

ÉCOLE SUPÉRIEURE EN SCIENCES ET TECHNOLOGIES DE
L'INFORMATIQUE ET DU NUMÉRIQUE



FUNDAMENTALS OF DATA SCIENCE AND DATA MINING

General Overview

Dr. Chemseddine Berbague

2023-2024

CONTACT

- Dr. Chemseddine Berbague
- Email: berbague@estin.dz
- Google class room: [FDS](#)

ABOUT

- **Teaching objectives:**
 - **Understanding the principals of data mining, data analytics, and data science.**
- **Teaching unit: UEF 1.3**
- **Module 1: Fundamentals of Data Science and Data Mining.**
- **Semester hourly volume : 67h30 (lectures: 1h30, DWs: 1h30, PW: 1h30)**
- **Credits : 5**
- **Coefficient : 3**

CONTENT

- Introduction to Data Science.
- Process of Data Science.
- Data Science Tools.
- Basis of Data Mining.
- Graph Mining
- Web Mining.

EVALUATION

- **Exam (60%)**
- **Continuous assessment (40%)**

IMPORTANT DATES

- Lectures starting date: 25/09/2022
- Exam date: ?!
- Module Project:
 - Assignment date: ?!
 - Due date: ?!
- Preliminary exam: ?!



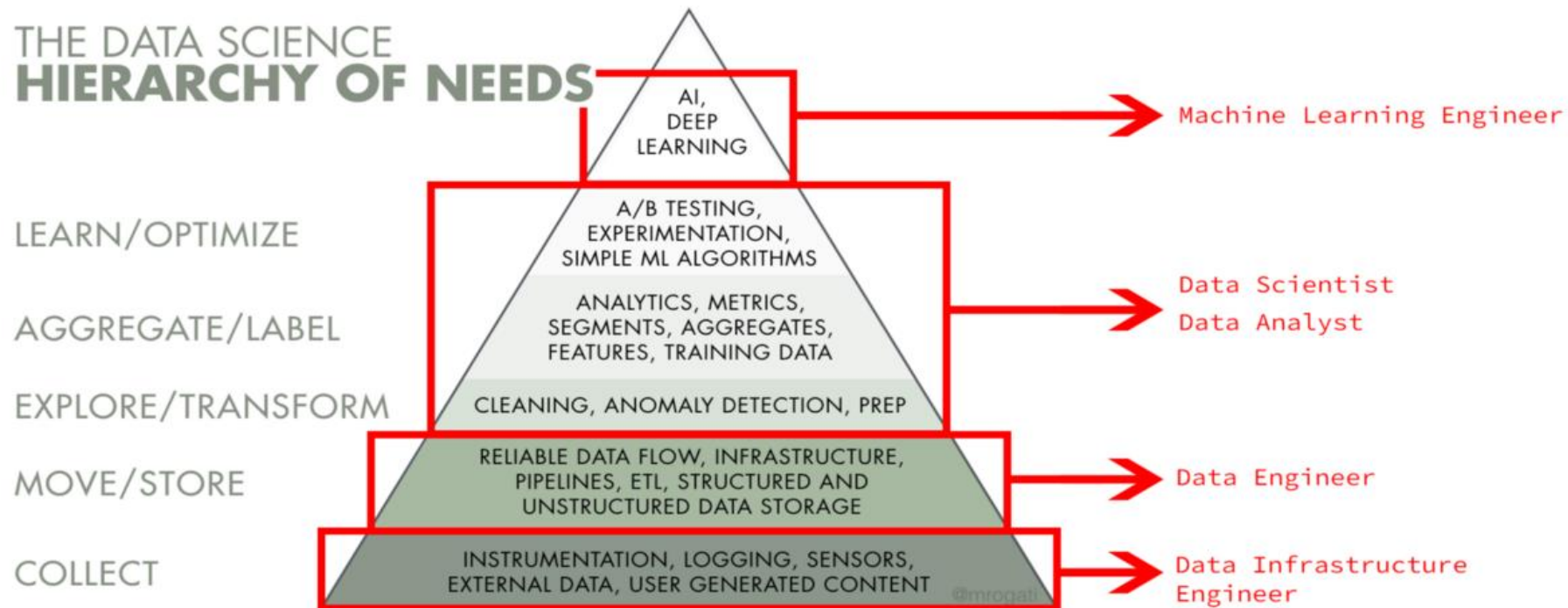
A QUICK LOOK

OVER THE MODULE AND A BIT MORE ...

VERY SHORTLY ...

- **Data Science** is the practice of extracting insights and knowledge from data to inform decision-making and solve problems.
- Throughout the semester, we will delve into the Fundamentals of Data Science: theoretical and practical sides.

DATA SCIENCE: PREREQUISITES, TASKS, AND ROLES



DATA SCIENCE VS STATISTICS VS SOFTWARE ENGINEERING



Josh Wills
@josh_wills



Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

5:55 PM · May 3, 2012 · Twitter Web Client

1,829 Retweets 92 Quote Tweets 2,133 Likes



Dan Nguyen ✓ @dancow · Sep 25, 2012



Replying to @josh_wills

@josh_wills Since scientists are usually not good at software engineering...isn't "Data Engineer" a better title?



1



5



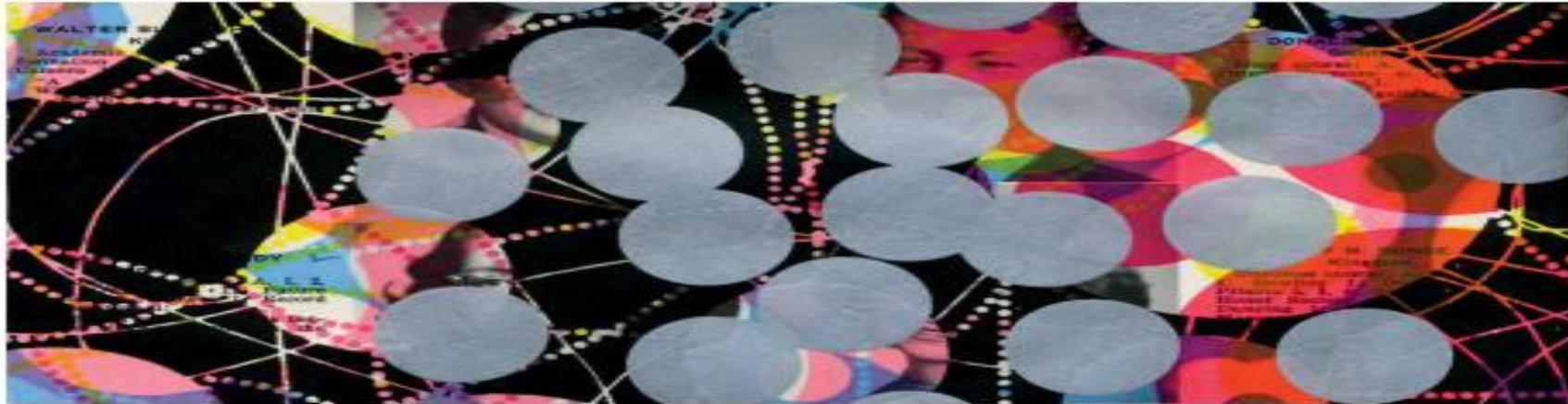
Josh Wills @josh_wills · Sep 25, 2012



@dancow data engineer has come to mean something else, sort of like an ETL developer who needs real software eng experience.

