Module 02
Data Thinking
Lecture 2

Wan Fang

Southern University of Science and Technology

Agenda

- Data, Narrative, Visualization
  - The 3-minute Story
  - The Big Idea
  - Examples & Exercises

- Four Types of Data Analytics
  - **Descriptive**, which answers the question, "*What happened?*"
  - **Diagnostic**, which answers the question, "*Why did this happen?*"
  - **Predictive**, which answers the question, "*What might happen in the future?*"
  - **Prescriptive**, which answers the question, "*What should we do next?*"
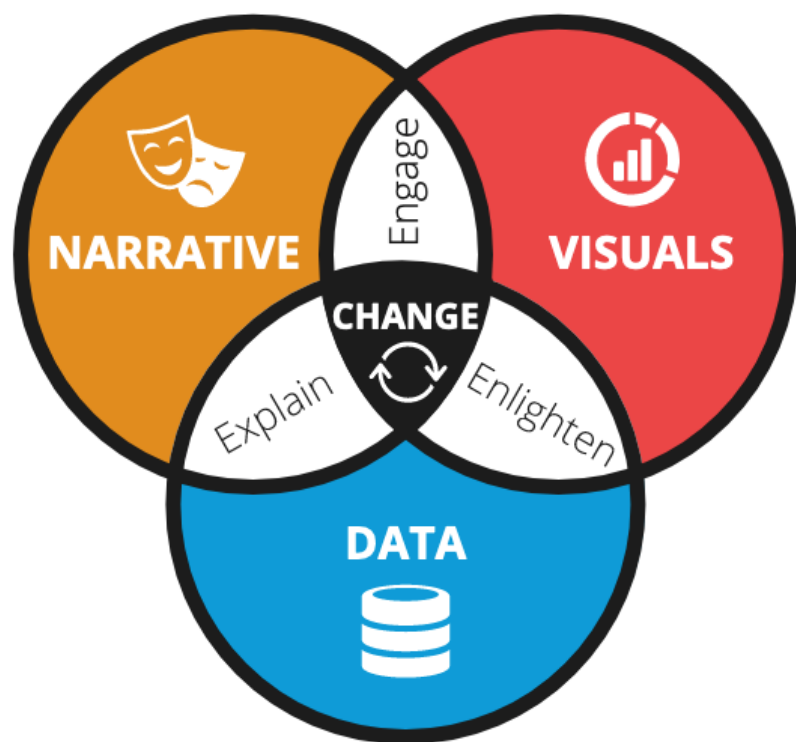

- Student Presentation

Data, Narrative, Visualization

Wan Fang

Southern University of Science and Technology

[Adapted from Storytelling with Data by Cole Nussbaumer Knaflic]

Data Stories that Drive Changes

Data stories are proven to be more memorable, persuasive, and engaging than statistics alone.
- *But how do we tell data stories?*

"Storytelling is the most powerful way to put ideas into the world today."

Today's corporate world is rife with poorly designed slides cluttered with inappropriate charts, and data storytelling is the key to converting data to insights that drive actions and value.

Narrative Structure

- **Most stories follow a common arc …**
  - A protagonist who faces a complication goes on a journey of resolving a difficulty before returning to their normal lives.
    - Aristotle's Tragedy Structure, Campbell's Hero's Journey, and Freytag's Pyramid are variations of these common patterns in stories.

- **Data stories stand to benefit from the age-old narrative structure.**
  - A data story begins by setting the scene of the current situation, proceeds by providing insights that lead up to the central insight, and ends with relevant recommendations.
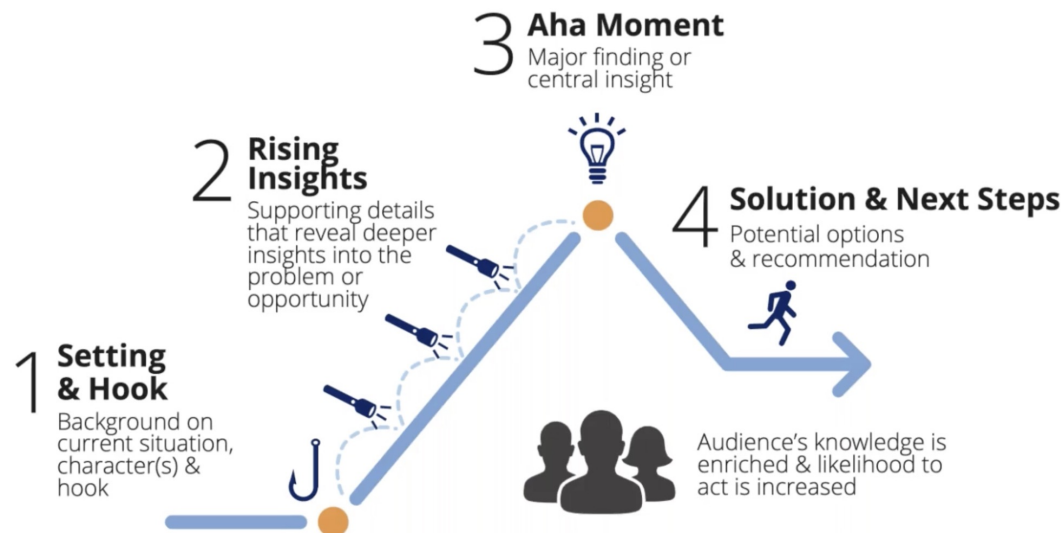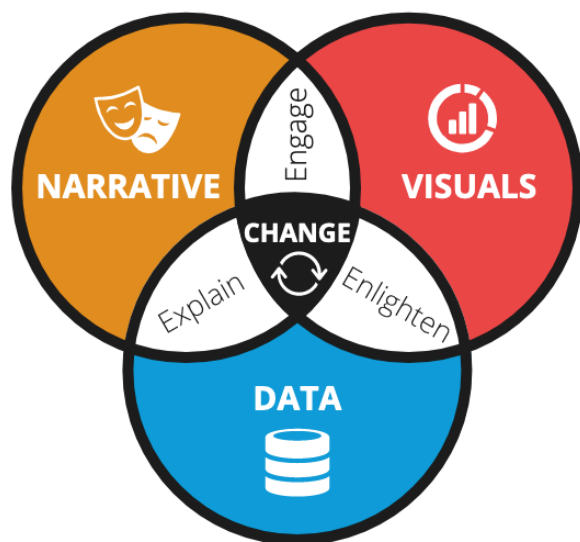
Data

- Analysts facing a deluge of data may be tempted to squeeze all the available data into one chart.
 - *Those who do so risk muddling their data stories with noise and confusing their audience.*

- That is why selecting the right data in a data story is crucial.
  - *Data storytellers should remove or aggregate secondary data points, as the examples below demonstrate.*

