Are you data science ready?

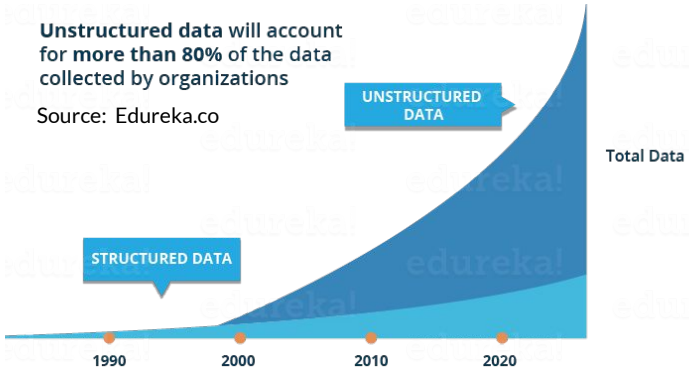
Skills for investment professionals

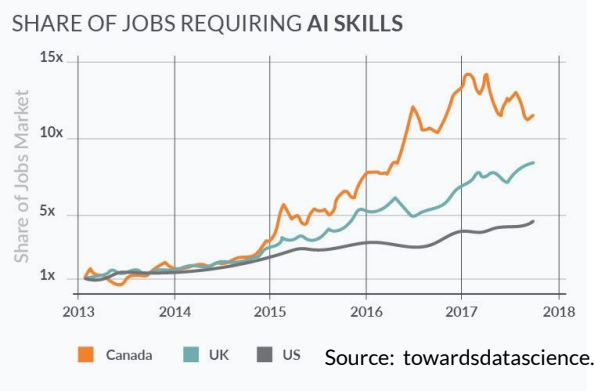
Dr. Jatin Thukral, CFA

Note: The views expressed are personal and not of speaker's current or past employers. All images are meant for exposition of ideas only and viewers are cautioned to conduct their own due diligence before arriving at any conclusions.

What is Data Science?

- **Data Science** is the ***science of extracting actionable insights from both structured and unstructured data sets*** using techniques from **Statistics, Mathematics** and **Computer Programming**
- Over the last 20 years, **concurrent advances in technology, telecommunications, social media and ML research** have lead to explosive growth in data creation and consumption by businesses, who in turn **need data scientists to make sense of this data**





- Given this unending demand for employees with data science skills, it is not surprising that the Harvard Business Review, in its October 2012 issue, called "***Data Scientist as the sexiest job of the 21st Century***"

How is Data Science used in Finance?

Customer Experience Enhancement

- Personalized Customer Service
- Financial Product Recommendations
- Chatbots and Voice Recognitions
- Real-Time Inquiry Response Modeling
- Churn Prediction

Risk Management

- Fraud Detection
- Money Laundering Detection
- Insurance and Claims Automation
- Loan Delinquency Prediction
- Risk Modeling (Market, Political etc.)

Asset Management

- Quant Investment Funds
- Algorithmic Trading
- High Frequency Trading
- Socially Responsible Investing
- Portfolio Optimization

FinTech and Data Monetization

- Credit Underwriting
- Payment Processing
- Cyber Security
- Advertising
- Natural Language Report Generation

How is Data Science changing Finance?

- **Data Science** is disrupting the financial services sector ***by increasing productivity*** and ***by giving birth to new business models***
- Most notable of these disruptions are getting manifested through one of the **following techniques**:

Automation	*E.g*, options market making requiring human judgement is now feasible in microseconds
Scalability of Insights	*E.g*, investors can '*machine-read*' 10,000s of conf. calls, earnings and broker reports daily
Big Data Engineering	*E.g.* entire micro-structure data can be analyzed to identify tell-tale signs of momentum
Use of Alternate Data	*E.g*, insights can be extracted from texts, records and images sold by 3rd party vendors
Unstructured Data	*E.g. using satellite images of soyabean plantations to predict soyabean futures price changes*
Consolidated Insights	*E.g.* weak signals - of little use on their own - can be combined into actionable insight

- To understand these techniques, let us look at an industry **(Quant Funds)** that **has changed with the advent of data science**

Case Study How Data Science changed Quant Funds?

