId='300021', page='Settings'),
Row(userId='300021', page='Submit Downgrade'),
Row(userId='126', page='NextSong'),
Row(userId='39', page='Add to Playlist'),
Row(userId='67', page='Roll Advert'),
Row(userId='85', page='Thumbs Down'),
Row(userId='88', page='Thumbs Down'),
Row(userId='82', page='Roll Advert'),
Row(userId='65', page='About'),
Row(userId='101', page='Error'),
Row(userId='133', page='Thumbs Up'),
Row(userId='147', page='Thumbs Up'),
Row(userId='3', page='About'),
Row(userId='154', page='Home'),
Row(userId='10', page='Save Settings'),
Row(userId='153', page='Thumbs Down'),
Row(userId='99', page='Help'),
Row(userId='100016', page='Add to Playlist'),
Row(userId='200002', page='Logout'),
Row(userId='300019', page='Logout'),
Row(userId='300020', page='Thumbs Down'),
Row(userId='300022', page='Save Settings'),
Row(userId='87', page='Add to Playlist'),
Row(userId='20', page='Home'),
Row(userId='9', page='Submit Upgrade'),
Row(userId='148', page='Settings'),
Row(userId='134', page='Add to Playlist'),
Row(userId='99', page='About'),
Row(userId='200009', page='Submit Downgrade'),
Row(userId='59', page='Roll Advert'),
Row(userId='124', page='Thumbs Up'),
Row(userId='6', page='Logout'),
Row(userId='138', page='NextSong'),
Row(userId='74', page='Add Friend'),
Row(userId='20', page='Add Friend'),
Row(userId='47', page='Thumbs Down'),
Row(userId='71', page='Upgrade'),
Row(userId='43', page='Upgrade'),
Row(userId='54', page='Cancel'),
Row(userId='100017', page='NextSong'),
Row(userId='300001', page='Thumbs Down'),
Row(userId='300009', page='About'),
Row(userId='15', page='NextSong'),
Row(userId='80', page='Thumbs Up'),
Row(userId='55', page='NextSong'),
Row(userId='32', page='Add to Playlist'),
Row(userId='99', page='Roll Advert'),
Row(userId='19', page='Settings'),
Row(userId='147', page='Logout'),
Row(userId='17', page='About'),
Row(userId='81', page='Submit Downgrade'),
Row(userId='13', page='Upgrade'),
Row(userId='73', page='Cancellation Confirmation'),
Row(userId='112', page='About'),
Row(userId='110', page='Thumbs Down'),
Row(userId='100001', page='Home'),
```

```
Row(userId='100001', page='Error'),
Row(userId='100007', page='Settings'),
Row(userId='100016', page='Error'),
Row(userId='200002', page='Roll Advert'),
Row(userId='200020', page='NextSong'),
Row(userId='200022', page='Logout'),
Row(userId='300017', page='Logout'),
Row(userId='70', page='Add to Playlist'),
Row(userId='82', page='Home'),
Row(userId='100', page='Thumbs Down'),
Row(userId='92', page='Add Friend'),
Row(userId='106', page='Settings'),
Row(userId='145', page='Logout'),
Row(userId='30', page='Help'),
Row(userId='82', page='About'),
Row(userId='49', page='Settings'),
Row(userId='67', page='About'),
Row(userId='13', page='Help'),
Row(userId='154', page='Add to Playlist'),
Row(userId='55', page='Submit Upgrade'),
Row(userId='100022', page='Cancel'),
Row(userId='200022', page='Roll Advert'),
Row(userId='300015', page='Submit Upgrade'),
Row(userId='78', page='Home'),
Row(userId='44', page='Add Friend'),
Row(userId='104', page='About'),
Row(userId='73', page='About'),
Row(userId='', page='Error'),
Row(userId='96', page='Help'),
Row(userId='72', page='Upgrade'),
Row(userId='86', page='Save Settings'),
Row(userId='100025', page='Help'),
Row(userId='200011', page='Roll Advert'),
Row(userId='200010', page='Add Friend'),
Row(userId='200006', page='Save Settings'),
Row(userId='300006', page='Roll Advert'),
Row(userId='51', page='Thumbs Up'),
Row(userId='46', page='NextSong'),
Row(userId='35', page='Help'),
Row(userId='118', page='Add Friend'),
Row(userId='69', page='Roll Advert'),
Row(userId='50', page='Logout'),
Row(userId='42', page='Logout'),
Row(userId='89', page='Add to Playlist'),
Row(userId='66', page='Add Friend'),
Row(userId='123', page='Logout'),
Row(userId='26', page='Logout'),
Row(userId='28', page='Submit Upgrade'),
Row(userId='155', page='NextSong'),
Row(userId='100018', page='Logout'),
Row(userId='100003', page='Home'),
Row(userId='300009', page='Upgrade'),
Row(userId='27', page='Help'),
Row(userId='113', page='Thumbs Up'),
Row(userId='136', page='NextSong'),
Row(userId='80', page='About'),
Row(userId='85', page='Home'),
```

```
Row(userId='85', page='Roll Advert'),
Row(userId='118', page='Home'),
Row(userId='112', page='Settings'),
Row(userId='44', page='Add to Playlist'),
Row(userId='85', page='Downgrade'),
Row(userId='88', page='Downgrade'),
Row(userId='56', page='Thumbs Down'),
Row(userId='13', page='Downgrade'),
Row(userId='100003', page='Thumbs Up'),
Row(userId='100014', page='Logout'),
Row(userId='200001', page='Roll Advert'),
Row(userId='300016', page='Error'),
Row(userId='300025', page='Add Friend'),
Row(userId='300014', page='Save Settings'),
Row(userId='92', page='Thumbs Down'),
Row(userId='129', page='Downgrade'),
Row(userId='9', page='Downgrade'),
Row(userId='77', page='Thumbs Down'),
Row(userId='43', page='Settings'),
Row(userId='139', page='Help'),
Row(userId='50', page='Submit Upgrade'),
Row(userId='200003', page='Submit Upgrade'),
Row(userId='300002', page='Upgrade'),
Row(userId='113', page='Thumbs Down'),
Row(userId='58', page='Thumbs Up'),
Row(userId='92', page='Submit Upgrade'),
Row(userId='111', page='Add Friend'),
Row(userId='56', page='Help'),
Row(userId='132', page='Downgrade'),
Row(userId='3', page='Downgrade'),
Row(userId='63', page='Thumbs Down'),
Row(userId='153', page='Settings'),
Row(userId='84', page='Add to Playlist'),
Row(userId='86', page='Downgrade'),
Row(userId='7', page='Save Settings'),
Row(userId='100025', page='Home'),
Row(userId='100008', page='Add to Playlist'),
Row(userId='200010', page='Error'),
Row(userId='300009', page='Home'),
Row(userId='300023', page='Submit Upgrade'),
Row(userId='95', page='Downgrade'),
Row(userId='54', page='Add to Playlist'),
Row(userId='53', page='Logout'),
Row(userId='53', page='Error'),
Row(userId='58', page='Add Friend'),
Row(userId='121', page='Add to Playlist'),
Row(userId='73', page='Logout'),
Row(userId='86', page='Settings'),
Row(userId='15', page='Thumbs Up'),
Row(userId='5', page='Roll Advert'),
Row(userId='114', page='Submit Upgrade'),
Row(userId='100005', page='Roll Advert'),
Row(userId='100001', page='Cancel'),
Row(userId='100024', page='Cancel'),
Row(userId='97', page='Roll Advert'),
Row(userId='32', page='Cancellation Confirmation'),
Row(userId='43', page='Add Friend'),
```

```
Row(userId='65', page='Thumbs Down'),
Row(userId='122', page='Submit Upgrade'),
Row(userId='21', page='Downgrade'),
Row(userId='100016', page='NextSong'),
Row(userId='100012', page='Help'),
Row(userId='100010', page='Logout'),
Row(userId='200014', page='Home'),
Row(userId='200018', page='Home'),
Row(userId='39', page='About'),
Row(userId='61', page='Add to Playlist'),
Row(userId='87', page='Logout'),
Row(userId='106', page='Logout'),
Row(userId='35', page='Save Settings'),
Row(userId='76', page='Home'),
Row(userId='93', page='Thumbs Down'),
Row(userId='77', page='Settings'),
Row(userId='66', page='Help'),
Row(userId='77', page='Roll Advert'),
Row(userId='37', page='Save Settings'),
Row(userId='134', page='Thumbs Down'),
Row(userId='147', page='Submit Upgrade'),
Row(userId='100', page='Submit Upgrade'),
Row(userId='200016', page='Add to Playlist'),
Row(userId='200018', page='Cancellation Confirmation'),
Row(userId='300006', page='Upgrade'),
Row(userId='29', page='Add to Playlist'),
Row(userId='39', page='Settings'),
Row(userId='39', page='Error'),
Row(userId='17', page='Add Friend'),
Row(userId='82', page='Thumbs Down'),
Row(userId='107', page='Logout'),
Row(userId='19', page='Downgrade'),
Row(userId='60', page='Logout'),
Row(userId='64', page='Roll Advert'),
Row(userId='98', page='Save Settings'),
Row(userId='100009', page='NextSong'),
Row(userId='100008', page='Help'),
Row(userId='100004', page='Submit Upgrade'),
Row(userId='200015', page='Add to Playlist'),
Row(userId='200017', page='Thumbs Up'),
Row(userId='300005', page='Upgrade'),
Row(userId='300005', page='Submit Upgrade'),
Row(userId='300021', page='Help'),
Row(userId='39', page='Add Friend'),
Row(userId='15', page='Help'),
Row(userId='44', page='About'),
Row(userId='122', page='Upgrade'),
Row(userId='', page='Register'),
Row(userId='25', page='Submit Downgrade'),
Row(userId='117', page='Save Settings'),
Row(userId='300003', page='NextSong'),
Row(userId='300021', page='Upgrade'),
Row(userId='140', page='Thumbs Up'),
Row(userId='127', page='Home'),
Row(userId='8', page='NextSong'),
Row(userId='27', page='Save Settings'),
Row(userId='128', page='Upgrade'),
```

```
Row(userId='103', page='Thumbs Down'),
Row(userId='129', page='Cancellation Confirmation'),
Row(userId='63', page='Upgrade'),
Row(userId='110', page='Add Friend'),
Row(userId='100007', page='Help'),
Row(userId='100021', page='About'),
Row(userId='100015', page='Cancellation Confirmation'),
Row(userId='200004', page='Thumbs Down'),
Row(userId='200003', page='Thumbs Down'),
Row(userId='200016', page='NextSong'),
Row(userId='200001', page='Cancellation Confirmation'),
Row(userId='300007', page='Home'),
Row(userId='27', page='NextSong'),
Row(userId='132', page='Upgrade'),
Row(userId='129', page='Home'),
Row(userId='96', page='Thumbs Up'),
Row(userId='67', page='Thumbs Up'),
Row(userId='143', page='Upgrade'),
Row(userId='42', page='Add Friend'),
Row(userId='19', page='Home'),
Row(userId='87', page='Help'),
Row(userId='97', page='Thumbs Up'),
Row(userId='128', page='Submit Upgrade'),
Row(userId='126', page='Submit Upgrade'),
Row(userId='138', page='Submit Upgrade'),
Row(userId='54', page='Submit Upgrade'),
Row(userId='139', page='About'),
Row(userId='100017', page='Cancellation Confirmation'),
Row(userId='100013', page='Cancellation Confirmation'),
Row(userId='200019', page='Settings'),
Row(userId='200016', page='Cancel'),
Row(userId='300002', page='Logout'),
Row(userId='300007', page='Thumbs Up'),
Row(userId='300020', page='Thumbs Up'),
Row(userId='9', page='Upgrade'),
Row(userId='127', page='Downgrade'),
Row(userId='105', page='Settings'),
Row(userId='7', page='NextSong'),
Row(userId='114', page='Home'),
Row(userId='129', page='Save Settings'),
Row(userId='2', page='Thumbs Down'),
Row(userId='109', page='Add to Playlist'),
Row(userId='86', page='Help'),
Row(userId='49', page='Roll Advert'),
Row(userId='200002', page='Add to Playlist'),
Row(userId='200009', page='NextSong'),
Row(userId='200006', page='About'),
Row(userId='200025', page='Submit Downgrade'),
Row(userId='300008', page='Help'),
Row(userId='300011', page='Upgrade'),
Row(userId='300007', page='Add Friend'),
Row(userId='300004', page='Downgrade'),
Row(userId='29', page='Add Friend'),
Row(userId='105', page='Home'),
Row(userId='96', page='Add Friend'),
Row(userId='49', page='Downgrade'),
Row(userId='115', page='Downgrade'),
```

```
Row(userId='91', page='Home'),
Row(userId='89', page='Logout'),
Row(userId='108', page='Add to Playlist'),
Row(userId='119', page='Thumbs Down'),
Row(userId='155', page='Add Friend'),
Row(userId='55', page='About'),
Row(userId='100005', page='Add Friend'),
Row(userId='100013', page='Thumbs Down'),
Row(userId='100006', page='Settings'),
Row(userId='300020', page='NextSong'),
Row(userId='29', page='Roll Advert'),
Row(userId='38', page='Submit Downgrade'),
Row(userId='16', page='Settings'),
Row(userId='106', page='Upgrade'),
Row(userId='150', page='Help'),
Row(userId='147', page='Thumbs Down'),
Row(userId='111', page='Help'),
Row(userId='155', page='Downgrade'),
Row(userId='56', page='Error'),
Row(userId='137', page='Thumbs Down'),
Row(userId='100022', page='About'),
Row(userId='200010', page='Home'),
Row(userId='200004', page='Add Friend'),
Row(userId='300013', page='Add to Playlist'),
Row(userId='300018', page='Home'),
Row(userId='300022', page='Thumbs Down'),
Row(userId='300004', page='Save Settings'),
Row(userId='300010', page='Thumbs Down'),
Row(userId='15', page='Add Friend'),
Row(userId='92', page='About'),
Row(userId='99', page='Thumbs Down'),
Row(userId='10', page='Thumbs Up'),
Row(userId='76', page='Roll Advert'),
Row(userId='60', page='Submit Upgrade'),
Row(userId='113', page='About'),
Row(userId='76', page='Add to Playlist'),
Row(userId='114', page='Downgrade'),
Row(userId='100015', page='Save Settings'),
Row(userId='200011', page='Help'),
Row(userId='200009', page='Thumbs Up'),
Row(userId='300018', page='Downgrade'),
Row(userId='300011', page='Submit Downgrade'),
Row(userId='141', page='Home'),
Row(userId='131', page='Settings'),
Row(userId='73', page='Submit Upgrade'),
Row(userId='70', page='Logout'),
Row(userId='113', page='Help'),
Row(userId='106', page='Downgrade'),
Row(userId='112', page='Add Friend'),
Row(userId='139', page='Submit Upgrade'),
Row(userId='155', page='About'),
Row(userId='136', page='Error'),
Row(userId='55', page='Help'),
Row(userId='100013', page='Save Settings'),
Row(userId='200008', page='Thumbs Up'),
Row(userId='300017', page='Save Settings'),
Row(userId='300024', page='Add to Playlist'),
```

