

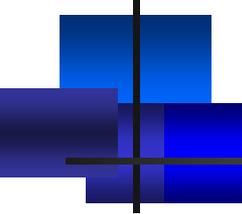
데이터사이언스

전 성 해

청주대학교 데이터사이언스학과

<http://delab.cju.ac.kr>

shjun@cju.ac.kr



데이터 사이언스 : 정의

Data science is to study data.

데이터 사이언스는 데이터를 연구(공부)하는 것이다.

데이터 사이언스

컴퓨터학

통계학

(하드웨어 + 소프트웨어)

(데이터분석 + 수리통계)

데이터사이언스

- 소프트웨어: 컴퓨터 언어(R, Python), 데이터베이스(SQL)
- 데이터분석: 확률통계, 머신러닝(인공지능)



Data Scientist Salary

Overall Score 6.6 / 10

#4 in Best Technology Jobs | #7 in Best STEM Jobs | #8 in 100 Best Jobs

Overview

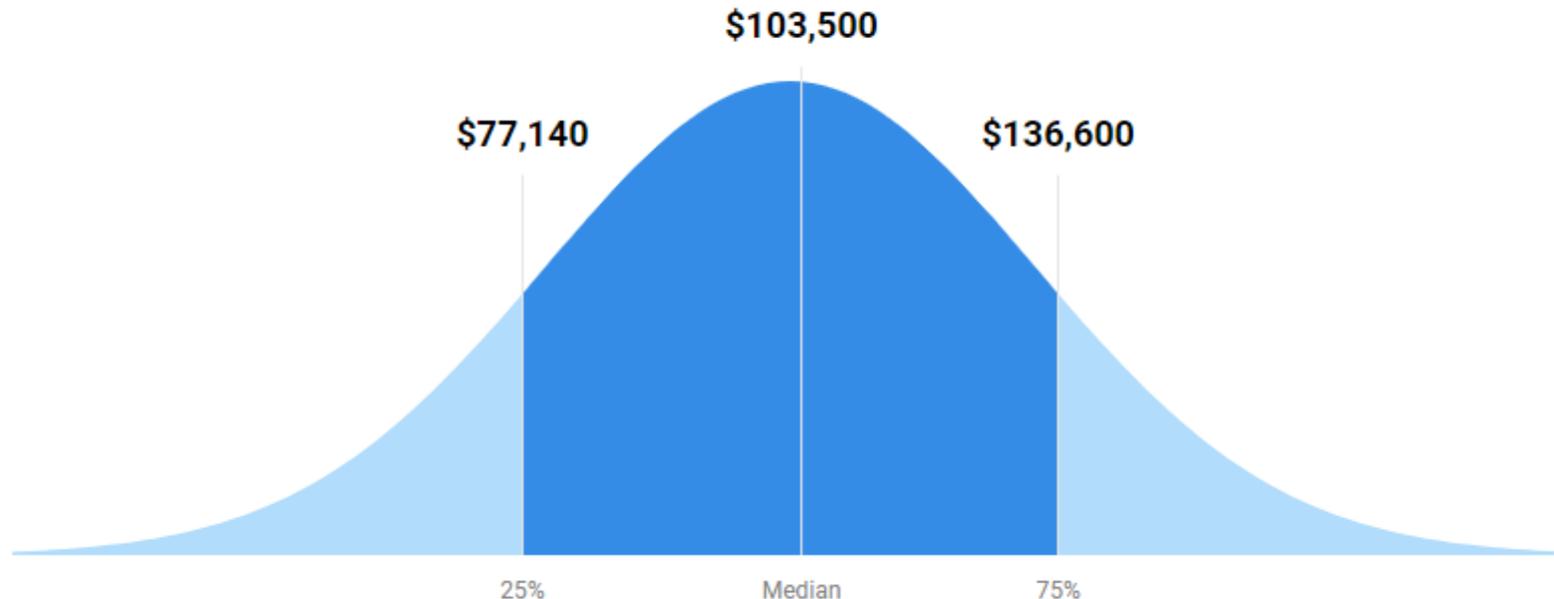
Salary

Reviews and Advice

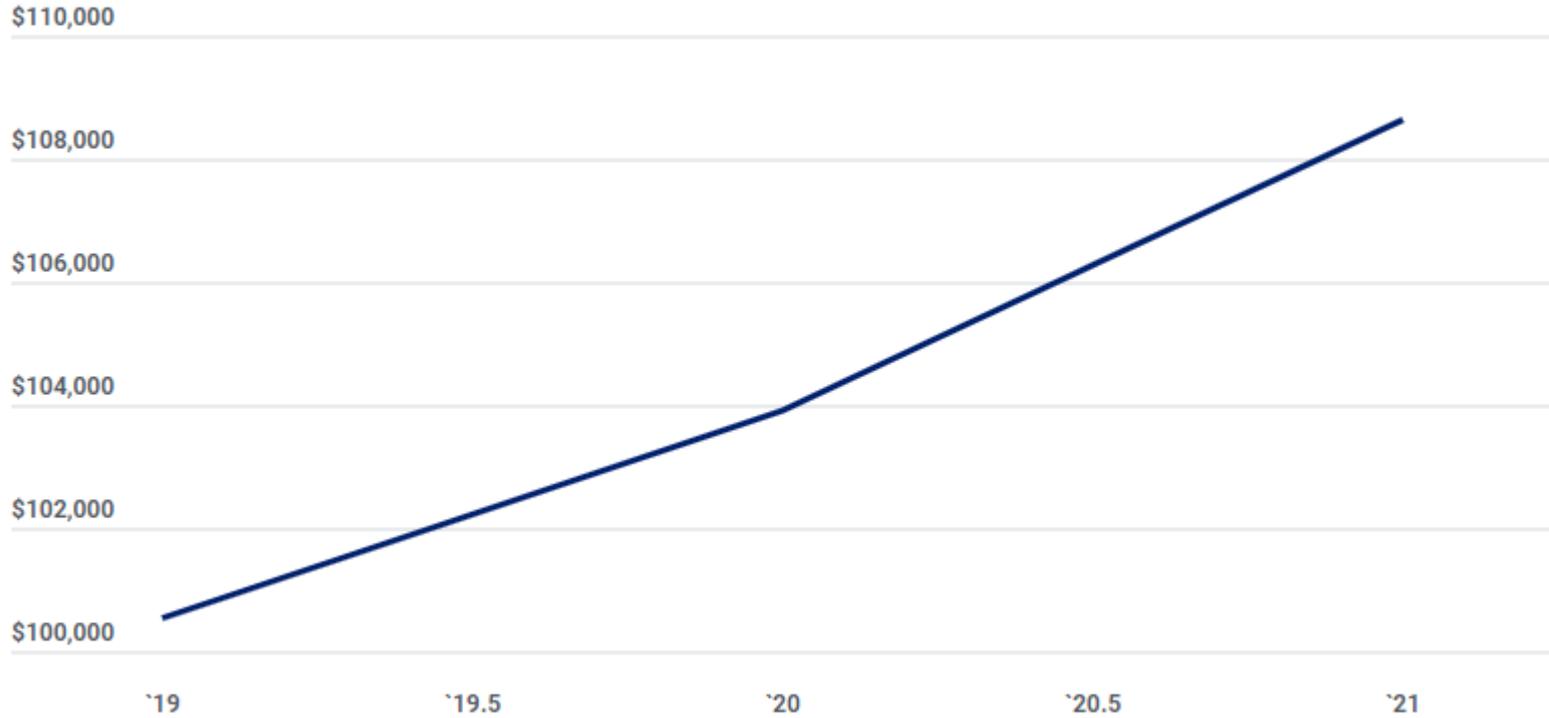
Job Openings

How Much Does a Data Scientist Make?

Data Scientists made a median salary of \$103,500 in 2022. The best-paid 25% made \$136,600 that year, while the lowest-paid 25% made \$77,140.



Salary Outlook



X is Year and Y is Mean Salary

<https://money.usnews.com/careers/best-jobs/data-scientist/salary>



데이터사이언티스트 연봉



전체 이미지 동영상 뉴스 쇼핑 도서 지도 : 더보기

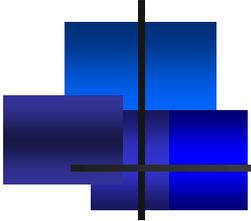
카카오

네이버

토스

미국에서는 데이터 사이언티스트의 평균 연봉이 11만 달러 이상이며, 경력이 쌓이면 15만 달러 이상의 연봉을 받는 경우가 많습니다. 한국에서는 데이터 사이언티스트의 평균 연봉이 6000만 원 이상이며, 경력이 쌓이면 1억 원 이상의 연봉을 받는 경우도 있습니다.

