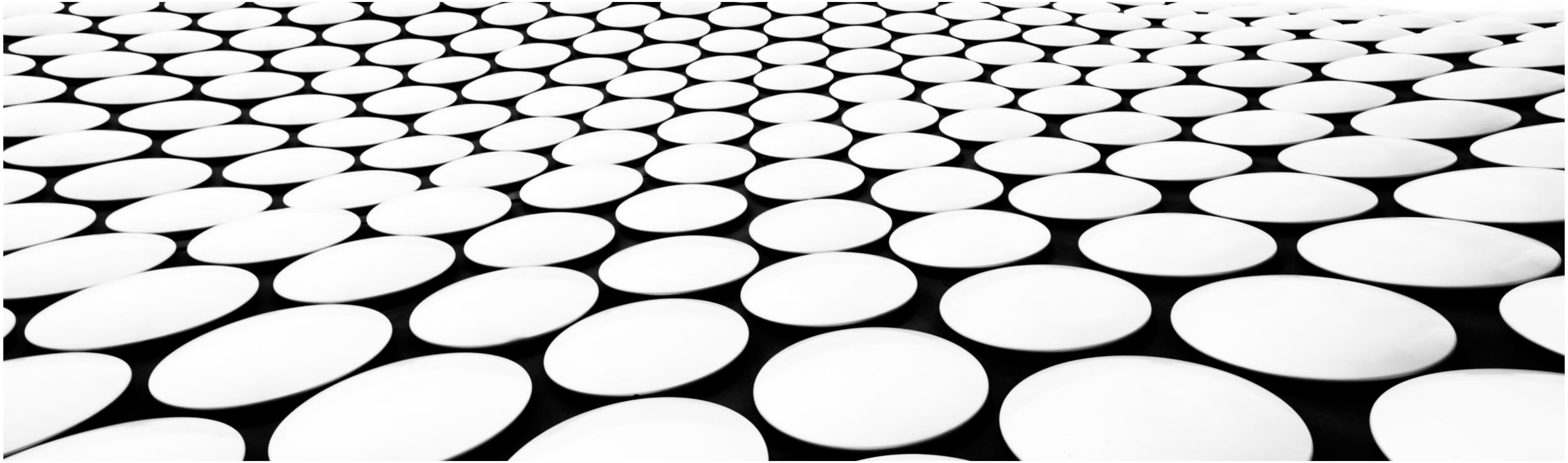
DERIVING INSIGHTS FROM DATA

Input/Output in Data Science



DERIVING INSIGHTS FROM DATA

- Types of insights
- Types of data

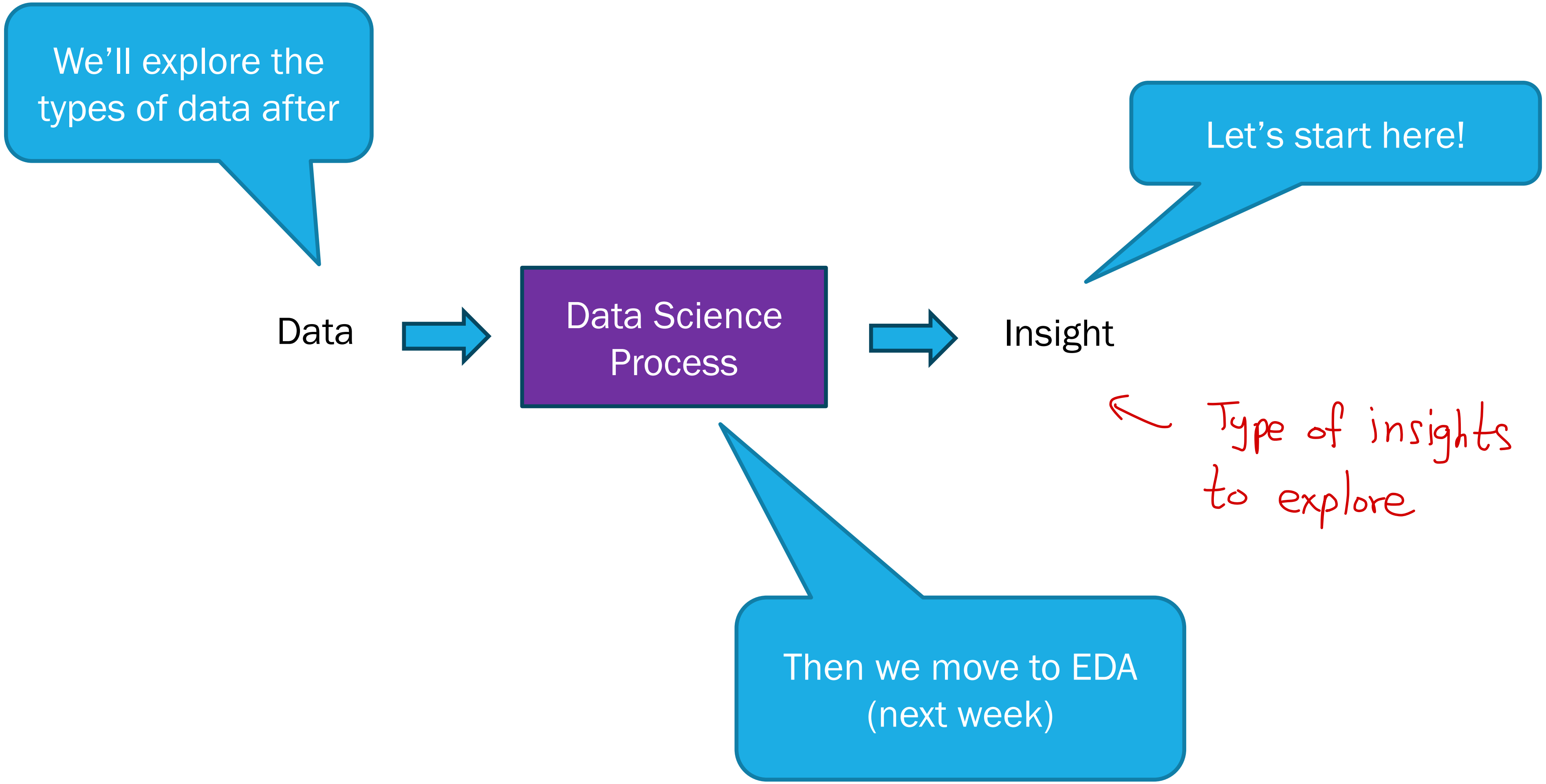
Back to Data Science Definition(s)

Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processing, scientific visualization, algorithms and systems to extract or extrapolate knowledge and insights from potentially noisy, structured, or unstructured data.

Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI) and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning.

Data science is the study of data to extract meaningful insights for business. It is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data. This analysis helps data scientists to ask and answer questions like what happened, why it happened, what will happen, and what can be done with the results.

Common: data as input, insights as output



We'll explore the types of data after

Let's start here!

Data

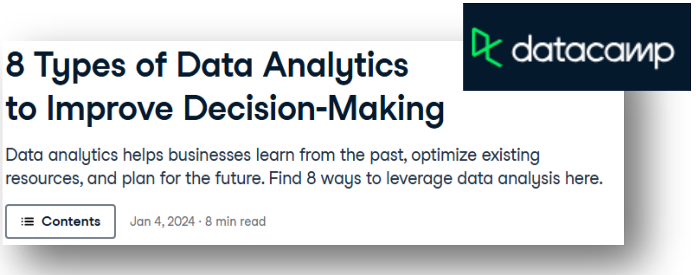
Data Science Process

Insight

Type of insights we want to explore

Then we move to EDA (next week)

Insights



datacamp

8 Types of Data Analytics to Improve Decision-Making

Data analytics helps businesses learn from the past, optimize existing resources, and plan for the future. Find 8 ways to leverage data analysis here.

Contents Jan 4, 2024 · 8 min read

The Four Main Types of Data Analytics For Decision-Making

Descriptive →	What happened? storytelling
Diagnostic →	Why did something happen?
Predictive →	Based on what we know now, what will happen in the future?
Prescriptive →	Should we implement some rules or regulation or provide advice based on what we know? modify future

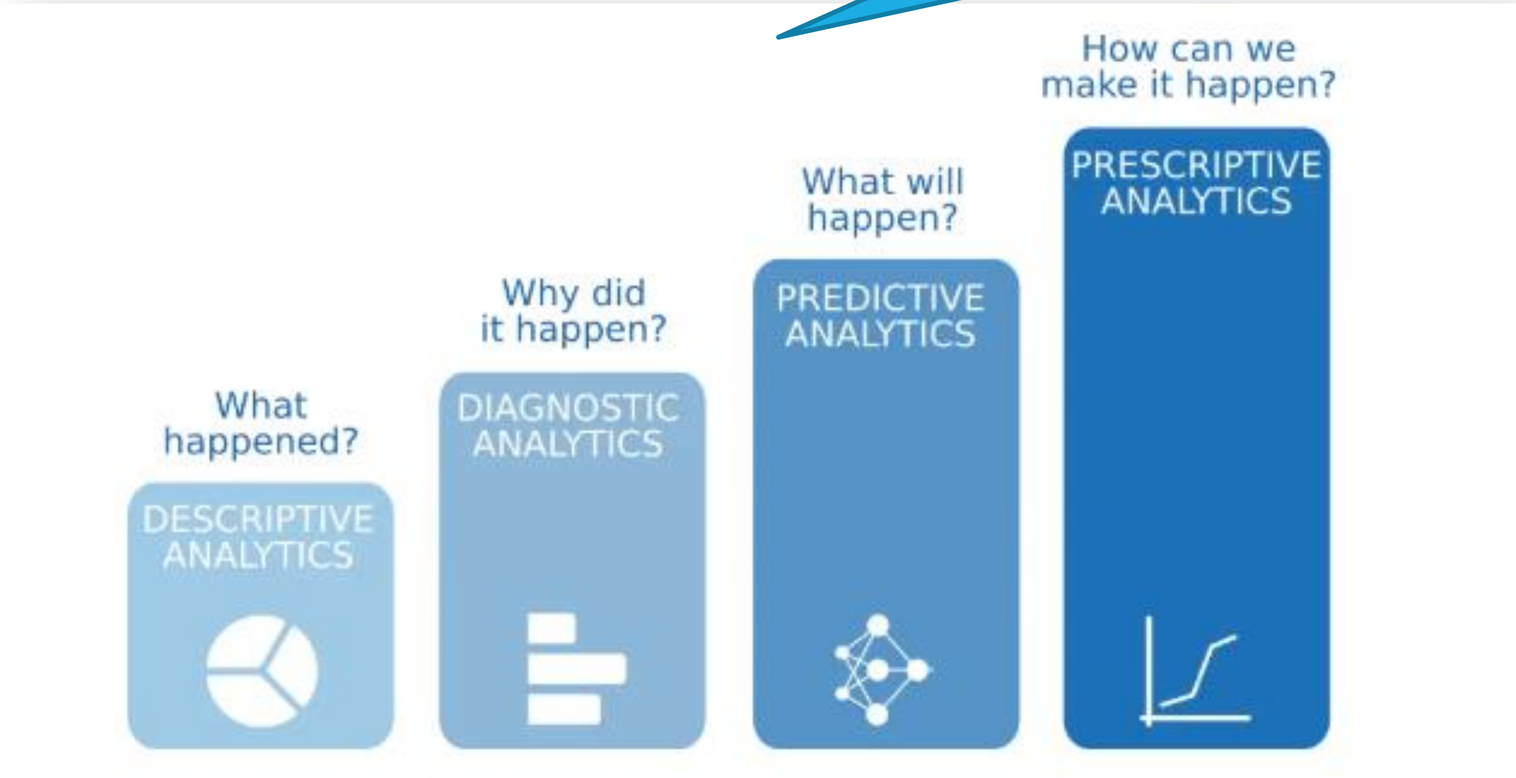
Source

What is advanced analytics?

Tutorials

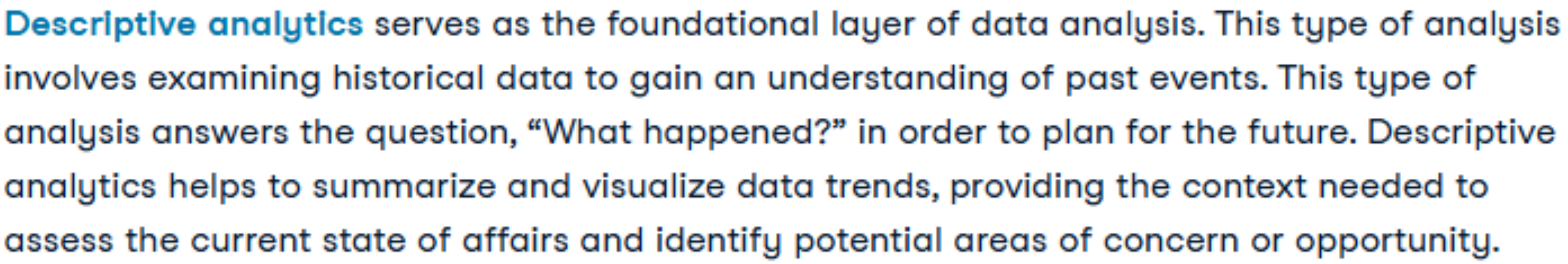
Pablo Martín August 31, 2023

Many sources (as this one) show these 4 types of analytics

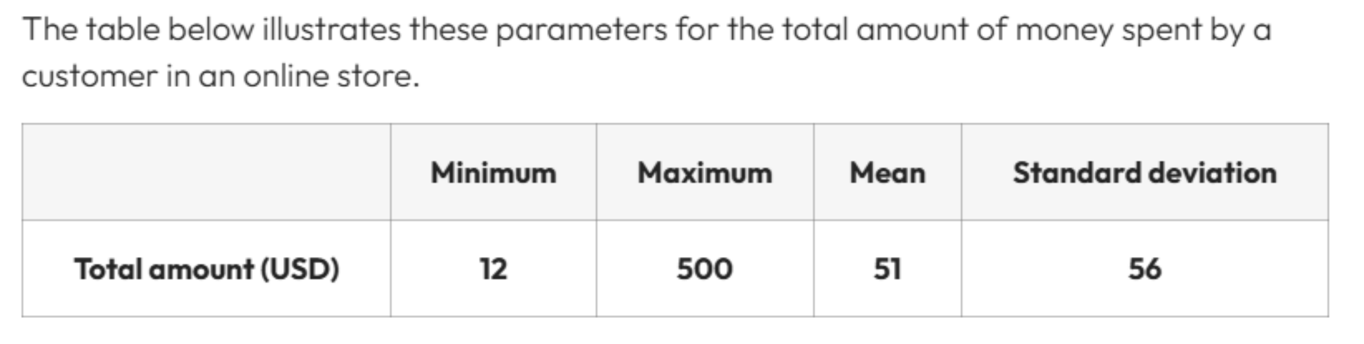


Source

- Descriptive
- Diagnostic
- Predictive
- Prescriptive



Descriptive analytics serves as the foundational layer of data analysis. This type of analysis involves examining historical data to gain an understanding of past events. This type of analysis answers the question, “What happened?” in order to plan for the future. Descriptive analytics helps to summarize and visualize data trends, providing the context needed to assess the current state of affairs and identify potential areas of concern or opportunity.



The table below illustrates these parameters for the total amount of money spent by a customer in an online store.

