

#### **Stats Central**

**Mark Wainwright Analytical Centre** 

## What can Data Science do for you?

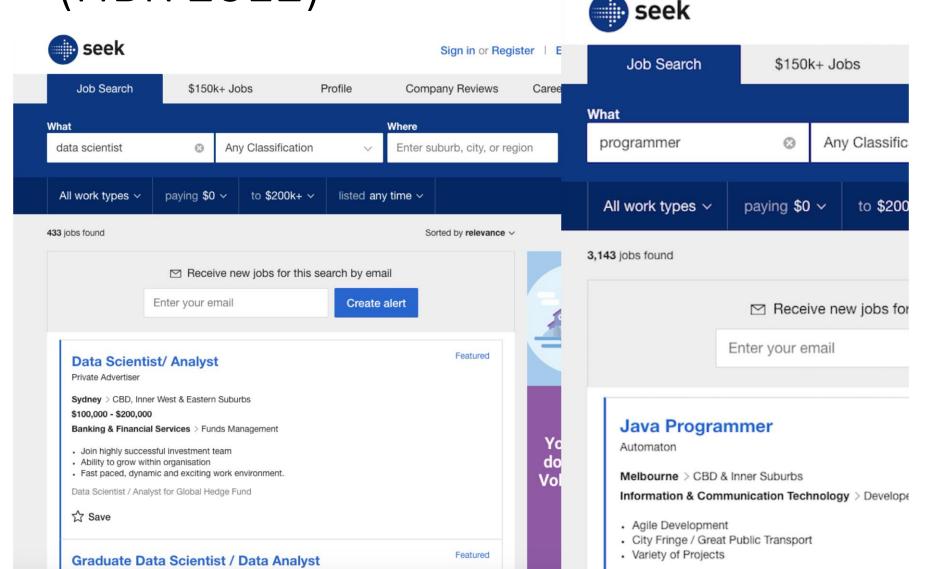


Raymond Wong
Stats Central and CSE UNSW

March 12, 2019

Why do you care?

# The sexiest job of the 21<sup>st</sup> century (HBR 2012)



What is Data Science?

Data science is a field of Big Data geared toward providing meaningful information based on large amounts of complex data. Data science, or data-driven science, combines different fields of work in statistics and computation in order to interpret data for the purpose of decision making.

Source: Investopedia 2018/3/2

## What can data science do?

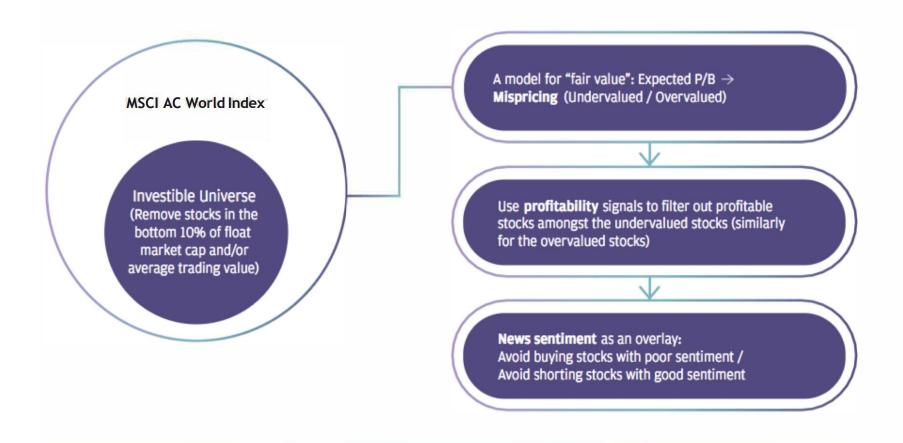
• To uncover findings from data (e.g., hidden behaviors / trends)



To produce "data product"