2023. 5. 13.



Data

People

소통

언어: R / Python

Data

Data – 라틴어 어원: datum (주다)

: 우리 눈 앞에 주어진 모든 것 (given things)

: 세상의 흔적

데이터 학습과 지능

Learning: 학습 (데이터 분석)



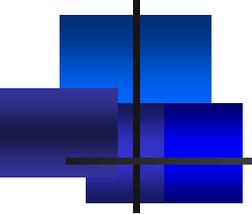
Big or Small Data

Artificial Intelligence

Statistics

Computer Science

Data Science



Data & Intelligence

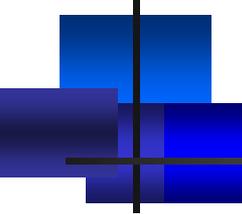
- Intelligence

- 지능(知能)

- 도전적인 새로운 과제를 성취하기 위하여 사전(과거)지식과 경험을 적용할 수 있는 능력
 - **새로운 대상이나 상황에** 부딪혀 그 의미를 이해하고 합리적인 적응 방법을 알아내는 지적 활동의 능력

**“Data Science enables AI,
AI justifies Data Science.”**

**데이터사이언스는 AI를 가능하게 하고,
AI는 데이터사이언스를 정당화합니다.**



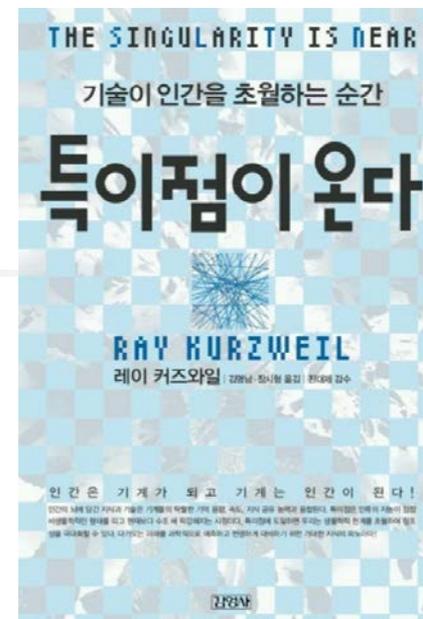
Google(구글링), ChatGPT

- 문제 해결

- 실력이 뛰어난 **전문가가** 문제 해결 ↓
- **일반인도** 구글 검색이나 ChatGPT를 통하여 문제 해결 ↑
 - **AI의 도움을 받아 자신의 업무를 수행 (ChatGPT, ...)**

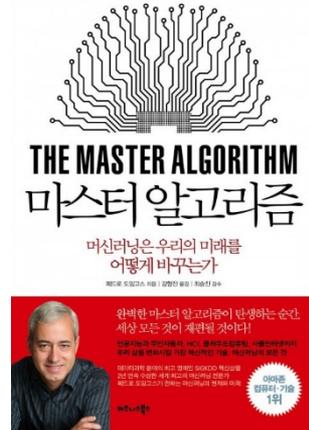
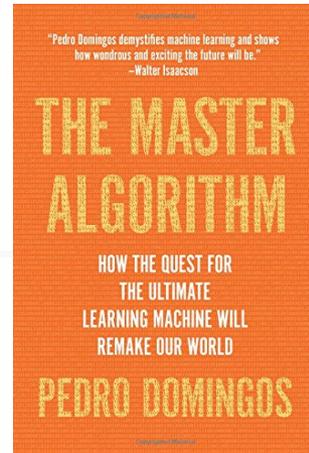
특이점 (Singularity)

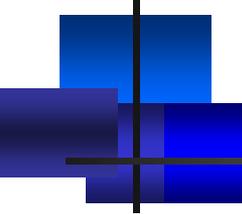
- 특이점
 - 상식적으로 이해할 수 없는 일이 벌어지는 상황 (순간)
- 레이 커즈와일 (Ray Kurzweil)
 - 발명가, 사상가, 미래학자
 - The Singularity Is Near (2005년)
- AI 특이점: **2045년**
 - 인간은 기계가 되고 **기계는 인간이 된다.**
 - AI → AGI (artificial general intelligence)
 - Weak AI → Strong AI