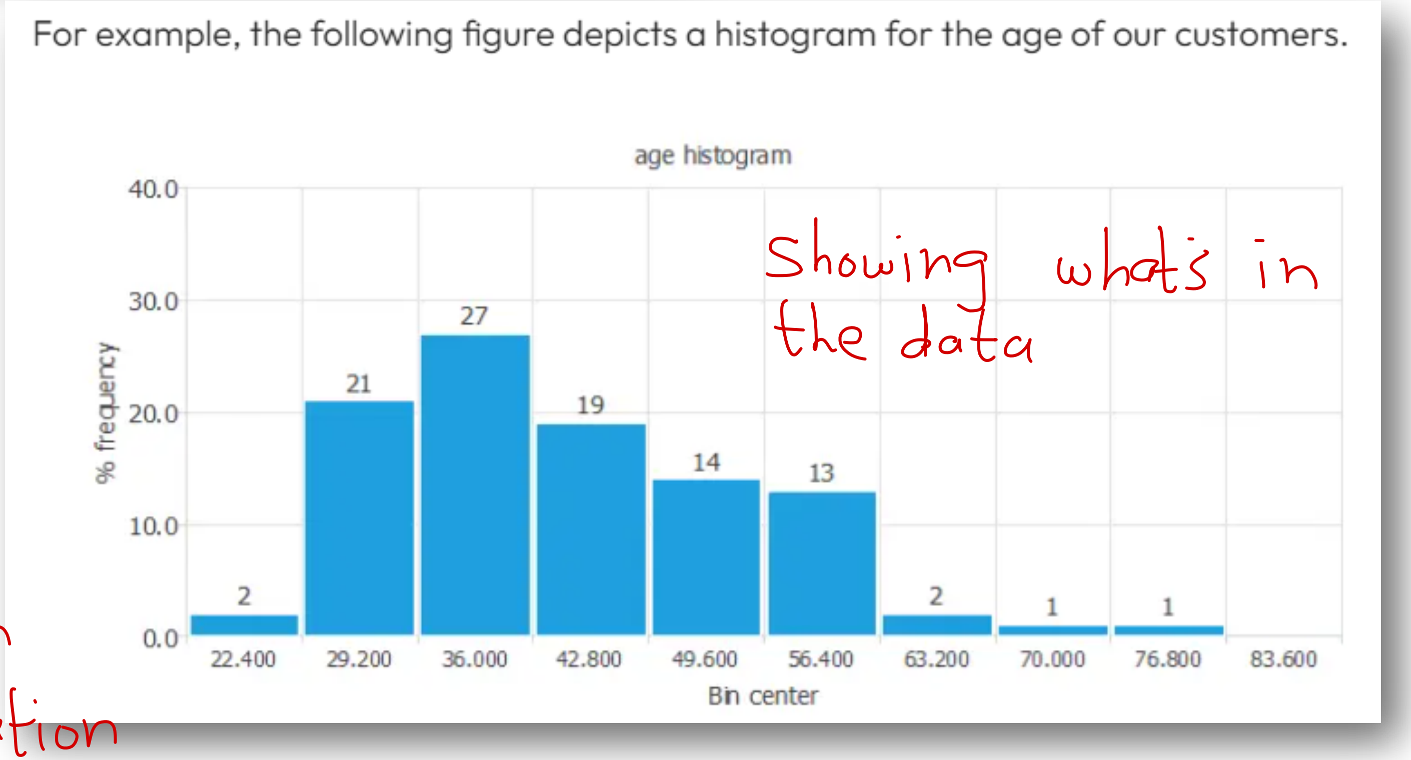
	Minimum	Maximum	Mean	Standard deviation
Total amount (USD)	12	500	51	56

Descriptive Statistics

What happened?

DESCRIPTIVE ANALYTICS

Example in sales

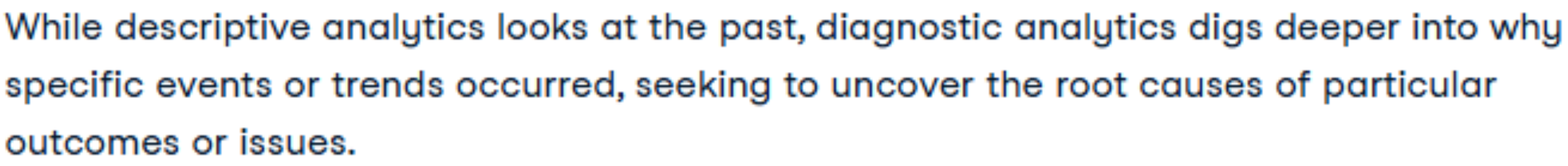


For example, the following figure depicts a histogram for the age of our customers.

Showing what's in the data

Histogram Visualization

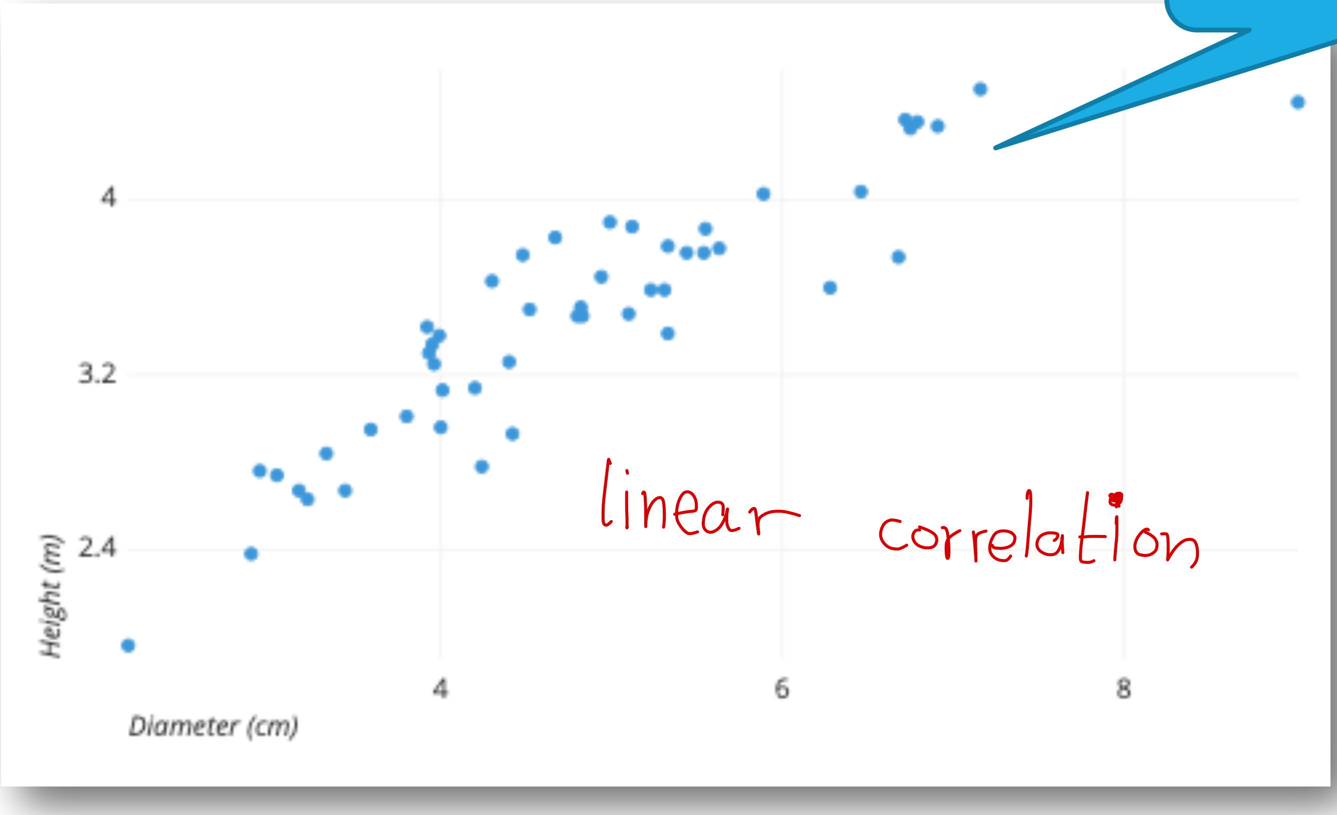
- Descriptive
- Diagnostic
- Predictive
- Prescriptive



While descriptive analytics looks at the past, diagnostic analytics digs deeper into why specific events or trends occurred, seeking to uncover the root causes of particular outcomes or issues.

Tricky sometimes

Example in environment (tree size)



linear correlation

Why did it happen?

DIAGNOSTIC ANALYTICS

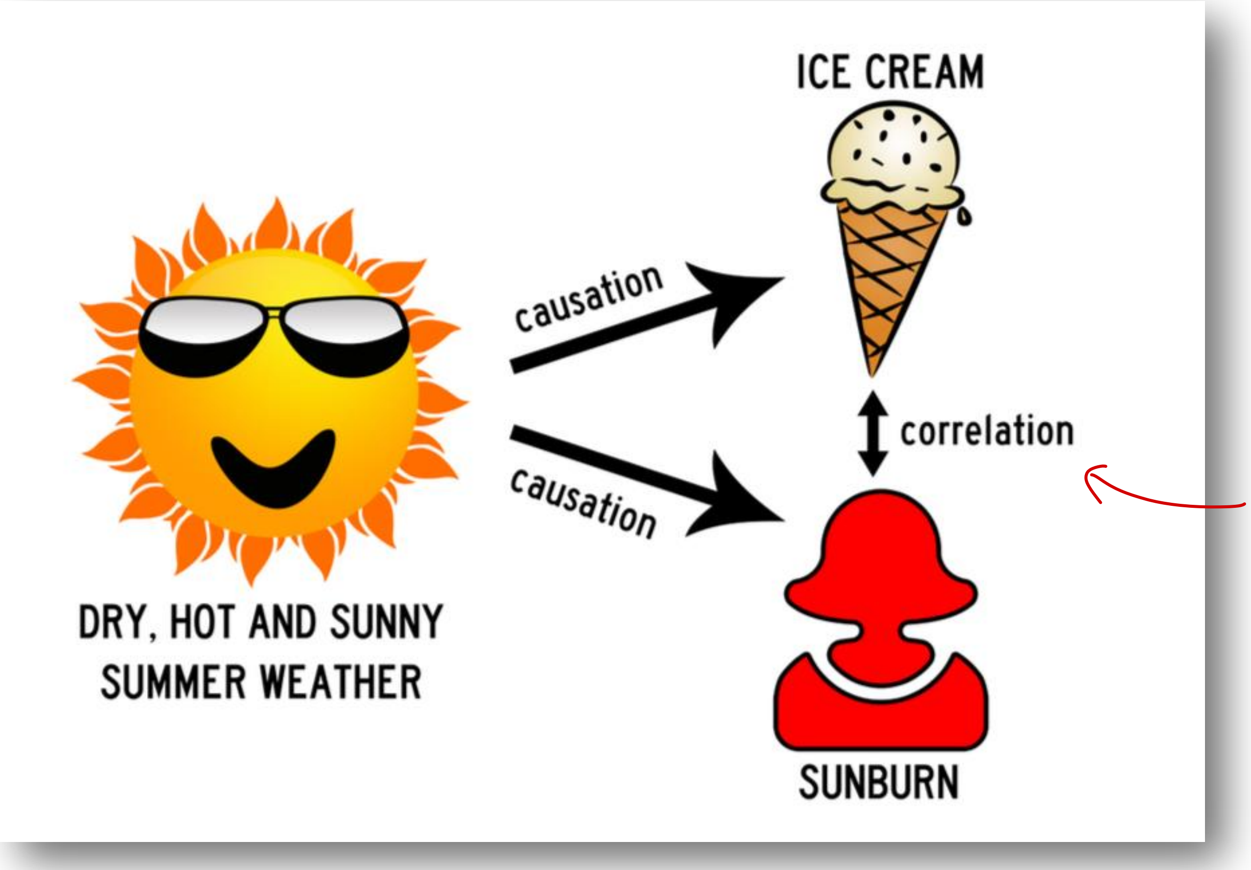
Look at correlation

ATLASSIAN

Source

Why did it happen?

