# Introduction to
# Python For Data Science

**April, 2018**

GSA

D2D
DATA TO DECISIONS

# Questions?

**Functions**

**Methods**

**Packages and modules**

**NumPy**

GSA

D2D
DATA TO DECISIONS

# Checkpoint

- Some helpful resources:
  - ➢ List of all functions:
    - https://docs.python.org/3/library/functions.html
  - ➢ Official Python Docs:
    - https://docs.python.org/3/index.html
  - ➢ Free eBook : O'REILLY's A WhirlWind Tour of Python:
    - http://www.oreilly.com/programming/free/a-whirlwind-tour-of-python.csp
  - ➢ Python Beginners Guide Wiki:
    - https://wiki.python.org/moin/BeginnersGuide/Programmers

- Google (or bing)

GSA

D2D
DATA TO DECISIONS

## Exercise 1

**How many arguments does hex() require, and how many are optional?**

# Exercise 1

## How many arguments does hex() require, and how many are optional?

1 Required , 0 Optional (Exactly 1 argument)

```
help(hex)
Help on built-in function hex in module builtins:
hex(number, /)
    Return the hexadecimal representation of an integer.
hex(12648430)
    '0xc0ffee'

?hex
Signature: hex(number, /)
Docstring:
Return the hexadecimal representation of an integer.
hex(12648430)
'0xc0ffee'
Type:      builtin_function_or_method
```

GSA

D2D
DATA TO DECISIONS

## Exercise 2 (Code Along)

Write 1 python statement to print the max between:

- The length of a newly created sorted list of [1,4,6,8]
- The Integer value of the String '3'

(Expected output is 4)

GSA

D2D
DATA TO DECISIONS

## Exercise 2 (Code Along)

**Write 1 python statement to print the max between:**

- **The length of a newly created sorted list of [1,4,6,8]**
- **The Integer value of the String '3'**

**(Expected output is 4)**

```
print(max(list([len(sorted(list([1,4,6,8]))),int('3')])))
```

GSA

D2D
DATA TO DECISIONS

## Exercise 3

**Which of the following statements use functions and which use methods?**

- min([3,5,False])
- list([3,5,”True”]).index(True)
- x=7
  x.bit_length()

GSA

D2D
DATA TO DECISIONS

## Exercise 3

**Which of the following statements use functions and which use methods?**

- min([3,5,False]) ← Valid , Function Calls Only
- list([3,5,”True”]).index(True) ← Valid, Error, Function and Method Calls
- x=7 ← Valid, No calls, just assignment

  x.bit_length() ← Valid, Method Call Only

GSA

D2D
DATA TO DECISIONS

# Checkpoint

- Use the help() function for function … help!
- Functions take input parameters and return an object
- Input and return types don't have to be the same
- You can nest function calls
  - ➢ i.e. A function can be an argument to another function
  - ➢ As long as inner calls return a type compatible with outer call's input parameters

- Function: collection of code to perform a certain function:
  - ➢ Called explicitly with input parameters
  - ➢ Can optionally return data (otherwise returns None)
- Method: Collection of code to perform a certain function:
  - ➢ Called on a specific object
  - ➢ Can access other class attributes (Don't worry about that now)

- You can define your own Functions and Methods!

GSA

D2D
DATA TO DECISIONS

## Exercise 4 (Code Along)

**Write a function my sum that takes 2 integers and prints out their sum**


## Exercise 4 (Code Along)

**Write a function my sum that takes 2 integers and prints out their sum**

```
def mysum(x,y):
    print(x+y)

mysum(4,6)
```

**Questions:**

- **What happens if I pass 2 strings or 2 lists?**
- **What about mysum(3,mysum(7,3))?**
- **Does this function return anything?**

GSA

D2D
DATA TO DECISIONS

## Exercise 5 (Code Along)

**Write a function dategreeting(name) that takes a String name and returns a greeting followed by the date**

**Example: Hello John, Today is 2018-04-12**

GSA

D2D
DATA TO DECISIONS

## Exercise 5 (Code Along)

**Write a function dategreeting(name) that takes a String name and returns a greeting followed by the date**

**Example: Hello John, Today is 2018-04-12**

```
import datetime

def dategreeting(name):
    return "Hello "+name+", Today is " +str(datetime.date.today())

print(dategreeting("John"))
```

**Questions:**

- **What happens if I pass an integer?**
- **Does this function return anything?**

GSA

D2D
DATA TO DECISIONS

# Exercise 6 (Code Along)

## Use the mysum() function in another python module


# Exercise 6 (Code Along)

## Use the mysum() function in another python module

```
import mymodule

mysum(3,55)
```

# Checkpoint

- Functions are defined using the keyword **def**
  - ➢def addition_function(x,y):
- Values are returned using the **return** keyword (even if not present!)
  - ➢None value
- Functions take arguments
- A function can be an argument to another function
  - ➢addition_function(3,addition_function(3,5))
- No types are defined for arguments or return types
- Functions can call other functions
- Creating commonly used functions in modules is a good idea
- Creating methods is a future topic (need to create classes first!)
- **Be careful with scopes!**

