

Introduction to Data Science Course

Introduction to Data Science

Le Ngoc Thanh
Inthanh@fit.hcmus.edu.vn
Department of Computer Science

Contents

- ◎ **What is Data Science?**
- ◎ Skills for Data Science
- ◎ Data Science Process
- ◎ Applications
- ◎ Why to learn?
- ◎ About course

What is Data Science?

◎ **Extracting useful conclusion** from large datasets with:

- Processes
- Problem definition
- Set of principles
- Analysis
- Algorithms
- Computer



Data Scientist vs all







Data Scientist?

Data Scientist vs all

There's a joke that says:

“A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician.”

Data Scientist also known as Data Managers, statisticians.	Data Engineers also known as database administrators and data architects.	Data Analysts also known as business Analysts.
		
A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.	They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.	They typically help people from across the company understand specific queries with charts.
Skills: Mathematics, Programming, Communication	Skills: Programming, Mathematics, Big data	Skills: Statistics, Communication, Business knowledge
  	  	  
Will use programmes such as: SQL, Python, R	Will use programmes such as: Hadoop, NoSQL, and Python	Will use programmes such as: Excel, Tableau, SQL

History of Data Science

◎ History of data collection

- 3200 BC....
- Record experiences, events through writing
- Example: the sale of an item, the issuing of an invoice, the delivery of goods, ...

◎ History of data analysis

- Statistics
 - Summary (descriptive) statistics
 - Probability theory, probability distributions
 - Statistical learning

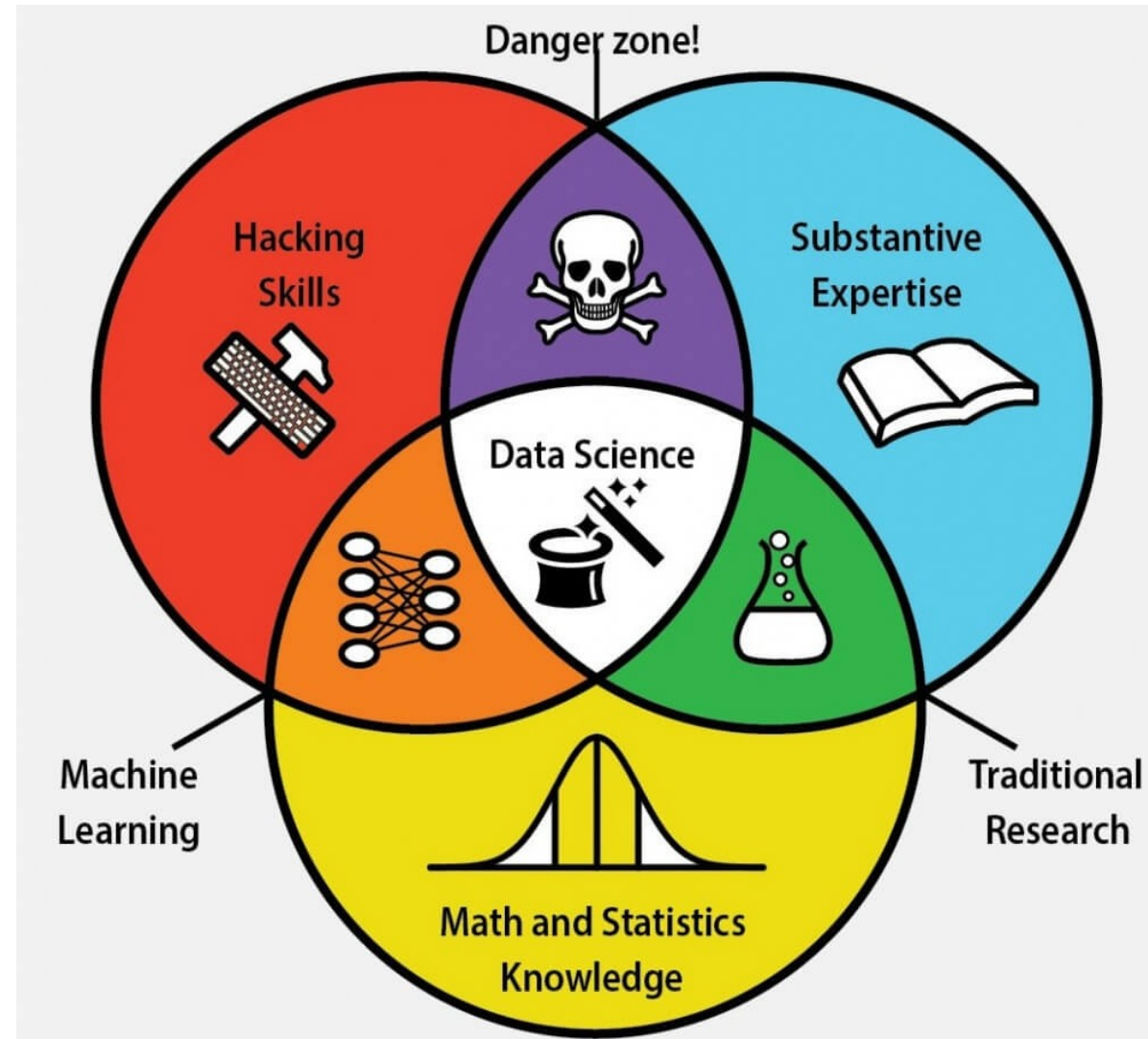
◎ 1990s: emergence and evolution of data science