WHY DATA SCIENCE ?

Harvard
Business
Review



DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

WHY DATA SCIENCE ?



**Harvard
Business
Review**

Analytics And Data Science | **Is Data Scientist Still the Sexiest Job of the 21st Century?**

Analytics And Data Science

Is Data Scientist Still the Sexiest Job of the 21st Century?

by Thomas H. Davenport and DJ Patil

July 15, 2022

WHY DATA SCIENCE ?

We have seen both continuity and change in the data science role. It has been remarkably successful in many ways, and some of its challenges — proliferation of related roles, the need for an ethical perspective — result in part from the widespread adoption of data science. The amount of data, analytics, and AI in business and society seem unlikely to decline, so the job of data scientist will only continue to grow in its importance in the business landscape.



DATA SCIENCE

Few statistics around the world !

DATA SCIENCE JOBS OVER THE WORLD

	Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1	Enterprise Architect	\$144,997	4.1/5	14,021
#2	Full Stack Engineer	\$101,794	4.3/5	11,252
#3	Data Scientist	\$120,000	4.1/5	10,071
#4	Devops Engineer	\$120,095	4.2/5	8,548
#5	Strategy Manager	\$140,000	4.2/5	6,977
#6	Machine Learning Engineer	\$130,489	4.3/5	6,801
#7	Data Engineer	\$113,960	4.0/5	11,821

WHY DATA SCIENCE ?



4. Machine Learning Engineer

What they do: Machine learning engineers develop and implement self-running artificial intelligence algorithms and systems for products and applications. | **Most common skills:** **Deep Learning**, **TensorFlow**, Natural Language Processing (NLP) | **Most common industries:** Internet, Computer Software, Information Technology & Services | **Top locations hiring:** San Francisco, Seattle, Los Angeles | **Salary range:** \$72,600-\$170,000 | **Current gender distribution:** 22.3% female; 77.7% male | **Median years of prior experience:** 4 | **Top roles transitioned from:** Software Engineer, **Data Scientist**, Artificial Intelligence Specialist | **Remote job availability:** 19.3%

DATA SCIENCE JOBS IN USA

Percentile	10%	25%	50% (Median)	75%	90%
Hourly Wage	\$ 28.57	\$ 37.32	\$ 48.52	\$ 62.87	\$ 80.31
Annual Wage (2)	\$ 59,430	\$ 77,620	\$ 100,910	\$ 130,770	\$ 167,040

DATA SCIENCE JOBS IN USA

Industry	Employment (1)	Percent of industry employment	Hourly mean wage	Annual mean wage (2)
Computer Systems Design and Related Services	16,620	0.73	\$ 55.91	\$ 116,300
Management of Companies and Enterprises	12,570	0.50	\$ 51.95	\$ 108,060
Management, Scientific, and Technical Consulting Services	7,270	0.47	\$ 51.99	\$ 108,150
Scientific Research and Development Services	5,840	0.73	\$ 54.53	\$ 113,430
Credit Intermediation and Related Activities (5221 and 5223 only)	5,690	0.28	\$ 56.34	\$ 117,190

Industry profile for Data Scientists

DATA SCIENCE JOBS IN USA

Industry	Employment (1)	Percent of industry employment	Hourly mean wage	Annual mean wage (2)
Other Information Services	3,720	1.03	\$ 67.12	\$ 139,600
Scientific Research and Development Services	5,840	0.73	\$ 54.53	\$ 113,430
Computer Systems Design and Related Services	16,620	0.73	\$ 55.91	\$ 116,300
Software Publishers	3,710	0.70	\$ 52.92	\$ 110,070
Data Processing, Hosting, and Related Services	2,570	0.68	\$ 60.65	\$ 126,160

Industries with the highest concentration of employment in Data Scientists

DATA SCIENCE JOBS IN USA

Industry	Employment (1)	Percent of industry employment	Hourly mean wage	Annual mean wage (2)
Computer and Peripheral Equipment Manufacturing	400	0.26	\$ 71.29	\$ 148,290
Semiconductor and Other Electronic Component Manufacturing	620	0.17	\$ 68.34	\$ 142,150
Other Information Services	3,720	1.03	\$ 67.12	\$ 139,600
Data Processing, Hosting, and Related Services	2,570	0.68	\$ 60.65	\$ 126,160
Accounting, Tax Preparation, Bookkeeping, and Payroll Services	1,110	0.11	\$ 59.83	\$ 124,440

Top paying industries for Data Scientists

DATA SCIENCE JOBS IN USA

State	Employment (1)	Employment per thousand jobs	Location quotient (9)	Hourly mean wage	Annual mean wage (2)
California	17,750	1.07	1.43	\$ 64.00	\$ 133,110
New York	10,050	1.16	1.54	\$ 58.91	\$ 122,540
Texas	5,850	0.48	0.64	\$ 54.98	\$ 114,360
North Carolina	5,050	1.14	1.52	\$ 54.80	\$ 113,980
Illinois	4,670	0.83	1.10	\$ 50.00	\$ 104,010