- **Quant Funds:** A quant fund is an investment fund that selects securities by utilizing the capabilities of advanced quantitative analysis. (Investopedia)
- Till about early 2000s, quant fund strategies were primarily reliant of signals based on simple data sets such as fundamental ratios (E/Ps, P/S, changes in dividend yields etc.) and economic measures
- **Starting mid 2000s, many quant funds started incorporating data science based signals**
- Post 2008, quant hedge funds increased their outperformance over fundamental hedge funds
- The **most successful quant hedge funds** (Two Sigma, Renaissance Technologies, Citadel etc.) **were the most aggressive users of data science**

Case Study How Data Science changed Quant Funds?

- Data science **changed** quant funds for good through the aforesaid techniques: **Scalability of Insights**, **Big Data Engineering**, **Alternate Data Sets** and **Unstructured Data**

DS Technique 1: Scalability of Insights

Example A: Use ***machines to 'read'*** millions of sell side research reports

Goldman Stanley

18 Sep 2015
Europe/United Kingdom
Equity Research
Commercial Banking

Shawbrook Group PLC (LSE: SHAW)

NEUTRAL RECOMMENDATION / COMPANY UPDATE

Modest Upside for Next 12 Months; Possible Economic Headwinds Ahead

Rating:	NEUTRAL*
Price (18 Sep 15, GBp):	327p
Target Price (GBp):	358p
52-Week Price Range:	301p – 381p
Market Cap. (GBP £ M):	818.5

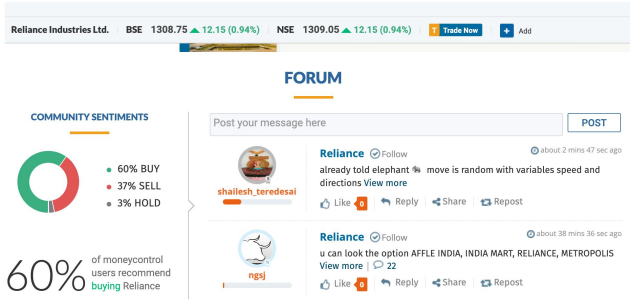
*Target Price is for 12 months.

Research Analysts:
Donnie Azoff, CFA
Mark Hanna, CFA
Billy Valentine

- **Mixed 1H15:** While Shawbrook reported lower-than-expected mortgage growth for the first half of the year, industry-wide growth rates were also down. However, the company's NIM remained stable, its Cost-to-Income Ratio fell to 48.9%, and its CET 1 Ratio rose to 15.2% as a result of the successful IPO in April.
- **Financial Guidance:** The Company recently reaffirmed its long-term and short-term financial guidance, including £2 billion in gross originations by FY 17, NIM remaining stable at current levels, a Cost-to-Income Ratio of 40% by FY 17, a pre-tax ROA of 3.5% by FY 17, annualised cost of risk of 0.5-0.6%, and a CET 1 Ratio of at least 13%. We believe the Gross Origination and NIM targets may be on the high side, given market headwinds and increased competition. However, Shawbrook's relatively diversified loan book (only ~22% BTL mortgages vs. 41% for ALD and 46% for OSB) may help it weather the storm of regulatory changes.

Price Performance Since IPO

Example B: 'Understand' millions of emails and messages of investors on forums/lists



Reliance Industries Ltd. BSE 1308.75 ▲ 12.15 (0.94%) | NSE 1309.05 ▲ 12.15 (0.94%) Trade Now Add

