

# 데이터사이언스 관점으로 이해하는 퀀트

이현열 (Henry Lee)

# Profile

- **EDUCATION**

박사과정, 한양대학교 재무금융  
석사, KAIST 금융공학  
학사, 한양대학교 경영학부

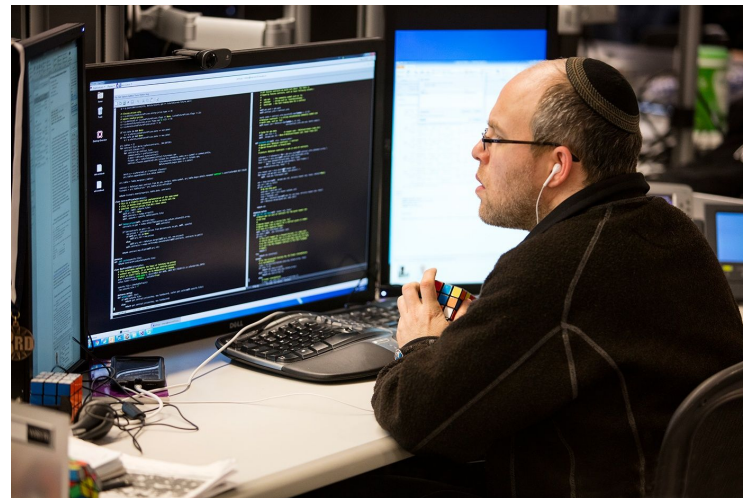
- **Job Experience**

메리츠화재, 장기전략팀, 데이터분석 (2019 ~ 현재)  
NH-Amundi 자산운용, 패시브솔루션본부, 퀀트 매니저 (2016 ~ 2019)  
한국투자증권, 고객자산운용부, 주식 매니저 (2014~2016)

- **Others**

스마트베타 (워터베어프레스, 2017)  
패스트캠퍼스, 퀀트투자 과정  
Henry's Quantopia (<http://henryquant.blogspot.kr>)  
<https://github.com/hyunyulhenry>  
[leebisu@gmail.com](mailto:leebisu@gmail.com)