States with the highest employment level

DATA SCIENCE JOBS IN USA

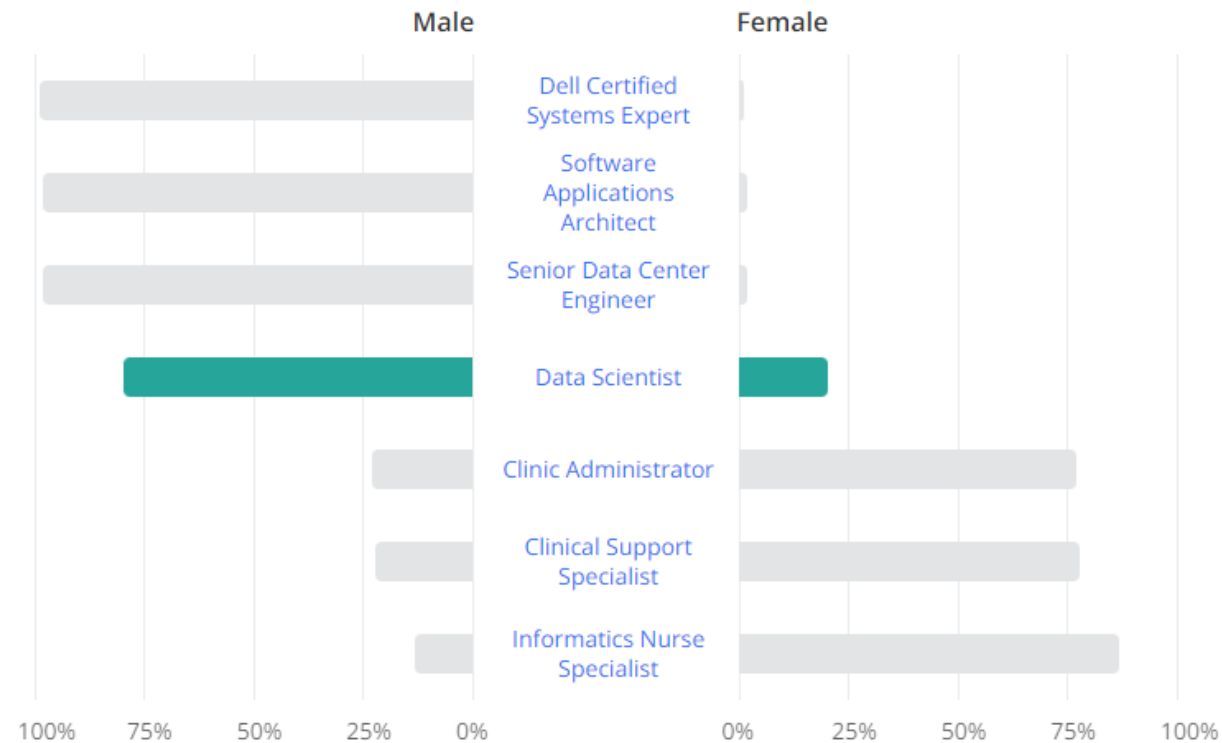
State	Employment (1)	Employment per thousand jobs	Location quotient (9)	Hourly mean wage	Annual mean wage (2)
District of Columbia	1,380	2.09	2.77	\$ 56.48	\$ 117,470
Colorado	3,070	1.17	1.56	\$ 50.86	\$ 105,790
New York	10,050	1.16	1.54	\$ 58.91	\$ 122,540
Utah	1,790	1.15	1.53	\$ 38.38	\$ 79,830
North Carolina	5,050	1.14	1.52	\$ 54.80	\$ 113,980

