



Women in Data 2019
Introduction to Python

Karina Marks, Data Scientist



www.linkedin.com/in/karinajmarks



kmarks@mango-solutions.com



What to expect?



Who am I / are we?



Why Python?



What's possible with Python?



Python code examples



How to learn Python?

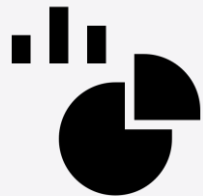
About me and Mango



About me



Mathematics (MMath) Cardiff University



Created by Fauzan Adima
from Noun Project

6 + years Python

4+ years R

Intern at Mango (2016)

2+ years Data Scientist at Mango

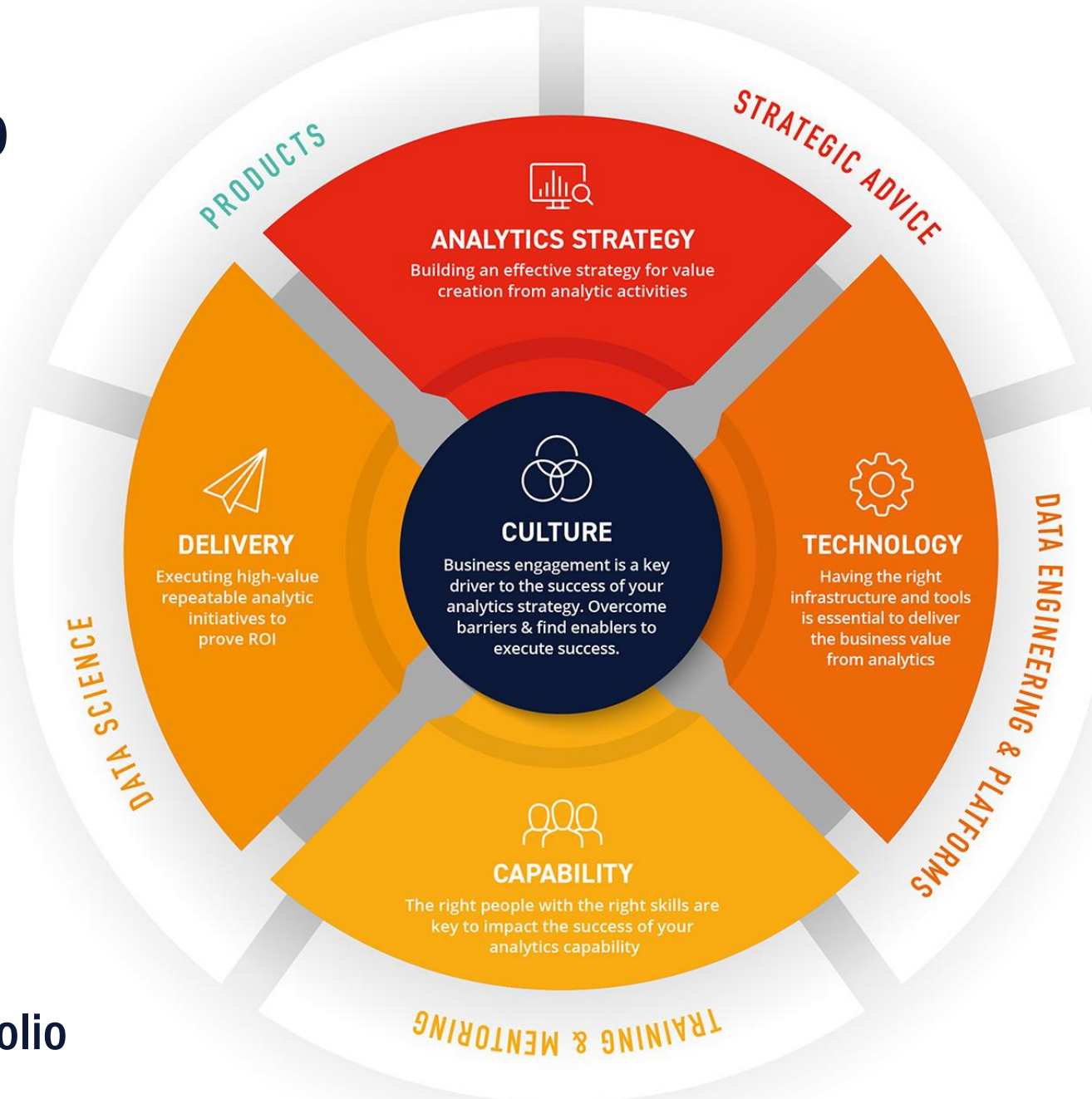


Mango is a trusted partner to enterprises that seek to become truly data-driven by embedding data science into decision making – we help them thrive

About Mango



Mango's products and services are aligned to data-driven transformation



INNER WHEEL :

Data-driven pillars

OUTER WHEEL :

Supported by Mango's product and services portfolio



Why Python?