```

Row(userId='88', page='Roll Advert'),
Row(userId='58', page='Roll Advert'),
Row(userId='41', page='Add to Playlist'),
Row(userId='12', page='Help'),
Row(userId='112', page='Roll Advert'),
Row(userId='54', page='Settings'),
Row(userId='103', page='Home'),
Row(userId='37', page='Upgrade'),
Row(userId='43', page='Thumbs Up'),
Row(userId='97', page='Upgrade'),
Row(userId='94', page='Save Settings'),
Row(userId='151', page='Logout'),
Row(userId='100011', page='NextSong'),
Row(userId='100013', page='NextSong'),
Row(userId='100014', page='Downgrade'),
Row(userId='200022', page='Home'),
Row(userId='200021', page='Add Friend'),
Row(userId='200004', page='Downgrade'),
Row(userId='200003', page='Roll Advert'),
Row(userId='200021', page='Cancel'),
Row(userId='27', page='Add to Playlist'),
Row(userId='65', page='Home'),
Row(userId='77', page='Home'),
Row(userId='70', page='Submit Upgrade'),
Row(userId='66', page='Logout'),
Row(userId='32', page='Downgrade'),
Row(userId='109', page='Add Friend'),
Row(userId='12', page='Submit Downgrade'),
Row(userId='20', page='Error'),
Row(userId='127', page='About'),
Row(userId='111', page='Thumbs Up'),
...]
```

In []:

Writing in Pandas

In [33]: `churn_pd=data_churn.toPandas()`

There are no NaN's


```
In [34]: churn_pd.isna().sum()
```

```
Out[34]: artist          0
auth          0
firstName     0
gender        0
itemInSession 0
lastName      0
length        0
level         0
location      0
method        0
page          0
registration  0
sessionId     0
song          0
status        0
ts            0
userAgent     0
userId        0
churn         0
hour          0
dtype: int64
```

```
In [35]: churn_pd.isnull().sum()
```

```
Out[35]: artist          0
auth          0
firstName     0
gender        0
itemInSession 0
lastName      0
length        0
level         0
location      0
method        0
page          0
registration  0
sessionId     0
song          0
status        0
ts            0
userAgent     0
userId        0
churn         0
hour          0
dtype: int64
```

There are no duplicates

```
In [36]: churn_pd.duplicated().sum()
```

```
Out[36]: 0
```

```
In [37]: churn_pd.shape
```

```
Out[37]: (286500, 20)
```

```
In [38]: df=churn_pd.copy()
```

```
In [39]: #df.dropna(inplace=True)
```

```
In [40]: df.isna().sum()
```

```
Out[40]: artist          0  
auth          0  
firstName     0  
gender        0  
itemInSession 0  
lastName      0  
length        0  
level         0  
location      0  
method        0  
page          0  
registration   0  
sessionId     0  
song          0  
status        0  
ts            0  
userAgent     0  
userId        0  
churn         0  
hour          0  
dtype: int64
```

```
In [41]: df.duplicated().sum()
```

```
Out[41]: 0
```

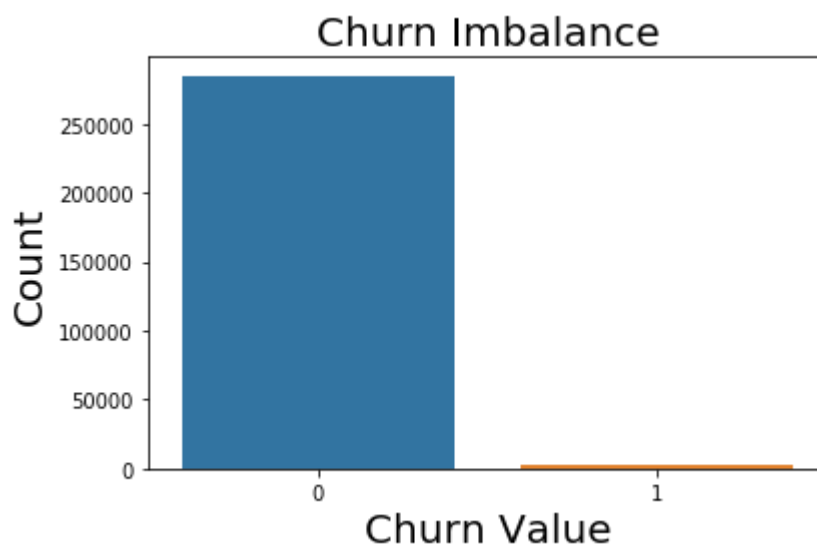
```
In [ ]:
```

Churn Imbalance

```
In [42]: df['churn'].value_counts()
```

```
Out[42]: 0    284278  
        1      2222  
        Name: churn, dtype: int64
```

```
In [43]: import seaborn as sns
from seaborn import barplot
#plt.hist(df['churn'].value_counts())
sns.countplot(x="churn",data=df)
plt.title('Churn Imbalance',size=20)
plt.xlabel('Churn Value',size=20)
plt.ylabel('Count',size=20)
plt.tight_layout()
plt.savefig('churn_imbalance.jpeg')
plt.show()
```



In []:

In []:

Creating the churn dataframe

```
In [44]: df_churn=df[df['churn']==1]['hour'].hist(bins=20);  
df_churn
```

Out[44]:

	artist	auth	firstName	gender	itemInSession	lastName	length	level
35	missing_val	Logged In	Alexi	F	4	Warren	249.117182	paid
357	missing_val	Logged In	Faigy	F	29	Howe	249.117182	paid
879	missing_val	Logged In	Faigy	F	115	Howe	249.117182	paid
929	missing_val	Logged In	Kael	M	46	Baker	249.117182	paid
930	missing_val	Logged In	Kael	M	47	Baker	249.117182	paid
932	missing_val	Logged In	Faigy	F	125	Howe	249.117182	paid
984	missing_val	Logged In	Grant	M	119	Flores	249.117182	paid
993	missing_val	Logged In	Grant	M	121	Flores	249.117182	paid
1079	missing_val	Logged In	Elizabeth	F	17	Washington	249.117182	paid
1149	missing_val	Logged In	Panav	M	41	Myers	249.117182	paid
1174	missing_val	Logged In	Grant	M	139	Flores	249.117182	paid
1461	missing_val	Logged In	Ethan	M	36	Johnson	249.117182	paid
1613	missing_val	Logged In	Ethan	M	55	Johnson	249.117182	paid

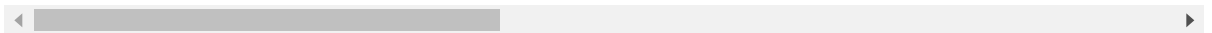
	artist	auth	firstName	gender	itemInSession	lastName	length	level
1649	missing_val	Logged In	Panav	M	88	Myers	249.117182	paid
1651	missing_val	Logged In	Panav	M	89	Myers	249.117182	paid
1701	missing_val	Logged In	Alexi	F	58	Warren	249.117182	paid
1885	missing_val	Logged In	Nicole	F	28	Beck	249.117182	paid
2346	missing_val	Logged In	Alexi	F	113	Warren	249.117182	paid
2537	missing_val	Logged In	Nicole	F	74	Beck	249.117182	paid
3011	missing_val	Logged In	Kael	M	229	Baker	249.117182	paid
3220	missing_val	Logged In	Ethan	M	209	Johnson	249.117182	paid
3221	missing_val	Logged In	Ethan	M	210	Johnson	249.117182	paid
3341	missing_val	Logged In	Ethan	M	246	Johnson	249.117182	paid
3929	missing_val	Logged In	Autumn	F	44	Martin	249.117182	paid
3946	missing_val	Logged In	Ethan	M	357	Johnson	249.117182	paid
4045	missing_val	Logged In	Aurora	F	6	Humphrey	249.117182	paid
4146	missing_val	Logged In	Aurora	F	21	Humphrey	249.117182	paid

	artist	auth	firstName	gender	itemInSession	lastName	length	level
4905	missing_val	Logged In	Ethan	M	520	Johnson	249.117182	paid
5071	missing_val	Logged In	Ethan	M	544	Johnson	249.117182	paid
5507	missing_val	Logged In	Calvin	M	16	Marshall	249.117182	paid
...
280929	missing_val	Logged In	Emilia	F	31	House	249.117182	paid
280978	missing_val	Logged In	Chloe	F	36	Deleon	249.117182	paid
281026	missing_val	Logged In	Emilia	F	11	House	249.117182	paid
281262	missing_val	Logged In	Emilia	F	41	House	249.117182	paid
281479	missing_val	Logged In	Antonina	F	310	Kennedy	249.117182	paid
281480	missing_val	Cancelled	Antonina	F	311	Kennedy	249.117182	paid
281544	missing_val	Logged In	Chloe	F	63	Deleon	249.117182	paid
281678	missing_val	Logged In	Noah	M	59	Whitehead	249.117182	paid
281766	missing_val	Logged In	Noah	M	147	Whitehead	249.117182	paid
281974	missing_val	Logged In	Chloe	F	81	Deleon	249.117182	paid
282008	missing_val	Logged In	Jade	F	8	Stewart	249.117182	paid

	artist	auth	firstName	gender	itemInSession	lastName	length	level
282342	missing_val	Logged In	Bryson	M	5	Watson	249.117182	paid
282364	missing_val	Logged In	Joseph	M	91	Green	249.117182	paid
282409	missing_val	Logged In	Emma	F	4	Allen	249.117182	paid
282436	missing_val	Logged In	Chloe	F	235	Deleon	249.117182	paid
283228	missing_val	Logged In	Carter	M	206	Tanner	249.117182	paid
283483	missing_val	Logged In	Carter	M	274	Tanner	249.117182	paid
283798	missing_val	Logged In	Lindyn	F	504	Patel	249.117182	paid
284465	missing_val	Logged In	Chloe	F	120	Deleon	249.117182	paid
284467	missing_val	Logged In	Chloe	F	121	Deleon	249.117182	paid
284468	missing_val	Logged In	Chloe	F	122	Deleon	249.117182	paid
284484	missing_val	Logged In	Alyssia	F	96	Campbell	249.117182	paid
284692	missing_val	Logged In	Lauren	F	7	Boone	249.117182	paid
285180	missing_val	Logged In	Emilia	F	93	House	249.117182	paid
285197	missing_val	Logged In	Emilia	F	101	House	249.117182	paid

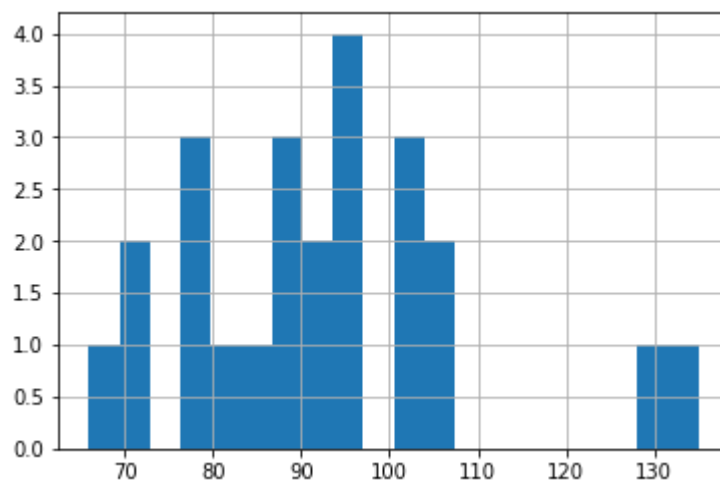
	artist	auth	firstName	gender	itemInSession	lastName	length	level
285266	missing_val	Logged In	Emilia	F	138	House	249.117182	paid
285312	missing_val	Logged In	Emilia	F	161	House	249.117182	paid
285552	missing_val	Logged In	Chloe	F	84	Deleon	249.117182	paid
285652	missing_val	Logged In	Joshua	M	44	White	249.117182	paid
286059	missing_val	Logged In	Joshua	M	81	Warner	249.117182	paid

2222 rows × 20 columns



Histogram of Churns by hour

```
In [45]: df_churn['hour'].value_counts().hist(bins=20)
plt.show()
```

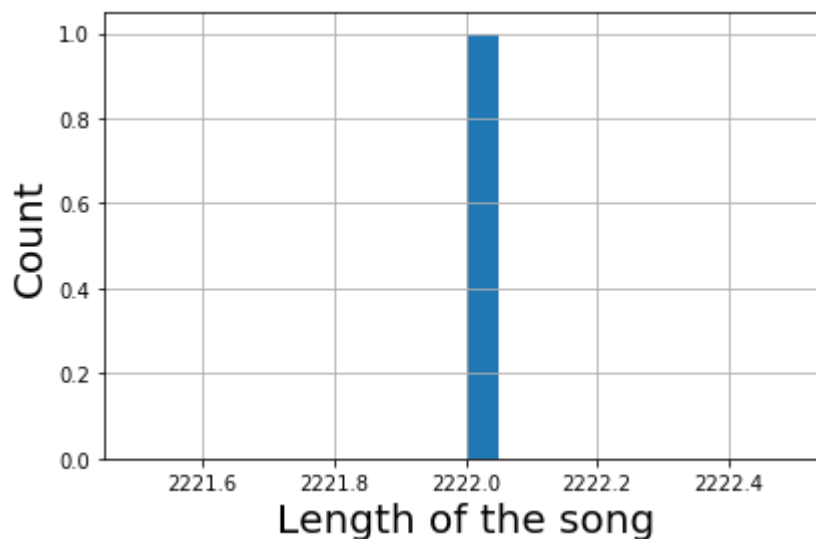


```
In [ ]:
```

Churn Length

The below histogram shows the length of the song for customers who churned. Interestingly, they all had the same length!! This is related to the fact that the artists and songs of customers that churned were all NaN's in the original given mini dataset.

