

# CPU Bound Programs: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2019

## Syntax

- Finding duplicate values in columns:

```
DataFrame.duplicated()
```

- Seeing how long a piece of code takes to run:

```
import time

start = time.time()

# We're timing how long this line takes to run.

duplicates = []

elapsed = time.time() - start

print("Took {} seconds to run.".format(elapsed))
```

- Calling your program from the command line using the time command:

```
time python script.py
```

- Using the cProfile function to display timing statistics of a program or a line of code:

```
import cProfile

cProfile.run('print(10)')
```

## Concepts

- A memory limitation is when a data set won't fit into memory available on the computer.
- A program bound is similar to a limitation in that it affects how you're able to process your data. The two primary ways a program can be bound are:
  - CPU-bound — The program will be dependent on your CPU to execute quickly. The faster your processor is, the faster your program will be.
  - I/O-bound — The program will be dependent on external resources, like files on disk and network services to execute quickly. The faster these external resources can be accessed, the faster your program will run.

- The more efficient you make your code, the less back-and-forth trips will need to be made, and the faster your code will run.
- Big O notation expresses time complexity in terms of the length of the input variable, represented as  $n$ .
- Big O notation is a great way for estimating the time complexity of algorithms where:
  - You can easily trace all of the function calls, and understand any nested time complexity.
  - The code is relatively straightforward.
- The general process behind refactoring code is:
  - Measure how long the current code takes to run.
  - Rewriting the code so that the algorithm you want is nicely isolated from the rest of the code.
  - Rewriting the algorithm to reduce time complexity.
  - Benchmarking the new algorithm to see if it's faster.
  - Rinse and repeat as needed.
- Space complexity indicates how much additional space in memory our code uses over and above the input arguments.

## Resources

- [Documentation for pandas duplicated method](#)
- [Big O notation](#)
- [cProfile](#)
- [Unix time command](#)
- [Contexttimer](#)
- [lineprofiler](#)
- [Complexity Analysis Introduction](#)



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2019