Identify Right Data for Your Data Story

Explicitly demonstrating that the revenue per customer is falling is a better choice than plotting the total revenue and customer.
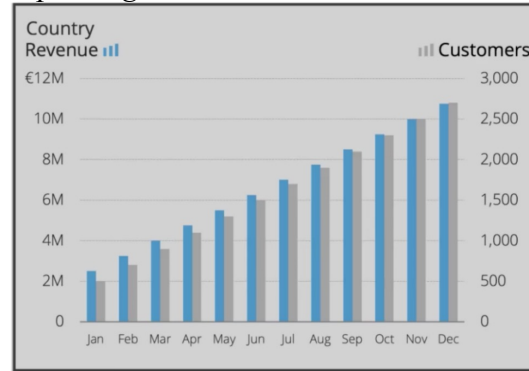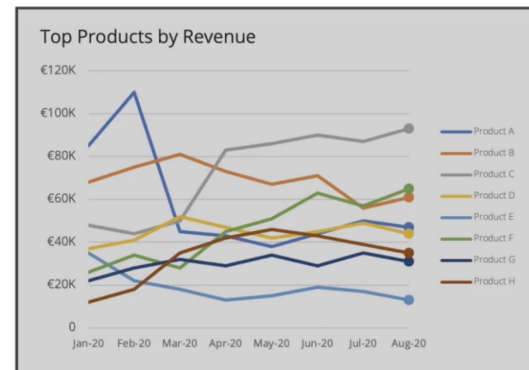
Calculated metrics may be more insightful than total values.
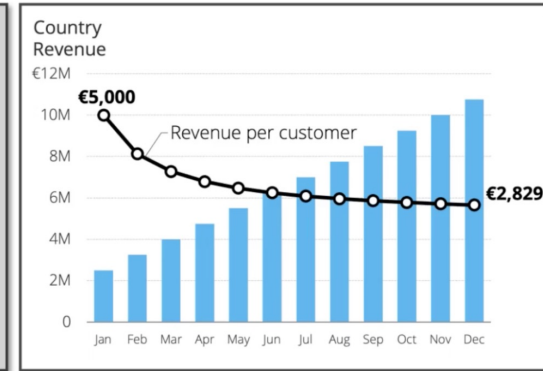
Remove Surplus Data that Isn't Needed

Often, less is more when it comes to showing multiple charts in one graph.

Ask yourself **what is essential** to making your point. Remove what's unnecessary.
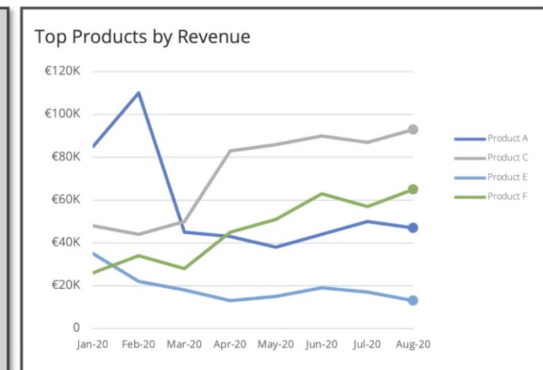
Visuals

1. Choose the appropriate visualization

- Visualizations differ in effectiveness based on the context of the data story.

- The key is to select the best chart based on the message.
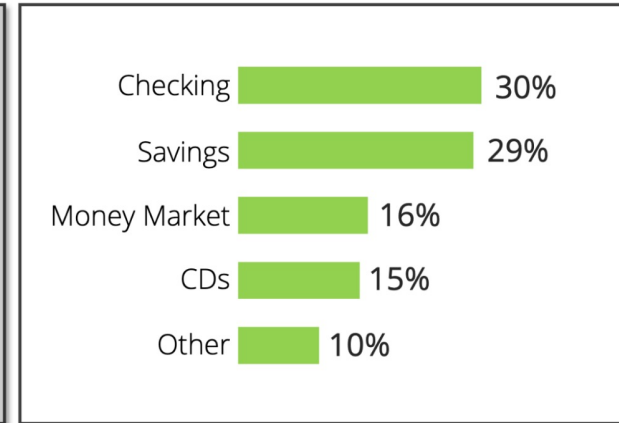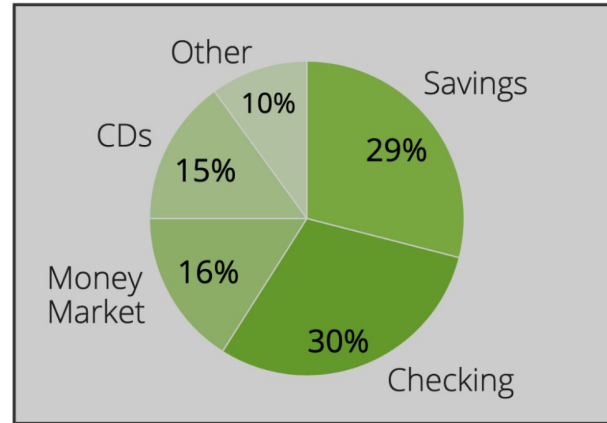
Bar charts are superior to pie charts in communicating differences in proportions
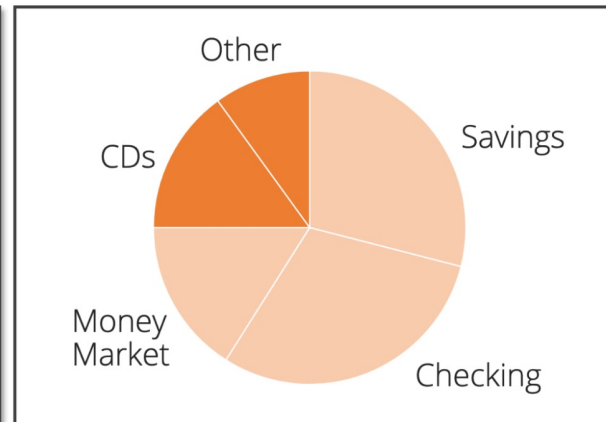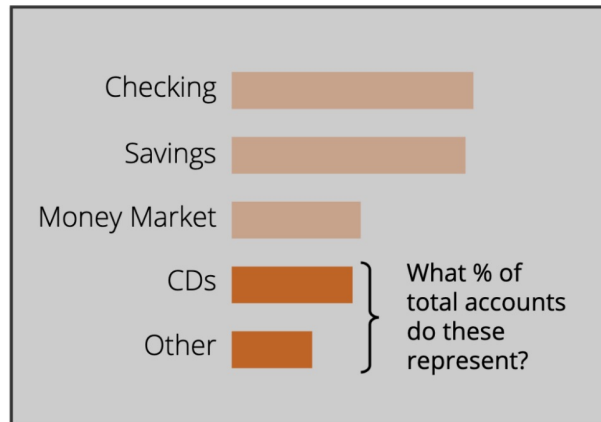
Pie charts are a better choice when illustrating the sum of proportions across categories

