



LIVE ONLINE TRAINING

Business Analytics with Python Bootcamp

Starting Jan 31, 2023
OREILLY.COM





About me



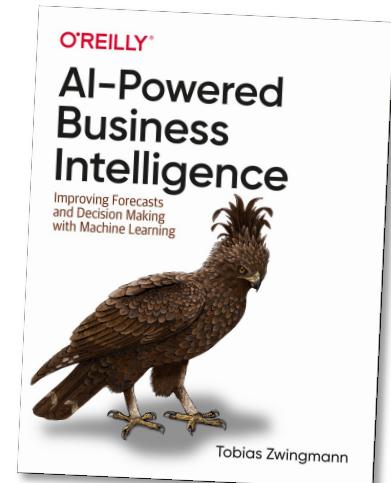
Co-Founder / Managing Partner @ RAPYD.AI

15+ years corporate experience

Data Science Mentor @ Springboard

Author: *AI-Powered Business Intelligence*

Connect: [in/tobias-zwingmann/](https://www.linkedin.com/in/tobias-zwingmann/)





Agenda

- 1. Kick-off**
(10 minutes)
- 2. Business analytics fundamentals**
(100 minutes, incl. break)
- 3. Hands-on business analytics**
(90 minutes, incl. break)
- 4. Wrap-up**
(10 minutes)



Kick-off



Bootcamp overview

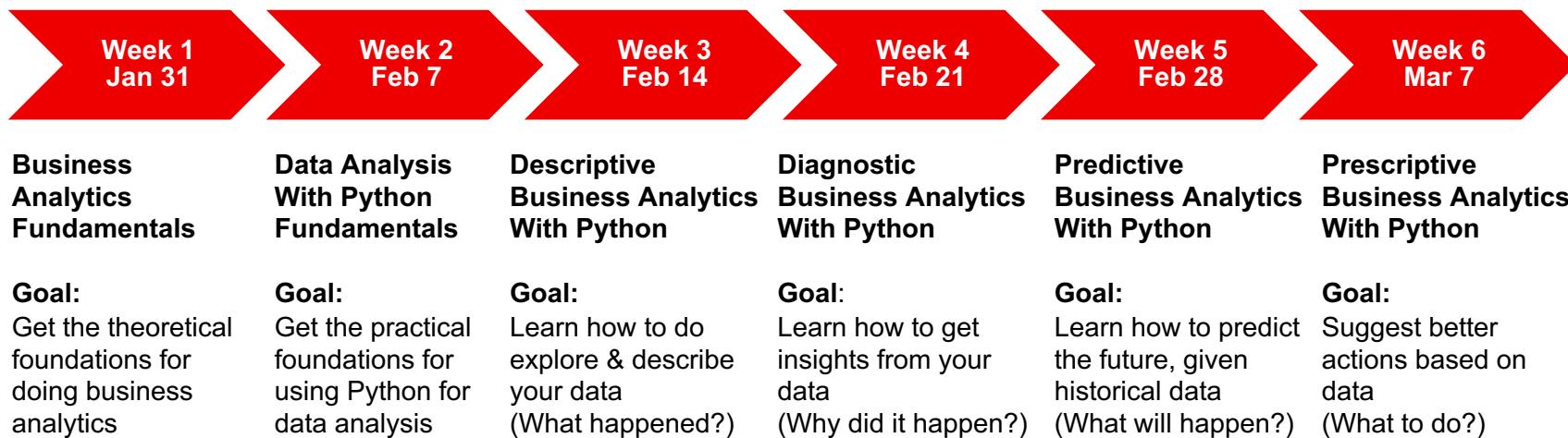
Learning goals:

- Derive actionable insights from data
- Perform exploratory data analysis and create meaningful visualizations
- Use value-based analysis techniques and create association rules for effective decision support
- Apply clustering techniques to discover segments in your data, e.g. different customer groups
- Build predictive models for regression and classification tasks
- Understand the key criteria for evaluating the performance of a predictive model
- Suggest specific business actions that will lead to better results



Bootcamp overview

Structure & Road map



NOTE: With today's registration, you'll be signed up for all six sessions. Although you can attend any of the sessions individually, it's recommended participating in all six weeks.



- 1. Share your background**
- 2. Share your goal**
- 3. Ask Questions**

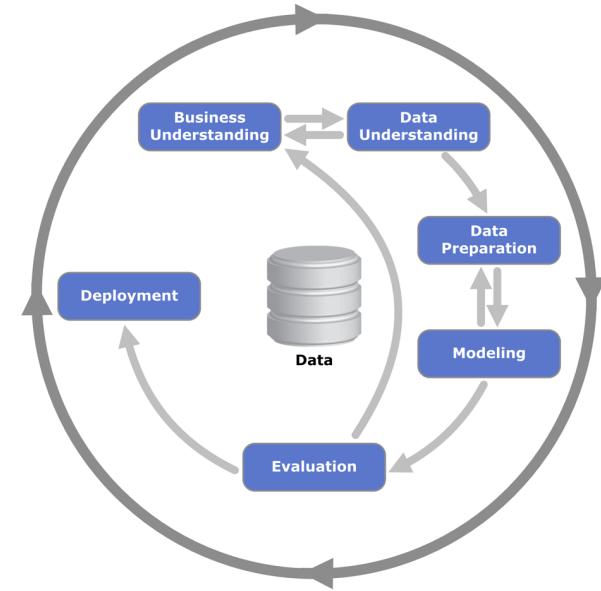


Business Analytics Fundamentals



What is Business Analytics?

- **Goal:** Solve **business** problems using data **analytics**
- This involves data exploration, statistical modeling, but also other qualitative and quantitative methods (e.g. surveys, interviews, etc.).
- Business analytics is an iterative, methodical approach of getting insights from an organization's data, with an emphasis to ultimately make better business **decisions**.
- Data analysis is often focused on certain **business processes** with the goal of making these more effective.



Source: CRISP-DM



Why should we care about Business Analytics?

Context:

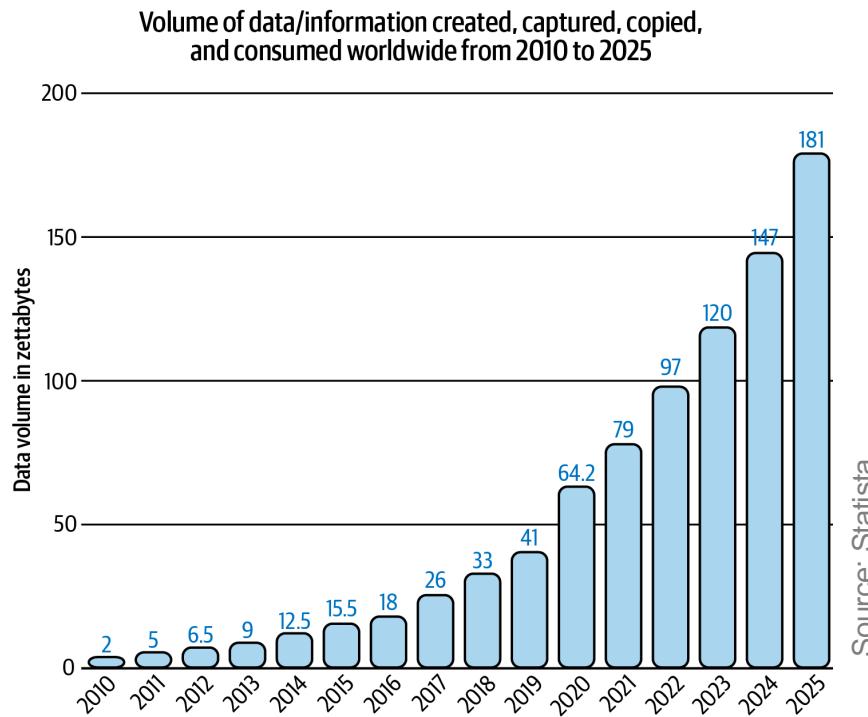
- IDC predicts that 50% of the world's GDP will be based on data-related products, services, and experiences by 2025.
 - Organizations need to leverage digital and data more than ever to stay competitive.
- **Top 5 market cap companies are digital & data centric organizations.**

Driving factor:

- The volume of data is growing rapidly.

Consequence:

- Being able to analyze & leverage data becomes an increasingly critical skill!

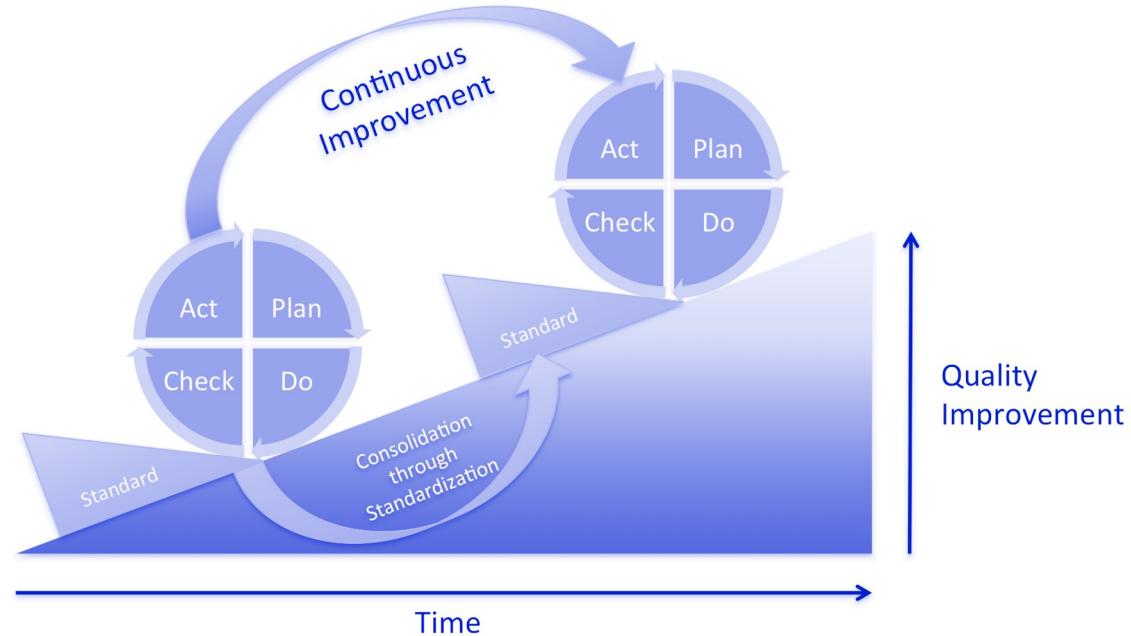




Measuring Business Performance

How can we measure, control and manage business performance?

- **PDCA** (plan–do–check–act) is an iterative design and management method used for the control and continual improvement of processes and products. (Sometimes also OPDCA – "Observe the current condition.")

