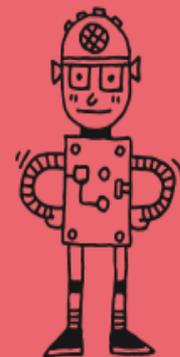


Data Scientist로 성장하기

인하대학교 공과대학
소프트웨어융합공학 연계전공

김승환



Data Science 란?

- 데이터 사이언스는 2000년대에 급부상한 개념임
- 컴퓨터의 발달로 적은 비용으로 많은 데이터를 저장, 관리가 가능함
- 보유한 데이터를 활용해보자~ 문제는 어떻게?
- 가장 Hot 한 직업군으로 볼 수 있음



- 데이터 사이언스는 통계학도 아니고 컴퓨터공학도 아님
- 다만, 통계학과 컴퓨터 공학에 대한 이해가 필수적임

빅데이터 수요 맞추려면 전문가 양성 서둘러야

Let's Master 빅데이터 (6) 전문인력 양성

데이터 개발·분석전문가 등
분야별로 세분화해 육성을

미국은 84개 프로그램에
총 2억달러 예산 투입

英·日도 전문가 양성 서둘러

미국계 빅데이터 전문업체인 맵알테크놀로지스 등 전문업체들이 잇따라 국내 시장에 상륙하면서 빅데이터 시장이 서서히 달궈지고 있다. 전문가들에 의하면, 양적으로도 2015년 3000억원, 2020년 1조원 규모로 성장할 전망이다. 국내 빅데이터 업체들도 치열한 한판 승부를 각오하고 있지만 애로사항이 점점산중이다.

보안신문 (주)파이터코리아

주요뉴스

- 1 롯데 유통
- 2 11.2조 일
- 3 "볼거리!"
- 4 치솟는 집
- 5 문재인 디
- 6 울산서 A
- 7 애플 '10.5
- 8 美 증시 기
- 9 위기의 퀘
- 10 "경유차

기업의 입장: 구슬이 서말이라도 꿰어야 보배다. 어떻게 꿰어야 할까?

수많은 정보를
분석하는 능력

해당 분야에
대한 풍부한
지식

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS



DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

데이터 관리,
처리 능력

알아낸 정보를
효과적으로
전달

1

Data Science 란?

미래 성장 산업을 주도할 인력이 필요함(지식형 인재 필요)



Data Science 란?

- 2019년 미국 Best Job 1위가 Data Scientist 임

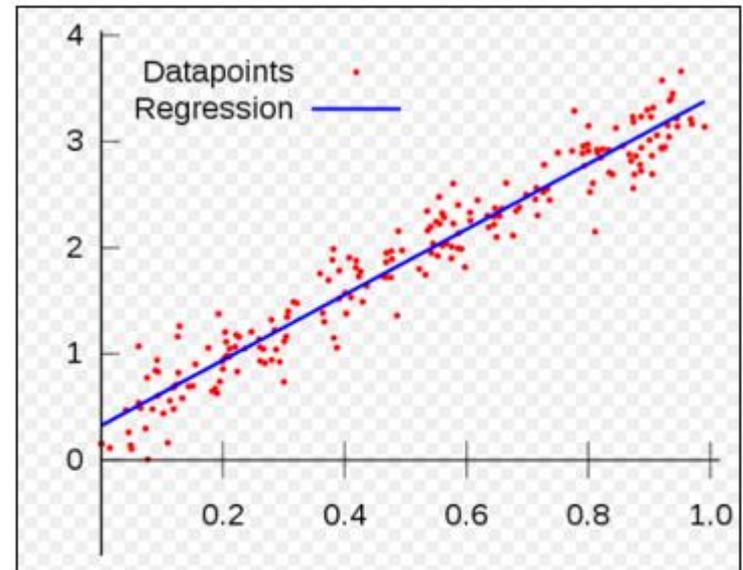
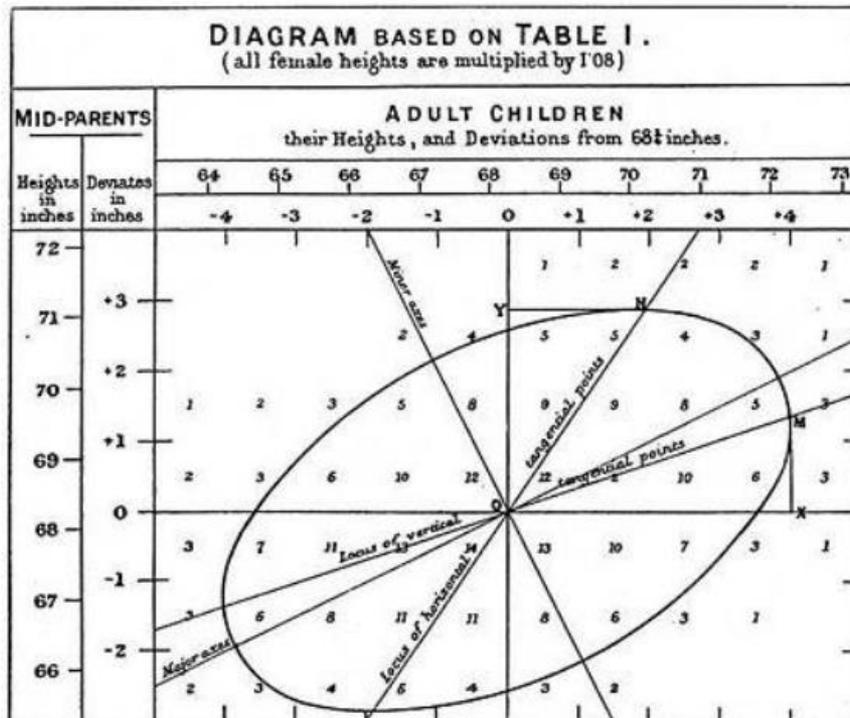
50 Best Jobs in America for 2019