Visuals

2. Calibrate the visuals to the message

- A proper data story arranges the visuals such that they meet the audience's needs.
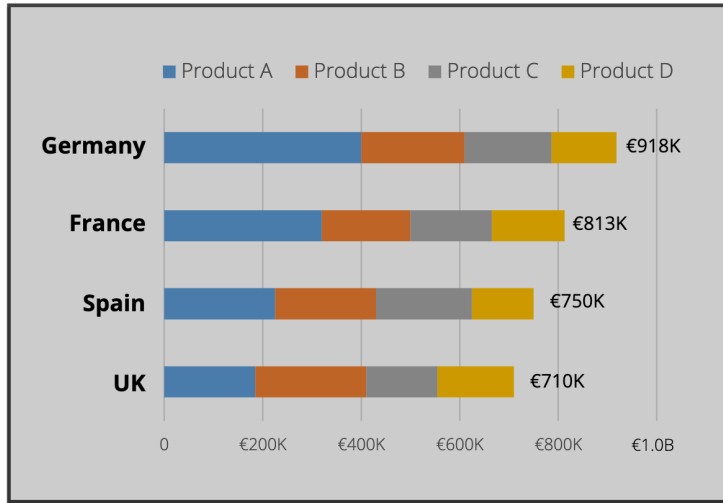
A comparison of the total revenue from each country

A comparison of the revenue of individual products across countries, especially for product B
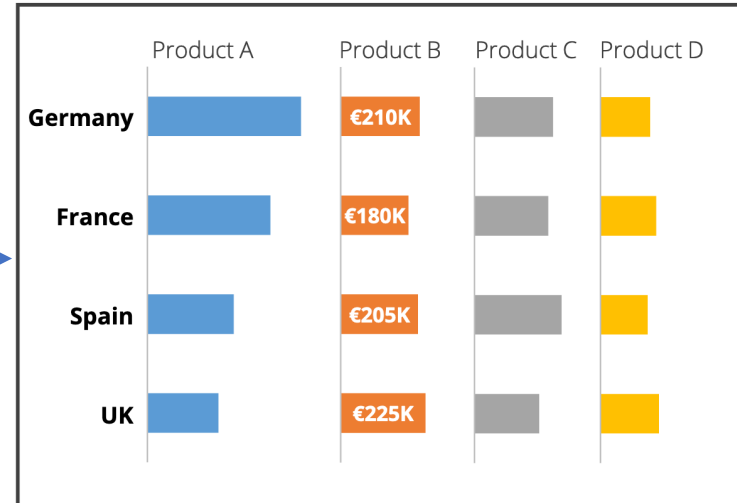
A comparison across products for each segment

A comparison across segments for each product
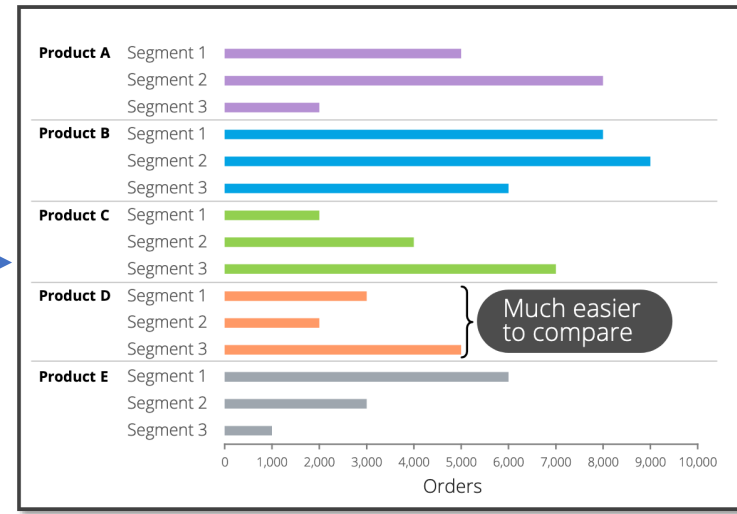
Visuals

3. Focusing the attention of the audience

• The audience pays attention to what stands out in a chart. Highlighting the key points distills the signal from the noise, allowing the audience to garner insights from charts quickly.
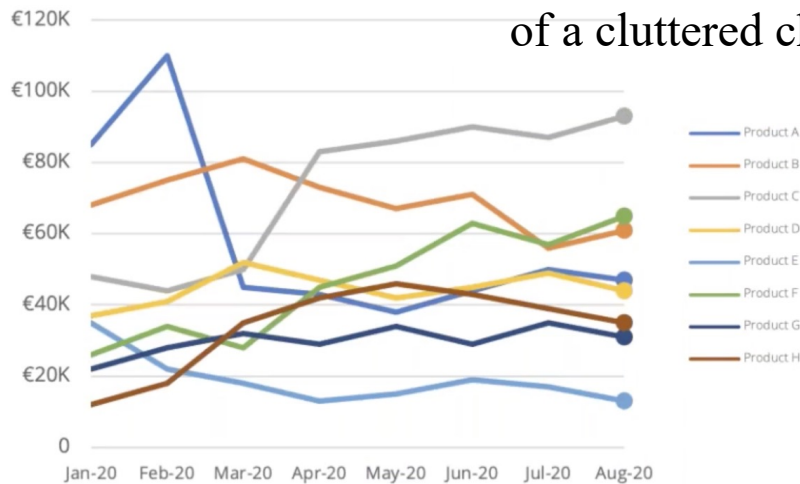
A common example of a cluttered chart
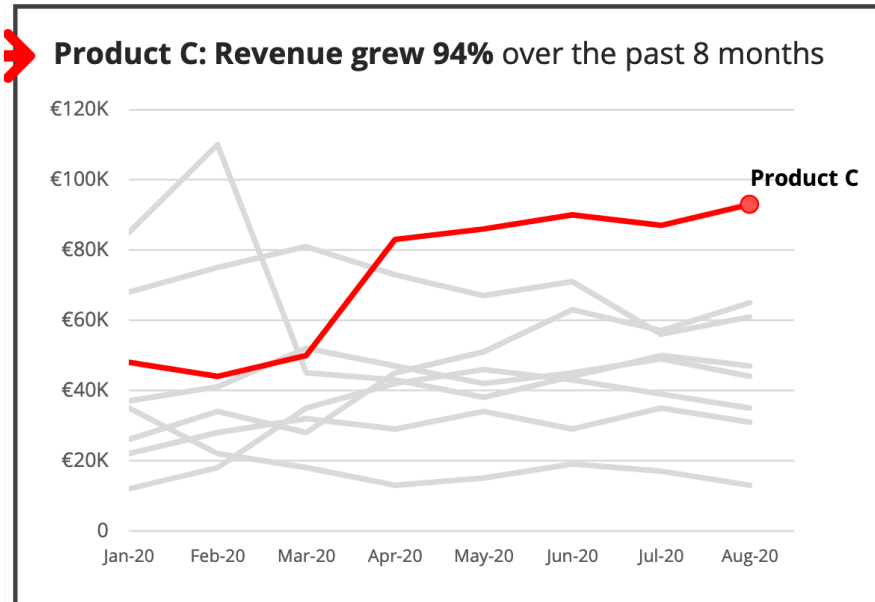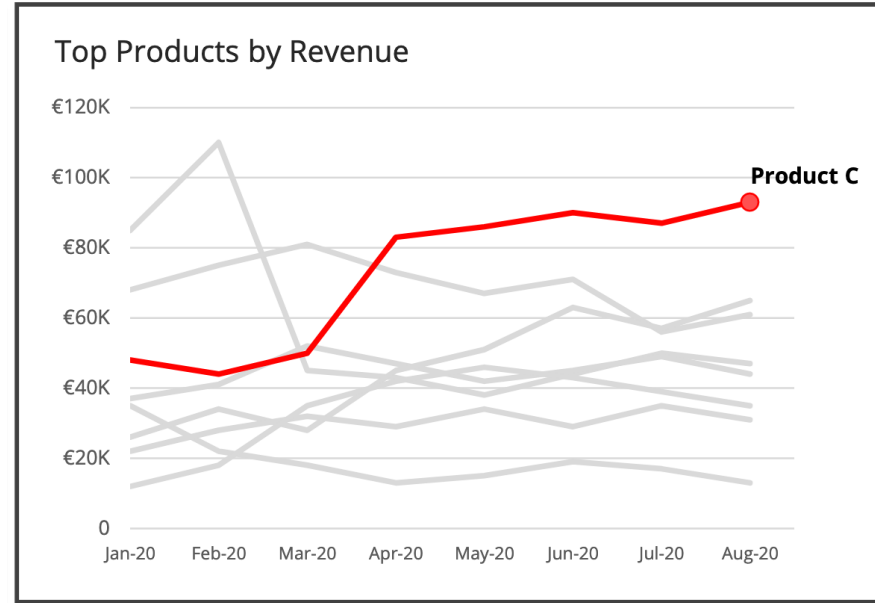
The 3-minute story & Big Idea