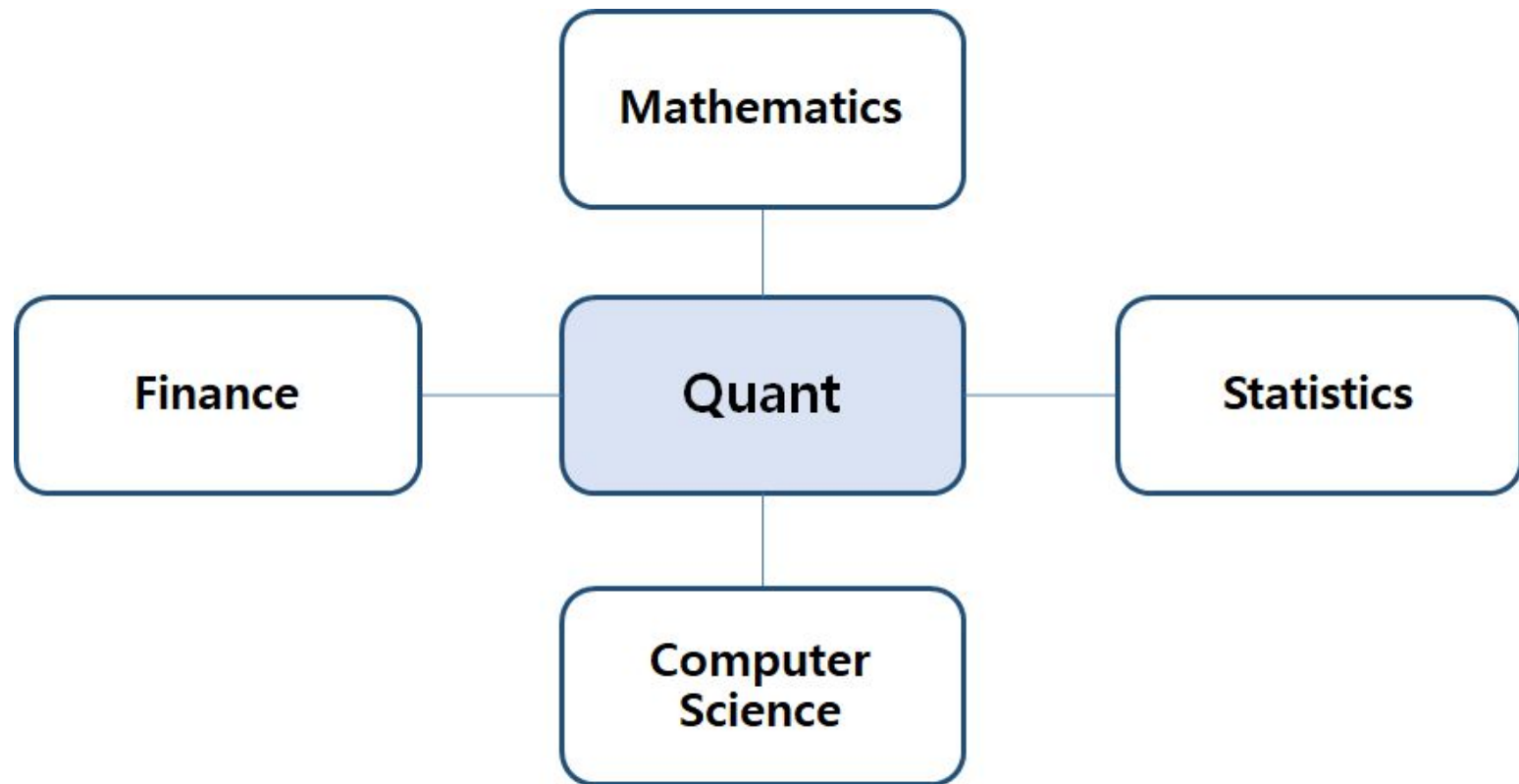
## 정성적 투자 vs 정량적 투자



# Quant

- 계량적(**Quant**itative)의 약자, 퀀트(Quant)
- 수학적 기술을 금융시장에 접목하고, 이를 이용해 투자하는 사람들
- 정성이 아닌, 시장에서 발생하는 현상에서 규칙을 찾아 투자

# Quant



# Quant

## (Fundamental) Quant Management

주식, 채권 등 투자자산을 대상으로 수익률과 연관된 유의한 설명변수를 찾고, 이를 이용하여 투자 경제학, 통계학 기반

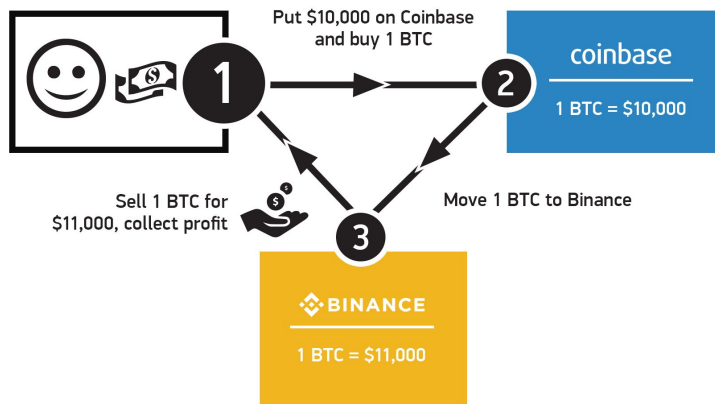
	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Portfolios Formed on Size												
Return	1.64	1.16	1.29	1.24	1.25	1.29	1.17	1.07	1.10	0.95	0.88	0.90
$\beta$	1.44	1.44	1.39	1.34	1.33	1.24	1.22	1.16	1.08	1.02	0.95	0.90
ln(ME)	1.98	3.18	3.63	4.10	4.50	4.89	5.30	5.73	6.24	6.82	7.39	8.44
ln(BE/ME)	-0.01	-0.21	-0.23	-0.26	-0.32	-0.36	-0.36	-0.44	-0.40	-0.42	-0.51	-0.65
ln(A/ME)	0.73	0.50	0.46	0.43	0.37	0.32	0.32	0.24	0.29	0.27	0.17	-0.03
ln(A/BE)	0.75	0.71	0.69	0.69	0.68	0.67	0.68	0.67	0.69	0.70	0.68	0.62
E/P dummy	0.26	0.14	0.11	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.01
E(+)/P	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.09
Firms	772	189	236	170	144	140	128	125	119	114	60	64