#### AN OVERVIEW OF OUR STOCK SELECTION STRATEGY

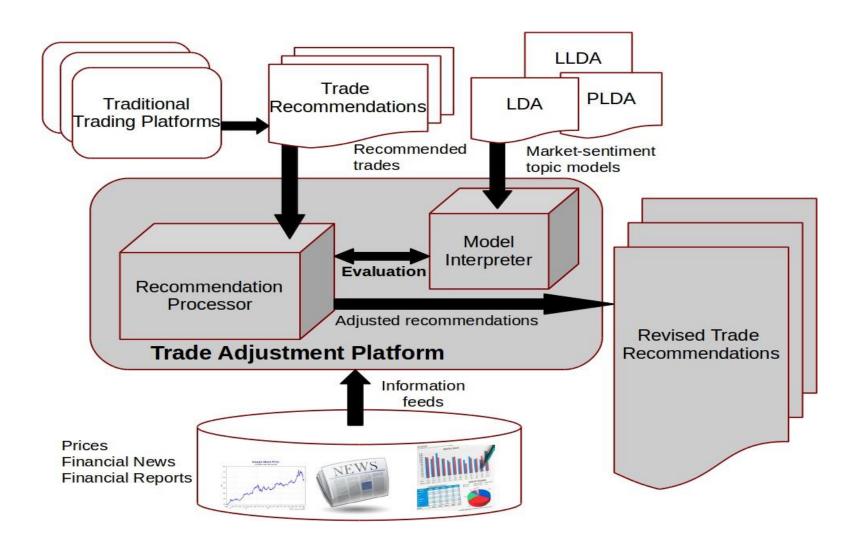


#### **FINAL STRATEGY**

Short: Overval ad stocks with poor pro italbility, removing stocks with good se timent

Long: Underval ed stocks with high profitability, removing stocks with poor sentiment

# Market sentiment analysis



## Portfolio revision

Revised down (sell less) due to +vs sentiment

#### **Revised Trades**

	Original track	Sentiment rank actual/estimated	Revisea des
Open	F D		00
	\$17		\$506, 64
Jul 12	-6,000	( +6/+4.01 )	-3,597
	-\$96,360		-\$57,720
Aug 12	8,000	0/0	8,000
	\$152,000		\$152,000
Close	8,000		12,199
	\$190,640		\$408,307



Submit...

# Main Steps in Data Science

Data Collection

Data Modeling and Analysis

Problem Solving and/or Decision Support

## The Process

- 1. Define an objective
- 2. Determine Data Requirements and Collection Plan
- 3. Collect and Review the Data
- 4. (Pre)Process the Data ("Munging")
- 5. Model and Analyze
- 6. Visualize and Communicate the Findings

# Big Data Challenges – The 4 Vs

- Volume: huge data size
- Variety: structured vs unstructured. Numeric, categorical, spatiotemporal. textual/tweets, videos/images...
- Veracity: data quality, source trustworthiness, volatility
- Velocity: data speed

### Case 1

# Legal AI tech



Users are guided to provide the scenarios and circumstances for the system to retrieve



#### Intelligence comes from:

- · Natural Language Processing (NLP)
- Topic Modelling
- · Intelligent Search | Retrieval
- Automated Summarisation



Wealth of knowledge created from knowledge bases, big data collections and fronted by an AI query and response engine



Knowledge

Base, Past

Precedence, etc

Trends,

# Legal AI tech



Matrimonial a w



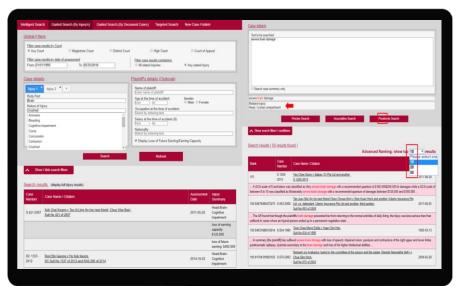
Tax aw

Maritime aw

**Property** aw

STATE COURTS

Corporate a w



Criminal aw





### Case 2



# Big Data + Good Models/Algorithms != Success

# Why Amazon & Netflix work well while others don't

- Have a purpose (a tangible objective)!
  - Not just a technical exercise
- Link "insight" to action (quickly, seamlessly & automatically)
- Push to business end-points (though infrastructure (e.g., Hadoop) is highly technical)
- Create feedback loops