The need for statisticians to join with computer scientists to bring mathematical rigor to the computational analysis of large data sets.































Contents

- © What is Data Science?
- © **Skills for Data Science**
- © Data Science Process
- © Applications
- © Why to learn?
- © About course

Skills-set for a data scientist



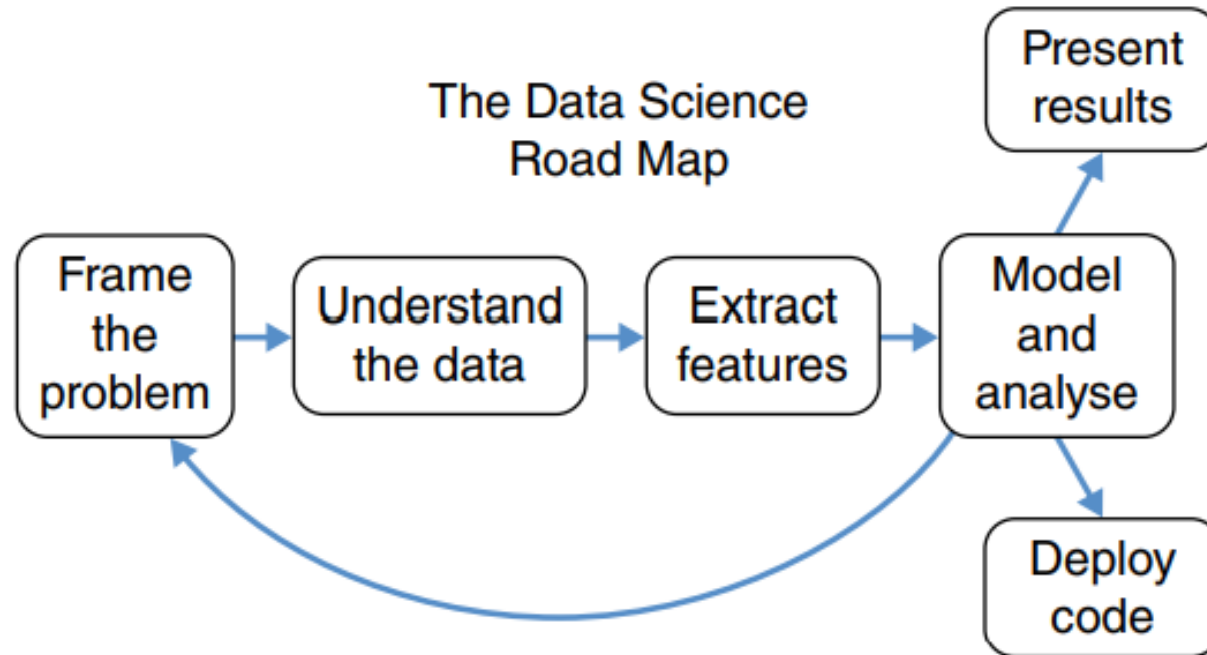
Skills-set for a data scientist

Roles/Skills	Advanced Analytics	Business Acumen	Data Visualization	Communication and Collaboration	Software Skills
Data Analyst					
Business Analyst					
Data Scientist					
Data Architect					
Data Engineer					
Data Visualization Engineer					

