
# Introduction to Python for Scientists / Graduate Students

CSC298(five weeks only), CSC315 and DSC615

# Who invented Python?

Guido van Rossum

https://gvanrossum.github.io/
https://www.linkedin.com/in/guido-van-rossum-4a0756

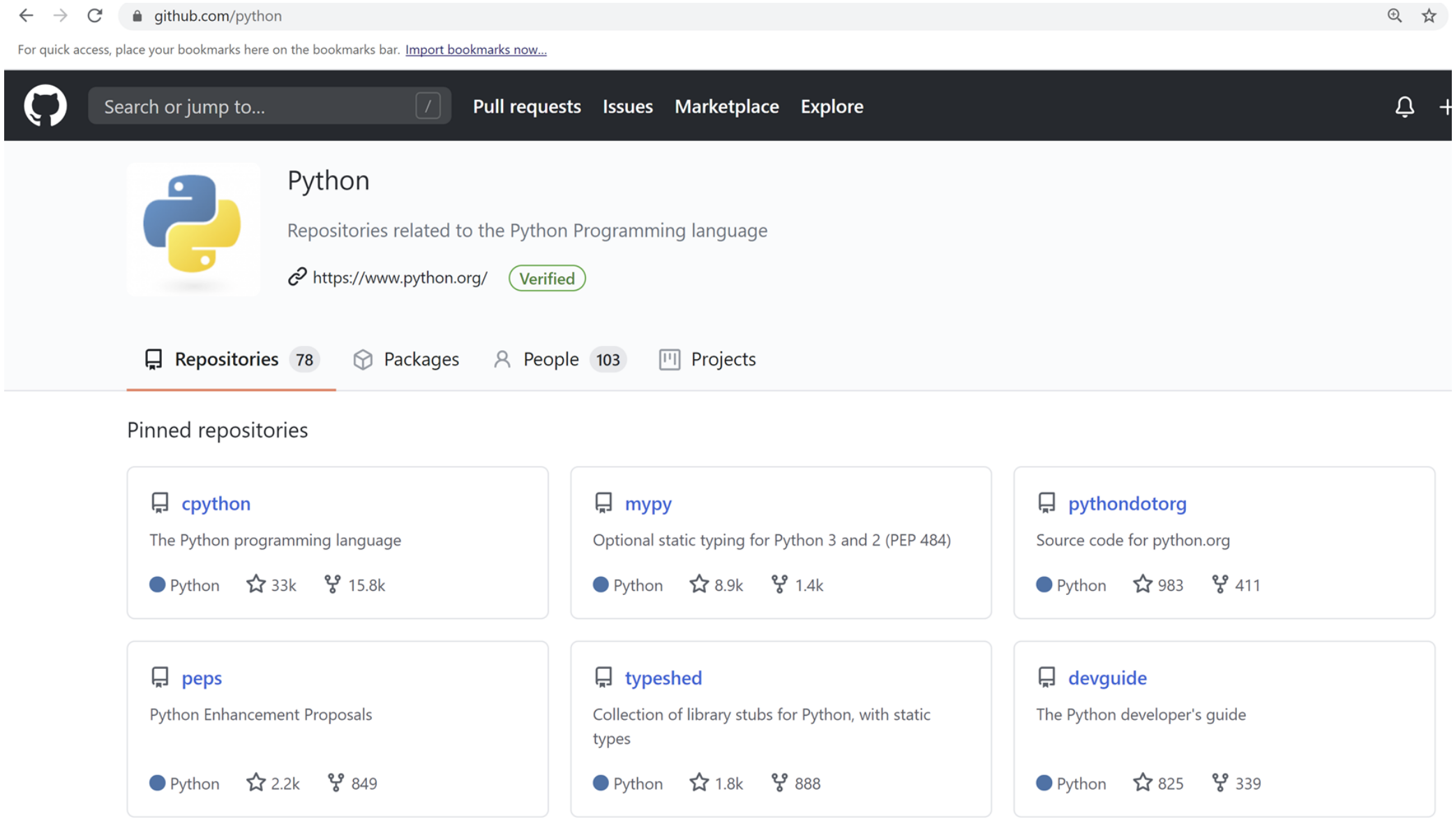
- He was born and raised in the Netherlands.
- He created Python in 1989 while working at a company in the Netherlands.
  https://www.python.org/doc/essays/foreword/

  " in December 1989, I was looking for a "hobby" programming project that would keep me occupied during the week around Christmas. "

- He worked for dropbox and retired.

https://en.wikipedia.org/wiki/Guido_van_Rossum

# Python is developed and maintained by a group of people through GitHub: https://github.com/python

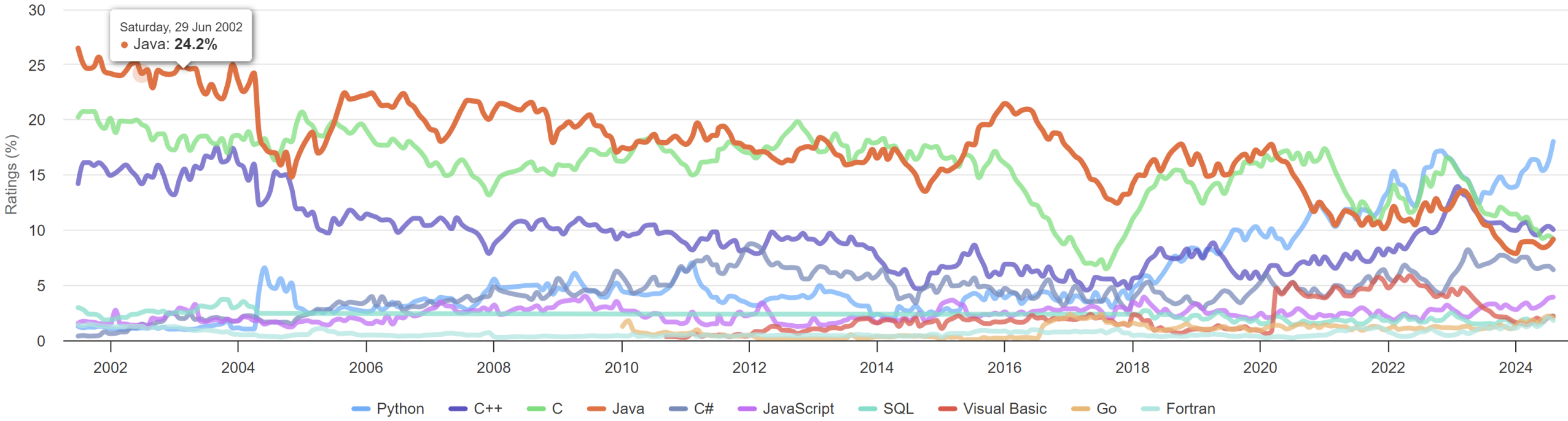


# Is Python a popular programming language?

TIOBE programming community index is a measure of popularity of programming languages.
The index is calculated from the number of search engine queries containing the name of the language.

https://www.tiobe.com/tiobe-index/

(August 2024)



# Which companies use Python?

https://wiki.python.org/moin/OrganizationsUsingPython




https://research.fb.com


https://aws.amazon.com/mxnet/


https://www.quora.com

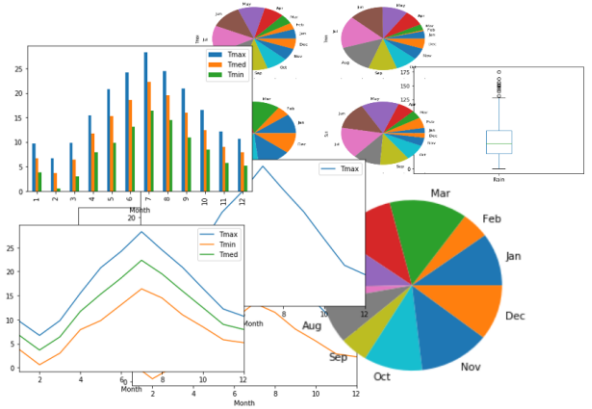
# What can Python be used for ?

