

InsureTech: Analyze Data, Innovate, Win!

BEST Hackathon

May 2025

A decorative graphic in the bottom right corner consisting of a yellow semi-circle and a white triangle pointing upwards.



Agenda

- Technical Introduction
 - Data
 - EDA & ETL
 - Artificial Intelligence
 - Data Visualization & Data Storytelling
- Business Introduction
 - Claims
- Case Study

Technical Introduction

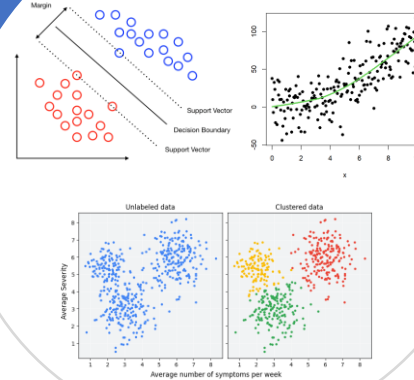
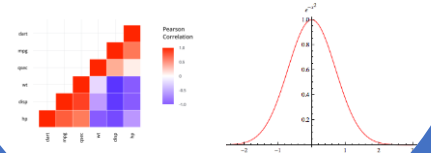
ETL & EDA



Data



x_{11}	x_{12}	...	x_{1d}	$y_1 \leftarrow \text{label}$
...
x_{i1}	x_{i2}	...	x_{id}	$y_i \leftarrow \text{label}$
...
x_{n1}	x_{n2}	...	x_{nd}	$y_n \leftarrow \text{label}$



AI

Data Visualization & Data Storytelling





Data, ETL & EDA



Data

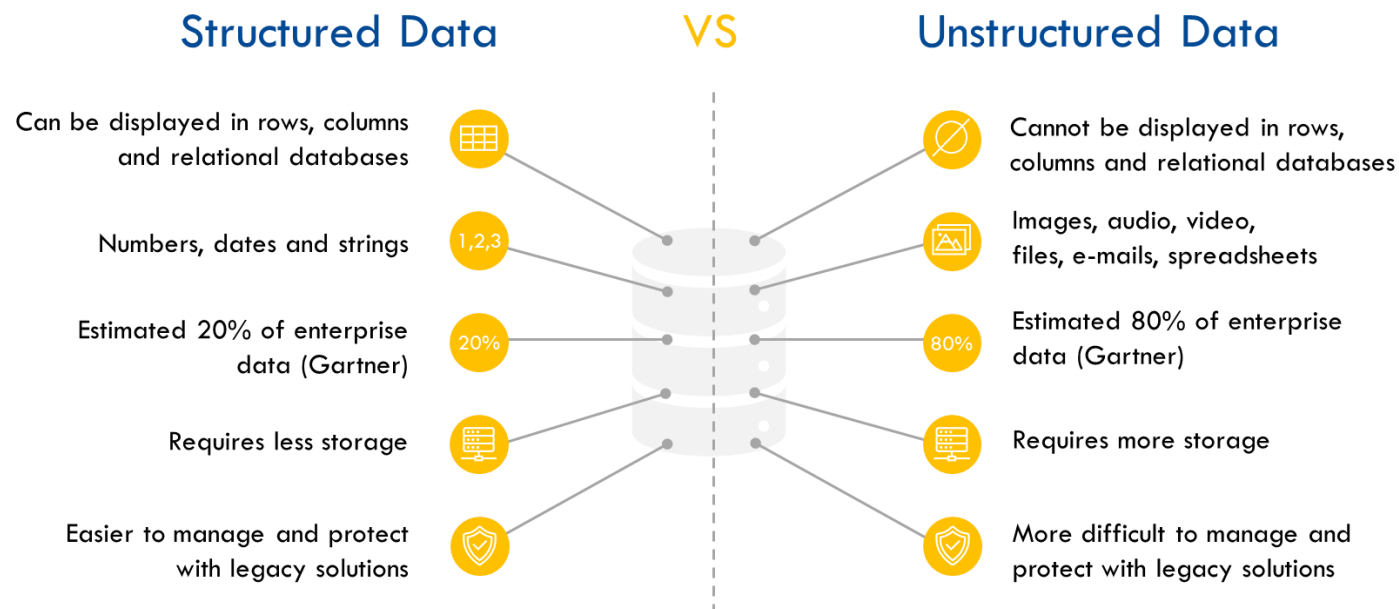
What is it?

Data refers to facts, figures, or any kind of information gathered through observations, measurements, or research, and then stored.

Why is important?

- 1 Informed decision making
- 2 Improvement in efficiency and productivity
- 3 Personalization and user experience
- 4 Innovation and development
- 5 Predictive analysis and forecasting

Data Types





What do we do with data now?



EDA

An **exploratory** and visual process aimed at **understanding** and preparing data for deeper analysis.

Typical preliminary steps before:

- Accurate statistical analysis
- Application of ML algorithms
- Building data models



ETL

A data **transformation** process focused on **integrating** and **preparing** data for large-scale loading and accessibility.