Best Jobs | 2019 | United States | Share |    

Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1 Data Scientist	\$108,000	4.3/5	6,510	View Jobs
#2 Nursing Manager	\$83,000	4/5	13,931	View Jobs
#3 Marketing Manager	\$82,000	4.2/5	7,395	View Jobs
#4 Occupational Therapist	\$74,000	4/5	17,701	View Jobs
#5 Product Manager	\$115,000	3.8/5	11,884	View Jobs
#6 Devops Engineer	\$106,000	4.1/5	4,657	View Jobs
#7 Program Manager	\$87,000	3.9/5	14,753	View Jobs

원조 Data Science, 통계학

- 1875년, Francis Galton은 아래의 그림으로 부모와 자식의 키의 관계를 설명했음
- 이 때 발표된 논문이 [“regression toward the mean”](#) ,데이터과학의 시초로 볼 수 있음
- 회귀분석(Regression) → 신경망 모형 → 딥러닝 모형으로 발전해 왔음

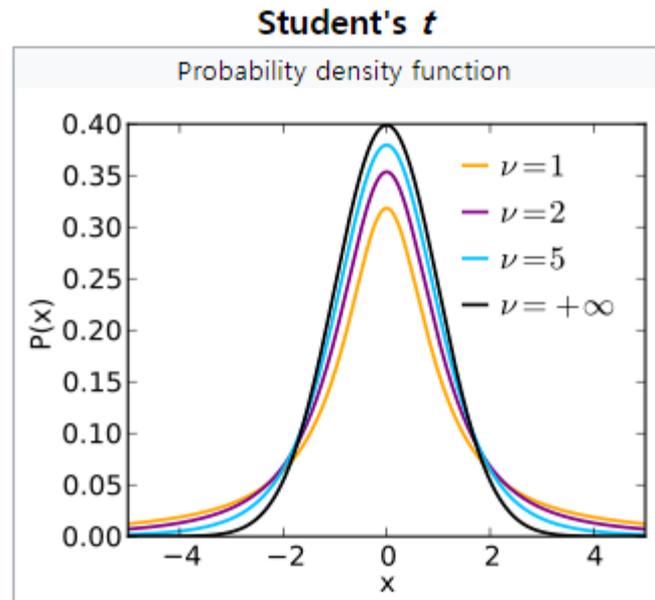


$$Y = \alpha + \beta x + \epsilon$$

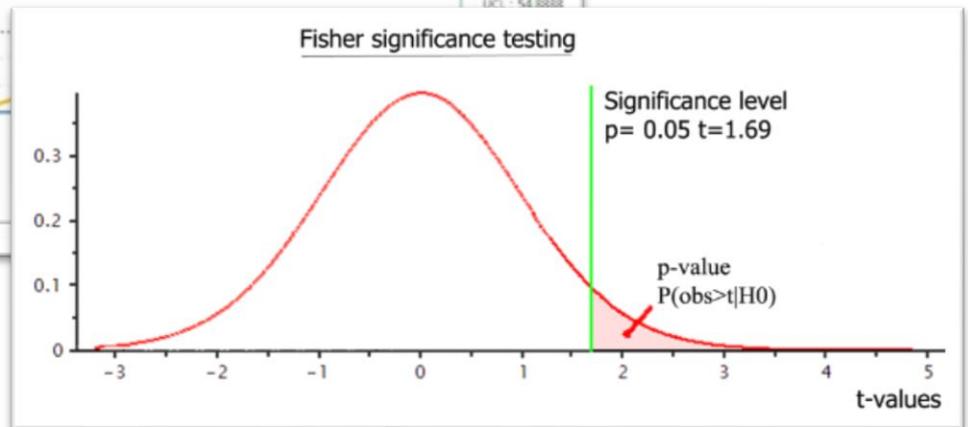
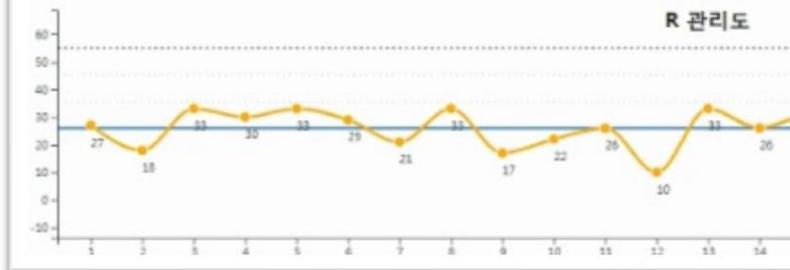