Python is used for countless applications …

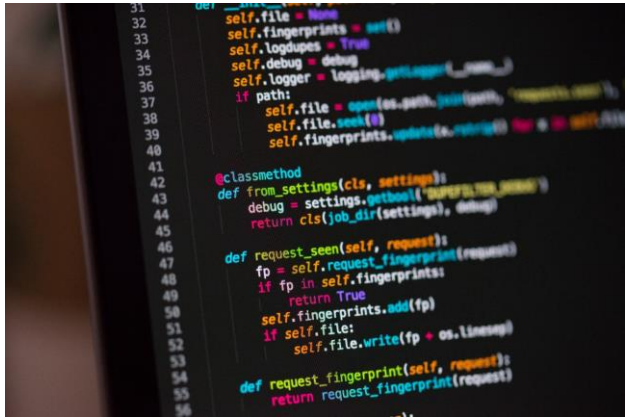
- Graphical User Interface (e.g. menus and buttons)
- network communication (e.g. emails)
- web development (e.g. websites)
- game development (e.g. video games)
- text processing (e.g. searching for keywords in a document)
- image processing (e.g. manipulating photos)
- data visualization (e.g. showing the temperature history)
- statistical analysis (e.g. performing t-test)
- machine learning (e.g. deep learning)
- data science

# Uses for Python

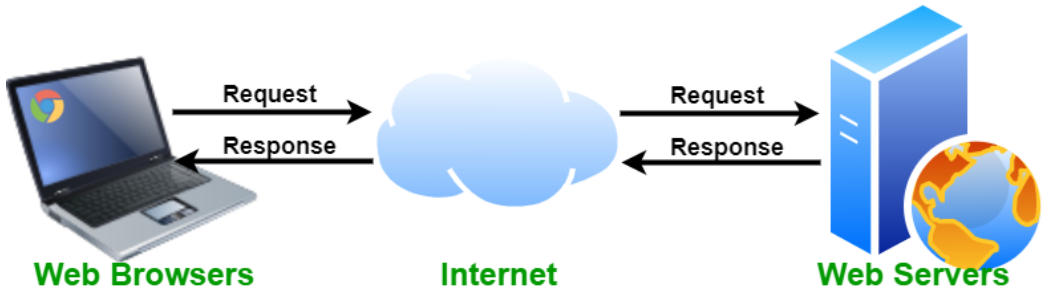
Machine Learning and Artificial Intelligence




Data Analysis and Plotting



Scripting

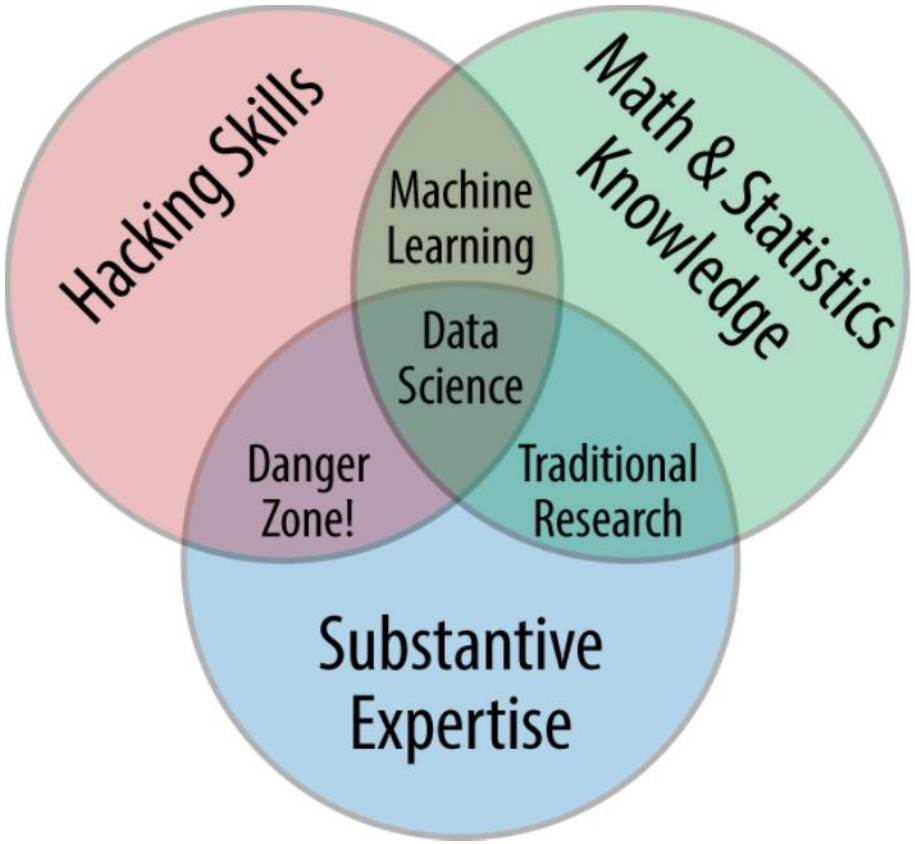



Web Servers / Interactive


Simple Video Games

# What is **Data Science** ?

The 'first' diagram
to define data science




**Data science** is an interdisciplinary field that combines computer programming (hacking), math, and machine learning to solve problems in a specific domain/field.

A **data scientist** needs to have:
(1) programming skills (hacking)
(2) knowledge of math, especially statistics
(3) knowledge of machine learning
(4) domain knowledge and expertise
e.g. biology, physics, psychology, …