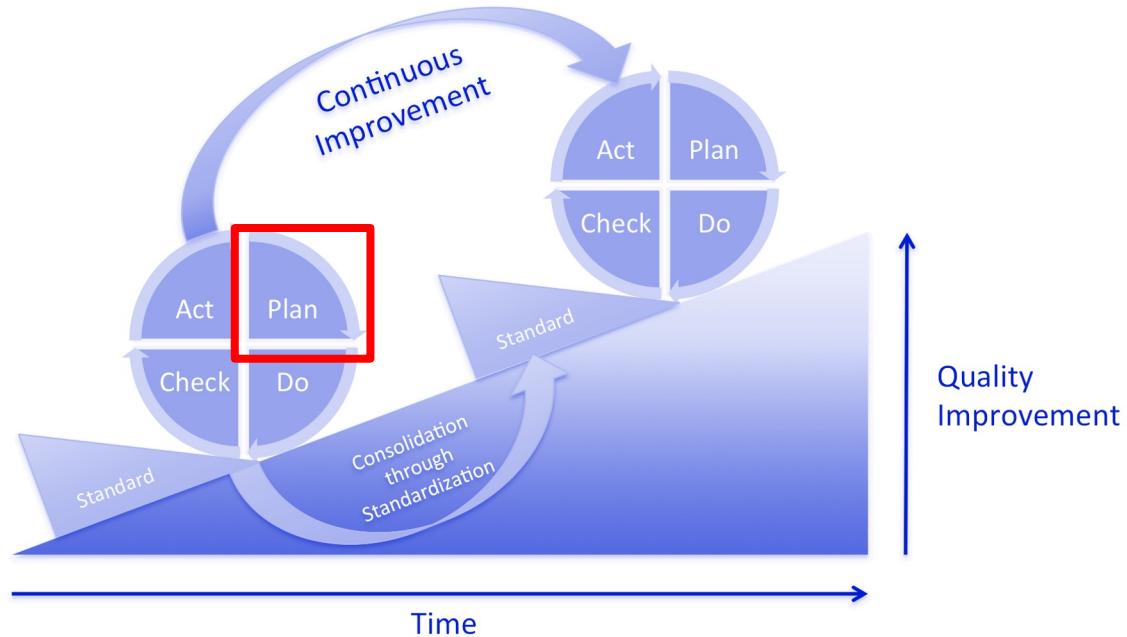




Measuring Business Performance

1. Plan:

- Establish objectives and processes required to deliver the desired results (**goals!**)

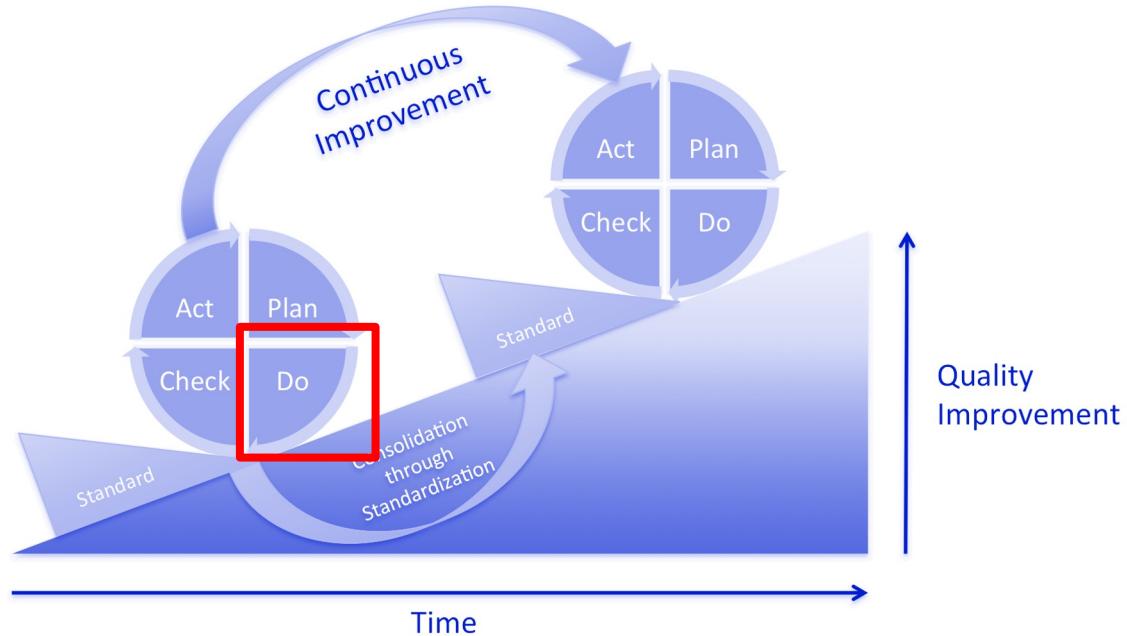




Measuring Business Performance

2. Do:

- Carry out the objectives from the previous step.

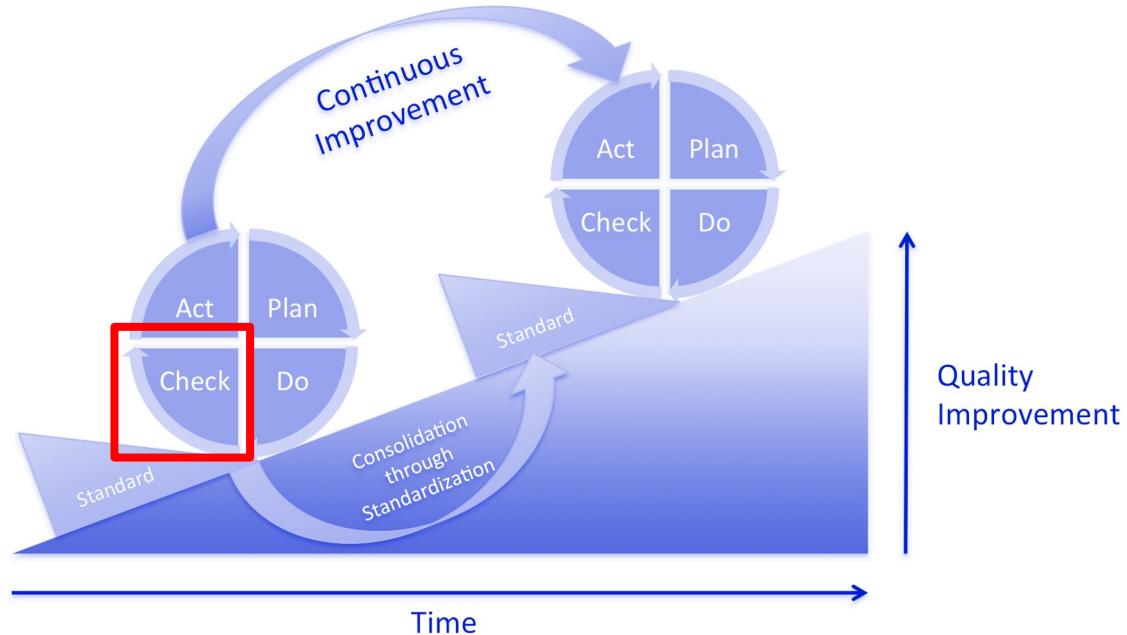




Measuring Business Performance

3. Check:

- **Data** and results gathered from the do phase are analyzed and evaluated.
- Compare data to the expected outcomes
- Suggest areas for improvement

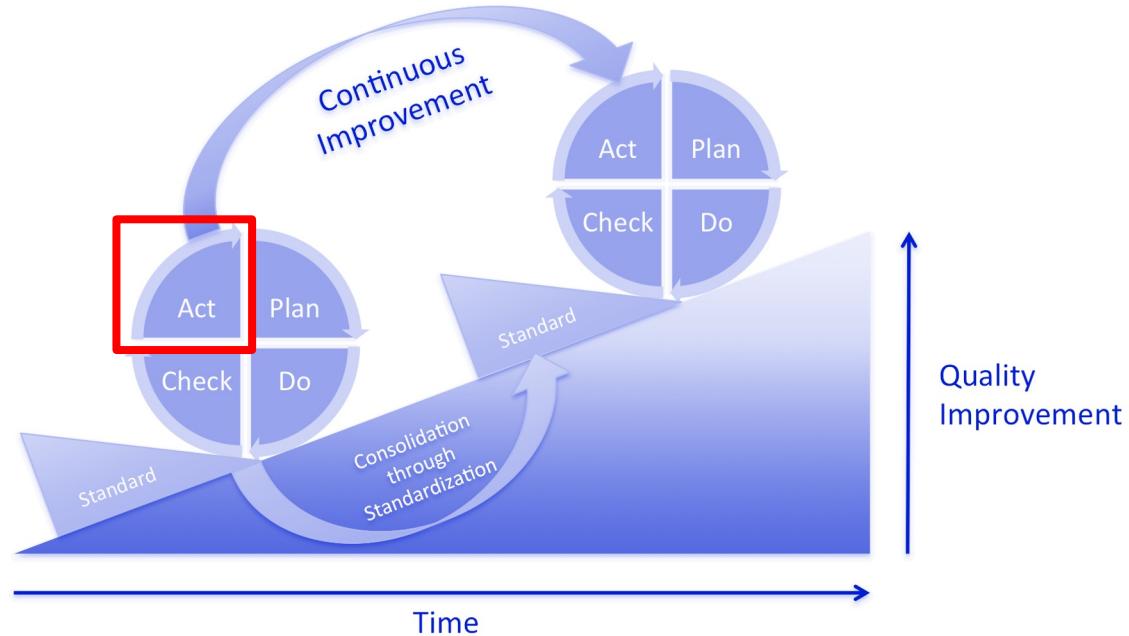




Measuring Business Performance

4. Act (Adjust):

- Improve a process based on insights from "Do" and "Check"
- Identify & solve issues, e.g. problems, opportunities, inefficiencies, etc.
- Start planning again for the next cycle with a better baseline.

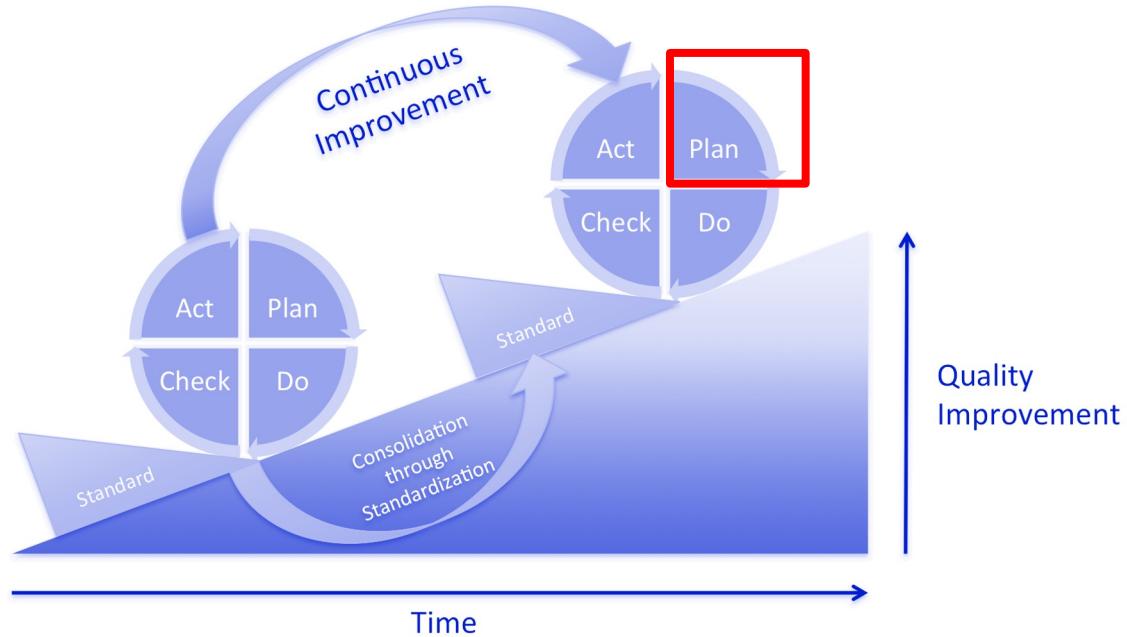




Measuring Business Performance

Repeat: Plan

- Start over again from a higher baseline.
- You should not run into the same issues again in the previous phase, otherwise the action were not effective!