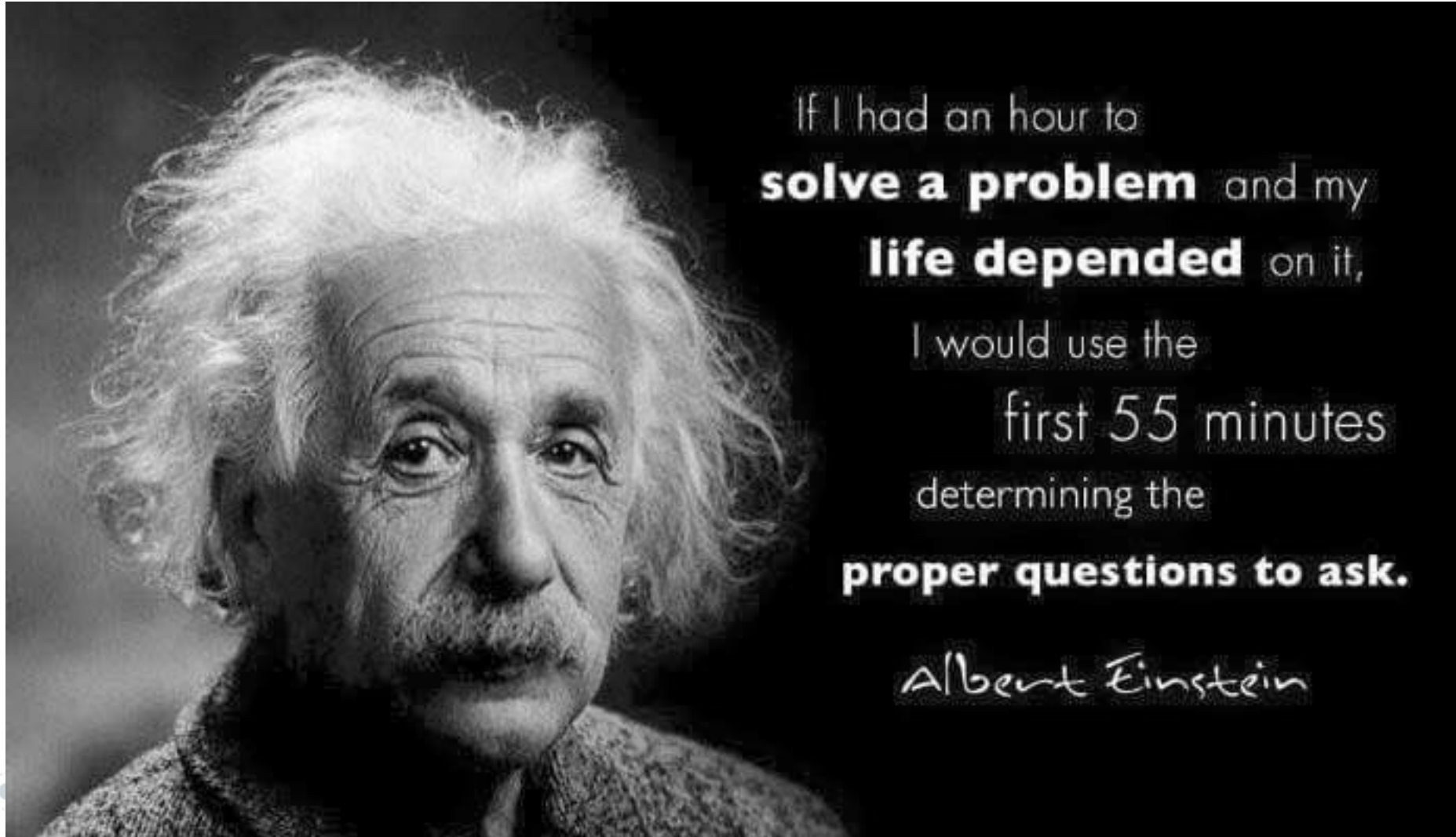
Contents

- © What is Data Science?
- © Skills for Data Science
- © **Data Science Process**
- © Applications
- © Why to learn?
- © About course

What is DS-ist do?