GSA

D2D
DATA TO DECISIONS

# Checkpoint

- Modules and Packages provide a way of code reuse
- Python comes with a library of standard modules/packages
  - ➢Such as datetime
  - ➢…or the statistics module
    - Import statistics
    - print(statistics.mean([1,2,3,4,5,6]))
- A package is a collection of modules
- You can import an entire package, or a module within the package
  - ➢Import matplotlib
  - ➢Import matplotlib.pyplot
- Additional packages can be installed using **pip**
  - ➢***To install a new package:*** pip install < package_name >
  - ➢***To uninstall a package:*** pip uninstall < package_name >
  - ➢***To list all installed packages:*** pip list
  - ➢***To see information about a package:*** pip show <package_name>
- **Anaconda has another package management system**: conda

GSA

D2D
DATA TO DECISIONS

## Exercise 7 (Code Along)

Given the following 3 lists to represent room lengths, width and heights:

lengths=[10,21,32,45,7]

widths=[11,4,21,7,18]

heights=[10,10,10,10,8]

Calculate and print the volumes of all rooms

GSA

D2D
DATA TO DECISIONS

## Exercise 7 (Code Along)

**Given the following 3 lists to represent room lengths, width and heights:**

**lengths=[10,21,32,45,7]**

**widths=[11,4,21,7,18]**

**heights=[10,10,10,10,8]**

**Calculate and print the volumes of all rooms**

```
import numpy as np
Lengths,widths,heights=[10,21,32,45,7], [11,4,21,7,18],[10,10,10,10,8]
lengthsarray=np.array(lengths)
widthssarray=np.array(widths)
heightssarray=np.array(heights)
volumes=lengthsarray*widthssarray*heightssarray
print(volumes)
```

GSA

D2D
DATA TO DECISIONS

# Exercise 8

## What is the output of each of the following expressions:

- `np.array([1,2])+np.array([3,4])`
- `list(np.array([1,2]))+list(np.array([3,4]))`
- `np.array([1,4]) + np.array([1,4,5])`
- `np.array([1,4])*4`
- `np.array([True,True])+np.array([3,4])`
- `np.array(["False",True])+np.array([True,False])`

GSA

D2D
DATA TO DECISIONS

## Exercise 8

**What is the output of each of the following expressions:**

- `np.array([1,2])+np.array([3,4])` ← `[3,6]`
- `list(np.array([1,2]))+list(np.array([3,4]))` ← `[1,2,3,4]`
- `np.array([1,4]) + np.array([1,4,5])` ← `Error`
- `np.array([1,4])*4` ← `[4,16]`
- `np.array([True,True])+np.array([3,4])` ← `[4,5]`
- `np.array(["False",True])+np.array([True,False])` ← `Error`

GSA

D2D
DATA TO DECISIONS

# Exercise 9

**Create a 100x100 Array of random numbers between 0 and 1, then use the array get the following values:**

- **Max first column**
- **Min last column**
- **mean first row**
- **median last row**
- **standard deviation first row + last row**
- **print the middle 4 elements**

GSA

D2D
DATA TO DECISIONS

# Exercise 9

**Create a 100x100 Array of random numbers between 0 and 1, then use the array get the following values:**

- **Max first column**
- **Min last column**
- **mean first row**
- **median last row**
- **standard deviation first row + last row**
- **print the middle 4 elements**

```
import numpy as np
array = np.random.rand(100,100)
print(np.max(array[:,0]))
print(np.min(array[:,-1]))
print(np.mean(array[0,:]))
print(np.median(array[-1,:]))
print(np.std(array[0,:]+array[-1,:]))
print(array[50:52,50:52])
```

GSA

D2D
DATA TO DECISIONS

## Checkpoint

- NumPy : Numeric Python (Useful for Numeric operations)
- Can be sliced like lists
- Allow for fast and efficient Arithmatic
- Don’t confuse operators for NumPy Arrays with Lists
- Has some statistical Functions
- Find a list of values that meet a condition with Operators
  - ➢ array>0.5
  - ➢ array[array>0.5]← Boolean array
  - ➢ Can also select elements by passing Boolean array
    - array = np.array([1,2,3,4,5])
      print(array[[True, False, True, True, False]])
- nan and inf
  - ➢ nan: not a number
  - ➢ Inf: infinity

GSA

D2D
DATA TO DECISIONS

# Exercise 10

**What is the output of each of the following expressions:**

- `print(np.nan+1)`
- `print(-1*np.inf)`
- `print(np.nan == np.nan)`
- `print(np.inf > np.nan)`
- `print(np.nan - np.nan)`
- `print(np.inf > np.nan)`
- `print(np.inf == np.inf)`

GSA

D2D
DATA TO DECISIONS

# Exercise 10

## What is the output of each of the following expressions:

- `print(np.nan+1)` ← `nan`
- `print(-1*np.inf)` ← `-inf`
- `print(np.nan == np.nan)` ← `False`
- `print(np.inf > np.nan)` ← `False`
- `print(np.nan - np.nan)` ← `nan`
- `print(np.inf > np.nan)` ← `False`
- `print(np.inf == np.inf)` ← `True`

GSA

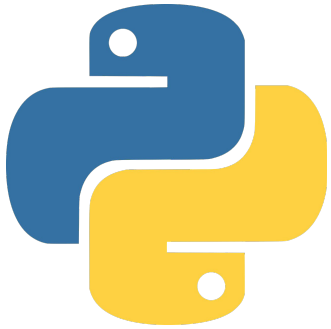
D2D
DATA TO DECISIONS

# Questions?

GSA

D2D
DATA TO DECISIONS

# Thank You



GSA

D2D
DATA TO DECISIONS