- The idea behind each of these concepts is that you are able to boil the "*so-what*" down to a paragraph and, ultimately, to a single, concise statement.
  - You must really know your stuff—*know what the most important pieces are* as well as *what isn't essential* in the most stripped-down version.

- While it sounds easy, being concise is often more challenging than being verbose.
  - Mathematician and philosopher Blaise Pascal recognized this in his native French, with a statement that translates roughly to "*I would have written a shorter letter, but I did not have the time*" (a sentiment often attributed to Mark Twain).

3-minute story

- The 3-minute story is exactly that:
  - ***if you had only three minutes to tell your audience what they need to know, what would you say?***

- A great way to ensure you are clear on and can articulate the story you want to tell.
  - Being able to do this removes you from dependence on your slides or visuals for a presentation.
  - This is useful in the situation where your boss asks you what you're working on or if you find yourself in an elevator with one of your stakeholders and want to give her the quick rundown.
  - Or if your half-hour on the agenda gets shortened to ten minutes, or to five.

- If you know exactly what it is you want to communicate, you can make it fit the time slot you're given, even if it isn't the one for which you are prepared.

Big Idea

- The Big Idea boils *the so-what* down even further: ***to a single sentence***.
  - This is a concept that Nancy Duarte discusses in her book, *Resonate* (2010).

- The Big Idea has three components:
  1. It must articulate your unique point of view;
  2. It must convey what's at stake; and
  3. It must be a complete sentence.

- Let's consider an illustrative 3-minute story and Big Idea,
  - *Leveraging the summer learning program on science example* that was introduced previously.
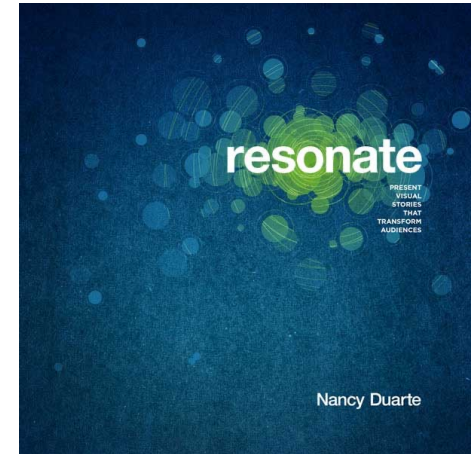
Let's consider a specific example

- **Imagine you are a fourth-grade science teacher.**

- You just wrapped up <u>an experimental pilot summer learning program</u> on science that was aimed at giving kids exposure to the unpopular subject.

- You <u>surveyed the children</u> at the onset and end of the program to understand whether and how perceptions toward science changed.

- You <u>believe</u> the data shows a great success story.

- You <u>would like to continue</u> to offer the summer learning program on science going forward.

Please identify the *Who*, *What*, and *How*

- *who we have identified as our audience,*
- *what we need them to know and do, and*
- *the data that will help us make our case*

Identify the *Who*, *What*, and *How*

- **Who**: The budget committee that can approve funding for continuation of the summer learning program.

- **What**: The summer learning program on science was a success; please approve budget of $X to continue.

- **How**: Illustrate success with data collected through the survey conducted before and after the pilot program.

The *Who*, *What*, and *How* identified
- *who we have identified as our audience,*
- *what we need them to know and do, and*
- *the data that will help us make our case*

3-minute story

- Try writing one yourself …

3-minute story: TLDR

- A group of us in the science department were brainstorming about how to resolve an ongoing issue we have with incoming fourth-graders. It seems that when kids get to their first science class, they come in with this attitude that it's going to be difficult and they aren't going to like it. It takes a good amount of time at the beginning of the school year to get beyond that. So we thought, what if we try to give kids exposure to science sooner? Can we influence their perception? We piloted a learning pro- gram doing just that. We invited elementary school students large group of second- and third-graders. Our goal was xposure to science in hopes of forming positive perception were successful, we surveyed the students before and after the program. We found that, going into the program, the biggest segment of students, 40%, felt just "OK" about science, whereas after the program, most of these shifted into positive perceptions, with nearly 70% of total students expressing some level of interest toward science. We feel that this demonstrates the success of the program and that we should not only continue to offer it, but also to expand our reach with it going forward.