Data Scientist ~ “Frame is problem”



Example 1

- Let's go shopping!

Milk, eggs, sugar,
bread



Customer1

Milk, eggs, cereal,
bread



Customer2

Eggs, sugar



Customer3



Example 2

Emmy-winning US TV Shows



Police Detective TV Dramas



Critically Acclaimed Witty TV Shows



Example 3



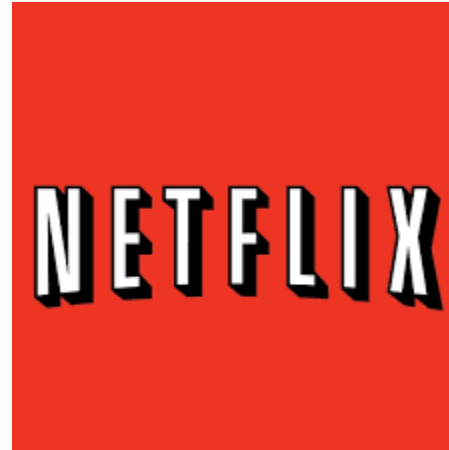
Contents

- ◎ What is Data Science?
- ◎ Skills for Data Science
- ◎ Data Science Process
- ◎ **Applications**
- ◎ Why to learn?
- ◎ About course

Where Is Data Science Used?

◎ Case studies that illustrate the impact of data science:

- **Consumer companies:** sales and marketing
 - Walmart
 - NetFlix
 - Amazon

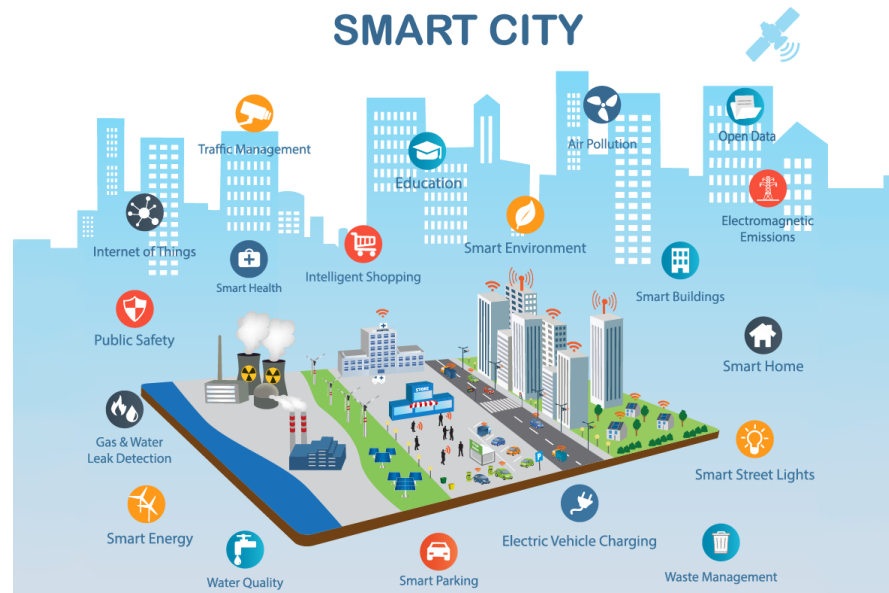


Where Is Data Science Used?

◎ Case studies that illustrate the impact of data science:

○ Governments

- Organize cities: track, analyze, and control environmental, energy, and transport systems and to inform long-term urban planning
- Design drugs for individual patients
- Predict crime hot spots and recidivism



Where Is Data Science Used?

