



# Welcome to: **STATS R FUN #1!!!**

Intro to Python + What is Data?

# What is Python ???

```
print("Hello World")
```

Commands



Hello World

Outcome



# Basic Things We Can Do in Python

Set Variables

```
x = 10  
x
```



```
10
```

Do Basic Math

```
2 + 2
```



```
4
```

# What is Data?

Data is EVERYTHING !!!

We can see this in surveys, a bunch of numbers, etc.

BUT... we must differentiate TYPES of data

We could hypothetically try doing things like `2 + Hello` in Python... so let's try it 🐱🐱

```
-----  
TypeError                                Traceback (most recent call last)  
<ipython-input-1-45c169148032> in <cell line: 1>()  
----> 1 1 + "Hello"  
  
TypeError: unsupported operand type(s) for +: 'int' and 'str'
```



!?

# So...What are Types?

Type	Name in Python	Description	Examples
Integers	'int'	Whole numbers	-1, 0, 1, 2, 3...
Floats	'float'	Decimals	2.5, 6.125, <u>1.0</u>
Strings	'str'	Letters, words, and sentences (or anything surrounded by quotes)	"A", "Hello!", <u>"12345"</u>
Booleans	'bool'	True/False	True, False
Lists	'list'	Collections of items surrounded by brackets []	[1, 2, 3] ["Hi", 1.5] [[1, 2, 3], "HI"]
Dictionaries	'dict'	Collections of items assigned to labels, called "keys"	{"Colors": ["red", "blue"], "Numbers": [1,2,3]}

# So...What are Types?

## Lists ('list'):

Collections of items surrounded by brackets []

```
my_list = [1, "Hello", 2.125]
```

## Dictionaries ('dict'):

Collections of items assigned to labels, called "keys". Surrounded by curly brackets {}

```
my_dict = {  
    'colors': ['red', 'blue', 'green'],  
    'numbers': [1, 2, 3],  
    'favorite food': 'pizza'  
}
```

# Indexing

**Indexing** means to retrieve an element from a collection (such as lists or dictionaries)

Indexing from a **List**:

```
my_list = [1, "Hello", 2.125]
```

```
my_list[0]
```



```
1
```

```
my_list[2]
```



```
2.125
```



# Indexing

**Indexing** means to retrieve an element from a collection (such as lists or dictionaries)

Indexing from a **Dictionary**:

```
my_dict = {  
    'colors': ['red', 'blue', 'green'],  
    'numbers': [1, 2, 3],  
    'favorite food': 'pizza'  
}
```

`my_dict['colors']`



`['red', 'blue', 'green']`

`my_dict['colors'][0]`



`'red'`

# Importing Data Into Python

```
nlsq_csv_file_sample_2021_04_08.csv - Notepad
File Edit Format View Help
Keyword,Min Monthly Volume,Max Monthly Volume,Specific Monthly Volume,Difficulty,Rank
perfect tower 2 improve chance to find modules,0,10,10.5370846,1,33
straight-forward analytics solution,0,10,,1,34
bi-soft,11,50,16.71429414,1,42
vantagens self service,11,50,27.24240562,1,46
turn key analytics,0,10,,2,26
southwest virtual agents,0,10,,3,34
coming out bi 7 watch online free,11,50,27.24240562,3,36
coming out bi 7 watch online,11,50,27.24240562,3,49
power music login,101,200,121.5019653,4,18
bi rapportages,51,100,58,4,38
legal dashboard,0,10,1,5,18
caso de exito big data netflix,11,50,27.24240562,5,21
consultora de bi,0,10,,5,32
power bi consulting london,0,10,,6,15
adobe analytics danismanlik,0,10,,6,16
spend analytics dashboard,0,10,,6,18
syscon online login,11,50,38.68984882,6,23
spend analysis dashboard,11,50,27.24240562,6,24
escalabilidade em saas,0,10,,6,27
usa ge cnovember 2019,0,10,10.5370846,6,34
Ln 6, Col 1 100% Windows (CRLF) UTF-8
```



Author	User Rating	Reviews	Price	Year	Genre
JJ Smith	4.7	17350	8	2016	Non Fiction
Stephen King	4.6	2052	22	2011	Fiction
Jordan B. Peterson	4.7	18979	15	2018	Non Fiction
George Orwell	4.7	21424	6	2017	Fiction
National Geographic Kids	4.8	7665	12	2019	Non Fiction
George R. R. Martin	4.4	12643	11	2011	Fiction
George R. R. Martin	4.7	19735	30	2014	Fiction
Amor Towles	4.7	19699	15	2017	Fiction
James Comey	4.7	5983	3	2018	Non Fiction
Fredrik Backman	4.6	23848	8	2016	Fiction
Fredrik Backman	4.6	23848	8	2017	Fiction
Larry Schweikart	4.6	460	2	2010	Non Fiction

# Importing Data Into Python

Author	User Rating	Reviews	Price	Year	Genre
JJ Smith	4.7	17350	8	2016	Non Fiction
Stephen King	4.6	2052	22	2011	Fiction
Jordan B. Peterson	4.7	18979	15	2018	Non Fiction
George Orwell	4.7	21424	6	2017	Fiction
National Geographic Kids	4.8	7665	12	2019	Non Fiction
George R. R. Martin	4.4	12643	11	2011	Fiction
George R. R. Martin	4.7	19735	30	2014	Fiction
Amor Towles	4.7	19699	15	2017	Fiction
James Comey	4.7	5983	3	2018	Non Fiction
Fredrik Backman	4.6	23848	8	2016	Fiction
Fredrik Backman	4.6	23848	8	2017	Fiction
Larry Schweikart	4.6	460	2	2010	Non Fiction

We usually have **tidy data**, this means:

- Each row is a new observation
- Each column is a feature/variable