*TL;DR or TLDR stands for "**Too Long; Didn't Read**."*

3-minute story: An Example

A group of us in the science department were brainstorming about how to resolve an ongoing issue we have with incoming fourth-graders.	General Background
It seems that when kids get to their first science class, they come in with this attitude that it's going to be difficult and they aren't going to like it.	Specific Elaboration
It takes a good amount of time at the beginning of the school year to get beyond that.	Setup the Conflict
So we thought, what if we try to give kids exposure to science sooner? Can we influence their perception?	**Raise the Question**
We piloted a learning program last summer aimed at doing just that.	Initiate the solution
We invited elementary school students and ended up with a large group of second- and third-graders.	What we did
Our goal was to give them earlier exposure to science in hopes of forming positive perception.	Why we did it
To test whether we were successful, we surveyed the students before and after the program.	How we did it
We found that, going into the program, the biggest segment of students, **40%**, felt just "OK" about science, whereas after the program, most of these shifted into positive perceptions, with nearly **70%** of total students expressing some level of interest toward science.	We found that …
We feel that this demonstrates the success of the program and that we should not only continue to offer it, but also to expand our reach with it going forward.	**What we want …**

Big Idea

- Once again …
- Try writing one yourself …

Big Idea

- The pilot summer learning program was successful at improving students' perceptions of science and, because of this success, we recommend continuing to offer it going forward; please approve our budget for this program.

Four Types of Data Analytics

Wan Fang

Southern University of Science and Technology

[Adapted from Harvard Business School Online by Catherine Cote]

Data Analytics To Improve Decision-making

- Data is a powerful tool that's available to organizations at a staggering scale.
  - When harnessed correctly, it has the potential to _drive decision-making_, _impact strategy formulation_, and _improve organizational performance_.

2020 GLOBAL STATE OF ENTERPRISE ANALYTICS

MINDING THE DATA-DRIVEN GAP

MicroStrategy
Intelligence Everywhere

94% SAY **DATA** AND **ANALYTICS** IS IMPORTANT TO THEIR BUSINESS **GROWTH** AND **DIGITAL TRANSFORMATION**

Question: What benefits has your organization realized through its analytics use?

Benefit	Percentage
Improved efficiency and productivity	64%
Faster, more effective decision making	56%
Better financial performance	51%
Identification and creation of new product and service revenue	46%
Improved customer acquisition and retention	46%
Improved customer experiences	44%
Competitive advantage	43%

Four Types of Analytics

- **Analytics** is used to extract meaningful insights from data that can drive decision-making and strategy formulation.
  - There are **four types of analytics** you can leverage depending on the data you have and the type of knowledge you'd like to gain.

Descriptive analytics
- looks at data to examine, understand, and describe something that's already happened.

Diagnostic analytics
- goes deeper than descriptive analytics by seeking to understand the "why" behind what happened.

Predictive analytics
- relies on historical data, past trends, and assumptions to answer questions about what will happen in the future.

Prescriptive analytics
- identifies specific actions an individual or organization should take to reach future targets or goals.

Descriptive Analytics

- **The simplest type of analytics and the foundation the other types are built on.**
  - It allows you to pull trends from raw data and succinctly describe what happened or is currently happening.
  - If you're new to the field of data analytics, descriptive analytics is an accessible and rewarding place to start.

- **Descriptive analytics answers the question, "*What happened?*"**
  - For example, imagine you're analyzing your company's data and find there's a seasonal surge in sales for one of your products: a video game console.
  - Here, descriptive analytics can tell you, "This video game console experiences an increase in sales in October, November, and early December each year."

- **Data visualization is a natural fit for communicating descriptive analysis**
  - because charts, graphs, and maps can show trends in data—as well as dips and spikes—in a clear, easily understandable way.

Descriptive Analytics: Examples

Traffic and Engagement Reports

- One example of descriptive analytics is reporting. If your organization tracks engagement in the form of social media analytics or web traffic, you're already using descriptive analytics.
  - These reports are created by taking raw data—generated when users interact with your website, advertisements, or social media content—and using it to compare current metrics to historical metrics and visualize trends.

- The three other analytics types can then be used to determine why traffic from each source increased or decreased over time, if trends are predicted to continue, and what your team's best course of action is moving forward.

Descriptive Analytics: Examples

Demand Trends

- Descriptive analytics can also be used to identify trends in customer preference and behavior and make assumptions about the demand for specific products or services.

变与不变：2020年中国短视频+教育发展展望   生成海报   报告下载

2020-06-09   短视频   教育

《变与不变：2020年中国短视频+教育发展展望》成文于短视频平台成为新的教育流量阵地，短视频+教育如火如荼发展的背景下。通过对短视频+教育生态的全面分析，探讨短视频的风靡对教育主体、教育内容、内容传播将产生的影响及短视频生态内的教育机会，短视频+教育的市场规模和发展趋势，供行业内企业、个人、投资机构等各方决策提供参考。

iResearch
艾 瑞 咨 询

变与不变
中国短视频+教育发展展望

2020年

Diagnostic Analytics

- The process of using data to determine the causes of trends and correlations between variables.
    - It can be viewed as a logical next step after using descriptive analytics to identify trends.
    - Diagnostic analysis can be done manually, using an algorithm, or with statistical software (such as Microsoft Excel).

- Several concepts to understand before diving into diagnostic analytics:
    - hypothesis testing,
    - the difference between correlation and causation, and
    - diagnostic regression analysis.

Diagnostic Analytics

Hypothesis Testing

- The statistical process of proving or disproving an assumption.
    - Having a hypothesis to test can guide and focus your diagnostic analysis.
    - The hypothesis directs your analysis and serves as a reminder of what you're aiming to prove or disprove.

- Hypotheses can be future-oriented, but these aid predictive or prescriptive analytics.
    - "*If we change our company's logo, more people in North America will buy our product.*"

- When conducting diagnostic analytics, hypotheses are historically-oriented.
    - "*I predict this month's decline in sales was caused by our product's recent price increase.*"

Diagnostic Analytics

Correlation vs. Causation

- When exploring relationships between variables, it's important to be aware of the distinction between correlation and causation.
  - If two or more variables are correlated, their directional movements are related.
  - If two variables are positively correlated, it means that as one goes up or down, so does the other.
  - Alternatively, if two variables are negatively correlated, one variable goes up while the other goes down.

- The key in diagnostic analytics is remembering that
  - ***just because two variables are correlated, it doesn't necessarily mean one caused the other to occur.***

- If your organization is able to dedicate resources to running controlled experiments, you may be able to determine causation between variables.
  - While determining causation is ideal, correlation can still offer the insight needed to make sense of your data and use it to make impactful decisions.

Diagnostic Analytics

Diagnostic Regression Analysis

- Regression analysis can be used to determine the relationship between two or more variables.
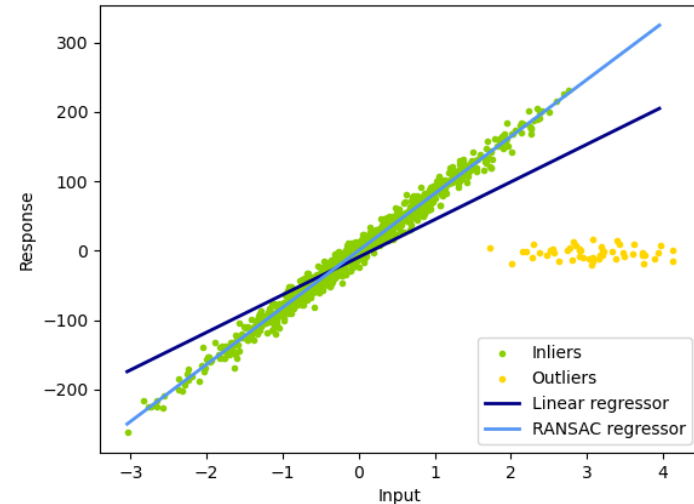
- Regression allows us to gain insights into the structure of that relationship and provides measures of how well the data fit that relationship.