```
In [46]: df_churn['length'].value_counts().hist(bins=20)
plt.ylabel('Count',size=20)
plt.xlabel('Length of the song',size=20)
plt.tight_layout()
#plt.savefig('./churn_length.jpeg')
plt.show()
```



```
In [47]: df_churn['length'].value_counts()
```

```
Out[47]: 249.117182    2222
         Name: length, dtype: int64
```

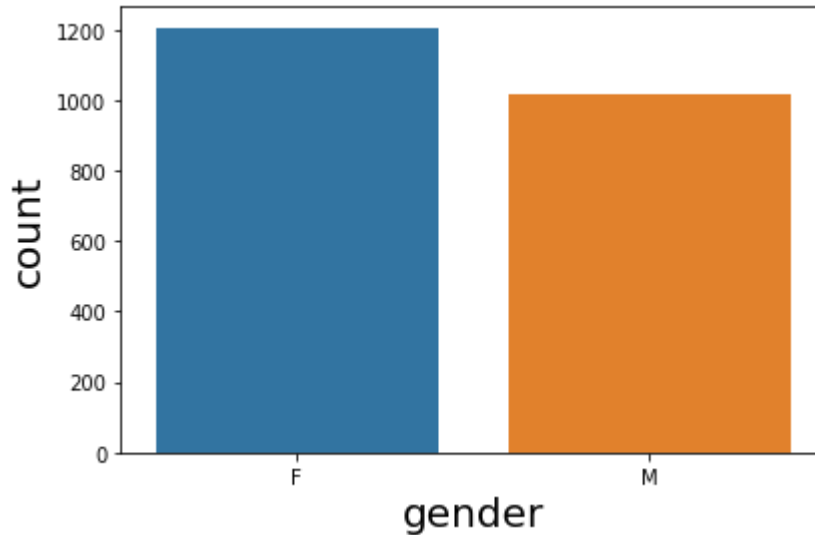
```
In [ ]:
```

Churn by gender

```
In [48]: df_churn['gender'].value_counts()
```

```
Out[48]: F    1207
         M    1015
         Name: gender, dtype: int64
```

```
In [49]: countplot(x='gender',data=df_churn)
plt.xlabel('gender',size=20)
plt.ylabel('count',size=20)
plt.tight_layout()
#plt.savefig('./churn_gender.jpeg')
plt.show()
```



As it can be seen, more females churned than males. It seems appropriate for Sparkify to announce a discount for females in order to diminish this pronounced "female churn".

In []:

Churn by location

In []:

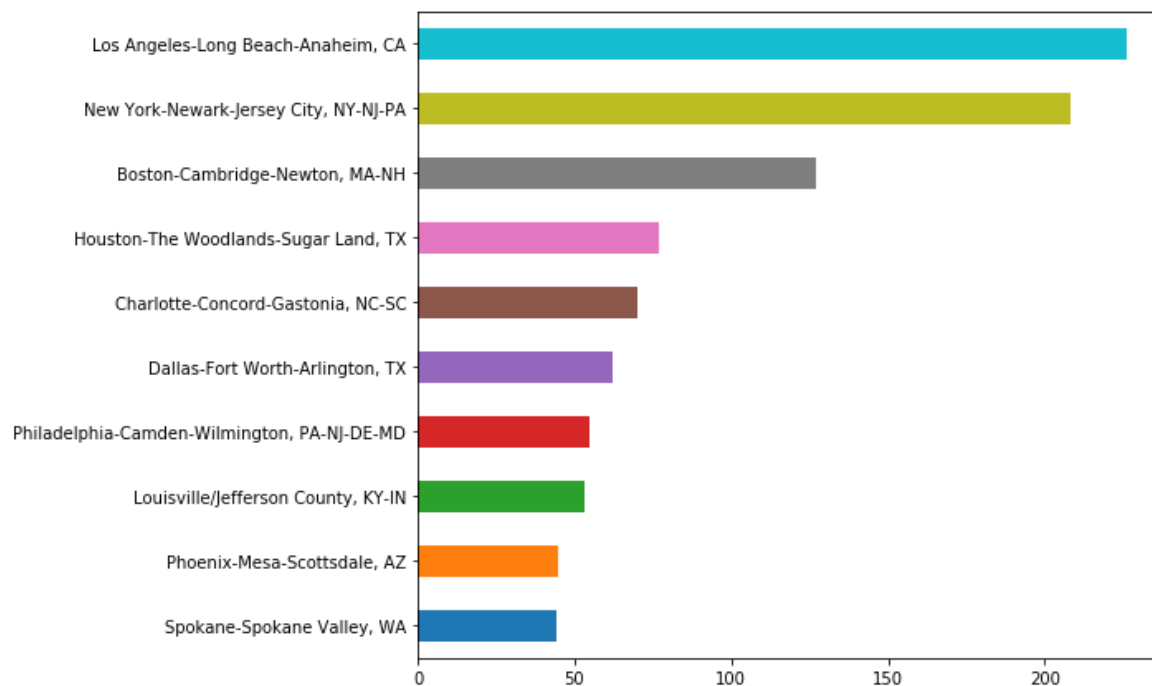
```
In [50]: df_churn['location'].value_counts()[:10]

#plt.tight_layout()
#plt.show()
```

```
Out[50]: Los Angeles-Long Beach-Anaheim, CA      226
New York-Newark-Jersey City, NY-NJ-PA      208
Boston-Cambridge-Newton, MA-NH      127
Houston-The Woodlands-Sugar Land, TX      77
Charlotte-Concord-Gastonia, NC-SC      70
Dallas-Fort Worth-Arlington, TX      62
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD      55
Louisville/Jefferson County, KY-IN      53
Phoenix-Mesa-Scottsdale, AZ      45
Spokane-Spokane Valley, WA      44
Name: location, dtype: int64
```

```
In [ ]:
```

```
In [51]: plt.figure(figsize=[10,6])
(df_churn['location'].value_counts()[:10].sort_values()).plot(kind='barh');
plt.tight_layout()
#plt.savefig('./churn_location.jpeg')
plt.show()
```



Offer a discount in the top ten place with more churn!

From this figure, it can be seen that Los Angeles, New York, and Boston have the highest churn numbers. Therefore, it is appropriate to offer a discount for Sparkify's plans in order to minimize customer churns in these places (the best it would be to apply this discount in the top 10 places shown in the figure).

In []:

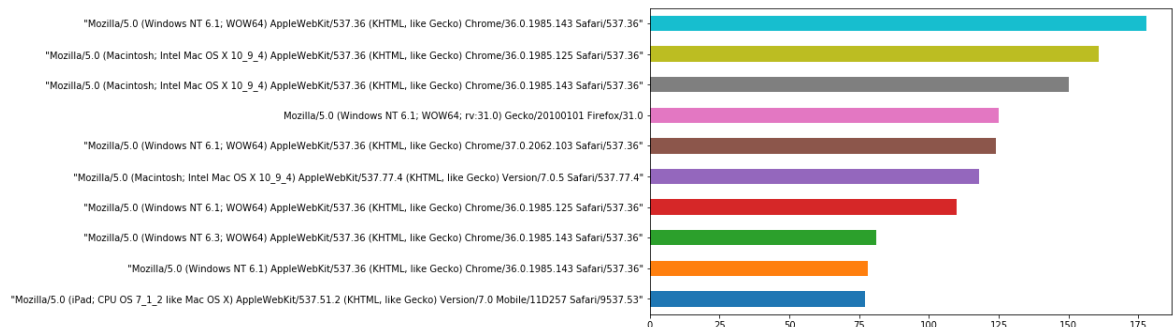
Churn by userAgent

In []:

```
In [52]: df_churn['userAgent'].nunique()
```

Out[52]: 51

```
In [53]: plt.figure(figsize=[10,6])
(df_churn['userAgent']).value_counts()[ :10].sort_values().plot(kind=
'barh');
#plt.tight_layout()
#plt.savefig('./churn_useragent.jpeg')
plt.show()
```



As previously discussed, we can work on the top 10 userAgent and try to minimize the customer churns.

In []:

```
In [54]: df_churn['artist'].unique()
```

Out[54]: array(['missing_val'], dtype=object)

```
In [55]: df_churn['song'].unique()
```

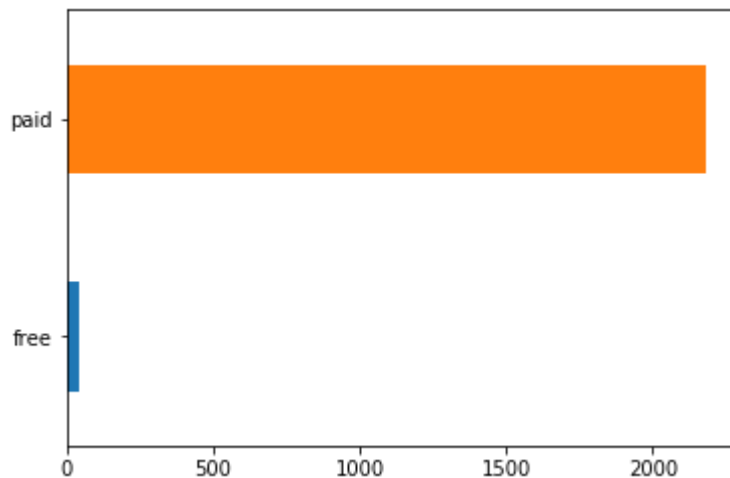
```
Out[55]: array(['missing_val'], dtype=object)
```

***** as in the churn the 'song' and 'artist' columns are just the missing values, we will drop these columns**

Churn by level

```
In [56]: df_churn['level'].value_counts().sort_values().plot(kind='barh')
```

```
Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea5e3eb38>
```



```
In [57]: df_churn['method'].value_counts()
```

```
Out[57]: GET      2107  
        PUT       115  
        Name: method, dtype: int64
```

```
In [ ]:
```

```
In [ ]:
```

The same analysis, but for customers who did not churn

```
In [ ]:
```

```
In [58]: df_nochurn=df[df['churn']==0]['hour'].hist(bins=20);  
df_nochurn
```

Out[58]:

	artist	auth	firstName	gender	itemInSession	lastName	length
0	Martha Tilston	Logged In	Colin	M	50	Freeman	277.89016
1	Five Iron Frenzy	Logged In	Micah	M	79	Long	236.09424
2	Adam Lambert	Logged In	Colin	M	51	Freeman	282.82730
3	Enigma	Logged In	Micah	M	80	Long	262.71302
4	Daft Punk	Logged In	Colin	M	52	Freeman	223.60771
5	The All-American Rejects	Logged In	Micah	M	81	Long	208.29995
6	The Velvet Underground / Nico	Logged In	Micah	M	82	Long	260.46649
7	Starflyer 59	Logged In	Colin	M	53	Freeman	185.44281
8	missing_val	Logged In	Colin	M	54	Freeman	249.11718
9	Frumpies	Logged In	Colin	M	55	Freeman	134.47791
10	Britt Nicole	Logged In	Micah	M	83	Long	229.87710
11	missing_val	Logged In	Micah	M	84	Long	249.11718
12	Edward Sharpe & The Magnetic Zeros	Logged In	Colin	M	56	Freeman	223.58159

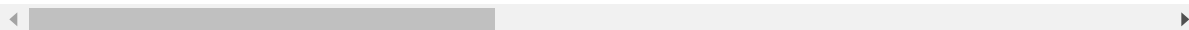
	artist	auth	firstName	gender	itemInSession	lastName	length
13	Tesla	Logged In	Micah	M	85	Long	201.06404
14	missing_val	Logged In	Micah	M	86	Long	249.11718
15	Stan Mosley	Logged In	Colin	M	57	Freeman	246.69995
16	Florence + The Machine	Logged In	Micah	M	87	Long	168.64608
17	Tokyo Police Club	Logged In	Ashlynn	F	0	Williams	166.11220
18	Orishas	Logged In	Colin	M	58	Freeman	222.22322
19	Ratatat	Logged In	Micah	M	88	Long	229.77261
20	Manolo Garcia	Logged In	Ashlynn	F	1	Williams	283.74159
21	Downhere	Logged In	Colin	M	59	Freeman	223.92118
22	Modjo	Logged In	Alexi	F	0	Warren	250.93179
23	Miley Cyrus	Logged In	Micah	M	89	Long	231.26159
24	David Bowie	Logged In	Ashlynn	F	2	Williams	174.41914
25	Skillet	Logged In	Colin	M	60	Freeman	233.32526

	artist	auth	firstName	gender	itemInSession	lastName	length
26	Edwyn Collins	Logged In	Alexi	F	1	Warren	216.84200
27	Telepopmusik	Logged In	Micah	M	90	Long	241.60608
28	Kings Of Leon	Logged In	Ashlynn	F	3	Williams	307.46077
29	Florence + The Machine	Logged In	Colin	M	61	Freeman	219.66322
...
286470	Shaggy	Logged In	Emilia	F	16	House	201.37751
286471	Natiruts	Logged In	Emilia	F	17	House	210.88608
286472	In Flames	Logged In	Emilia	F	18	House	217.65179
286473	Eminem	Logged In	Emilia	F	19	House	250.82730
286474	Blink-182	Logged In	Emilia	F	20	House	173.76608
286475	missing_val	Logged In	Emilia	F	21	House	249.11718
286476	Jason Mraz	Logged In	Emilia	F	22	House	218.40934

	artist	auth	firstName	gender	itemInSession	lastName	length
286477	Jessica Lea Mayfield	Logged In	Emilia	F	23	House	180.74077
286478	Jimi Hendrix	Logged In	Emilia	F	24	House	380.68200
286479	missing_val	Logged In	Emilia	F	25	House	249.11718
286480	Jamie Foxx featuring Twista	Logged In	Emilia	F	26	House	258.79465
286481	Aaliyah	Logged In	Emilia	F	27	House	228.28363
286482	missing_val	Logged In	Emilia	F	28	House	249.11718
286483	Hermano	Logged In	Emilia	F	29	House	115.90485
286484	Saliva	Logged In	Emilia	F	30	House	222.56281
286485	Lupe Fiasco	Logged In	Emilia	F	31	House	273.94567
286486	Harmonia	Logged In	Emilia	F	32	House	655.77751
286487	The Rolling Stones	Logged In	Emilia	F	33	House	225.30567

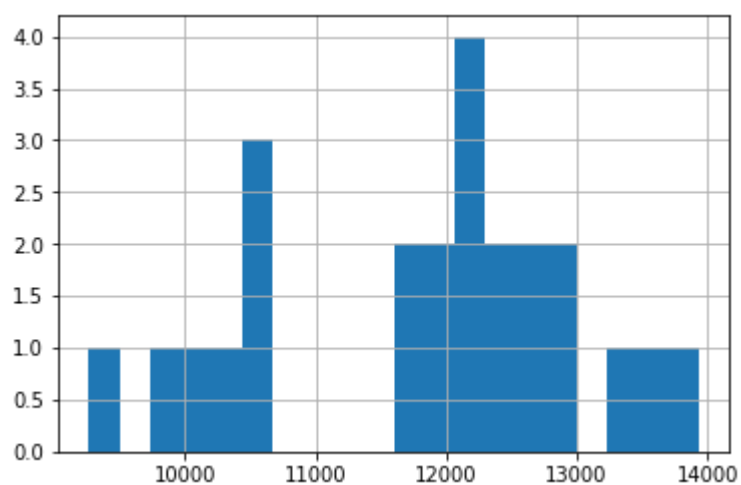
	artist	auth	firstName	gender	itemInSession	lastName	length
286488	Alejandro Sanz	Logged In	Emilia	F	34	House	241.52771
286489	Valley of the Giants	Logged In	Emilia	F	35	House	420.46649
286490	missing_val	Logged In	Emilia	F	36	House	249.11718
286491	Olive	Logged In	Emilia	F	37	House	264.12363
286492	Iron Maiden	Logged In	Emilia	F	38	House	258.66404
286493	missing_val	Logged In	Emilia	F	39	House	249.11718
286494	missing_val	Logged Out	missing_val2	missing_val2	40	missing_val2	249.11718
286495	missing_val	Logged Out	missing_val2	missing_val2	41	missing_val2	249.11718
286496	missing_val	Logged Out	missing_val2	missing_val2	42	missing_val2	249.11718
286497	missing_val	Logged In	Emilia	F	43	House	249.11718
286498	missing_val	Logged In	Emilia	F	44	House	249.11718
286499	Camera Obscura	Logged In	Emilia	F	45	House	170.89261

284278 rows × 20 columns