http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

# What is **Machine Learning (ML)** ?

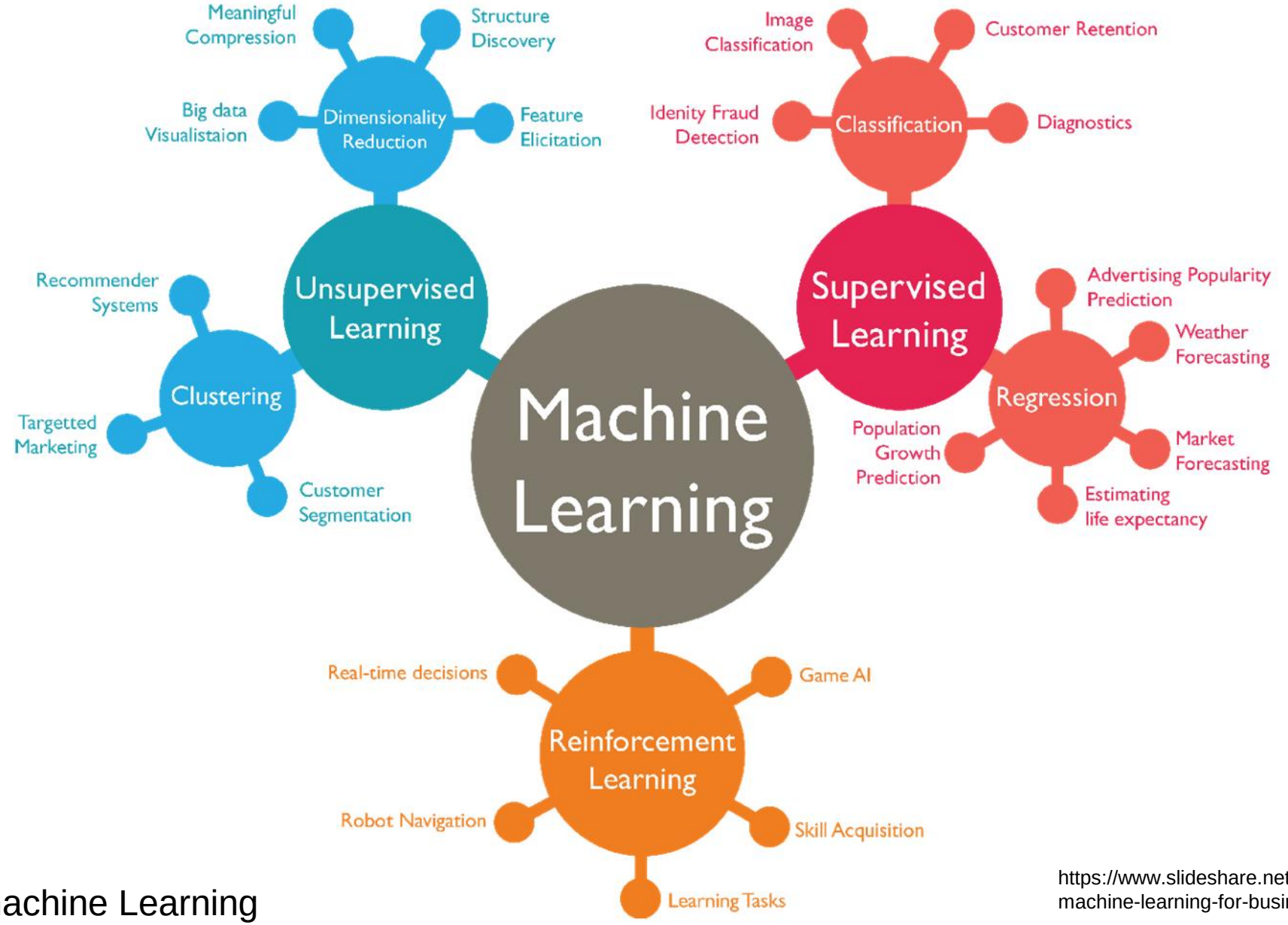
- Machine Learning is a sub-field of Artificial Intelligence.

  It has many definitions if you google it ….

  - Machine Learning is to extract patterns from data.
  - Machine Learning is to give computers the ability to learn without being explicitly learned.
  - Study of algorithms that improve their performance at some task with experience
  - Machine Learning is the study of (computer) algorithms that can learn something from data and apply the learned knowledge to perform some tasks.

- **ML** algorithms can keep improving their performance by using more data. - More Data, Better Performance.

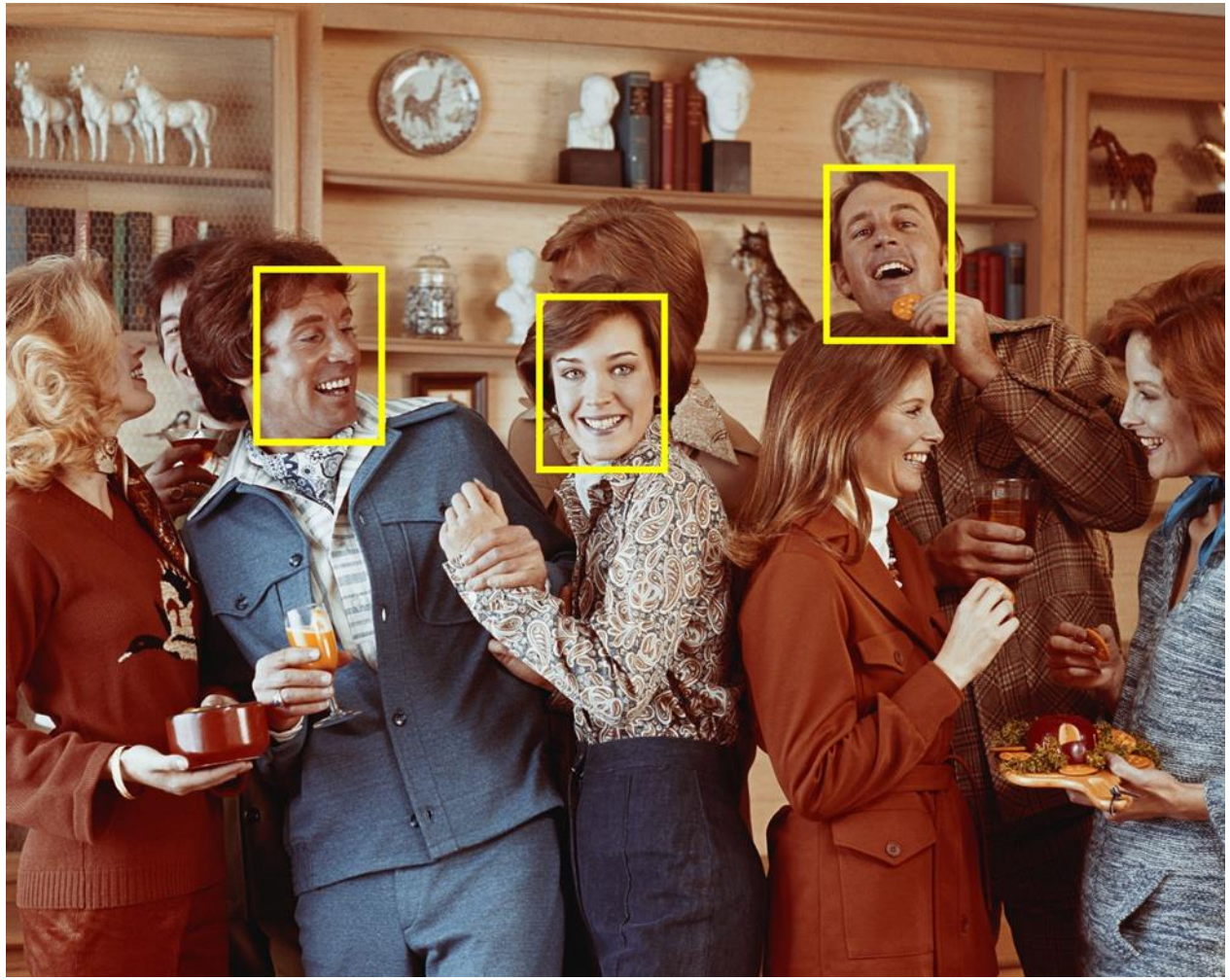
# What is **Machine Learning** ?




Three types of machine Learning

https://www.slideshare.net/awahid/big-data-and-machine-learning-for-businesses

# Companies using Machine Learning: **Facebook**: for face detection in photos

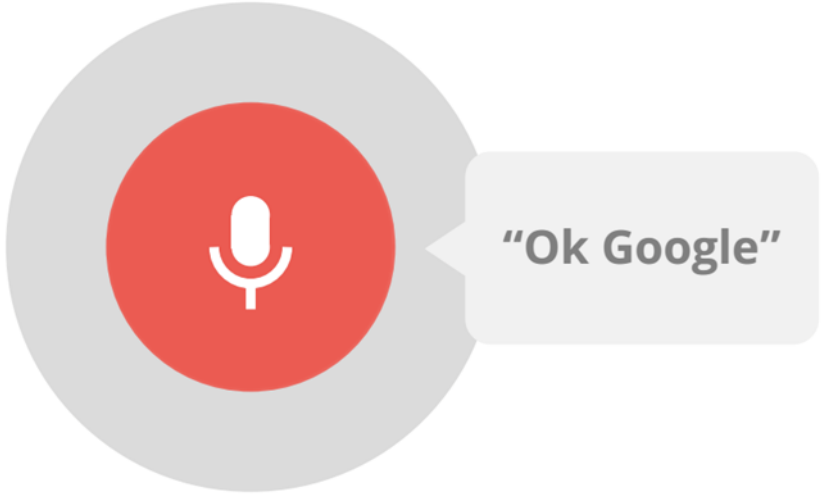


Facebook provides ML-based tools to automatically tag human faces in photos.