- When regression analysis is used to explain the relationships between variables in a historical context, that's an example of diagnostic analytics.
  - The regression can then be used to develop forecasts for the future, which is an example of predictive analytics.

Diagnostic Analytics: Examples

Examining Market Demand

- One use case of diagnostic analytics is determining the reasons behind product demand.
    - For example, take meal kit subscription company **HelloFresh**.
    - The company gathers millions of data points from global users, including information about geographic location, disclosed demographic data, meal type, flavor preferences, and typical order cadence and timing.

- HelloFresh's team uses data to **identify relationships between trends in customer attributes and behavior**.
    - As a hypothetical example, imagine the HelloFresh team identifies a spike in fish-based recipe orders.
    - After conducting diagnostic analysis, they find that the attributes most highly correlated with ordering fish recipes are identifying as female and living in the northeastern United States.

- From there, the team could conduct market research with that specific demographic to learn more about the demand for fish recipes.
    - Was it caused by a recent scientific study touting the health benefits of fish for women?
    - Perhaps people who live in the northeastern United States have a refined palate for seafood because they live relatively close to the Atlantic Ocean.
    - Their reasoning could provide impactful insights to HelloFresh.

- Dipping into the other types of analytics,
    - the team could also consider whether the trend is expected to continue (predictive analytics) and if it's worth the effort and money to create more fish-based recipes to cater to this audience's preference (prescriptive analytics).

Predictive Analytics

- **<u>The use of data to predict future trends and events.</u>**

  - It uses historical data to forecast potential scenarios that can help drive strategic decisions.

- The predictions could be for the **<u>near future</u>** or the **<u>more distant future</u>**, such as predicting your company's cash flows for the upcoming year.

  - For instance, predicting the malfunction of a piece of machinery later that day

- Predictive analysis can be conducted **<u>manually</u>** or **<u>using machine-learning algorithms</u>**.

  - Either way, historical data is used to make assumptions about the future.

Predictive Analytics: Examples

Manufacturing: Preventing Malfunction

不同类型的维护

- 响应式维护 —— 出问题了再进行维护
  - 例子：替换汽车蓄电池
  - 问题：非预期的失效可能导致高昂的代价，潜在危险

- 定期维护 —— 按照固定的时间间隔定定期维护
  - 例子：5000 英里汽车更换机油
  - 问题：不必要的维护导致浪费；不能消除所有的失效问题

- 预测式维护 —— 预测问题出现的的时机
  - 例子：预测某种ＧＭ汽车模型的潜在问题，通过检测电池，油泵，电机等信息
  - 问题：提高复杂设备的预测精度非常困难

Predictive Analytics: Examples

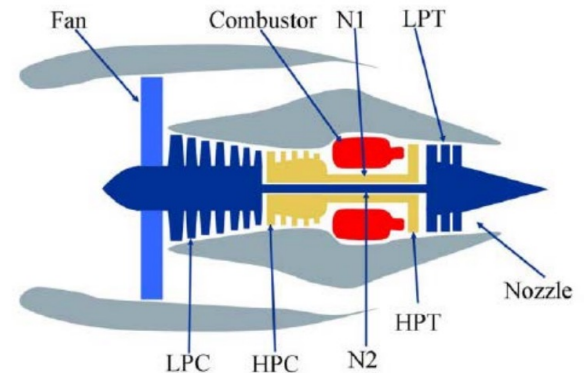**Manufacturing: Preventing Malfunction**

涡轮发动期的预测式维护

从 100 台相同型号发动机上采集的传感器数据

在问题出现之前，预测和修正潜在问题
- 引入和分析历史传感数据
- 训练模型预测失效的时机
- 部署模型到实时传感数据之中
- 实时预测失效事件

Data provided by NASA PCoE
http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

Fan   Combustor   N1   LPT
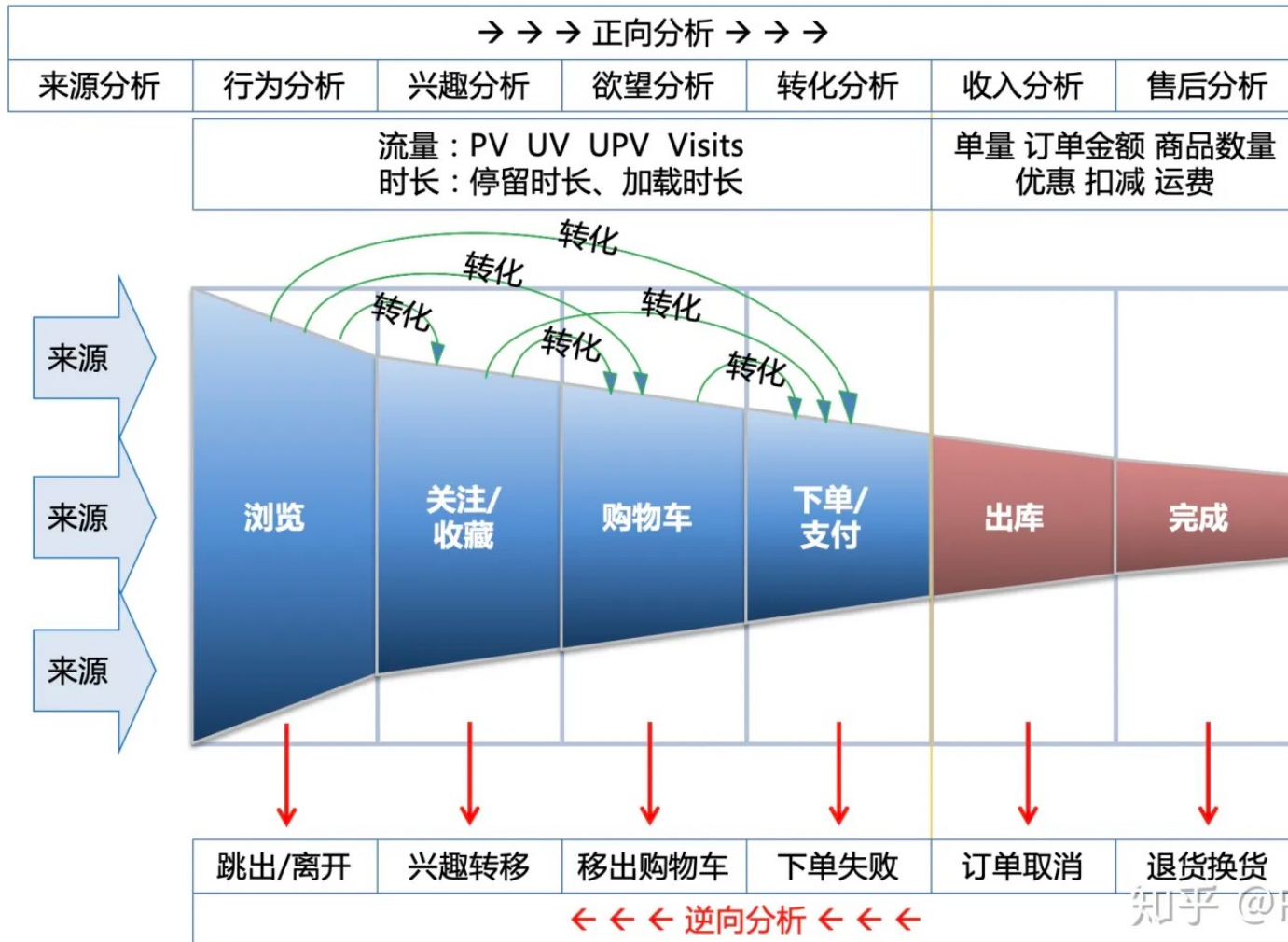
LPC   HPC   N2