```
In [59]: df_nochurn['hour'].value_counts().hist(bins=20)
```

```
Out[59]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea3a59668>
```



```
In [60]: df_nochurn['length']
```

```
Out[60]: 0      277.890160
          1      236.094240
          2      282.827300
          3      262.713020
          4      223.607710
          5      208.299950
          6      260.466490
          7      185.442810
          8      249.117182
          9      134.477910
         10      229.877100
         11      249.117182
         12      223.581590
         13      201.064040
         14      249.117182
         15      246.699950
         16      168.646080
         17      166.112200
         18      222.223220
         19      229.772610
         20      283.741590
         21      223.921180
         22      250.931790
         23      231.261590
         24      174.419140
         25      233.325260
         26      216.842000
         27      241.606080
         28      307.460770
         29      219.663220
          ...
        286470    201.377510
        286471    210.886080
        286472    217.651790
        286473    250.827300
        286474    173.766080
        286475    249.117182
        286476    218.409340
        286477    180.740770
        286478    380.682000
        286479    249.117182
        286480    258.794650
        286481    228.283630
        286482    249.117182
        286483    115.904850
        286484    222.562810
        286485    273.945670
        286486    655.777510
        286487    225.305670
        286488    241.527710
        286489    420.466490
        286490    249.117182
        286491    264.123630
        286492    258.664040
        286493    249.117182
        286494    249.117182
        286495    249.117182
```

```

286496    249.117182
286497    249.117182
286498    249.117182
286499    170.892610
Name: length, Length: 284278, dtype: float64

```

```

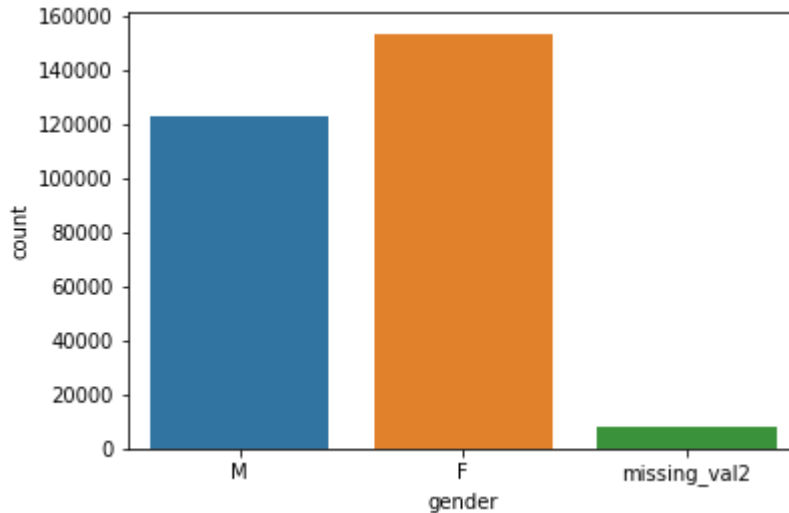
In [61]: df_nochurn['gender'].value_counts()
countplot(x='gender', data=df_nochurn)

```

```

Out[61]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea37ae438>

```



```

In [ ]:

```

```

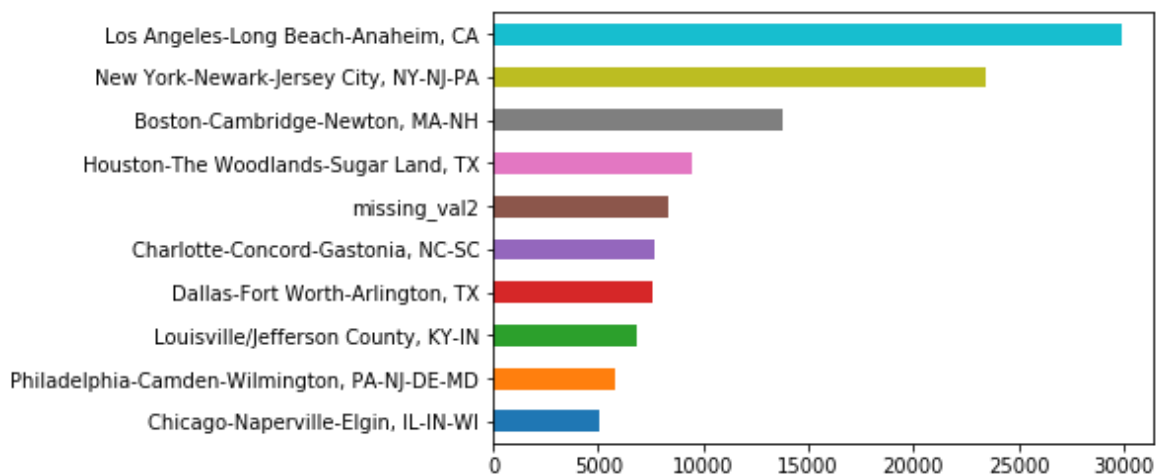
In [62]: df_nochurn['location'].value_counts()[:10].sort_values().plot(kind='barh')

```

```

Out[62]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea2af6390>

```



```

In [63]: df_nochurn['location'].nunique()

```

```

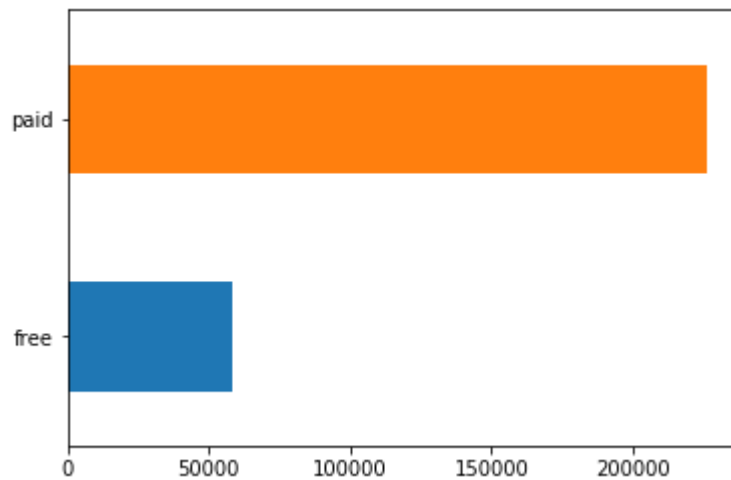
Out[63]: 115

```



```
In [64]: df_nochurn['level'].value_counts().sort_values().plot(kind='barh')
```

```
Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea0a11b00>
```



```
In [ ]:
```

```
In [ ]:
```

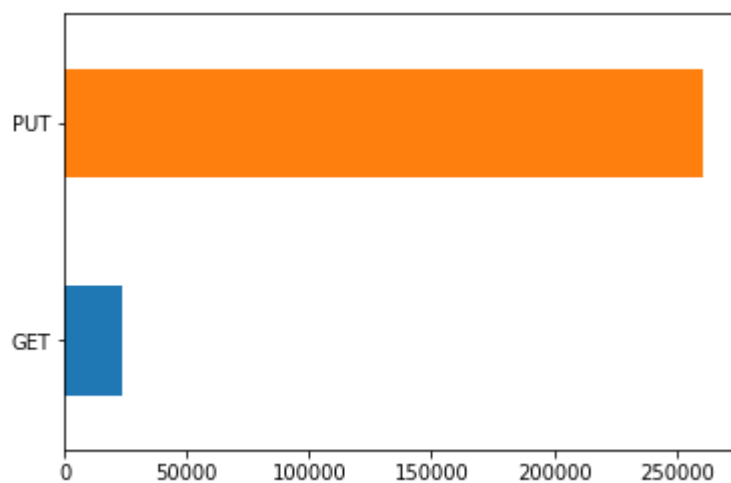
```
In [65]: #
```

```
In [66]: df_nochurn['method'].unique()
```

```
Out[66]: array(['PUT', 'GET'], dtype=object)
```

```
In [67]: df_nochurn['method'].value_counts().sort_values().plot(kind='barh')
```

```
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9e9fd37898>
```



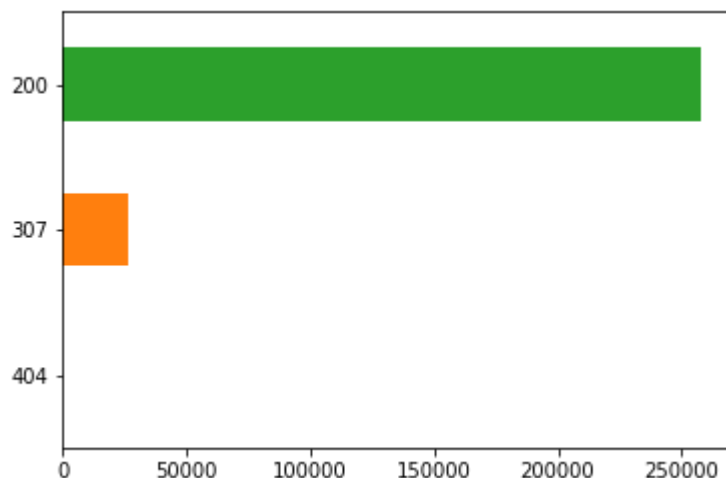
```
In [ ]:
```

No Churn by status

```
In [68]: df_nochurn['status'].nunique()
```

```
Out[68]: 3
```

```
In [69]: df_nochurn['status'].value_counts().sort_values().plot(kind='barh');
```



```
In [70]: df_nochurn['status'].value_counts()
```

```
Out[70]: 200    257705  
        307    26315  
        404      258  
        Name: status, dtype: int64
```

auth not useful at all

```
In [71]: df_nochurn['auth'].value_counts()
```

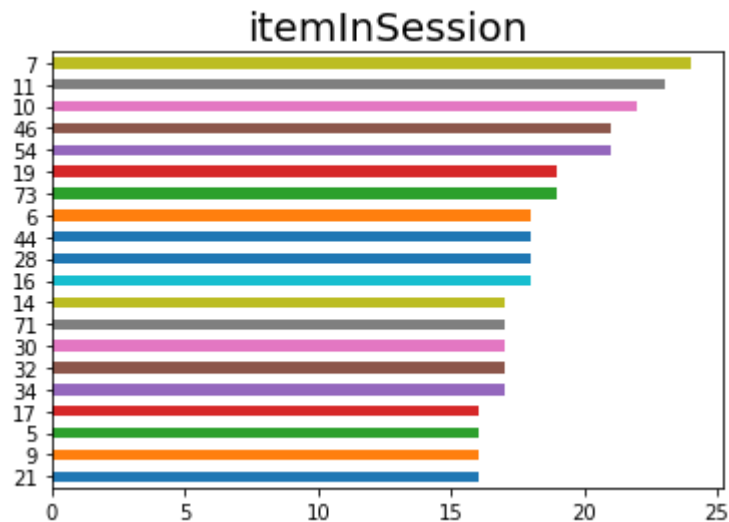
```
Out[71]: Logged In    275932  
        Logged Out    8249  
        Guest         97  
        Name: auth, dtype: int64
```

```
In [ ]:
```

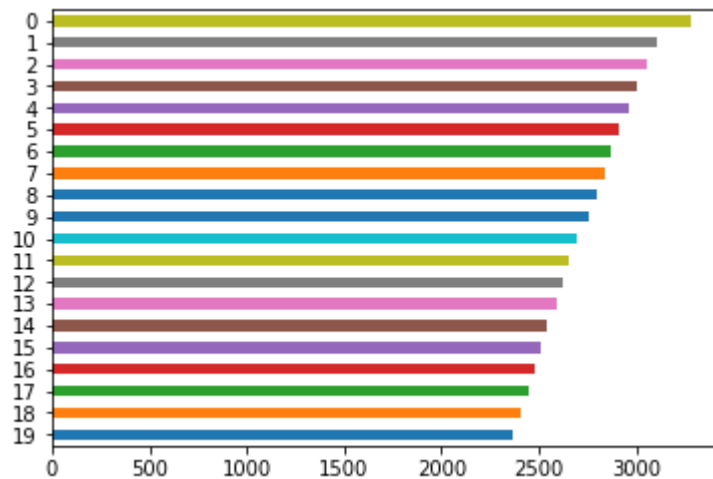
```
In [ ]:
```

Coming back to Churns!