DIAGNOSTIC ANALYTICS



Don't cause each other

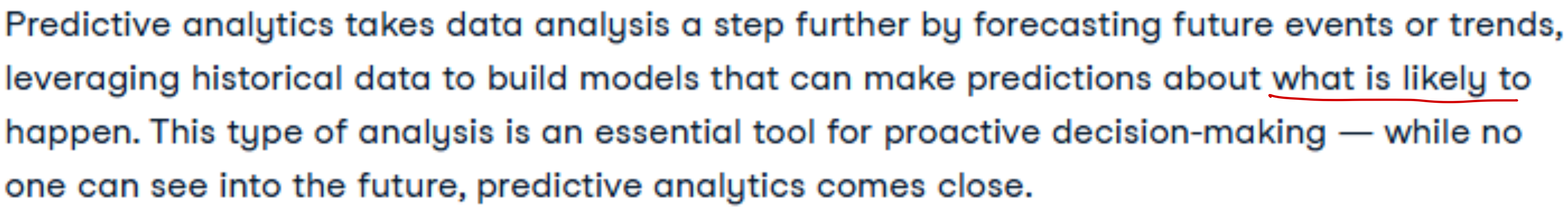
Correlation ≠ Causality

Descriptive

Diagnostic

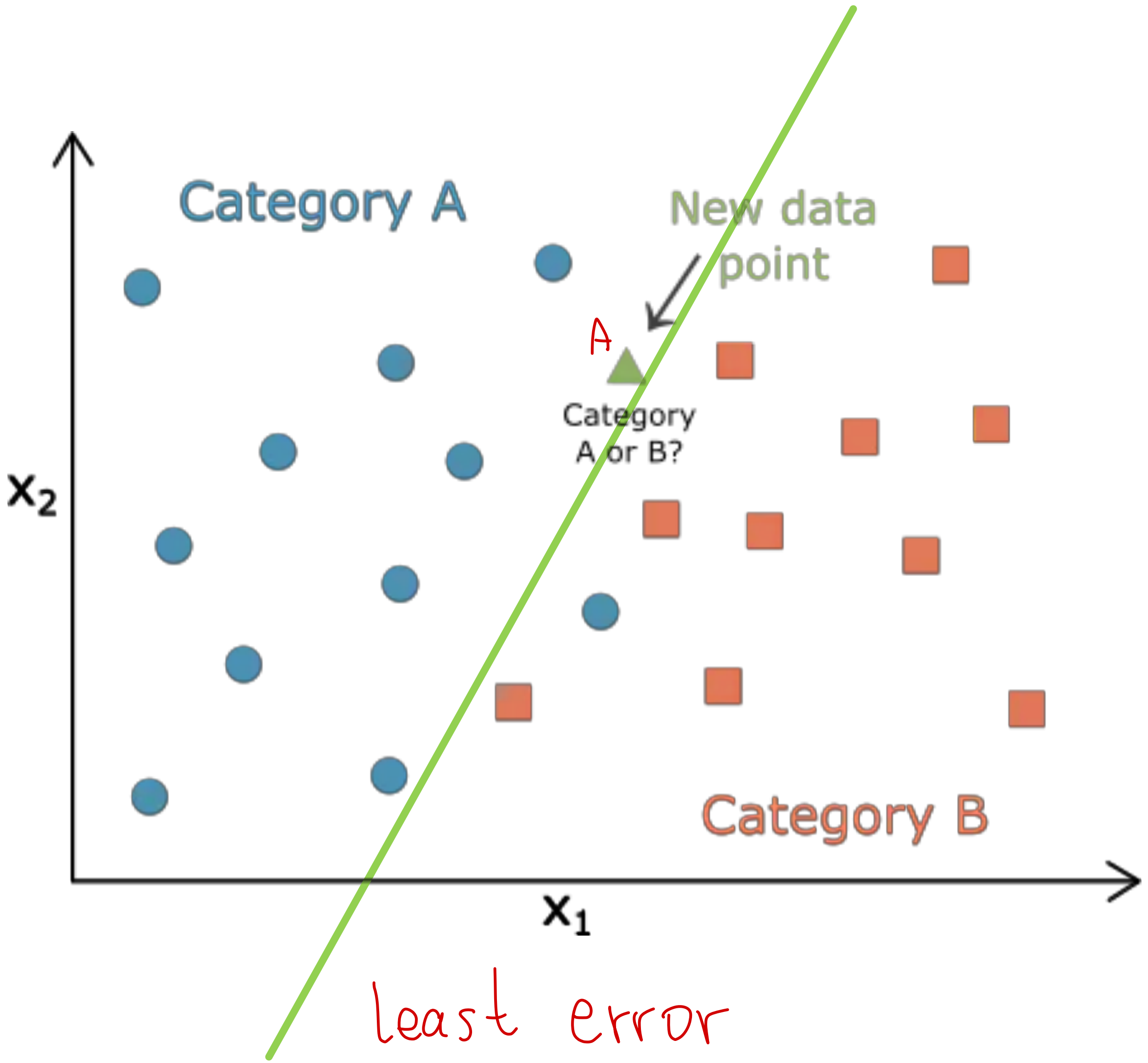
Predictive

Prescriptive

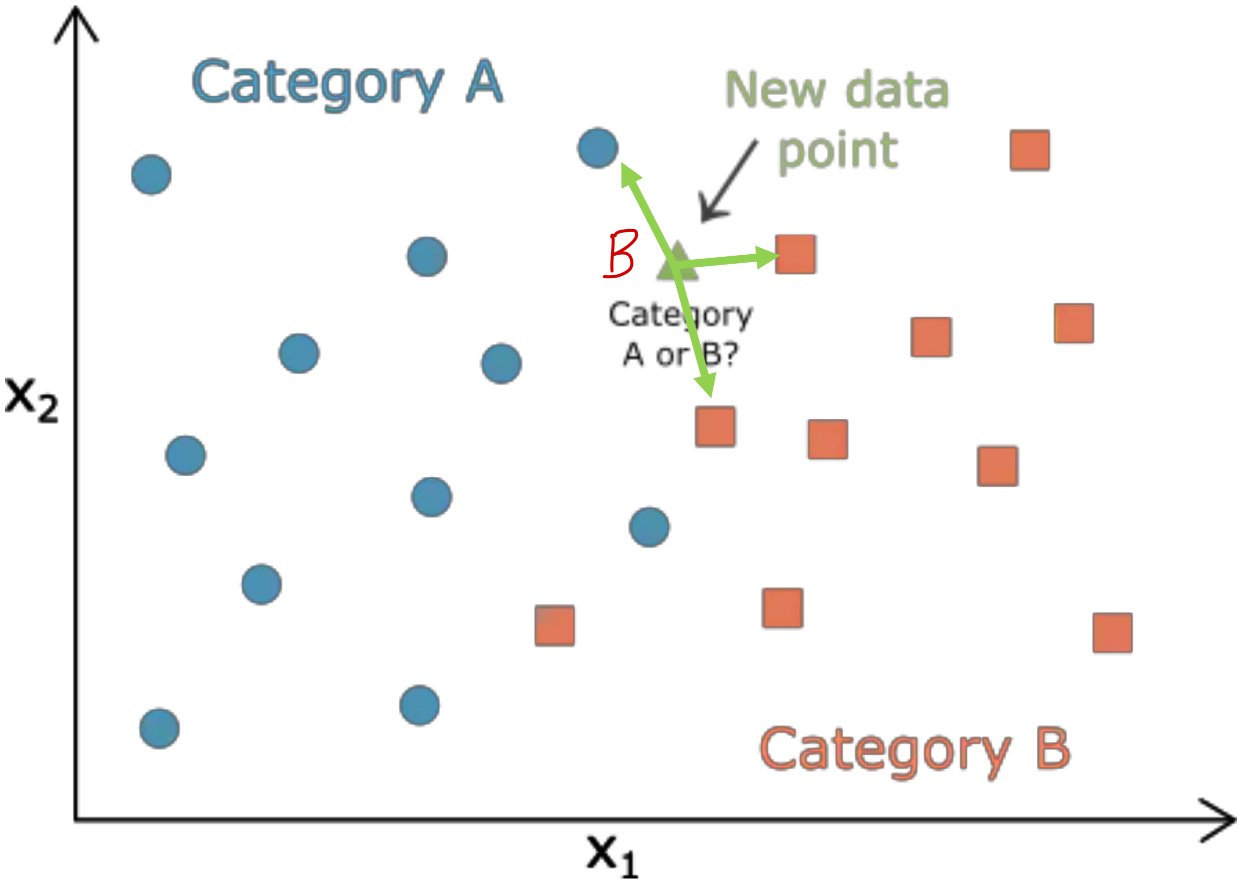


Predictive analytics takes data analysis a step further by forecasting future events or trends, leveraging historical data to build models that can make predictions about what is likely to happen. This type of analysis is an essential tool for proactive decision-making — while no one can see into the future, predictive analytics comes close.

Logistic Regression



K – Nearest neighbors



What will happen?

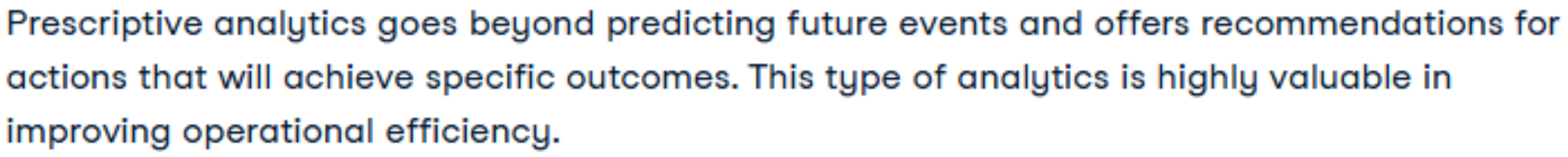
PREDICTIVE ANALYTICS

Descriptive

Diagnostic

Predictive

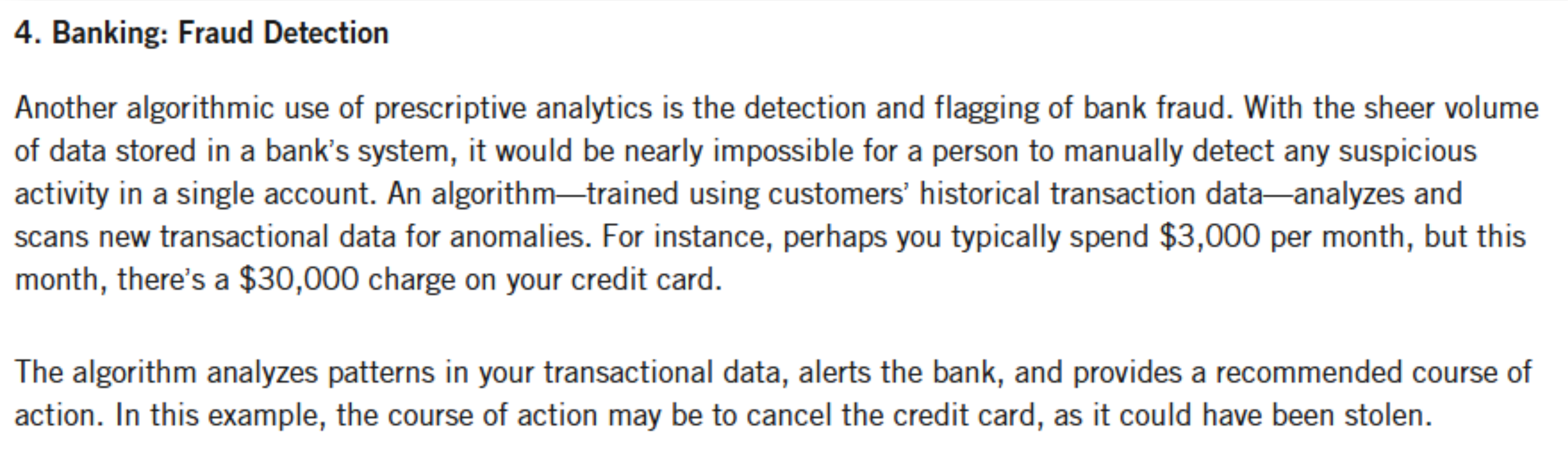
Prescriptive



Prescriptive analytics goes beyond predicting future events and offers recommendations for actions that will achieve specific outcomes. This type of analytics is highly valuable in improving operational efficiency.

Harvard Business School Online

6 EXAMPLES OF PRESCRIPTIVE ANALYTICS IN ACTION



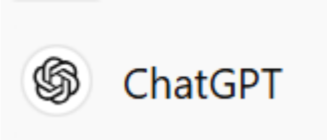
4. Banking: Fraud Detection

Another algorithmic use of prescriptive analytics is the detection and flagging of bank fraud. With the sheer volume of data stored in a bank's system, it would be nearly impossible for a person to manually detect any suspicious activity in a single account. An algorithm—trained using customers' historical transaction data—analyzes and scans new transactional data for anomalies. For instance, perhaps you typically spend $3,000 per month, but this month, there's a $30,000 charge on your credit card.

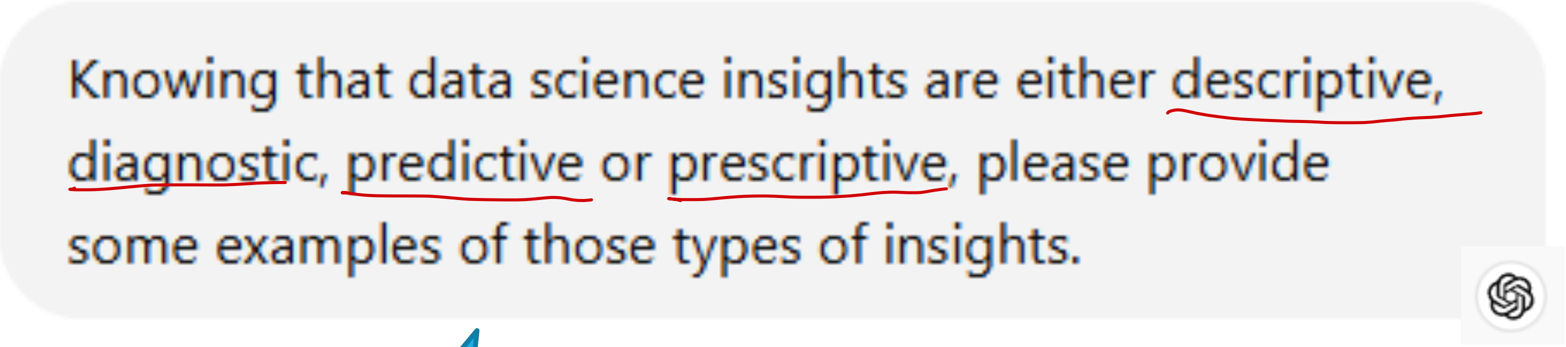
The algorithm analyzes patterns in your transactional data, alerts the bank, and provides a recommended course of action. In this example, the course of action may be to cancel the credit card, as it could have been stolen.

Source

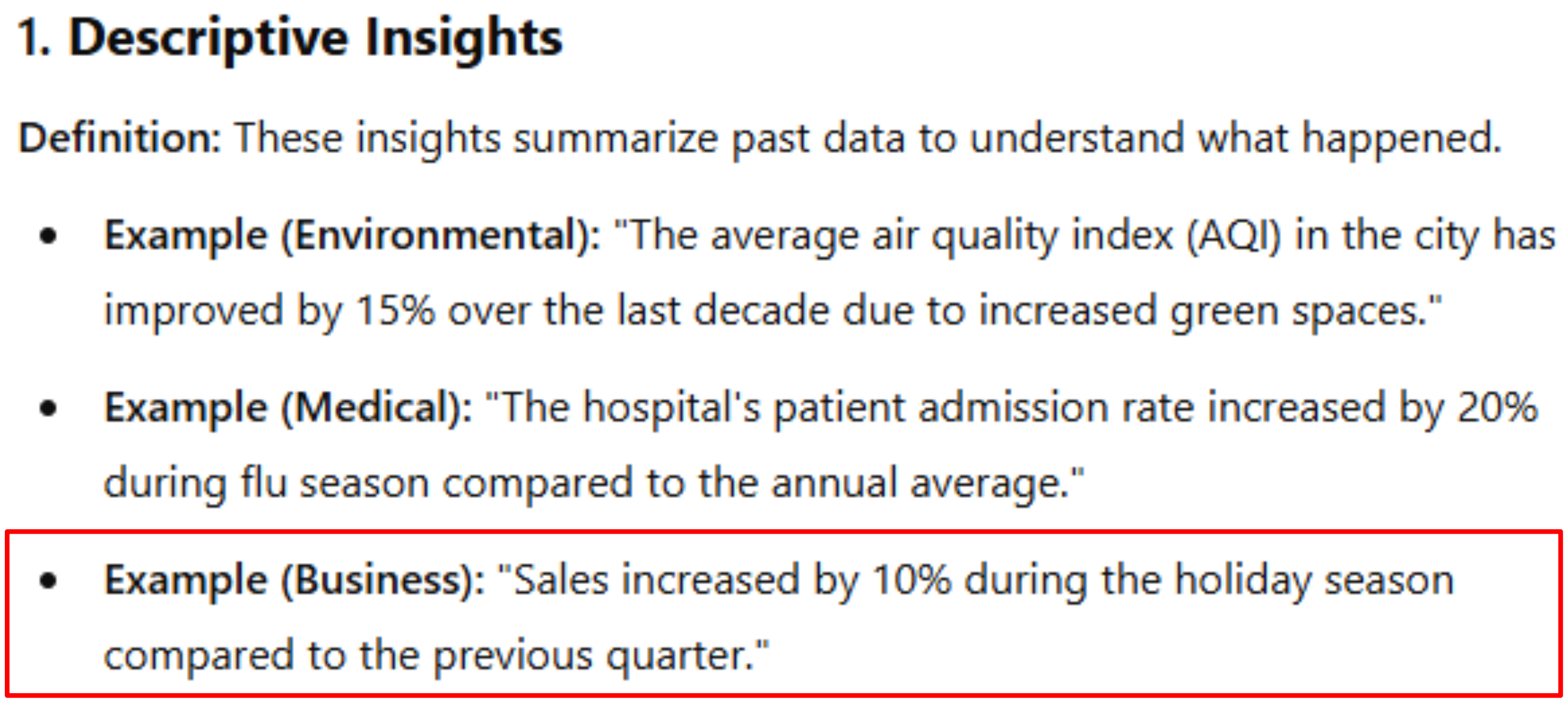
To help you learn… ask for examples.


ChatGPT

To help you learn the terms… explicitly type the different insights to remember them.


Knowing that data science insights are either descriptive, diagnostic, predictive or prescriptive, please provide some examples of those types of insights.

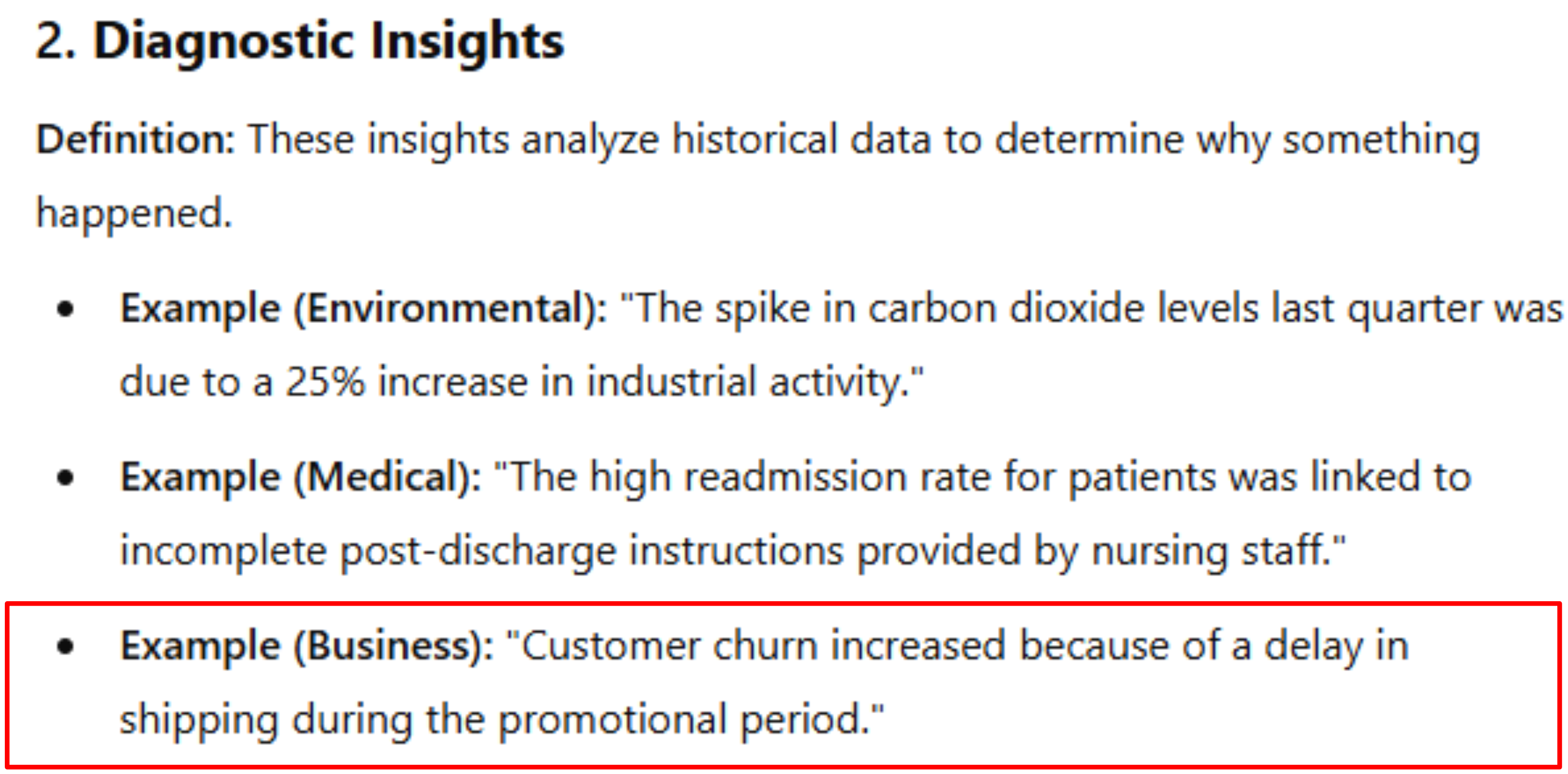
It’s not an oracle.. It can make mistakes.