HPT

Nozzle

Predictive Analytics: Examples

Marketing: Behavioral Targeting

- In marketing, consumer data is abundant and leveraged to create content, advertisements, and strategies to better reach potential customers where they are.
    - By examining historical behavioral data and using it to predict what will happen in the future, you engage in predictive analytics.

- Predictive analytics can be applied in marketing to **forecast sales trends** at various times of the year and plan campaigns accordingly.

- Additionally, historical behavioral data can help you predict a lead's likelihood of moving down the funnel **from awareness to purchase.**

Predictive Analytics: Examples

Marketing: Behavioral Targeting

			→ → → 正向分析 → → →			
来源分析	行为分析	兴趣分析	欲望分析	转化分析	收入分析	售后分析

流量：PV UV UPV Visits 　　　 单量 订单金额 商品数量
时长：停留时长、加载时长 　　　　优惠 扣减 运费

转化
转化
转化
转化
转化
转化

来源						
来源	浏览	关注/收藏	购物车	下单/支付	出库	完成	售后
来源						

跳出/离开	兴趣转移	移出购物车	下单失败	订单取消	退货换货

← ← ← 逆向分析 ← ← ←

知乎 @Raman

Prescriptive Analytics

- The process of using data to determine an optimal **course of action**.

  - By considering **all relevant factors**, this type of analysis yields recommendations for next steps.

 - Prescriptive analytics is a valuable tool for data-driven decision-making.

Divergence →

Rehoboam

Source
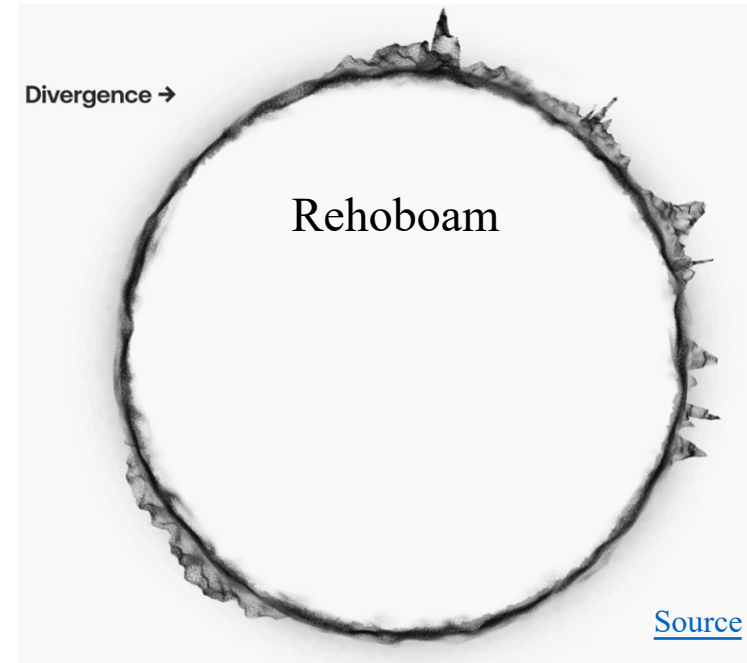
- It's important to note:

  - **While algorithms can provide data-informed recommendations, they can't replace human discernment**.

  - Prescriptive analytics is a tool to inform decisions and strategies and should be treated as such.

  - Your judgment is valuable and necessary to provide context and guard rails to algorithmic outputs.

Prescriptive Analytics: Examples

Content Curation: Algorithmic Recommendations

- If you've ever scrolled through a social media platform or dating app, you've likely experienced prescriptive analytics firsthand through algorithmic content recommendations.
    - Gather data based on your engagement history on their platforms (and potentially others, too).
    - Release a specific recommendation.

- On social media, TikTok's "For You" feed is one example of prescriptive analytics in action.
    - The company's website explains that a user's interactions on the app, much like lead scoring in sales, are weighted based on indication of interest.
    - Videos are then ranked to determine how likely you'll be interested in each video and delivered to each unique 'For You' feed."

- Prescriptive analytics use case can make for higher customer engagement rates, increased customer satisfaction, and the potential to retarget customers with ads based on their behavioral history.