```
In [72]: df_churn['itemInSession'].value_counts()[:20].sort_values().plot(kind
        = 'barh');
        plt.title('itemInSession',size=20)
        #plt.savefig('./churn_iteminsession.jpeg')
        plt.show()
```



```
In [73]: df_nochurn['itemInSession'].value_counts()[:20].sort_values().plot(ki
        nd= 'barh');
```



It changes from customers who churn to customers which did not churn!!

In []:

In []:

Temporal EDA

In [74]: `df_churn.head(2)`

Out[74]:

	artist	auth	firstName	gender	itemInSession	lastName	length	level	location
35	missing_val	Logged In	Alexi	F	4	Warren	249.117182	paid	Spot Valley
357	missing_val	Logged In	Faigy	F	29	Howe	249.117182	paid	Pho M Scotts

Let us set the initial date as 1970-01-01

In [75]: `df_churn['date']=pd.to_timedelta(df_churn['ts'],unit="ms")+ pd.Timestamp('1970-01-01')`

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>
 """Entry point for launching an IPython kernel.

In [76]: `df_nochurn['date']=pd.to_timedelta(df_nochurn['ts'],unit="ms")+ pd.Timestamp('1970-01-01')`

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>
 """Entry point for launching an IPython kernel.

In [77]: `df_churn.tail(2)`

Out[77]:

	artist	auth	firstName	gender	itemInSession	lastName	length	level	
285652	missing_val	Logged In	Joshua	M	44	White	249.117182	paid	E
286059	missing_val	Logged In	Joshua	M	81	Warner	249.117182	paid	S St (

2 rows × 21 columns

Creating the column 'day'

In [78]: `df_churn['day']=df_churn.date.dt.day`

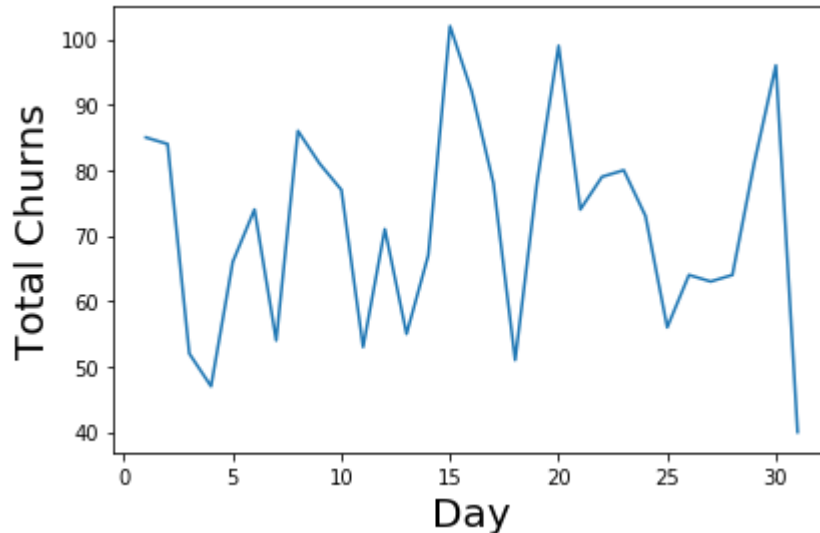
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

"""Entry point for launching an IPython kernel.

```
In [79]: df_churn.groupby(['day']).sum()['churn'].plot()  
plt.ylabel('Total Churns',size=20)  
plt.xlabel('Day',size=20)  
plt.tight_layout()  
#plt.savefig('./churn_totalchurnbyday.jpeg')  
plt.show()
```



This time series is interesting, it has an almost well-defined pattern, with the number of total churns reaching maximums and minimums almost periodically.

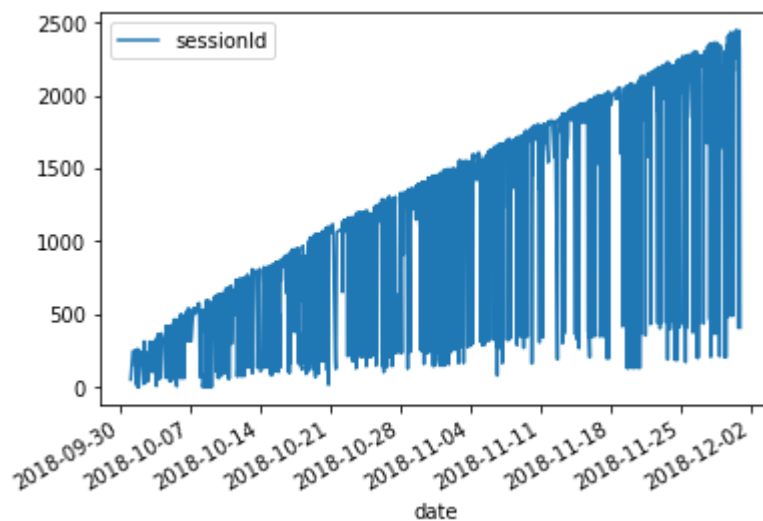
In []:

In []:

Plotting sessionId by date

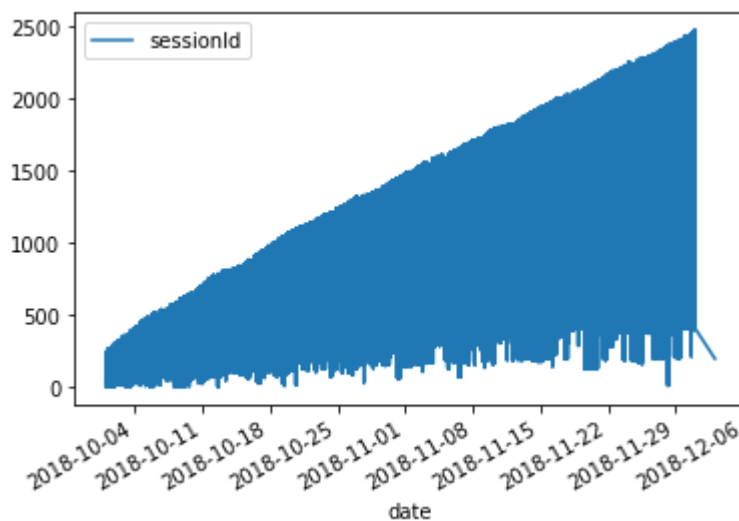
```
In [80]: df_churn.plot(x='date',y='sessionId')
```

```
Out[80]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9e9aa73e48>
```



```
In [81]: df_nochurn.plot(x='date',y='sessionId')
```

```
Out[81]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9ea589ecc0>
```



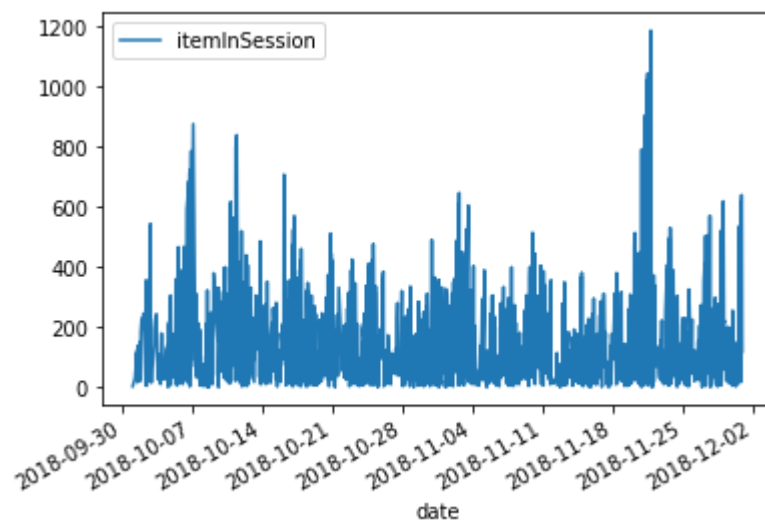
It is ok for churn and no-churn

```
In [ ]:
```

Plotting itemInSession by Date

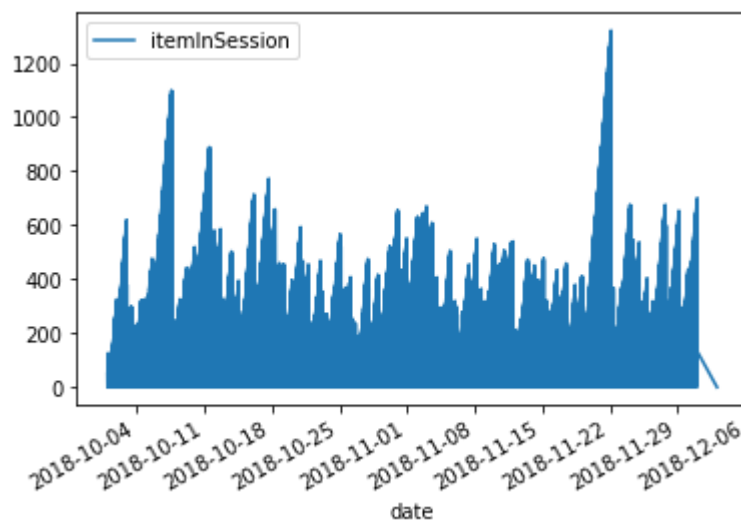
```
In [82]: df_churn.plot(x='date',y='itemInSession')
```

```
Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9eac3fbda0>
```



```
In [83]: df_nochurn.plot(x='date',y='itemInSession')
```

```
Out[83]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9e9c504400>
```



Seems to be ok!

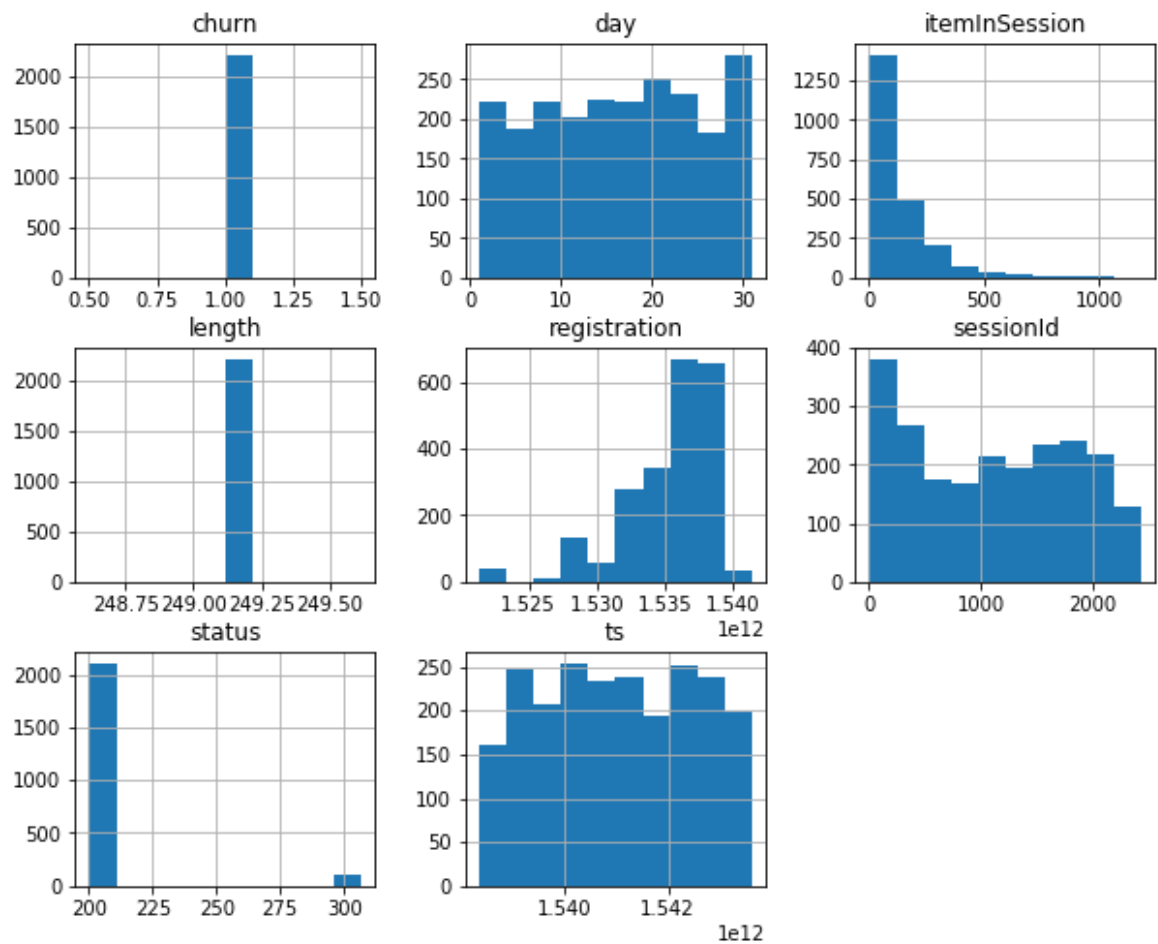
```
In [ ]:
```