Types of Business Processes

There are three major types of business processes:



Operational Processes

Represent the core business and primary value stream,
e.g., taking orders from customers, manufacturing goods, etc.



Management processes

Oversees operational processes
e.g., controlling, governance, budget, strategy, etc.



Supporting processes

Support operational processes
e.g., accounting, recruitment, tech support, safety training, etc.



Top-Level Business Processes



Operational Processes

Marketing to Lead (M2L)

Opportunity to Order (O2O)

Issue to Complete (I2C)



Management processes

Record to Report (R2R)

Concept to Launch (C2L)

Strategy to Operations (S2O)



Supporting processes

Order to Cash (O2C)

Purchase to Pay (P2P)

Hire to Retire (H2R)



Example Processes

Order to cash (OTC)

- Represents the business process for receiving and processing **customer orders** and revenue recognition.
- Keeps track of payment deadlines, missed payments, dunning, etc.
- **Business lever:** Faster cash collection / better liquidity





Example Processes

Purchase to Pay (P2P)

- Represents the business process for linking **procurement** department to the account payable department.
- Examples: Supply management, accounting, purchase orders, etc.
- **Business lever:** Improve cash-flow and track financial commitments, error and fraud reduction





Example Processes

Record to Report (R2R)

- Collect, process and deliver relevant, timely and accurate **information** used for strategic, financial and operational feedback to understand how a business is performing
- Examples: Company reports, dashboards, key performance indicators, etc.
- **Business lever:** Provide valuable insights (information and knowledge) on how the business is performing and discover areas for improvement.





How these concepts play together

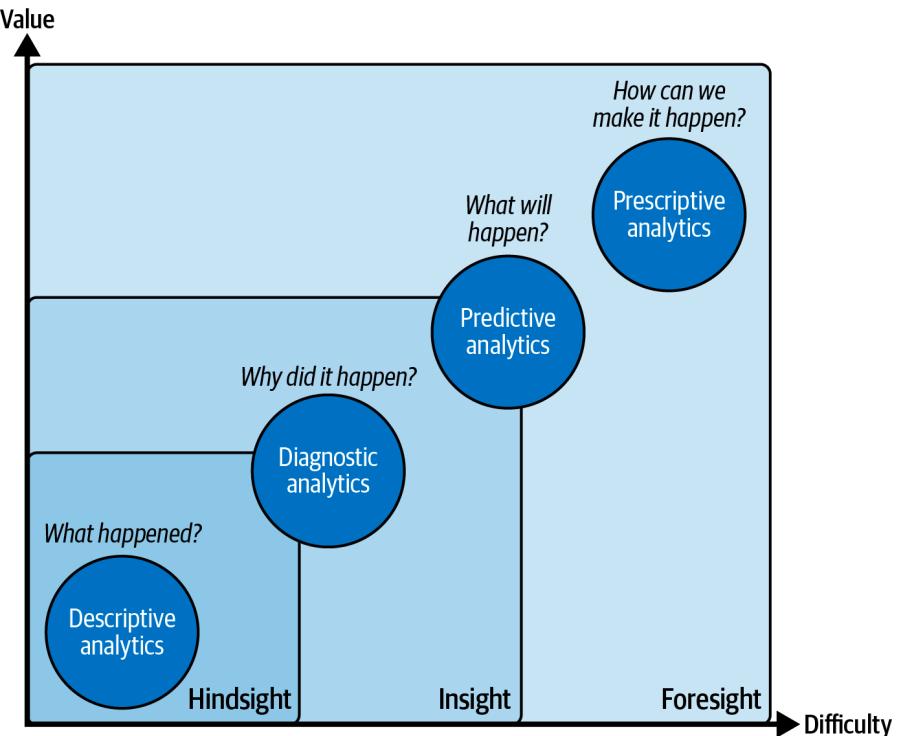
“The Business”





From Data to Better Decisions

- Going from data to better decisions typically involves multiple steps:
 - **Descriptive:** Understand what happened
 - **Diagnostic:** Understand why it happened (and what levers you can pull)
 - **Predictive:** Understand what will happen in the future
 - **Prescriptive:** Suggest actions to achieve a desired outcome
- These steps build in top of each other!





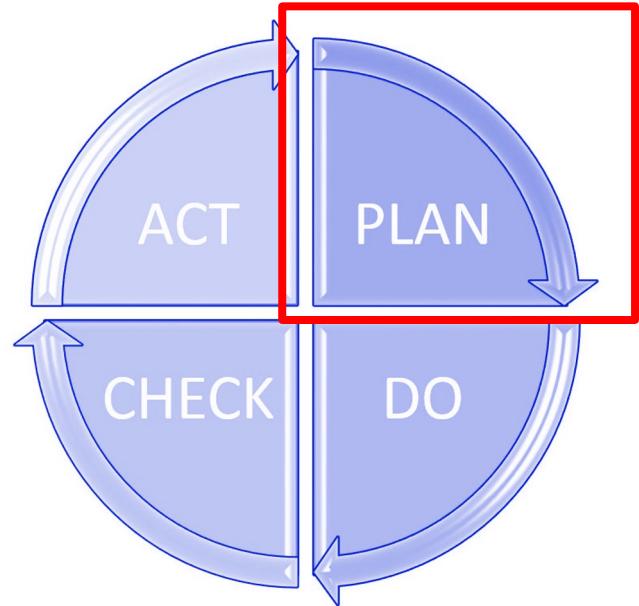
A small example...

- Imagine we're business analysts for an education platform **Edu X** that sells online courses to industry professionals.
- The platform acquires leads through various channels such as online ads, search engines, and referrals.
- When these people fill out a form and provide their email address or phone number, they're classified as leads.
- At some point leads convert into paying customers when they buy a course.



Example: Record to Report

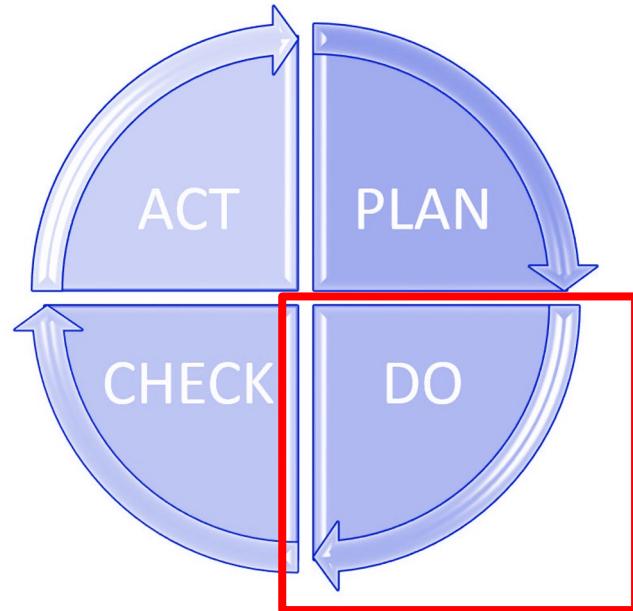
- PDCA – Cycle
- Management **planned** to sell 1,000 courses for the next quarter





Example: Record to Report

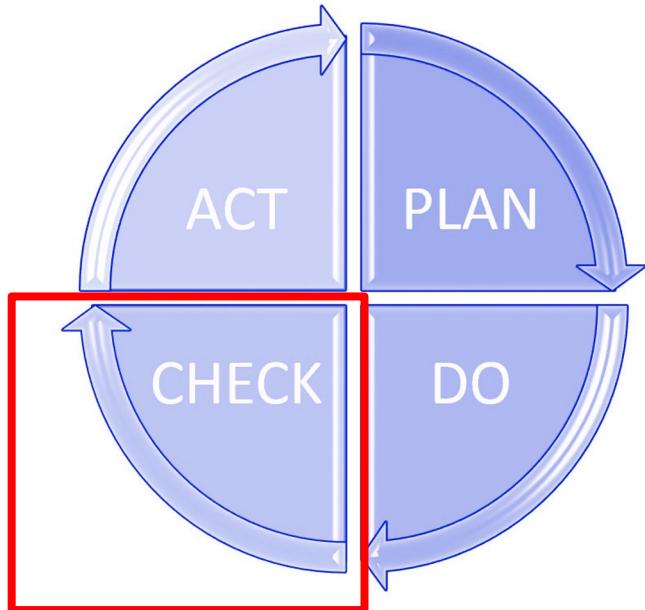
- PDCA – Cycle
- Management **planned** to sell 1,000 courses for the next quarter
- Business continued as usual. In the meantime, we collected data.





Example: Record to Report

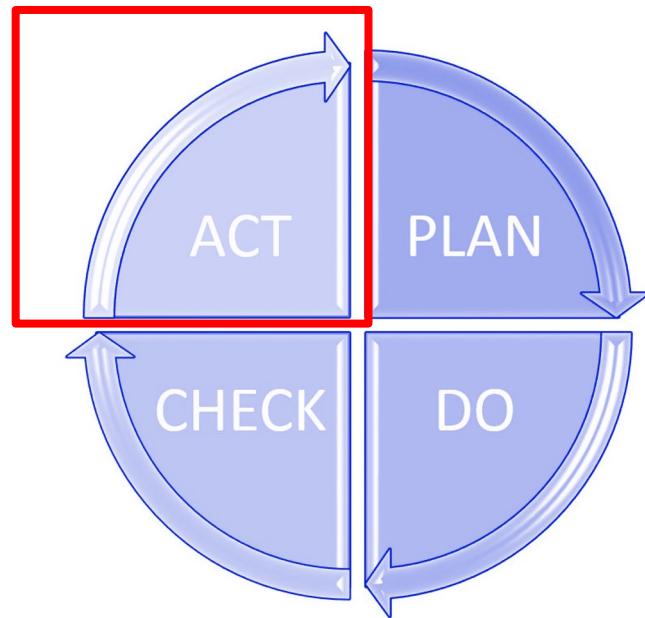
- PDCA – Cycle
- Management planned to sell 1,000 courses for the next quarter
- Business continued as usual. In the meantime, we collected data.
- After the end of the quarter, we **checked** the results:
- Only 800 courses were sold 😞
- We do some **descriptive analytics** to see that last quarter we had 40,000 visitors on our website





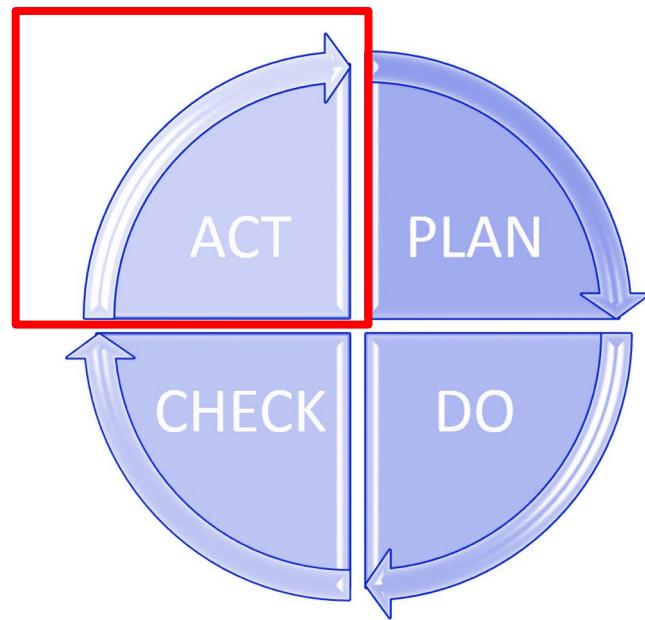
Example: Record to Report

- By now, it's clear we need to change something about our process, because otherwise we will not reach next quarter's target.
- In this phase we might find a couple of things we could **adjust** to convert more website visitors into paying customers.



