

Intro to text analytics

Dr. Mayank Kejriwal

University of Southern California

A bit about me

- Research Assistant Professor in Industrial and Systems Engineering
- Research Lead in USC Information Sciences Institute
- Previously:
 - PhD and MS in Computer Science (University of Texas at Austin), BS in Computer Engineering and Physics (University of Illinois at Urbana Champaign)
- Research focus is on using AI for social good (fighting human trafficking, resource mobilization after natural disasters...)
 - I'm always looking for good researchers in my group!

Analytics: some broad views

- From Oxford Dictionary:
 - the **systematic** computational analysis of **data or statistics**
 - information resulting from the **systematic** analysis of **data or statistics**

Analytics: some broad views

- From Oxford Dictionary:
 - the **systematic** computational analysis of **data or statistics**
 - information resulting from the **systematic** analysis of **data or statistics**
- From Techopedia:
 - Analytics is the **scientific process** of discovering and **communicating** the meaningful patterns which can be found in data. It is concerned with **turning raw data into insight for making better decisions**. Analytics relies on the application of statistics, computer programming, and operations research in order to quantify and gain insight to the meanings of data. It is especially useful in areas which record a lot of data or information

Analytics: some broad views

- From Oxford Dictionary:
 - the **systematic** computational analysis of **data or statistics**
 - information resulting from the **systematic** analysis of **data or statistics**
- From Techopedia:
 - Analytics is the **scientific process** of discovering and **communicating** the meaningful patterns which can be found in data. It is concerned with **turning raw data into insight for making better decisions**. Analytics relies on the application of statistics, computer programming, and operations research in order to quantify and gain insight to the meanings of data. It is especially useful in areas which record a lot of data or information.
- From 'Innovating with Analytics', MIT Sloan Management Review:
 - "...there is a strong correlation between driving competitive advantage and innovation with analytics and a company's effectiveness at managing the **information transformation cycle**, that is: **capturing data, analyzing information, aggregating and integrating data, using insights to guide future strategy and disseminating information and insights.**"

Analytics is an emerging power

- Analytics 'Have' and 'Have-nots' exist at the levels of departments, companies and personnel
 - Almost every modern **industry** has some company investing in analytics



At MillerCoors, analytics is a central part of a corporate program to change the way the organization conducts business.

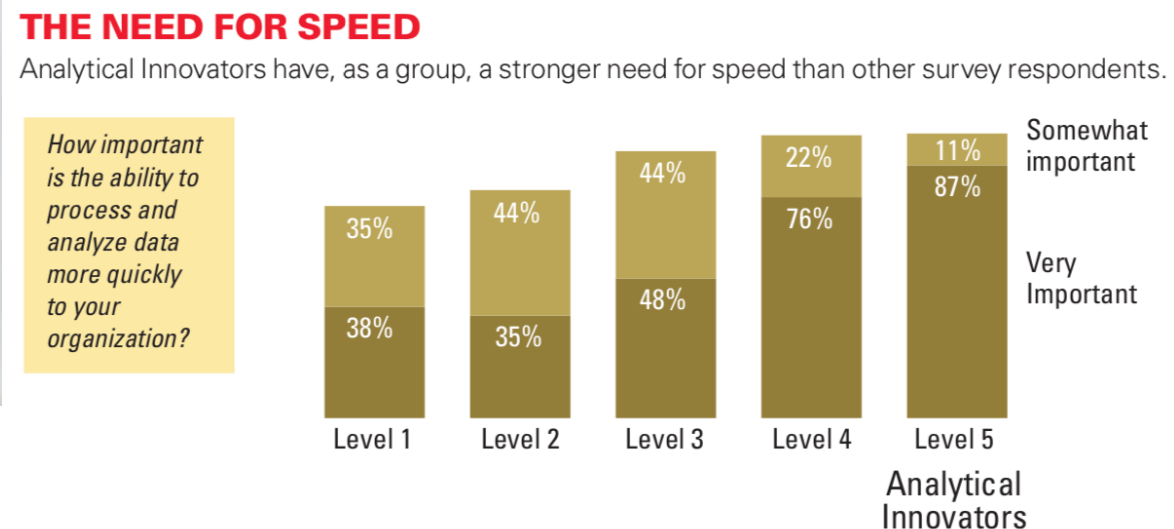
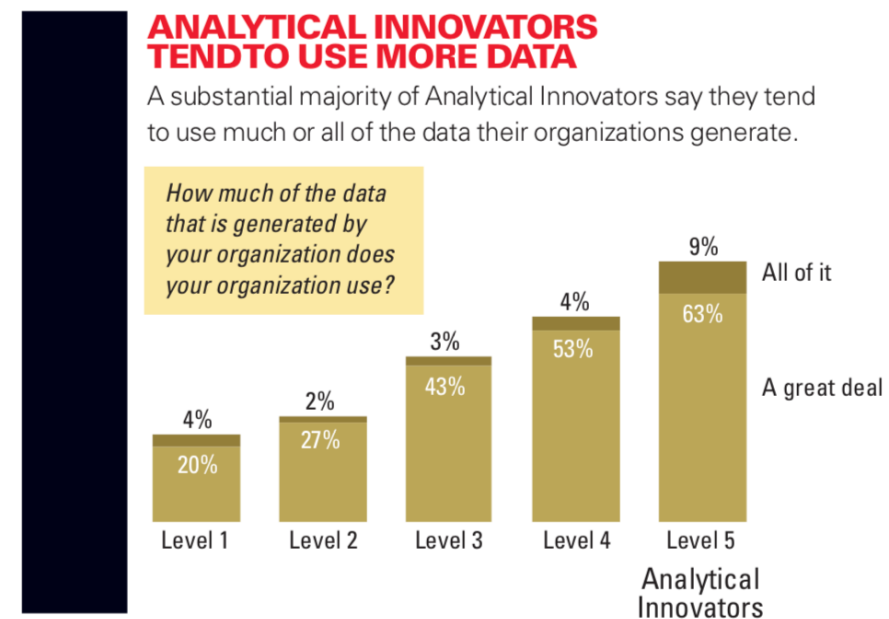
?