1. Descriptive Insights

Definition: These insights summarize past data to understand what happened.

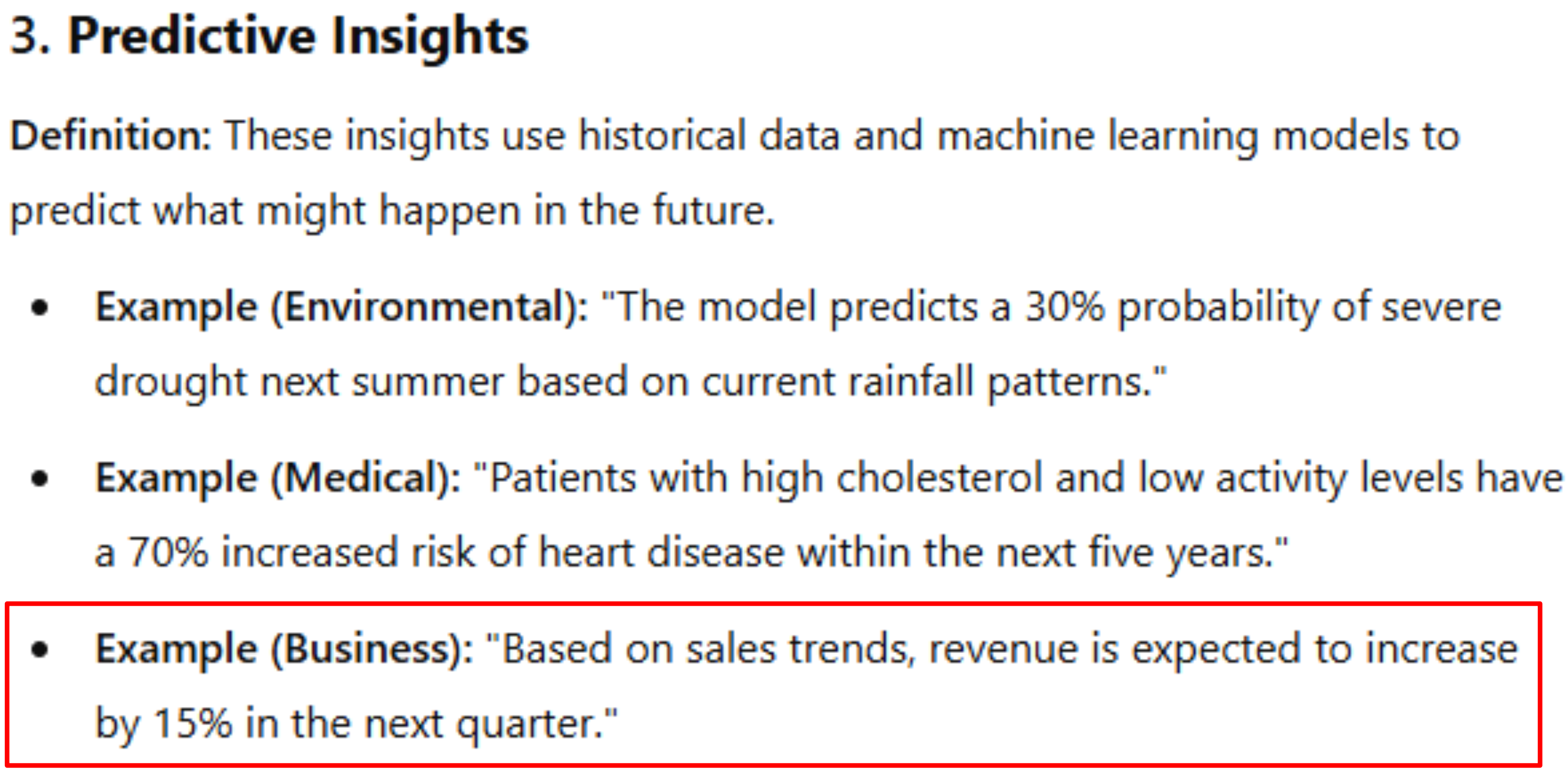
- **Example (Environmental):** "The average air quality index (AQI) in the city has improved by 15% over the last decade due to increased green spaces."
- **Example (Medical):** "The hospital's patient admission rate increased by 20% during flu season compared to the annual average."
- **Example (Business):** "Sales increased by 10% during the holiday season compared to the previous quarter."



2. Diagnostic Insights

Definition: These insights analyze historical data to determine why something happened.

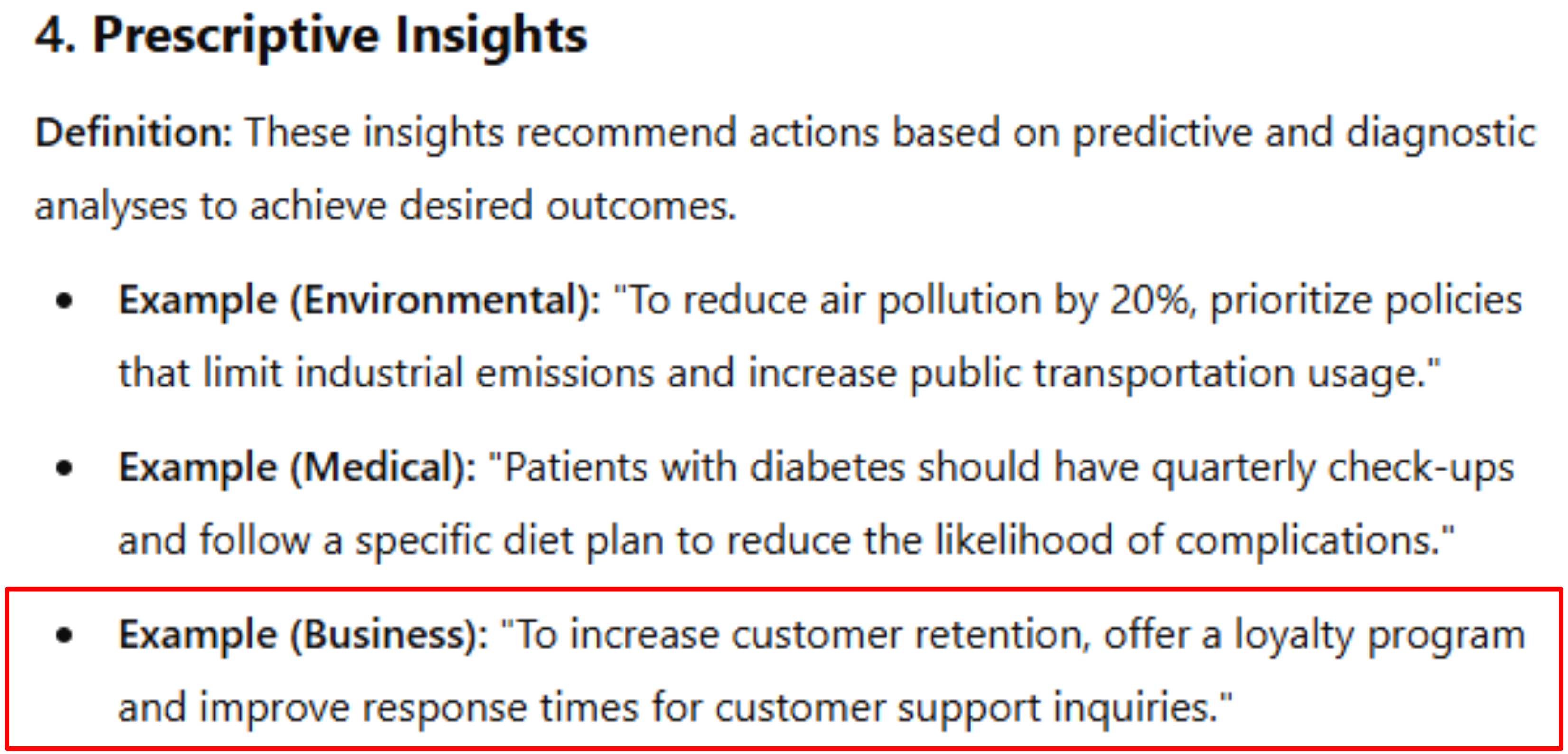
- **Example (Environmental):** "The spike in carbon dioxide levels last quarter was due to a 25% increase in industrial activity."
- **Example (Medical):** "The high readmission rate for patients was linked to incomplete post-discharge instructions provided by nursing staff."
- **Example (Business):** "Customer churn increased because of a delay in shipping during the promotional period."



3. Predictive Insights

Definition: These insights use historical data and machine learning models to predict what might happen in the future.

- **Example (Environmental):** "The model predicts a 30% probability of severe drought next summer based on current rainfall patterns."
- **Example (Medical):** "Patients with high cholesterol and low activity levels have a 70% increased risk of heart disease within the next five years."
- **Example (Business):** "Based on sales trends, revenue is expected to increase by 15% in the next quarter."



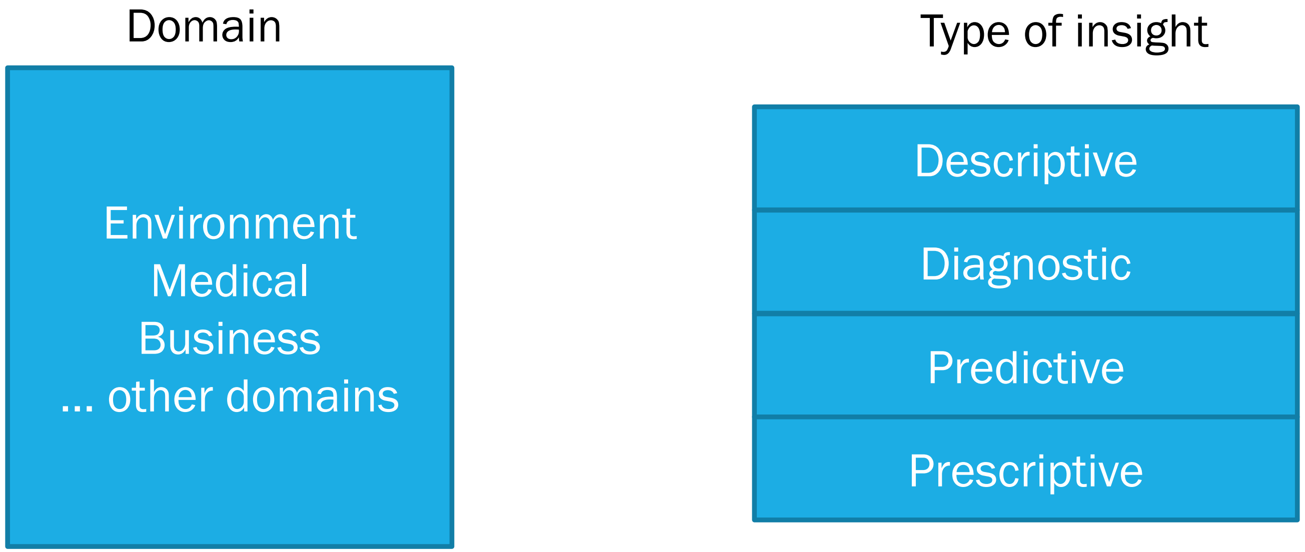
4. Prescriptive Insights

Definition: These insights recommend actions based on predictive and diagnostic analyses to achieve desired outcomes.

- **Example (Environmental):** "To reduce air pollution by 20%, prioritize policies that limit industrial emissions and increase public transportation usage."
- **Example (Medical):** "Patients with diabetes should have quarterly check-ups and follow a specific diet plan to reduce the likelihood of complications."
- **Example (Business):** "To increase customer retention, offer a loyalty program and improve response times for customer support inquiries."

Beyond our role as data scientists

Looking at many examples can be overwhelming… let’s fix one variable.

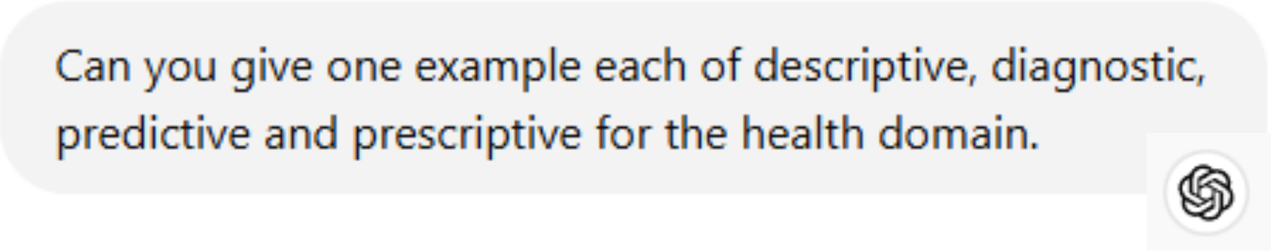


Domain

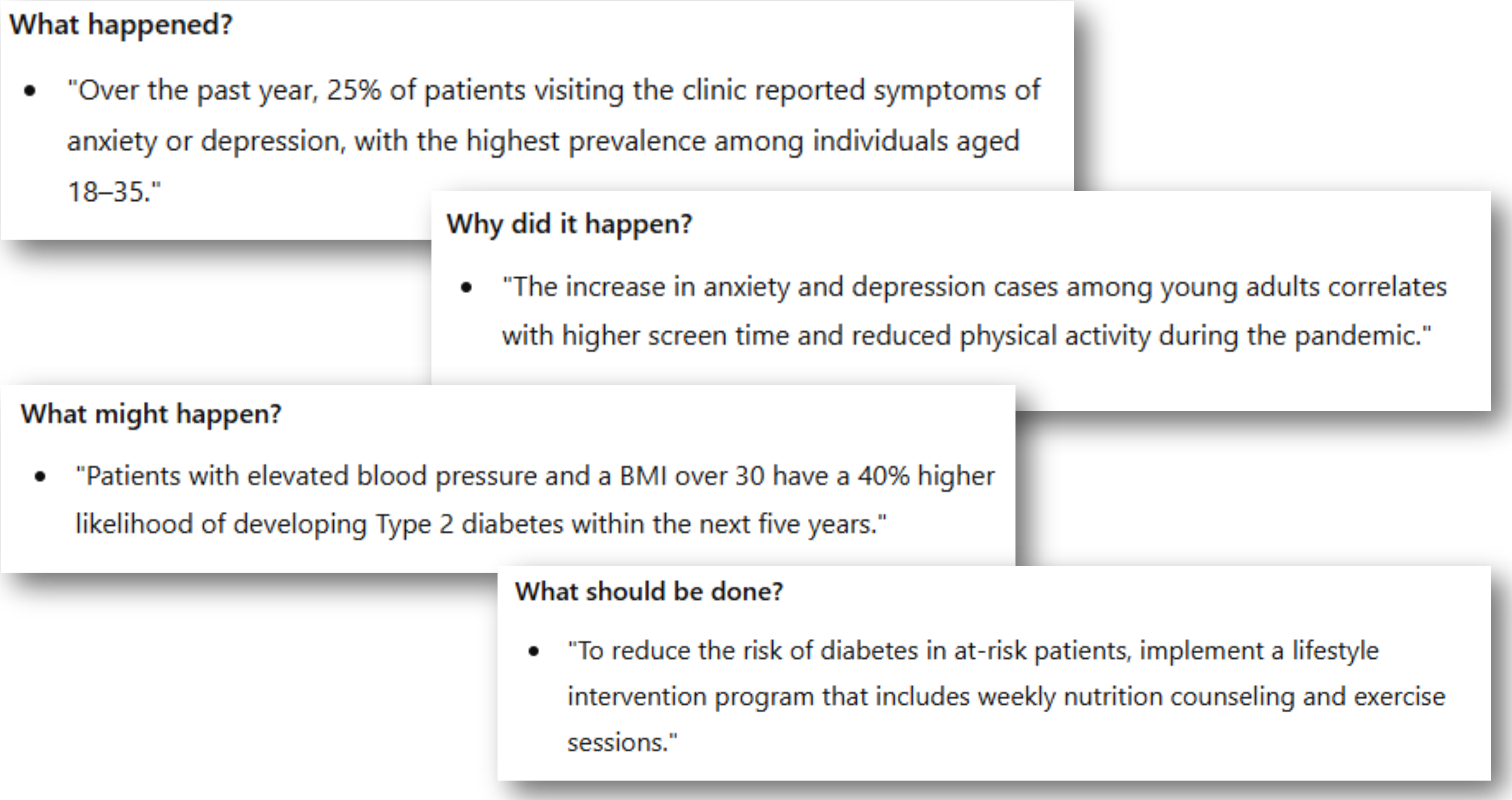
Environment
Medical
Business
… other domains

Type of insight

Descriptive
Diagnostic
Predictive
Prescriptive



Can you give one example each of descriptive, diagnostic, predictive and prescriptive for the health domain.