States with the highest concentration of jobs and location quotients



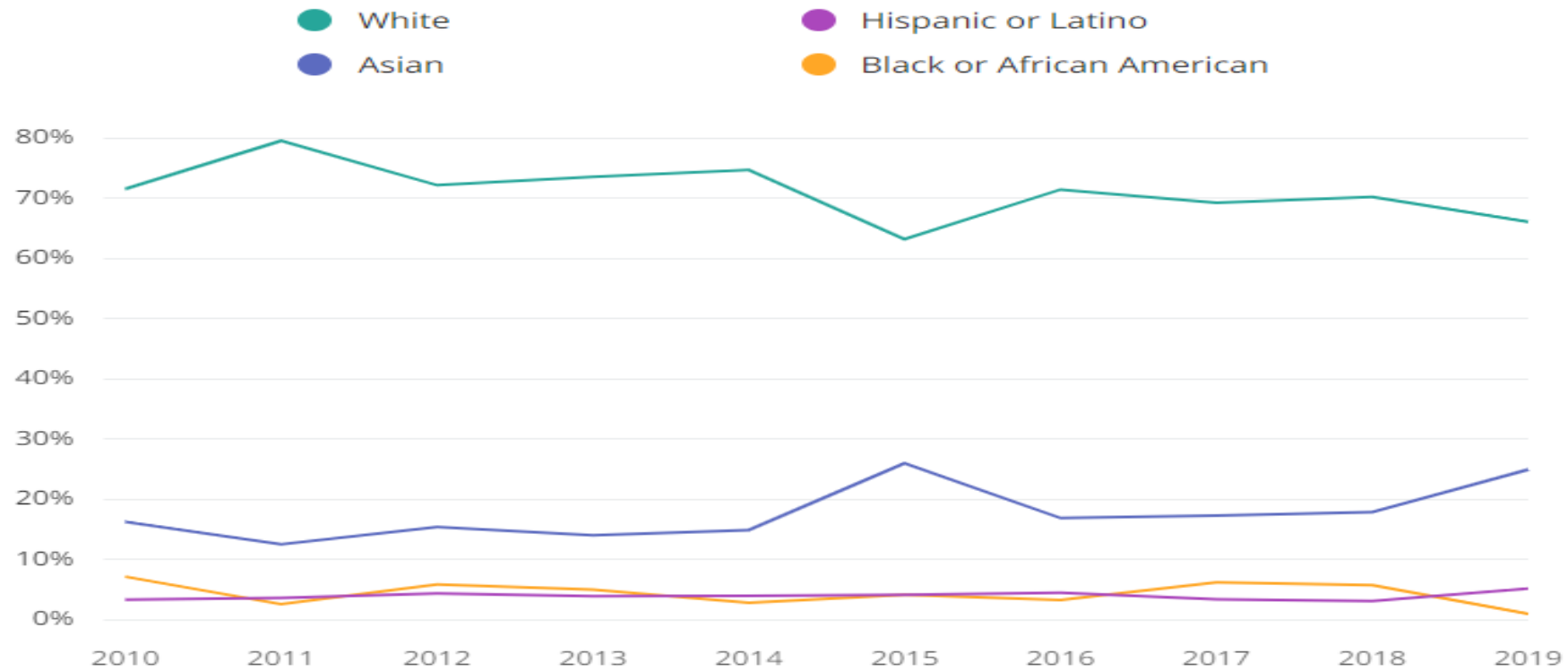
FOR FULL DETAILS, JUST CHECK THIS ARTICLE FROM :
BUREAU OF LABOR STATISTICS

DEMOGRAPHIC OF DATA SCIENCE



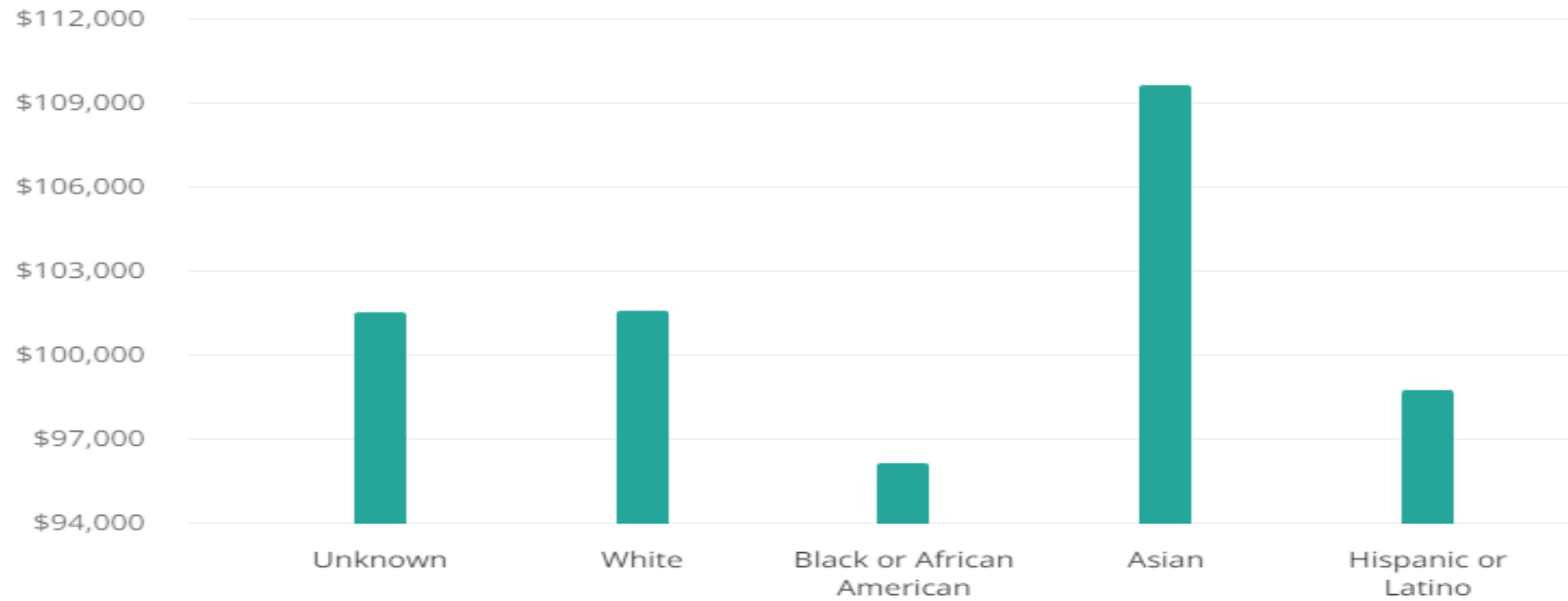
Male to Female Ratio

DEMOGRAPHIC OF DATA SCIENCE



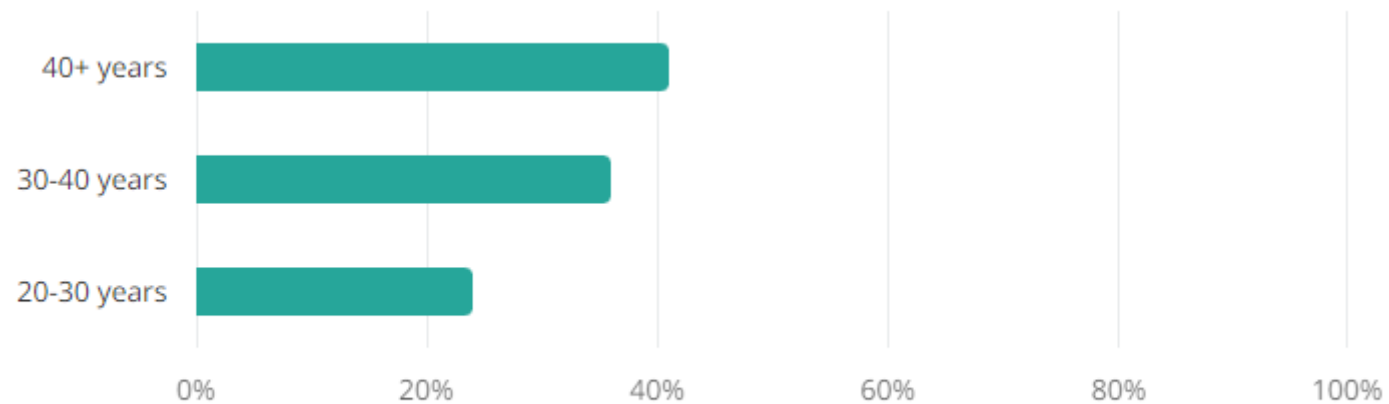
Race And Ethnicity Over Time

DEMOGRAPHIC OF DATA SCIENCE