THE LEADING QUESTION

How are companies using analytics?

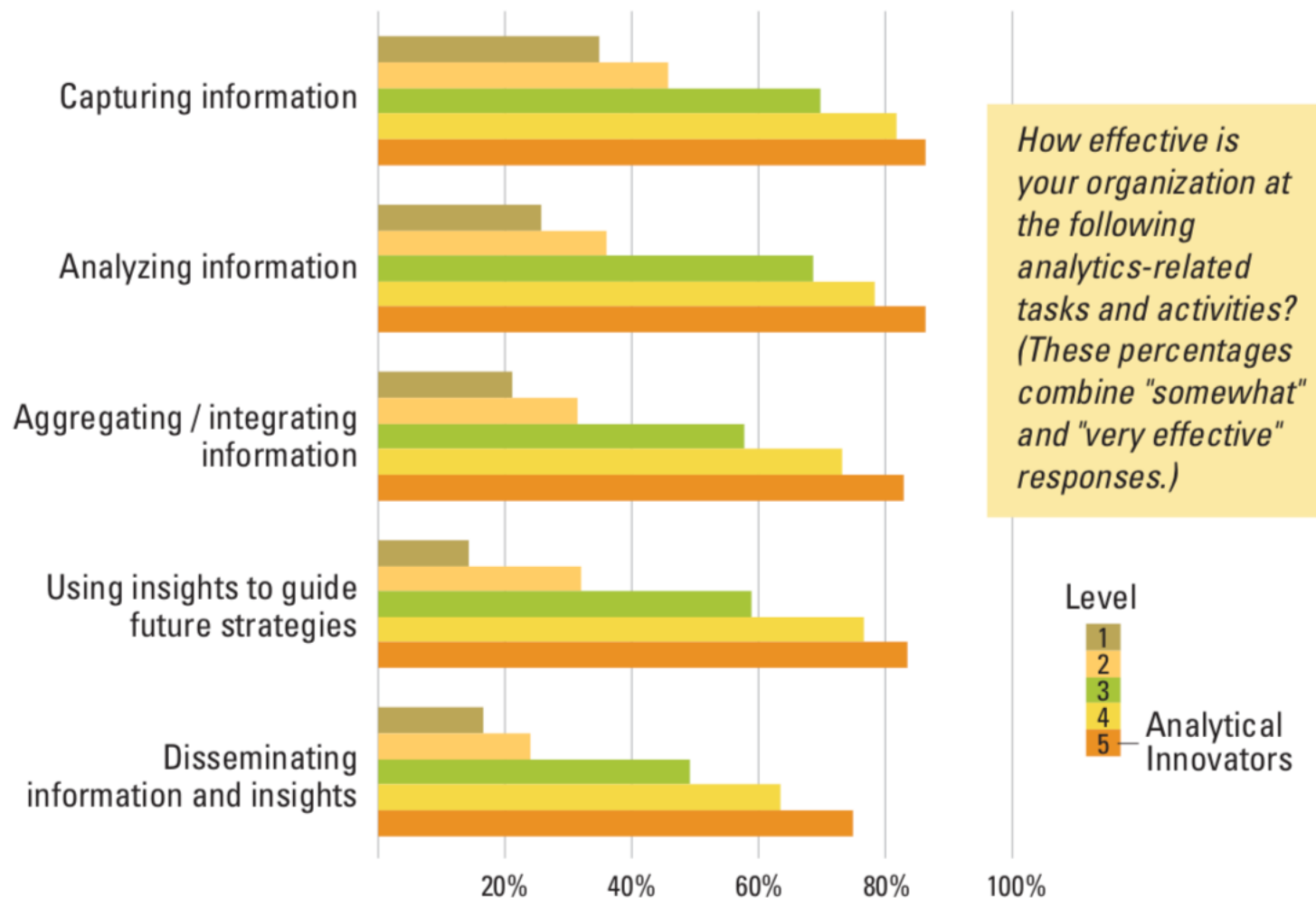
FINDINGS

- ▶ **Sixty-seven percent of survey respondents say analytics gives at least a moderate competitive advantage.**
- ▶ **More than half agree somewhat or strongly that analytics improves the organization's innovation abilities.**
- ▶ **Some respondents report that analytics is shifting the organization's power structure.**



MANAGING THE INFORMATION TRANSFORMATION CYCLE

Analytical Innovators report that they are more effective at managing all stages of what we call the information transformation cycle, that is: capturing data, analyzing information, aggregating and integrating data, using insights from analytics to guide future strategy and disseminating information and insights.



1. Start with a **headline** that clearly names your product (or service) and states what it is going to do
2. Write an **opening paragraph** that gives the most important details about your product (or service), how consumers will benefit from it, and the problem that it's going to solve. You should write this as if it's the only part anyone's going to read
3. Include **quotes** from your company spokesperson and from a hypothetical satisfied customer
4. **Conclude** by briefly explaining how the customer can learn more or begin using your product (or service)
5. Also, keep your Amazon press release simple! 3–4 sentences for most paragraphs. Don't use technical terms, which are difficult to understand. Finally, no more than one page and integrate with a **F.A.Q.** section, which I personally find very important, as it is useful to anticipate challenging questions

<https://www.sec.gov/Archives/edgar/data/1018724/000119312516530910/d168744dex991.htm>



<https://medium.com/@IndianaStyle/amazon-press-release-how-to-55d61188ecdd>

“If you can't come up with a compelling press release,

the product you want to launch probably isn't worth making”

Analytics is inherently applied

- In industry, analytics forces you to ask who your customer is (which in turn forces you to ask what the *problem* is)
