Hands-On Introduction to Data-Science in Investment Management

CFA Society of Columbus
October 27th 2020

Tal Sansani, CFA

Data-scientist, quantitative analyst, and investment management consultant.

tsansani@gmail.com
https://github.com/talsan

- **UCLA,** 2001-2005
  - Applied Mathematics, Specialization in Computing

- **American Century Investments,** 2005-2019
  - VP & Senior Quantitative Analyst
  - Quantitative Equity Group: Managing ~ $12B AUM, across a range of Long-Short, Long-Only, Domestic, International products
  - Built & Managed the Data Science & Analytics Team
  - Research Focus: Sentiment signals, economic networks, trend-detection, intangible assets, and other non-financial indicators

- **Investment Management Consultant,** Present
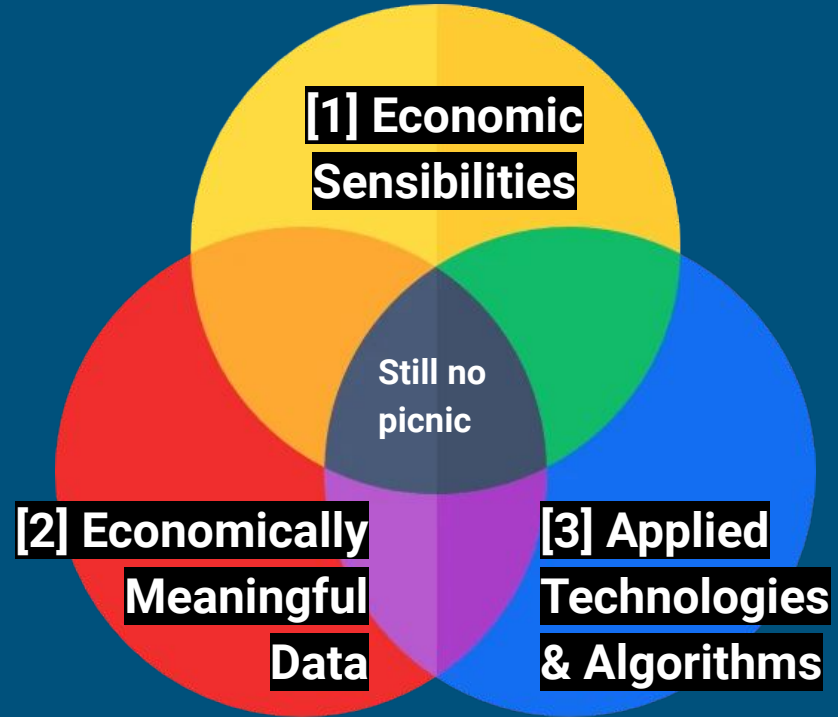 - Help investment teams get more out of data, analytics, and technology.

Contents

Practical Perspectives on Investment Data-Science

(but mostly just personal opinions)

Start with the Investment Problem First
Data-Science & Technology are a Means not an End

- **Without economic sensibilities there is no anchor for understanding performance, incorporating feedback, and communicating with clients.**

- **Overfitting is the enemy**

- **A Data-Set is not a stock-selection idea**
  - Like any other piece of information: it's how about how you contextualize it into your unique process

[1] Economic Sensibilities

Still no picnic

[2] Economically Meaningful Data

[3] Applied Technologies & Algorithms

My Evolving Quant Research Stack

Big Data Storage, Processing & Power Tools:	**AWS** (S3, Athena, RDS, DynamoDB, EC2, Lambda, ElasticSearch)... with nothing against Google or Azure, but as you do more, it becomes more cost-effective to pick one.	AWS
Data Engineering, Web Interfacing & Modelling:	**Python** (but that doesn't mean R can't handle it, it's just a bit more work to make it play nice with other systems).	
Interactive Development Environment (IDE):	**Pycharm** for python data/model work; **rStudio** for everything R	PC
Analytics & Data-Exploration:	**RStudio** is great (tidyverse makes analysis easy), but I still use python because I don't want to manage two environments.	R
Reproducible Research Reports:	**Jupyter Notebook** & **R Markdown** (note: these are amazing tools, but not development environments!!!)	jupyter
Interactive Dashboards:	**Tableau** for traditional dashboards & canned reports (powerful charts in no-time); **Shiny/Dash/Flask** for highly customized research applications (endlessly flexible);	+ableau

How to Get Started?
Little Steps That Add Up

1. **Champion a data-driven culture** and awareness of modern investment applications of data & technology
   - Meaningfully connect data technologists with investment management
2. **Start with a small, targeted, actionable investment idea** or business-process-improvement
   - Avoid big splashes, mass hirings, and giant meetings
3. **Evaluate the value**, the costs, and learn from the pain points
4. **Repeat Steps 1 through 3**
   - Adding resources (talent, data, and technology) along the way