- 1908년, 기네스 맥주회사에 근무하던 Gosset은 정규분포로부터 t 분포를 만들
- t분포를 이용한 t-Test는 두 집단의 데이터가 같다고 볼 수 있는지 알아내는 과학적 접근으로 많은 학문의 발전에 기여하였음
- 이는 소규모 실험에서 얻은 정보로 최적의 의사결정을 할 수 있도록 하는 내용임
- 현재까지 과학에서 가장 많이 사용하는 기법으로 볼 수 있음



Statistician William Sealy Gosset, known as "Student"



- 품질관리는 데이터로부터 생산 공정의 이상여부를 탐지하는 과학으로 대표적인 원조 데이터 사이언스로 볼 수 있음
- 이 역시, 통계학임



Computer 분야 발전

- Computer와 통신기술의 발달은 많은 데이터를 생산/유통 가능케 함
- 데이터는 클라우드 시스템을 통해 공유되어 빅데이터 화
- 빅데이터는 집단지성의 힘으로 작용하여 미래 Biz.의 핵심적 경쟁요소로 Position 됨
- 빠르게, 정확하게 계산 → 빅데이터를 통한 지능화(A.I 로 발전)

1964년

IBM 360 Computer
개발

1993년

인터넷 보급
(PC 통신)

2007년

애플 아이폰 개발

1983년

IBM PC 개발

1998년

초고속 인터넷 &
Ubiquitous 시대

2010년 이후

IoT, 빅데이터,
클라우드, A.I

Computer

통신기술

Data

Computer 분야 발전

- 구글, 페이스북, 유튜브 등 빅데이터를 보유한 기업이 성장을 주도함
- 빅데이터를 관리하고 운영하는 시스템 아키텍처 분야
- 빅데이터를 분석하고 이를 통해 지식을 창출하는 분야