Why Python?

Python Principles

The Zen of Python

Beautiful is better than ugly.

Explicit is better than implicit.

Simple is better than complex.

Complex is better than complicated.

Flat is better than nested.

Sparse is better than dense.

Readability counts.

Special cases aren't special enough to break the rules.

Although practicality beats purity.

Errors should never pass silently.

Unless explicitly silenced.

In the face of ambiguity, refuse the temptation to guess.

There should be one-- and preferably only one --obvious way to do it.

Although that way may not be obvious at first unless you're Dutch.

Now is better than never.

Although never is often better than **right** now.

If the implementation is hard to explain, it's a bad idea.

If the implementation is easy to explain, it may be a good idea.

Namespaces are one honking great idea -- let's do more of those!



ANACONDA®

Why Python?

Easy to Install



ANACONDA DISTRIBUTION

Most Trusted Distribution for Data Science

ANACONDA NAVIGATOR

Desktop Portal to Data Science

ANACONDA PROJECT

Portable Data Science Encapsulation

DATA SCIENCE LIBRARIES

Data Science IDEs



Analytics & Scientific Computing



Visualization



Machine Learning



...and many more!

CONDA®

Data Science Package & Environment Manager



Why Python?

Connects with Other Languages



Why Python?

Documentation & Community

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

home // about // get pandas // documentation // community // talks // donate

Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

pandas is a [NumFOCUS](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

A Fiscally Sponsored Project of
NI IM FOCUS

VERSIONS

- Release 0.25.3 - November 2019
[download](#) // [docs](#) // [pdf](#)
- Development 1.0.0 - September 2019
[github](#) // [docs](#)
- Previous Releases
0.25.2 - [download](#) // [docs](#) // [pdf](#)
0.25.1 - [download](#) // [docs](#) // [pdf](#)
0.25.0 - [download](#) // [docs](#) // [pdf](#)

seaborn 0.9.0 Gallery Tutorial API Site Page Search

seaborn: statistical data visualization

Seaborn is a Python data visualization library based on [matplotlib](#). It provides a high-level interface for drawing attractive and informative statistical graphics.

For a brief introduction to the ideas behind the library, you can read the [introductory notes](#). Visit the [installation page](#) to see how you can download the package. You can browse the [example gallery](#) to see what you can do with Seaborn, and then check out the [tutorial](#) and [API reference](#) to find out

Contents

- Introduction
- Release notes
- Installing
- Example gallery
- Tutorial

Features

- Relational: [API](#) | [Tutorial](#)
- Categorical: [API](#) | [Tutorial](#)
- Distributions: [API](#) | [Tutorial](#)
- Regressions: [API](#) | [Tutorial](#)

matplotlib

Version 3.1.1

Installation Documentation Examples Tutorials Contributing

home | contents | modules | index

matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and [iPython](#) shells, the [Jupyter](#) notebook, web application servers, and four graphical user interface toolkits.

Matplotlib 3.0 is Python 3 only. For Python 2 support, Matplotlib 2.2.x will be continued as a LTS release and updated with bugfixes until January 1, 2020.

Support Matplotlib

SciPy.org

SciPy (pronounced "Sigh Pie") is a Python-based ecosystem of open-source software for mathematics, science, and engineering. In particular, these are some of the core packages:

- NumPy: Base N-dimensional array package
- SciPy library: Fundamental library for scientific computing
- Matplotlib: Comprehensive 2D Plotting
- IPython: Enhanced Interactive Console
- Sympy: Symbolic mathematics
- pandas: Data structures & analysis

NUMFOCUS Large parts of the SciPy ecosystem (including all six projects above) are fiscally sponsored by NUMFOCUS.

CORE PACKAGES:
NumPy | SciPy library | Matplotlib

python

Python 3: Fibonacci series up to n

```
>>> def fib(n):
>>>     a, b = 0, 1
>>>     while a < n:
>>>         print(a, end=' ')
>>>         a, b = b, a+b
>>>     print()
>>> fib(1000)
```

0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987

Functions Defined

The core of extensible programming is defining functions. Python allows mandatory and optional arguments, keyword arguments, and even arbitrary argument lists. [More about defining functions in Python 3](#)

Python is a programming language that lets you work quickly and integrate systems more effectively. [Learn More](#)

NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the [BSD license](#), enabling reuse with few restrictions.