FORUM

COMMUNITY SENTIMENTS

- 60% BUY
- 37% SELL
- 3% HOLD

60% of moneycontrol users recommend buying Reliance

Post your message here POST

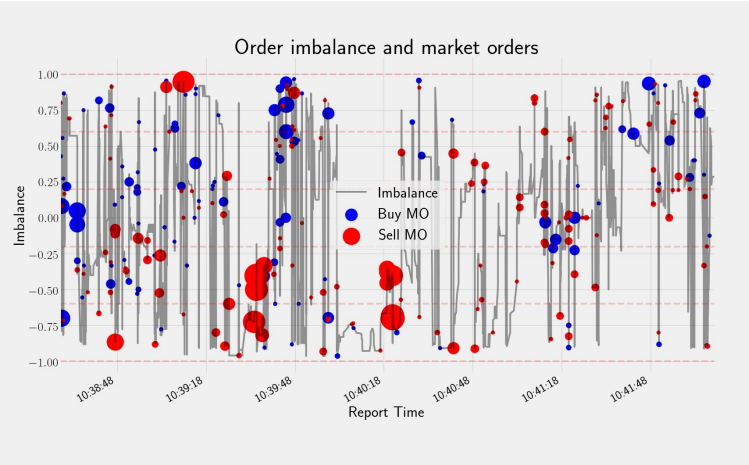
shailesh_teredesai — Reliance Follow — about 2 mins 47 sec ago
already told elephant move is random with variables speed and directions View more
Like 0 Reply Share Repost

ngsj — Reliance Follow — about 38 mins 36 sec ago
u can look the option AFFLE INDIA, INDIA MART, RELIANCE, METROPOLIS View more | 22
Like 0 Reply Share Repost

Case Study How Data Science changed Quant Funds?

DS Technique 2: Big Data Engineering

Example A: Find patterns (e.g. order imbalances) across billions of orders in order books



Source https://davidsevangelista.github.io/

Example B: Map security-level flow information to their sources and identify biases (herding etc)

	Date	Typical Price	Up or Down	Volume*	Raw Money Flow	1-period Positive Money Flow	1-period Negative Money Flow	14-period Positive Money Flow	14-period Negative Money Flow	14-period Money Flow Ratio	14-period Money Flow Index
1	3-Dec-10	24.63		18,730							
2	6-Dec-10	24.69	1	12,272	302,976	302,976	-				
3	7-Dec-10	24.99	1	24,691	617,060	617,060	-				
4	8-Dec-10	25.36	1	18,358	465,534	465,534	-				
5	9-Dec-10	25.19	-1	22,954	578,388	-	578,388				
6	10-Dec-10	25.17	-1	15,919	400,628	-	400,628				
7	13-Dec-10	25.01	-1	16,067	401,796	-	401,796				
8	14-Dec-10	24.96	-1	16,568	413,516	-	413,516				
9	15-Dec-10	25.08	1	16,019	401,810	401,810	-				
10	16-Dec-10	25.25	1	9,774	246,780	246,780	-				
11	17-Dec-10	25.21	-1	22,573	569,130	-	569,130				
12	20-Dec-10	25.37	1	12,987	329,483	329,483	-				
13	21-Dec-10	25.61	1	10,907	279,371	279,371	-				
14	22-Dec-10	25.58	-1	5,799	148,335	-	148,335				
15	23-Dec-10	25.46	-1	7,395	188,250	-	188,250	2,643,012	2,700,043	0.98	49.47
16	27-Dec-10	25.33	-1	5,818	147,351	-	147,351	2,340,037	2,847,394	0.82	45.11
17	28-Dec-10	25.09	-1	7,165	179,766	-	179,766	1,722,977	3,027,160	0.57	36.27
18	29-Dec-10	25.03	-1	5,673	141,978	-	141,978	1,257,443	3,169,137	0.40	28.41
19	30-Dec-10	24.91	-1	5,625	140,138	-	140,138	1,257,443	2,730,887	0.46	31.53
20	31-Dec-10	24.89	-1	5,023	125,057	-	125,057	1,257,443	2,455,317	0.51	33.87
21	3-Jan-11	25.13	1	7,457	187,374	187,374	-	1,444,817	2,053,521	0.70	41.30
22	4-Jan-11	24.64	-1	11,798	290,653	-	290,653	1,444,817	1,930,658	0.75	42.80
23	5-Jan-11	24.51	-1	12,366	303,090	-	303,090	1,043,007	2,233,748	0.47	31.83
24	6-Jan-11	24.15	-1	13,295	321,133	-	321,133	796,227	2,554,881	0.31	23.76
25	7-Jan-11	23.98	-1	9,257	221,953	-	221,953	796,227	2,207,703	0.36	26.51
26	10-Jan-11	24.07	1	9,691	233,205	233,205	-	699,949	2,207,703	0.32	24.07
27	11-Jan-11	24.36	1	8,870	216,084	216,084	-	636,663	2,207,703	0.29	22.38
28	12-Jan-11	24.35	-1	7,169	174,567	-	174,567	636,663	2,233,936	0.28	22.18
29	13-Jan-11	24.14	-1	11,356	274,191	-	274,191	636,663	2,319,877	0.27	21.53
30	14-Jan-11	24.81	1	13,379	331,942	331,942	-	968,605	2,172,526	0.45	30.84