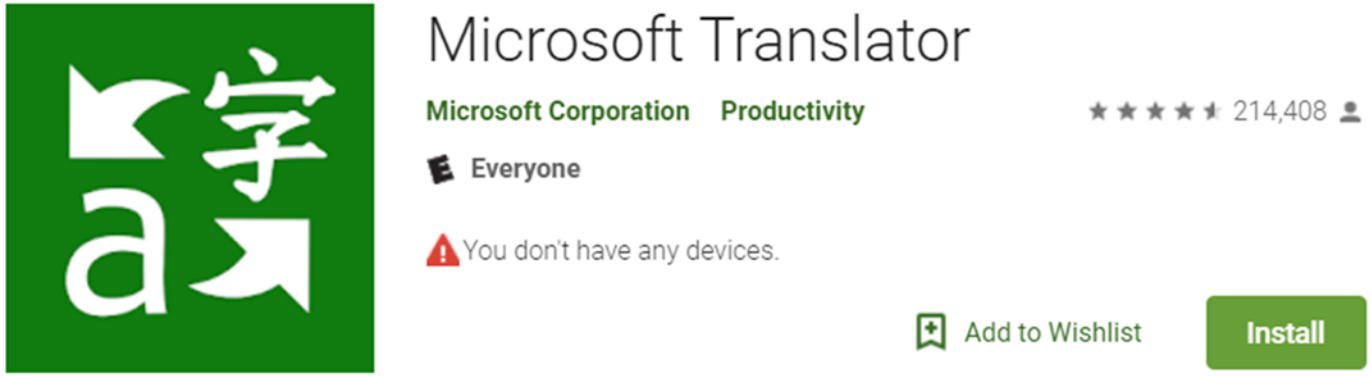
The same technology can be used in the video surveillance industry(e.g. looking for criminals who escaped from jails).

https://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/

# Companies using Machine Learning:
# **Google**, **Microsoft**, **Apple**: for speech recognition




ML algorithms have been developed to recognize human voice commands.




# Companies using Machine Learning:
# **Google**: for language translation

Google has developed ML algorithms to translate one language to another.

## Test: Chinese to English

Human Translation :

At Yellow Crane Tower in the west,
My old friend said farewell.
In the misty March, when flowers are blooming, he went down to Yangzhou.
Lonely sail, distant shadow.
Vanish in blue emptiness.
All I saw is the Yangtze river,
flowing into the far horizon.

Chinese sentences

Chinese - detected

故人西辞黄鹤楼，烟花三月下扬州。

孤帆远影碧空尽，唯见长江天际流。

Gùrén xī cí huáng hè lóu, yānhuā sān yuè xià yángzhōu.

Machine (Google) Translation

English

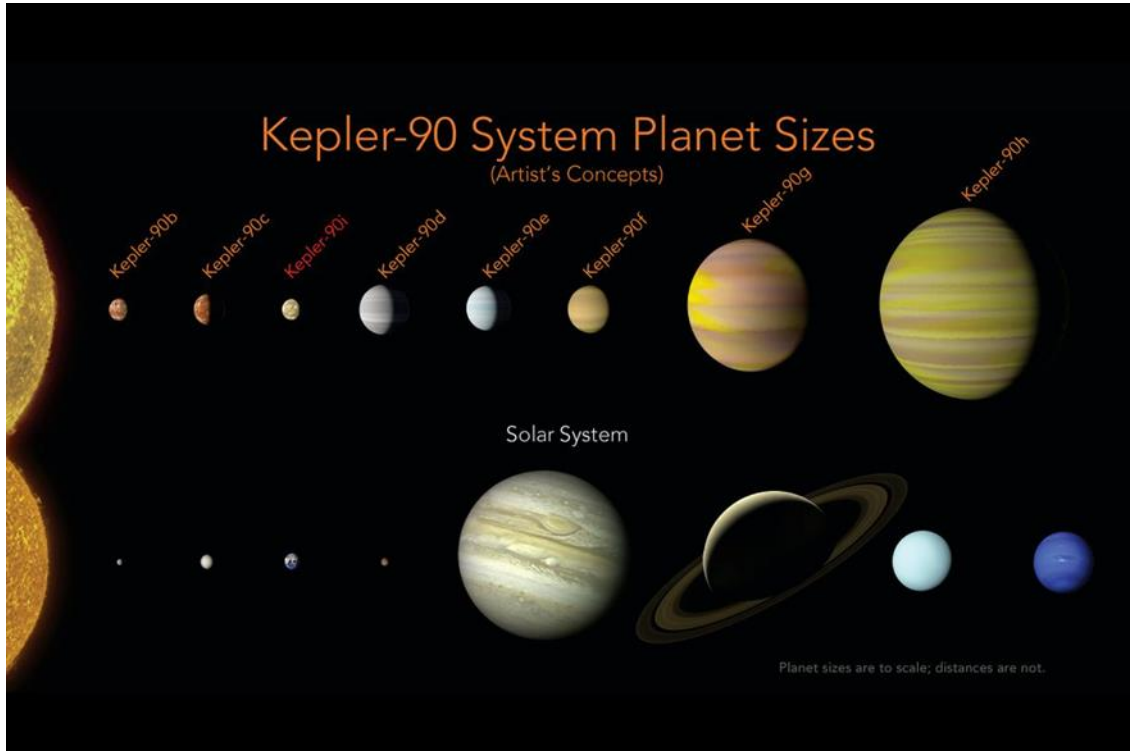
The old man resigned from the Yellow Crane Tower and the fireworks went down to Yangzhou in March.

The lonely sail is far away from the sky, only the Yangtze River is flowing.