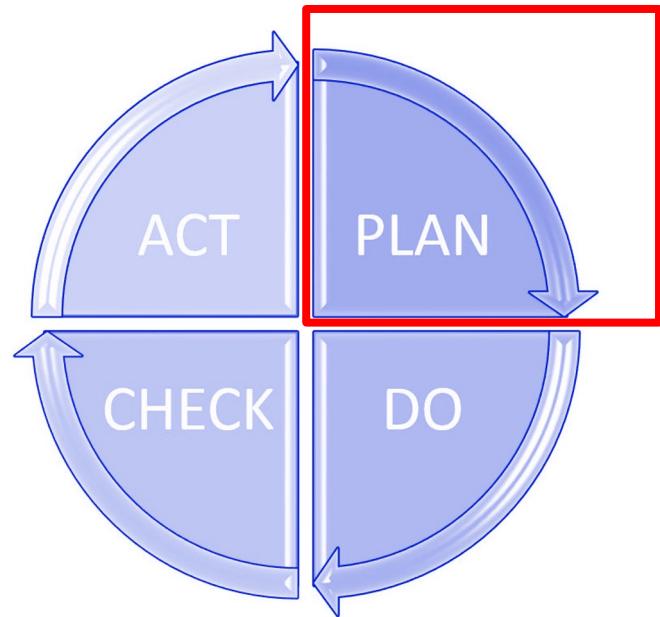
Example: Record to Report

1. We could use **diagnostic analytics** to find out that website visitors acquired via social are much more likely to become a customer.
→ Optimize ad spend
2. With **predictive analytics** we could assign each lead a score based on various factors (e.g. source, time spent on website, etc.)
→ Follow-up on hot leads
3. Using **prescriptive analytics** we might find out which kind of communication channels (phone, SMS, email, etc.) work best for each lead
→ Suggest channels to marketing



Example: Strategy to Operations

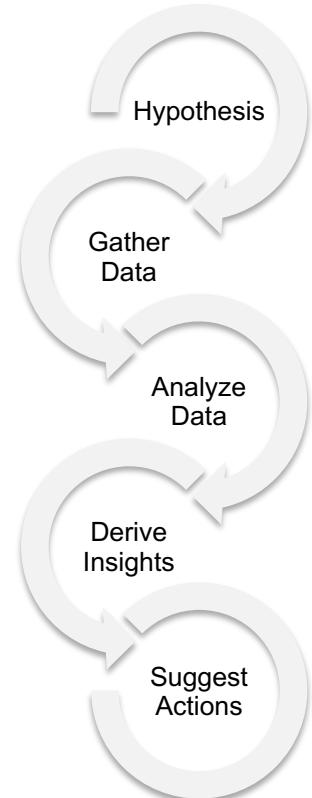
- We adjust the action & budget **plan**.
- Hopefully, we meet our goals next quarter
- To be continued...





Analytical Thinking

- How did we get to the insight that we need to increase our ad spend for the social media ad budget?
- **Bad approach:** Start with a loose problem definition and analyze random data sources you think are interesting
- **Good approach:** Follow a proven methodology / framework!





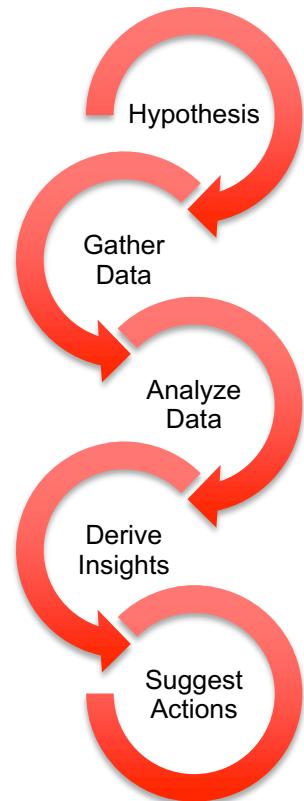
Hypothesis-Driven Analytics Framework

Don't start with data first!

Start with the problem!

→ Hypothesis-Driven Framework

1. Formulate a Hypothesis
2. Gather Data
3. Analyze Data
4. Derive Insights
5. Suggest Actions

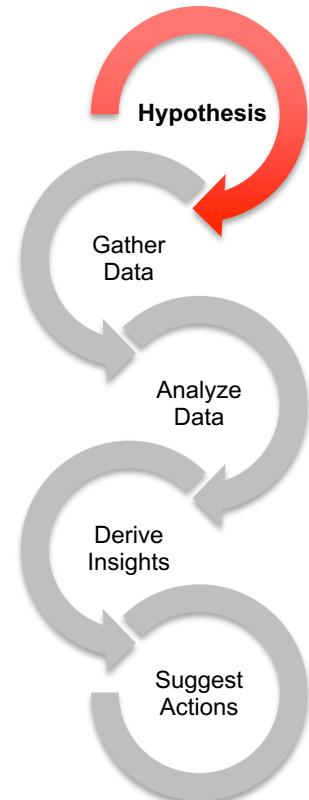




Hypothesis-Driven Analytics Framework

Step 1: Hypothesis

- What is the problem you're facing?
- Understand the unique challenges
- How do you describe the problem?
- Gather expert knowledge
- **Avoid analysis paralysis:** Gather your thinking around a centralized theme (hypothesis)
 - SMART Problem Statement
 - Issue Trees

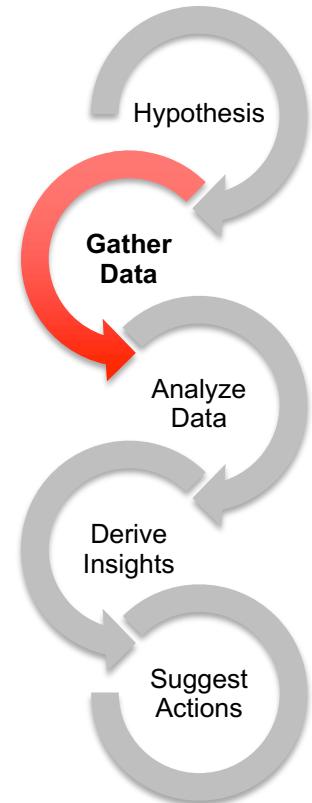




Hypothesis-Driven Analytics Framework

Step 2: Gather data

- Which data sources are needed to answer questions in the hypothesis?
- Get access to data
- Collect data if necessary

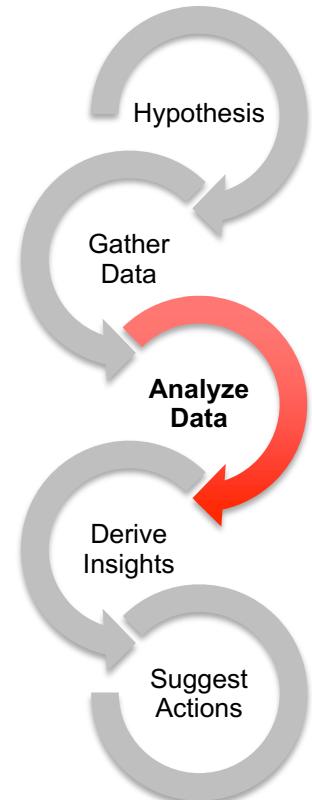




Hypothesis-Driven Analytics Framework

Step 3: Analyze data

- What does your data indicate?
- Which problems exist within your data?
- Do you need to take a step back, i.e. get more or other data?
- Descriptive, Diagnostic, Predictive, Prescriptive Analytics

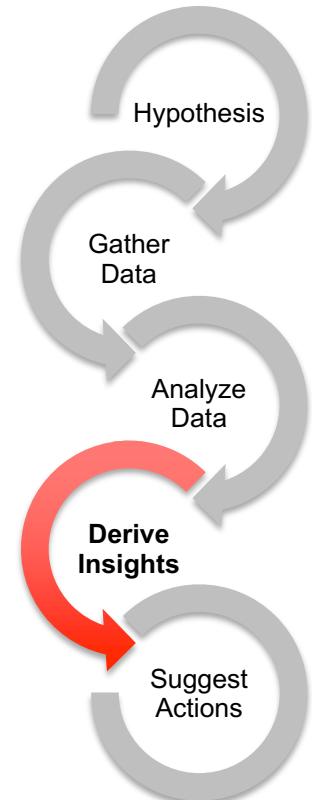




Hypothesis-Driven Analytics Framework

Step 4: Derive Insights

- Which conclusions can you draw from your data analysis with regards to the hypothesis / problem you're looking at?
- What might be the limits / constraints of your conclusions (try to quantify uncertainty)
- Which assumptions did you make?

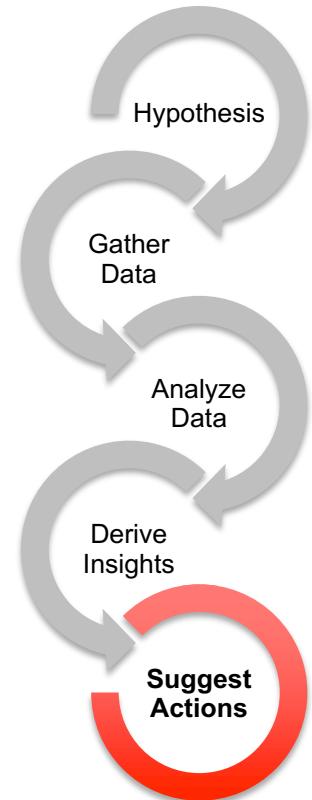




Hypothesis-Driven Analytics Framework

Step 5: Suggest Actions

- Present your findings in a coherent story and suggest actions to specific stakeholders
- Lay out different options
- Explain the associated risks





Problem Statements & Hypothesis

Try to understand and define the problem before you even start thinking about data!

Context: Why are you working on this problem?