◎ Case studies that illustrate the impact of data science:

- Consumer companies: sales and marketing
- Governments: improve health, criminal justice, and urban planning
- Professional sporting:
 - Strategies



Where Is Data Science Used?

Case studies that illustrate the impact of data science:

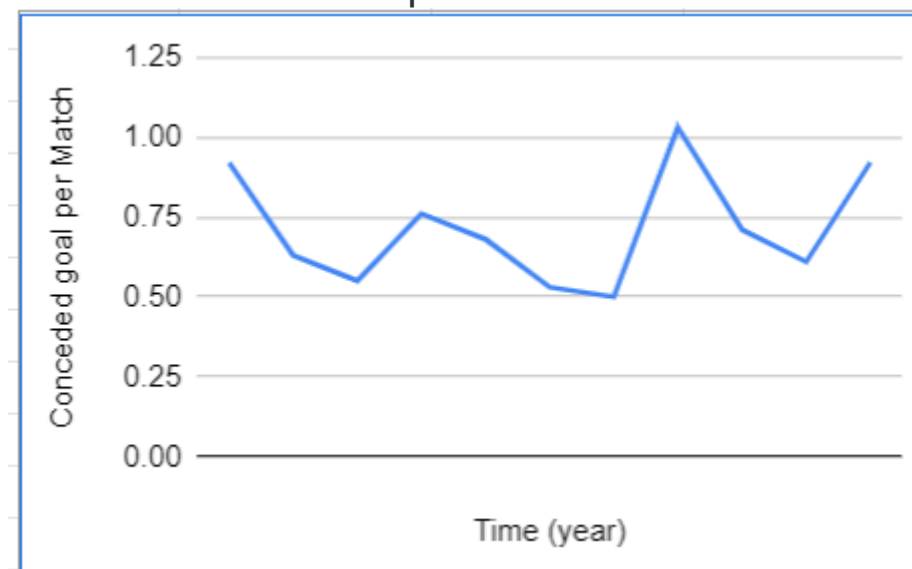
Professional sporting:

- Strategies
- Player recruitment

Team	Year	Conceded goal	Matches	Conceded/match
Barcelona (Spain)	2008	35	38	0.92
	2009	24	38	0.63
	2010	21	38	0.55
	2011	29	38	0.76
Bayern Munich (Germany)	2013	23	34	0.68
	2014	18	34	0.53
	2015	17	34	0.5
Manchester City (England)	2016	39	38	1.03
	2017	27	38	0.71
	2018	23	38	0.61
	2019	35	38	0.92



