Content

- Concepts of Data X
  - Data Science vs. Data Analytics
  - Data Eco-system & Lifecycle
  - Data Privacy & Ethics
  - Data Integrity
  - Data & Analytics Skills

- 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?”

Data Science vs. Data Analytics

- **Data science** is the process of building, cleaning, and structuring datasets to analyze and extract meaning.
- **Data analytics**, on the other hand, refers to the process and practice of analyzing data to answer questions, extract insights, and identify trends.

  - *You can think of data science as a precursor to data analysis. If your dataset isn’t structured, cleaned, and wrangled, how will you be able to draw accurate, insightful conclusions?*

- Every analysis should be a feedback loop that deepens your learning.

  - *What can I learn from the results of this analysis about the underlying context, about competition, about customers, about suppliers?*
  - *How do the results of this analysis validate or reinforce hypotheses I had before I did the analysis?*

Data Science *in Business*

- To collect, organize, and maintain data—often to write algorithms that make large-scale analysis possible.
  - When designed correctly and tested thoroughly, algorithms can catch information or trends that humans miss.
  - They can also significantly speed up the processes of gathering and analyzing data.

Gain customer insights
- Data about your customers can reveal details about their habits, demographics, preferences, and aspirations.
- A foundational understanding of data science can help you make sense of and leverage it to improve user experiences and inform retargeting efforts.

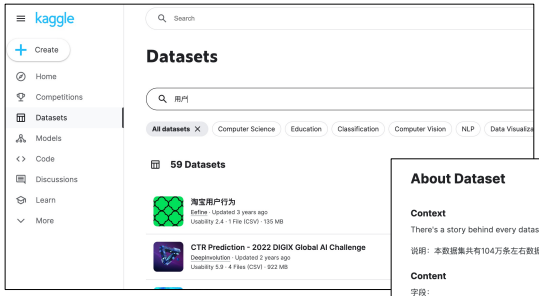
Increase security
- You can also use data science to increase your business’s security and protect sensitive information.
- For example, machine-learning algorithms can detect bank fraud faster and with greater accuracy than humans, simply because of the sheer volume of data generated every day.

Inform internal finances
- Your organization’s financial team can utilize data science to create reports, generate forecasts, and analyze financial trends.
- Data on a company’s cash flows, assets, and debts is constantly gathered, which financial analysts use to manually or algorithmically detect trends in financial growth or decline.

Streamline manufacturing
- Manufacturing machines gather data from production processes at high volumes.
- In cases where the volume of data collected is too high for a human to manually analyze it, an algorithm can be written to clean, sort, and interpret it quickly and accurately to gather insights that drive cost-saving improvements.

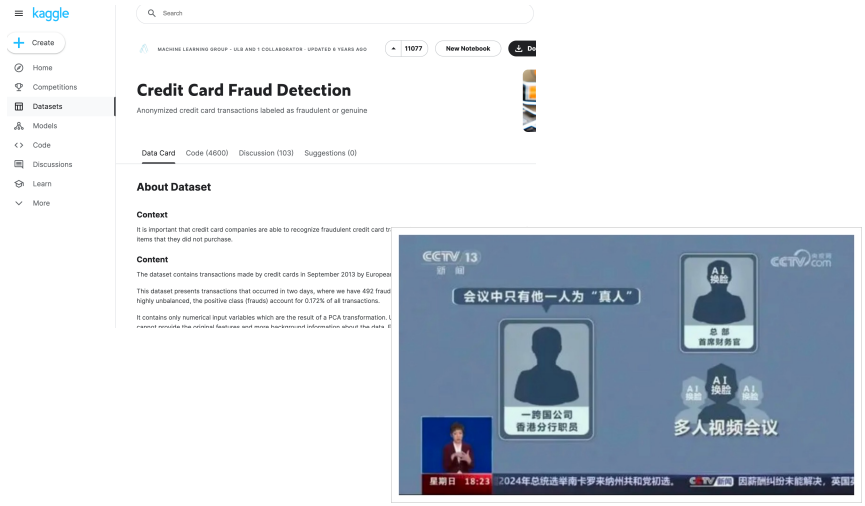
Predict future market trends
- Collecting and analyzing data on a larger scale can enable you to identify emerging trends in your market.
- By staying up to date on the behaviors of your target market, you can make business decisions that allow you to get ahead of the curve.

Gain customer insights



- Data about your customers can reveal details about their habits, demographics, preferences, and aspirations.
- A foundational understanding of data science can help you make sense of and leverage it to improve user experiences and inform retargeting efforts.

Increase security



Streamline manufacturing

SIEMENS NVIDIA

Data Analytics *in Business*

- The main goal of business analytics is *to extract meaningful insights from data that an organization can use to inform its strategy and, ultimately, reach its objectives*.
- Business analytics can be used for:

Budgeting and forecasting

- By assessing a company's historical revenue, sales, and costs data alongside its goals for future growth, an analyst can identify the budget and investments required to make those goals a reality.

Risk management

- By understanding the likelihood of certain business risks occurring—and their associated expenses—an analyst can make cost-effective recommendations to help mitigate them.

Marketing and sales

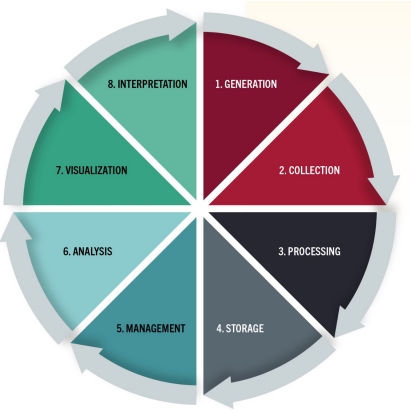
- By understanding key metrics, such as lead- to-customer conversion rate, a marketing analyst can identify the number of leads their efforts must generate to fill the sales pipeline.