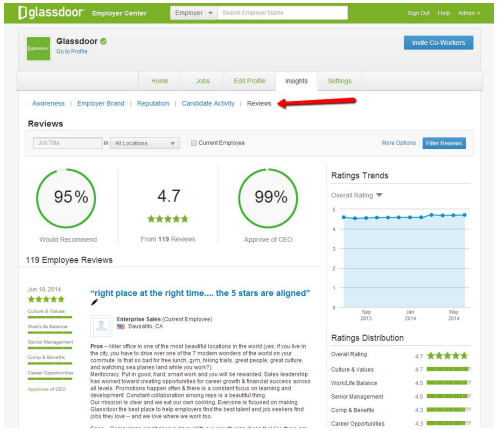
* Volume shown in 1000s

Source: Money Flow Index - StockCharts

Case Study How Data Science changed Quant Funds?

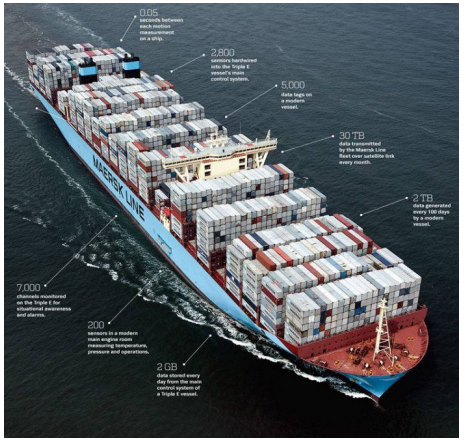
DS Technique 3: Leverage Alternate Data Sets

Example A: Scrape online employees' reviews to assess the sentiments of its staff



Source www.glassdoor.com

Example B: Use ship bookings to assess mgmt's confidence in its expected future sales

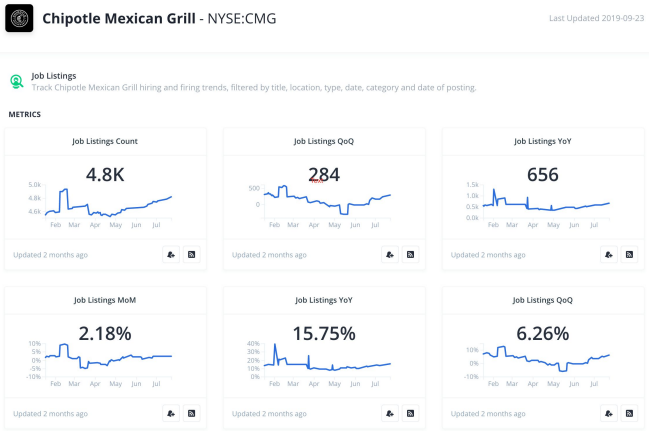


Source digital.hbs.edu

Case Study How Data Science changed Quant Funds?

DS Technique 3: Leverage Alternate Data Sets (cont'd)

Example C: Use data on job openings to assess mgmt's confidence in its future growth



Source www.thinknum.com

Example D: Use retail footfall data to find trends in cx visits to different brand stores

Source www.retailsensing.com

Case Study How Data Science changed Quant Funds?