데이터 크기의 증가

비용도 증가



Computer 분야 발전

Northwind Customers - Microsoft Excel

ID	Company	Last Name	First Name	E-mail	Adi.	Job Title	Business Home Phc	Mobile Fax	Numb	Address	City	State/Province	ZIP/Postal Code	Court
1	Company A	Bedecks	Anna			Owner	(123555-0100			123 1st Street	Seattle	WA	98999	USA
2	Company B	Grabosky	Antonio			Owner	(123555-0100			123 2nd Street	Boston	MA	98999	USA
3	Company C	Awen	Thomas			Purchasin	(123555-0100			123 3rd Street	Los Angeles	CA	98999	USA
4	Company D	Lee	Christina			Purchasin	(123555-0100			123 4th Street	New York	NY	98999	USA
5	Company E	O'Donnell	Marin			Owner	(123555-0100			123 5th Street	Minneapolis	MN	98999	USA
6	Company F	Pérez-Ola	Francisco			Purchasin	(123555-0100			123 6th Street	Milwaukee	WI	98999	USA
7	Company G	Xie	Ming-Yang			Owner	(123555-0100			123 7th Street	Boise	ID	98999	USA
8	Company H	Andersson	Elisabeth			Purchasin	(123555-0100			123 8th Street	Portland	OR	98999	USA
9	Company I	Mortensen	Sven			Purchasin	(123555-0100			123 9th Street	Salt Lake City	UT	98999	USA
10	Company J	Wascker	Poland			Purchasin	(123555-0100			123 10th Street	Chicago	IL	98999	USA
11	Company K	Krschne	Peter			Purchasin	(123555-0100			123 11th Street	Miami	FL	98999	USA
12	Company L	Edwards	John			Purchasin	(123555-0100			123 12th Street	Las Vegas	NV	98999	USA
13	Company M	Ludick	Andre			Purchasin	(123555-0100			123 13th Street	Memphis	TN	98999	USA
14	Company N	Gilo	Carlos			Purchasin	(123555-0100			123 14th Street	Denver	CO	98999	USA
15	Company O	Kupkova	Helena			Purchasin	(123555-0100			123 15th Street	Honolulu	HI	98999	USA
16	Company P	Goldschm	Daniel			Purchasin	(123555-0100			123 16th Street	San Francisco	CA	98999	USA
17	Company Q	Bogel	Jean Philippe			Owner	(123555-0100			123 17th Street	Seattle	WA	98999	USA
18	Company R	Aulter	Michael			Purchasin	(123555-0100			123 18th Street	Boston	MA	98999	USA
19	Company S	Egger	Alexander			Accountn	(123555-0100			123 19th Street	Los Angeles	CA	98999	USA
20	Company T	Li	George			Purchasin	(123555-0100			123 20th Street	New York	NY	98999	USA
21	Company U	Tham	Bernard			Accountn	(123555-0100			123 21th Street	Minneapolis	MN	98999	USA
22	Company V	Ramos	Luciana			Purchasin	(123555-0100			123 22th Street	Milwaukee	WI	98999	USA
23	Company W	Eatin	Michael			Purchasin	(123555-0100			123 23th Street	Portland	OR	98999	USA
24	Company X	Hasselbe	Jonas			Owner	(123555-0100			123 24th Street	Salt Lake City	UT	98999	USA
25	Company Y	Rodman	John			Purchasin	(123555-0100			123 25th Street	Chicago	IL	98999	USA
26	Company Z	Liu	Run			Accountn	(123555-0100			123 26th Street	Miami	FL	98999	USA
27	Company AA	Toh	Karen			Purchasin	(123555-0100			123 27th Street	Las Vegas	NV	98999	USA
28	Company BB	Ragheb	Amr			Purchasin	(123555-0100			123 28th Street	Memphis	TN	98999	USA
29	Company CC	Lee	Soo Jung			Purchasin	(123555-0100			123 29th Street	Denver	CO	98999	USA