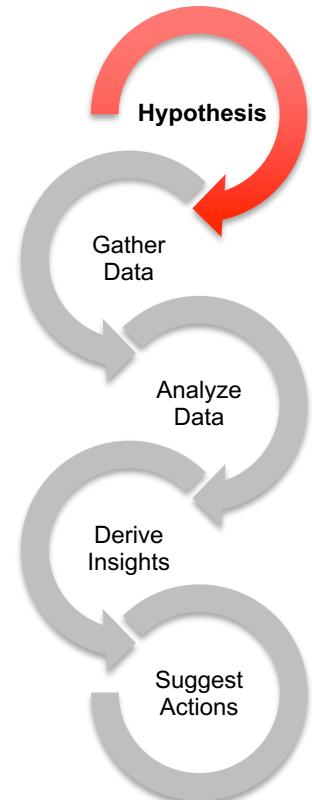
Success criteria: What is the key criteria that will deem this work a success?

Scope of solution space: What's the focus?

Constraints: Which constraints exist within your solution space?

Stakeholders: Who do you need to talk to?

→ SMART Problem Statement





Problem Statements & Hypothesis

Define the problem (SMART)

Specific – Clearly identify the problem

Measurable – Quantify the focus

Actionable – Provide actions to tackle the problem

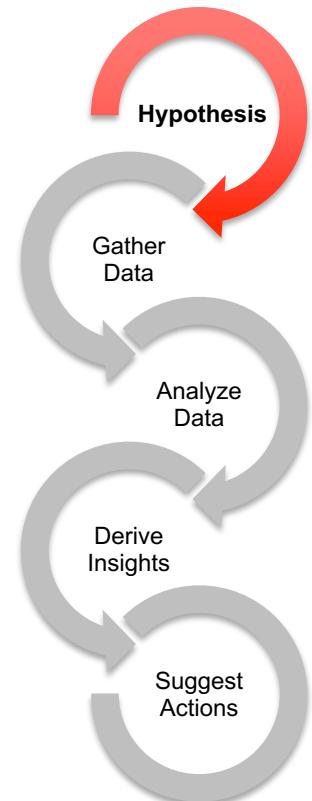
Relevant – Identify actions that will solve the problem

Timebound – Set a firm date by which to solve this problem

Bad example:

Edu X is only converting 30% of all lead on their platform and misses their goal on selling 1,000 courses per quarter.

→ Factual statement, not a SMART problem statement





Problem Statements & Hypothesis

Define the problem (SMART)

Specific – Clearly identify the problem

Measurable – Quantify the focus

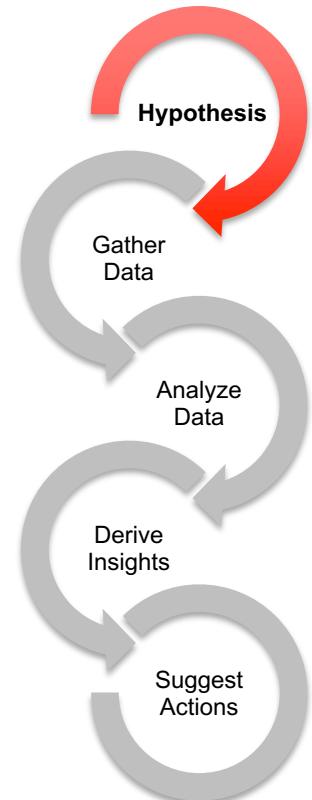
Actionable – Provide actions to tackle the problem

Relevant – Identify actions that will solve the problem

Timebound – Set a firm date by which to solve this problem

Good example:

What opportunities exist for Edu X to increase its lead conversion rate to 45% over the next 3 months through an improved marketing strategy in alignment with the business objective of selling more than 1,000 courses per quarter.





Problem Statements & Hypothesis

Define the problem (SMART)

Specific – Clearly identify the problem

Measurable – Quantify the focus

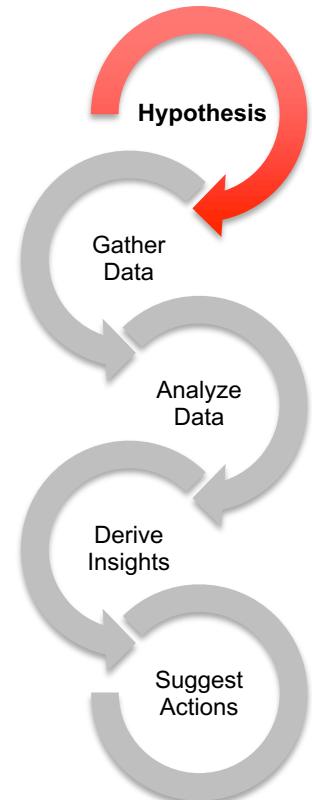
Actionable – Provide actions to tackle the problem

Relevant – Identify actions that will solve the problem

Timebound – Set a firm date by which to solve this problem

Good example:

What opportunities exist for Edu X to increase its lead conversion rate to 45% over the next 3 months through an improved marketing strategy in alignment with the business objective of selling more than 1,000 courses per quarter.





Problem Statements & Hypothesis

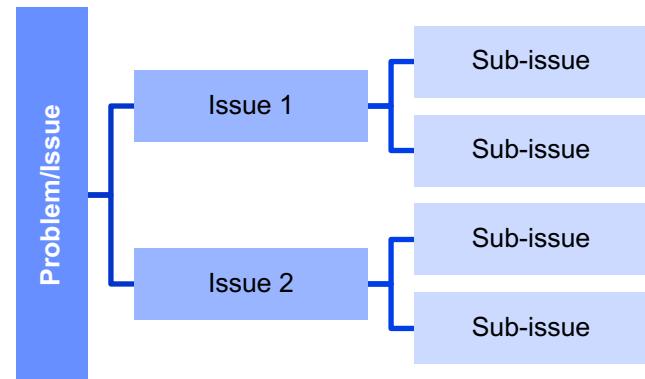
Practical tips:

- **Collaborate:** Team up with other stakeholders on this project
- **Start with your intuition:** Write down your initial thoughts and start iterating over them. It's much easier to iterate from a starting point (no matter how wrong) instead of staring at a blank page.
- **Ask questions:** What's the context and background to this problem? Why is it relevant for the business? How would a good outcome look like?



Issue Trees

- Once you have a problem statement, your next step is to break this problem down into multiple sub-problems (sub-issues) for further exploration
- Issue Trees are a great way to do that!
- **Background:**
 - Issue Trees are a popular problem structuring tool used daily by big consulting companies like McKinsey, BCG, etc.
 - Issue Trees allow you do decompose a problem into different sub-problems without overlaps while still exploring all possible options for the main hypothesis

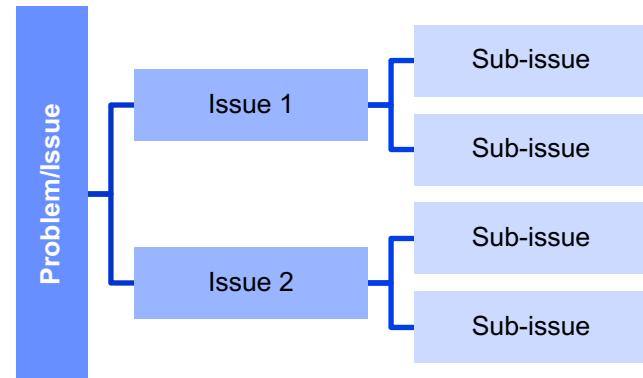




Issue Trees

What are Issue Trees?

- Start with a problem statement
 - Divide problems into sub-problems and components
 - Consists of actions, assertions, criteria, questions, topics as branches
- Sometimes also called hypothesis trees or logic trees

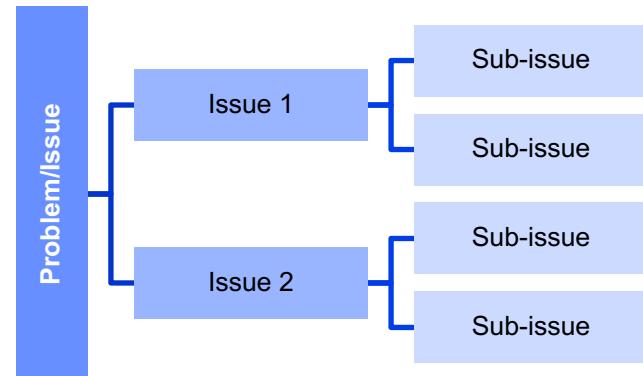




Issue Trees

Benefits:

- 1. Make the problem easier to manage:**
 - Smaller pieces are easier to manage intellectually
 - Prioritize and allocate to different resources
- 2. Maintain integrity of problem solving**
 - Solving the parts will solve the problem
 - Consider all options and test hypothesis
 - No overlaps, no gaps
- 3. Communication**
 - Build a common understanding of the problem and the approach being taken





Issue Trees

"Secret Sauce":

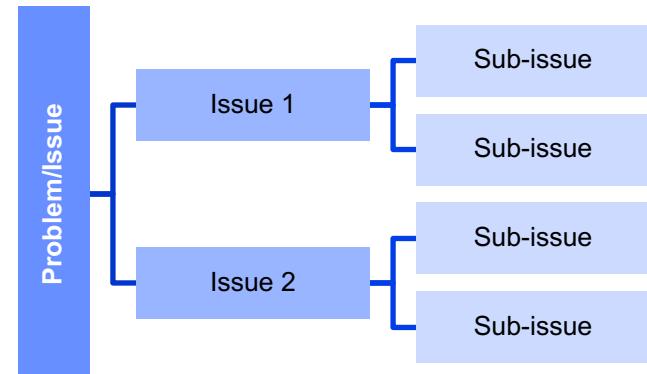
MECE criteria: Issue Trees are mutually exclusive and collectively exhaustive

1. Mutually exclusive (ME)

→ No overlaps between different parts of the tree

2. Collectively exhaustive (CE)

→ You consider all possible options (no gaps)

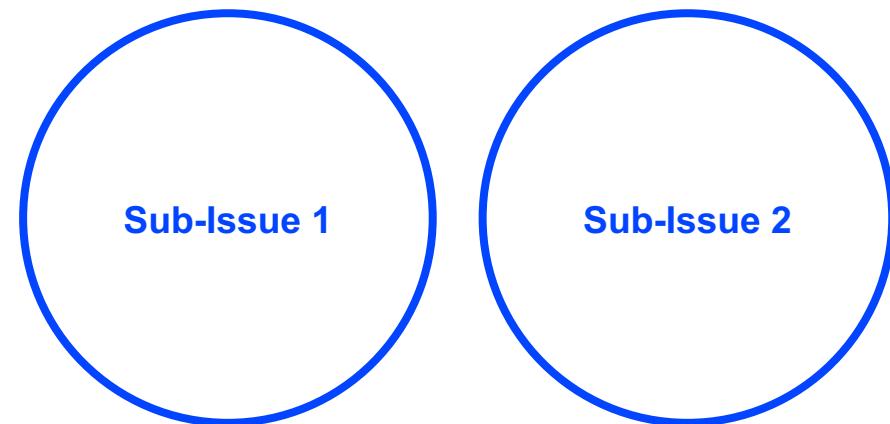




Issue Trees

1. Mutually exclusive (ME)

- Mutually exclusive means that components of an Issue Tree are completely independent of each other.
- Mutually exclusive entities do not include each other.
- (Think of it as a Venn diagram where the circles don't overlap.)