 - Inevitably, you must be able to connect the dots between the customer for your ‘analytical product’ and the actual customer

Former Apple CEO Steve Jobs, *“When you start looking at a problem and it seems really simple ..., you don’t really understand the complexity of the problem. And your solutions are way too oversimplified. Then you get into the problem, and you see it’s really complicated. And you come up with all these convoluted solutions ... That’s where most people stop. The really great person will keep on going and find ... the **key, underlying principle of the problem** and come up with a **beautiful elegant solution** that works.”*

The role of text

Applications

- Security
 - Monitoring and analysis of online plain text sources such as [Internet news](#), [blogs](#) (and so on) for [national security](#) purposes
 - Software for tracking and monitoring terrorist activities
- Biomedical
 - Use large patient textual datasets in clinical fields, along with datasets of demographic information in population studies and adverse event reports, to facilitate studies and precision medicine
 - PubGene: a search engine that combines biomedical text mining with network visualization
- Online media
 - Media companies such as the Tribune Company using text analytics to provide readers with better search experience, in turn increasing stickiness and revenue
 - Editors also benefit: by being able to share, associate and package news across properties, significantly increasing opportunities to monetize content

Cont'd

- Business/Marketing
 - Heavily in use, but still in the wild west (many applications waiting to be discovered)
 - Customer Relationship Management (CRM): Research has found a 5% increase in customer retention boosts lifetime customer profits by 50% on average across multiple industries, as well as a boost of up to 90% within specific industries such as insurance
- Digital humanities and computational social science
 - Key enabling technologies have been parsing, machine translation, topic categorization, and machine learning
 - Analysis of readability, gender bias and topic bias using content analysis and big data
- Many more!