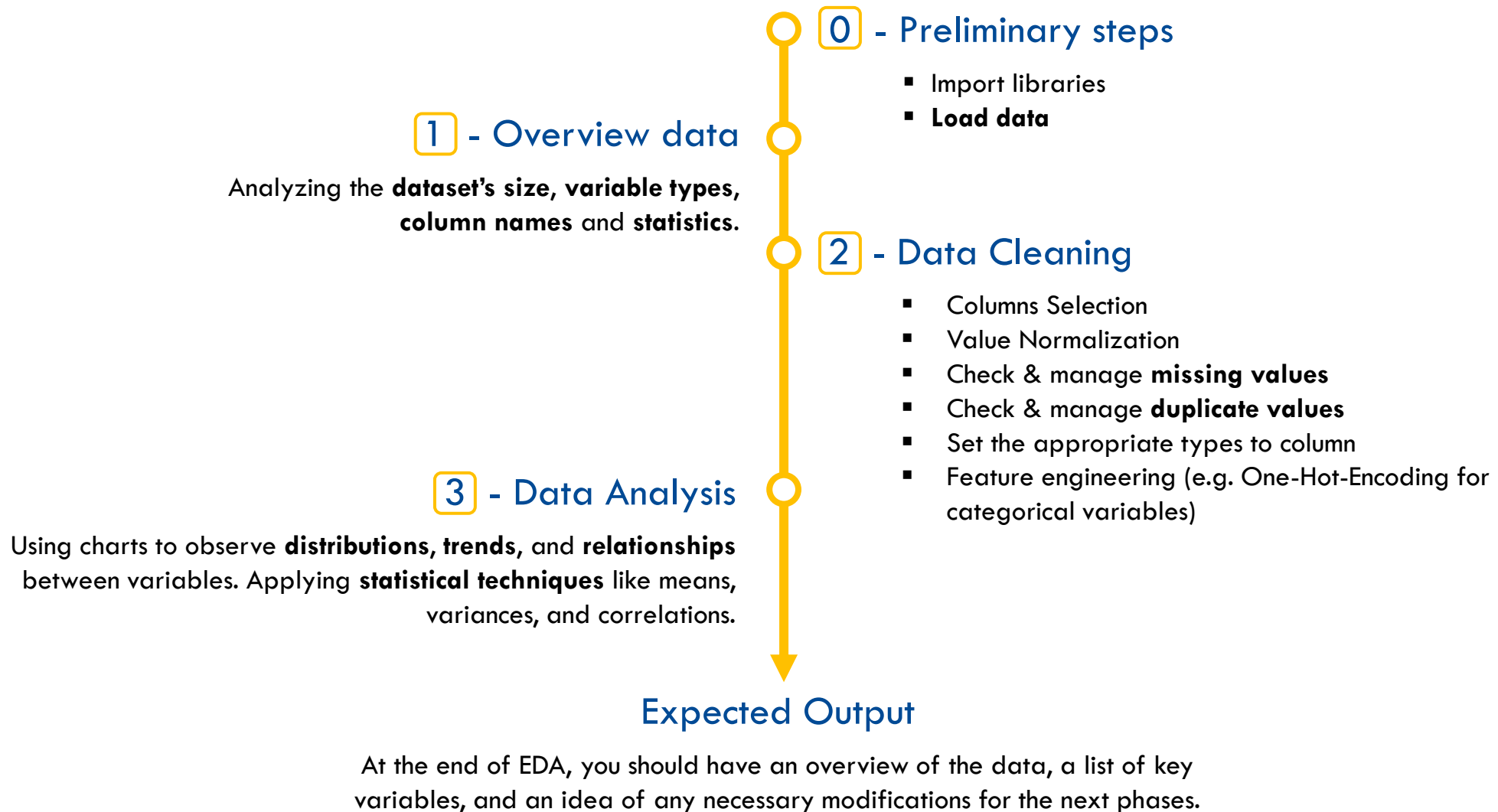
EDA

What is it?

EDA is the initial approach to analyzing a dataset, identifying patterns, outliers, and relationships among variables using visual and statistical methods.

Why is important?

- 1 Identify errors and inconsistencies
- 2 Visualize distributions, correlations, and outliers
- 3 Develop a preliminary intuition about the data
- 4 Prevent errors in subsequent modeling



ETL

What is it?

ETL, short for Extract, Transform, Load, is a **fundamental data engineering process** where data is gathered from a source, modified to suit specific needs, stored in a target location. Supports **data-driven decisions**, creating **quality datasets** for machine learning, and **ensuring structured storage** for compliance.



Why is it important?

- to process data with real-time ingestion and for immediate decision-making
- adaptability across diverse applications - from batch to streaming data
- to integrate within different architectures, like novel concepts like the Data Lakehouse



Extract

Transfer data A → B

- **Frequency**
batch, streaming, message services
- **Volume**
amount of data to be processed, for scalability, cost, latency and retention
- **Shape & Quality**
how to deal with the variety of data correctly as not to compromise data integrity
- **Format**
structured, semi-structured, non-structured

Transform

Turn raw data into valuable insights through various stages of the data lifecycle.

- **Enrichment** (additional sources)
- **Joining** (combine one or more datasets)
- **Filtering**
- **Structuring** (change format)
- **Conversion** (change datatype)
- **Aggregation**
- **Anonymization**
- **Splitting**
- **Deduplication** (remove duplicates)

Load

Write transformed data into a destination for easy access by users and applications.

- **Overwrite**
- **Insert**
Append
- **Upsert**
Update + insert = merge
- **Delete**
Soft or hard



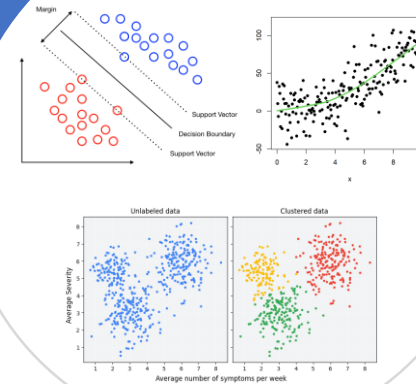
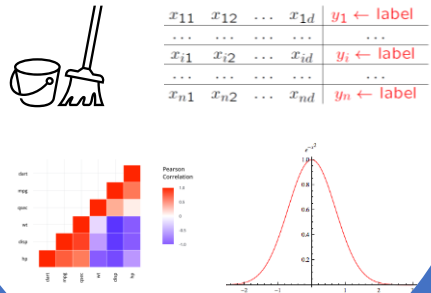
Artificial Intelligence

Technical Introduction

ETL & EDA



Data



AI

Data Visualization & Data Storytelling



AI Technology

1956

Artificial Intelligence

The field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

1997

Machine Learning

Subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions.

2017

Deep Learning

A machine learning technique in which layers of neural networks are used to process data and make decisions.

2021

Generative AI

Create new written, visual, and auditory content given prompts or existing data.