DS Technique 4: Process Unstructured Datasets (Text and Voice)

Example A: Decode and quantify unconscious biases hidden in the texts of conference calls

Source www.basf.com

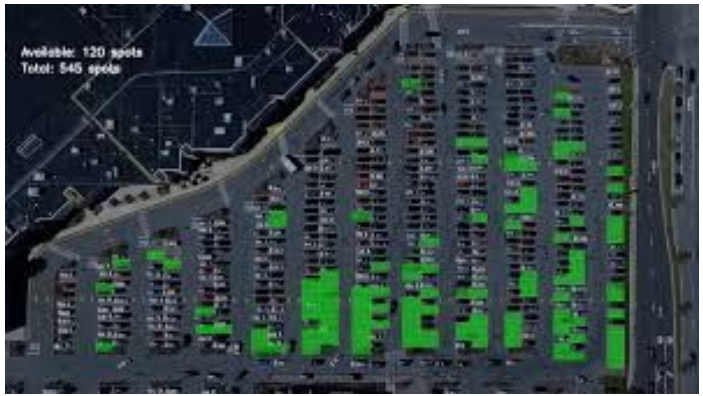
Example B: Detect the signals for future actions in central bank statements

Source www.livemint.com

Case Study: How Data Science changed Quant Funds?

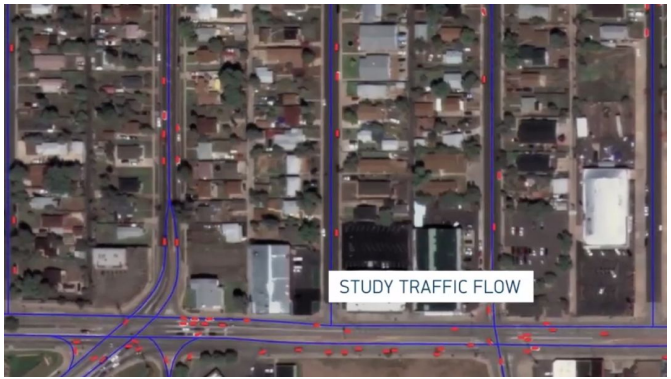
DS Technique 4: Process Unstructured Datasets (Images and Videos)

Example A: Changes in number of cars parked in Walmart stores may predict changes in sales



Source www.towardsdatascience.com

Example B: Analyze traffic patterns to assess real estate / infrastructure demand

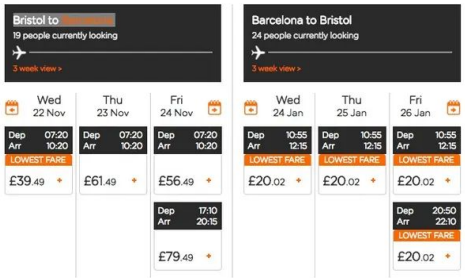


Source www.digitalglobe.com

Disruption in Quant Funds was Just a Precursor!

- In the early years of data science adoption in Quant funds, it was tempting for finance professionals from other sub-industries to underestimate the future impact of data science to their respective sub-industries
- For instance, a few **fundamental investors** were quick to dismiss data science as *just another technology fad*, until of course a group of their own peers proved them wrong by **leveraging data science in novel ways**:

Example A: Use of airline pricing data to assess respective ***pricing powers*** of different airlines



Source www.theguardian.com

Example B: Analysis of social media to assess demand for new products (Iphones, Juul etc.)

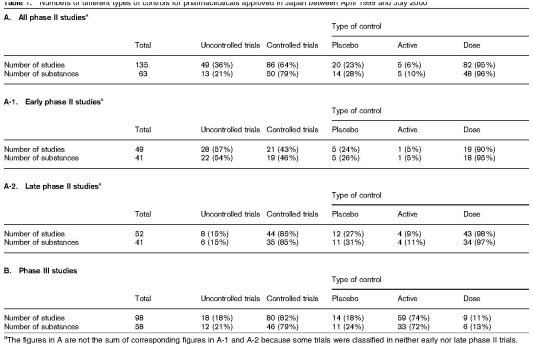
Source www.digitalglobe.com

Disruption in Quant Funds was Just a Precursor!

- The enterprising value investors went one step further to do some advanced predictive data science!