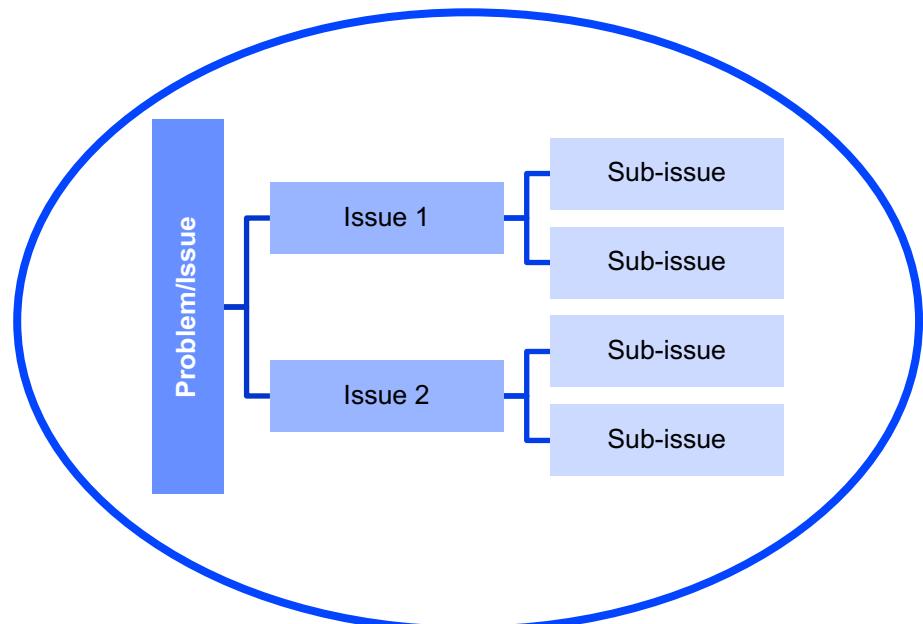


Issue Trees

1. Collectively exhaustive (CE)

- This means the entire set of possible solutions is considered.
- Essentially, you are "exhausting" the set of things to look at.
- If you looked at each component of a collectively exhaustive issue tree, there isn't an area left to look at with respect to the problem.

Entire Solution Space

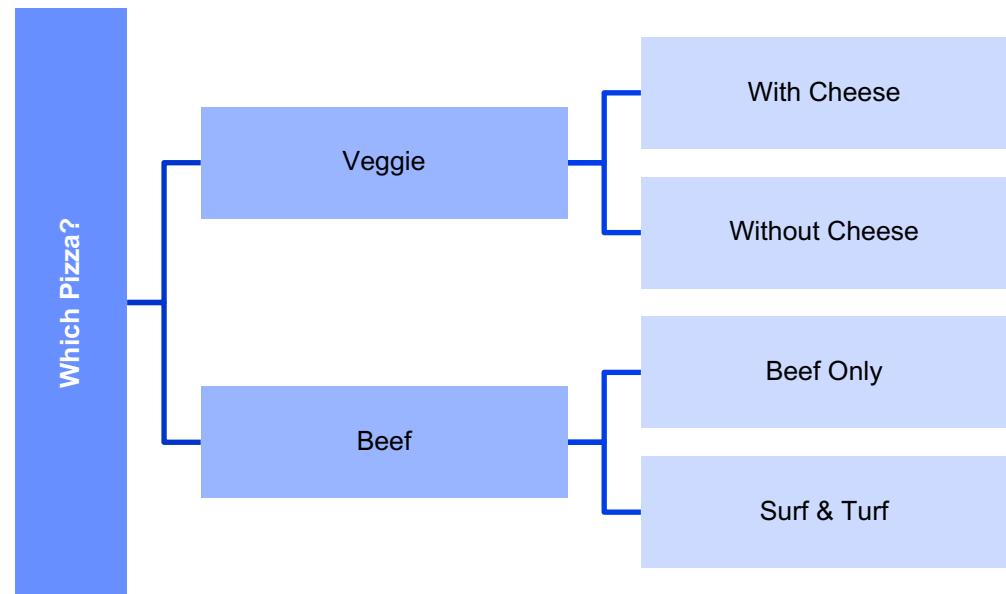


Issue Trees

Let's say we want to decide which Pizza to order.

A **BAD** issue tree would look like this:

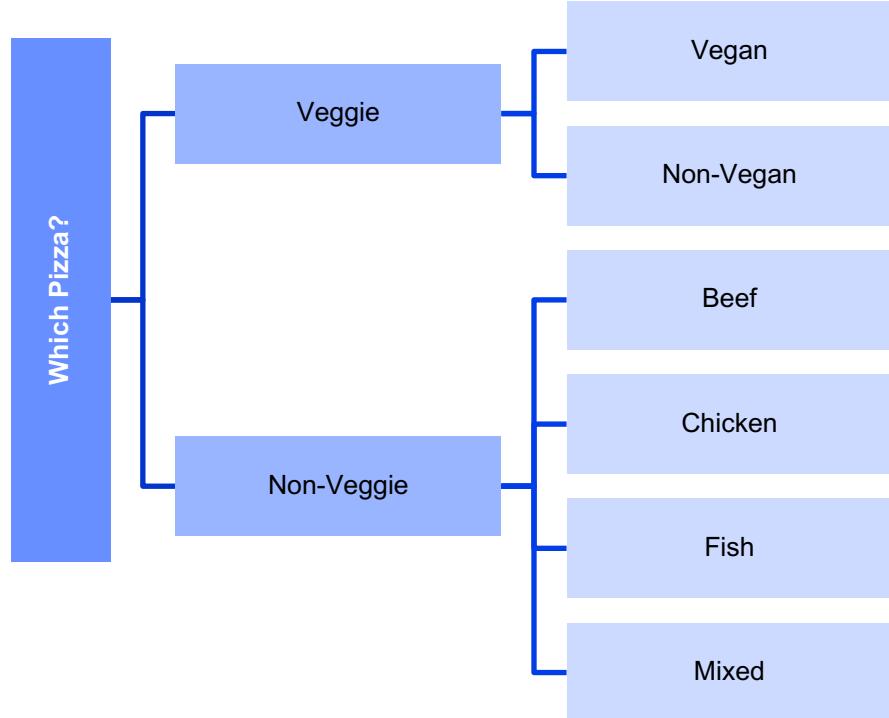
- It's not ME (There could also be beef Pizza without cheese)
- It's not CE (There are more options, e.g. chicken)



Issue Trees

Example of a **GOOD** issue tree:

- It's ME because there's no overlap in the different categories.
- It's CE because there's no other option left.
- This tree helps us to break down the problem fast. (If you're vegetarian, you can immediately drop the bottom part of the tree)
- Note: There might be even more types of meat. Looking at a specific problem is critical!





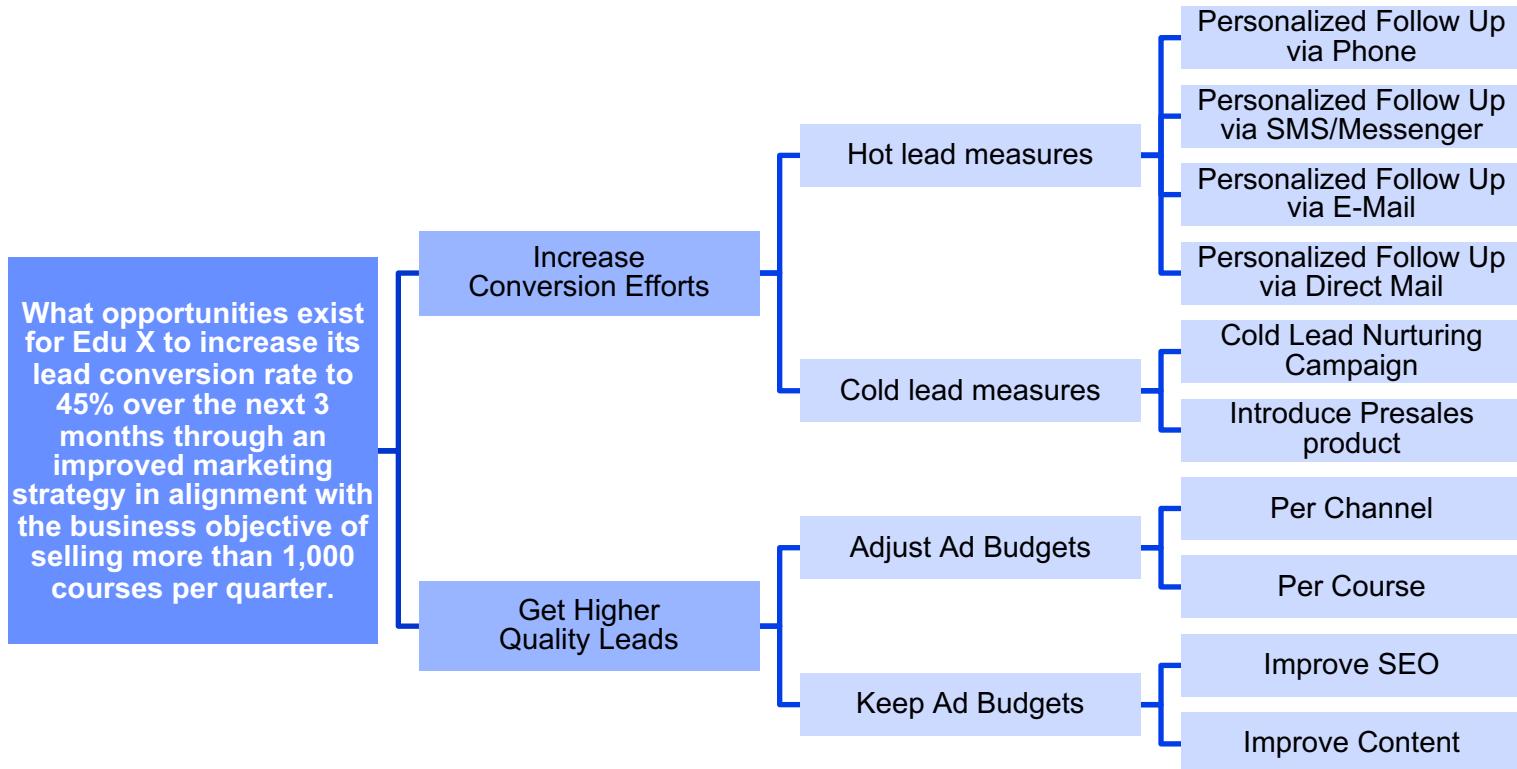
Issue Trees

Real examples where Issue Trees are applied:

- How can company X increase its market share by 5% over the next 6 month?
- Where should we open a new store location to increase revenue by 10% in the next year?
- How can we reach the goal of reducing operational expenses by \$5M in the next fiscal quarter?
- ...



Issue Trees - Example





Exercise: Create an issue tree
→ https://tiny.cc/ba_week1



KPIs and performance metrics

A crash course in **good KPI design**:

- KPIs show progress towards a goal.
- KPIs come with a threshold and a target.
- KPIs compound multiple metrics into one.
- KPIs are created backwards from the business.
- KPIs can be operational or strategic.
- KPIs can be leading or lagging.
- KPIs need definitions.
- KPIs need reviews.
- KPIs need owners.
- KPIs are vital!



KPIs and performance metrics

Strategic vs. operational KPIs

- Strategic KPIs measure the progress towards a long-term goal.
- Strategic KPIs are typically used by senior management for strategic decision making
- Strategic KPIs can typically not be influenced directly, but are a product of various processes
- Examples: Market share, customer satisfaction, revenue growth, etc.