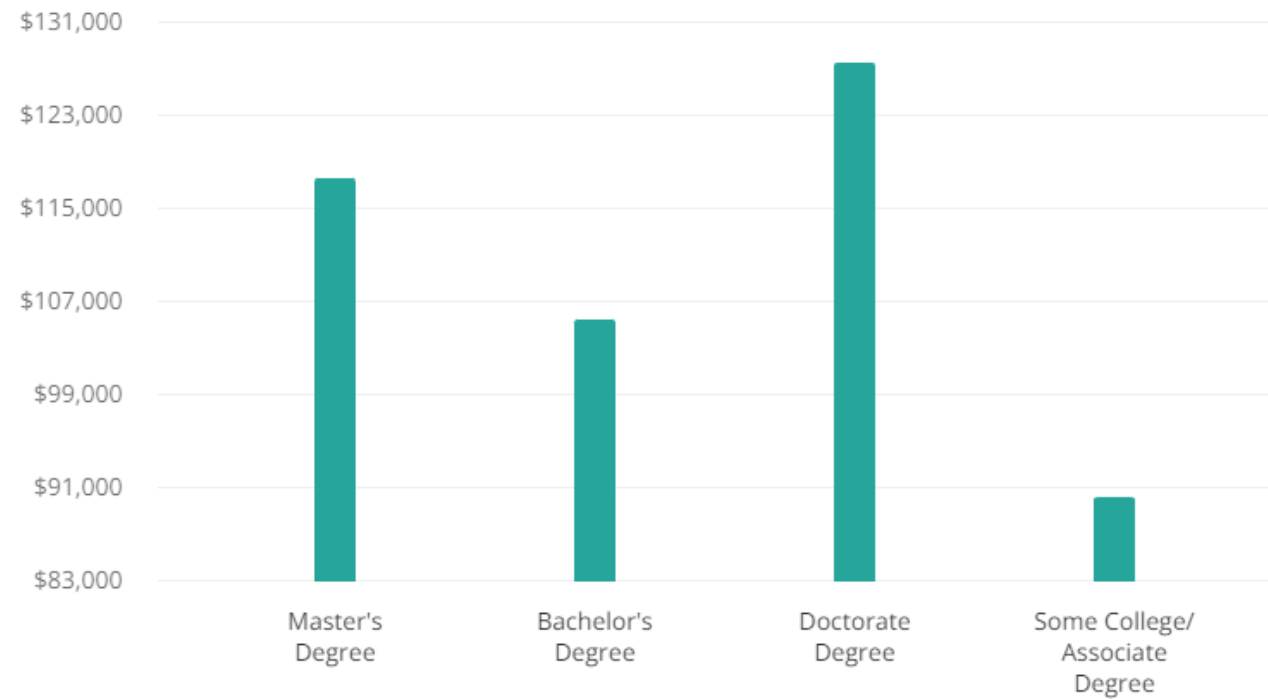
Wage Gap By Race

DEMOGRAPHIC OF DATA SCIENCE



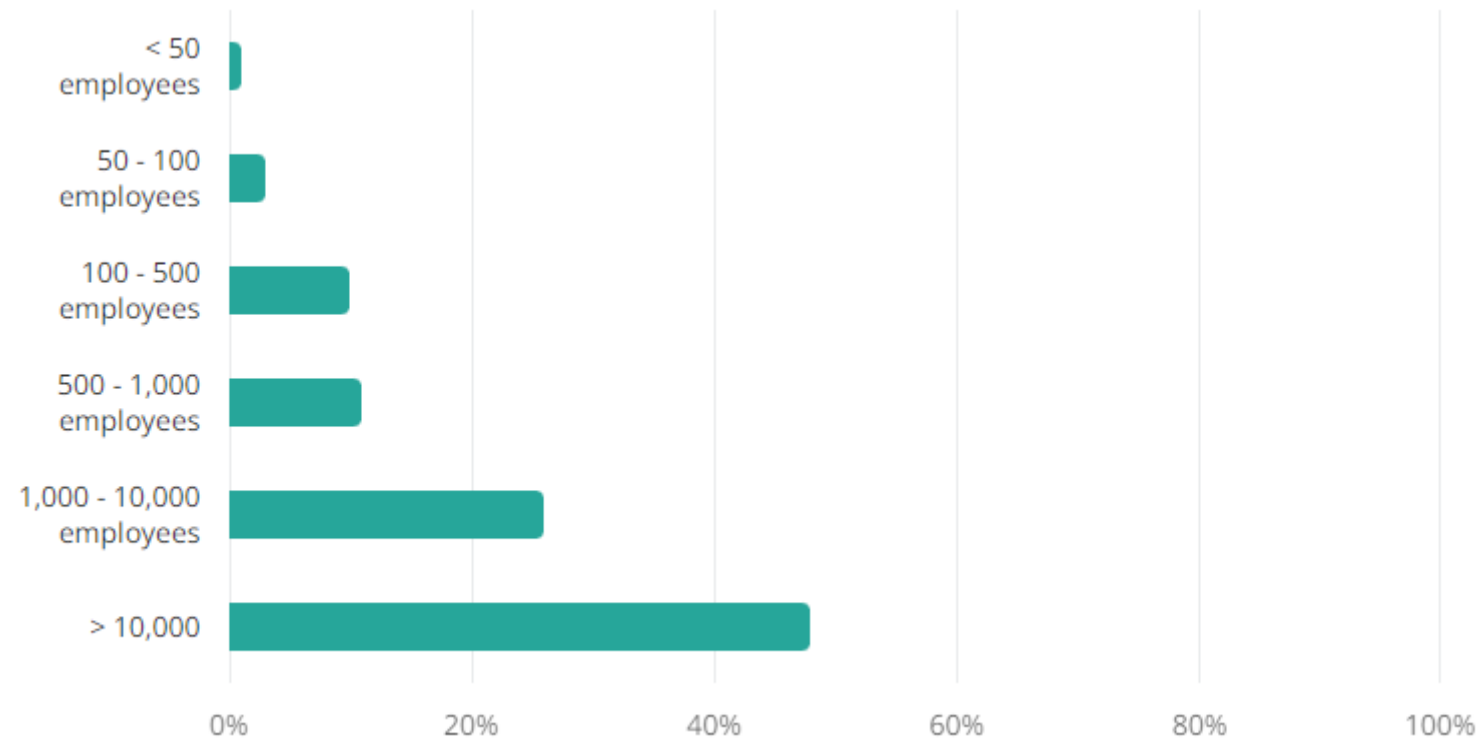
Age Breakdown

DEMOGRAPHIC OF DATA SCIENCE



Wage Gap By Education

DEMOGRAPHIC OF DATA SCIENCE



Jobs By Employer Size

DEMOGRAPHIC OF DATA SCIENCE

Industry ↕	Percentages ↕
Fortune 500	22%
Technology	19%
Start-up	10%
Finance	7%
Internet	5%
Manufacturing	5%
Professional	5%
Media	4%
Health Care	4%

