

# PyData入門編

## Pythonデータ活用術

(Pythonエンジニア養成読本 第4章 +  $\alpha$ )

2015/05/30

PyData.Okinawa  
Masaaki Saito

# 自己紹介

👉 さいとう まさあき

👉 沖縄 読谷村

👉 Perl好き (Pythonも好きです！)

👉 PyData.Okinawa 共同主催者

👉 Yomitan.pm (Perl Mongers Group) 主催者

👉 Twitter @masakyst

# アジェンダ

- ① PyDataとは
- ② ツール、モジュールの初歩的な解説
- ③ オープンデータを利用した分析事例

# 1. PyDataとは？

データ分析の需要の高まりとともに、Python+Dataを組み合わせた「PyData」と称される領域に注目が集まっています。

データ加工や集計をはじめ、数値計算、画像処理、自然言語処理など、データを様々な方法で扱い活用するためのツールとしてPythonを利用するのが、PyDataです。

## 2. ツール、モジュールの初歩的な解説

IPython, Jupyter, NumPy, Pandas, Matplotlibのデモ

👉 pydata\_chapter4.ipynb

👉 pandas sample.ipynb

👉 matplotlib sample.ipynb

# 今回の実行環境について

👉 Python 2.7系 推奨

👉 Anaconda 推奨

👉 Jupyter 推奨

Anaconda

<https://store.continuum.io/cshop/anaconda/>

### 3. オープンデータを利用した分析事例

養成読本4-4 実践編のサンプル

利用するデータは、神奈川県横浜市のオープンデータ

【推計人口】 男女別人口及び世帯数\_\_行政区  
です。

<http://www.city.yokohama.lg.jp/ex/stat/opendata/suikei01.html>

このデータは、クリエイティブ コモンズ ライセンスの元で  
提供されています。

 opendata sample.ipynb

# $\alpha$ . MovieLens 映画評価データ分析事例

MovieLensとは、映画評価システムの1つで、ミソネタ大学のコンピュータ理工学研究科のGroupLens研究所が運営を続けています。

今回の評価データは、1990年代後半から2000年代前半にかけて、MovieLensユーザを対象に収集されてきました。

4000本の映画に対する6000人の評価をまとめたものを利用します

<http://grouplens.org/datasets/movielens/#attachments>