Getting Started

To install NumPy, we strongly recommend using a *scientific Python distribution*. See [Installing the SciPy Stack](#) for details.

Many high quality online tutorials, courses, and books are available to get started with NumPy. For a quick introduction to NumPy we provide the [NumPy Tutorial](#). We also recommend the [SciPy Lecture Notes](#) for a broader introduction to the scientific Python ecosystem.

For more information on the SciPy Stack (for which NumPy provides the fundamental array data structure), see [c10v.000](#)

About PyData

An educational program of NumFOCUS, a 501(c)(3) nonprofit charity



PyData Mission

PyData is a community of data analysis tools to share the mission of best practices, new techniques, and visualization. PyData is limited to Python, Julia, and R.



Explore Meetup

Find events hosted by local groups where you can meet new people, try something new, or just do more of what you love.

Get started

Python within 50 miles of London, G

London Python
6,216 Pythonistas

London Python Spring
1,767 Pythonistas

London Python Amateurs
842 Members

Oxford Python
917 Pythonistas



Home

PUBLIC

Stack Overflow

Tags

Users

Jobs

TEAMS What's this?

First 25 Users Free

Products Customers Use cases

[python]

Questions tagged [python]

Ask Question

Python is a multi-paradigm, dynamically typed, multipurpose programming language, designed to be quick (to learn, to use, and to understand), and to enforce a clean and uniform syntax. Two similar but incompatible versions of Python are commonly in use, Python 2.7 and 3.x. For version-specific Python questions, add the [python-2.7] or [python-3.x] tag. When using a Python variant or library (e.g. Jython, PyPy, Pandas, Numpy), please include it in the tags.

Learn more... Top users Synonyms (4) python jobs

1,290,704 questions

Newest Active Bountied 42 Unanswered More Filter

9634 votes
38 answers
2.1m views

What does the "yield" keyword do?
What is the use of the yield keyword in Python? What does it do? For example, I'm trying to understand this code: `def _get_child_candidates(self, distance, min_dist, max_dist): if self...`

python iterator generator yield coroutine

asked Oct 23 '08 at 22:21
Alex. S. 109k 15 47 61

5569 votes
23 answers
1.8m views

Does Python have a ternary conditional operator?
If Python does not have a ternary conditional operator, is it possible to simulate one using other language constructs?

python operators ternary-operator conditional-operator

community wiki
20 revs, 16 users 43% Devoted

5514 votes
29 answers
2.6m views

What does if __name__ == "__main__": do?
What does the `if __name__ == "__main__":` do? # Threading example `import time, thread def myfunction(string, sleeptime, lock, *args): while True: lock.acquire() time.sleep(...)`

python namespaces main python-module idioms

asked Jan 7 '09 at 4:11
Devoted 76.1k 39 79 109

5394 votes

What are metaclasses in Python?
What are metaclasses and what do we use them for?

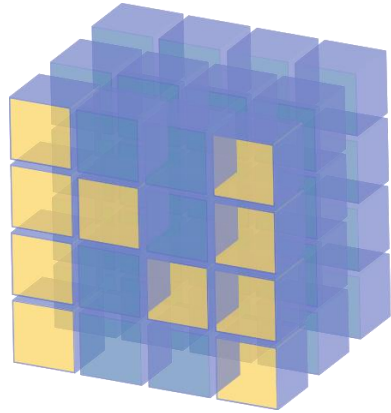
Why Python?