### Case 3

```
· · · · 20865510 2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海游保税港区7栋746
                                                                                 *** 广东省佛山市三水区乐平镇范湖经济开发区三花公路23号广州工商学院
····•50877236 2017-12-31 23:55:1( 杨秋起
                                     胜浦镇
                                                                        ***12762734 *** 河南省鹤壁市淇县淇县红旗路与东环路交叉口向北20米路西老夫子(***12762734 ***
……(11332438 2017-12-31 23:55:1(黄珊珊
                                     浦江国际
                                                                        ***91436676 *** 内蒙古自治区呼伦贝尔市海拉尔区车站街1号金海商务宾馆财务部 ***91436676 *
••••6tmt.236
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                 *** 山东省青岛市富春江路1443号华夏银行
.....20865511
          2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                 *** 西藏自治区拉萨市坡关区粤玛贡泰办事外
- .. 20865512 2017-12-31 23:55:1(小紅书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                     上海上海市宝山区长江西路***0号泗塘3村82弄6号502室
----20865513
          2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                    浙江省丽水市缙云县新建镇葛湖小学
•..-.•20865514 2017-12-31 23:55:1(小紅书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                    广东省深圳市福田区福田区园岭街道泥岗西路体育馆西门射击馆
... ..66774237
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                    四川省成都市锌江区王家坝街20号4栋2-4
··.. •89J0131 / 2017-12-31 23:55:1(黄春治
                                     福建泉州市南安市洪濑镇98
                                                                         ***60441121 ***
                                                                                    福建泉州市南安市丰州镇桥中路福源花苑B栋402
                                                                                                                              • 60441121 ..
.....20865515
          2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                 *** 福建省泉州市鲤城区百远干休所1-303
.....ao33160
          2017-12-31 23:55:1(谢青
                                     湖北省武汉市汉阳区七里晴川
                                                                        ***86295559 ***
                                                                                    广东省揭阳市普宁市池尾街道梅园81幢106号祥祥荟庄
                                                                                                                              ... 86795559
...-66(74238
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                     东省深圳市龙岗区深圳市龙岗区五和和磡村住宅二区19-3
...*'66714-239
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                    河北省邯郸市磁县世纪城15号楼
....66774240
          2017-12-31 23:55:1( 小紅书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                 *** 浙江省宁波市海曙区青林湾东A区2幢408
"•••667742A1
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                    上海上海市杨浦区国定路777号上海财经大学
--- 72187532
          2017-12-31 23:55:1(孙娟
                                     山东省临沂市河东区东旭第一城
                                                                        ***63359957 ***
                                                                                    黑龙江省伊春市铁力市桃山林业局
                                                                                                                               ...,53359957 ••
··*'·83991838
          2017-12-31 23:55:1(叶君
                                     中梁国宾府售楼处(公安局对面)
                                                                        ***61276600 ***
                                                                                                                              m6J276600 ···
                                                                                    江苏省盐城市东台市东台市龙晶花园22幢4单元501
•--*20865516
          2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                 *** 福建省福州市台江区义洲新兴小区D9座202
••-20860017
          2017-12-31 23:55:1(小红书福利社
                                     广东省深圳市南山区临海大道53号前海湾保税港区7栋746
                                                                                 *** 湖南省娄底市娄星区众一桂府1期
.-70724131
          2017-12-31 23:55:1(牛奶工作室
                                     广州市海文路11号
                                                                        ***2230639 *** 江苏省苏州市苏州工业园区唯亭镇唯文路5号
                                                                                                                              ***2230639
...,70724132
          2017-12-31 23:55:1(牛奶工作室
                                     广州市海文路11号
                                                                        ***2230639 *** 福建省漳州市云霄县莆美镇宝洋里186号
                                                                                                                              ***2230639
.....70724133
          2017-12-31 23:55:1(牛奶工作室
                                     广州市海文路11号
                                                                        ***2230639 ***
                                                                                    广东省东莞市横沥镇合宝电器有限公司
                                                                                                                              ***2230639