What happened?

- "Over the past year, 25% of patients visiting the clinic reported symptoms of anxiety or depression, with the highest prevalence among individuals aged 18–35."

Why did it happen?

- "The increase in anxiety and depression cases among young adults correlates with higher screen time and reduced physical activity during the pandemic."

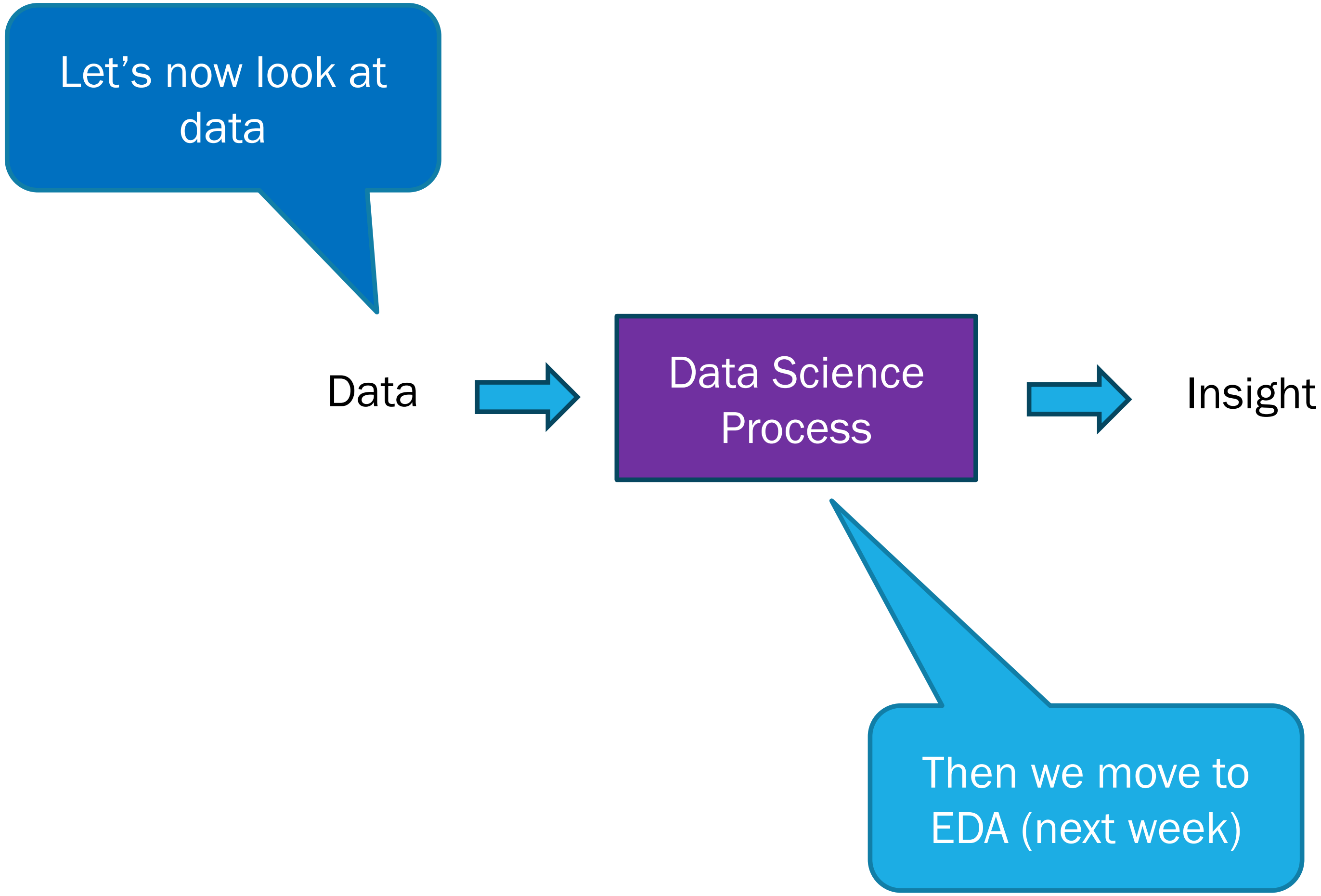
What might happen?

- "Patients with elevated blood pressure and a BMI over 30 have a 40% higher likelihood of developing Type 2 diabetes within the next five years."

What should be done?

- "To reduce the risk of diabetes in at-risk patients, implement a lifestyle intervention program that includes weekly nutrition counseling and exercise sessions."

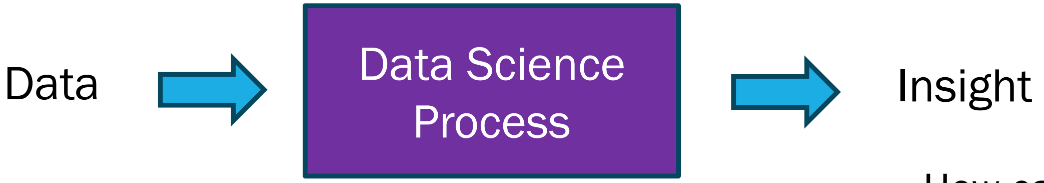
Data



Let’s now look at data

Data → Data Science Process → Insight

Then we move to EDA (next week)



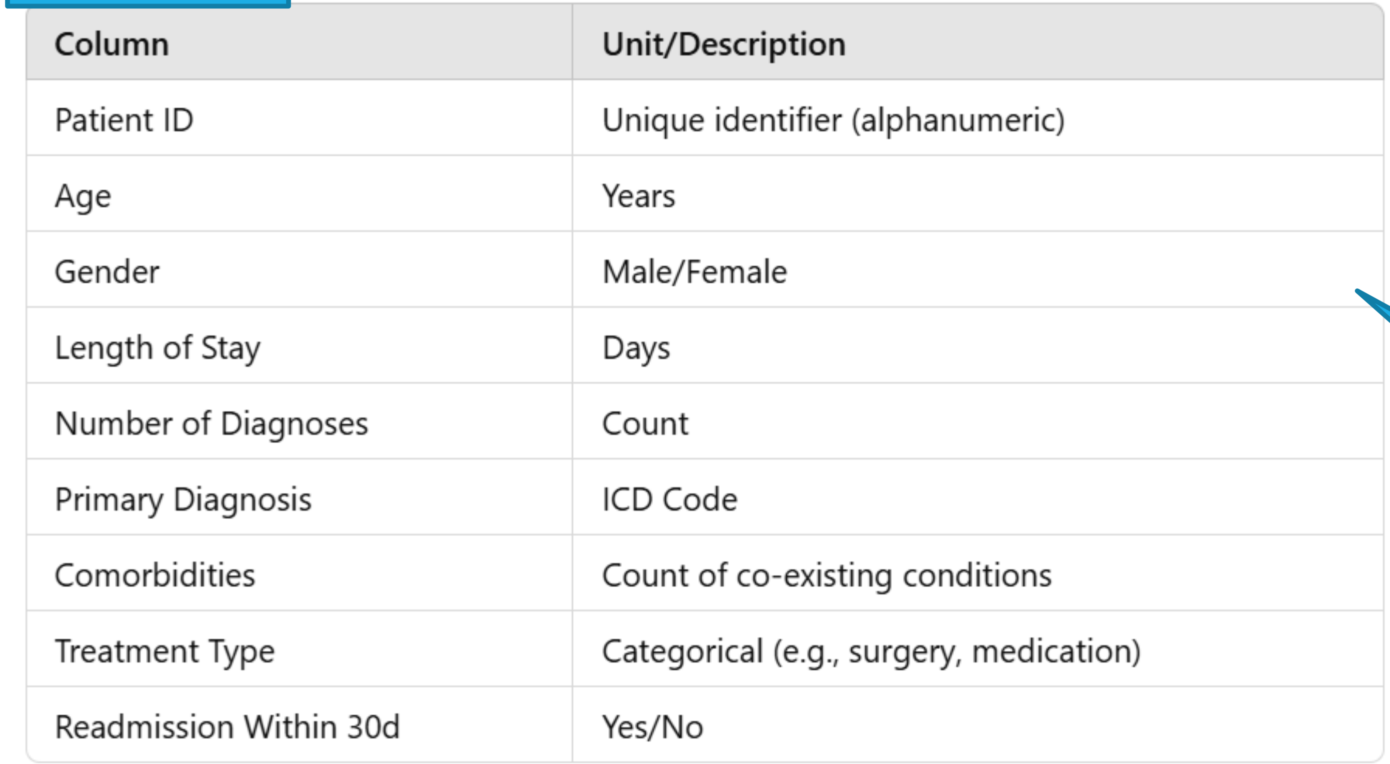
Data

Data Science Process

Insight

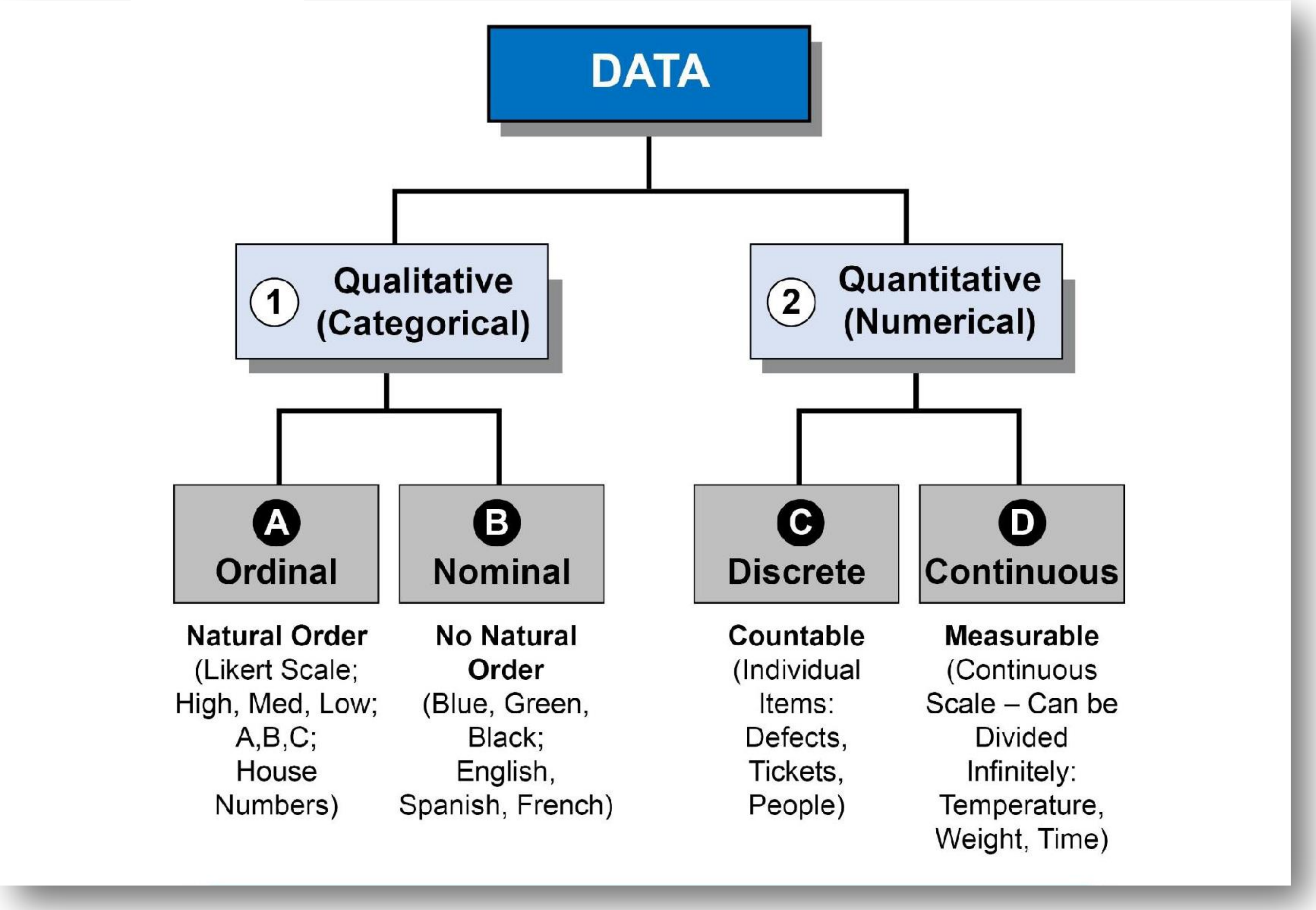
How can we predict the likelihood of hospital readmissions within 30 days based on patient demographics, medical history, and treatment details?

Dataset schema



Column	Unit/Description
Patient ID	Unique identifier (alphanumeric)
Age	Years
Gender	Male/Female
Length of Stay	Days
Number of Diagnoses	Count
Primary Diagnosis	ICD Code
Comorbidities	Count of co-existing conditions
Treatment Type	Categorical (e.g., surgery, medication)
Readmission Within 30d	Yes/No

Can we characterize the data? We will not be able to perform the same analysis on the different kinds of data...



DATA

1 Qualitative (Categorical)

2 Quantitative (Numerical)

A Ordinal

Natural Order (Likert Scale; High, Med, Low; A,B,C; House Numbers)

B Nominal

No Natural Order (Blue, Green, Black; English, Spanish, French)

C Discrete

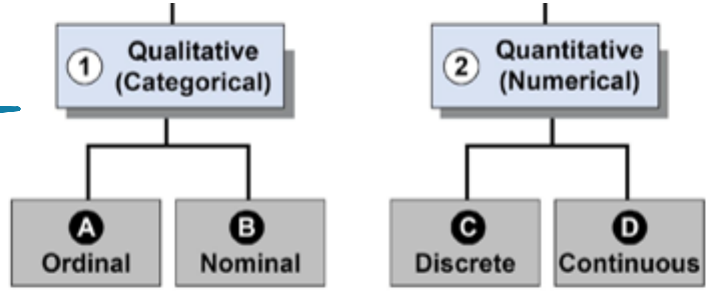
Countable (Individual Items: Defects, Tickets, People)