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI

Machine Learning

INPUT

Data with labels

Data without labels

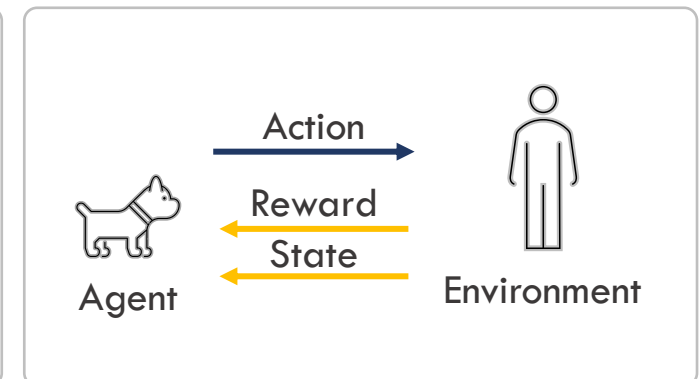
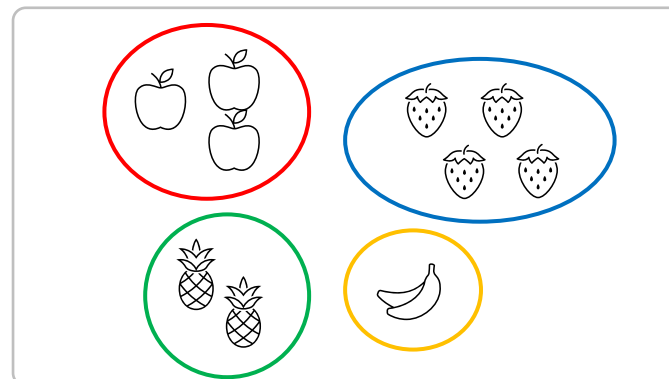
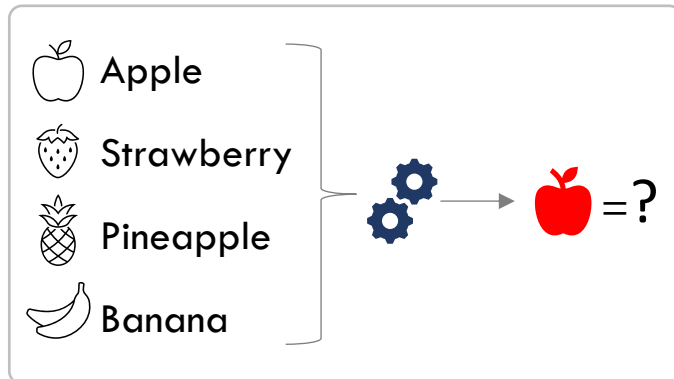
States and Actions

MODEL

Supervised Learning

Unsupervised Learning

Reinforcement Learning



OUTPUT

Label

Cluster

Action

Deep Learning

What is it?

Deep Learning is a field within ML where intelligence is induced using algorithms that have been **inspired by** the biological functioning of the **human brain**.

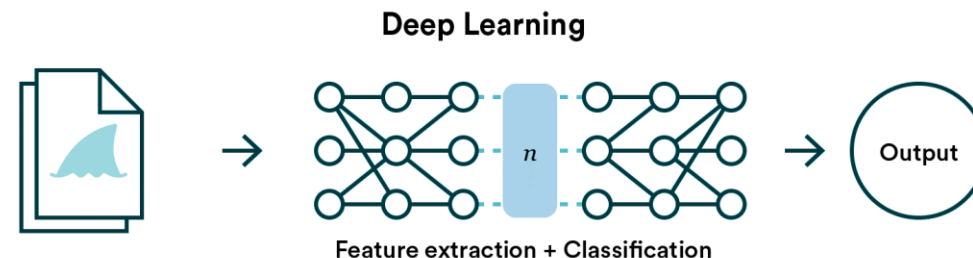
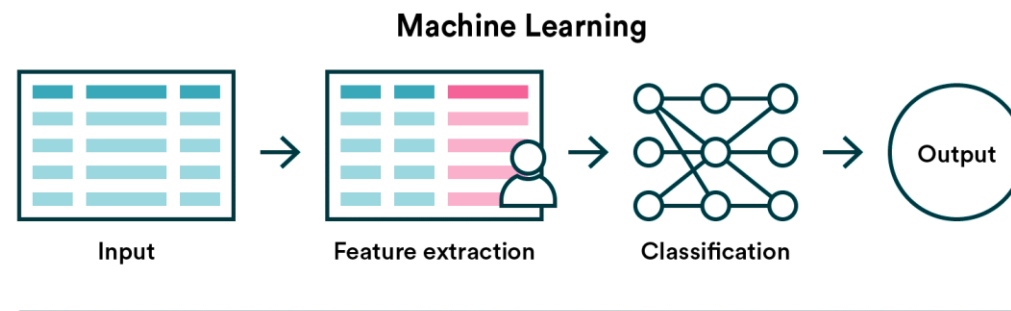
It refers to the usage of neural networks with **numerous layers**, hence the term "deep". Deep neural networks can have tens, hundreds, or even thousands of hidden layers, in contrast to conventional shallow networks that only have one or two.

Why is it important?

- 1 Speech Recognition
- 2 Natural Language Processing
- 3 Image Recognition at a High Level



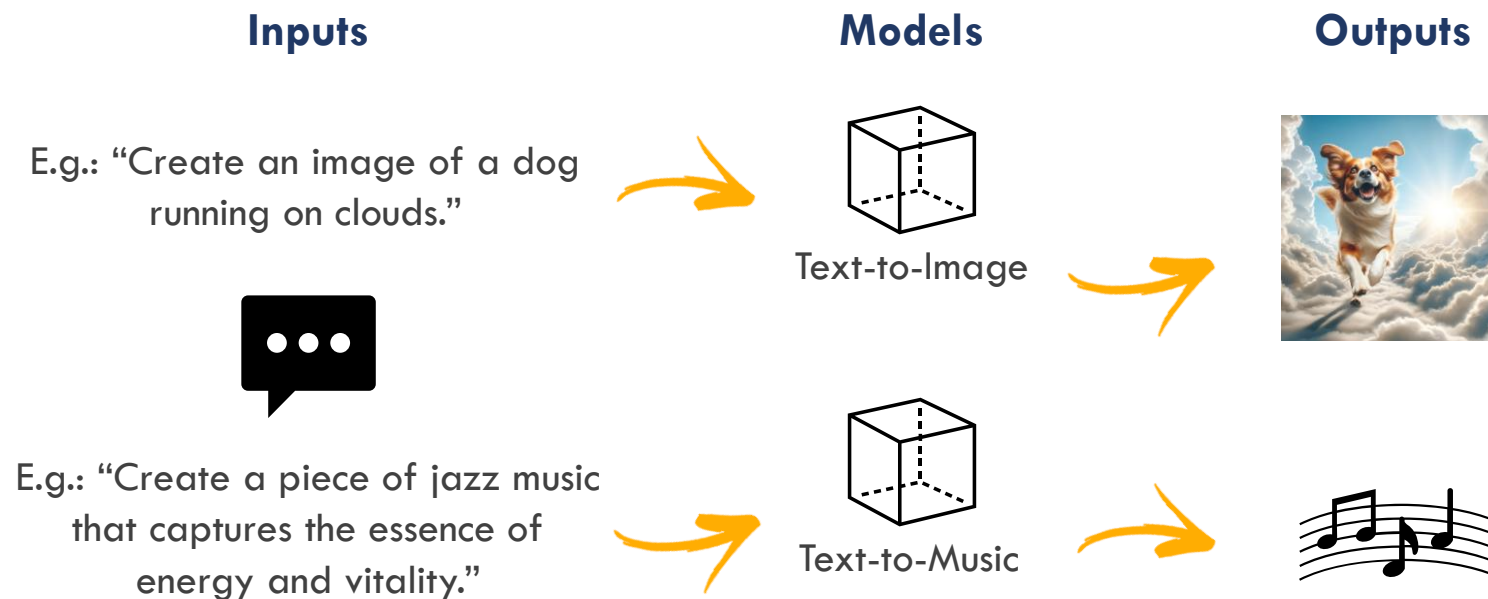
It requires **a lot of data to learn**, sometimes millions of observations



Generative AI

What is it?