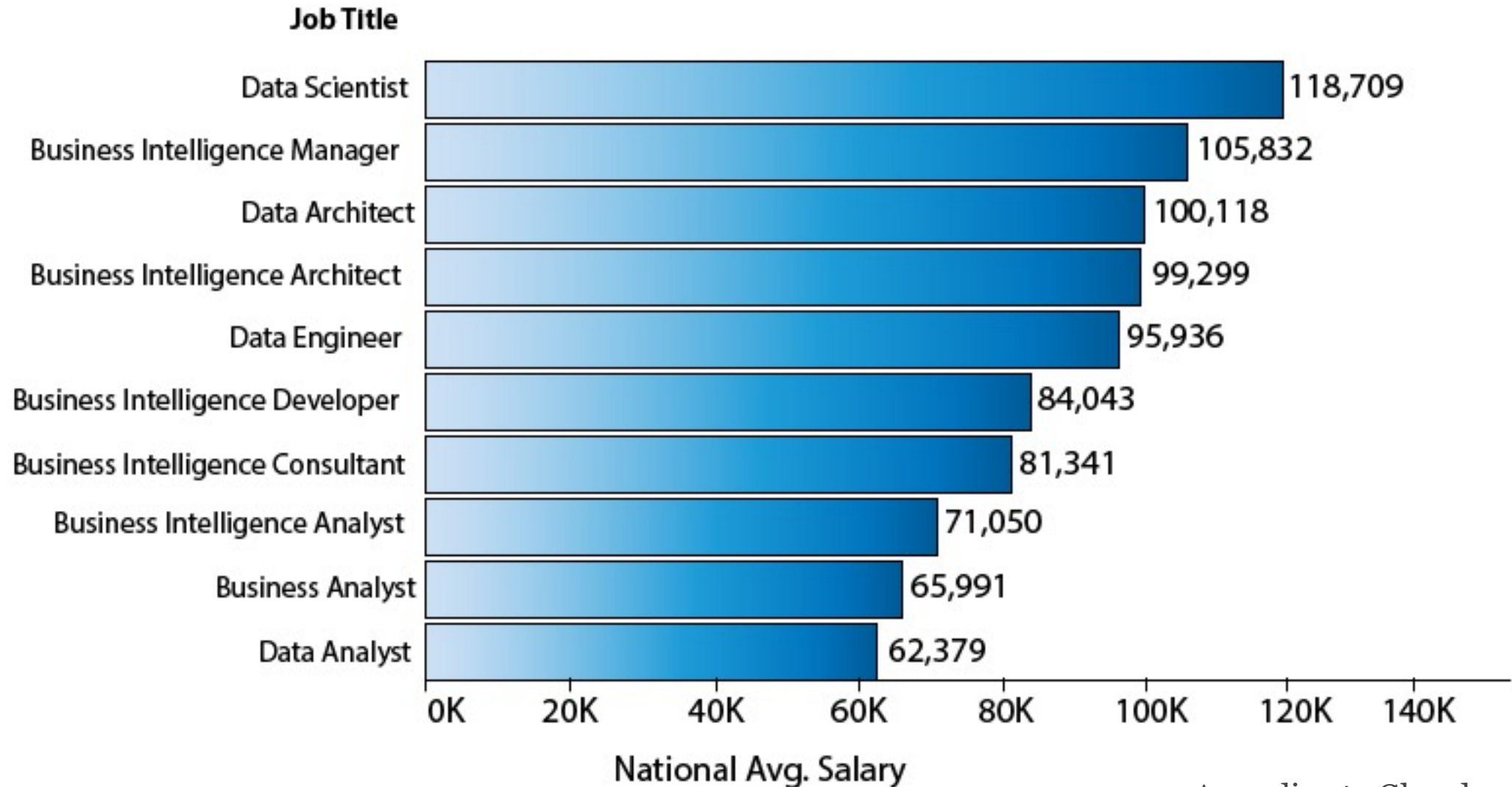
Pep Guardiola



Contents

- ◎ What is Data Science?
- ◎ Skills for Data Science
- ◎ Data Science Process
- ◎ Applications
- ◎ **Why to learn?**
- ◎ About course

Why Learn Data Science?




According to Glassdoor

Why Learn Data Science?

CAREERBUILDER
Global Leader in Recruitment Solutions


data scientist Tất cả ngành nghề Tất cả địa điểm Work from home ☐

29 việc làm Sắp xếp theo ▾




Data Scientist
Công Ty Cổ Phần Giáo Dục Lớ...
\$ Lương: Dưới 55 Tr VND
📍 Hồ Chí Minh
📅 Xem video
📅 Lưu việc làm
📅 14/09/2021

📄 Chế độ bảo hiểm ✈ Du Lịch \$ Chế độ thưởng



Data Scientist
GalaxyOne
\$ Lương: Trên 22 Tr VND
📍 Hồ Chí Minh
📅 Lưu việc làm
📅 07/09/2021

💻 Laptop 📄 Chế độ bảo hiểm ✈ Du Lịch



Senior PHP Developer
Công Ty TNHH MTV Wacontre
\$ Lương: 28,6 Tr - 33 Tr VND
📍 Hồ Chí Minh
📅 Lưu việc làm
📅 29/08/2021

✈ Du Lịch 🚗 Xe đưa đón \$ Chế độ thưởng

Why Learn Data Science?