Portfolio	0	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Stocks Sorted on Book-to-Market Equity (BE/ME)													
Return	0.30	0.67	0.87	0.97	1.04	1.17	1.30	1.44	1.50	1.59	1.92	1.83	
$\beta$	1.36	1.34	1.32	1.30	1.28	1.27	1.27	1.27	1.27	1.29	1.33	1.35	
ln(ME)	4.53	4.67	4.69	4.56	4.47	4.38	4.23	4.06	3.85	3.51	3.06	2.65	
ln(BE/ME)	-2.22	-1.51	-1.09	-0.75	-0.51	-0.32	-0.14	0.03	0.21	0.42	0.66	1.02	
ln(A/ME)	-1.24	-0.79	-0.40	-0.05	0.20	0.40	0.56	0.71	0.91	1.12	1.35	1.75	
ln(A/BE)	0.94	0.71	0.68	0.70	0.71	0.71	0.70	0.68	0.70	0.70	0.70	0.73	
E/P dummy	0.29	0.15	0.10	0.08	0.08	0.08	0.09	0.09	0.11	0.15	0.22	0.36	
E(+)/P	0.03	0.04	0.06	0.08	0.09	0.10	0.11	0.11	0.12	0.12	0.11	0.10	
Firms	89	98	209	222	226	230	235	237	239	239	120	117	

# Quant

## Statistical Arbitrage (Pair Trading)

$(1 + 2) \neq 3$  or  $A \neq A$ , 경제적 이론 보다는 차익을 발견하여 트레이딩

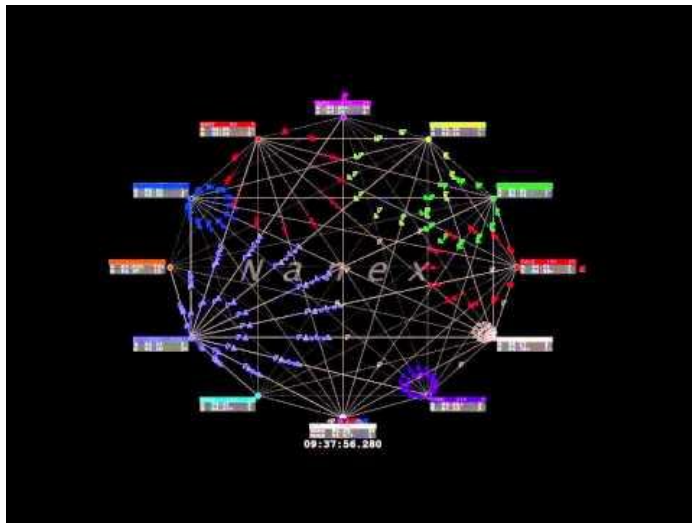


# Quant

## High Frequency Trading (HFT)

시장 조성자(Market Making), 거래소간 시간차를 이용한 선행매매 (플래시 보이스)

[https://www.youtube.com/watch?v=B\\_k\\_elbBz8c](https://www.youtube.com/watch?v=B_k_elbBz8c)





# Quant

	(Fundamental) Quant Management	Statistical Arbitrage
공통점	Rule – based Investment	
학문적 기반	경제학, 통계학	재정거래 관계, 통계학
회전율	Days to Months	Hours to Days
투자 예시	팩터 투자, 글로벌 매크로	페어 트레이딩, 차익거래 (합병, 현선물 등)
투자대상	Firm	Timing