...../0"/24134
          2017-12-31 23:55:1(牛奶工作室
                                     广州市海文路11号
                                                                                    江西省南昌市西湖区丁公路与广场东路交汇处盈石im售楼中心
                                                                        ***2230639 ***
                                                                                                                              ***2230639
•••.•77774726
          2017-12-31 23:55:1( 刘璇璇
                                     广东深圳市宝安区西乡街道麻布新村5巷6号1101
                                                                                    湖南株洲市攸县城关镇雪花社区翻身巷42号(近新天地商业街)
                                                                        ***43117223 ***
                                                                                                                              ***43117223
·.,, ·.179--88
          2017-12-31 23:55:1(张馨
                                     广东深圳市罗湖区南湖街道湘满天宾馆(罗湖口岸店)房均***91806846 ***
                                                                                    安徽芜湖市绿影新村64栋104室《绿影猪蹄店》联系电话
                                                                                                                              ***91806846
          2017-12-31 23:55:1( 白如
•....68030899
                                     河北省廊坊市大厂回族自治县英国宫
                                                                        ***0***2147 ***
                                                                                    云南省大理白族自治州大理市下关镇龙溪路护民善93号(市防疫站)***0***2147
.....55442193
          2017-12-31 23:55:1( 晏丹琳
                                     广东省深圳市龙岗区横岗信义御城
                                                                                 *** 浙江省绍兴市上虞区曹娥工业功能区越窑路(自来水厂斜对面)(兰剑)
          2017-12-31 23:55:1(江苏苏州吴中区郭巷公江苏苏州市吴中区苏州市吴中区通园路1号2号库三楼
'*''*8·•-5337
                                                                                    上海上海市嘉定区和平街254号
·h H8•1t-.-5338
          2017-12-31 23:55:1(江苏苏州吴中区郭春公江苏苏州市吴中区苏州市吴中区通园路1号2号库三楼
                                                                                    上海上海市嘉定区和平街254号
n•o8••·5339
          2017-12-31 23:55-1(江苏苏州吴中区郭巷公江苏苏州市吴中区苏州市吴中区通园路1号2号库三楼
                                                                                    上海上海市寨定区和平街254号
.....4....2993
          2017-12-31 23:55:1(白白叶叶
                                                                        18n •90563 •••
                                     化妆品
                                                                                    成都市郫都区红光镇学府南路8号天邑宏御花园宏园一号门化妆品 18***905631
....4.-2994
          2017-12-31 23:55:1(白白叶叶
                                     化妆品
                                                                        18.. •90563 ...-.
                                                                                    安徽省蚌埠市怀远县淝南乡徐万村化妆品
                                                                                                                              18---905631
... "75985279
          2017-12-31 23:55:1(耿
                                                                        -23679779 •••
                                                                                    浙江省嘉兴市桐乡市一舟大道1号原华实业
                                                                                                                              ***23679779
                                                                        •H8S092521 ...•
••••80166815
          2017-12-31 23:55:1( 海帝童装
                                     福建省宁德市福安市城北富春商城1号楼6号店面
                                                                                    福建省宁德市福安市赛岐镇十二米街永安西路85号海帝童装
                                                                                                                              ***85092521
                                                                        18•••905631 ...
., ... 4•...-2995
          2017-12-31 23:55:1(白白叶叶
                                     化妆品
                                                                                    安徽省黄山市黟县新街恒通电子化妆品
                                                                                                                              18 --- 905631
UIH4• '4'2996
          2017-12-31 23:55:1(白白叶叶
                                                                        :8 • • • 905631
                                     化妆品
                                                                                    广东省东莞市虎门镇白沙三村油烟鸭加工基地恒峰辣味前进三百米(18~~905631
••••4•tt299/
                                                                        lB•-905631 ...
          2017-12-31 23:55:1(白白叶叶
                                     化妆品
                                                                                    陕西省渭南市潼关县太要镇许家城村化妆品
                                                                                                                              18***905631
,..-.771u727
          2017-12-31 23.55:1(刘璇璇