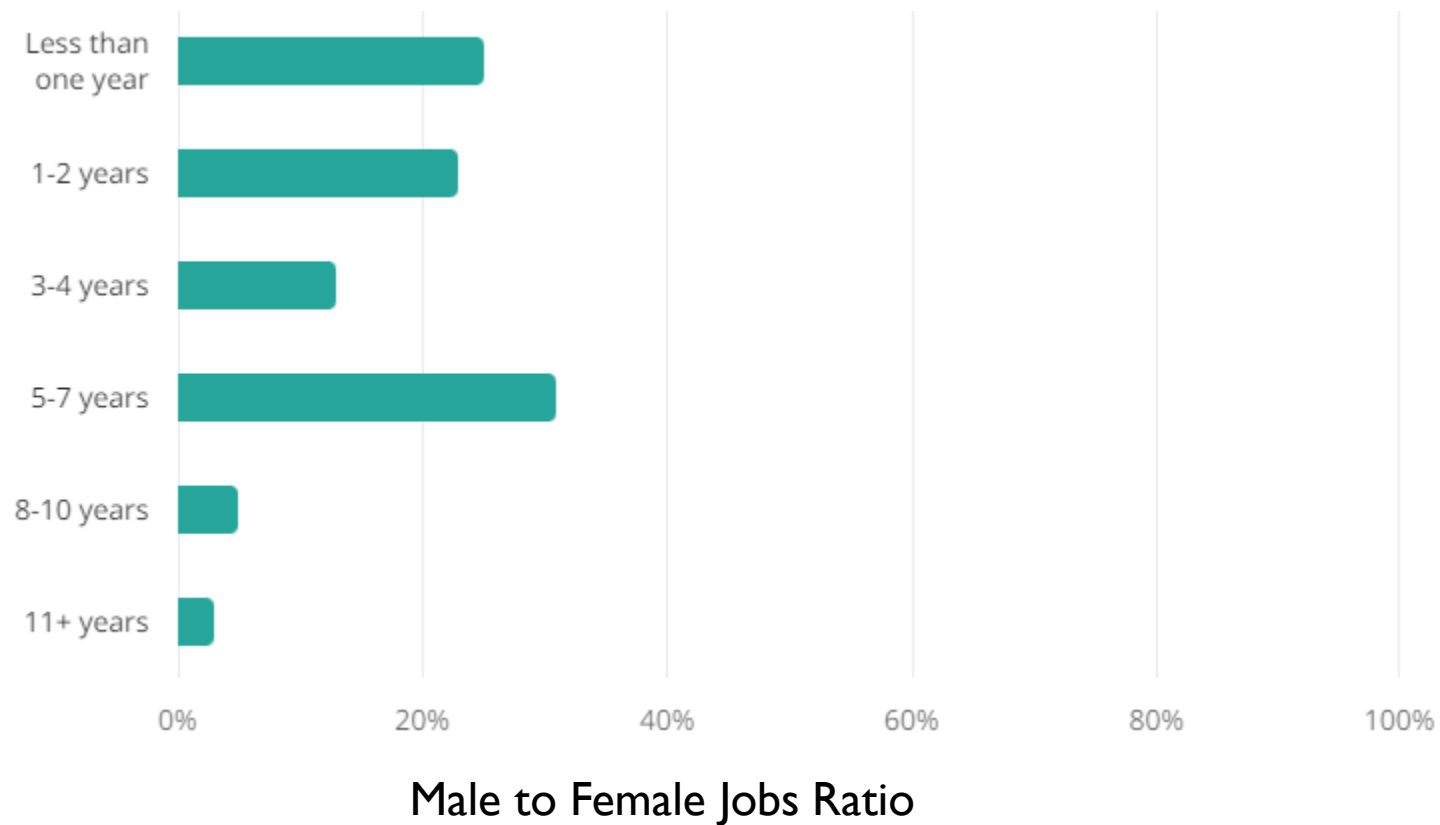
Top Industries Hiring Data Scientists

DEMOGRAPHIC OF DATA SCIENCE

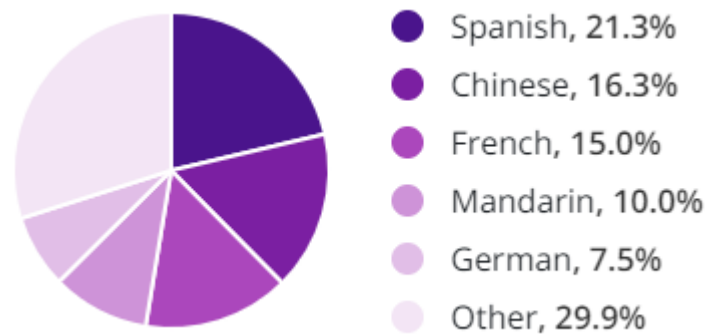


Unemployment Rate Over Time

DEMOGRAPHIC OF DATA SCIENCE



DEMOGRAPHIC OF DATA SCIENCE



Languages spoken by data scientists



FOR FULL DETAILS CHECK THIS ARTICLE FROM


ZIPPIA











DATA SCIENCE JOBS

IN ALGERIA

INQUIRING ON LINKEDIN (09/2023)





Jobs

Date Posted

Experience Level

Company

Job Type

On-site/Remote


Easy Apply

All filters

Data scientist jobs in Algeria

15 results

Set alert ☐




Freelance Data Engineer

Toptal

Algeria (Remote)

☒ You have a preferred skill badge


3 hours ago



Freelance Embedded C++ Developer

Toptal


Algeria (Remote)


 1 connection works here


3 hours ago


Freelance Data Engineer

Toptal · Algeria (Remote) 3 hours ago · 2 applicants

 Contract · Mid-Senior level

 1,001-5,000 employees · Software Development

 1 connection

 See recent hiring trends for Toptal - Second Feed. [Try Premium for free](#)

☒ You have a preferred skill badge

Apply

Save

INQUIRING ON EMPLOITIC (09/2023)

The screenshot displays the Emploitic website interface. At the top, the navigation bar includes the logo, links for 'Offres d'emploi', 'Découvrez les entreprises', and 'Blog', along with buttons for 'Se connecter' and 'Espace recruteur'. The main section is titled 'Emploi data scientist' and shows '70 Offres d'emploi trouvées'. It features a search bar with 'data scientist' entered, a location dropdown set to 'Région, wilaya ...', and a blue 'Rechercher' button. A sidebar on the left contains a 'Votre avis' button. Below the search bar, there's a 'Filtrez par' section with a 'Réinitialisez' link. The first job listing is for 'Data Scientist Expert' at Ooredoo, located in 'Alger, Algérie', posted on '05 Septembre', and marked as 'Confirmé / Expérimenté (6 à 10 ans)'.

emploitic. Offres d'emploi Découvrez les entreprises Blog

Se connecter Espace recruteur

Emploi data scientist

70 Offres d'emploi trouvées

data scientist

Région, wilaya ...

Rechercher

Créer Alerte emploi

Filtrez par Réinitialisez







Métier / Fonction

Data Scientist Expert

Ooredoo

Alger, Algérie 05 Septembre Confirmé / Expérimenté (6 à 10 ans)

CERTIFICATIONS IN DATA SCIENCE

<p>EXAM</p> <p>Exam MB-260: Microsoft Customer Data Platform Specialist</p> <p>Azure • Data Engineer • Intermediate</p>  <p>+ Save</p>	<p>CERTIFICATION</p> <p>Microsoft Certified: Customer Data Platform Specialty</p> <p>Exam MB-260 Azure • Data Engineer • Intermediate</p>  <p>+ Save</p>	<p>EXAM</p> <p>Exam DP-100: Designing and Implementing a Data Science Solution on Azure</p> <p>Azure • Data Scientist • Intermediate</p>  <p>+ Save</p>
<p>CERTIFICATION</p> <p>Microsoft Certified: Azure Data Scientist Associate</p> <p>Exam DP-100 Azure • Data Scientist • Intermediate</p>  <p>+ Save</p>	<p>CERTIFICATION</p> <p>Microsoft Certified: Azure AI Fundamentals</p> <p>Exam AI-900 Azure • AI Engineer • Beginner</p>  <p>+ Save</p>	<p>EXAM</p> <p>Exam AI-900: Microsoft Azure AI Fundamentals</p> <p>Azure • AI Engineer • Beginner</p>  <p>+ Save</p>

Microsoft Azure certifications in data science

CERTIFICATIONS IN DATA SCIENCE

Most popular Trending



Data Science A-Z™: Real-Life Data Science Exercises Included

Kirill Eremenko, Ligency I Team, Ligency Team

4.6 ★★★★★ (32,209)

\$11.99 ~~\$84.99~~



Deep Learning Prerequisites: Linear Regression in Python

Lazy Programmer Inc.