Documentation & Community

**What's possible with
Python?**



Python in the World of Analytics

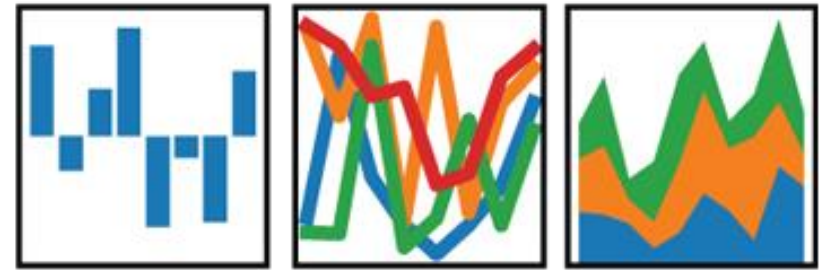


NumPy

pandas



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



matplotlib



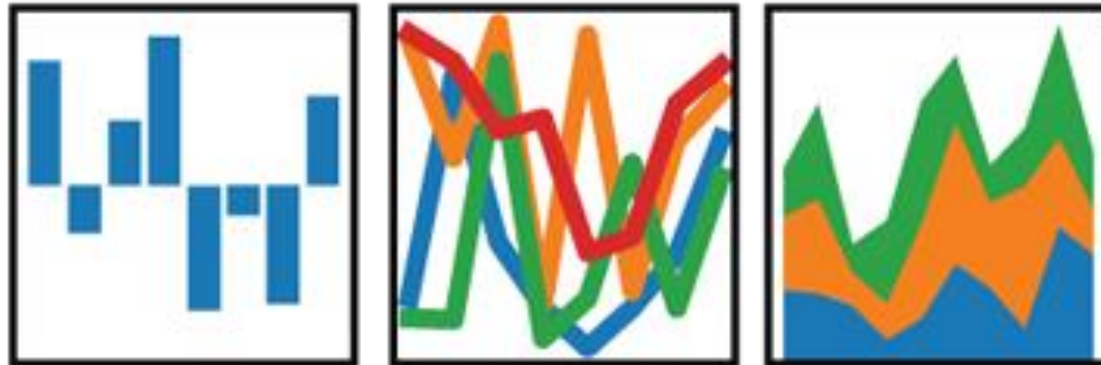
SciPy

Python in the World of Analytics



pandas

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



Python in the World of ML

Machine Learning, Deep Learning and Artificial Intelligence



Keras

A deep learning library



TensorFlow



PYTORCH

Deep Learning with PyTorch

**Lets Learn Some
Python!**



Data Import





Data Wrangling



country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M
C	1999	212K/1T
C	2000	213K/1T

country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172
B	2000	80K	174
C	1999	212K	1T
C	2000	213K	1T

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



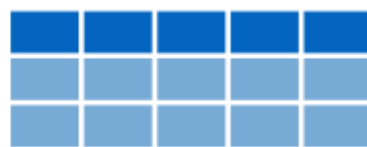
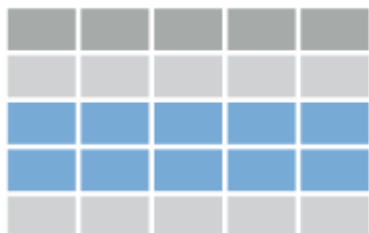
country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K



country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T



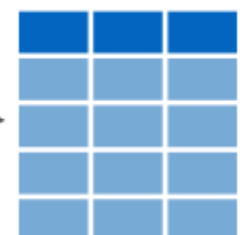
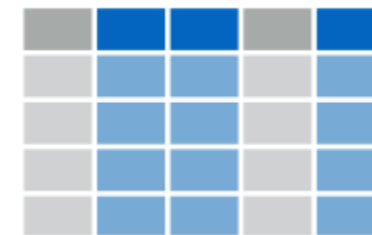
country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T