정형 Data

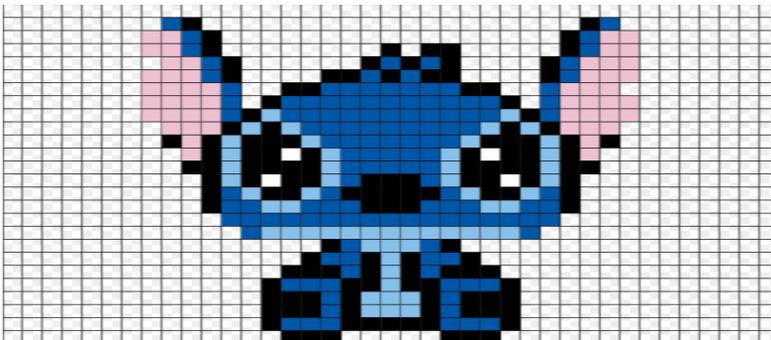


Image Data

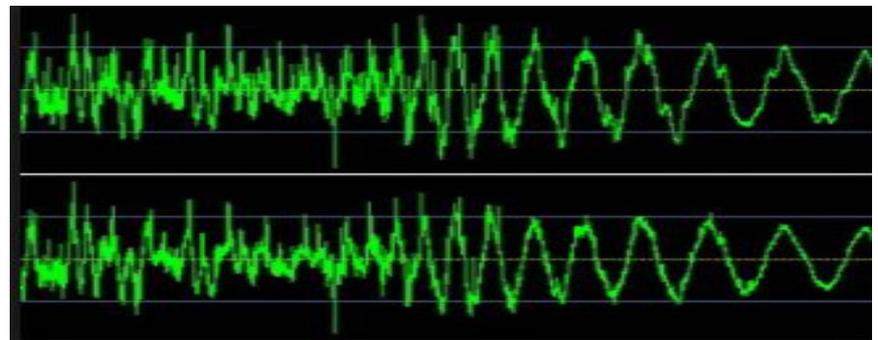
samplelog.log

```

1 #Software: Microsoft Internet Information Services X.X-
2 #Version: X-
3 #Date: 2010-03-24 07:00:01-
4 #Fields: date time s-sitename s-computername s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username
5 2010-03-24 07:00:01 ZZZZC941948879 RUFFLES 222.222.222.222 GET / - 80 - 220.181.7.113 HTTP/1.1
6 2010-03-24 07:00:23 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/12/im_not_mean_im_just_ar
7 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-blank.gif - 80 - 217.
8 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /grep-options.gif - 80 - 217.25
9 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-cat.gif - 80 - 217.25
10 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-pwd-cd.gif - 80 - 217.
11 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /robots.txt - 80 - 95.55.207.95
12 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /rss-short.xml - 80 - 173.45.25
13 2010-03-24 07:00:43 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/08/22-things-you-dont-knc
14 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /screen.css - 80 - 98.88.35.133
15 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/rss-header-red.gif - 80 -
16 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/logo.jpg - 80 - 98.88.35.1
17 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/input-emailsend.jpg - 80 -
18 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /images/cm-ebook-banner.gif - 8
19 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg.jpg - 80 - 98.88.35.133
20 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg-top.jpg - 80 - 98.88.35.
21 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/checkout-login.gif -
22 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/topnav-contact.jpg - 80 -
23 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/portent-email-sub.gif
24 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /rss-header.jpg - 80 - 98.88.35

```

Log Data



Sound Data

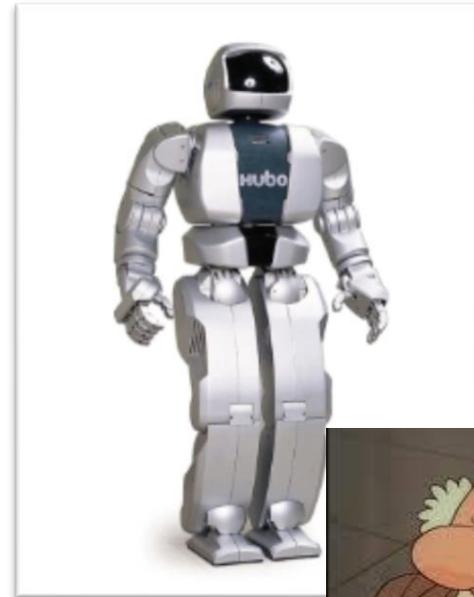
Computer 분야 발전

Computer는 복잡한 계산, 단순반복적 계산을 정확하고 빠르게 수행하는 기능에서 인간의 고유영역인 판단, 의사결정 영역까지 자신의 영역을 확장하고 있음