4.7 ★★★★★ (5,421)

\$12.99 ~~\$89.99~~

Bestseller



[2022] Machine Learning and Deep Learning Bootcamp in...

Holczer Balazs

4.2 ★★★★★ (944)

\$16.99 ~~\$119.99~~



Taming Big Data with Apache Spark and Python - Hands On!

Sundog Education by Frank Kane, Frank Ka...

4.5 ★★★★★ (13,567)

\$11.99 ~~\$24.99~~

Bestseller



Recursion, Backtracking and Dynamic Programming in Python

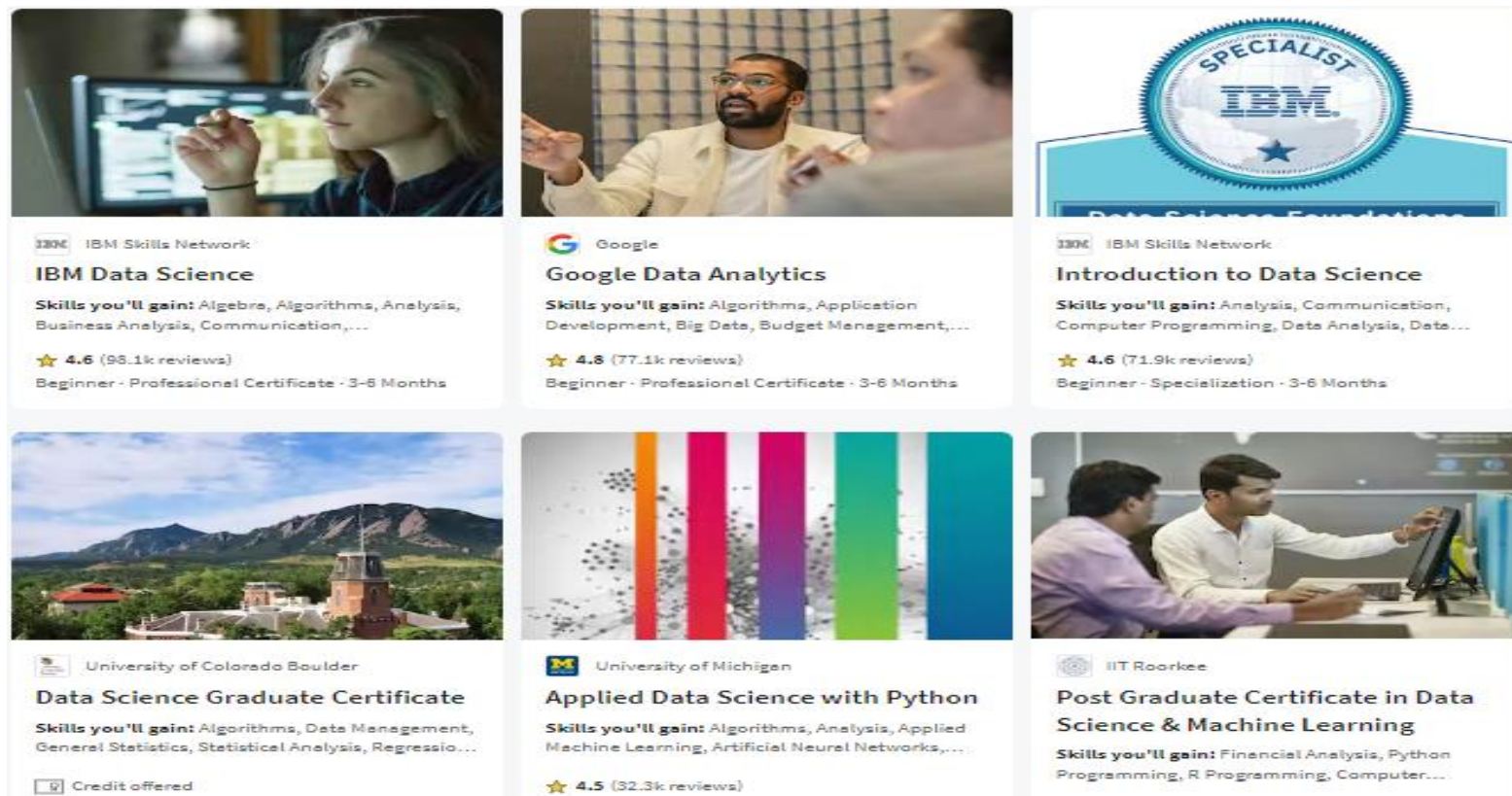
Holczer Balazs

4.6 ★★★★★ (692)