Generative AI is a branch of artificial intelligence focused on **creating new and personalized content**, miming human creation. It uses Machine Learning techniques to learn patterns from **vast amounts of data** and generate realistic and creative outputs. Generative AI can produce high-quality content in various domains, such as **images, text, videos, and audio**.

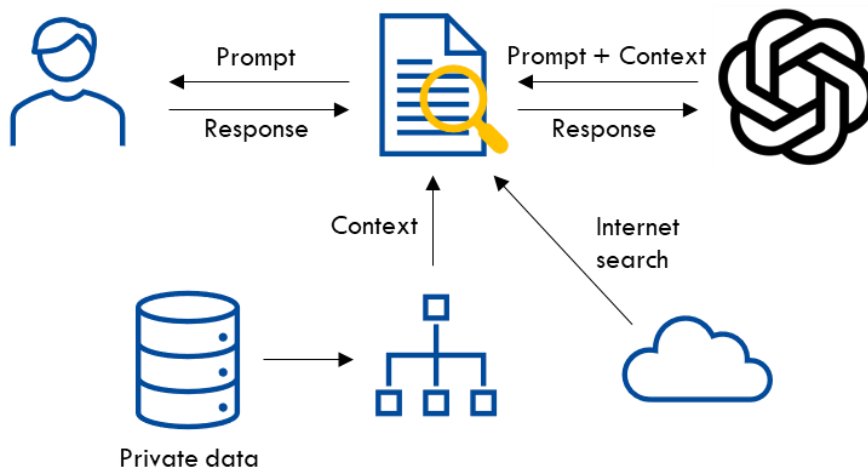
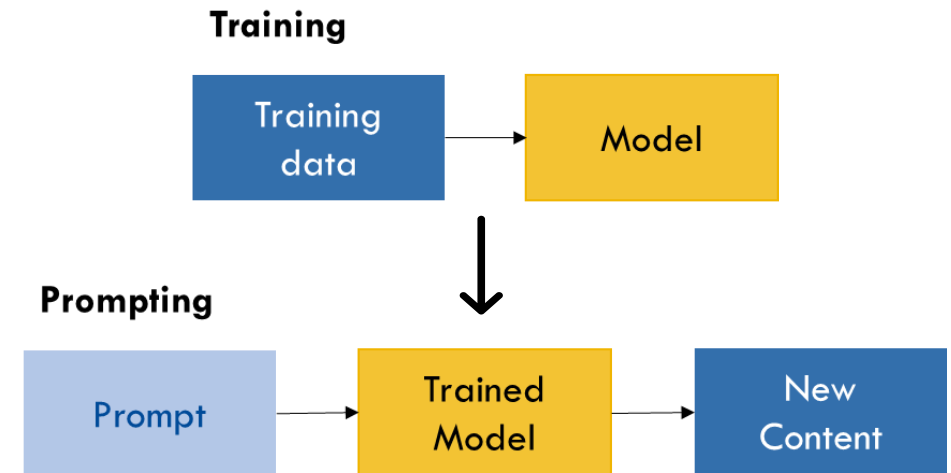


Generative AI

How does it work?

GenAI tools are fundamentally **sequence-prediction machines**. When they are prompted, they complete a sequence with whatever is **most likely to appear**.

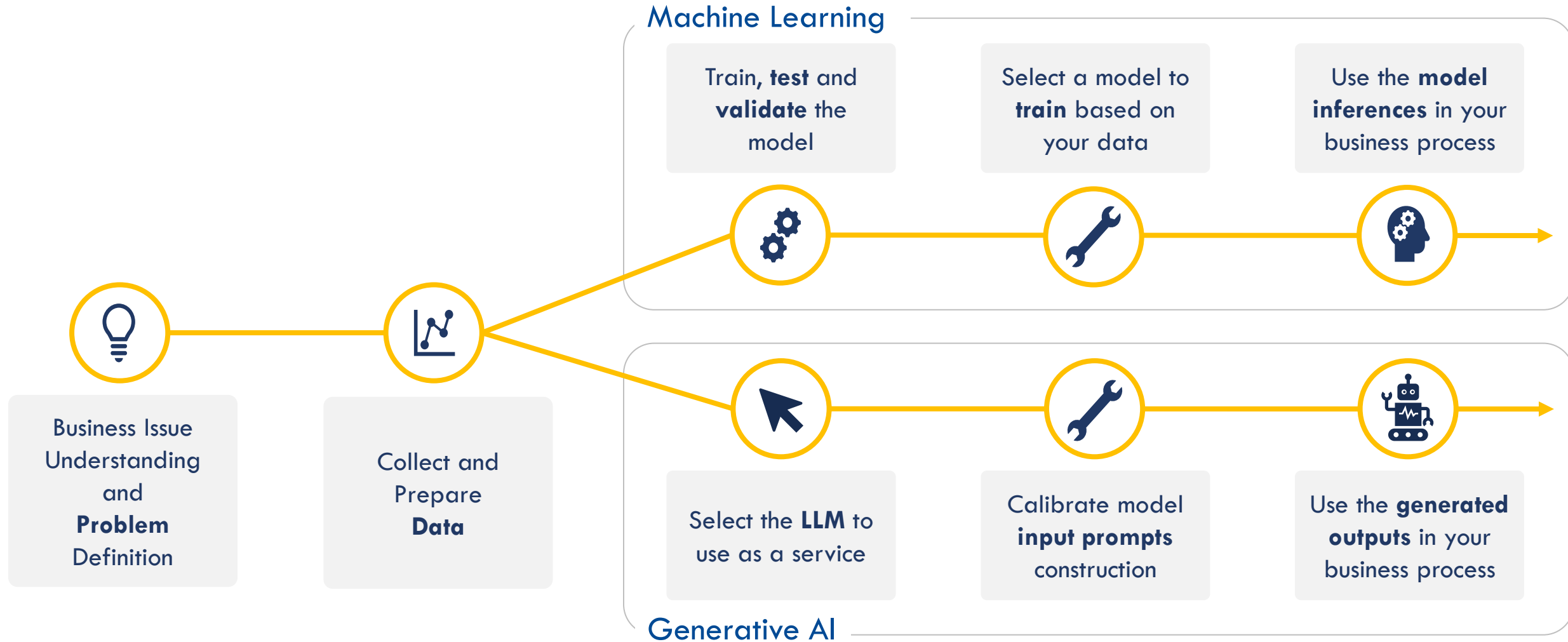
Remember, prompts are not limited to text!