Example C: Use of clinical trial records to predict success in pharma drug research

Example D: Analysis of hotel booking records to segregated differential threats of Airbnb



A. All phase II studies[a]

				Type of control		
	Total	Uncontrolled trials	Controlled trials	Placebo	Active	Dose
Number of studies	135	49 (36%)	86 (64%)	20 (23%)	5 (6%)	82 (95%)
Number of substances	63	13 (21%)	50 (79%)	14 (28%)	5 (10%)	48 (96%)

A-1. Early phase II studies[a]

				Type of control		
	Total	Uncontrolled trials	Controlled trials	Placebo	Active	Dose
Number of studies	49	28 (57%)	21 (43%)	5 (24%)	1 (5%)	19 (90%)
Number of substances	41	22 (54%)	19 (46%)	5 (26%)	1 (5%)	18 (95%)

A-2. Late phase II studies[a]

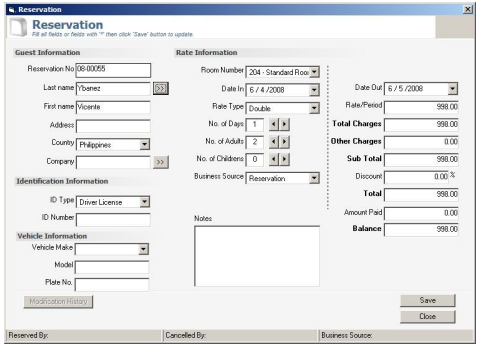
				Type of control		
	Total	Uncontrolled trials	Controlled trials	Placebo	Active	Dose
Number of studies	52	8 (15%)	44 (85%)	12 (27%)	4 (9%)	43 (98%)
Number of substances	41	6 (15%)	35 (85%)	11 (31%)	4 (11%)	34 (97%)

B. Phase III studies

				Type of control		
	Total	Uncontrolled trials	Controlled trials	Placebo	Active	Dose
Number of studies	98	18 (18%)	80 (82%)	14 (18%)	59 (74%)	9 (11%)
Number of substances	58	12 (21%)	46 (79%)	11 (24%)	33 (72%)	6 (13%)

[a]The figures in A are not the sum of corresponding figures in A-1 and A-2 because some trials were classified in neither early nor late phase II trials.

