

Data Science & ML Course

Lesson #3 [Part #1] Python Introduction

Ivanovitch Silva
September, 2018



Agenda

- Modules
- Iterations
- List comprehension



top100.csv



Update from repository

```
git clone https://github.com/ivanovitchm/datascience2machinelearning.git
```

Or

```
git pull
```

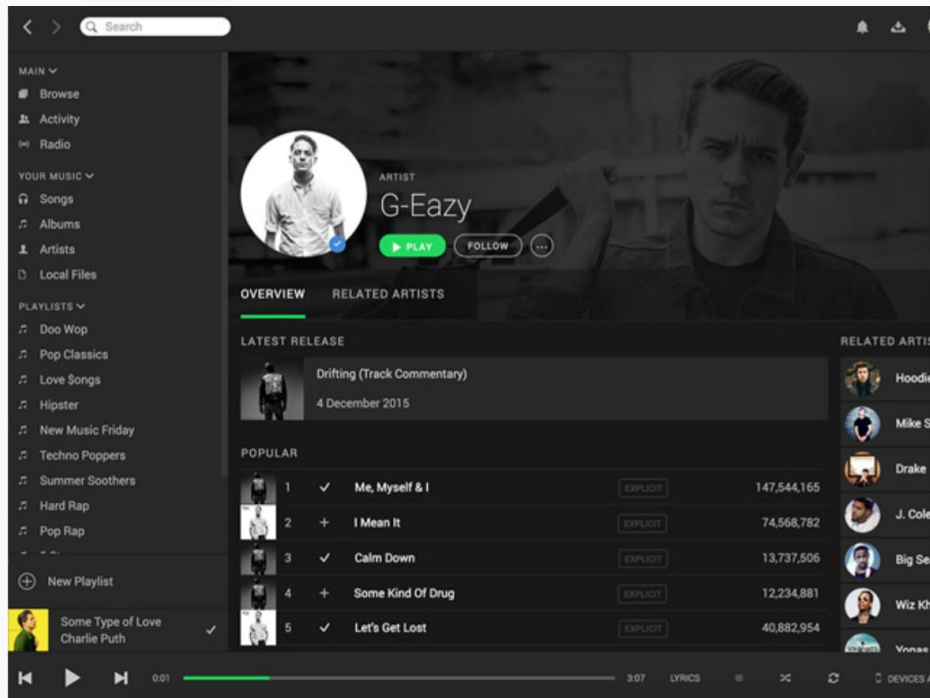


Spotify

As of January 2018, Spotify has over 70 million paying users

What are the average total streams for each song in the top 100?

Which song was the most popular song of 2017?



Spotify's Worldwide Daily Song Ranking

kaggle

	Track Name	Artist	Position	Streams
77	Sign of the Times	Harry Styles	756325	503894417
92	Photograph	Ed Sheeran	1525708	441132246
70	Look What You Made Me Do	Taylor Swift	335837	562562226
36	Scared to Be Lonely	Martin Garrix	1074560	866104216
13	Attention	Charlie Puth	560536	1112777364

Introduction to Modules

Math

math.py

```
def total(input):  
    total = 0  
    for num in input:  
        total += num  
    return total  
  
def exp(input):  
    return 2.718281**input
```

Script

script.py

```
import math  
  
l = [1,2,3]  
print(math.total(l))  
print(math.exp(3))
```

Output

```
6  
20.085536923187668
```

Introduction to Modules

Non-Modularized

test.py

```
l = [1,2,3,4,5,6,7,8,9,10]
total = 0
for num in l:
    total += num

length = 0
for num in l:
    length += 1

print(total)
print(length)
```

Modularized

test.py

```
l = [1,2,3,4,5,6,7,8,9,10]

def sum(input):
    total = 0
    for num in input:
        total += num
    return total

def length(input):
    count = 0
    for num in input:
        count += 1
    return count

print(sum(l))
print(length(l))
```

Importing functions

script.py

```
import test
t = [3,4,5,6]

test.sum(t)
test.length(t)
```

Local and Global Variables

script.py

```
l = [1,2,3,4,5,6,7,8,9,10]
```

```
def sum(input):
```

```
    total = 0
```

```
    for num in input:
```

```
        total += num
```

```
    return total
```

```
def length(input):
```

```
    count = 0
```

```
    for num in input:
```

```
        count += 1
```

```
    return count
```

```
print(sum(l))
```

```
print(length(l))
```



Accessible
Area

script.py

```
l = [1,2,3,4,5,6,7,8,9,10]
```

```
def sum(input):
```

```
    total = 0
```

```
    for num in input:
```

```
        total += num
```

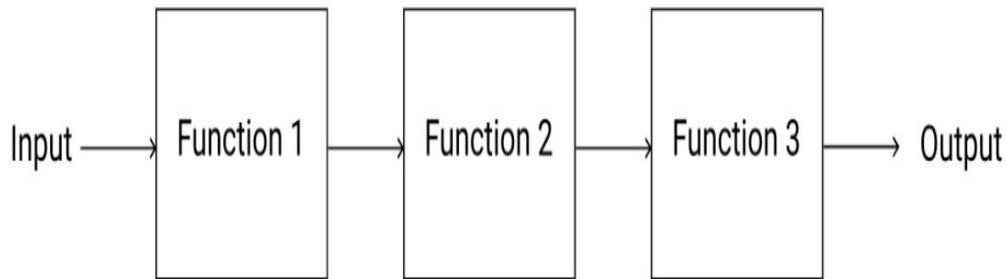
```
    return total
```

```
print(total)
```