Product development (or research and development)

- By understanding how customers reacted to product features in the past, an analyst can help guide product development, design, and user experience in the future.

Data Ecosystem & Lifecycle

- ***Data ecosystem*** refers to the programming languages, packages, algorithms, cloud-computing services, and general infrastructure an organization uses to collect, store, analyze, and leverage data.



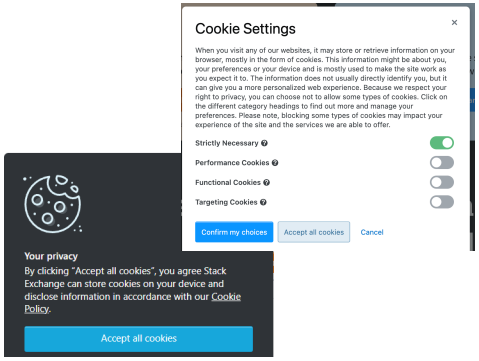
- ***Data life cycle*** describes the path data takes from when it's first generated to when it's interpreted into actionable insights. This life cycle can be split into eight steps: generation, collection, processing, storage, management, analysis, visualization, and interpretation.

- A data project's steps are often described as a cycle because the lessons learned, and insights gleaned from one project typically inform the next. In this way, the final step of the process feeds back into the first, enabling you to start again with new goals and learnings.

Data Privacy & Ethics

- ***Data privacy***, also known as *information privacy*, is a subcategory of data protection that encompasses the ethical and legal obligation to protect access to personally identifiable information (PII), which is any information that can be linked to a specific individual.
  - Some examples of PII include full name, address, ID number, and passport number.

- Data privacy is made up of three key questions:
  - 1. What data is collected?
  - 2. How is the data stored?
  - 3. Who can access the data?



Data Privacy & Ethics

- ***The ethics of data privacy*** can be boiled down to the fact that
  - an **individual's consent** is necessary to collect, store, and use their personal information.

- As a data handler, you have a responsibility to be transparent with your subjects about
  - your intentions,
  - what their data will be used for, and
  - who will have access to it.

- In addition, you need to ensure your use of data doesn't cause harm to an individual or group of people.
  - This is referred to as ***disparate impact*** and is unlawful.

Data Integrity

- Data integrity is the accuracy, completeness, and quality of data as it's maintained over time and across formats.
  - Preserving the integrity of your company's data is a constant process.

- Threats to a dataset's integrity include:
  - **Human error:**
    - For instance, accidentally deleting a row of data in a spreadsheet.
  - **Inconsistencies across formats:**
    - For instance, a dataset in Microsoft Excel that relies on cell referencing may not be accurate in a different format that doesn't allow those cells to be referenced.
  - **Collection error:**
    - For instance, data collected is inaccurate or lacking information, creating an incomplete picture of the subject.
  - **Cybersecurity or internal privacy breaches:**
    - For instance, someone hacks into your company's database with the intent to damage or steal information, or an internal employee damages data with malicious intent.

- To maintain your datasets' integrity,
  - diligently check for errors in the collection, formatting, and analysis phases,
  - monitor for potential breaches, and
  - educate your team about the importance of data integrity.

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.

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 in diagnostic analytics:
  - hypothesis testing
  - difference between correlation and causation
  - 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.
- 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.

Predictive Analytics

- **The use of data to predict future trends and events.**
  - It uses historical data to forecast potential scenarios that can help drive strategic decisions.
- The predictions could be for the **near future** or the **more distant future**, 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 **manually** or **using machine-learning algorithms**.
  - 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

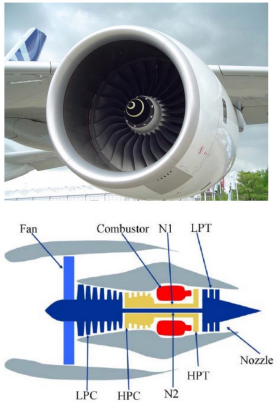
涡轮发动期的预测式维护

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/



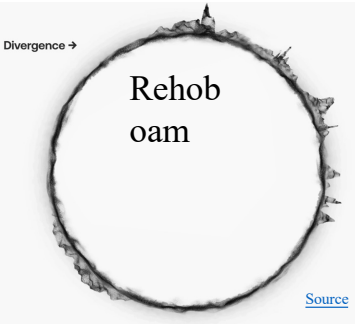
Predictive Analytics: Examples

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.**

Prescriptive Analytics

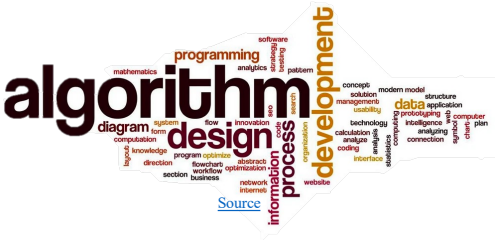
- The process of using data to determine an optimal **course of action**.
  - By considering **all relevant factors**, this type of analysis yields recommendations for next steps.
  - Prescriptive analytics is a valuable tool for data-driven decision-making.



Source

- It's important to note:
  - **While algorithms can provide data-informed recommendations, they can't replace human discernment**.
  - Your judgment is valuable and necessary to provide context and guard rails to algorithmic outputs.

Content Curation: Algorithmic Recommendations



Source

Source

Source