◎ Main reasons:

- Gain problem solving skills
- High demand
- Analytics is everywhere, everything is data-driven
- Lots of high paying job opportunities



Contents

- ◎ What is Data Science?
- ◎ Skills for Data Science
- ◎ Data Science Process
- ◎ Applications
- ◎ Why to learn?
- ◎ **About course**

Course Assessment

◎ Contacts:

- Le Ngoc Thanh and tutors
- Communication channel: Moodle, Email, Social Network

◎ Course Assessment:

- Final Examination: 50%
- Exercises/Homeworks: 10%
- Projects: 40%

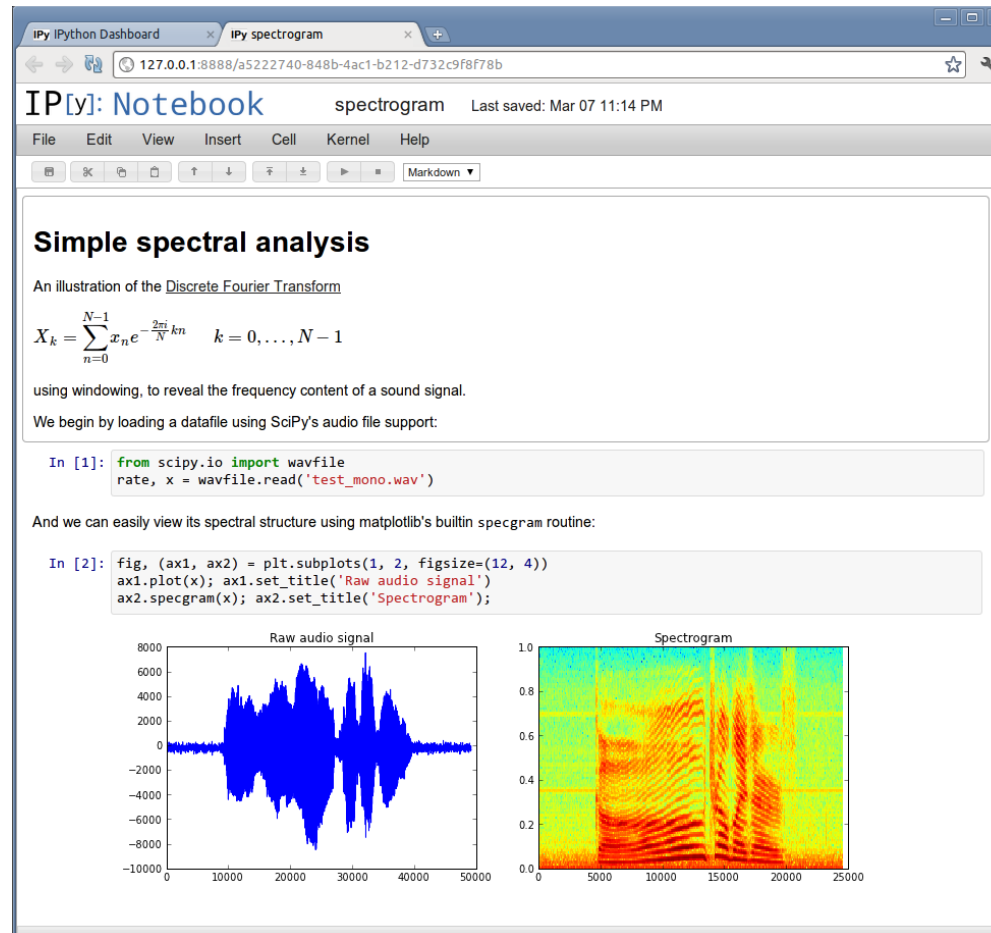
◎ Regularizations:

- Attend classes regularly and have discussions
- All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.

Any plagiarism will be 0 points for course grade.

Tools

- Python and the Libraries
- IPython Notebook



References

- © [MIT \(Edx\), The Analytics Edge \(2018\)](#)
- © [Harvard, Introduction to Data Science \(2018\)](#)
- © [Berkeley, The Foundations of Data Science](#)
- © [Berkeley, Principles and Techniques of Data Science](#)
- © [Andrew Ng \(Coursera\), AI For Everyone](#)
- © [Caltech, Learning from Data](#)