Text-based GenAI tools (e.g., Copilot and ChatGPT) learn patterns in language from a **massive amount of text data available on the internet**.

To build responses, an **AI agent** can also retrieve data from **private data sources** and perform **internet researches**.

Machine Learning vs Generative AI Workflow





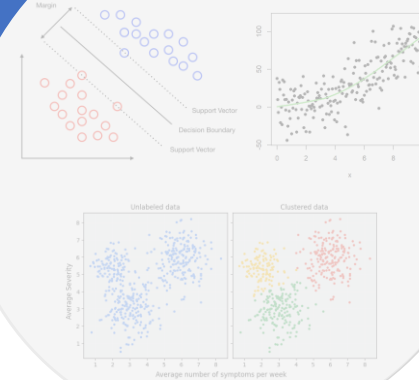
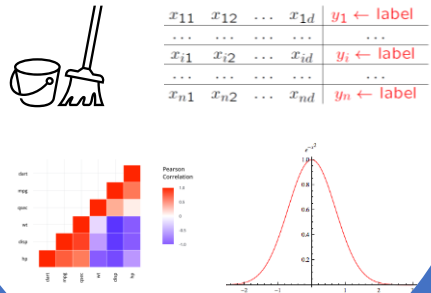
Data Visualization & Data Storytelling

Technical Introduction

ETL & EDA



Data



AI

Data Visualization & Data Storytelling





How can we turn data into stories that inspire action?



DATA VISUALIZATION

The **graphical representation** of data to help people understand patterns, trends, and insights through charts, graphs, and other visual tools



DATA STORYTELLING

The use of **data** combined with **narrative techniques** to communicate insights in an engaging and clear way

Data Visualization is one of the main tools of Data Storytelling: it helps make the data's story visually appealing and easily understandable.



Data Visualization

What is it?

Data visualization is the **graphical representation** of information and data through **visual elements** like charts and graphs.

It **simplifies** complex datasets, making **patterns**, trends, and outliers **easily interpretable**. By encoding data visually, it transforms raw numbers into **actionable insights**, leveraging design and human perception for effective communication.

Why is it important?

- 1 Enhanced Understanding
- 2 Data-Driven Decision Making
- 3 Effective Communication
- 4 Improved Engagement
- 5 Tracking KPIs



Data Visualization



DASHBOARD

A dashboard is a **visual tool** that collects and displays key information, including KPIs, to **facilitate monitoring and decision-making**. Its primary purpose is to clearly visualize data and highlight key business performance metrics.



KPI

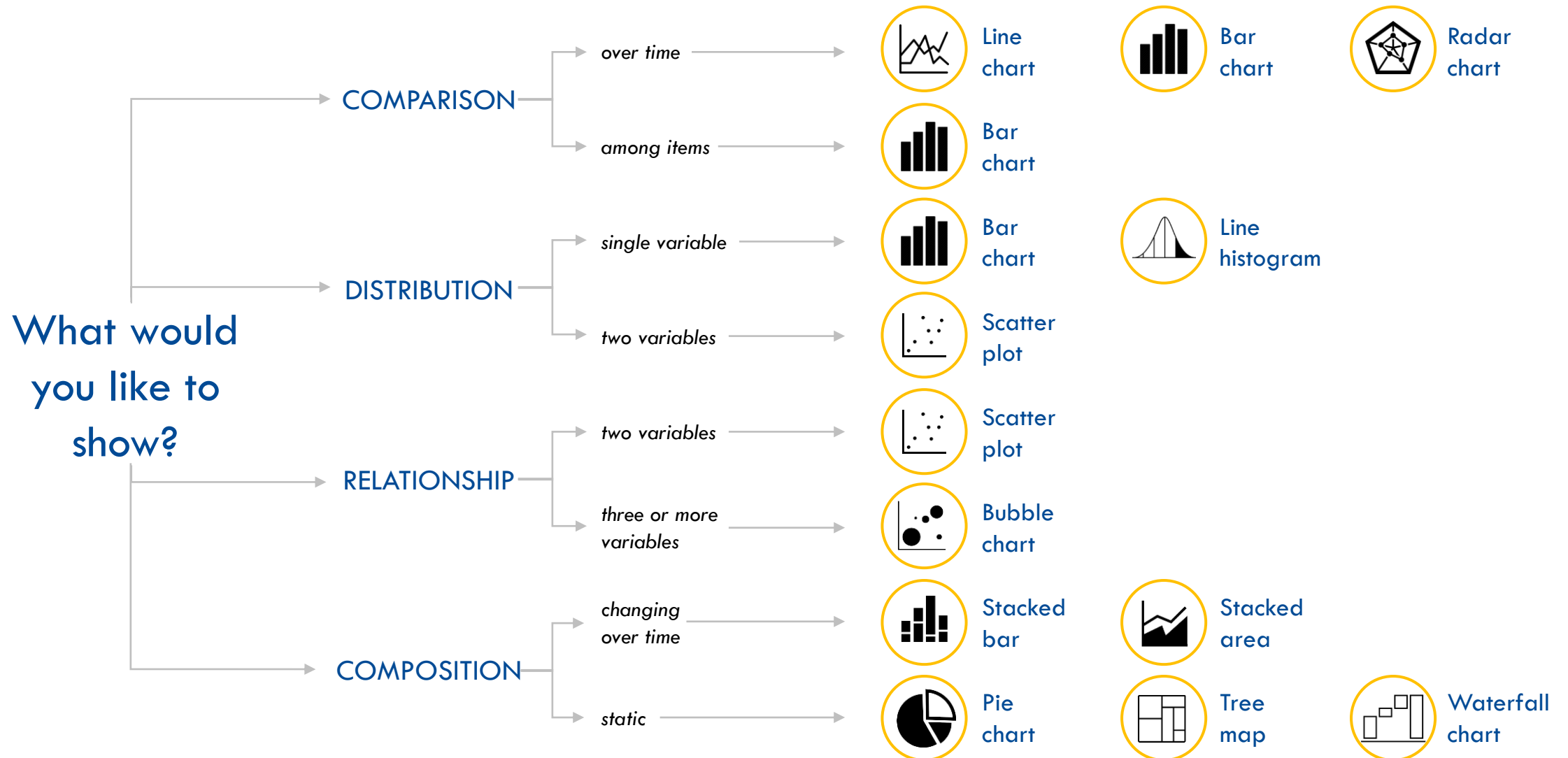
KPIs are **measurable values** that indicate how effectively an individual, team, or organization is **achieving specific objectives**. Examples include sales revenue, customer satisfaction scores, or website traffic.