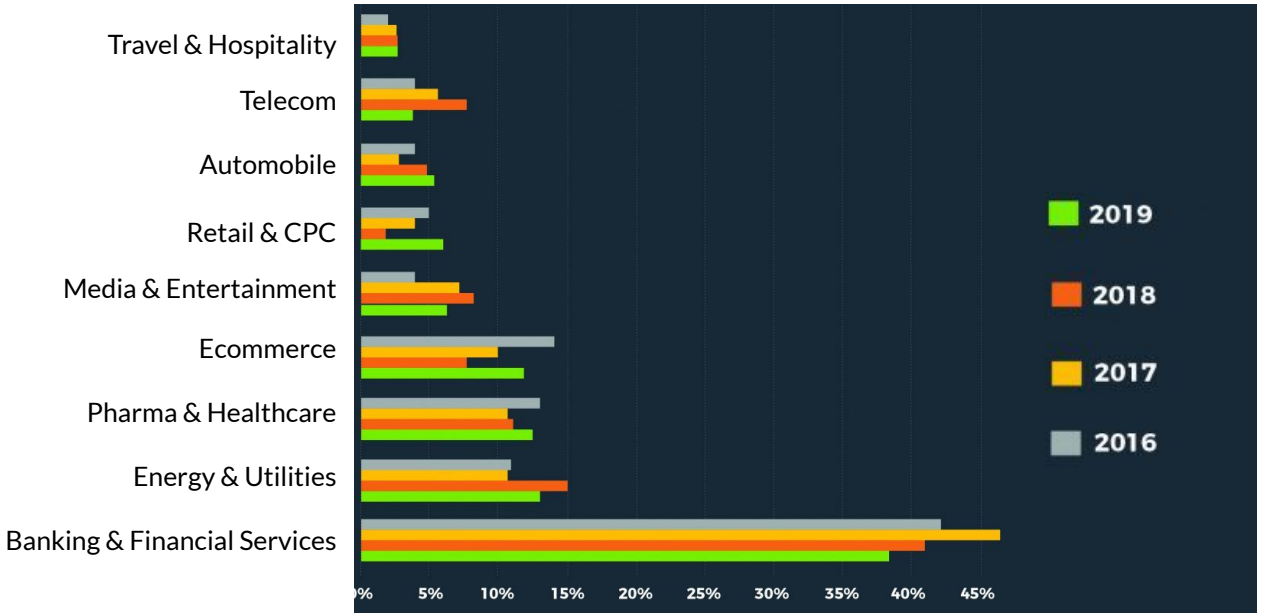
Source semanticscholar.org



Source www.sourcecodester.com

- The learning from the experience of value investors was clear for financial professionals from all sub-industries: ***Better to leverage data science for your career than to limit yourself to good old wisdom!***

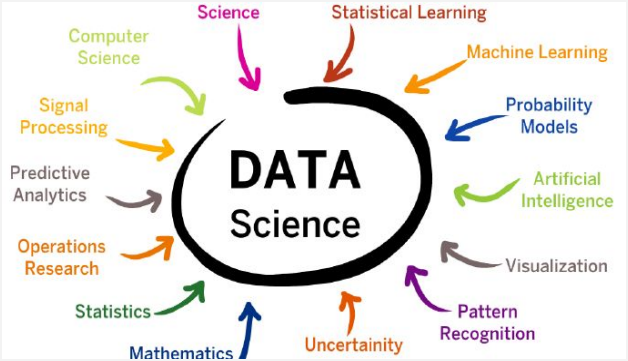
Is it then a surprise that FIs now lead in hiring data science talent?



Source: www.analyticsindiamag.com

But, what are the skills needed for data science jobs in Finance?

Data science is a vast field comprising of an exceptionally **large number of skill sets**!



Source www.digitalglobe.com

To learn this broad ranging skill set **can be daunting** for a non data scientist

So, where to start?

Two groups of skills:

Computer Skills

- Python / R + SQL
- HDFS + NoSQL DBs
- Spark
- Kafka

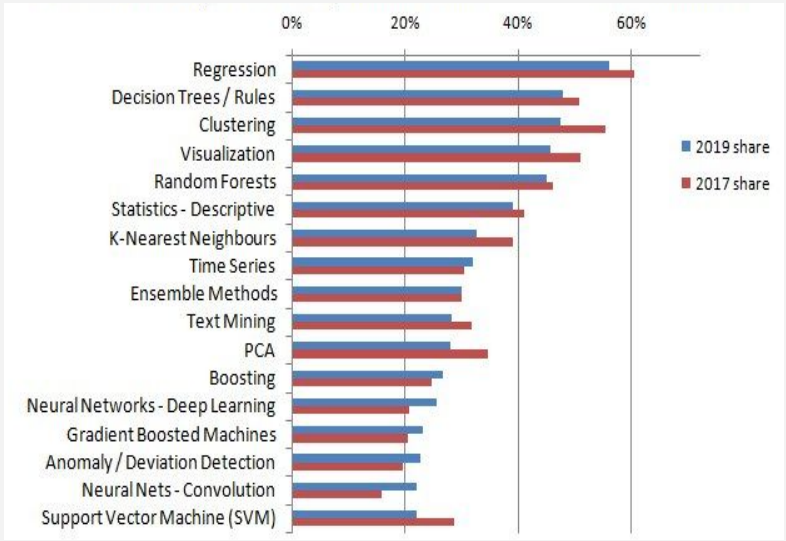
Mathematics

- Statistics
- Machine Learning
- Deep Learning
- Artificial Intelligence

Within the two groups, focus on the most popular skills!