```
In [ ]:
```

Plotting histograms of the time series

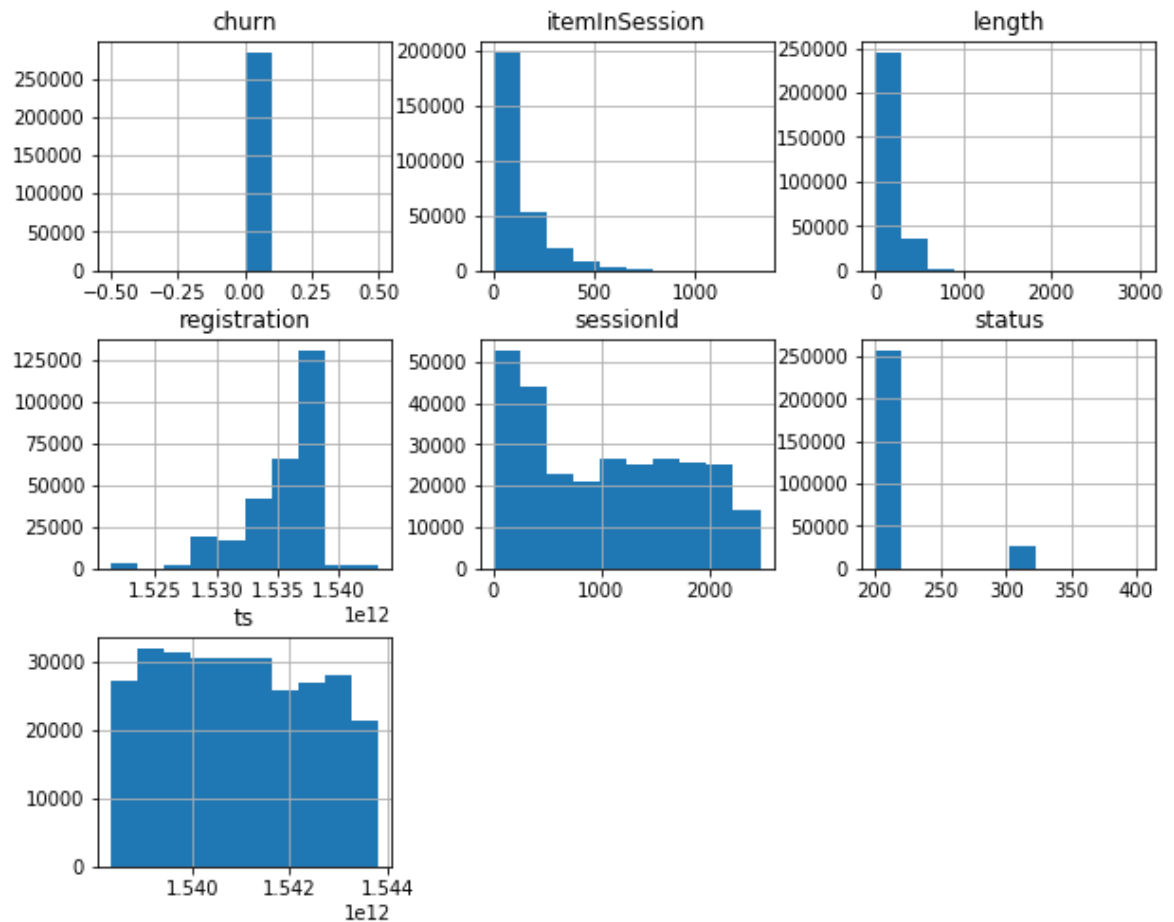

```
In [84]: plt.figure()
df_churn.hist(figsize=[10,8]);
plt.show()
```

<matplotlib.figure.Figure at 0x7f9e9ae99828>



```
In [85]: #plt.plot(df_nochurn[df_nochurn['gender']=='M']['date'],df_nochurn[df
_
_nochurn['gender']=='M']['gender'])
import matplotlib.pyplot as plt
df_nochurn.hist(figsize=[10,8])
#df_churn.hist(ax=axes[1],figsize=[10,8])
plt.suptitle('No churn', x=0.5, y=1.05, ha='center', fontsize='xx-lar
ge')
plt.show()
```

No churn



In []:

In []:

Feature Engineering

Once you've familiarized yourself with the data, build out the features you find promising to train your model on. To work with the full dataset, you can follow the following steps.

- Write a script to extract the necessary features from the smaller subset of data
- Ensure that your script is scalable, using the best practices discussed in Lesson 3
- Try your script on the full data set, debugging your script if necessary

If you are working in the classroom workspace, you can just extract features based on the small subset of data contained here. Be sure to transfer over this work to the larger dataset when you work on your Spark cluster.

Splitting into train and test!

In []:

In []:

In []:

In []:

In []:

In []:

In [86]: df.head(2)

Out[86]:

	artist	auth	firstName	gender	itemInSession	lastName	length	level	location
0	Martha Tilston	Logged In	Colin	M	50	Freeman	277.89016	paid	Bakersfield, CA
1	Five Iron Frenzy	Logged In	Micah	M	79	Long	236.09424	free	Boston-Cambridge-Newton, MA-NH

In [87]: to_drop_cols=['artist','auth','firstName','lastName','page','registration','sessionId','song','hour']

In [88]: data_churn=data_churn.drop(*to_drop_cols)

```
In [89]: data_churn.printSchema()
```

```
root
|-- gender: string (nullable = false)
|-- itemInSession: long (nullable = true)
|-- length: double (nullable = false)
|-- level: string (nullable = true)
|-- location: string (nullable = false)
|-- method: string (nullable = true)
|-- status: long (nullable = true)
|-- ts: long (nullable = true)
|-- userAgent: string (nullable = false)
|-- userId: string (nullable = true)
|-- churn: integer (nullable = true)
```

```
In [90]: data_churn = data_churn.withColumn("userId", data_churn["userId"].cast(IntegerType()))
```

```
In [91]: data_churn.printSchema()
```

```
root
|-- gender: string (nullable = false)
|-- itemInSession: long (nullable = true)
|-- length: double (nullable = false)
|-- level: string (nullable = true)
|-- location: string (nullable = false)
|-- method: string (nullable = true)
|-- status: long (nullable = true)
|-- ts: long (nullable = true)
|-- userAgent: string (nullable = false)
|-- userId: integer (nullable = true)
|-- churn: integer (nullable = true)
```

```
In [92]: num_cols=['itemInSession', 'length', 'status', 'ts', 'userId', 'churn']
```

```
In [ ]:
```

```
In [ ]:
```

```
In [93]: data_churn = data_churn.withColumn("rank", percent_rank().over(Window
.partitionBy().orderBy("ts")))
train_data = data_churn.where("rank <= .8").drop("rank")
test_data = data_churn.where("rank > .8").drop("rank")
```

One hot encoder

```
In [94]: stringInd_gender = StringIndexer(inputCol="gender", outputCol="gender_index",handleInvalid='skip')
#model = stringIndexer.fit(df2)
#indexed = model.transform(df2)
#encoder_gender = OneHotEncoder(dropLast=False, inputCol="gender_index", outputCol="gender_vec")
encoder_gender = OneHotEncoder(inputCol="gender_index", outputCol="gender_vec")

stringInd_level = StringIndexer(inputCol="level", outputCol="level_index",handleInvalid='skip')
#model = stringIndexer.fit(df2)
#indexed = model.transform(df2)
#encoder_level = OneHotEncoder(dropLast=False, inputCol="level_index", outputCol="level_vec")
encoder_level = OneHotEncoder(inputCol="level_index", outputCol="level_vec")

stringInd_method = StringIndexer(inputCol="method", outputCol="method_index",handleInvalid='skip')
#model = stringIndexer.fit(df2)
#indexed = model.transform(df2)
#encoder_method = OneHotEncoder(dropLast=False, inputCol="method_index", outputCol="method_vec")
encoder_method = OneHotEncoder(inputCol="method_index", outputCol="method_vec")
```

Just StringIndexer

```
In [95]: stringInd_location = StringIndexer(inputCol="location", outputCol="location_index",handleInvalid='skip')
#model = stringIndexer.fit(df2)
#indexed = model.transform(df2)
#encoder_location = OneHotEncoder(dropLast=False, inputCol="location_index", outputCol="location_vec")
encoder_location = OneHotEncoder(inputCol="location_index", outputCol="location_vec")

stringInd_Agent = StringIndexer(inputCol="userAgent", outputCol="userAgent_index",handleInvalid='skip')
#model = stringIndexer.fit(df2)
#indexed = model.transform(df2)
#encoder_Agent = OneHotEncoder(dropLast=False, inputCol="userAgent_index", outputCol="userAgent_vec")
encoder_Agent = OneHotEncoder(inputCol="userAgent_index", outputCol="userAgent_vec")
```

```
In [96]: vec_assem=VectorAssembler(inputCols=['itemInSession','length','status','ts','userId',"gender_vec","level_vec","method_vec","location_vec","userAgent_vec"],outputCol='features',handleInvalid='skip')
```

```
In [97]: #pipeline=Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_level,encoder_level,stringInd_method,encoder_method,stringInd_location,encoder_location,stringInd_Agent,encoder_Agent,vec_assem])
```

```
In [98]: #train_data=pipeline.fit(train_data).transform(train_data)
```

```
In [99]: #test_data=pipeline.fit(test_data).transform(test_data)
```

```
In [100]: train_data.show()
```

```
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
|gender|itemInSession|length|level|location|me
thod|status|ts|userAgent|userId|churn|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
|      M|          50| 277.89016| paid| Bakersfield, CA|
PUT|    200|1538352117000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          79| 236.09424| free| Boston-Cambridge-...|
PUT|    200|1538352180000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          51| 282.8273| paid| Bakersfield, CA|
PUT|    200|1538352394000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          80| 262.71302| free| Boston-Cambridge-...|
PUT|    200|1538352416000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          52| 223.60771| paid| Bakersfield, CA|
PUT|    200|1538352676000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          81| 208.29995| free| Boston-Cambridge-...|
PUT|    200|1538352678000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          82| 260.46649| free| Boston-Cambridge-...|
PUT|    200|1538352886000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          53| 185.44281| paid| Bakersfield, CA|
PUT|    200|1538352899000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          54|249.1171819778458| paid| Bakersfield, CA|
PUT|    200|1538352905000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          55| 134.47791| paid| Bakersfield, CA|
PUT|    200|1538353084000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          83| 229.8771| free| Boston-Cambridge-...|
PUT|    200|1538353146000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          84|249.1171819778458| free| Boston-Cambridge-...|
GET|    200|1538353150000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          56| 223.58159| paid| Bakersfield, CA|
PUT|    200|1538353218000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          85| 201.06404| free| Boston-Cambridge-...|
PUT|    200|1538353375000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          86|249.1171819778458| free| Boston-Cambridge-...|
PUT|    307|1538353376000|"Mozilla/5.0 (Win...|    9|    0|
|      M|          57| 246.69995| paid| Bakersfield, CA|
PUT|    200|1538353441000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          87| 168.64608| free| Boston-Cambridge-...|
PUT|    200|1538353576000|"Mozilla/5.0 (Win...|    9|    0|
|      F|          0| 166.1122| free| Tallahassee, FL|
PUT|    200|1538353668000|"Mozilla/5.0 (Mac...|    74|    0|
|      M|          58| 222.22322| paid| Bakersfield, CA|
PUT|    200|1538353687000|Mozilla/5.0 (Wind...|    30|    0|
|      M|          88| 229.77261| free| Boston-Cambridge-...|
PUT|    200|1538353744000|"Mozilla/5.0 (Win...|    9|    0|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```

```
In [101]: #data_trans.printSchema()
```

In []:

In [102]: `#train_data=data_trans.limit(int(int(data_trans.count())*0.8))`In [103]: `#test_data=data_trans.subtract(train_data)`

In []:

In []:

In []:

In []:

In []:

In []:

In []:

Modeling

Split the full dataset into train, test, and validation sets. Test out several of the machine learning methods you learned. Evaluate the accuracy of the various models, tuning parameters as necessary. Determine your winning model based on test accuracy and report results on the validation set. Since the churned users are a fairly small subset, I suggest using F1 score as the metric to optimize.

In []:

In []:

In []:

In [112]:

Logistic Regression


```

In [104]: lr = LogisticRegression(featuresCol='features',labelCol='churn',maxI
ter=2)#, regParam=0.0, elasticNetParam=0)
paramGrid = ParamGridBuilder().addGrid(lr.regParam, (0.05, 0.15))\
                                .addGrid(lr.elasticNetParam, (0.25,0.75
))\
                                .addGrid(lr.maxIter, (5, 10))\
                                .build()
#                                .addGrid(lr.elasticNetParam, (0.25,0.7
5))\
#                                .addGrid(lr.maxIter, (5, 10))\
#                                .addGrid(lr.tol, (1e-4, 1e-5))\
#pipeline = Pipeline(stages=[stringInd_gender,encoder_gender,stringInd
d_level,encoder_level,stringInd_method,encoder_method,stringInd_locat
ion,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,lr])
#pipeline = Pipeline(stages=[lr])

pipeline=Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_l
evel,encoder_level,stringInd_method,encoder_method,stringInd_location
,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,lr])

cv = CrossValidator(estimator=pipeline,
                    estimatorParamMaps=paramGrid,
                    evaluator=MulticlassClassificationEvaluator(label
Col='churn',metricName="f1"),
#                    evaluator=MulticlassClassificationEvaluator(labe
lCol='churn',predictionCol='prediction'),
                    numFolds=4)

```

```

In [105]: lrModel=cv.fit(train_data)

```

```

In [106]: lrModel.avgMetrics

```

```

Out[106]: [0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097]

```

```

In [107]: lrModel.bestModel

```

```

Out[107]: PipelineModel_d01d7ca15680

```

```

In [101]: #'lrPredictions = lrModel.transform(test_data)

```

getting the best parameters

```
In [114]: lrModel.getEstimatorParamMaps()[ np.argmax(lrModel.avgMetrics) ]
```

```
Out[114]: {Param(parent='LogisticRegression_677982e4770d', name='regParam', doc
='regularization parameter (>= 0).'): 0.05,
  Param(parent='LogisticRegression_677982e4770d', name='elasticNetParam', doc='the ElasticNet mixing parameter, in range [0, 1]. For alpha
= 0, the penalty is an L2 penalty. For alpha = 1, it is an L1 penalt
y.'): 0.25,
  Param(parent='LogisticRegression_677982e4770d', name='maxIter', doc
='max number of iterations (>= 0).'): 5}
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [123]: #eval_f1 = MulticlassClassificationEvaluator(labelCol="churn", predic
tionCol="prediction", metricName="f1")
eval_f1 = MulticlassClassificationEvaluator(labelCol="churn", predict
ionCol="prediction")
```

```
In [124]: eval_f1
```

```
Out[124]: MulticlassClassificationEvaluator_34f78acd3407
```

```
In [147]: #f1score = eval_f1.evaluate(lrPredictions)
```

```
In [151]: train_data.printSchema()
```

```
root
|-- gender: string (nullable = false)
|-- itemInSession: long (nullable = true)
|-- length: double (nullable = false)
|-- level: string (nullable = true)
|-- location: string (nullable = false)
|-- method: string (nullable = true)
|-- status: long (nullable = true)
|-- ts: long (nullable = true)
|-- userAgent: string (nullable = false)
|-- userId: integer (nullable = true)
|-- churn: integer (nullable = true)
|-- gender_index: double (nullable = false)
|-- gender_vec: vector (nullable = true)
|-- level_index: double (nullable = false)
|-- level_vec: vector (nullable = true)
|-- method_index: double (nullable = false)
|-- method_vec: vector (nullable = true)
|-- location_index: double (nullable = false)
|-- location_vec: vector (nullable = true)
|-- userAgent_index: double (nullable = false)
|-- userAgent_vec: vector (nullable = true)
|-- features: vector (nullable = true)
```

```
In [150]: test_data.printSchema()
```

```
root
|-- gender: string (nullable = false)
|-- itemInSession: long (nullable = true)
|-- length: double (nullable = false)
|-- level: string (nullable = true)
|-- location: string (nullable = false)
|-- method: string (nullable = true)
|-- status: long (nullable = true)
|-- ts: long (nullable = true)
|-- userAgent: string (nullable = false)
|-- userId: integer (nullable = true)
|-- churn: integer (nullable = true)
|-- gender_index: double (nullable = false)
|-- gender_vec: vector (nullable = true)
|-- level_index: double (nullable = false)
|-- level_vec: vector (nullable = true)
|-- method_index: double (nullable = false)
|-- method_vec: vector (nullable = true)
|-- location_index: double (nullable = false)
|-- location_vec: vector (nullable = true)
|-- userAgent_index: double (nullable = false)
|-- userAgent_vec: vector (nullable = true)
|-- features: vector (nullable = true)
```

```
In [ ]: #lrModel.avgMetrics
```