GRAPHS AND CHARTS

Graphs are essential for **visualizing and interpreting data** quickly and understandably. Without them, the information could be hard to analyze.

Graphs cheat sheet





Design Guidelines

Best Practices

- **Proportionality** and fairness to the represented number
- **Utility** to convey useful information
- **Clarity** for straightforward messages
- **Use Contrast, Colors, Textual elements** to provide support to understanding and convey specific messages
- **Efficiency** and **efficacy** to obtain correct perception of data

Most Common Errors

- **Misleading representations** like improper scaling or alignment
- **Volume or area misrepresentation** when comparing areas
- **Overuse of color** and ignore accessibility to guarantee optimal readability
- Be aware to **cultural sensitivity**:
 - Directionality in reading
 - Calendar and date formats
 - Color association and symbol interpretation

Data Storytelling

What is it?

Data storytelling is the process of **transforming data analyses into an understandable storyline** for a wider audience in order to influence the decisions of business users and other stakeholders.

It combines data visualization, contextual information, and narrative techniques to **explain findings, reveal patterns, and support decision-making.**

Why is it important?

- Bridge the **gap** between **raw** data and **meaningful insights**
- **Enabling stakeholders** to understand the implications of the data and take informed actions.

Components of Data Storytelling

Visualization

Even simple datasets are hard to read. Represent the data **using** a **graph** with **labels, colors** and **targets**

Narrative

What am I looking at?
No visualization goes **undescribed**

Context

Why is this happening?
The data alone is not enough to **tell the story**

Emotional Triggers

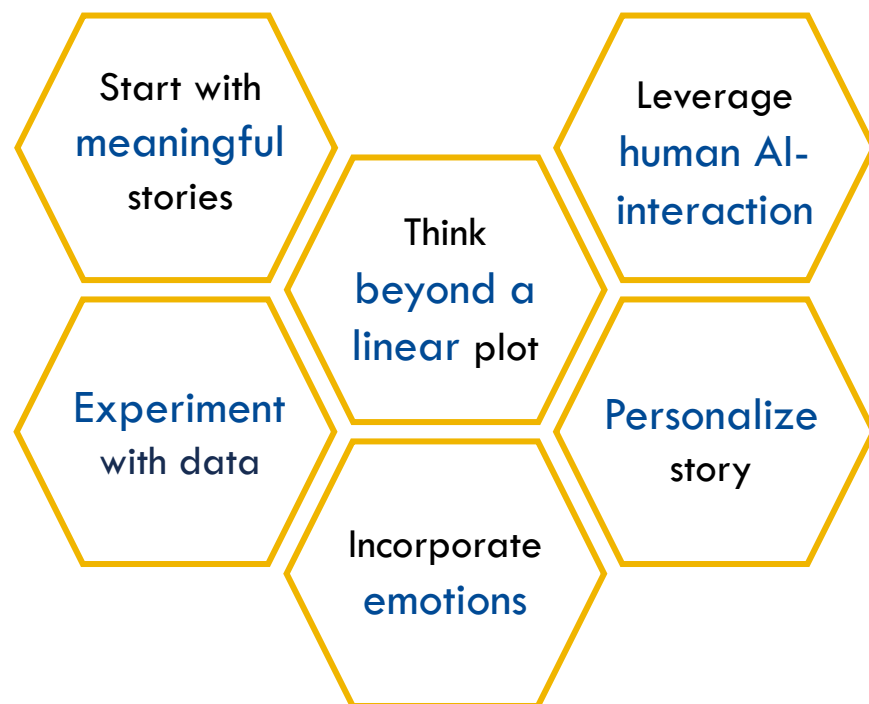
Why is this happening?
What are we going to do?
Data story must **lead** to **insight** and **action**