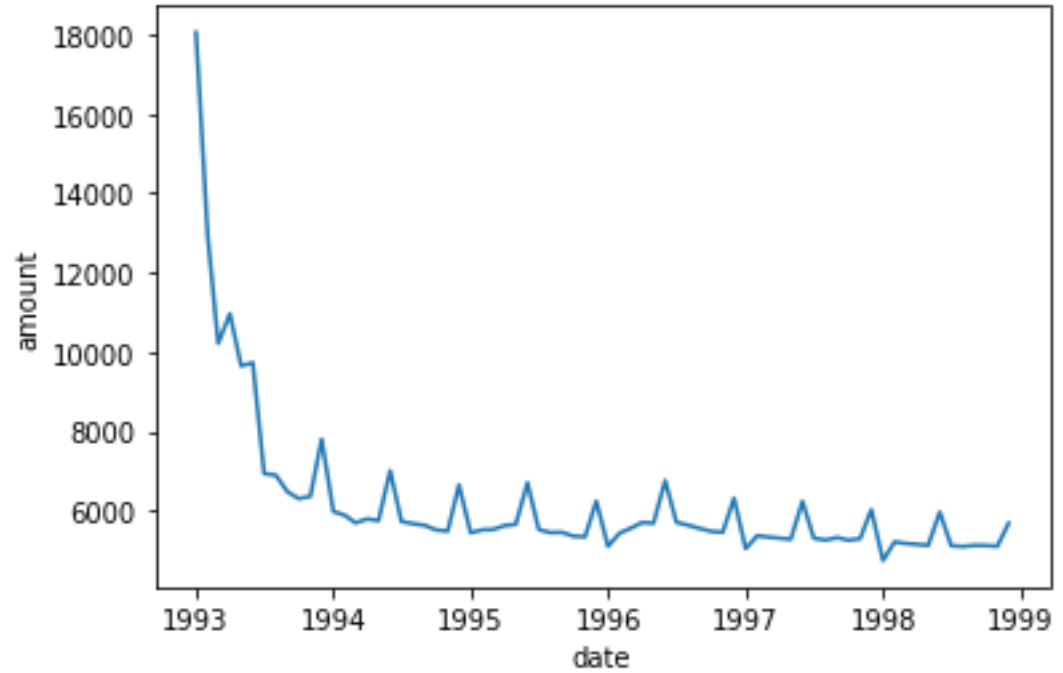
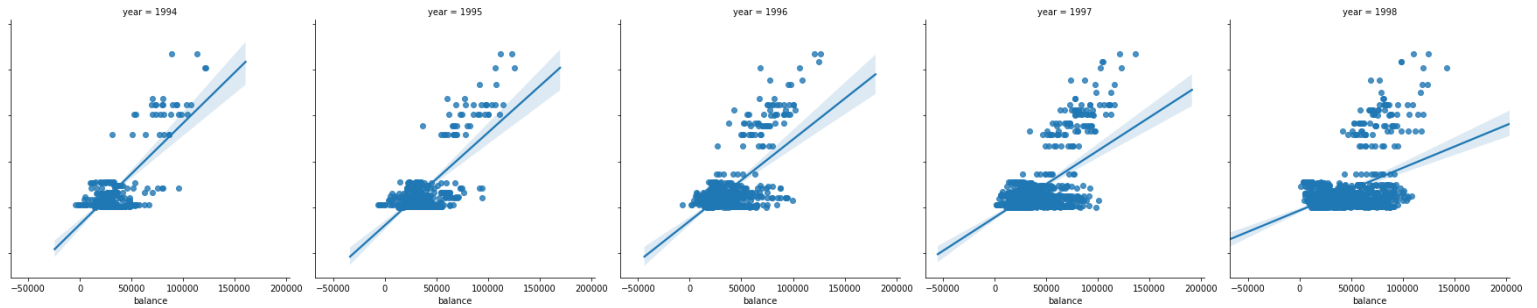
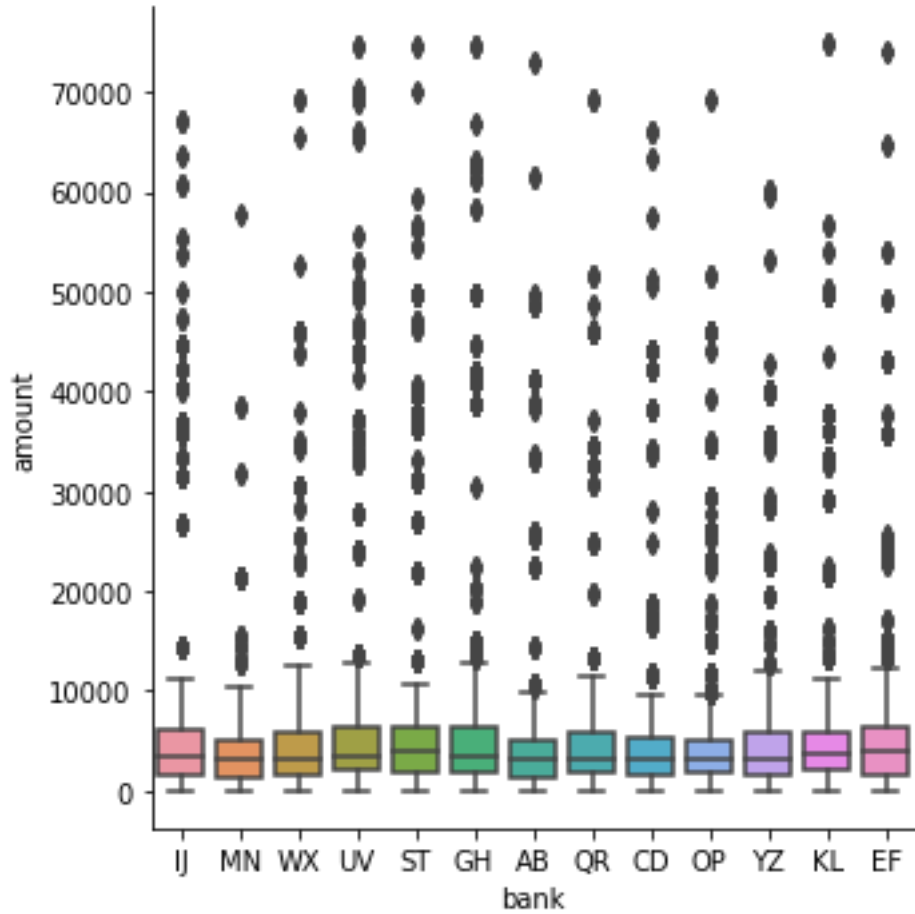
country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M
C	1999	212K/1T
C	2000	213K/1T