통계학



컴퓨터
공학



데이터 사이언스는 종합과학이다

데이터 사이언스는 과거, 특정 학문이라기 보다는 기존 학문을 이용하여
현대사회의 문제를 해결하는 종합과학으로 볼 수 있음

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS



DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

데이터사이언티스트 취업

- 데이터 사이언티스트 직종은 대기업, 고연봉 직종임
- 연구소, 컨설팅 회사 등에서도 선호하는 직종임



(주)카카오페이
은행/금융업 · 평점 4.6 · <https://www.kakaopay.com>

팔로우 인재풀 등록

마감 2019.02.25 09:00 ~ 2019.03.18 17:00 2019.01.25 등록

데이터 사이언티스트(Data Scientist) 영입

[SK 바이오팜] Bioinformatics 데이터 분석 경력사원 모집

[SK 바이오팜] Bioinformatics 데이터 분석 경력사원 모집	
· 공고번호	R1904D120003
· 채용인원	0명
· 고용형태	정규



[하나금융융합기술원] Data Scientist
석,박사 연구원
하나금융티아이
인천광역시 서구



**[인텔리전스랩스] 데이터분석가(게임
어뷰징 탐지)**
넥슨코리아(NEXON)
경기도 성남시

2018
SK텔레콤 DT추진단
분석/개발직군 모집

[국내 최고의 데이터 전문 조직에서
여러분들을 기다립니다]

직종별 직무 - 바이오 분야

유전정보 분석, 바이오센서, 스마트 헬스장비, 신약 개발 등 다양한 분야에서 데이터 사이언티스트를 필요로 하고 있음



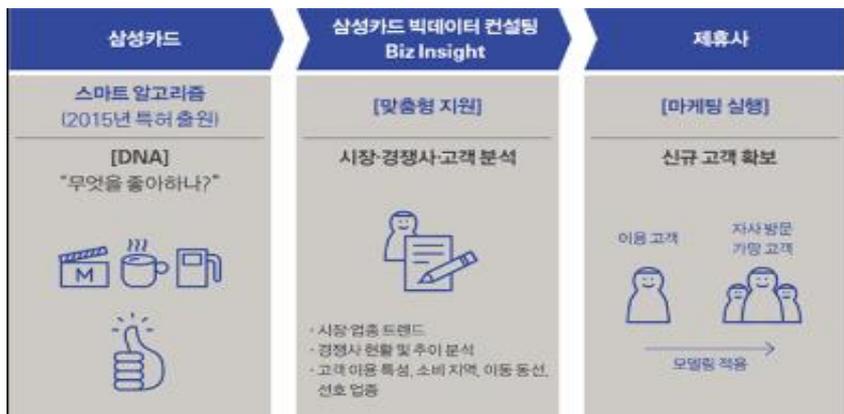
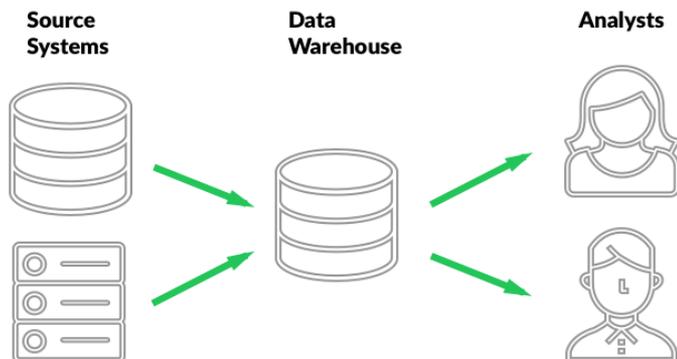
직종별 직무 - 제조 분야

생산 공정의 모든 정보가 센서에 의해 수집/관리되고 이를 통한 의사결정이 다시, 로봇에 의해 반영되는 과정에 필요한 여러 업무에 데이터 사이언티스트가 필요함



직종별 직무 - 경영 분야

- 기업이 보유한 막대한 고객 사용정보를 바탕으로 회사 경영/마케팅 활용
- 제휴사/가맹점 매출 증대에 도움을 주는 활동으로 통해 가맹점 유치



직종별 직무 - A.I 분야

- 다양한 분야에서 머신러닝을 이용한 서비스를 개발하고, 이를 수행할 인력을 필요로 하고 있음(머신러닝은 통계학이라고 볼 수 있음)

Google  Scree...120727.jpg x   

전체 **이미지** 지도 쇼핑 더보기 설정 도구

검색결과 약 1개 (2.75초)



이미지 크기:
802 × 1207

다른 크기의 해당 이미지를 찾을 수 없습니다.

관련 있을 가능성이 높은 검색어: **water bottle**

[The Best Water Bottles: Reviews by Wirecutter | A New York Times ...](#)
<https://thewirecutter.com/reviews/best-water-bottle/> ▾ 이 페이지 번역하기
 2018. 12. 4. - The Hydro Flask 21oz Standard Mouth makes the most reliable, versatile bottle for carrying water, we found. It's an insulated stainless steel ...

유사한 이미지