D Continuous

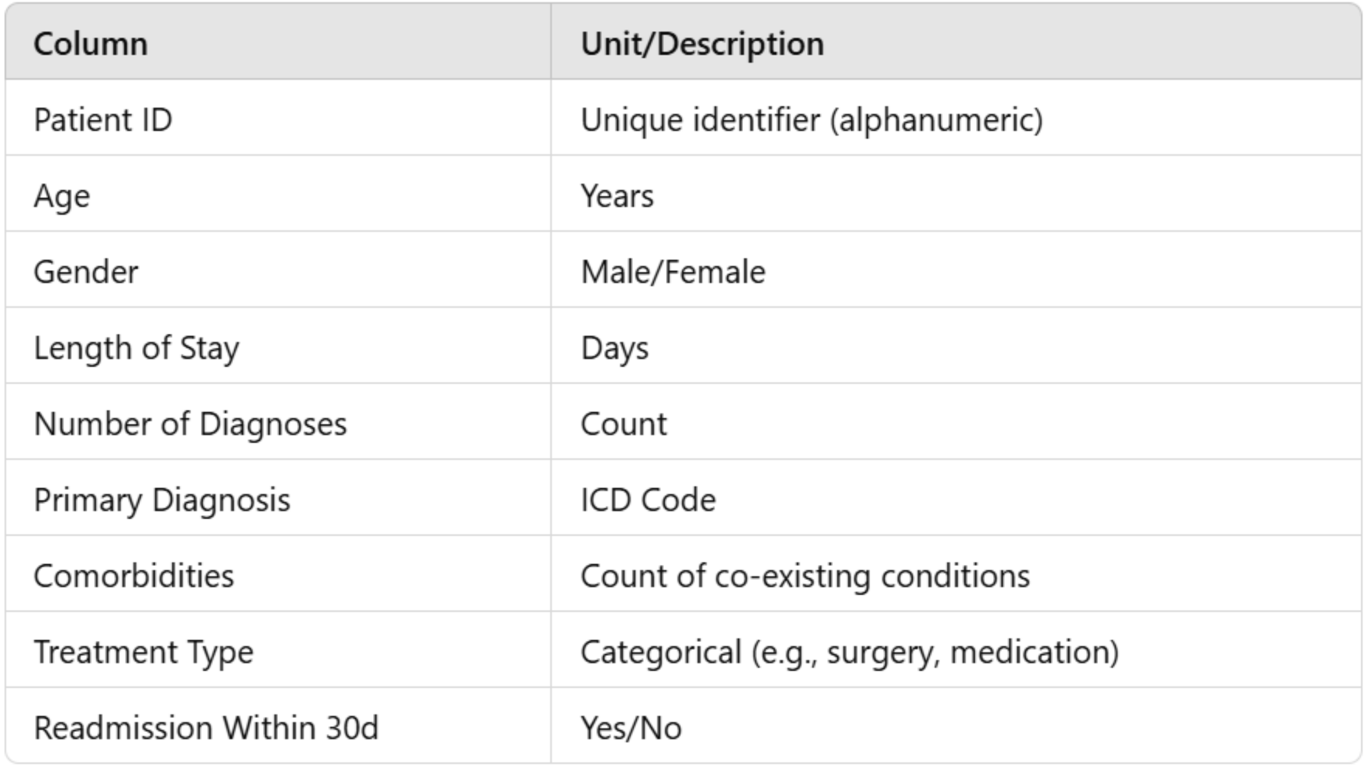
Measurable (Continuous Scale – Can be Divided Infinitely: Temperature, Weight, Time)

Source

For each row, decide on type:

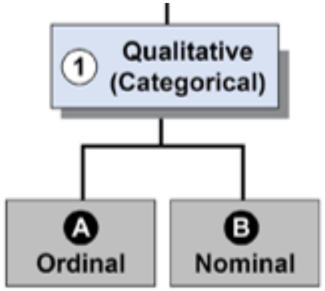


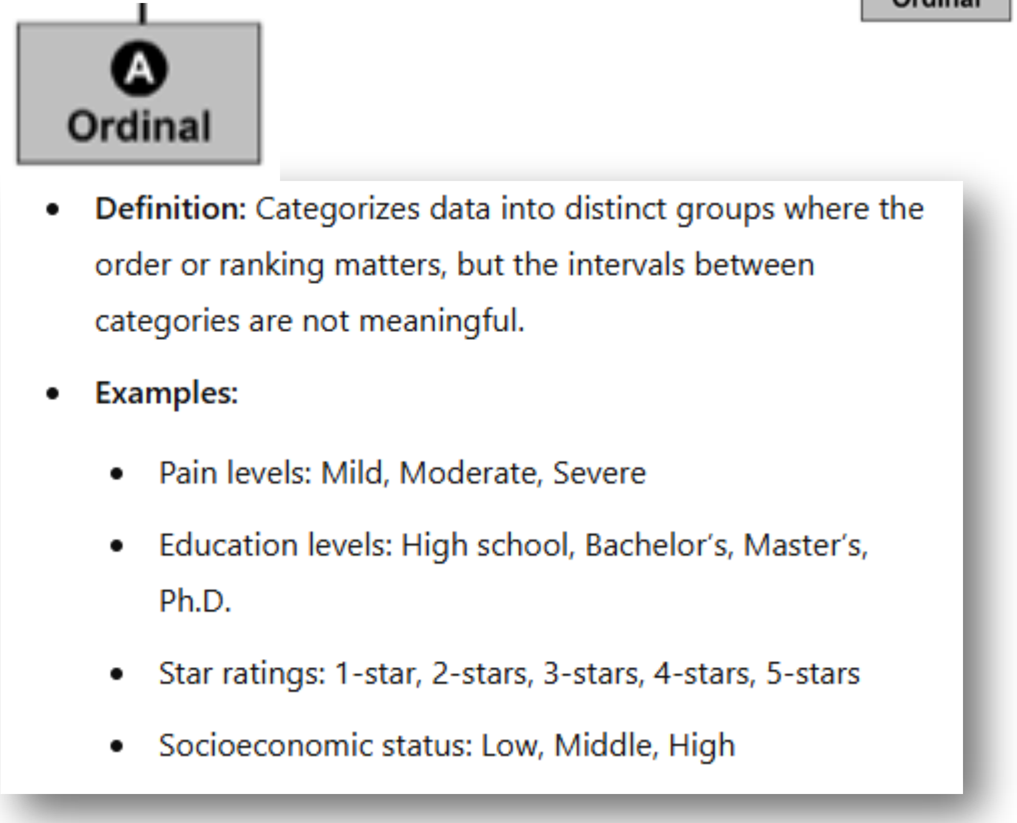
Dataset schema



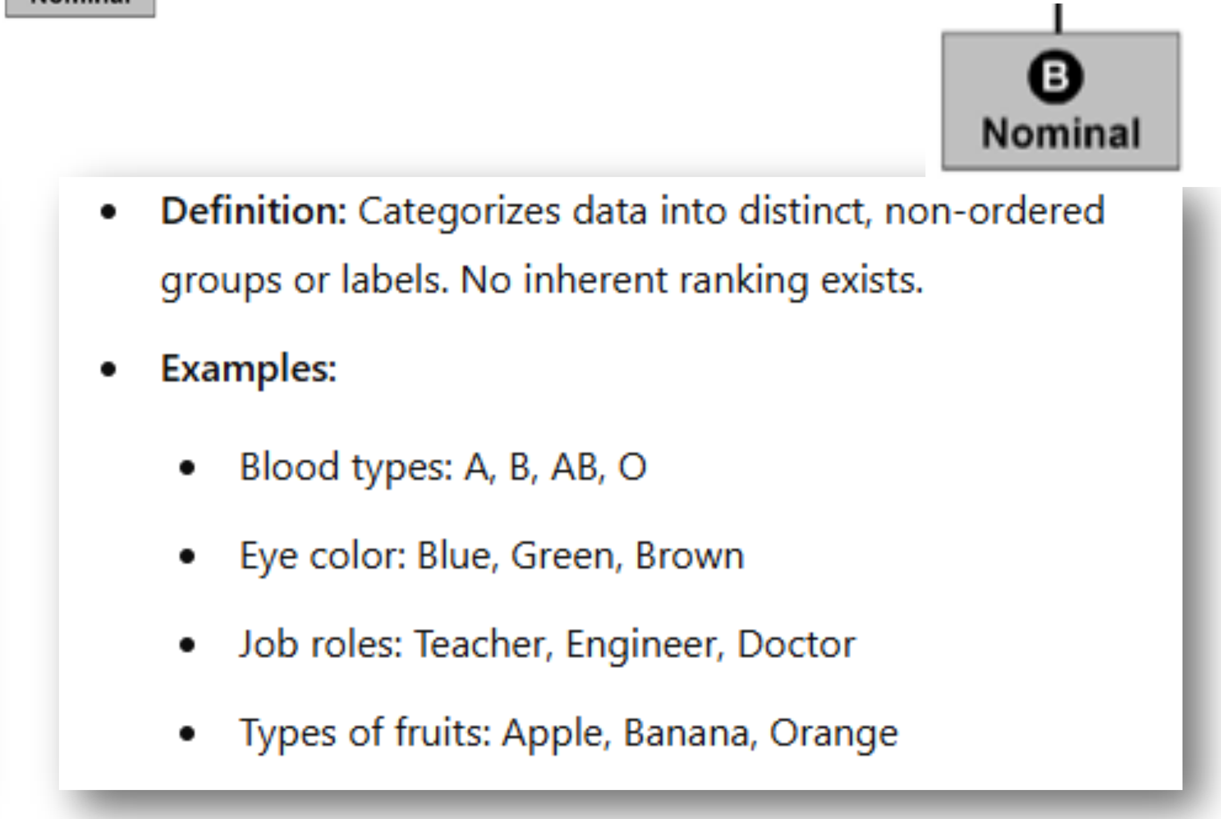
Column	Unit/Description
Patient ID	Unique identifier (alphanumeric)
Age	Years
Gender	Male/Female
Length of Stay	Days
Number of Diagnoses	Count
Primary Diagnosis	ICD Code
Comorbidities	Count of co-existing conditions
Treatment Type	Categorical (e.g., surgery, medication)
Readmission Within 30d	Yes/No

This is a good place to ask ChatGPT for additional examples!

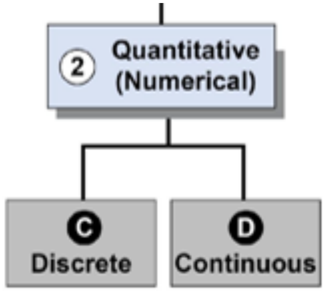


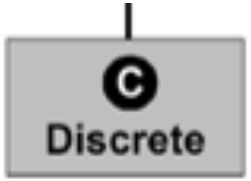


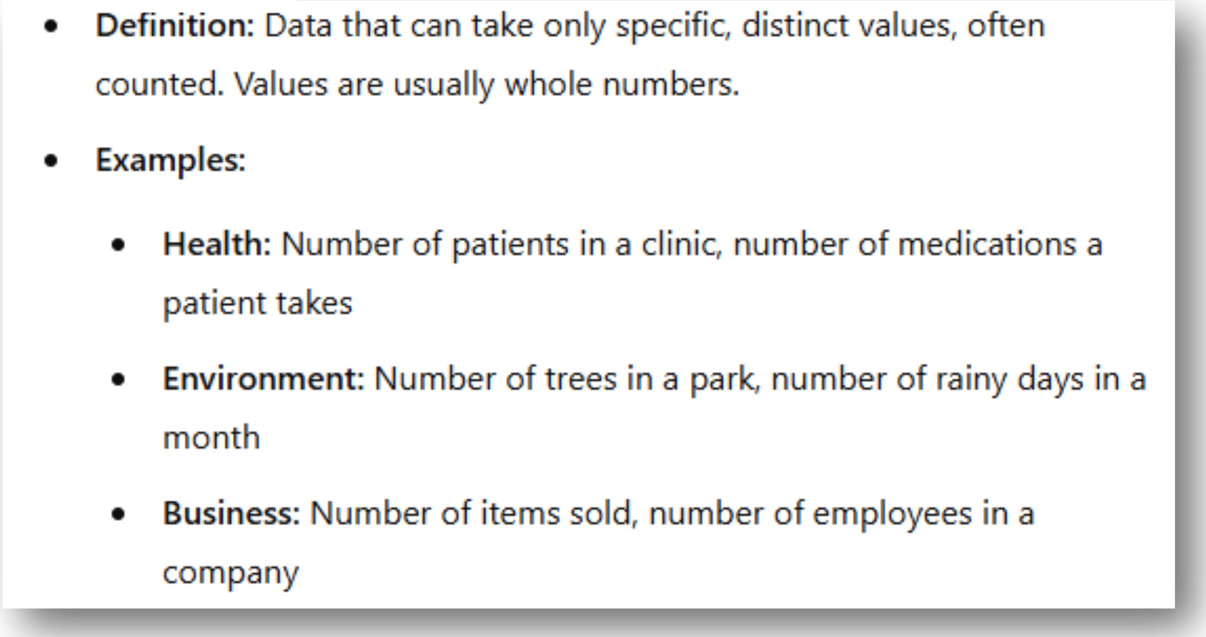
- **Definition:** Categorizes data into distinct groups where the order or ranking matters, but the intervals between categories are not meaningful.
- **Examples:**
  - Pain levels: Mild, Moderate, Severe
  - Education levels: High school, Bachelor's, Master's, Ph.D.
  - Star ratings: 1-star, 2-stars, 3-stars, 4-stars, 5-stars
  - Socioeconomic status: Low, Middle, High



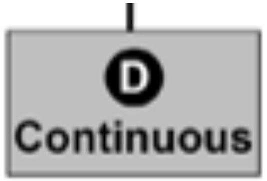
- **Definition:** Categorizes data into distinct, non-ordered groups or labels. No inherent ranking exists.
- **Examples:**
  - Blood types: A, B, AB, O
  - Eye color: Blue, Green, Brown
  - Job roles: Teacher, Engineer, Doctor
  - Types of fruits: Apple, Banana, Orange

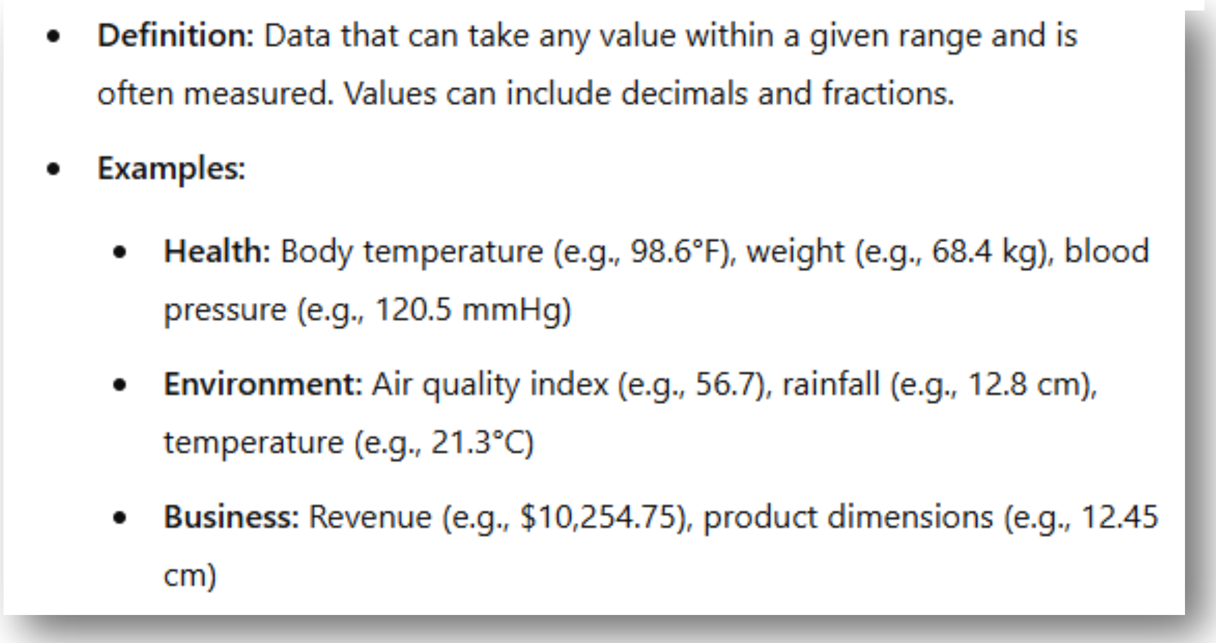




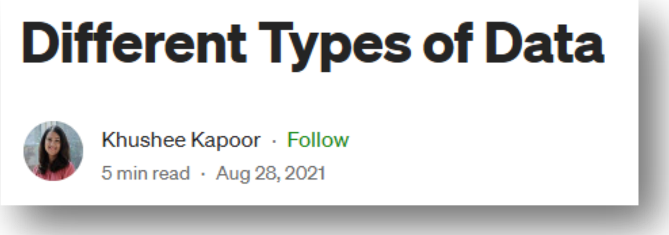


- **Definition:** Data that can take only specific, distinct values, often counted. Values are usually whole numbers.
- **Examples:**
  - **Health:** Number of patients in a clinic, number of medications a patient takes
  - **Environment:** Number of trees in a park, number of rainy days in a month
  - **Business:** Number of items sold, number of employees in a company





- **Definition:** Data that can take any value within a given range and is often measured. Values can include decimals and fractions.
- **Examples:**
  - **Health:** Body temperature (e.g., 98.6°F), weight (e.g., 68.4 kg), blood pressure (e.g., 120.5 mmHg)
  - **Environment:** Air quality index (e.g., 56.7), rainfall (e.g., 12.8 cm), temperature (e.g., 21.3°C)
  - **Business:** Revenue (e.g., $10,254.75), product dimensions (e.g., 12.45 cm)