                                     广东深圳市宝安区西乡街道麻布新村5巷6号1101
                                                                        .. ,431J7;>23 "·
                                                                                    江苏苏州市太仓市经济开发区江苏省太仓市陆渡镇东城花苑5园94-1 *** 43117223
.....66774242
          2017-12-31 23 55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                    湖南省长沙市雨花区新星小区竹星苑2栋406
-667742-tJ
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                   山东省潍坊市潍城区宝通西街7166号潍坊医学院浮爆山校区
.....5677,:244 2017-12-31 23 55 1(小红书福利社
                                     郑州市经开区九大街保税物流中心
                                                                                   云南省昆明市官渡区云南省昆明市官渡区昆船工业区昆船花园
          2017-12-31 23:55:1(小红书福利社
                                     郑州市经开区九大街保税物流中心
...-.n66714245
                                                                                   四川省成都市成华区成都理工大学对面上城国际后门中铁隧道股份项目部
......667747-16 2017-12-31 23 55 1(小紅 书福和計
                                     郑州市经开区九大街保税物流中心
                                                                                   山东省济宁市兖州市山拖小区4号楼
          2017-12-31 23:55:1(小红书福利社
.. ,.. 6677'1247
                                     郑州市经开区九大街保税物流中心
                                                                                   辽宁省辽阳市弓长岭区圣典浴都温泉酒店
          2017-12-31 23:55:1( 小红书福利社
                                     郑州市经开区九大街保税物流中心
.....66714248
                                                                                   天津天津市西青区杨柳青柳口路99号
          2017-12-31 23:55 1(小红书福利社
·•....-66774749
                                     郑州市经开区九大街保税物流中心
                                                                                    江西省抚州市临川区环城北路64号水岸明珠
          2017-12-31 23:55:1( 小紅书福利社
.....66774250
                                     郑州市经开区九大街保税物流中心
                                                                                    黑龙江省齐齐哈尔市建华区双华路49号建华医院血液透析室
          2017-12-31 23:55:1(小红书福利社
-66774251
                                     郑州市经开区九大街保税物流中心
                                                                                   广西壮族自治区百色市右江区城东进站路5号市民政局
          2017-12-31 23:55:1( 小紅书福利社
...66774252
                                     郑州市经开区九大街保税物流中心
                                                                                *** 黑龙江省大庆市让胡路区阳光喜城三期E3-2-202
          2017-12-31 23:55:1( 小紅书福利社
                                     郑州市经开区九大街保税物流中心
.....55774253
                                                                                *** 山东省烟台市芝罘区红旗中路186号鲁东大学北区
          2017-12-31 23:55:11 天津配送中心
                                     天津市东南区空港经济区西六道普洛斯物流园85-1号
                                                                                    河北唐山市丰南区丰南镇河北省唐山市丰南区唐山鑫开泰机械厂
.....1901133
          2017-12-31 23:55:1(叶君
                                                                                                                              ..61276600 ..
                                                                        •••6177€r>()O ...
.....83941839
                                     中梁国宾府售楼处(公安局对面)
                                                                                    江苏省盐城市东台市东台市金海饭店东街老曹不锈钢
          2017-12-31 23:55:1(上海配送中心(虹桥)
                                     上海市青浦区徐泾镇华徐公路589号普洛斯虹桥园区
...083569275
                                                                      (ffli!EJJJ!i-£M..•
                                                                                    浙江温州市文成县大峃镇文成县职业中专
          2017-12-31 23:55:1(张先生
                                                                        ... 04817115 •••
...../4076392
                                     黑龙江省哈尔滨市道外区北17三号
                                                                                    山西省大同市城区向阳里街道魏都大道當临宝城8座19号奥斯卡俱乐***04817115 **
          2017-12-31 23:55:1(上海配送中心(虹桥)
                                     上海市青浦区徐泾镇华徐公路589号普洛斯虹桥园区
                                                                      1tttf.mlf/l:i;l-J'liJ...
                                                                                    浙江宁波市海曙区段塘街道南苑街南苑社区229弄100号706室
. • ... 83569276
          2017-12-31 23:55:1(韩飞
                                                                       0371.552496."
                                                                                                                             037J 5524963...
                                     河南省郑州市中原区西元国际广场2207
.....71818319
                                                                                    广东省茂名市信宣市教育城初中(直接放保安亭)
          2017-12-31 23:55:1(韩飞
                                                                       0371-52496...
                                     河南省郑州市中原区西元国际广场2207
                                                                                   辽宁省铁岭市银州区银州路25号铁岭市国税局稽查局
                                                                                                                             :m1.5524953...
.....71818380
```