마스터 알고리즘: 머신러닝

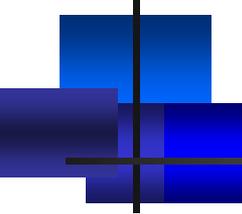
- 페드로 도밍고스 (Pedro Domingos)
 - 워싱턴 대학교 컴퓨터학과 교수
 - Machine learning과 우리의 미래
- Book: The Master Algorithm (마스터 알고리즘)
 - 여러 가지 학습 알고리즘의 통합**
 - 데이터 (data science) + 마스터 알고리즘 (machine learning)
→ AGI, Strong AI





Google vs. OpenAI

- 구글 (1998년) – 래리 페이지, 세르게이 브린
 - 구글의 생각: **Do not be evil**. 사악해 지지 말자 → **Do the right thing**.
 - 구글의 AI 빅데이터: Google Data Center, 클라우드 컴퓨팅, 알파고, Youtube, 지메일, ...
- OpenAI (2015년) – 샘 올트먼, 일론 머스크
 - Level 1 (Chatbots) 인간과 소통할 수 있는 AI
 - **Level 2 (Reasoners) 박사수준의 전문가 만큼 문제해결 능력을 갖춘 AI**
 - Level 3 (Agents) 인간을 대신하여 수 일동안 작업을 수행할 수 있는 AI
 - Level 4 (Innovations) 새로운 생각으로 혁신을 이끌어낼 수 있는 AI
 - Level 5 (Organizations) 조직의 업무를 수행할 수 있는 AI



Machine Learning

- **Machine learning**

- Machine learning is a field of computer science that uses statistical techniques to give computer systems **the ability to learn from data, without being explicitly programmed.**

- 명시적 프로그램 없이 데이터로부터 학습할 수 있는 능력을 컴퓨터에게 제공하는 통계적 기법

- 머신러닝의 목적은 기계(컴퓨터)가 지능을 갖도록 하는 것

- Artificial Intelligence

- AI is to make computers intelligent.

- 머신러닝은 AI 구현의 핵심적 역할을 수행함

통계학 정의

- Statistics is the art of **learning from data.**

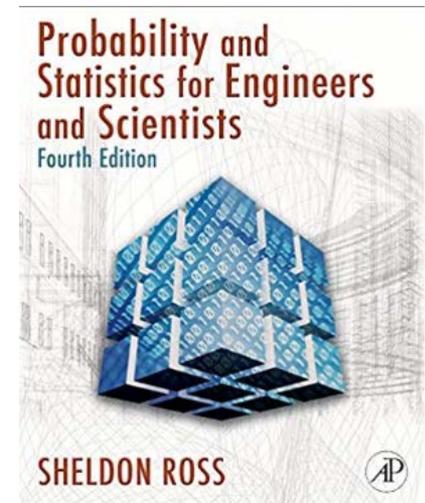
Professor S. M. Ross (UC Berkeley)

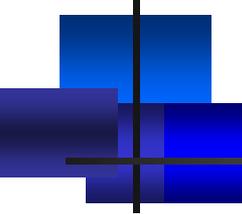
Statistician

Data Miner

Data Scientist

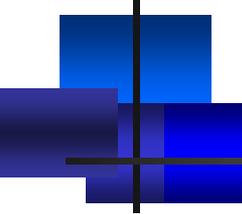
Data Artist





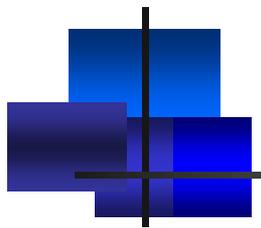
AI 시대 인간의 위치

- Job이 있는 사람
 - AI에 영향을 받지 않는 사람
 - AI를 부리는 사람 (AI를 지배(개발, 유지보수)하는 전문가)
 - AI에 의해 부림(지배)을 당하는 사람
- Job이 없는 사람



청주대 - 데이터사이언스 학과

- 대학교 (학부)
 - 공과대학
 - 데이터사이언스학과
- 대학원
 - 일반대학원
 - 빅데이터통계학과
 - 빅데이터/데이터사이언스전공
 - 응용통계전공



$$\int_0^2 (4x - 2x^2) dx = \left[2x^2 - \frac{2}{3}x^3 \right]_{x=0}^{x=2} = 8 - \frac{16}{3} = \frac{8}{3} = 2.666667$$

f=function(x) 4*x-2*x^2
integrate(f,0,2)

```
> f=function(x) 4*x-2*x^2  
> integrate(f, 0, 2)  
2.666667
```