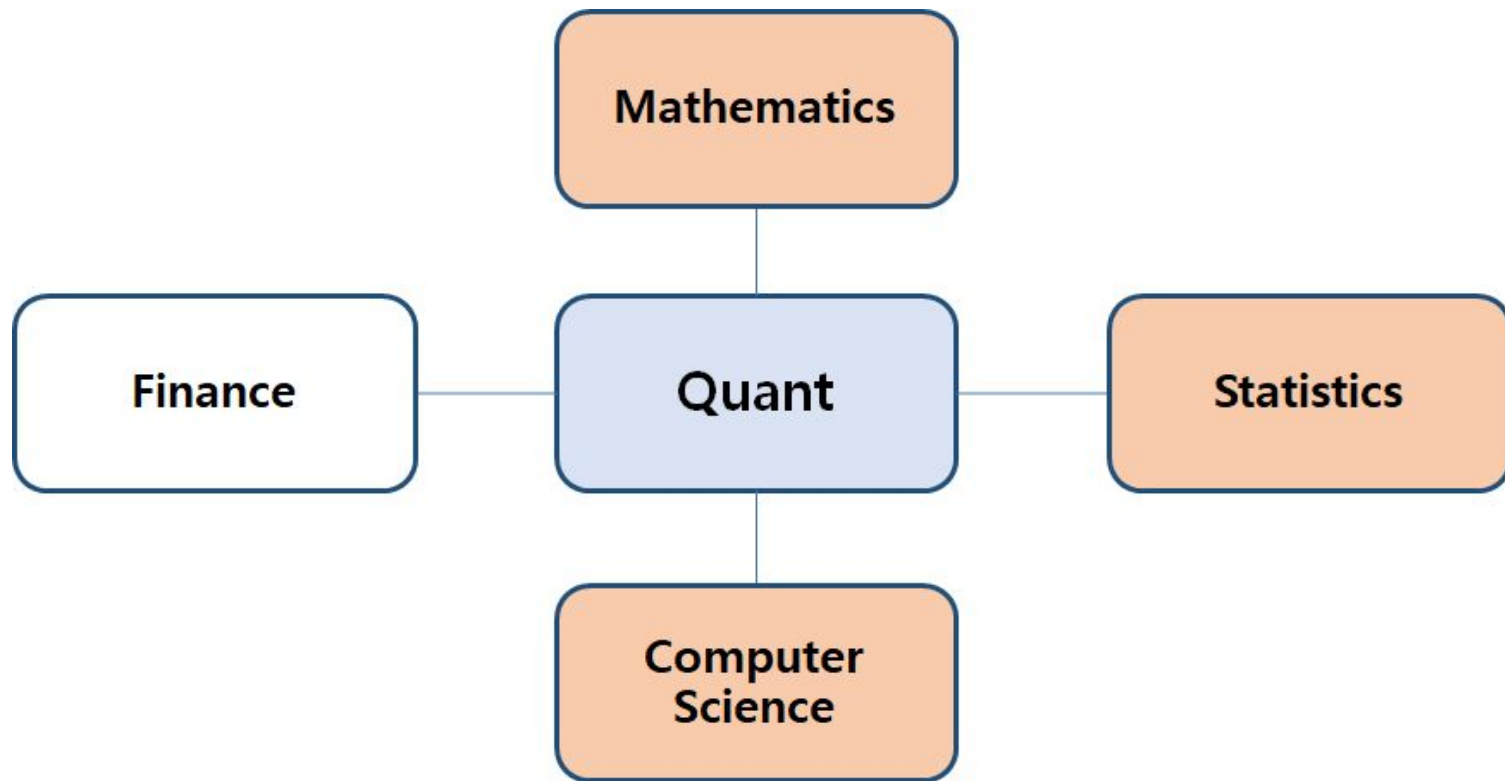
# Quant

## 전통적 퀀트의 한계점

- 글로벌로 확장된 투자 유니버스 → 데이터의 증가
- 기존 정형데이터(주가, 재무 등)가 아닌 비정형데이터를 이용한 투자 필요성
- HFT와 같은 단순 속도 전쟁의 한계