\$11.99 ~~\$84.99~~

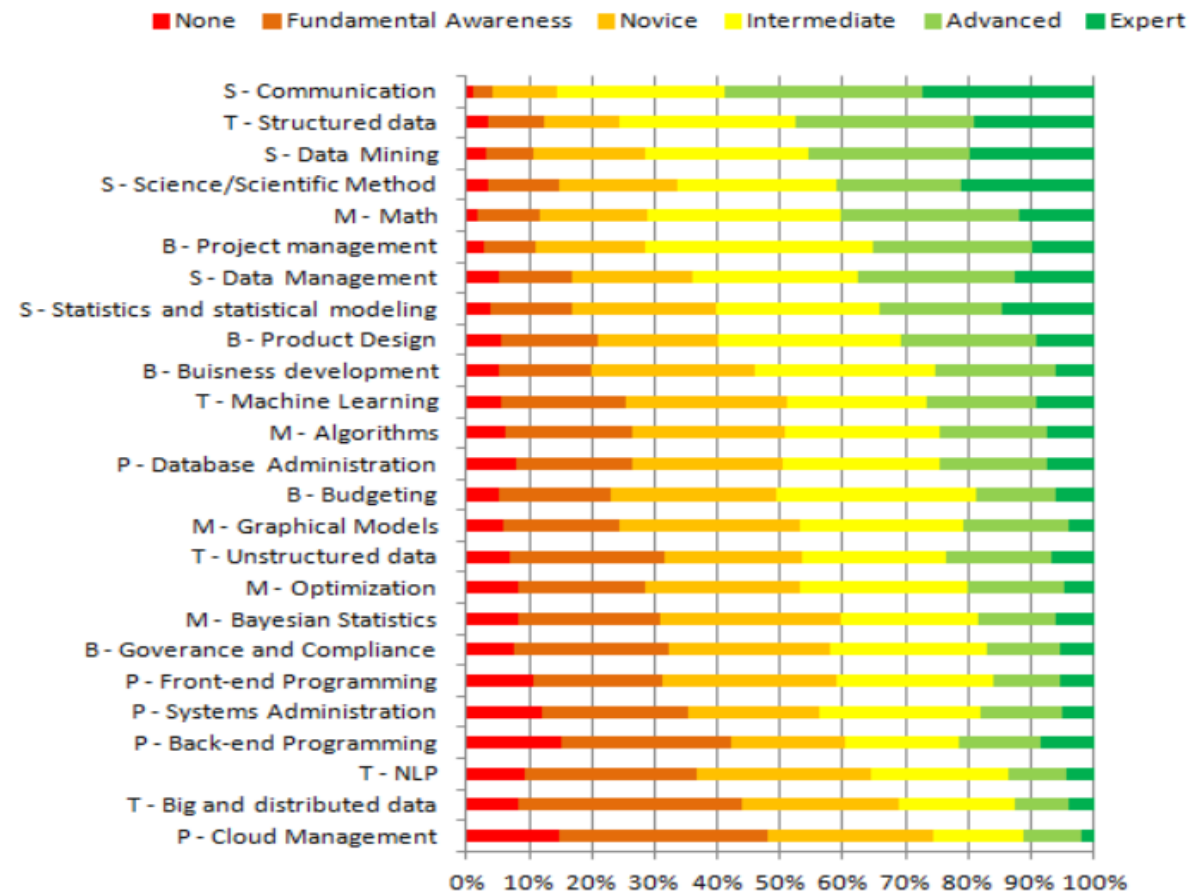
Udemy certifications in data science

CERTIFICATIONS IN DATA SCIENCE

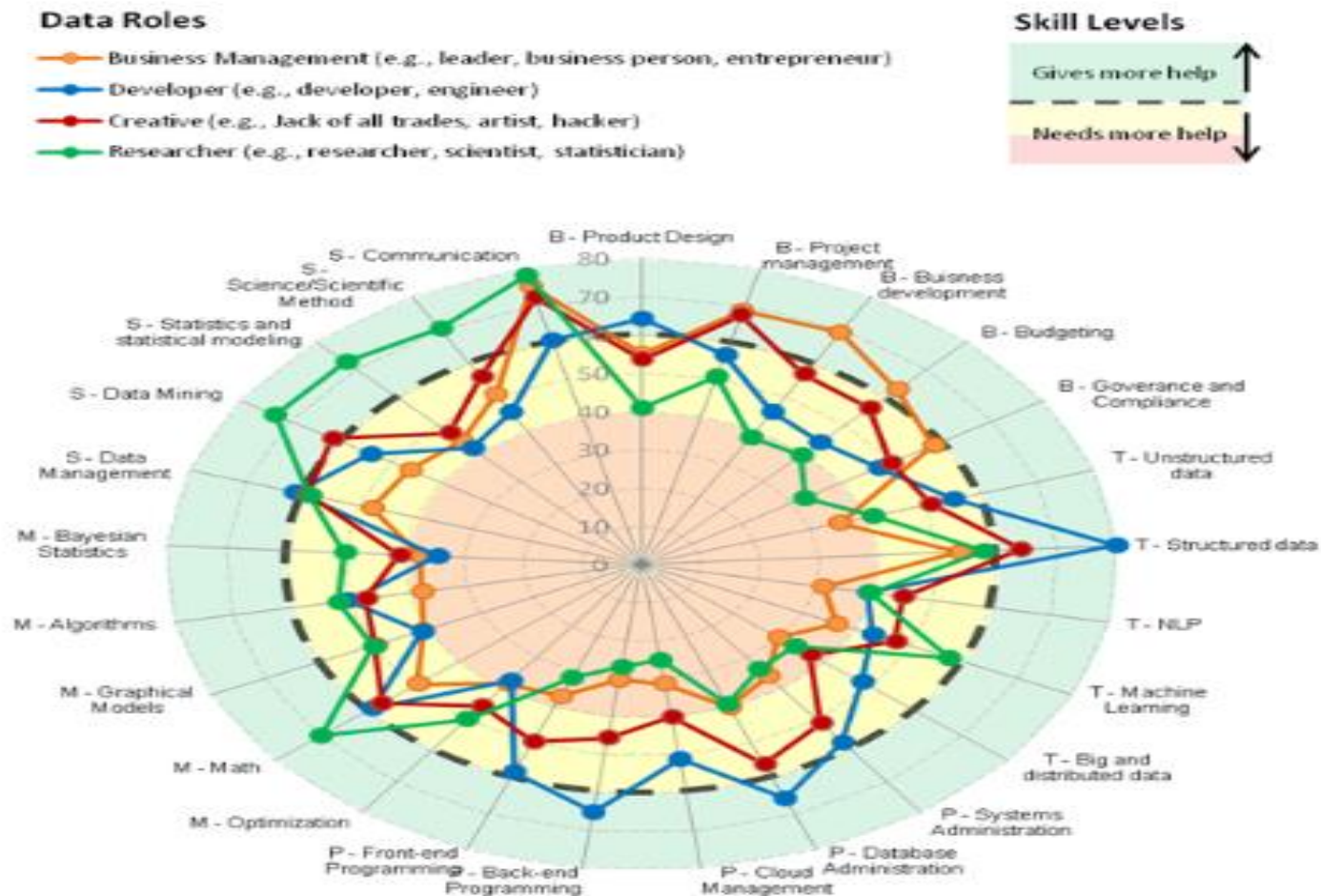


Coursera certifications in data science

PROFICIENCY IN DATA SCIENCE SKILLS



PROFICIENCY IN DATA SCIENCE SKILLS BY JOB ROLE



MAIN REFERENCES

- <https://kaggle.com/learn/>
- <https://linktr.ee/datasciencetools>

REFERENCES

- Dietrich, D., “Data science & big data analytics: discovering, analyzing, visualizing and presenting data”, Wiley, 2015.
- Lutz, M., & Biernat, E., “Data Science: fondamentaux et études de cas: Machine Learning avec Python et R”, Editions Eyrolles, 2015
- <http://data8.org/>
- <https://learn.microsoft.com/en-us/azure/architecture/>