# Companies using Machine Learning: **Google**: to discover new exoplanets

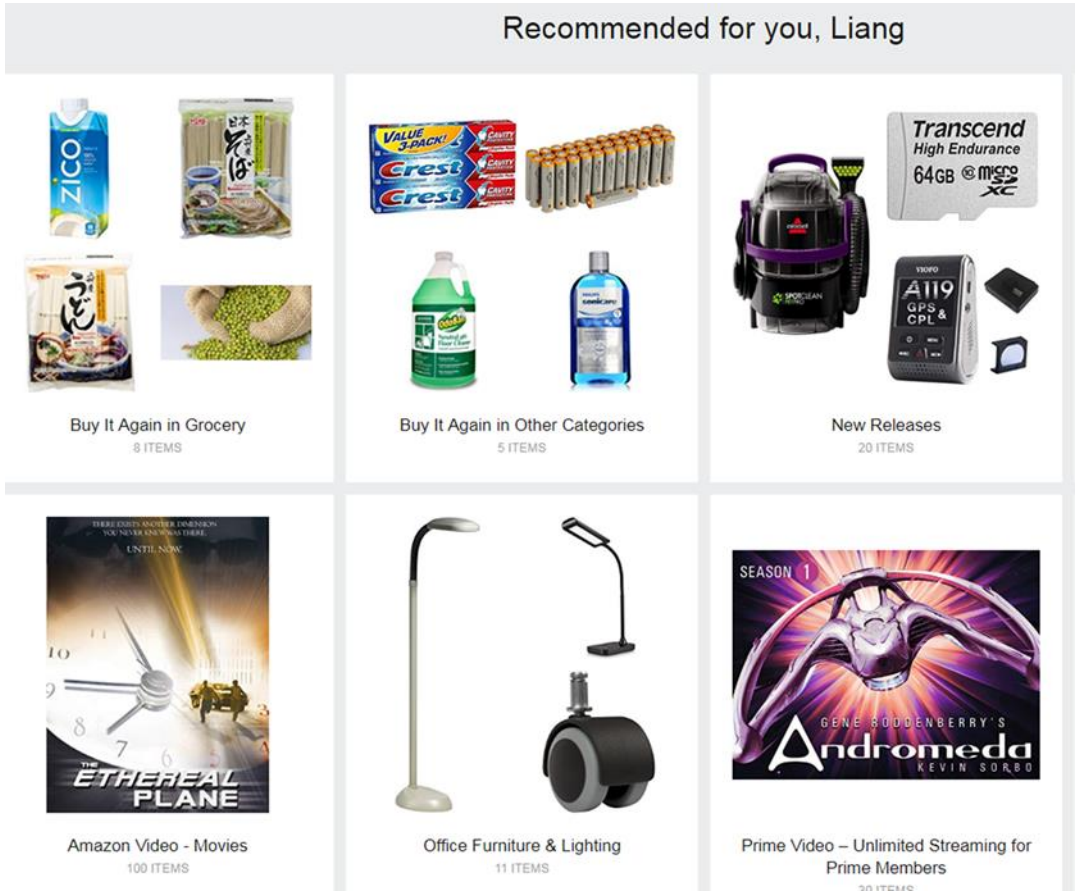
Researchers at Google in 2017 discovered two exoplanets by using ML algorithms to analyze data from NASA's Kepler space telescope and accurately identify the most promising planet signals.

https://ai.googleblog.com/2018/03/open-sourcing-hunt-for-exoplanets.html



# Companies using Machine Learning: **Amazon**: to make recommendations

https://www.amazon.com



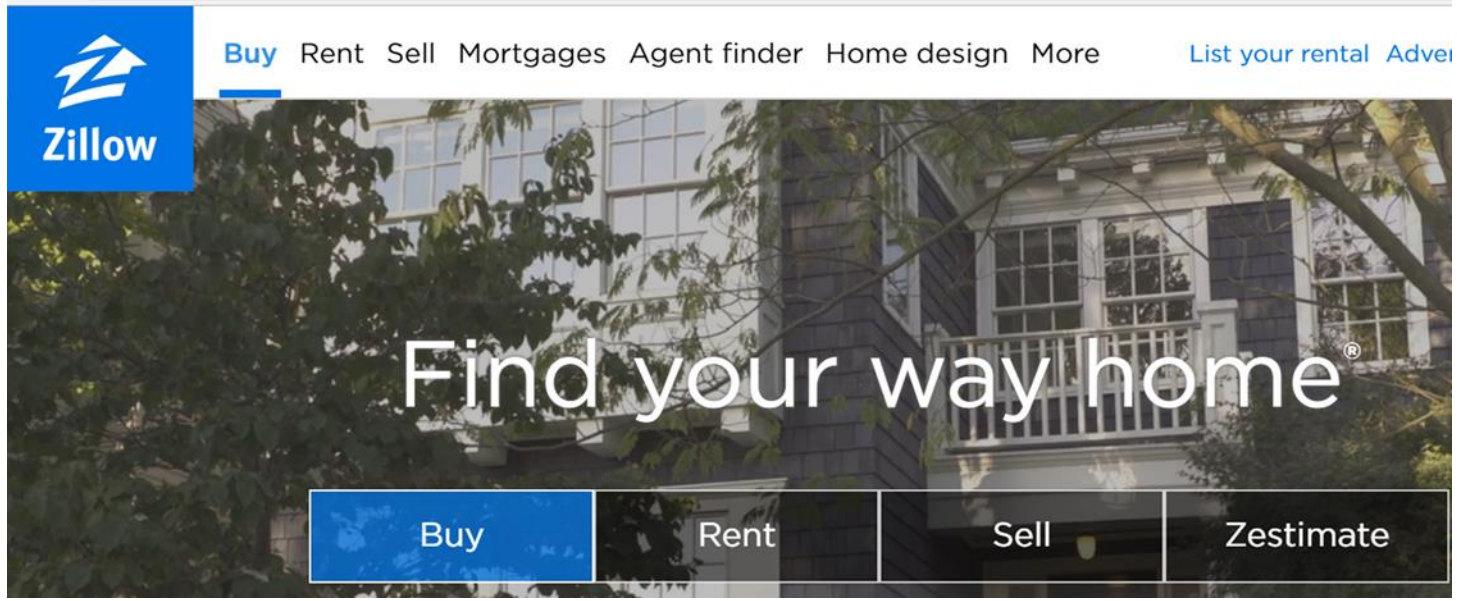

The recommendation system keeps tracking users' actions for days, months, or longer, and use ML to analyze users' preferences and then recommend products.

Big Brother is Watching You!

Amazon somehow figured out that I need lamp in my office

# Companies using Machine Learning:
# **Zillow**: to predict future sale prices of homes

https://www.zillow.com




The company Zillow is trying to use ML- algorithms to predict future sale prices of homes.

It was offering **$1,000,000 USD** to anyone who could develop ML algorithms for price prediction.

https://www.kaggle.com/c/zillow-prize-1

Zillow is the leading real estate and rental marketplace (online platform).
Through Zillow, people can buy, sell, and rent homes.

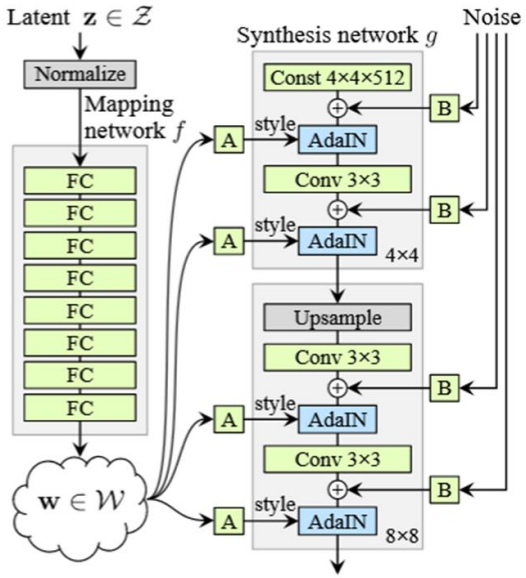
# Human Face Image Generation using a GAN

## A Style-Based Generator Architecture for Generative Adversarial Networks


Tero Karras
NVIDIA
tkarras@nvidia.com

Samuli Laine
NVIDIA
slaine@nvidia.com

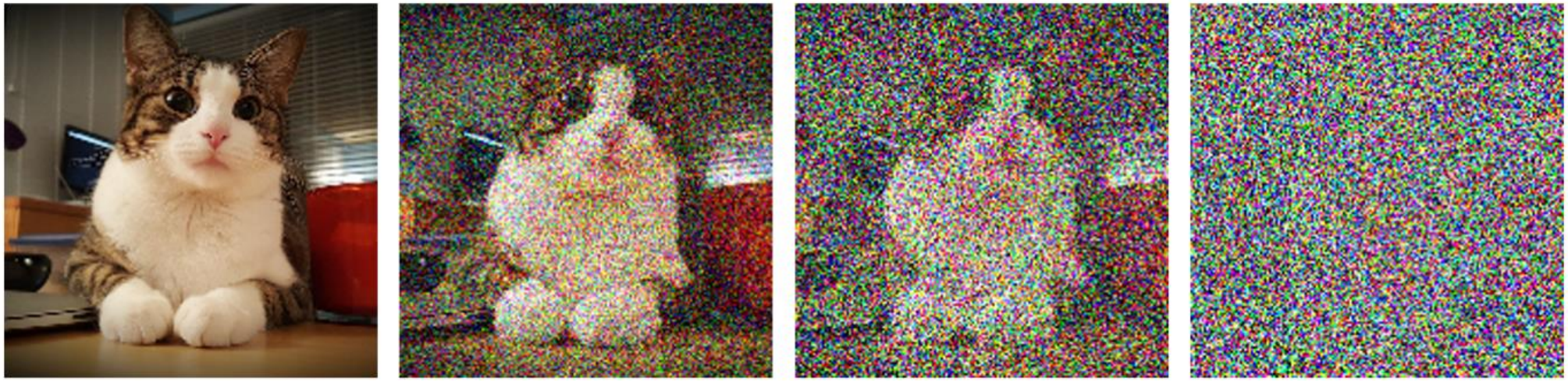
Timo Aila
NVIDIA
taila@nvidia.com


disable it for LSUN. Our training time is approximately one week on an NVIDIA DGX-1 with 8 Tesla V100 GPUs.

# Diffusion: a method to generate realistic images/videos

Forward: adding noises to the image in multiple iterations



Inverse: removing noises to recover the original image, which can be achieved by a neural network

https://openai.com/sora

https://arxiv.org/pdf/2208.11970.pdf

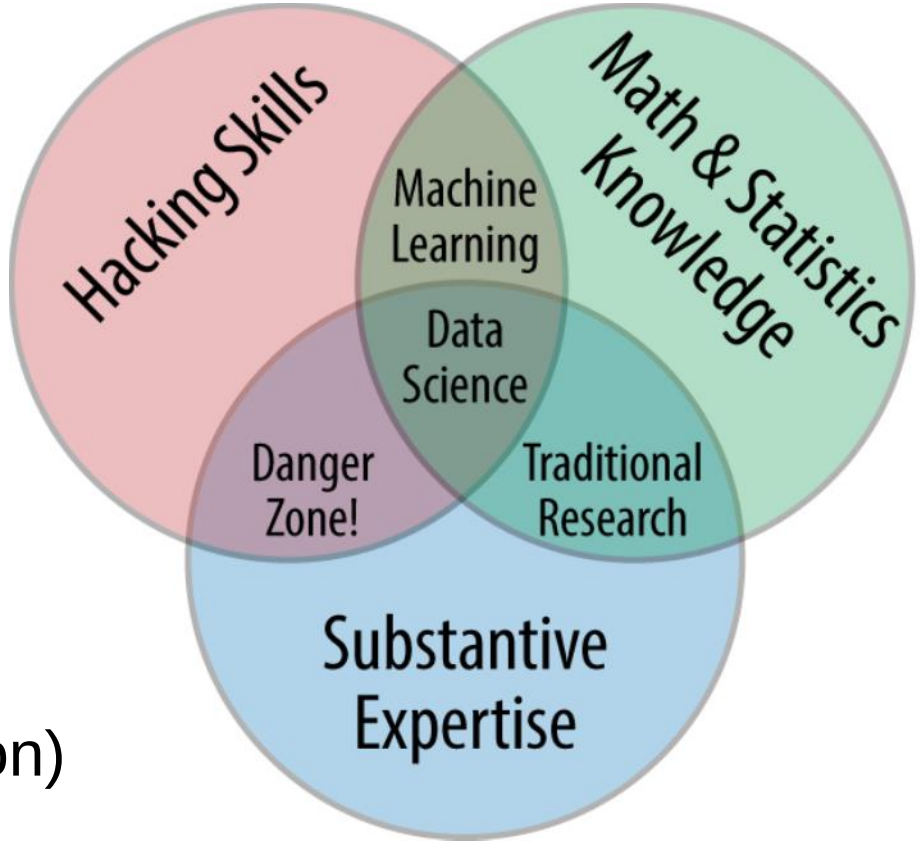
# Machine Leaning (ML) needs lots of mathematics

Basics (if you want to develop ML applications):

- Calculus
- Linear Algebra
- Probability and Statistics

Advanced (if you want to be a ML researcher):

- Information Theory
- Numerical Method and Optimization
- Signal Processing (speech and image recognition)
- Stochastic Process (reinforcement learning)
- Control Theory (reinforcement learning)



# Machine Leaning (ML) needs Python

- Python is #1 programming language for developing ML applications

Three major open source software packages for machine learning




Each package is written by using a mixture of different programming languages: C/C++ and Python.

The user (a ML-application developer) can use the packages though Python.

It may be difficult to use the packages though C/C++ (lack of documentation and examples)

# Data Science needs Python

- Python is #1 programming language for Data Science

Data Science combines programming, math, and machine learning to solve problems in a specific field.

Numpy: store data and manipulate data


Scipy and StatsModels for Math (linear algebra and stats)

Pandas to process tabular data

Scikit-image to process image data

Matplotlib to visualize data

There are many other packages…

# Outline

- Python basics (the first 5 weeks)

  learn: variables, functions, loops, strings, objects, classes, …

  develop: simple projects using Python

- Python packages (the remaining weeks)

  learn: basic concepts in data analysis and machine learning

  apply: Python packages for data processing, data visualization, and machine learning

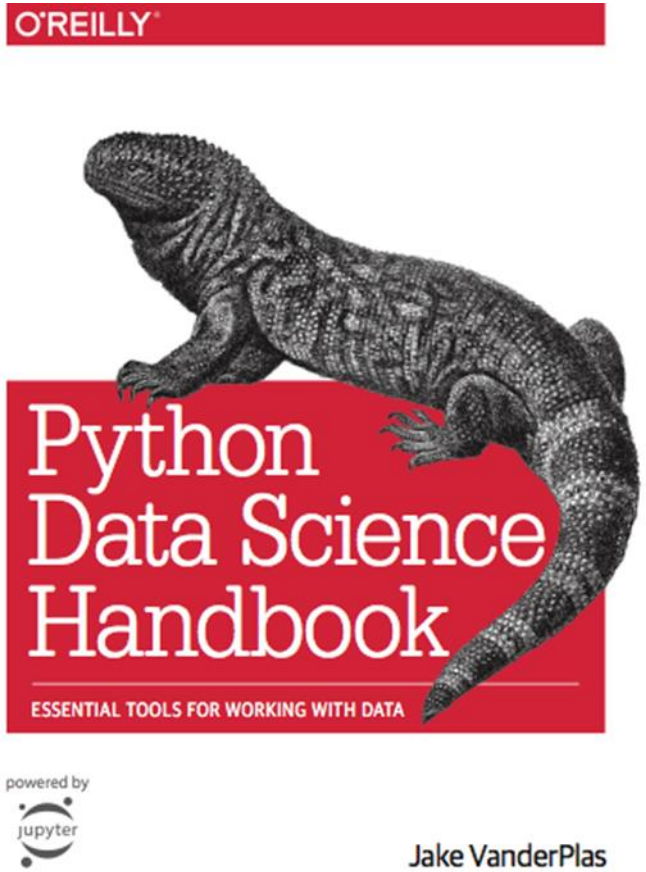
# Textbook – Python Basics

**A Byte of Python** at https://python.swaroopch.com/

More free books:
https://github.com/pamoroso/free-python-books

# Textbook – Python for Data Science

Python Data Science Handbook: **free** book with code
https://jakevdp.github.io/PythonDataScienceHandbook/



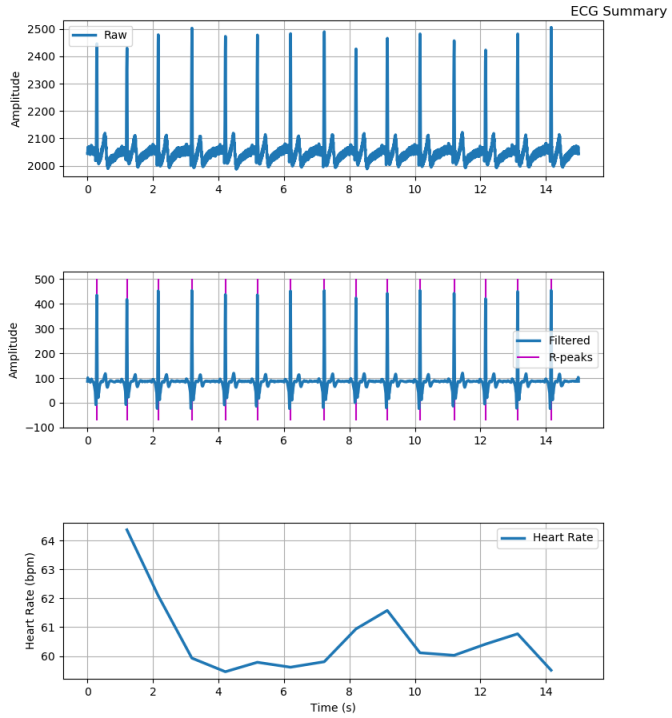
# Homework:  Labs and Projects

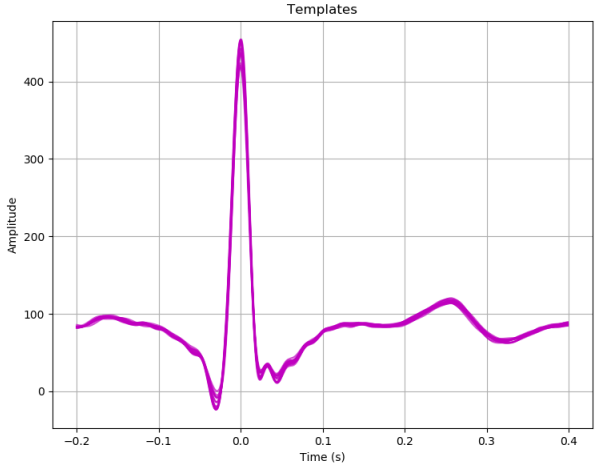
- **Labs**
- A lab has simple programming tasks or questions. For example, write a function to find the maximum value in a list.
- Most of the lectures will be accompanied by labs.

- **Programming Projects**
- There will be 4 projects: 2 for basics and 2 for data analysis.

# Sample applications in lectures and/or homework: 1D (dimensional) Data Processing and Analysis

- Temperature recordings over days or years (trend)
- ECG signals which are the recordings of the electrical activity of the heart over a period of time



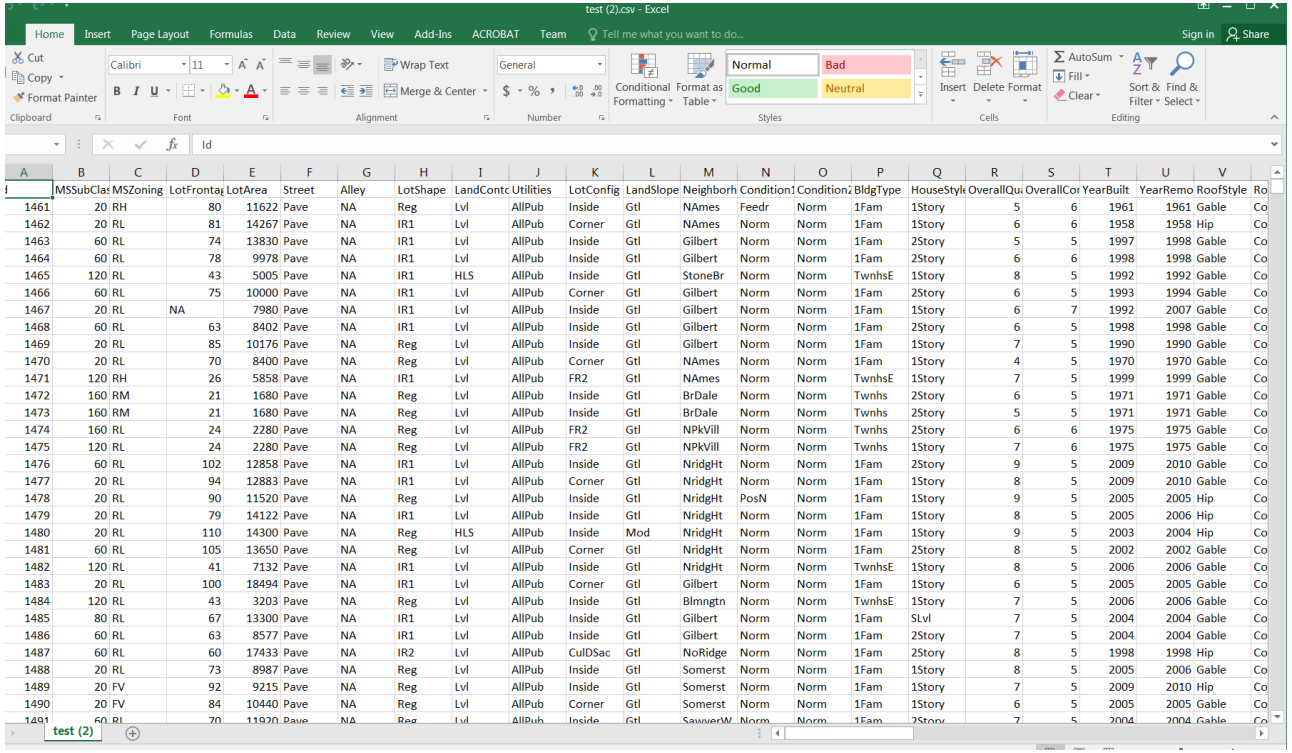

Heart rate can be extracted from ECG signals. Irregular heart rate could indicate some cardiovascular disease (atrial fibrillation).

# Sample applications in lectures and/or homework: Tabular Data Processing and Analysis

sample applications in lectures and/or homework :

- house price prediction (not Zillow)

The tabular data of house prices




| Id | MSSubCla | MSZoning | LotFronta | LotArea | Street | Alley | LotShape | LandCont | Utilities | LotConfig | LandSlope | Neighborh | Condition1 | Condition2 | BldgType | HouseStyle | OverallQu | OverallCo | YearBuilt | YearRemo | RoofStyle | Ro |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1461 | 20 | RH | 80 | 11622 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | NAmes | Feedr | Norm | 1Fam | 1Story | 5 | 6 | 1961 | 1961 | Gable | Co |
| 1462 | 20 | RL | 81 | 14267 | Pave | NA | IR1 | Lvl | AllPub | Corner | Gtl | NAmes | Norm | Norm | 1Fam | 1Story | 6 | 6 | 1958 | 1958 | Hip | Co |
| 1463 | 60 | RL | 74 | 13830 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 2Story | 5 | 5 | 1997 | 1998 | Gable | Co |
| 1464 | 60 | RL | 78 | 9978 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 2Story | 6 | 6 | 1998 | 1998 | Gable | Co |
| 1465 | 120 | RL | 43 | 5005 | Pave | NA | IR1 | HLS | AllPub | Inside | Gtl | StoneBr | Norm | Norm | TwnhsE | 1Story | 8 | 5 | 1992 | 1992 | Gable | Co |
| 1466 | 60 | RL | 75 | 10000 | Pave | NA | IR1 | Lvl | AllPub | Corner | Gtl | Gilbert | Norm | Norm | 1Fam | 2Story | 6 | 5 | 1993 | 1994 | Gable | Co |
| 1467 | 20 | RL | NA | 7980 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 1Story | 6 | 7 | 1992 | 2007 | Gable | Co |
| 1468 | 60 | RL | 63 | 8402 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 2Story | 6 | 5 | 1998 | 1998 | Gable | Co |
| 1469 | 20 | RL | 85 | 10176 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 1Story | 7 | 5 | 1990 | 1990 | Gable | Co |
| 1470 | 20 | RL | 70 | 8400 | Pave | NA | Reg | Lvl | AllPub | Corner | Gtl | NAmes | Norm | Norm | 1Fam | 1Story | 4 | 5 | 1970 | 1970 | Gable | Co |
| 1471 | 120 | RH | 26 | 5858 | Pave | NA | IR1 | Lvl | AllPub | FR2 | Gtl | NAmes | Norm | Norm | TwnhsE | 1Story | 7 | 5 | 1999 | 1999 | Gable | Co |
| 1472 | 160 | RM | 21 | 1680 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | BrDale | Norm | Norm | Twnhs | 2Story | 6 | 5 | 1971 | 1971 | Gable | Co |
| 1473 | 160 | RM | 21 | 1680 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | BrDale | Norm | Norm | Twnhs | 2Story | 5 | 5 | 1971 | 1971 | Gable | Co |
| 1474 | 160 | RL | 24 | 2280 | Pave | NA | Reg | Lvl | AllPub | FR2 | Gtl | NPkVill | Norm | Norm | Twnhs | 2Story | 6 | 6 | 1975 | 1975 | Gable | Co |
| 1475 | 120 | RL | 24 | 2280 | Pave | NA | Reg | Lvl | AllPub | FR2 | Gtl | NPkVill | Norm | Norm | Twnhs | 1Story | 7 | 6 | 1975 | 1975 | Gable | Co |
| 1476 | 60 | RL | 102 | 12858 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | NridgHt | Norm | Norm | 1Fam | 2Story | 9 | 5 | 2009 | 2010 | Gable | Co |
| 1477 | 20 | RL | 94 | 12883 | Pave | NA | IR1 | Lvl | AllPub | Corner | Gtl | NridgHt | Norm | Norm | 1Fam | 1Story | 8 | 5 | 2009 | 2010 | Gable | Co |
| 1478 | 20 | RL | 90 | 11520 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | NridgHt | PosN | Norm | 1Fam | 1Story | 9 | 5 | 2005 | 2005 | Hip | Co |
| 1479 | 20 | RL | 79 | 14122 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | NridgHt | Norm | Norm | 1Fam | 1Story | 8 | 5 | 2005 | 2006 | Hip | Co |
| 1480 | 20 | RL | 110 | 14300 | Pave | NA | Reg | HLS | AllPub | Inside | Mod | NridgHt | Norm | Norm | 1Fam | 1Story | 9 | 5 | 2003 | 2004 | Hip | Co |
| 1481 | 60 | RL | 105 | 13650 | Pave | NA | Reg | Lvl | AllPub | Corner | Gtl | NridgHt | Norm | Norm | 1Fam | 2Story | 8 | 5 | 2002 | 2002 | Gable | Co |
| 1482 | 120 | RL | 41 | 7132 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | NridgHt | Norm | Norm | TwnhsE | 1Story | 8 | 5 | 2006 | 2006 | Gable | Co |
| 1483 | 20 | RL | 100 | 18494 | Pave | NA | IR1 | Lvl | AllPub | Corner | Gtl | Gilbert | Norm | Norm | 1Fam | 1Story | 6 | 5 | 2005 | 2005 | Gable | Co |
| 1484 | 120 | RL | 43 | 3203 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | Blmngtn | Norm | Norm | TwnhsE | 1Story | 7 | 5 | 2006 | 2006 | Gable | Co |
| 1485 | 80 | RL | 67 | 13300 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | SLvl | 7 | 5 | 2004 | 2004 | Gable | Co |
| 1486 | 60 | RL | 63 | 8577 | Pave | NA | IR1 | Lvl | AllPub | Inside | Gtl | Gilbert | Norm | Norm | 1Fam | 2Story | 7 | 5 | 2004 | 2004 | Gable | Co |
| 1487 | 60 | RL | 60 | 17433 | Pave | NA | IR2 | Lvl | AllPub | CulDSac | Gtl | NoRidge | Norm | Norm | 1Fam | 2Story | 8 | 5 | 1998 | 1998 | Hip | Co |
| 1488 | 20 | RL | 73 | 8987 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | Somerst | Norm | Norm | 1Fam | 1Story | 8 | 5 | 2005 | 2006 | Gable | Co |
| 1489 | 20 | FV | 92 | 9215 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | Somerst | Norm | Norm | 1Fam | 1Story | 7 | 5 | 2009 | 2010 | Hip | Co |
| 1490 | 20 | FV | 84 | 10440 | Pave | NA | Reg | Lvl | AllPub | Corner | Gtl | Somerst | Norm | Norm | 1Fam | 1Story | 6 | 5 | 2005 | 2005 | Gable | Co |
| 1491 | 60 | RL | 70 | 11920 | Pave | NA | Reg | Lvl | AllPub | Inside | Gtl | SawyerW | Norm | Norm | 1Fam | 2Story | 7 | 5 | 2004 | 2004 | Gable | Co |

The data has many attributes including:
LotShape: General shape of property
LandContour: Flatness of the property
Utilities: Type of utilities available
Heating: Type of heating
GarageArea: Size of garage
….

You will learn how to pre-process the data and use ML algorithms to predict the sale price of a house based on the above information.

# The course focuses on Python Programming…

- The second half of the course is a tour of data science. The goal is to build your Python programming skills through applications and get familiar with the basic Python packages for data analysis.
- The examples in each lecture will be relatively simple, just to show the basic concepts and algorithms.
- This course is Not Machine Learning: you will not be asked to develop new algorithms to discover new planets, or win 1million $ prize from Zillow. (But you are welcome to try…)
- If you are interested in Machine Learning:
  - “Introduction to Machine Learning with Applications” CSC546/646
  - “Neural Networks and Deep Learning” CSC746

# The course is suitable for students who…

- have some programming skills in other programming languages
- want to learn Python basics
- want to learn Python packages for data science
- want to learn the basics of data analysis and machine learning
- are willing to spend lots of time programming

Programming is fun if you like it.

# How do we learn Python ?

- Attend lectures to get a rough understanding of Python
- Complete the labs (to be posted on Blackboard)
- Complete the projects (to be posted on Blackboard)
- When you get stuck, ask for help (in person or via email)

Instructor: liang.liang@miami.edu

TA: Sam Pass <sap215@miami.edu>