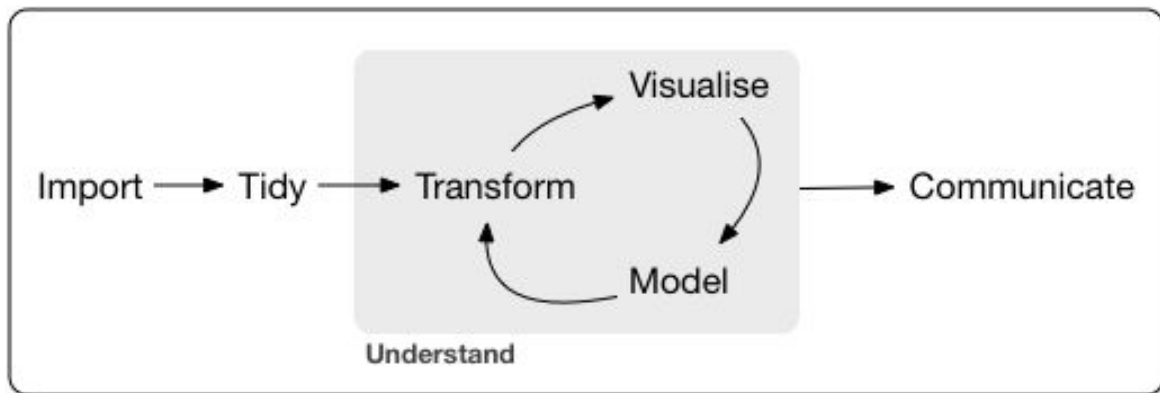
→ 빅데이터, 비정형데이터, 머신러닝(딥러닝) 활용이 가능한 데이터사이언스에 대한 수요 증가

# Data Science for Quant



# Data Science for Quant

1. Import and Tidy (데이터 구하기 및 가공)
2. Transform, Visualize, and Model (모델링)
3. Communicate (투자 집행 및 평가)

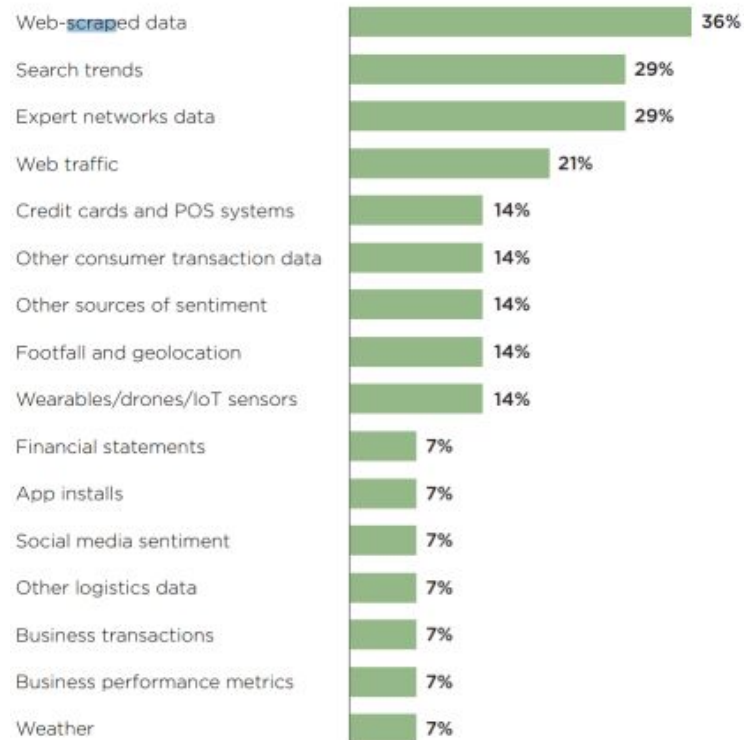


## Import and Tidy

- 크롤링, **API**를 이용하여 뉴스, 공시의 실시간 전송
- 기존 투자에 사용되던 데이터  
정형 데이터: 주가, 재무제표, 이익 예상치 등  
전처리에 큰 어려움이 없음
- 최근 투자에 사용되는 데이터  
비정형 데이터: 음성, 텍스트, 이미지 등  
기존 경제학자 및 대부분의 퀀트는 전처리 불가능

# Import and Tidy

## ALTERNATIVE DATA SETS CURRENTLY USED



Note: Based on 14 respondents.

