# Python for DataScience

Everyday is a NEW #BEGINNING

# Why do we Live ?

In Early 90's
-We live in planet called Earth …

# Why do we Live ?

In Early 21’st Century
-We live in
**Data driven**
planet called Earth

“Without data you’re just another person with an opinion.

# Who Am I ?

I’m not progrmmer by birth. I learned it.
I have no close relative who is programmer
-Dedication, Practice , Self-motivation

HELLO!

# I am Santosh Purbey

CTO – BitsInnovation

I am here because I love Python

- Data Anylistics
- Django
- Machine Learning
- Data Scientist
- IOT
- Cloud Computing
- Full stack Developer

# Python tools for data scientist

scikit learn

machine learning in Python

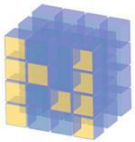
matplotlib

scikit-image

image processing in python

PyBrain

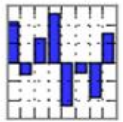
The Python Machine Learning Library

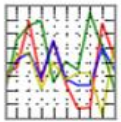


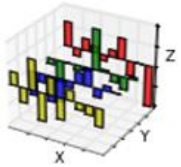
NumPy

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$








scipy

scrapy

# What is Data Scientist

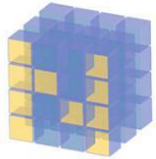
A Person who is better at Statistics than any software engineer and better at software engineering than any Statistician

# Why Data Science


The Sexiext job of the 21’st Century

# Why Data Science

The Sexiext job of the  21’st Century

# Pandas =  +  NumPy + 

-Python

Easy Syntex,

Helpful community

-Numpy

Fast, Memory Efficient Calc

well tested algorithms

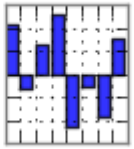
-R

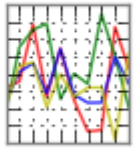
DataFrame Column labels

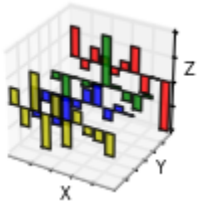
Indexes to align rows

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$








http://pandas.pydata.org

```
>>> import pandas as pd
#To read CSV File
>>> csvObj = pd.read_csv("file.csv")
>>> len(pd.DataFrame)
        425
```

# Pandas Data Structures : Series

| Index | Values |
|---|---|
| 0 | 3 |
| 1 | 5 |
| 2 | 11 |
| 3 | -1 |
| 4 | 4.6 |

One Dimentional array like object containing data and index(or labels)

```
>>> Import pandas as pd
>>> s= pd.Series([3, 5, 11, -1, 4.6])
```

# Pandas Data Structures : Series

| Index | Values |
|---|---|
| 0 | a |
| 1 | b |
| 2 | c |
| 3 | d |
| 4 | e |

Lots of ways to build Series

```
>>> Import pandas as pd
>>> s= pd.Series(list('abcde'))
```

# Pandas Data Structures : Series

| Index | Values |
|---|---|
| A | 1 |
| B | 4 |
| C | 9 |
| D | -4 |
| E | 4.6 |

A series index can be specified

```
>>> Import pandas as pd
>>> s= pd.Series([1, 4, 9, -4, 4.6],
index=['A', 'B', 'C', 'D', 'E'])

>>>s['A']
1
>>>s['A', 'D']
A       1
D       -4
```

# Pandas Data Structures : DataFrame

Spreadsheet-like data Structure
Has both row and column
Consider as dict of Series

```
>>>data = {'country':['Nepal', 'India', 'China'],
          'Population':[22224444, 2111111, 2222222],
          'Year':[1991, 1991, 1991,],
          'Growth':[2.6, 14.8, 11.5]
          }
>>>frame = pd.DataFrame(data)
>>>frame
```

| | Country | Population | Year | Growth |
|---|---|---|---|---|
| 0 | Nepal | 22,224,444 | 1991 | 2.6 |
| 1 | India | 2,111,1111 | 1991 | 14.8 |
| 2 | China | 222,222,222 | 1991 | 11.5 |

# “Life is short Use Python”

Do not waste decade to learn Programming

**THANKS!**

# Any questions?

You can find me at @ fb.com/sp3411 &
Email: purbey.santoh@hotmail.com
Github: @santoshpurbey
Mobile: 9840061616