Hands-On Introduction
Textual-Analysis w/ Conference Calls

Part I: Data Sourcing & Structuring
Part II: Three Sentiment Models
Part III: Exploratory Analysis

https://github.com/talsan/calltone

Top / Bottom Sentiment Conference Calls
Over the Last Two Years
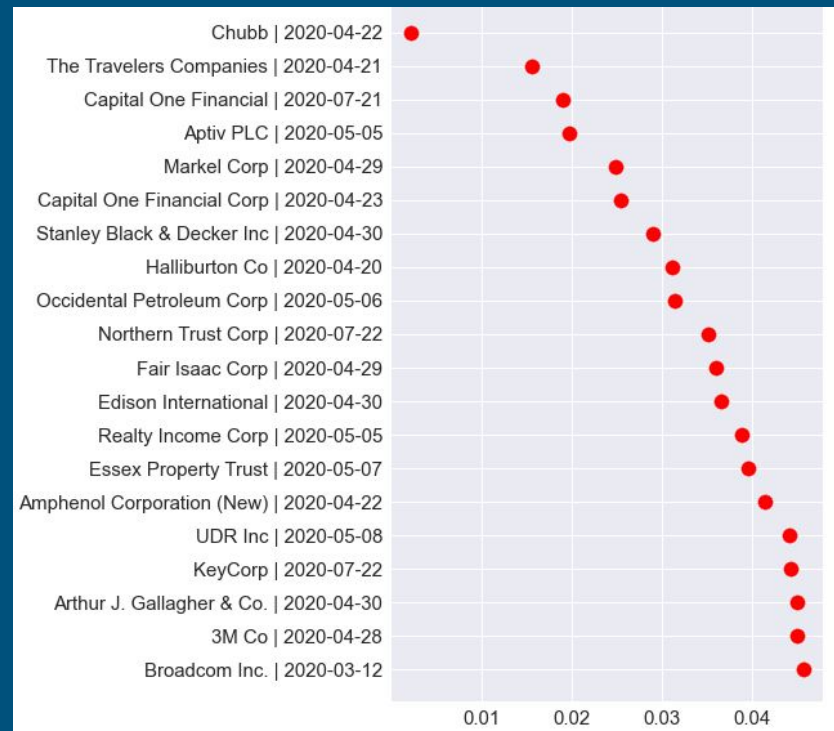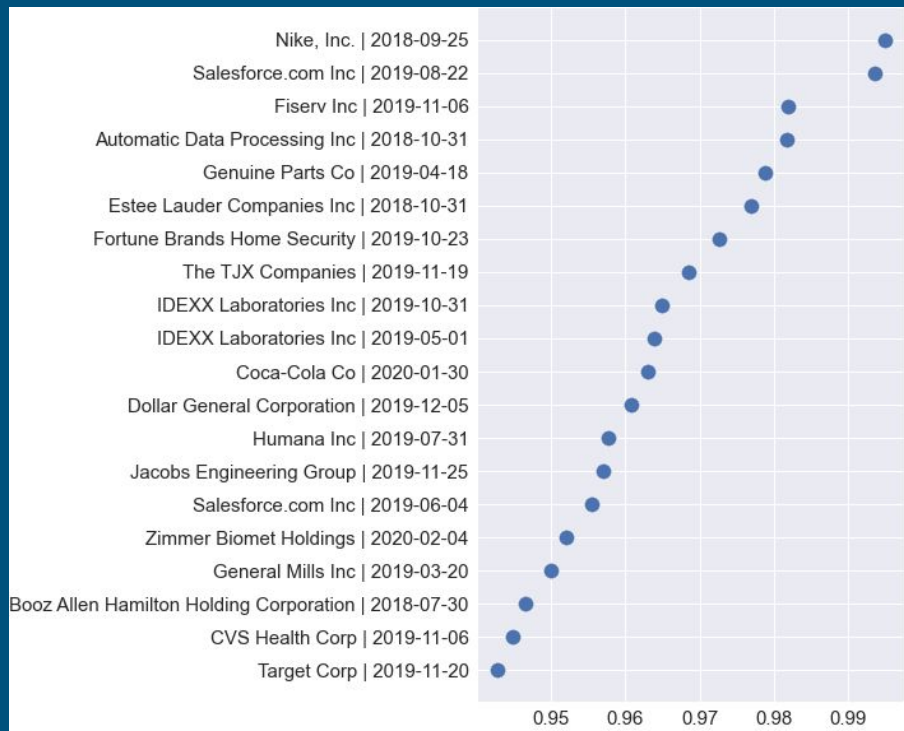
... so how did we get here?