Decision Options

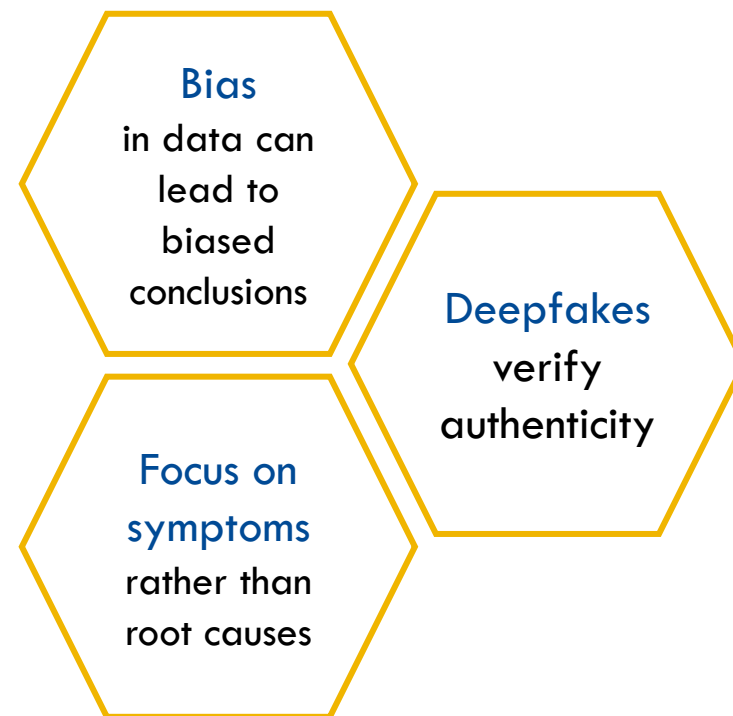


Storytelling Guidelines

Best Practices



Most Common Errors

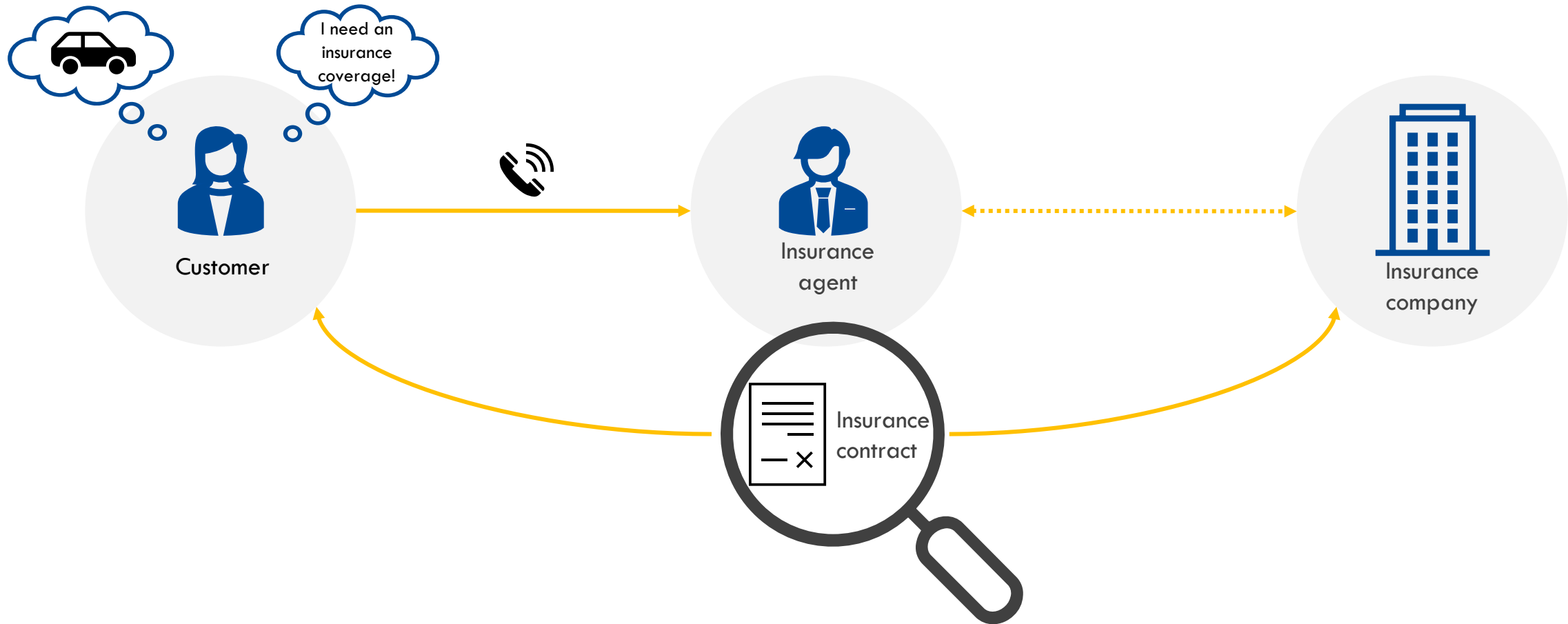




Claims

The insurance lifecycle

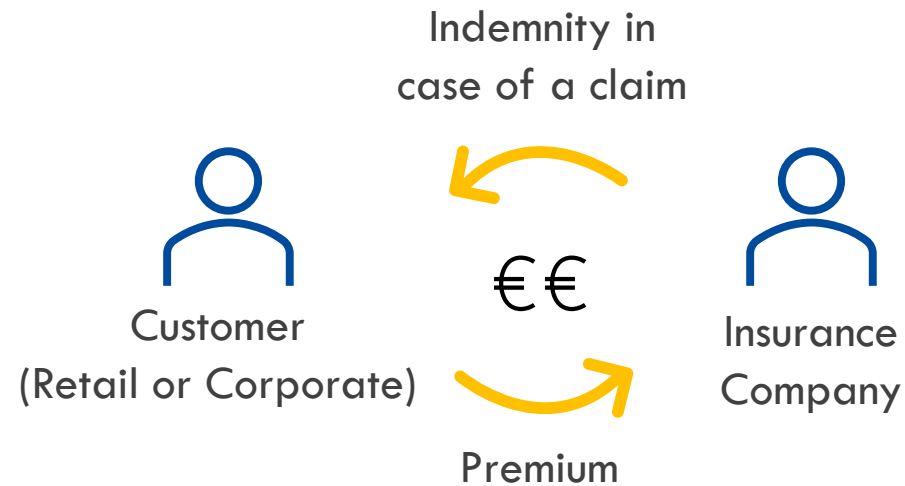
How everything starts...



Insurance Contract

The insurance contract is a **legal agreement** between the insurer and the customer.

The insurer, in exchange of a payment called **premium**, is required to pay an **indemnity** to the insured in case of a **claim** covered by the contract.



How can you calculate the premium?

- Customer's information
- Details about the good
- Risk profile
- Chosen warranties

Warranties

The warranties specify the risks that your insurance policy covers. Depending on the type of insurance policy, different warranties are defined. Below are some examples of warranties for various types of insurance:

Car insurance

- Total or partial theft
- Fire
- Vandalism
- Natural disasters
- Collision with wild animals
- Glass



House insurance

- Theft
- Fire
- Liability insurance (i.e. pet custody)
- Glass
- Legal protection



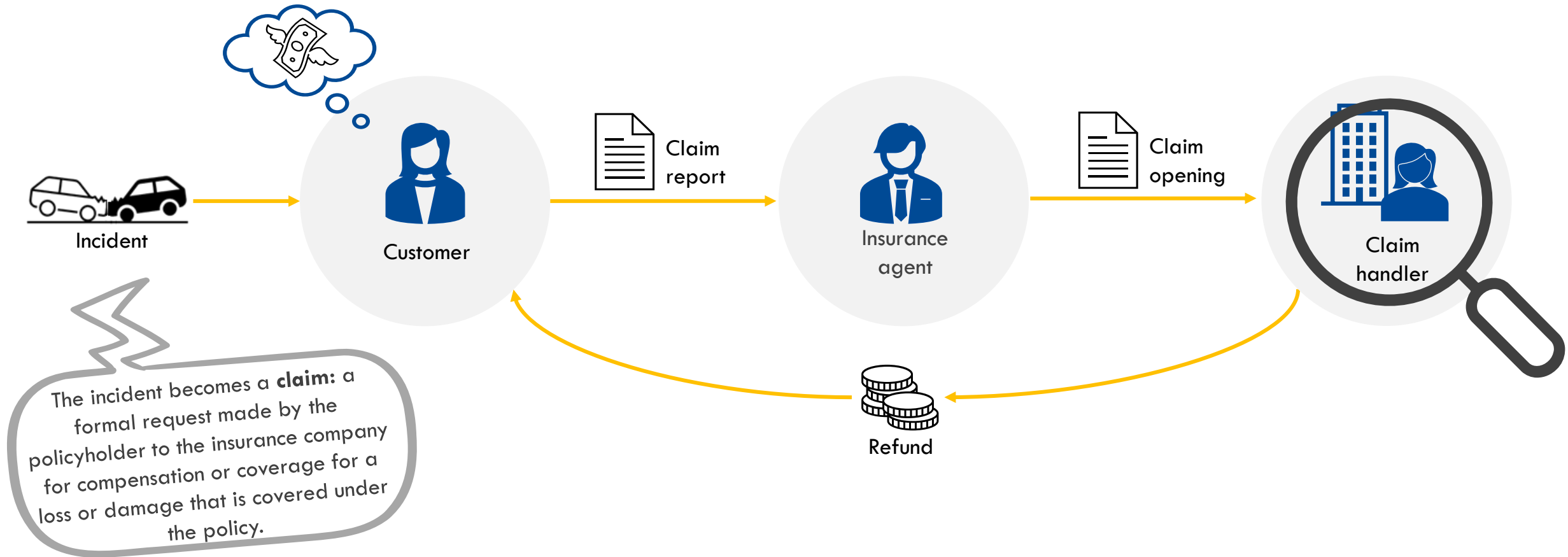
Health insurance

- Medical exams for suspected illness
- Incident/illness medical treatments
- Hospitalization
- Medicines purchase
- Virtual Hospital