Source: Greenwich Associates 2018 Future of Investment Research Study

# Import and Tidy



## 최근공시

## 유가증권시장

최근공시 > 유가증권시장

RSS 주소복사 도움말

유가증권시장 101건 (2019년 05월 24일)

시간	공시대상회사	보고서명	제출인	접수일자	비고
17:40	유 녹십자	[발행조건확정] 증권신고서(채무증권)	녹십자	2019.05.24	
17:39	유 LS	최대주주등소유주식변동신고서	LS	2019.05.24	
17:35	유 삼양통상	최대주주등소유주식변동신고서	삼양통상	2019.05.24	
17:33	유 SK네트웍스	최대주주등소유주식변동신고서	SK네트웍스	2019.05.24	
17:30	유 한화생명	특수관계인에대한채권매도	한화생명	2019.05.24	
17:27	유 동서	최대주주등소유주식변동신고서	동서	2019.05.24	
17:22	유 신영증권	[기재정정] 투자설명서(일괄신고)	신영증권	2019.05.24	
17:21	유 롯데손해보험	투자관단관련주요경영사항	롯데손해보험	2019.05.24	
17:21	유 한화생명	투자관단관련주요경영사항	한화생명	2019.05.24	
17:20	유 신영증권	[기재정정] 일괄신고추가서류(파생결합증권-주가연계증권)	신영증권	2019.05.24	
17:19	유 잇츠한불	유상증자결정(중속회사의주요경영사항)	잇츠한불	2019.05.24	
17:18	유 잇츠한불	타법인주식및출자증권취득결정	잇츠한불	2019.05.24	
17:11	유 미래에셋대우	투자설명서(일괄신고)	미래에셋대우	2019.05.24	
17:09	유 금호건기	최대주주등소유주식변동신고서	금호건기	2019.05.24	
17:07	유 미래에셋대우	일괄신고추가서류(파생결합증권-주가연계증권)	미래에셋대우	2019.05.24	
17:05	유 일정실업	최대주주등소유주식변동신고서	일정실업	2019.05.24	
16:58	유 서훈	[발행조건확정] 증권신고서(채무증권)	서훈	2019.05.24	
16:55	유 서훈	[기재정정] 증권신고서(채무증권)	서훈	2019.05.24	

[한두 안핵] 실적/주요공시

참가자 449명

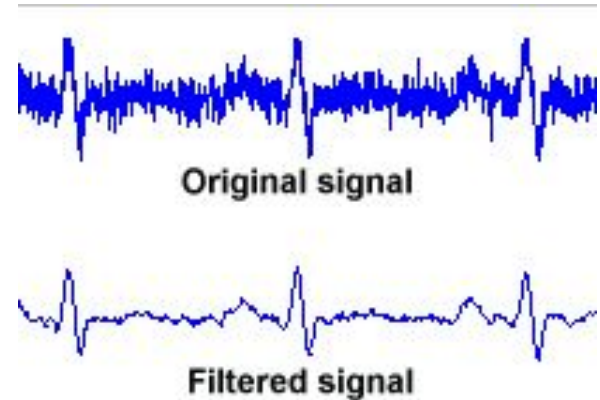
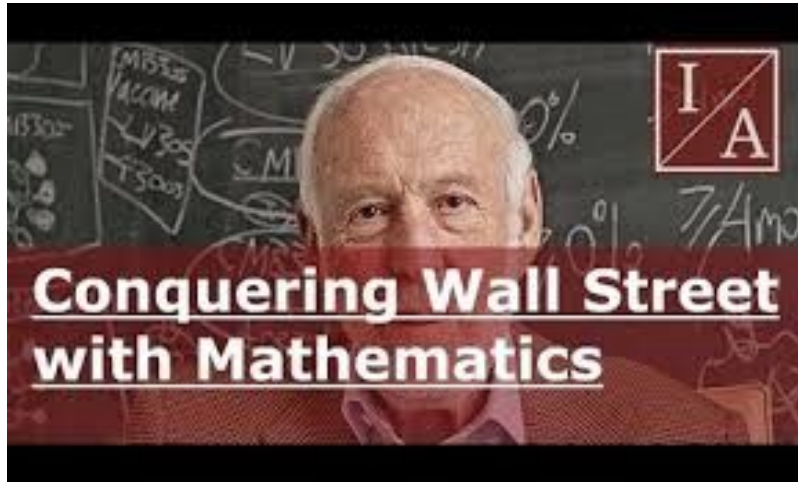