- Operational KPIs measure the performance of a specific process or activity
- Operational KPIs are typically used by managers and employees to monitor the day-to-day work
- Operational KPIs can typically be influenced directly by making changes to the process
- Examples: Production outputs, conversion rates, ad spend, etc.



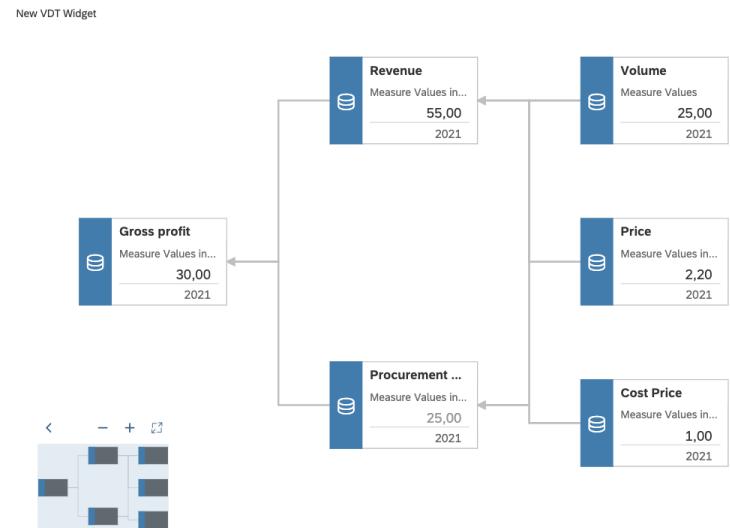
Value Driver Trees

What is a Value Driver Tree? (VDT)

- Visual representation of the factors that contribute to a given metric or KPI
- Hierarchical diagram that shows how different factors interrelate and can be combined to drive

Benefits:

- Break down high level value concept (revenue, cost, performance) into a simple equation that shows all connected drivers in a clear, accountable manner
- Help focus towards „optimizing“ drivers that have the greatest impact the main KPI.

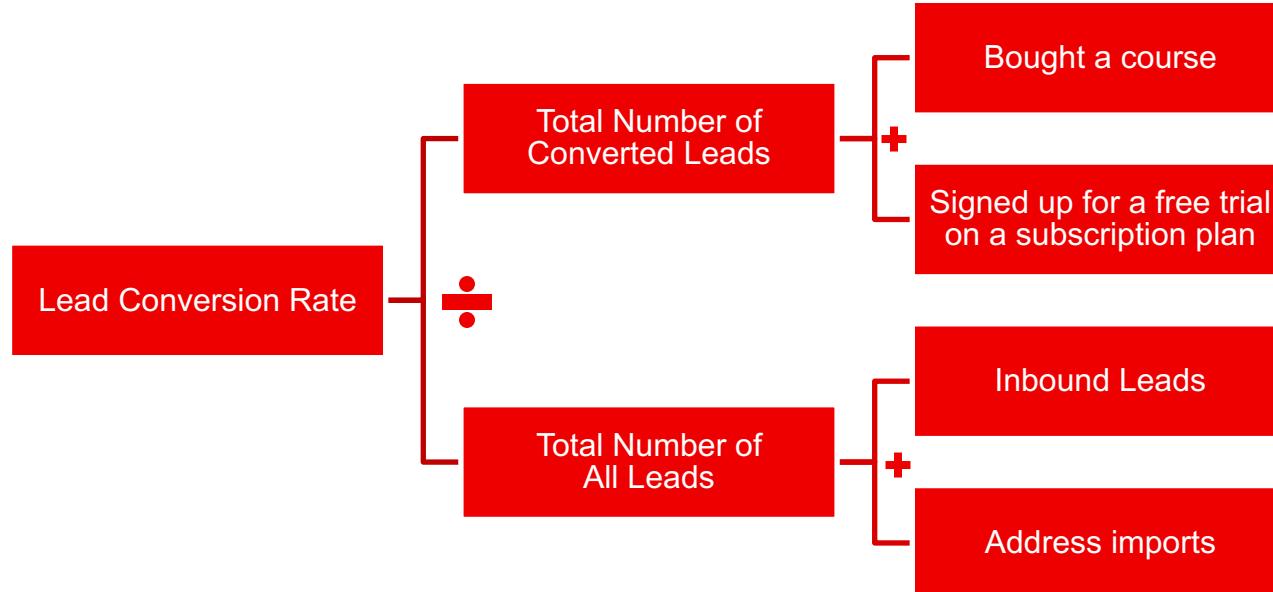


Source: SAP



Value Driver Trees

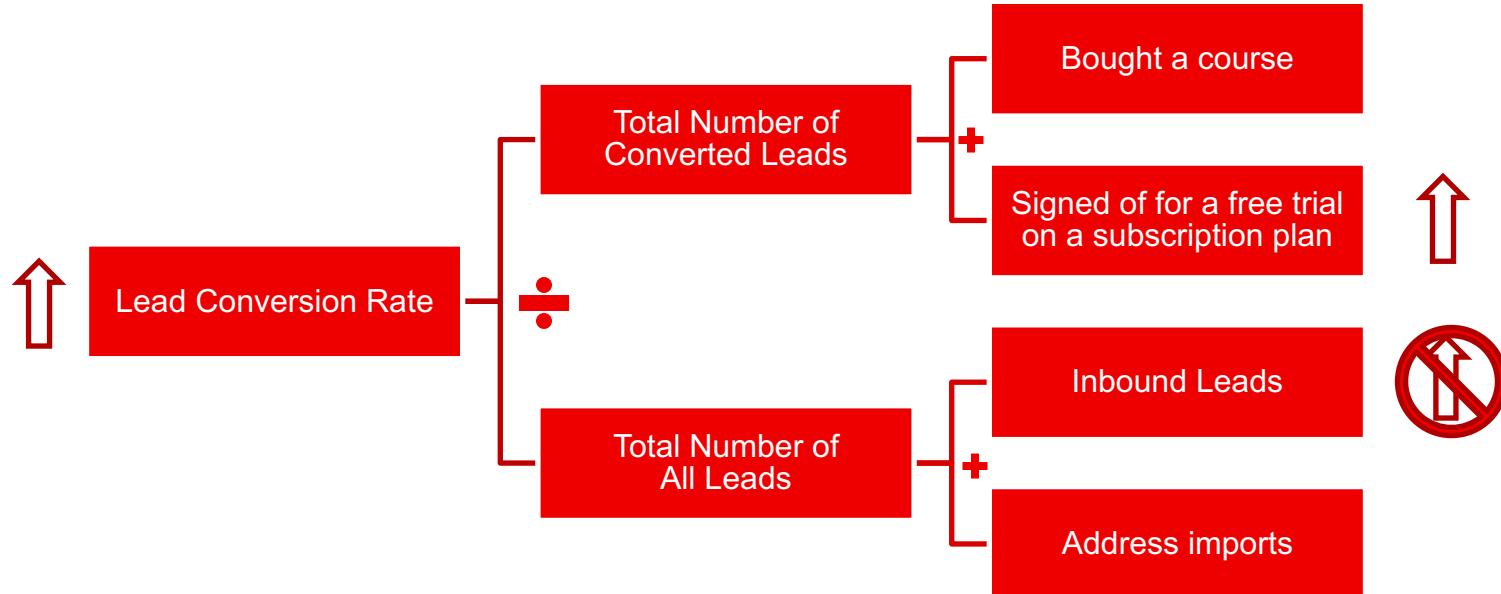
Example 1: Break KPI down into a simple equation





Value Driver Trees

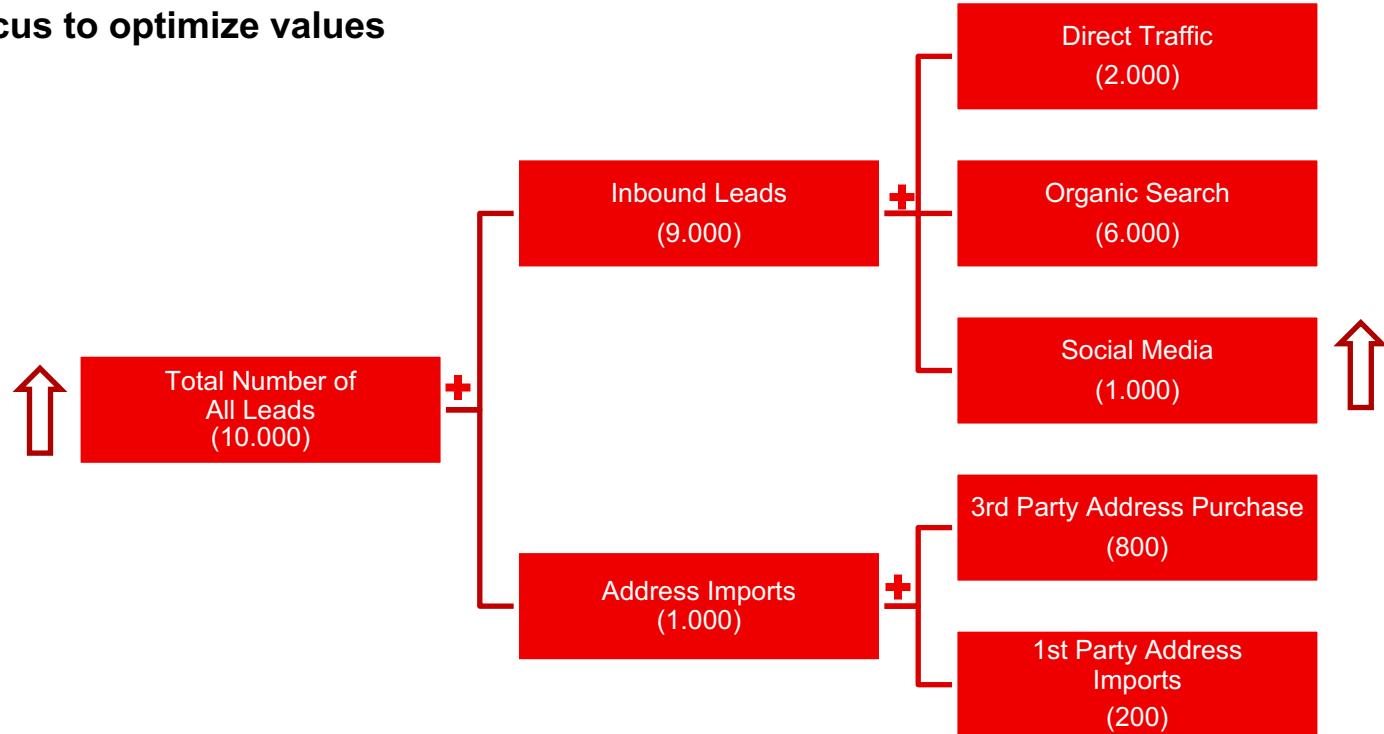
Example 2: Apply logical checks to influence KPIs