Prescriptive Analytics: Examples
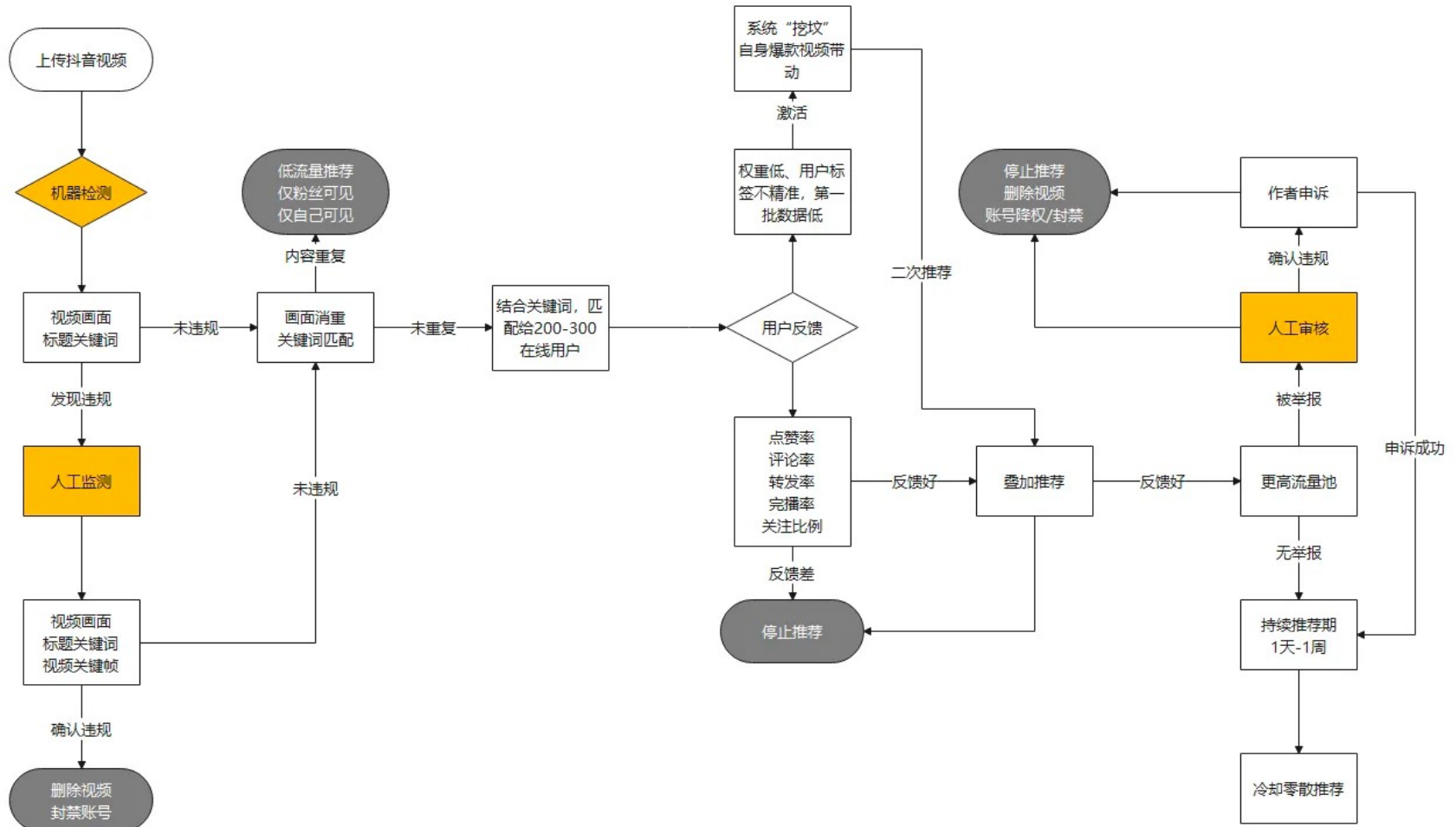
Prescriptive Analytics: Examples

视频信息:
aweme_id: 视频ID
desc: 视频描述
create_time: 创建时间
height: 视频高度
width: 视频宽度
dynamic_cover: 动态封面
ratio: 清晰度
has_watermark: 是否有水印
bit_rate_gear_name: 码率清晰度（视频清晰度可能影响视频评分？）
quality_type: 质量类型
bit_rate: 视频比率
is_h265: 是否为H265视频
duration: 时长
user_digged: 作者自己点赞
allow_share: 是否允许分享
allow_comment: 是否允许评论
with_goods: 包含商品
is_top: 是否置顶
is_vr: 是否VR视频
is_ads: 是否广告视频
risk_infos: 风险信息
position: 位置

is_pgcshow: 是否PGC内容
interaction_stickers: 互动贴纸

音乐信息:
id: 音乐id
title: 音乐标题
author: 音乐作者
album: 音乐专辑
cover: 音乐封面
duration: 音乐时长
has_edited: 音乐是否被编辑过
user_count: 使用人数
collect_stat: 收藏人数
owner_id: 作者id
owner_nickname: 作者昵称
is_original: 是否原创
binded_challenge_id: 绑定挑战赛
strong_beat: 节拍数据（可绘制波形图）
is_commerce_music: 是否商业音乐
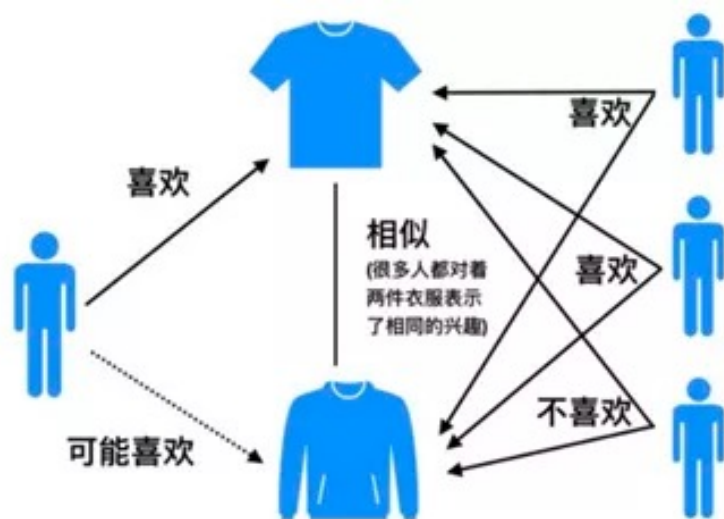is_original_sound: 是否原声
shoot_duration: 视频使用时长

用户信息:
nickname: 创作者昵称
avatar: 头像
signature: 签名
total_favorited: 总获赞
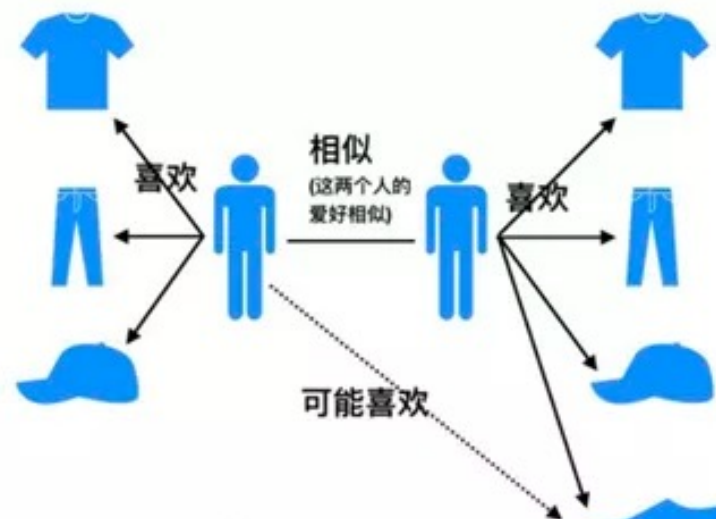fans_count: 粉丝数
following_count: 关注数
dongtai_count: 动态数

用户的行为数据:
who，唯一的用户标识
when，具体时间
where，地理位置
what，交互的内容（包括上面的使用、音乐和创作人）
which，用户的行为（上滑、下滑、左滑、右滑、点赞，关注，点评……）
environment，网络环境、运营商画像、设备品牌型号…

Prescriptive Analytics: Examples

Student Presentation

Thank you~

Wan Fang
Southern University of Science and Technology