Output

```
NameError: name 'total' is  
not defined
```


Using programming paradigms (functional)



```
import math
```

```
x = 7
```

```
def exp(x):  
    return math.exp(x)
```

```
def fraction(x):  
    return 1/x
```

```
x = exp(x)
```

```
x = fraction(x)
```

Lesson_03_part_1.ipynb

Sections 1.1 - 1.11



List Comprehension

```
streams = [57, 62, 63, 99, 142]
```

```
average = 84
```

```
diff = []
```

```
for num in streams:
```

```
    diff.append(num - average)
```

```
diff = [(num-average) for num in streams]
```

Getting the Artist Count Using a Function

```
# put your code here
def counter(names):
    dict_names = {}
    for i in names:
        if i in dict_names:
            dict_names[i] += 1
        else:
            dict_names[i] = 1
    return dict_names

counter(artists)
```

```
{'Ed Sheeran': 5,
 'Luis Fonsi': 2,
 'The Chainsmokers': 3,
 'Kendrick Lamar': 2,
 'French Montana': 1,
 'Post Malone': 3,
 'DJ Khaled': 2,
 'Kygo': 1,
 'Lil Uzi Vert': 1,
 'Bruno Mars': 2,
 'Dua Lipa': 1,
 'ZAYN': 2,
```

Getting the Artist Count Using a Function

Solution #1

```
dict_names = {}  
[dict_names.update({i:dict_names[i]+1}) if i in dict_names  
 else dict_names.update({i:1}) for i in artists]
```

Solution #2

```
counts_artists = {i:artists.count(i) for i in artists}
```

Getting the Artist Count Using a Function

```
from collections import Counter
```

```
artists = ['Ed Sheeran',  
           'Luis Fonsi',  
           'Luis Fonsi',  
           'The Chainsmokers',  
           'Kendrick Lamar']
```

```
count_artists = Counter(artists)  
count_artists
```

```
Counter({'Ed Sheeran': 1,  
         'Kendrick Lamar': 1,  
         'Luis Fonsi': 2,  
         'The Chainsmokers': 1})
```

Sorting a list of lists

```
artists_counts_lol.sort()  
artists_counts_lol
```

Alphabetical order



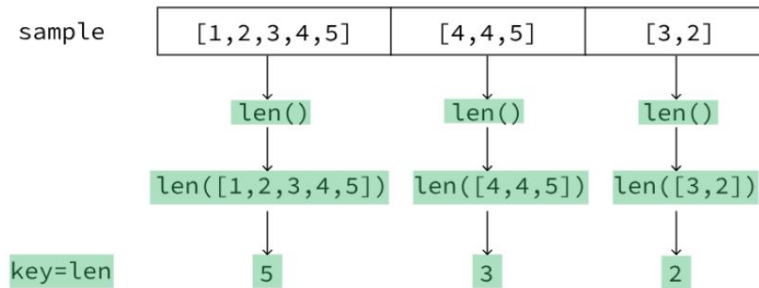
```
[['21 Savage', 1],  
 ['Alessia Cara', 1],  
 ['Avicii', 1],  
 ['Axwell /\ Ingrosso', 1],  
 ['Big Sean', 1],  
 ['Bruno Mars', 2],  
 ['CNCO', 1],  
 ['Calvin Harris', 2],  
 ['Camila Cabello', 1],  
 ['Cardi B', 1],
```

Customizing sort()

Code

```
sample.sort(key=len)
```

sample



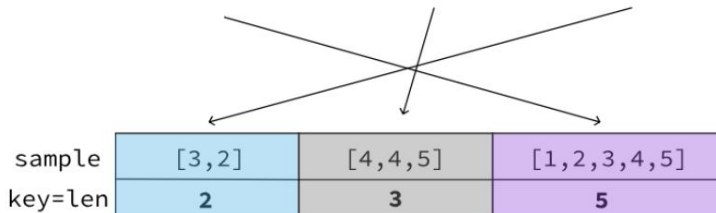
After calculating the length for each value, each value will be sorted:

Code

```
sample.sort(key=len)
```

sample
key=len

[1,2,3,4,5]	[4,4,5]	[3,2]
5	3	2



Creating a anonymous function

```
f = open("top100.csv", "r")
music = list(csv.reader(f))

artists = [row[1] for row in music[1:]]

artist_dict = Counter(artists)
artist_counts = [[key, value] for key, value in artist_dict.items()]

artist_counts.sort(key = lambda x: x[1], reverse=True)
```

How to deal with errors

```
In [1]: streams = [53, 33, 57, 21, "NULL", 47]
```

```
In [2]: total = 0
```

```
In [3]: for s in streams:
...:     total += s
...:
```

```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-3-9f5cd44a4b06> in <module>()
      1 for s in streams:
----> 2     total += s
      3
```

```
TypeError: unsupported operand type(s) for +=: 'int' and 'str'
```

How to deal with errors

```
>>> streams = [53, 33, 57, 21, "NULL", 47]
>>> total = 0
>>> for s in streams:
...     try:
...         total += s
...     except:
...         print("Error occurred")
Error occurred
>>> print(total)
211
>>> █
```

Lesson_03_part_1.ipynb

Sections 2.1 - 2.11