In [192]: `lrPredictions.show()`

```

+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+
|gender|itemInSession|length|level|location|method|status|ts|userAgent|userId|churn|gender_index|gender_vec|level_index|level_vec|method_index|method_vec|location_index|location_vec|userAgent_index|userAgent_vec|features|rawPrediction|probability|prediction|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+
|F|216|201.79546|free|Tampa-St. Petersburg|PUT|200|1542568361000|"Mozilla/5.0 (Mac...|25|0|0.0|(2,[0],[1.0]))|1.0|(1,[],[])|0.0|(1,[0],[1.0])|27.0|(113,[27],[1.0])|2.0|(56,[2],[1.0])|(178,[0,1,2,3,4,5...|[4.95028235330991...|[0.99296838455876...|0.0|
|F|217|249.1171819778458|free|Tampa-St. Petersburg|PUT|307|1542568362000|"Mozilla/5.0 (Mac...|25|0|0.0|(2,[0],[1.0]))|1.0|(1,[],[])|0.0|(1,[0],[1.0])|27.0|(113,[27],[1.0])|2.0|(56,[2],[1.0])|(178,[0,1,2,3,4,5...|[4.95028235270377...|[0.99296838455453...|0.0|
|F|220|249.1171819778458|free|Tampa-St. Petersburg|GET|200|1542568444000|"Mozilla/5.0 (Mac...|25|0|0.0|(2,[0],[1.0]))|1.0|(1,[],[])|1.0|(1,[],[])|27.0|(113,[27],[1.0])|2.0|(56,[2],[1.0])|(178,[0,1,2,3,4,5...|[3.86529941326765...|[0.97947351940438...|0.0|
|M|301|233.37751|paid|Concord, NH|PUT|200|1542568480000|Mozilla/5.0 (X11;...|65|0|1.0|(2,[1],[1.0]))|0.0|(1,[0],[1.0])|0.0|(1,[0],[1.0])|36.0|(113,[36],[1.0])|35.0|(56,[35],[1.0])|(178,[0,1,2,3,4,6...|[4.95028228117821...|[0.99296838405512...|0.0|
|F|60|242.18077|paid|Boston-Cambridge-...|PUT|200|1542568505000|"Mozilla/5.0 (iPa...|92|0|0.0|(2,[0],[1.0]))|0.0|(1,[0],[1.0])|0.0|(1,[0],[1.0])|2.0|(113,[2],[1.0])|10.0|(56,[10],[1.0])|(178,[0,1,2,3,4,5...|[4.95028226602449...|[0.99296838394932...|0.0|
|F|32|238.52363|free|Los Angeles-Long ...|PUT|200|1542568512000|"Mozilla/5.0 (Win...|39|0|0.0|(2,[0],[1.0]))|1.0|(1,[],[])|0.0|(1,[0],[1.0])|0.0|(113,[0],[1.0])|5.0|(56,[5],[1.0])|(178,[0,1,2,3,4,5...|[4.95028226178145...|[0.99296838391969...|0.0|
|F|61|249.1171819778458|paid|Boston-Cambridge-...|GET|200|1542568512000|"Mozilla/5.0 (iPa...|92|0|0.0|(2,[0],[1.0]))|0.0|(1,[0],[1.0])|1.0|(1,[],[])|2.0|(113,[2],[1.0])|10.0|(56,[10],[1.0])|

```

```

(178,[0,1,2,3,4,5...|[3.86529937204953...|[0.97947351857568...|
0.0|
|      F|              8|          480.39138| paid|      Pittsburgh, PA|
PUT|    200|1542568525000|"Mozilla/5.0 (X11...|    145|    0|
0.0|(2,[0],[1.0])|          0.0|(1,[0],[1.0])|          0.0|(1,[0],[1.
0])|
          42.0|(113,[42],[1.0])|          28.0|(56,[28],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028225390152...|[0.99296838386467...|
0.0|
|      F|          221|          212.4273| free|Tampa-St. Petersb...|
PUT|    200|1542568562000|"Mozilla/5.0 (Mac...|    25|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          0.0|(1,[0],[1.
0])|
          27.0|(113,[27],[1.0])|          2.0| (56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028223147401...|[0.99296838370808...|
0.0|
|      F|          222|249.1171819778458| free|Tampa-St. Petersb...|
PUT|    307|1542568563000|"Mozilla/5.0 (Mac...|    25|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          0.0|(1,[0],[1.
0])|
          27.0|(113,[27],[1.0])|          2.0| (56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028223086786...|[0.99296838370385...|
0.0|
|      F|          223|249.1171819778458| free|Tampa-St. Petersb...|
PUT|    307|1542568564000|"Mozilla/5.0 (Mac...|    25|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          0.0|(1,[0],[1.
0])|
          27.0|(113,[27],[1.0])|          2.0| (56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028223026172...|[0.99296838369961...|
0.0|
|      F|          224|249.1171819778458| free|Tampa-St. Petersb...|
GET|    200|1542568683000|"Mozilla/5.0 (Mac...|    25|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          1.0|    (1,[],
[])|
          27.0|(113,[27],[1.0])|          2.0| (56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[3.86529926839810...|[0.97947351649175...|
0.0|
|      M|          302|          216.65914| paid|      Concord, NH|
PUT|    200|1542568713000|Mozilla/5.0 (X11;...|    65|    0|
1.0|(2,[1],[1.0])|          0.0|(1,[0],[1.0])|          0.0|(1,[0],[1.
0])|
          36.0|(113,[36],[1.0])|          35.0|(56,[35],[1.0])|
(178,[0,1,2,3,4,6...|[4.95028213994555...|[0.99296838306901...|
0.0|
|      F|          62|          158.87628| paid|Boston-Cambridge-...|
PUT|    200|1542568747000|"Mozilla/5.0 (iPa...|    92|    0|
0.0|(2,[0],[1.0])|          0.0|(1,[0],[1.0])|          0.0|(1,[0],[1.
0])|
          2.0| (113,[2],[1.0])|          10.0|(56,[10],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028211933649...|[0.99296838292511...|
0.0|
|      F|          33|          225.90649| free|Los Angeles-Long ...|
PUT|    200|1542568750000|"Mozilla/5.0 (Win...|    39|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          0.0|(1,[0],[1.
0])|
          0.0| (113,[0],[1.0])|          5.0| (56,[5],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028211751805...|[0.99296838291242...|
0.0|
|      F|          225|          277.15873| free|Tampa-St. Petersb...|
PUT|    200|1542568774000|"Mozilla/5.0 (Mac...|    25|    0|
0.0|(2,[0],[1.0])|          1.0|    (1,[],[])|          0.0|(1,[0],[1.
0])|
          27.0|(113,[27],[1.0])|          2.0| (56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028210297048...|[0.99296838281084...|
0.0|
|      F|          63|249.1171819778458| paid|Boston-Cambridge-...|

```

```

GET|      200|1542568789000|"Mozilla/5.0 (iPa...|      92|      0|
0.0|(2,[0],[1.0])|      0.0|(1,[0],[1.0])|      1.0|(1,[],
[])|      2.0|(113,[2],[1.0])|      10.0|(56,[10],[1.0])|
(178,[0,1,2,3,4,5...|[3.86529920414633...|[0.97947351519996...|
0.0|
|      F|      64|      278.41261| paid|Boston-Cambridge-...|
PUT|      200|1542568905000|"Mozilla/5.0 (iPa...|      92|      0|
0.0|(2,[0],[1.0])|      0.0|(1,[0],[1.0])|      0.0|(1,[0],[1.
0])|      2.0|(113,[2],[1.0])|      10.0|(56,[10],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028202356499...|[0.99296838225642...|
0.0|
|      F|      65|249.1171819778458| paid|Boston-Cambridge-...|
PUT|      307|1542568906000|"Mozilla/5.0 (iPa...|      92|      0|
0.0|(2,[0],[1.0])|      0.0|(1,[0],[1.0])|      0.0|(1,[0],[1.
0])|      2.0|(113,[2],[1.0])|      10.0|(56,[10],[1.0])|
(178,[0,1,2,3,4,5...|[4.95028202295884...|[0.99296838225219...|
0.0|
|      F|      226|249.1171819778458| free|Tampa-St. Petersb...|
GET|      200|1542568907000|"Mozilla/5.0 (Mac...|      25|      0|
0.0|(2,[0],[1.0])|      1.0|(1,[],[])|      1.0|(1,[],
[])|      27.0|(113,[27],[1.0])|      2.0|(56,[2],[1.0])|
(178,[0,1,2,3,4,5...|[3.86529913262078...|[0.97947351376193...|
0.0|
+-----+-----+-----+-----+-----+-----+
- - - + - - - + - - - + - - - + - - - + - - - + - - - +
- - - + - - - + - - - + - - - + - - - + - - - + - - - +
- + - - - + - - - + - - - + - - - + - - - + - - - + - - - +
- - - - - + - - - - - + - - - - - + - - - - - + - - - - - +
- +
only showing top 20 rows

```

```
In [102]: predictionAndLabels=lrPredictions.select('prediction','churn')
```

```
In [106]: predictionAndLabels.show()
```

```
+-----+-----+
|prediction|churn|
+-----+-----+
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
|          |0.0|  0|
+-----+-----+
only showing top 20 rows
```

```
In [136]: from pyspark.sql.types import FloatType
```

```
predictionAndLabels = predictionAndLabels.withColumn("churn", predictionAndLabels["churn"].cast(FloatType()))
predictionAndLabels = predictionAndLabels.withColumn("prediction", predictionAndLabels["prediction"].cast(FloatType()))
```



```
In [137]: predictionAndLabels.show()
```

[illegible]

```
In [144]: results = lrModel.transform(test_data)
```

```
evaluator = MulticlassClassificationEvaluator(labelCol='churn', predictionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.metricName: "f1"})))
```

F1 Score:0.988297408226171

In []:

Linear Regression

In []:

In [199]:

```

In [115]: lreg = LinearRegression(maxIter=10, regParam=0.3, elasticNetParam=0.8
, labelCol='churn')
paramGrid = ParamGridBuilder().addGrid(lr.regParam, (0.0, 0.2, 0.5, 0.8
))\
                                .addGrid(lr.elasticNetParam, (0.0, 0.25,
0.5, 0.75))\
                                .build()
#                                .addGrid(lr.maxIter, (5, 10))\
#                                .addGrid(lr.elasticNetParam, (0.25, 0.7
5))\
#                                .addGrid(lr.maxIter, (5, 10))\
#                                .addGrid(lr.tol, (1e-4, 1e-5))\
#pipeline = Pipeline(stages=[stringInd_gender, encoder_gender, stringInd
d_level, encoder_level, stringInd_method, encoder_method, stringInd_locat
ion, encoder_location, stringInd_Agent, encoder_Agent, vec_assem, lr])
#pipeline = Pipeline(stages=[lr])

pipeline=Pipeline(stages=[stringInd_gender, encoder_gender, stringInd_l
evel, encoder_level, stringInd_method, encoder_method, stringInd_location
, encoder_location, stringInd_Agent, encoder_Agent, vec_assem, lreg])

cv = CrossValidator(estimator=pipeline,
                    estimatorParamMaps=paramGrid,
                    evaluator=MulticlassClassificationEvaluator(label
Col='churn', metricName="f1"),
#                    evaluator=MulticlassClassificationEvaluator(labe
lCol='churn', predictionCol='prediction'),
                    numFolds=4)

```

```

In [116]: lregModel=cv.fit(train_data)

```

```

In [117]: lregModel.avgMetrics

```

```

Out[117]: [0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0,
0.0]

```

```
In [118]: results = lregModel.transform(test_data)

evaluator = MulticlassClassificationEvaluator(labelCol='churn', predictionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.metricName: "f1"})))
```

F1 Score:0.0

In []:

Getting the best parameters

```
In [119]: lregModel.getEstimatorParamMaps()[ np.argmax(lregModel.avgMetrics) ]
```

```
Out[119]: {Param(parent='LogisticRegression_677982e4770d', name='regParam', doc='regularization parameter (>= 0).'): 0.0,
  Param(parent='LogisticRegression_677982e4770d', name='elasticNetParam', doc='the ElasticNet mixing parameter, in range [0, 1]. For alpha = 0, the penalty is an L2 penalty. For alpha = 1, it is an L1 penalty.'): 0.0}
```

In []:

In []:

In []:

In []:

Random Forest

```
In [120]: rf = RandomForestClassifier(labelCol="churn", featuresCol="features", numTrees=100)
```

```

In [121]: paramGrid = ParamGridBuilder() \
            .addGrid(rf.numTrees, [50,100]) \
            .addGrid(rf.maxDepth, [15]) \
            .build()
            # .addGrid(rf.maxDepth, [3,6,9]) \
            # .addGrid(rf.numTrees, [int(x) for x in np.linspace(start = 10, s
            top = 50, num = 3)]) \
            # .addGrid(rf.maxDepth, [int(x) for x in np.linspace(start = 5, st
            op = 25, num = 3)]) \
            # .addGrid(lr.elasticNetParam, (0.25,0.7
            5))\
            # .addGrid(lr.maxIter, (5, 10))\
            # .addGrid(lr.elasticNetParam, (0.25,0.7
            5))\
            # .addGrid(lr.maxIter, (5, 10))\
            # .addGrid(lr.tol, (1e-4, 1e-5))\
            #pipeline = Pipeline(stages=[stringInd_gender,encoder_gender,stringIn
            d_level,encoder_level,stringInd_method,encoder_method,stringInd_locat
            ion,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,lr])
            #pipeline = Pipeline(stages=[lr])

            pipeline=Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_l
            evel,encoder_level,stringInd_method,encoder_method,stringInd_location
            ,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,rf])

            cv = CrossValidator(estimator=pipeline,
                                estimatorParamMaps=paramGrid,
                                evaluator=MulticlassClassificationEvaluator(label
            Col='churn',metricName="f1"),
                                # evaluator=MulticlassClassificationEvaluator(labe
            lCol='churn',predictionCol='prediction'),
                                numFolds=4)