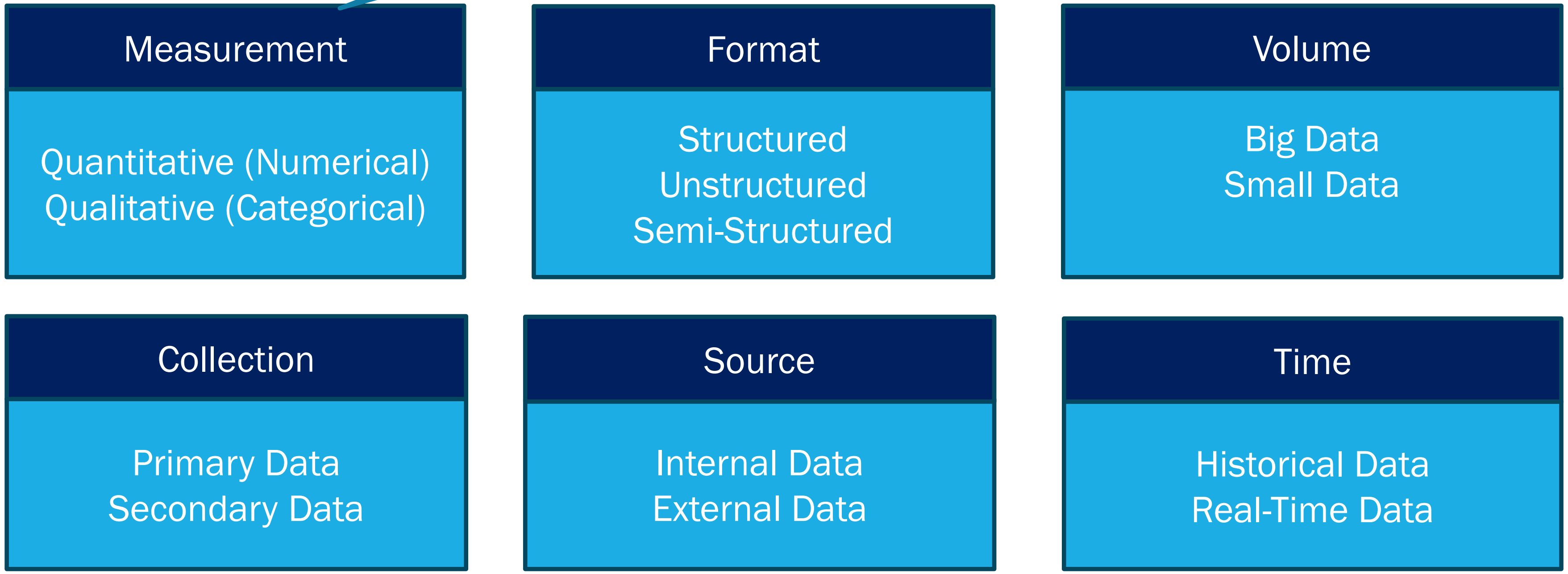


Different Types of Data

Khushee Kapoor · Follow

5 min read · Aug 28, 2021

There are many other characterizations



Measurement

Quantitative (Numerical)
Qualitative (Categorical)

Format

Structured
Unstructured
Semi-Structured

Volume

Big Data
Small Data

Collection

Primary Data
Secondary Data

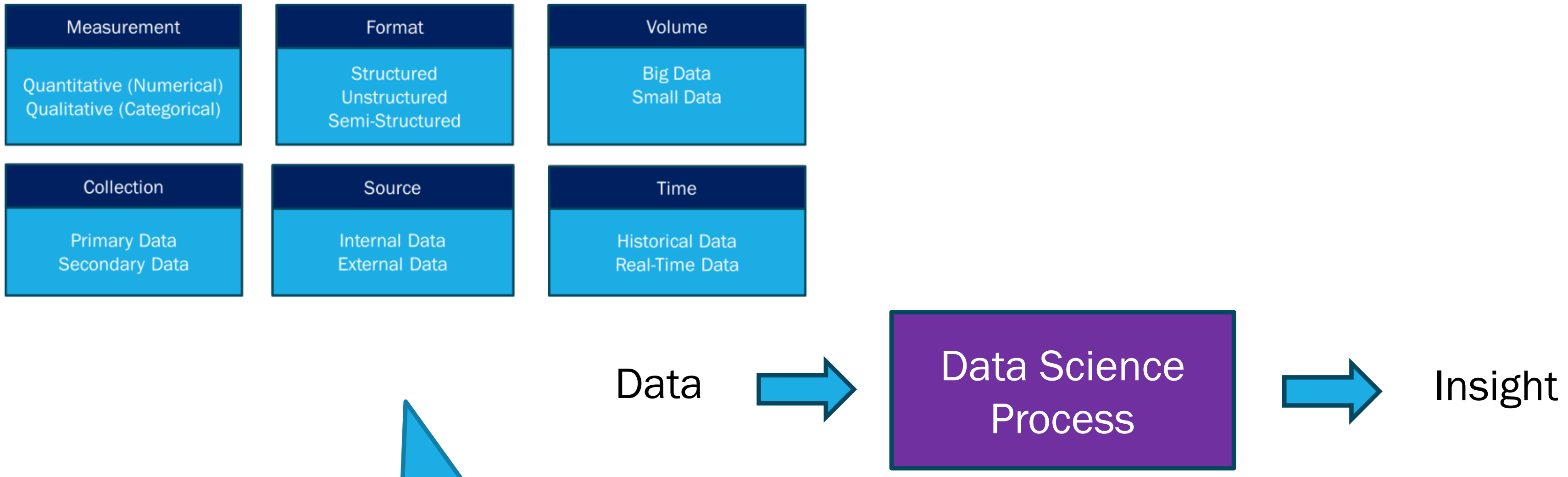
Source

Internal Data
External Data

Time

Historical Data
Real-Time Data

Source



Measurement
Quantitative (Numerical)
Qualitative (Categorical)

Format
Structured
Unstructured
Semi-Structured

Volume
Big Data
Small Data

Collection
Primary Data
Secondary Data

Source
Internal Data
External Data

Time
Historical Data
Real-Time Data

Let's explore these dimensions further + ask ChatGPT for examples in support to the provided insight

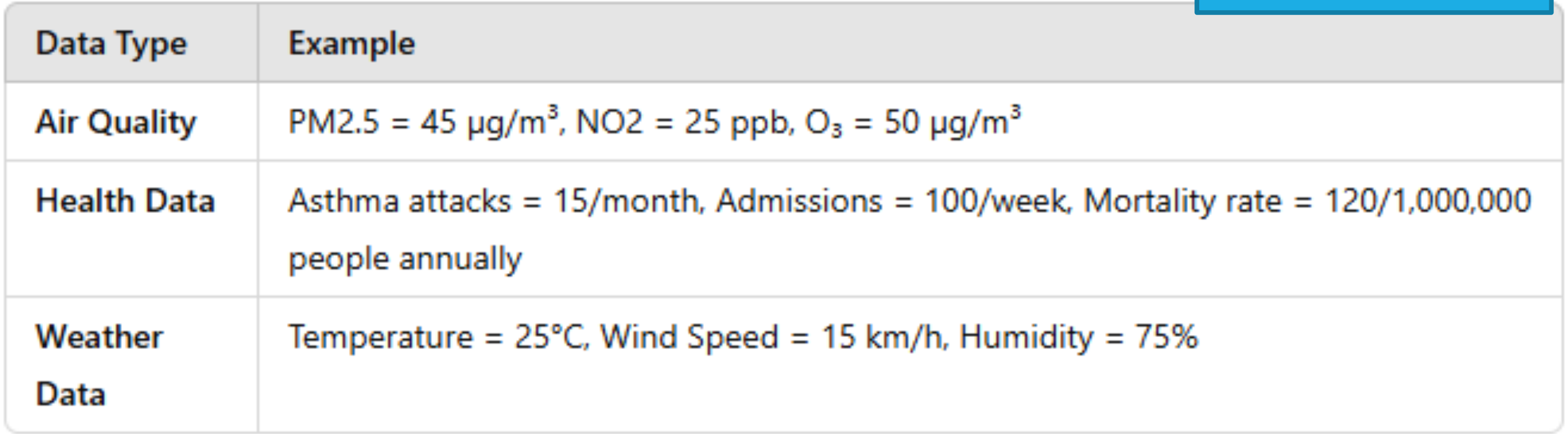
Insight: Predicting the Impact of Air Quality on Health Outcomes in Urban Areas

Predictive analysis

Measurement

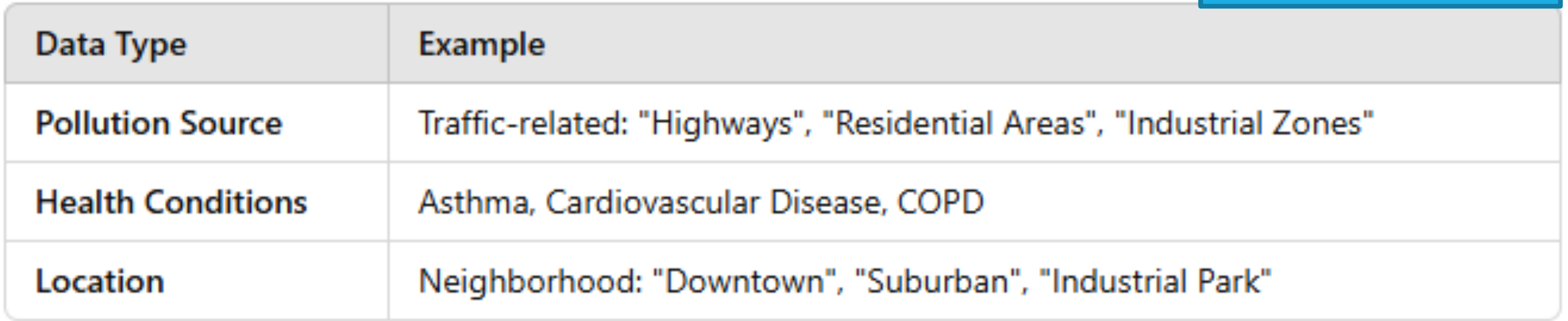
Quantitative (Numerical)
Qualitative (Categorical)

Numerical



Data Type	Example
Air Quality	PM2.5 = 45 $\mu g/m^3$, NO2 = 25 ppb, O_3 = 50 $\mu g/m^3$
Health Data	Asthma attacks = 15/month, Admissions = 100/week, Mortality rate = 120/1,000,000 people annually
Weather Data	Temperature = 25°C, Wind Speed = 15 km/h, Humidity = 75%

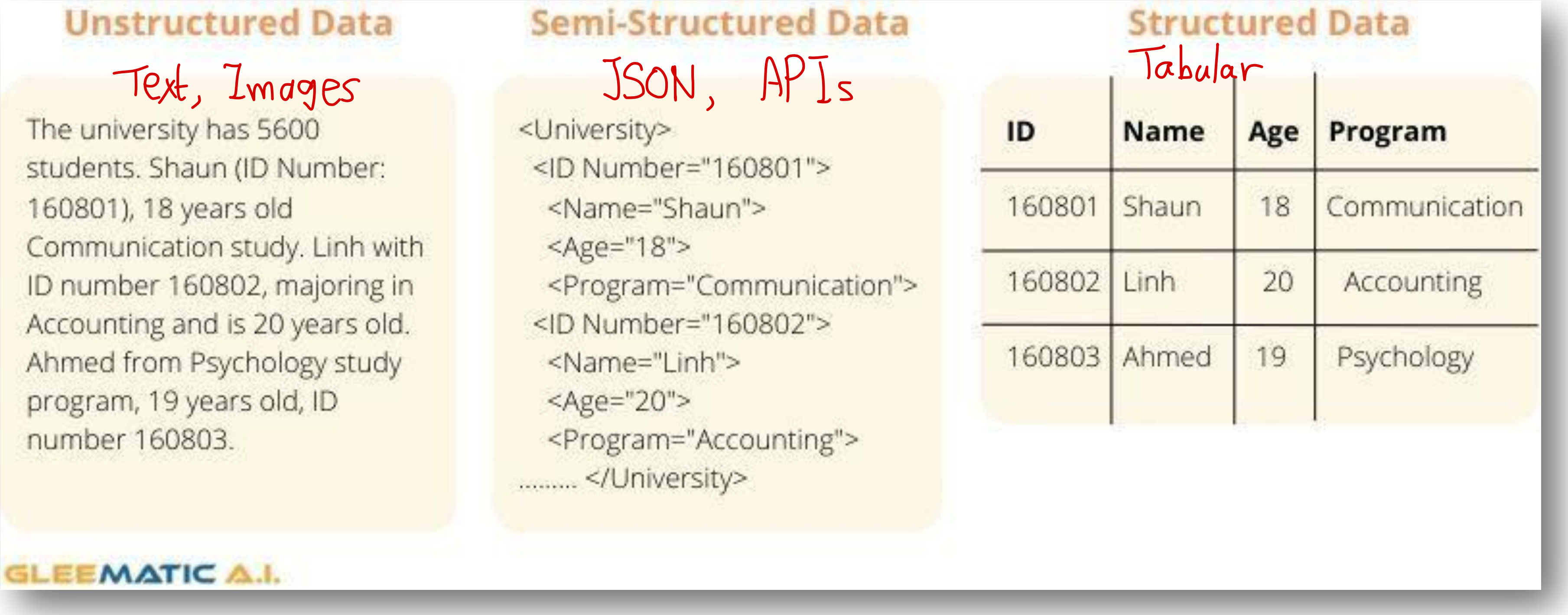
Categorical



Data Type	Example
Pollution Source	Traffic-related: "Highways", "Residential Areas", "Industrial Zones"
Health Conditions	Asthma, Cardiovascular Disease, COPD
Location	Neighborhood: "Downtown", "Suburban", "Industrial Park"

Format

Structured
Unstructured
Semi-Structured



Unstructured Data

Text, Images

The university has 5600 students. Shaun (ID Number: 160801), 18 years old Communication study. Linh with ID number 160802, majoring in Accounting and is 20 years old. Ahmed from Psychology study program, 19 years old, ID number 160803.

Semi-Structured Data

JSON, APIs

```
<University>
 <ID Number="160801">
   <Name="Shaun">
   <Age="18">
   <Program="Communication">
 <ID Number="160802">
   <Name="Linh">
   <Age="20">
   <Program="Accounting">
......... </University>
```

Structured Data

Tabular

ID	Name	Age	Program
160801	Shaun	18	Communication
160802	Linh	20	Accounting
160803	Ahmed	19	Psychology

GLEEMATIC A.I.

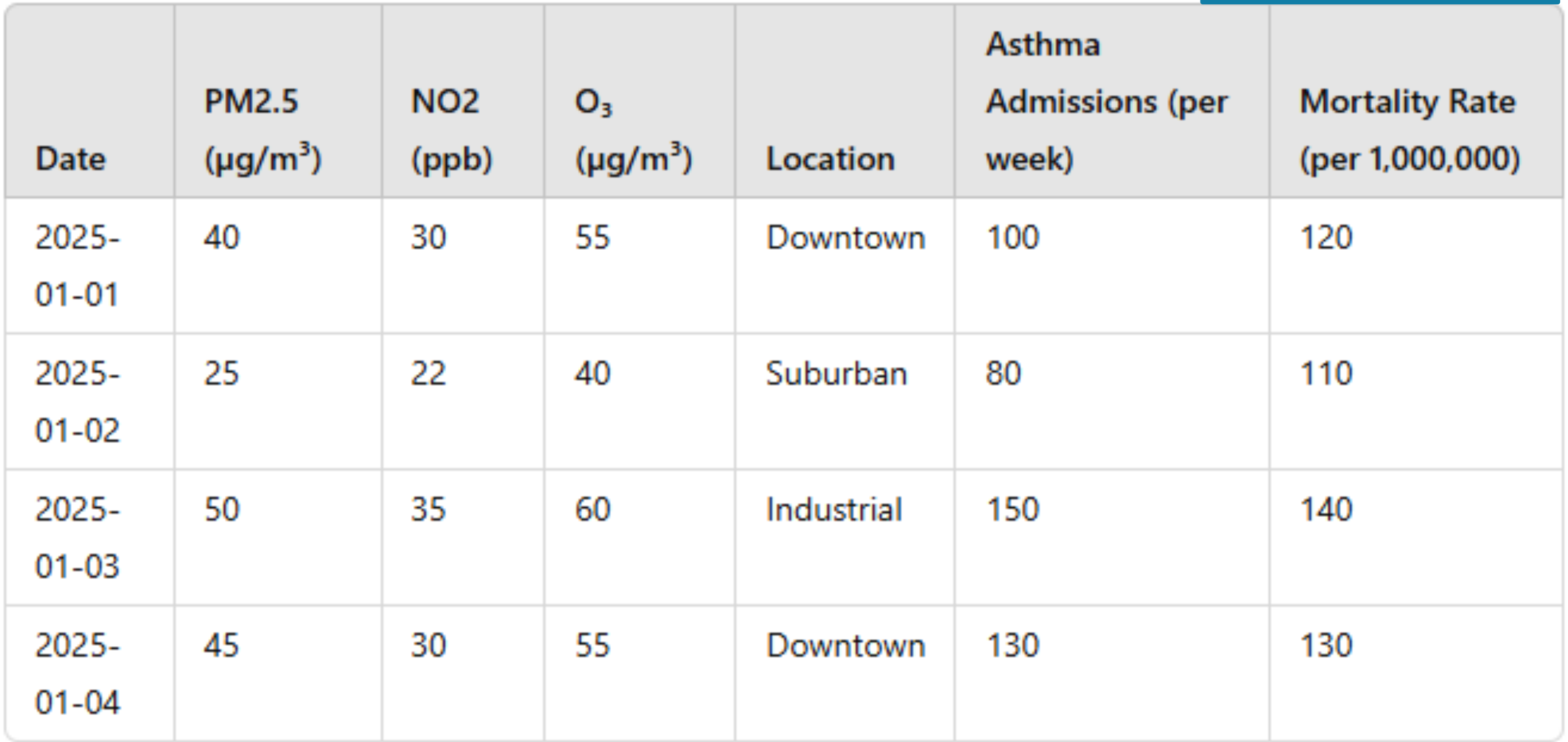
GLEEMATIC A.I.