## Import and Tidy





# Import and Tidy

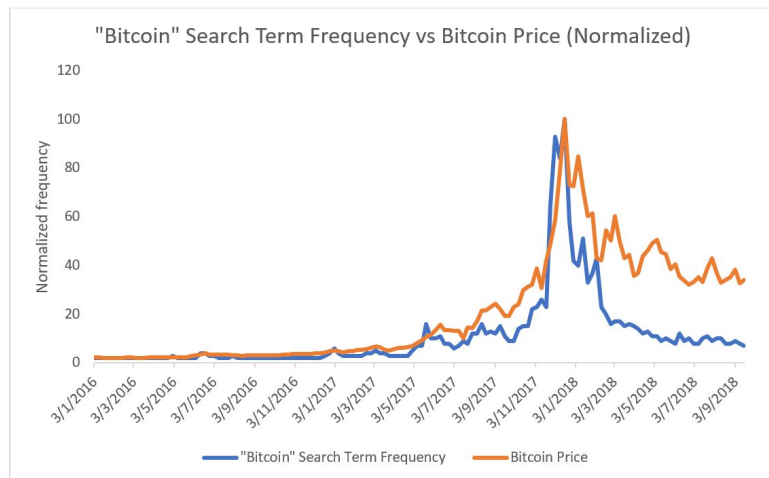


# Import and Tidy

**텍스트 데이터**는 비정형 데이터 중 투자에서 가장 활발하게 연구되는 분야

- 상대적으로 처리가 간단함
- 특정 종목 관련 검색량, 게시글 숫자 등 정량적 내용 뿐만 아니라 어조, 문구 등 정성적 내용도 주식 수익률과 연관됨
- 유명인의 **SNS** 활동이 주식 수익에 많은 영향을 미침

# Import and Tidy



## Simplified Income Statement for Company ABC

	2007	2006	Percentage change
Sales (\$)	95	100	-5.0
Expenses (\$)	74	80	-7.5
Net income (\$)	21	20	5.0
Return on sales (%)	22.1	20.0	

### Alternative textual descriptions, each of which is factually correct:

- a. In 2007, sales totaled \$95, and net income was \$21.
- b. In 2007, sales and net income were \$95 and \$21, respectively, compared to \$100 and \$20 for 2006.
- c. In 2007, sales decreased by 5%.
- d. In 2007, net income increased by 5%. Profitability improved, with return on sales increasing from 20% to over 22%.

# Import and Tidy



**Elon Musk**   
@elonmusk 팔로우 

Am considering taking Tesla private at \$420. Funding secured.

 트윗 번역하기

오전 9:48 · 2018년 8월 7일

15,999 리트윗 87,843 마음에 들어요



 6,322  15,999  87,843 

 내 답글을 트윗합니다.

**Elon Musk**  @elonmusk · 8월 7일 

Shareholders could either to sell at 420 or hold shares & go private

 트윗 번역하기

 1,387  2,201  19,411 

# Import and Tidy

이미지 데이터(인공위성 사진 등)를 이용한 투자 역시 이미 보편적 추세



# Transform, Visualize, and Model

## 전통적 주식 브로커 & 트레이더

- 브로커가 고객(운용사 등)으로부터 주문을 받음
- 해당 주문을 트레이더가 수행
- 감에 의존한 매매, 잦은 실수
- Best Execution에 대한 의문