```

```

In [122]: rfModel=cv.fit(train_data)

```

```

In [123]: rfModel.avgMetrics

```

```

Out[123]: [0.9879676156261097, 0.9879676156261097]

```

```

In [118]: #lrModel.avgMetrics

```

Getting the best parameters!

```

In [124]: rfModel.getEstimatorParamMaps()[ np.argmax(rfModel.avgMetrics) ]

```

```

Out[124]: {Param(parent='RandomForestClassifier_7b858e8cac14', name='numTrees',
doc='Number of trees to train (>= 1).'): 50,
Param(parent='RandomForestClassifier_7b858e8cac14', name='maxDepth',
doc='Maximum depth of the tree. (>= 0) E.g., depth 0 means 1 leaf nod
e; depth 1 means 1 internal node + 2 leaf nodes.'): 15}

```

In []:

In []:

```
In [119]: results = rfModel.transform(test_data)

evaluator = MulticlassClassificationEvaluator(labelCol='churn', predictionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.metricName: "f1"})))

F1 Score:0.988297408226171
```

In []:

```
In [120]: bestPipeline = rfModel.bestModel
bestModel = bestPipeline.stages[-1]
importances = bestModel.featureImportances
```

```
In [130]: #importances#.toArray()
```

```
In [122]: feature_list=np.array([i for i in range(178)])
```

```
In [123]: #feature_list
```

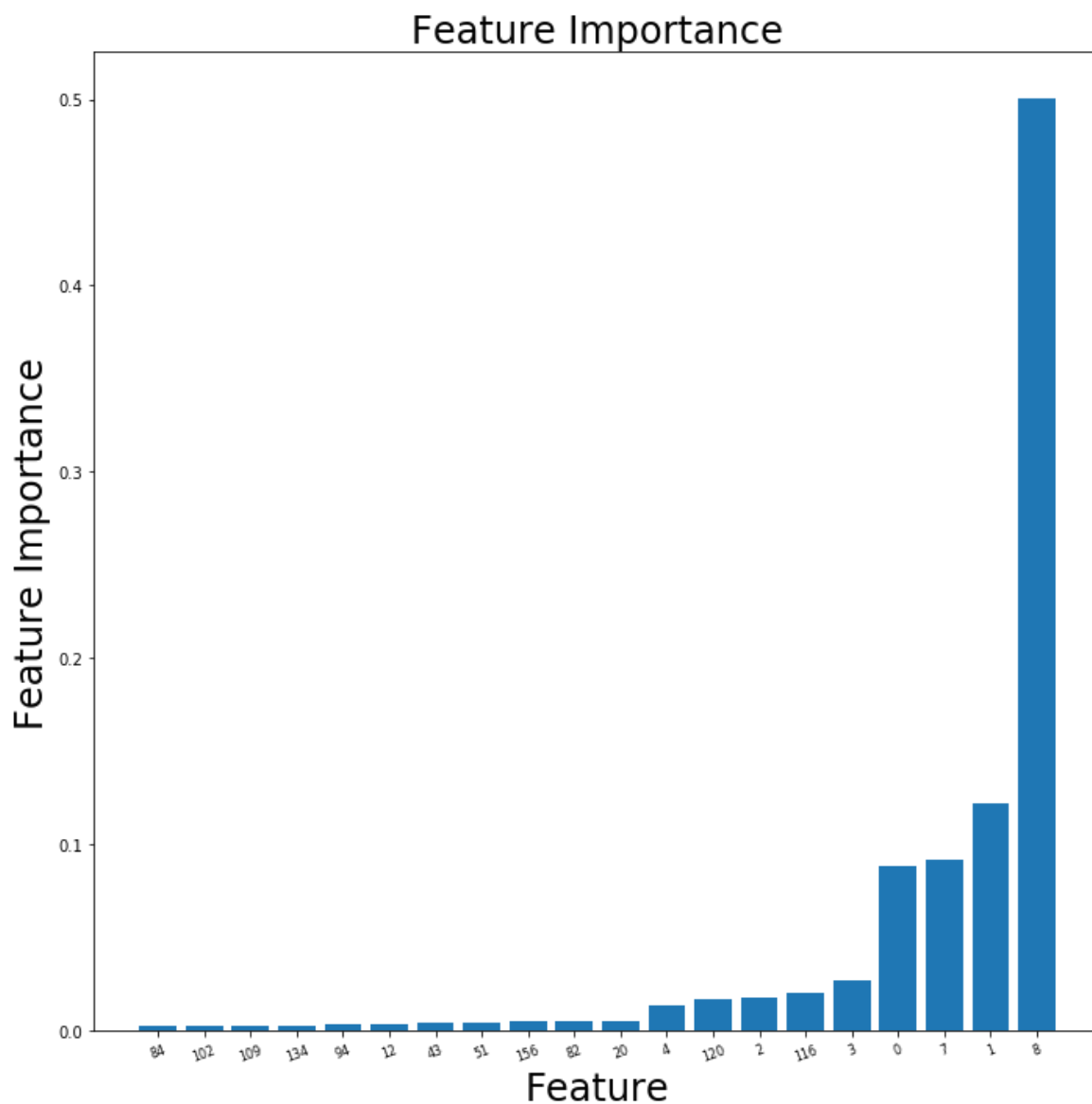
```

In [127]: num_features = 10
indices = np.argsort(importances)
indices2 = np.argsort(importances)

plt.figure(figsize=(10,10))
plt.title('Feature Importances')

# only plot the customized number of features
plt.bar(range(158,178), importances.toArray()[indices][158:178])
# Add feature names as x-axis labels
plt.xticks(range(158,178), feature_list[indices2][158:178], rotation=
20, fontsize = 8)
# Create plot title
plt.title("Feature Importance",size=24)
plt.xlabel('Feature',size=24)
plt.ylabel('Feature Importance',size=24)
plt.tight_layout()
#plt.savefig('./churn_rf_featimp.jpeg')
# Show plot
plt.show()

```



In []:

The three most important features are:

In [129]: `train_data.columns[8],train_data.columns[1],train_data.columns[7]`

Out[129]: `('userAgent', 'itemInSession', 'ts')`

In []:

In []:

In []:

In []:

In []:

In []:

OneVsRest

```

In [131]: lr = LogisticRegression(featuresCol='features',labelCol='churn',maxI
ter=2)#, regParam=0.0, elasticNetParam=0)
paramGrid = ParamGridBuilder().addGrid(lr.regParam, (0.05, 0.15))\
                                .addGrid(lr.elasticNetParam, (0.25,0.75
))\
                                .addGrid(lr.maxIter, (5, 10))\
                                .build()
#                                .addGrid(lr.elasticNetParam, (0.25,0.7
5))\
#                                .addGrid(lr.maxIter, (5, 10))\
#                                .addGrid(lr.tol, (1e-4, 1e-5))\
#pipeline = Pipeline(stages=[stringInd_gender,encoder_gender,stringInd
d_level,encoder_level,stringInd_method,encoder_method,stringInd_locat
ion,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,lr])
#pipeline = Pipeline(stages=[lr])

ovr=OneVsRest(classifier=lr,labelCol='churn')

pipeline=Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_l
evel,encoder_level,stringInd_method,encoder_method,stringInd_location
,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,ovr])

cv = CrossValidator(estimator=pipeline,
                    estimatorParamMaps=paramGrid,
                    evaluator=MulticlassClassificationEvaluator(label
Col='churn',metricName="f1"),
#                    evaluator=MulticlassClassificationEvaluator(labe
lCol='churn',predictionCol='prediction'),
                    numFolds=4)

```

```

In [132]: ORModel=cv.fit(train_data)

```

```

In [133]: ORModel.avgMetrics

```

```

Out[133]: [0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097,
0.9879676156261097]

```

```

In [134]: results = ORModel.transform(test_data)

evaluator = MulticlassClassificationEvaluator(labelCol='churn',predic
tionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.met
ricName: "f1"})))

```

```

F1 Score:0.988297408226171

```



```
In [135]: ORModel.getEstimatorParamMaps()[ np.argmax(ORModel.avgMetrics) ]
```

```
Out[135]: {Param(parent='LogisticRegression_d6e38016029d', name='regParam', doc
='regularization parameter (>= 0).'): 0.05,
  Param(parent='LogisticRegression_d6e38016029d', name='elasticNetParam', doc='the ElasticNet mixing parameter, in range [0, 1]. For alpha
= 0, the penalty is an L2 penalty. For alpha = 1, it is an L1 penalt
y.'): 0.25,
  Param(parent='LogisticRegression_d6e38016029d', name='maxIter', doc
='max number of iterations (>= 0).'): 5}
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

SVM

```
In [185]:
```

```
In [ ]:
```

```
In [136]: lsvc = LinearSVC(maxIter=10, regParam=0.1, labelCol='churn')
paramGrid = ParamGridBuilder().addGrid(lsvc.regParam, (0.05, 0.1, 0.2))\
    .build()
#                               .addGrid(lr.elasticNetParam, (0.25, 0.75))\
#                               .addGrid(lr.maxIter, (5, 10))\
#                               .addGrid(lr.tol, (1e-4, 1e-5))\
#pipeline = Pipeline(stages=[stringInd_gender, encoder_gender, stringInd_level, encoder_level, stringInd_method, encoder_method, stringInd_location, encoder_location, stringInd_Agent, encoder_Agent, vec_assem, lr])
#pipeline = Pipeline(stages=[lr])

pipeline=Pipeline(stages=[stringInd_gender, encoder_gender, stringInd_level, encoder_level, stringInd_method, encoder_method, stringInd_location, encoder_location, stringInd_Agent, encoder_Agent, vec_assem, lsvc])

cv = CrossValidator(estimator=pipeline,
                    estimatorParamMaps=paramGrid,
                    evaluator=MulticlassClassificationEvaluator(labelCol='churn', metricName="f1"),
#                    evaluator=MulticlassClassificationEvaluator(labelCol='churn', predictionCol='prediction'),
                    numFolds=4)
```

```
In [137]: SVMModel=cv.fit(train_data)
```

```
In [138]: SVMModel.avgMetrics
```

```
Out[138]: [0.9816338809740289, 0.9879676156261097, 0.985034316190955]
```

```
In [139]: results = SVMModel.transform(test_data)

evaluator = MulticlassClassificationEvaluator(labelCol='churn', predictionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.metricName: "f1"})))
```

```
F1 Score:0.988297408226171
```

```
In [ ]:
```

Getting the best parameters!

```
In [140]: SVMModel.getEstimatorParamMaps()[ np.argmax(SVMModel.avgMetrics) ]
```

```
Out[140]: {Param(parent='LinearSVC_ca47bf918ce0', name='regParam', doc='regularization parameter (>= 0).'): 0.1}
```

```
In [ ]:
```

In []:

In []:

GBTClassifier

In [204]:

```
In [125]: gbt = GBTClassifier(labelCol="churn", featuresCol="features", maxIter=10)
```

```
In [126]: paramGrid = ParamGridBuilder() \
    .build()
    # .addGrid(gbt.numFolds, [50,100]) \
    # .addGrid(gbt.MaxDepth, [12,15]) \
    # .addGrid(rf.maxDepth, [3,6,9]) \
    # .addGrid(rf.numTrees, [int(x) for x in np.linspace(start = 10, s
top = 50, num = 3)]) \
    # .addGrid(rf.maxDepth, [int(x) for x in np.linspace(start = 5, st
op = 25, num = 3)]) \
    #                                     .addGrid(lr.elasticNetParam, (0.25,0.7
5))\
    #                                     .addGrid(lr.maxIter, (5, 10))\
    #                                     .addGrid(lr.elasticNetParam, (0.25,0.7
5))\
    #                                     .addGrid(lr.maxIter, (5, 10))\
    #                                     .addGrid(lr.tol, (1e-4, 1e-5))\
    #pipeline = Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_l
evel,encoder_level,stringInd_method,encoder_method,stringInd_locat
ion,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,lr])
    #pipeline = Pipeline(stages=[lr])

    pipeline=Pipeline(stages=[stringInd_gender,encoder_gender,stringInd_l
evel,encoder_level,stringInd_method,encoder_method,stringInd_location
,encoder_location,stringInd_Agent,encoder_Agent,vec_assem,gbt])

    cv = CrossValidator(estimator=pipeline,
                        estimatorParamMaps=paramGrid,
                        evaluator=MulticlassClassificationEvaluator(label
Col='churn',metricName="f1"),
    #                               evaluator=MulticlassClassificationEvaluator(labe
lCol='churn',predictionCol='prediction'),
                        numFolds=4)
```

```
In [127]: GBTModel=cv.fit(train_data)
```

```
In [128]: GBTModel.avgMetrics
```

```
Out[128]: [0.9879450988392603]
```

```
In [129]: results = GBTModel.transform(test_data)

evaluator = MulticlassClassificationEvaluator(labelCol='churn',predictionCol = "prediction")
print('F1 Score:{}'.format(evaluator.evaluate(results, {evaluator.metricName: "f1"})))
```

F1 Score:0.988297408226171

In []:

Getting best parameters!

```
In [130]: GBTModel.getEstimatorParamMaps()[ np.argmax(GBTModel.avgMetrics) ]
```

Out[130]: {}

Because they are the default parameters!

In []:

In []:

Summarizing the results in a DataFrame

In []:

```
In [103]: results_df=pd.DataFrame([0.98829,0.0,0.98829,0.98829,0.98829,0.98799],index=['LR','LReg','RF','OVR','SVM','GBT'],columns=['f1-score'])
```

```
In [104]: results_df
```

Out[104]:

	f1-score
LR	0.98829
LReg	0.00000
RF	0.98829
OVR	0.98829
SVM	0.98829
GBT	0.98799

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

Final Steps

Clean up your code, adding comments and renaming variables to make the code easier to read and maintain. Refer to the Spark Project Overview page and Data Scientist Capstone Project Rubric to make sure you are including all components of the capstone project and meet all expectations. Remember, this includes thorough documentation in a README file in a Github repository, as well as a web app or blog post.

In []: