INSEAD The Business School for the World®

Machine Learning and Optimization Lecture 1

Introduction to the course and Python

Professor Georgina Hall

Agenda for today

Learning objectives:

- Introduction to the course: vocabulary, major concepts
- Logistics of the course
- Become familiar with tools used in the classroom: quizzes, poll everywhere, google colab
- Python recap

How will we get there?

- Leverage your general knowledge

and what you learnt during P0!

About me…

- Professor Georgina Hall

mlo.mim@insead.edu

- Assistant Professor of Decision Sciences
- Research topic: Optimization (algorithms, applications, etc.)
- French, British, Canadian
- PhD from Princeton University in Operations Research
- MS/BS from Ecole Centrale Paris, France

Course objective (1/2)

Goals: At the end, you will:

- know what “machine learning”, “optimization” and “coding” are in concrete terms
- understand at a high level what data analysts/scientists do so you can interact with them easily in the future
- understand what machine learning and optimization can do (and cannot do) for a company’s $$$

How will we do this?

- Hands-on approach to understanding machine learning algorithms and optimization
- Many business cases, datasets, exercises

Course objective (2/2)

Am I going to find this class hard?

- **NO,** not if you work.
- The class is not designed for programmers. Had complete newbies in Python get amazing grades last year.
- Work starts **NOW** though (not right before the final)

- Tips to acing this course:
  - **Always** learn your lectures
  - **Do** the homework regularly
  - **Practice** Python
  - **Come to office hours** & email me if you are still struggling

Installment teaching

- When I do pure content (i.e., no activities), I will teach in **10-15 mins installments** (to make it more digestible)

- I want everyone to be **fully concentrated** during these segments

- I will always give 2-3 mins **"quiet" breaks** between two installments.

- Permissible activities in the breaks:
  - → Asking clarification questions
  - → Being on your phone / reading over your notes / day-dreaming
  - → **No talking please**

- **Breaks contingent on getting back to work immediately afterwards & no talking.**

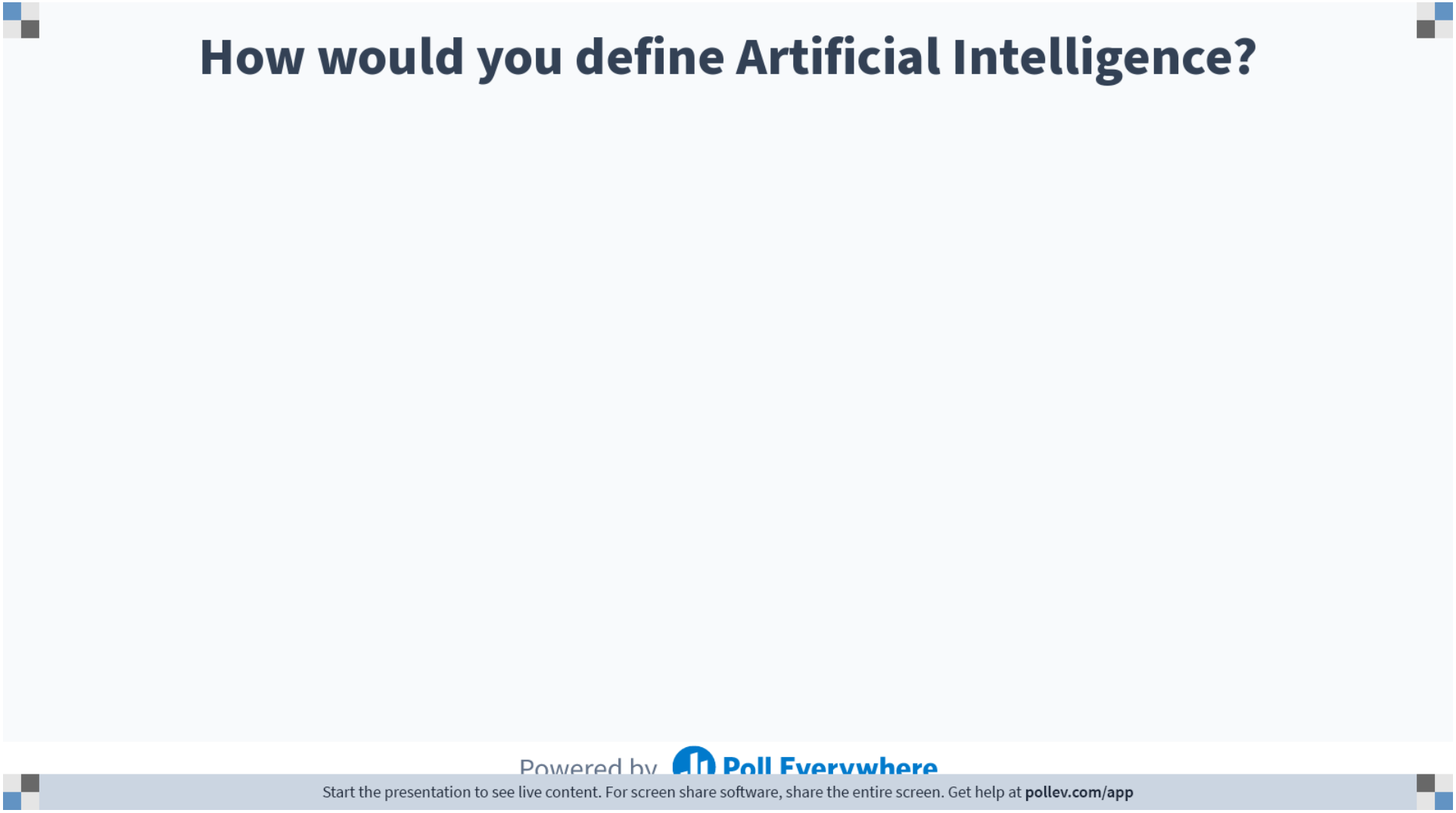
Course content

Activity on Poll Everywhere

- We are going to use **Poll Everywhere** for an activity
- One of my go-to tools in class
- Please save **pollev.com/mlo** as one of your favorites in your browser
- Please **indicate your full first and last name** as you sign in to get recognition of your participation for the participation grade

Before getting started…

Where are we at?

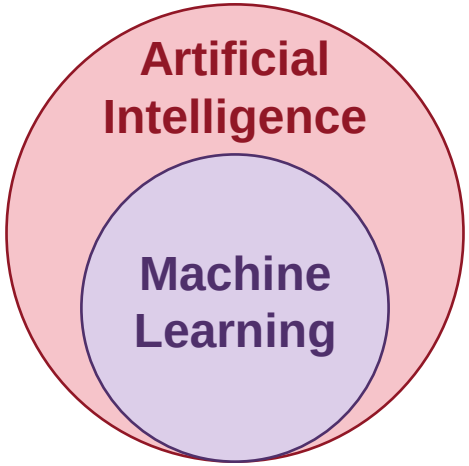


How would you define Machine Learning?

ML vs AI

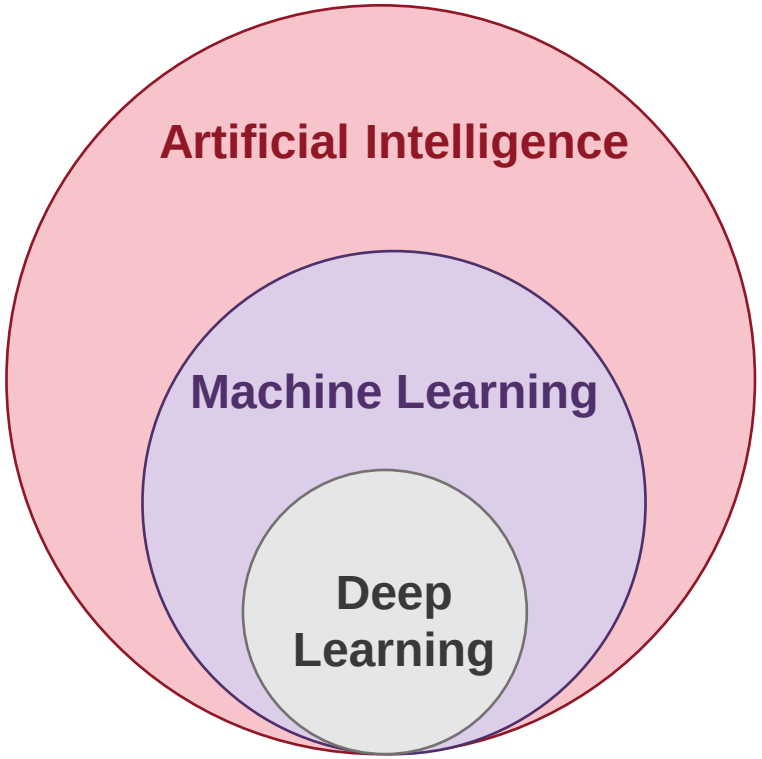
Machine Learning (ML):
Well-defined concept. Set of algorithms (i.e., procedures) that take data as an input and learn patterns or predictions from this data without being told explicitly how to build these patterns or predictions.

Artificial Intelligence (AI):
Not as clear. Typically set of tasks performed by a computer that are commonly associated with intelligent beings (includes interaction with environment and making decisions to maximize goals)



And what about Deep Learning?

Deep Learning:
Set of techniques for supervised machine learning. Involves fitting neural networks to input data to predict an output variable.



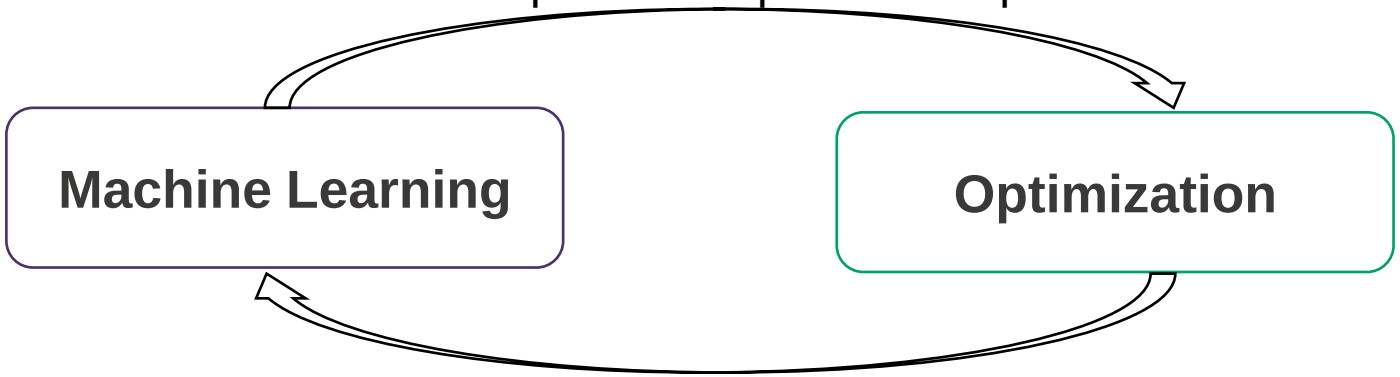
How would you define Optimization?

And Optimization?

Optimization
Find the best decision to make
when there are many different options to choose from and constraints to satisfy

Why does this course contain both Machine Learning and Optimization?
What is their link?

Data outputted by a machine learning algorithm can be used as input to an optimization problem

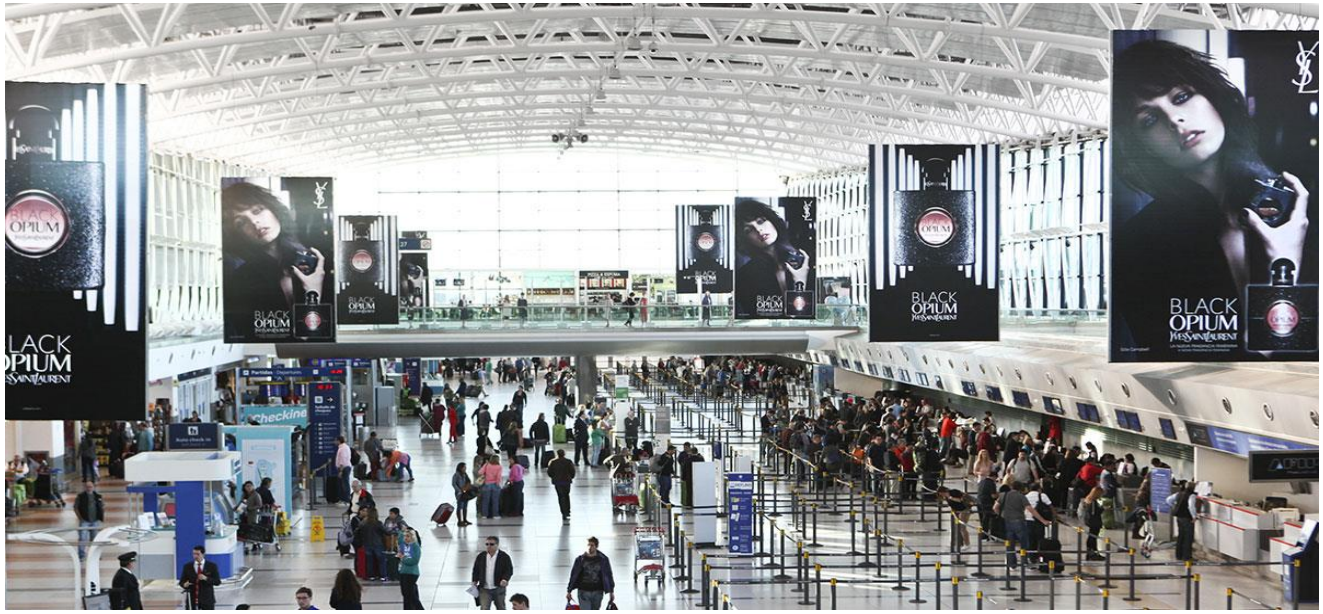


(Nearly) all machine algorithms solve an optimization problem in the background

INSEAD

Some applications of Machine Learning & Optimization to Business

Example 1: Travel Retail



A large cosmetics brand owns displays in a duty-free boutique in an airport.

Questions:

- How many people and of which nationality are likely to come by my display?
- How do spending habits differ between a client from the Middle East and a client from Asia?
- How should I best display my products to maximize profit?

Example 2: Medical Sales

A pharmaceutical company has to decide how to best dispatch its medical sales representatives.

Questions:

- Based on past data, which doctors would be most open to listening to the representatives' sales pitch?
- What is the shortest route a given representative can take to visit all the doctors they need to visit?

Class organization

How is the class organized?

Introduction to the course		Machine Learning		Optimization	
Python Recap	Data Engineering	Supervised Machine learning	Unsupervised Machine Learning	Linear programming	Integer Programming
Sessions 1 and 2		Sessions 3, 4, 5, 6, 8, 9		Sessions 11 and 12	

Case Sessions	
“Regular” cases	Capstone case
Sessions 7, 10, 13, 14, 15	

Introduction to the course

Introduction to the Course and Python Recap [Lecture 1]

- What is Machine Learning and Optimization?
- How do I use data structures in Python?
- How do I plot in Python?

Data Engineering [Lecture 2]

- How do I go from “raw” data to processed data?
- How do I cope with missing values, outliers, duplicates, categorical variables?

Machine Learning

Supervised learning
[Lectures 3,4,5,6]

Data is labeled (e.g., this email is spam/this email is not)

Goals:

- *Regression*: based on past data, what sales will we have tomorrow?
- *Classification*: based on past data, will we be able to retain this client?

Techniques:

- *Regression*: linear/polynomial regression, tree-based methods
- *Classification*: logistic regression, tree-based methods

Unsupervised learning
[Lectures 8,9]

Data is unlabeled

Goals:

- *Dimension reduction*: how can I go from many features of a client to a couple of main ones that describe his/her spending?
- *Clustering*: how can I cluster clients with similar spending habits?

Techniques:

- *Dimension reduction*: PCA
- *Clustering*: k-means clustering, hierarchical clustering

Other techniques [Lecture 10, briefly]

Optimization

- What is an optimization problem?
- Where do they appear?

Linear optimization [Lecture 11]:

- What is linear optimization?
- How to solve a linear optimization problem in Excel?

Integer optimization [Lecture 12]:

- What is integer optimization?
- Why is it useful? Where would I use it?

Datasets, Cases, and Exercises

Business Applications:

- How much is a diamond worth?
- Will this customer default on his/her credit card loans?
- Will a travel company retain their customers next year?
- How does poor reporting on gender lead to unfair outcomes?*
- How can a boats company segment its customer base?
- How should a tomato-product company decide to allocate the tomato crops it buys?
- How should a company configure its distribution network?
- How can we use analytics for headcount reduction?

"**Fun" Applications:**

- Was the movie Titanic realistic?
- What are the main "dimensions" of a McDonald's menu?
- How does image compression work?*
- Which countries are similar, and which are different?

Course logistics & norms

Homework

Between two classes:

- **Always learn** the previous **lecture**

→ **Quiz** at the beginning of most lessons to check your understanding

- **Always** finish up the material you were not able to complete in class

- **Sometimes**, there will be **short assignments** for you to do to prepare you for the final

- **For Sessions 7, 10, 13, 14 & 15**: you will have a case to read and sometimes work to prepare as a group.

Your grades

- **Quizzes (Individual):** at the beginning of most lectures, 10% of the grade

- **Participation in class (individual)**: 15% of the grade

Contributes positively to the grade: presence, in-class participation (including on PollEv), helping classmates, timely return to class after BORs, etc.

Contributes negatively to the grade: being disengaged, being late, being on your phone/computer except when authorized, not being tuned into the class when I cold-call, etc.

- **Work handed in for cases (group-based):** 15% of the grade

- **Theoretical final (individual)**: 30% of the grade

- **Python practical final (individual):** 30% of the grade

Classroom norms (1/2)

In class:

- **Please arrive on time:** The quiz starts on the dot and stops on the dot – you will not get an extension if you are late. **Swiss train policy holds.**
- **Masks on!**
- Scan the **QR code** for attendance.
- Sit according to seating chart.
- **Please only use your computers for the task at hand**: no social media, no browsing.

Missing class:

- You can miss a **maximum of 2 sessions** per course **for justified reasons only (sickness, interviews, family reasons).**
- A third absence may **lead to a failing grade.**
- **A heads-up** is appreciated but not mandatory

Classroom norms (2/2)

On Zoom:

- **Only if** you cannot make it to campus due to Covid reasons.
- Need to **seek approval** from the MIM team first (mim.programme@insead.edu)
- Please **turn on camera on Zoom** so I get to know you!

Lectures and final:

- Please come prepared to class: lecture learnt, homework done
- No make-up finals

Discussion around norms is ongoing and subject to change.

Who to reach out to in MLO?

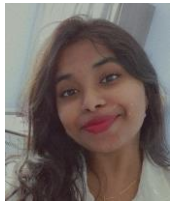
Meet the team

Georgina Hall (FBL) - Lecturer

- Lectures
- **Office hours: Monday & Wednesday, 7:30pm to 8:30pm over Zoom**

Piyush Gulati (SGP) - Tutor

- Will help out in certain lectures & has helped me write the final case
- **Office hours: Saturdays, 9am to 11am over Zoom**
- Mostly Python focused



Ida-Jananni Augustin Iroudayadasse (FBL) – Course Assistant

- Can help you with any logistics query

Dmitrii Sumkin (SGP) – Grader

- Will grade everything you hand in

Python Taskforce Section Dependent

One unique email address

mlo.mim@insead.edu

All 4 of us receive the emails sent to this address.

Coding

Logistics of coding

Coding

Coding is **central to this class**

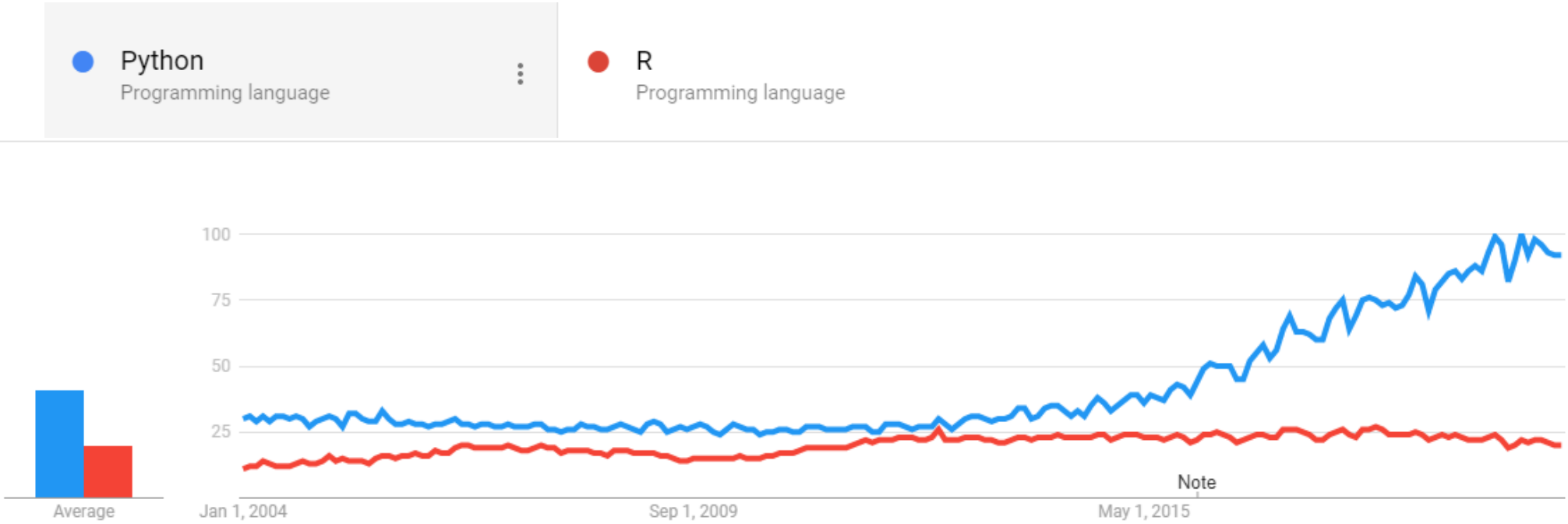
because it enables us to **deploy what we learn.**

- Not asking you to become a class A programmer: **just the basics.**
- The beginning can be a little hard: hang on in there. It gets easier!
- The best way to get better at coding is **practice, practice, practice**.

The choice of a language

You have worked with two languages already: Python and R.

In this course, I have decided to go with Python.



Source: google trends

Logistics of coding

- We will be using **Jupyter Notebooks** for the class (unlike Datacamp)

- You can open/modify these either in **Jupyter/Anaconda** or on **Google Colab**

- Each lecture will have an **“Exercise Book” notebook** for you to fill in and a **“Solutions” notebook** that I post a couple of days later

- **Let’s get started using Google Colab (or Jupyter Notebooks)! Run Section 1.**

Quiz: coding & you

Please log onto the LMS course platform (Module 1) and complete Quiz Lecture 1.

[No grades on this one]

This is what you will have to do at the beginning of each lecture so do get familiar with the process at the same time!

After that: start working on the Python Exercise Sheet!

Please call me/the Python Taskforce if you're encountering issues and/or write in the chat.

Quick reminder of some coding basics

Installing packages:

- Google Colab already has all the major packages installed (pandas, numpy, matplotlib – otherwise conda install)
- Don’t forget to import the package:

```
import numpy as np
```

Basic operations and types in Python:

- +: sum, / : division, **: exponentiation
- # to comment something out
- int / float / string / Boolean

Quick reminder of some coding basics

Different data types:

Python Lists

```
L=["One", 1 , True]
print(L[0])
print(L[:1])
print(L[-1])
```

Numpy Arrays

```
import numpy as np
Ln=np.array([1,2,3])
print(max(Ln))
Mn=np.array([4,5,6])
print(Ln+Mn)
```

```
3
[5 7 9]
```

Python dictionaries

```
dict={"key1":["Paris"], "key2":["Rabat"], "key3":["Milan"]}
dict["key1"][0]
```

```
'Paris'
```

Pandas dataframes

```
import pandas as pd
Dataframe1=pd.DataFrame(dict)
Dataframe1
```

	key1	key2	key3
0	Paris	Rabat	Milan

Quick reminder of some coding basics

Comparison operators:

- >=,<=,==, !=
- and (&), or (|), not

If/then/else loops:

```
school="INSEAD"

if school=="INSEAD":
  print("It's the best!")
elif school=="LSE":
  print("Not as good")
else:
  print("What's your FT ranking?")
```

```
It's the best!
```

For loops:

```
for i in Ln:
  print(i)
```

```
1
2
3
```

Quick reminder of some coding basics

Plotting:

- Use Matplotlib and the pyplot package in Matplotlib
- plt.plot(x,y), plt.show(), plt.scatter(), plt.hist()
- Customization: plt.xlabel(), plt.ylabel(), plt.title(), s=, c=

Wrap-up & Next time

Today, we saw:

- **Different concepts:** AI / Machine Learning / Deep Learning / Optimization
- **Applications** in different areas
- Course logistics
- A **review of coding** in Python

Next time:

- Introduction to the course: Data pre-processing and feature engineering
- Learn lecture + finish Python exercises
- Quiz at the beginning of next lecture

The Business School
for the World®

EUROPE | ASIA | MIDDLE EAST | NORTH AMERICA