country	year	rate
A	1999	0.7K
A	1999	19M
A	2000	2K
A	2000	20M
B	1999	37K
B	1999	172M
B	2000	80K
B	2000	174M
C	1999	212K
C	1999	1T
C	2000	213K
C	2000	1T



Data Visualisation

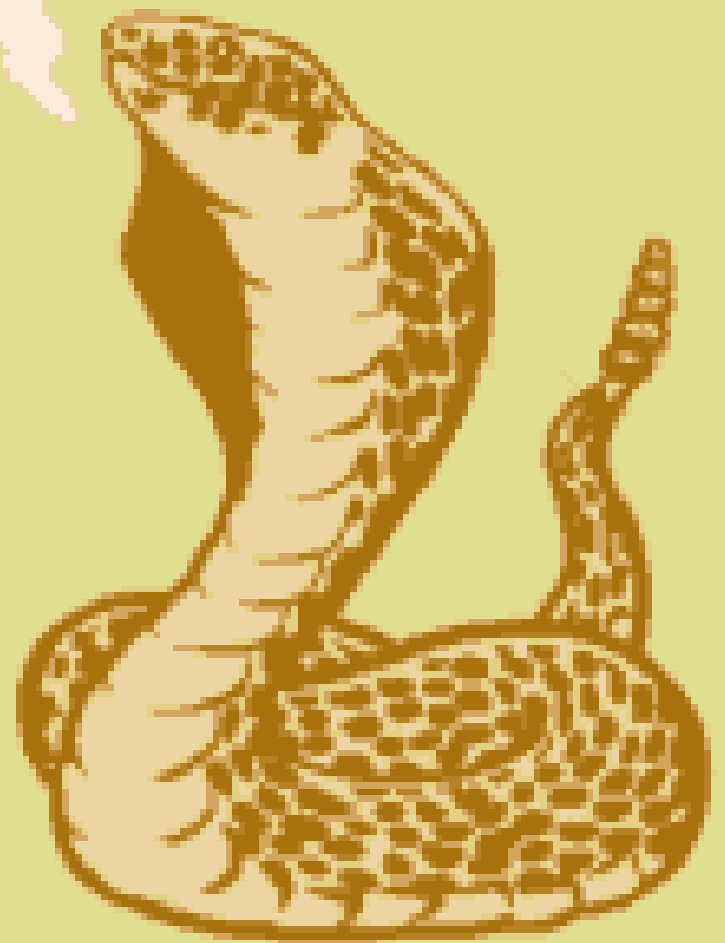


Statistics / Machine Learning



**You've Learnt Some
Python!**

YA





How to Learn More Python?

Free resources

- <https://www.learnpython.org/>
- Google 'Learn Python for free'
- Blog posts
- YouTube
- Meetups (PyData, Python User Groups)
- Stackoverflow



Take-home message

- There are lots of good reasons to learn Python
- Python is easy to read, broad in its applications and powerful in its performance
- There are lots of free resources out there
- Nothing should stop you from starting to use it NOW!



Questions?





We empower organisations to make **informed decisions**, using advanced **analytics** and **AI/ML** techniques to meet their objectives and deliver **data-driven value**. We do this through our unique combination of people, values and strategic priorities.



@MangoTheCat



www.mango-solutions.com



info@mango-solutions.com



+44 1249 705450

Chippenham Office

Mango Solutions
2 Methuen Park
Chippenham
SN14 0GB

London Office

Dawson House
5 Jewry Street
London
EC3N 2EX