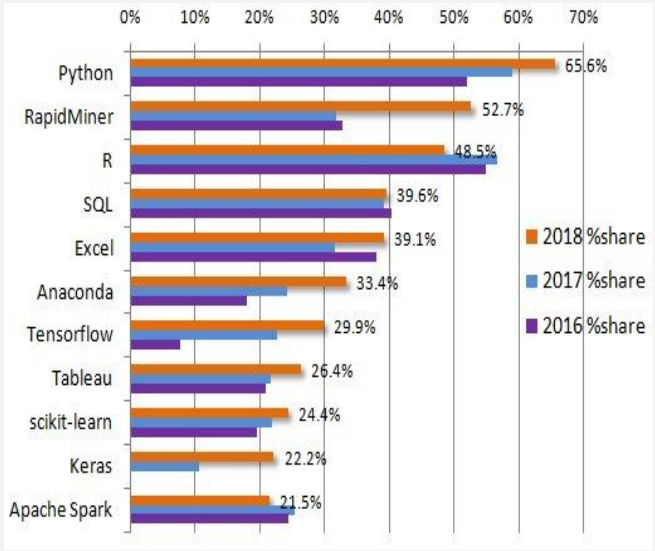
- The world of open source technology is democratic => **Skills are often popular because of a valid reason!**

Popular Algorithms



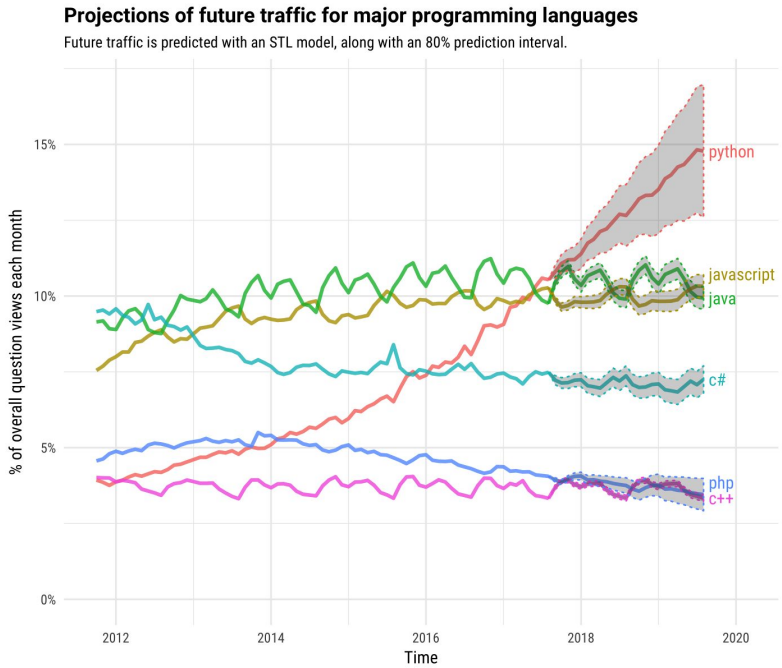
Source www.kdnuggets.com Poll Sep 2019

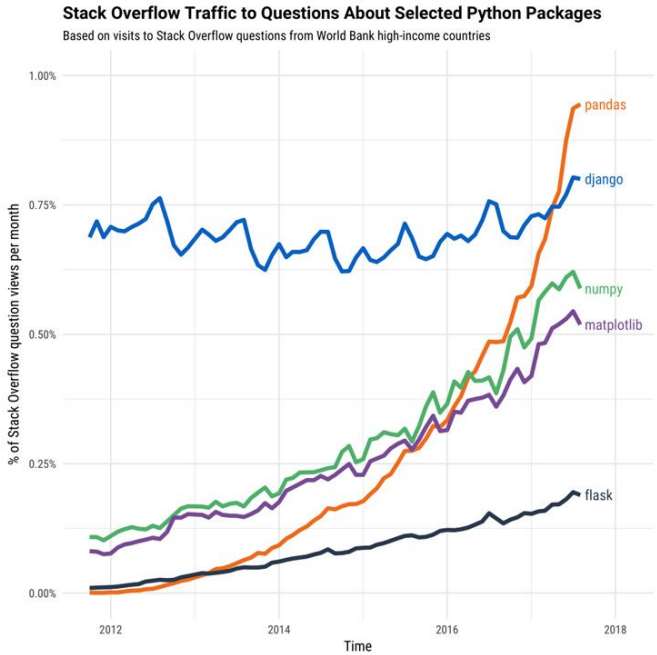
Popular Programming Languages



Source www.kdnuggets.com Poll 2016-18

One last tip, Python is the language of data science!





Thank You