Value Driver Trees

Example 3: Focus to optimize values





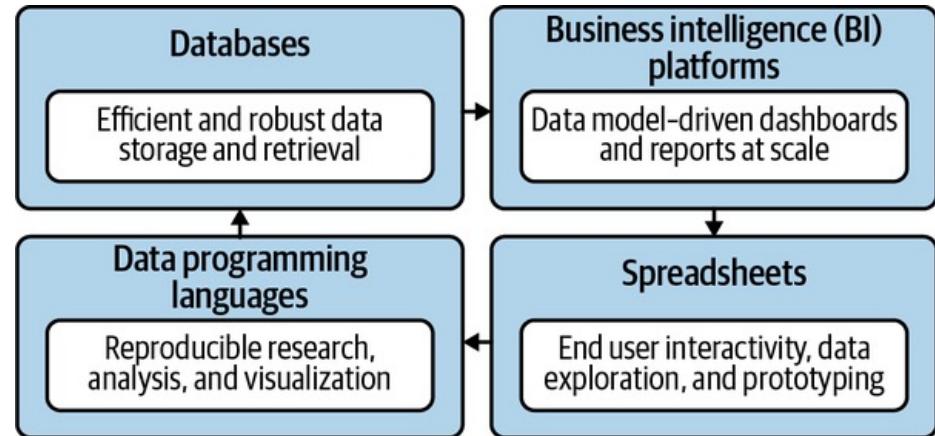
Q&A



Hands-on Business Analytics

Business Analytics tool stack

- Multiple tools are available for doing business analytics
- Each have their own strengths & weaknesses
- The stack you use depends on your setup, your business, but also your knowledge
- Ideally, you choose the right tool – deliberately – for the job at hand



Source: Advancing Into Analytics (G. Mount, O'Reilly 2021)



Business Analytics tool stack

Databases

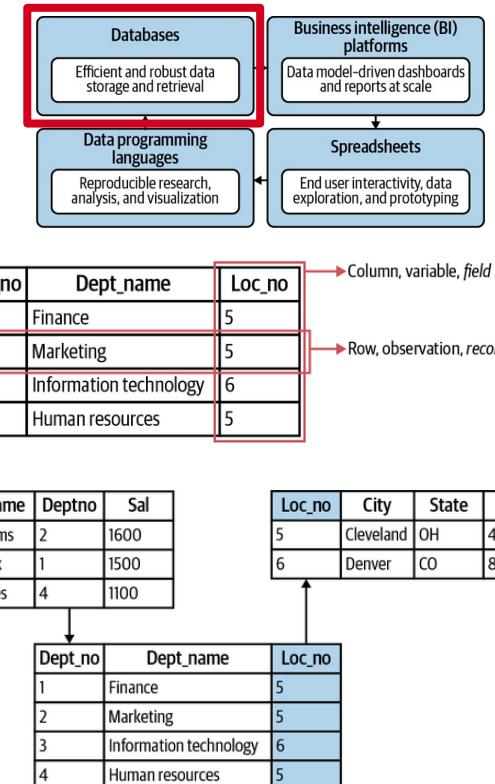
- Most popular for analytics: Relational databases
- Tables are basic building blocks with columns (variables) and rows (observations)
- Tables share relations
- Access data using SQL
- Allow CRUD operations (Create, Read, Update, Delete)

Pro:

- Powerful data analysis using SQL

Con:

- Need to know SQL



Source: Advancing Into Analytics (G. Mount, O'Reilly 2021)



Business Analytics tool stack

Spreadsheets

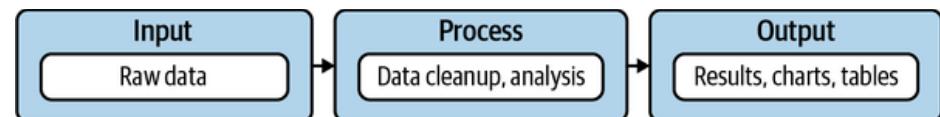
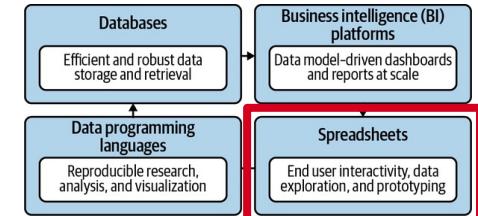
- Examples: Excel, Google Sheets

Pros:

- Quick data exploration
- Quick calculations
- Easy to use & flexible
- Quick distribution

Cons:

- Original data gets altered
- Data processing steps not documented (Lack of reproducibility)
- Don't follow the input– process – output paradigm



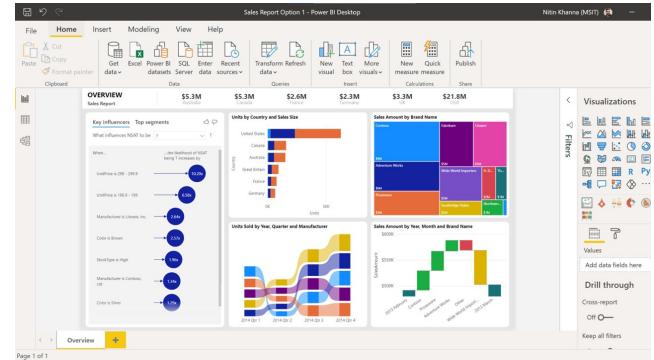
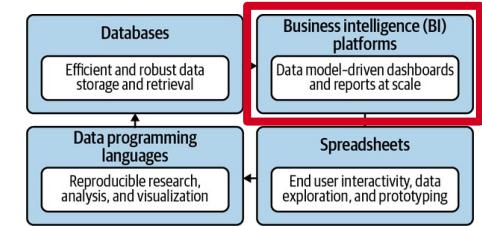
Source: Advancing Into Analytics (G. Mount, O'Reilly 2021)



Business Analytics tool stack

BI platforms

- Examples: Power BI, Tableau, Qlik, etc.
- Enterprise tools that allow users to collect, model, analyze and display data



Source: Advancing Into Analytics (G. Mount, O'Reilly 2021)

Pro

- Reliable, routine reporting at enterprise scale
- (Limited) Interactive data analysis for pre-defined data sets (self-service analytics)
- “Easy to use”

Cons

- Limited flexibility for in-depth data analysis (depending on data model)



Business Analytics tool stack

Data Programming Languages

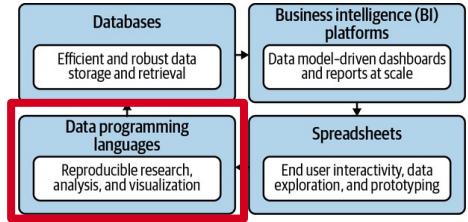
- Examples: Python, R

Pros

- High flexibility
- High capability
- Reproducibility
- Large ecosystem
- Free to use

Cons

- Programming knowledge required
- Hard to grasp for non-technical stakeholders



```
In [1]: import matplotlib.pyplot as plt
import numpy as np

Code cells show both code input and output:
In [2]: 6*0 - 3*2
Out[2]: 42

Markdown cells such as this one can contain text and LaTeX equations such as  $c(a, b) = \sqrt{a^2 + b^2}$ . We can use code to define the corresponding functions:
In [3]: def c(a, b):
    return a**2 + b**2**0.5

Let us compute  $c(a, b)$  as a function of  $a$  and  $b$ , and plot multiple lines each for a fixed  $b$ :
In [4]: a = np.linspace(0, 4, 10)
plt.figure(figsize=(8, 4))
for b in [0, 2, 4, 6]:
    c = c(a, b)
    plt.plot(a, c, linestyle='--', label='b=%s' % str(b))

plt.xlabel('a')
plt.ylabel('c(a, b)')
plt.legend()


```



Tools comparison

→ When to choose which?

Some rules of thumb:

	Databases	Spreadsheets	Programming	BI
Store Data	+++	+/-	---	---
Explore Data	+/-	++	+++	++
Analyze Data	++	+	+++	++
Report Data	---	+/-	++	+++
Share Data	+	+++	---	++



Python for data analysis

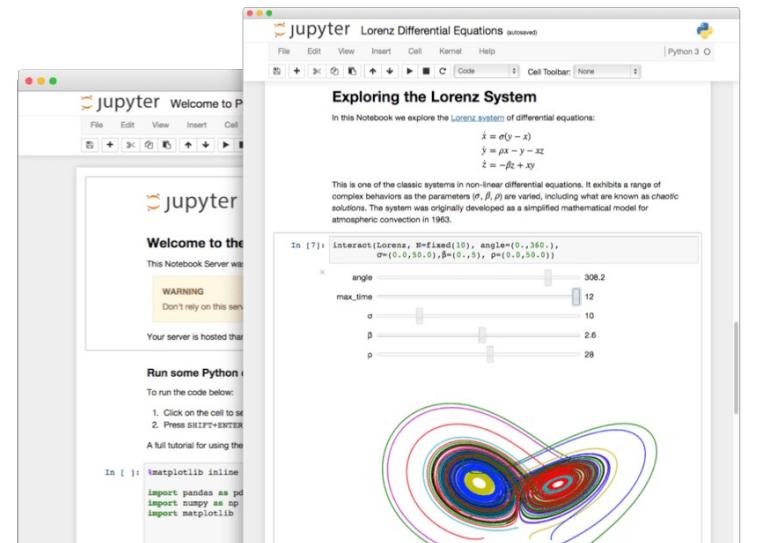
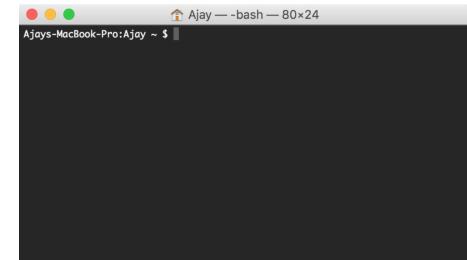
- Most popular language for data analysis
- Very large ecosystem
- R is also fine 😊
- Week 2





Jupyter Notebooks

- Notebooks are a development environment for a programming language (a way to write & run code)
- Combines data annotation and data analysis in one interactive HTML document
- Easier to use than traditional coding IDEs
- Highly interactive
- Can be read by less technical stakeholders as well
- Support multiple programming languages
- “**Swiss Army Knife for Data Analysts**”: Easy to use, easy to read, easy to share, easy to tell a linear story (top-down), all-round use





**Exercise: Explore a dataset with
Excel, then with Python**

→ https://tiny.cc/ba_week1



Q&A

How do you feel?



Wrap-up



What did we learn today?

- ❑ What is Business Analytics and why should we care?
- ❑ PDCA management cycle
- ❑ Different types of business processes
- ❑ Different types of business analytics (descriptive, diagnostic, predictive, prescriptive)
- ❑ Hypothesis-Driven Analytics Framework
- ❑ Smart Problem Statements
- ❑ Issue Trees
- ❑ KPIs 101
- ❑ Value Driver Trees
- ❑ Business Analytics tool stack and comparison
- ❑ Python & Jupyter Notebooks
- ❑ Compared a data analysis process : Excel vs. Python



Outlook for next week

Week 2: Data Analysis with Python Fundamentals

- Getting started with Python – what is it, why is it so popular?
- The Python Ecosystem
- Popular packages for data analysis
- Notebooks vs. script files
- Install Anaconda & Jupyter Notebook
- Essential coding best practices (versioning, clean code, functions, ...)
- Build your first code repository
- Organizing your code files
- Running scripts from the command line



Thank you!

Keep in touch:

linkedin.com/in/tobias-zwingmann
tobias@rapyd.ai