Part II: Quant Research Data Flow

Data Sourcing & Structuring

AWS S3 Data Lake

https://github.com/talsan/foolcalls

Transcripts, metadata, speaker info, etc.

AWS Athena Analytics (SQL-Like)

https://github.com/talsan/ishares

universe, sectors, style, size, cusips, sedols, etc.

Conference Call Sentiment & Analysis
https://github.com/talsan/calltone

https://github.com/talsan/alphavantage

split/dividend adjusted asset returns
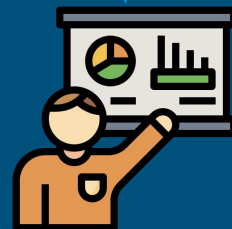
Appendix

Appendix A: Building & Cultivating
an Investment Data-Science Team

Hiring Talent

- Find candidates who show genuine interest in the investing domain
- Set realistic expectations
- Don't get too hung up in specific programming languages or technologies

Cultivating Talent

- Tightly integrate Data scientists and technologists with their investment peers
- Align incentives and blur the multi-disciplinary lines because **it's all connected** -- data, research, portfolio management, analytics -- and serving a common goal

Appendix B: Applications of Conference Calls

Investment Applications

1. **Tone:**
   Reflects management's optimism (or pessimism) about future prospects

2. **Communication Quality / Manipulation:**
   Are they exaggerating? Obfuscating? Omitting key information?

3. **Linkages & Text-Based Peer Groups**

4. **Emerging Themes / Topic Analysis**

Contextual Integration

- **How might this help a quant?**
  - Adds much needed context/"color" to a framework that (usually) invests on numeric/trailing information.
  - Ability to scale systematic insights

- **How might this help a fundamental team?**
  - Bubble up "needle in the haystack" red-flags or insights (especially among less covered names)
  - Streamline research process
  - Provide a "2nd opinion" and potentially reveal biases

Appendix C: Fool.com Structuring Example

KRUS earnings call for the period ending May 31, 2020.

Motley Fool Transcribers (MFTranscribers)
Jul 15, 2020 at 5:30AM

Kura Sushi USA Inc (NASDAQ:KRUS)
Q3 2020 Earnings Call
Jul 14, 2020, 4:30 p.m. ET

Contents:
- Prepared Remarks
- Questions and Answers
- Call Participants

Prepared Remarks:
Operator

Good day, ladies and gentlemen, and thank you for standing by. Welcome to the Kura Sushi USA, Inc. Third Quarter 2020 Earnings Conference Call. At this time, all participants have been placed in a listen-only opened for your questions following the presentation. Please note that this c today, July 14th, 2020.

On the call today we have Jimmy Uba, President and Chief Executive Officer; Officer; and Benjamin Porten, Investor Relations Manager.

I would now like to turn the conference over to Mr. Porten. Thank you. You m

Benjamin Porten -- Investor Relations Manager

Thank you, operator. Good afternoon everyone and thank you all for joining. access to our fiscal third quarter 2020 earnings release. It can be found at ww Relations section. A copy of the earnings release has also been included in ar

"publication_author": "Motley Fool Transcribers (MFTranscribers)",
"publication_time_published": "2020-07-15 05:30:00",
"publication_time_updated": "2020-07-15 05:30:00",
"call_title": "Kura Sushi USA Inc (KRUS) Q3 2020 Earnings Call Transcript",
"call_subtitle": "KRUS earnings call for the period ending May 31, 2020.",
"period_end": "2020-05-31",
"ticker": "KRUS",
"ticker_exchange": "NASDAQ",
"company_name": "Kura Sushi USA Inc",
"fool_company_id": "341476",
"fiscal_period_year": "2020",
"fiscal_period_qtr": "Q3",
"call_short_title": "Q3 2020 Earnings Call",
"call_date": "2020-07-14",
"call_time": "2020-07-14 16:30:00",
"duration_minutes": "47",

Call-Level Meta-Data

"call_transcript": [
  {
    "statement_num": 1,
    "section": "pres",
    "statement_type": "O",
    "speaker": "Operator",
    "role": "",
    "affiliation": "",
    "text": "Good day, ladies and gentlemen, and thank you for standing by.
  },
  {
    "statement_num": 2,
    "section": "pres",
    "statement_type": "P",
    "speaker": "Benjamin Porten",
    "role": "Investor Relations Manager",
    "affiliation": "",
    "text": "Thank you, operator. Good afternoon everyone and thank you all
  },

Transcripts
(with speaker and statement metadata)

Appendix D: AWS Athena Screen Shot