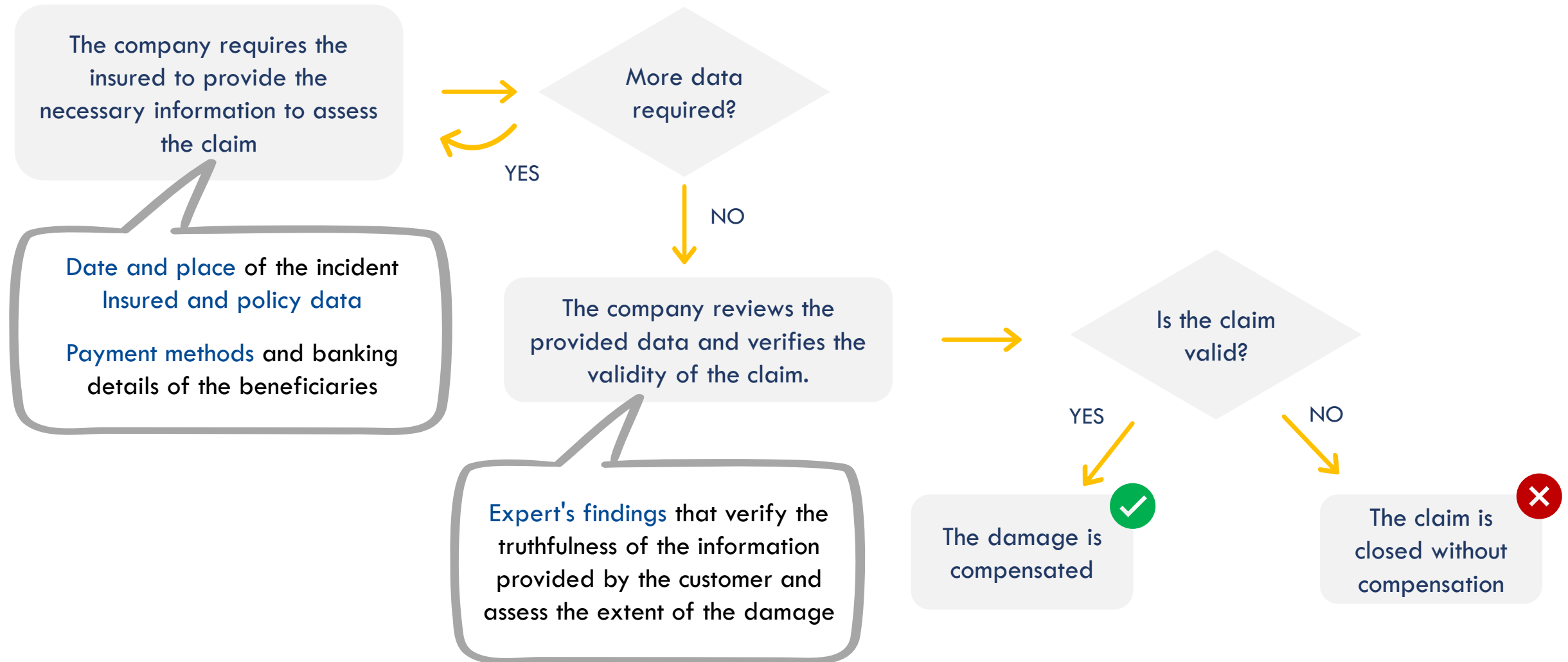


The insurance lifecycle

... and how it continuous



Claim Handling Lifecycle





Case Study



Case Study - Input

You will receive a dataset related to VEHICLE Claims. Below you can find the description of the columns.

COLUMN NAME	COLUMN DESCRIPTION
CLAIM_ID	A unique identifier assigned to each claim, used to distinguish each record in the database
POLICYHOLDER_AGE	The age of the policyholder
POLICYHOLDER_GENDER	The gender of the policyholder
WARRANTY	The type of warranty affected by the claim, such as vehicle damage, third-party liability, etc.
CLAIM_DATE	The date the claim was officially recorded with the insurance company
CLAIM_REGION	The region where the accident occurred
CLAIM_PROVINCE	The province where the accident occurred
VEHICLE_BRAND	The brand of the vehicle of the policyholder involved in the claim
VEHICLE_MODEL	The model of the vehicle of the policyholder involved in the claim
CLAIM_AMOUNT_PAID	The total amount paid by the insurance company to settle the claim, including damages, compensation, or repair costs.
PREMIUM_AMOUNT_PAID	The amount the policyholder pays to the insurance company in exchange for coverage of the risk specified in the contract



Case study - Output

Start with this dataset and try to extract some **meaningful insights** that can be used to improve any process in the insurance industrial context.

Feel free to apply any techniques related to Data Exploration and AI algorithms.

Need more data? No problem! You can integrate relevant open-source data to supercharge your analysis.

Let's see which amazing things you can discover!



REPORT STRUCTURE

- Presentation in pdf format
- Max 10 slides
- In English language
- Insert name, surname, and student id of all team members
- Give a title to the project



EVALUATION

- 30% - Clarity & Presentation
- 30% - Insightfulness & Actionability
- 20% - Technical Implementation
- 20% - Creativity & Innovation



Report Guideline

Use the following questions to guide your solution.



DATA ANALYSIS

- How did you decide to approach the dataset? What is the **aim** of your analysis?
- Which kind of techniques did you decide to apply and why?
You don't need to cover all theoretical concepts apply them as needed (for example, machine learning algorithms are optional).
- Which kind of visualizations have been exploited to perform the analysis and why?



INSIGHTS

- Do the insights extracted from the data support your thesis/the objective of your analysis?
- Do the insights extracted can be translated in concrete actions?
- Which aspects of your analysis are innovative?



BENEFITS

- Do you think your project could be applied to a real-case scenario?
- Which is the impact of the analysis performed?
- What is the value generated?



Today Scheduling