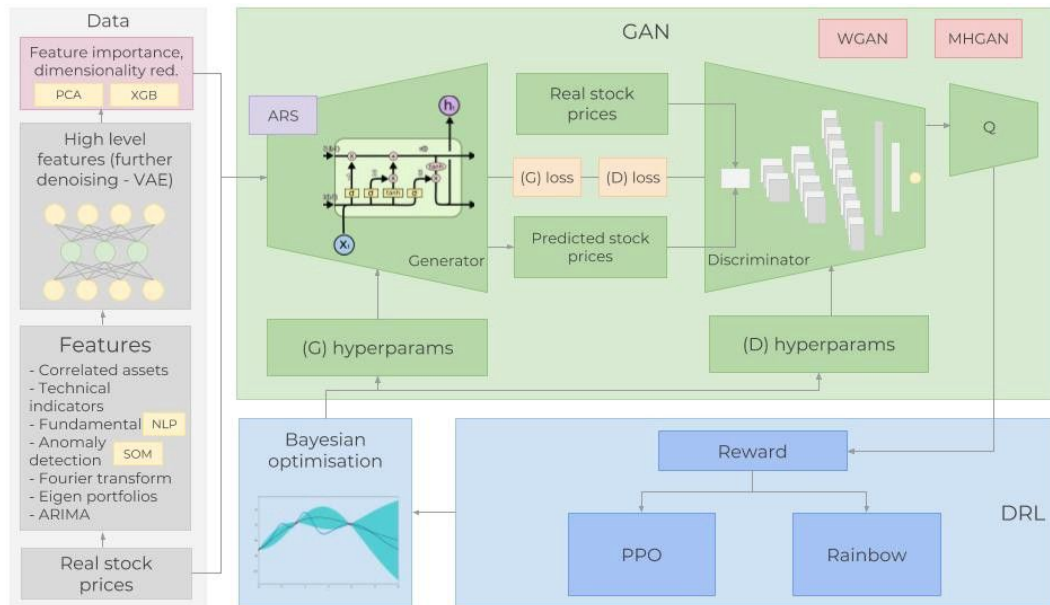
# Transform, Visualize, and Model

## 골드만삭스의 교훈

- 2000년 600명의 트레이더 → 현재 2명
- 200명의 개발자가 트레이딩 시스템 개발
- 회사의 전체 인원 중 1/3 (9000명)이 개발자
- 데빈 웨니그 이베이 CEO: *“AI를 받아들이지 않으면 어떤 산업에 있든 당신의 기업은 어려워질 것”*

# Transform, Visualize, and Model

딥러닝을 이용하여 마켓 임팩트를 최소화시키는 알고리즘이 적극적으로 연구 중





## Communicate

- 클라우드, 가상 서버를 통해 공간의 제약없이 여러명이 작업 가능
- markdown + git을 통해 **재생가능한 투자 의사결정** 및 공유
- shiny, 대시보드를 이용하여 투자상황에 대한 정보를 쉽게 확인 가능

# Job Qualification

<https://www.efinancialcareers.com>

Quant 모집 공고의 Requirements를 통해 요구능력 파악

- PhD in quantitative fields such as mathematics, statistics or equivalent
- In depth understanding of financial mathematics including stochastic differential equations, probability theory, statistical analysis, interest rates and credit risk modeling.
- Well organized, detail-oriented with good communication skills (both written and verbal)
- Strong coding ability in R, Python and C++ is a plus
- 5y work experience is required in quantitative modeling and/or validation in one of the following areas XVA/IM/CCR/IMM or Value at Risk models

## Requirements:

- A MS or PhD in finance, statistics, math, physics or computer science is strongly preferred along with 5+ years of work experience.
- Must have direct experience in TCA
- Current development experience with Python (Numpy, Scipy, Pandas, SKLearn)
- Strong Statistical and data management skills
- Experience modeling large unstructured datasets
- Deep understanding of US equity market microstructure
- Experience with market data: US real time and historical.

## Qualifications

- PhD or Masters in statistics, computer science, physics, finance or a related field.
- Strong data analysis skills analytical and problem solving skills are a prerequisite.
- 5+ years working in a systematic macro or global equities investment management firm.
- Strong familiarity with global macro products (equity, fixed income, futures, commodities, fx)
- Strong coding experience in Python, R, or Matlab.
- Entrepreneurial mindset needed.

## Requirements:

Technical skills: Java or C++, Perl/Python, Unix, SQL

Quantitative skills: Statistics, Experience with data analysis, Time Series, R, Kdb/Q

Soft skills: Strong communication, motivated, team player

Related Experience: 2-4 years