Source

Format

Structured
Unstructured
Semi-Structured

Structured

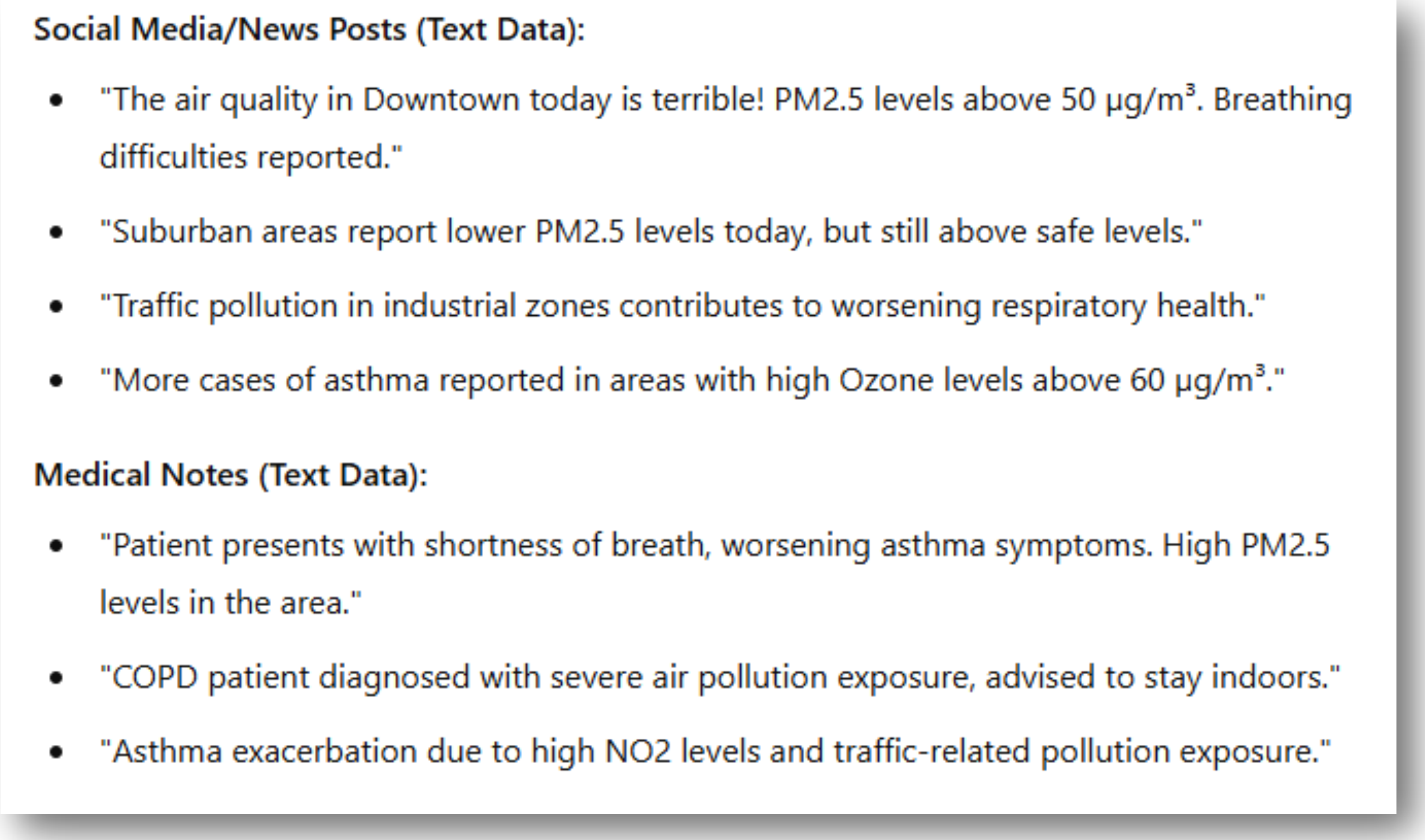


Date	PM2.5 (µg/m³)	NO2 (ppb)	O_3 (µg/m³)	Location	Asthma Admissions (per week)	Mortality Rate (per 1,000,000)
2025-01-01	40	30	55	Downtown	100	120
2025-01-02	25	22	40	Suburban	80	110
2025-01-03	50	35	60	Industrial	150	140
2025-01-04	45	30	55	Downtown	130	130

Format

Structured
Unstructured
Semi-Structured

Unstructured



Social Media/News Posts (Text Data):

- "The air quality in Downtown today is terrible! PM2.5 levels above 50 µg/m^3. Breathing difficulties reported."
- "Suburban areas report lower PM2.5 levels today, but still above safe levels."
- "Traffic pollution in industrial zones contributes to worsening respiratory health."
- "More cases of asthma reported in areas with high Ozone levels above 60 µg/m^3."

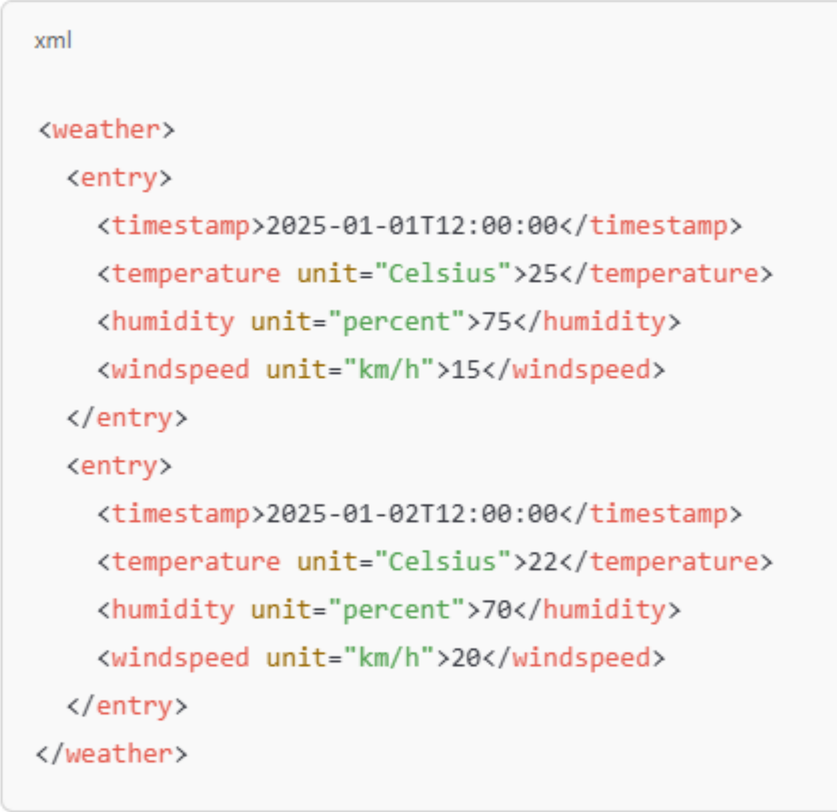
Medical Notes (Text Data):

- "Patient presents with shortness of breath, worsening asthma symptoms. High PM2.5 levels in the area."
- "COPD patient diagnosed with severe air pollution exposure, advised to stay indoors."
- "Asthma exacerbation due to high NO2 levels and traffic-related pollution exposure."

Format

Structured
Unstructured
Semi-Structured

XML (Weather Data):

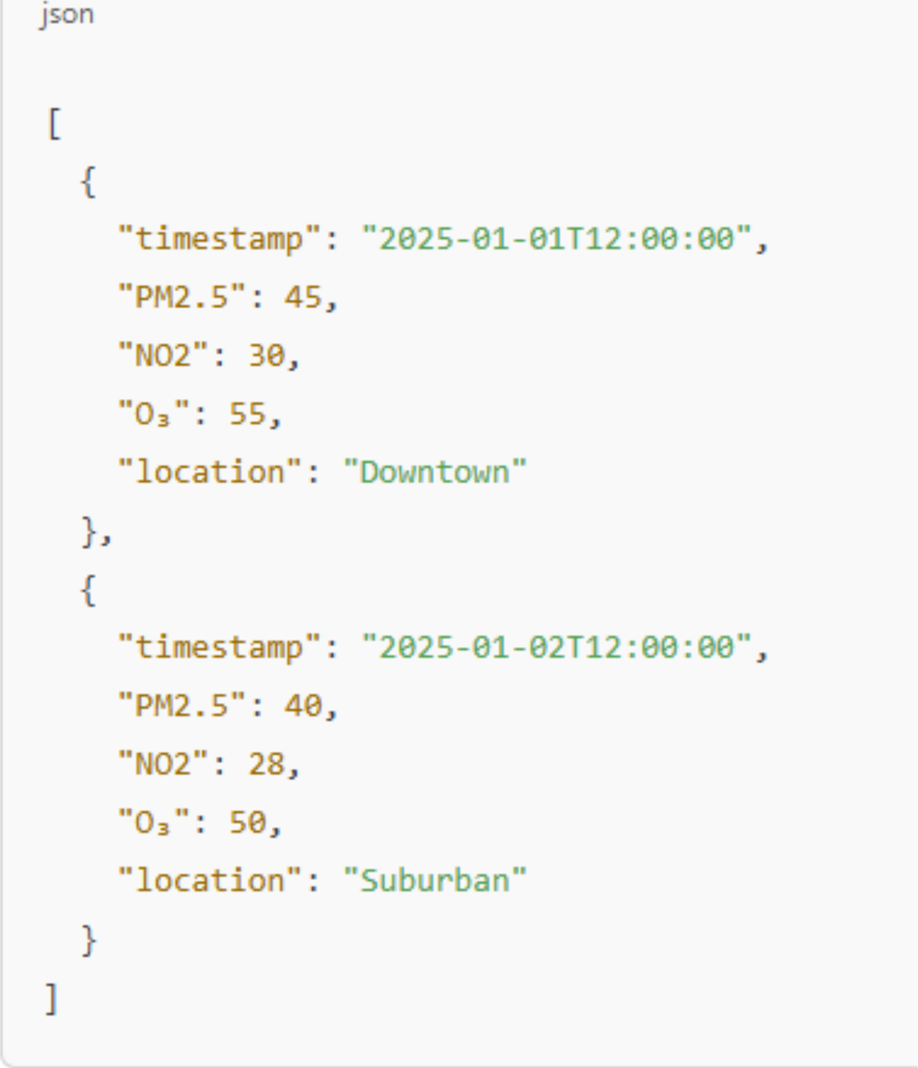


xml

```xml
<weather>
  <entry>
    <timestamp>2025-01-01T12:00:00</timestamp>
    <temperature unit="Celsius">25</temperature>
    <humidity unit="percent">75</humidity>
    <windspeed unit="km/h">15</windspeed>
  </entry>
  <entry>
    <timestamp>2025-01-02T12:00:00</timestamp>
    <temperature unit="Celsius">22</temperature>
    <humidity unit="percent">70</humidity>
    <windspeed unit="km/h">20</windspeed>
  </entry>
</weather>
```

JSON (Air Quality Sensor Data):

Semi-structured



json

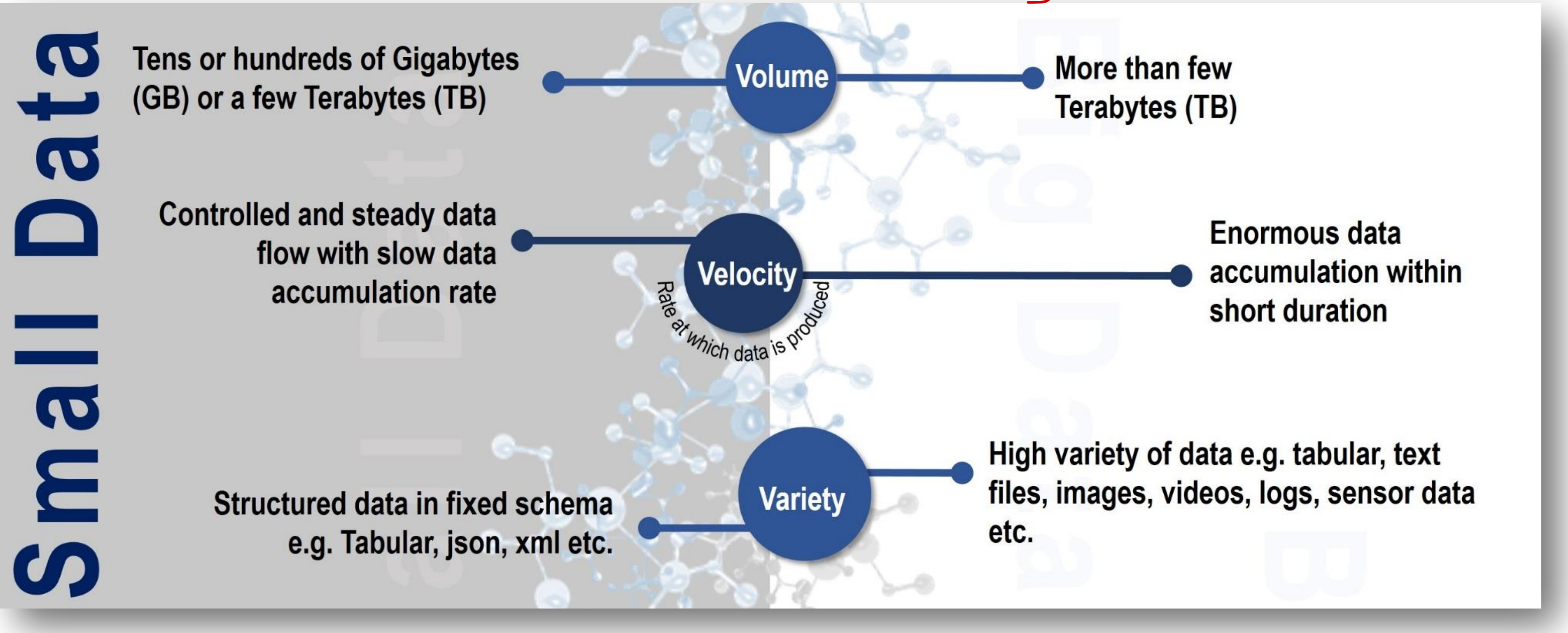
```json
[
  {
    "timestamp": "2025-01-01T12:00:00",
    "PM2.5": 45,
    "NO2": 30,
    "O₃": 55,
    "location": "Downtown"
  },
  {
    "timestamp": "2025-01-02T12:00:00",
    "PM2.5": 40,
    "NO2": 28,
    "O₃": 50,
    "location": "Suburban"
  }
]
```

Volume

Big Data
Small Data

Small

big



Small Data
Big Data
Tens or hundreds of Gigabytes (GB) or a few Terabytes (TB)
Volume
More than few Terabytes (TB)
Controlled and steady data flow with slow data accumulation rate
Velocity
Rate at which data is produced
Enormous data accumulation within short duration
Structured data in fixed schema e.g. Tabular, json, xml etc.
Variety
High variety of data e.g. tabular, text files, images, videos, logs, sensor data etc.

SCI-TECH NEWS

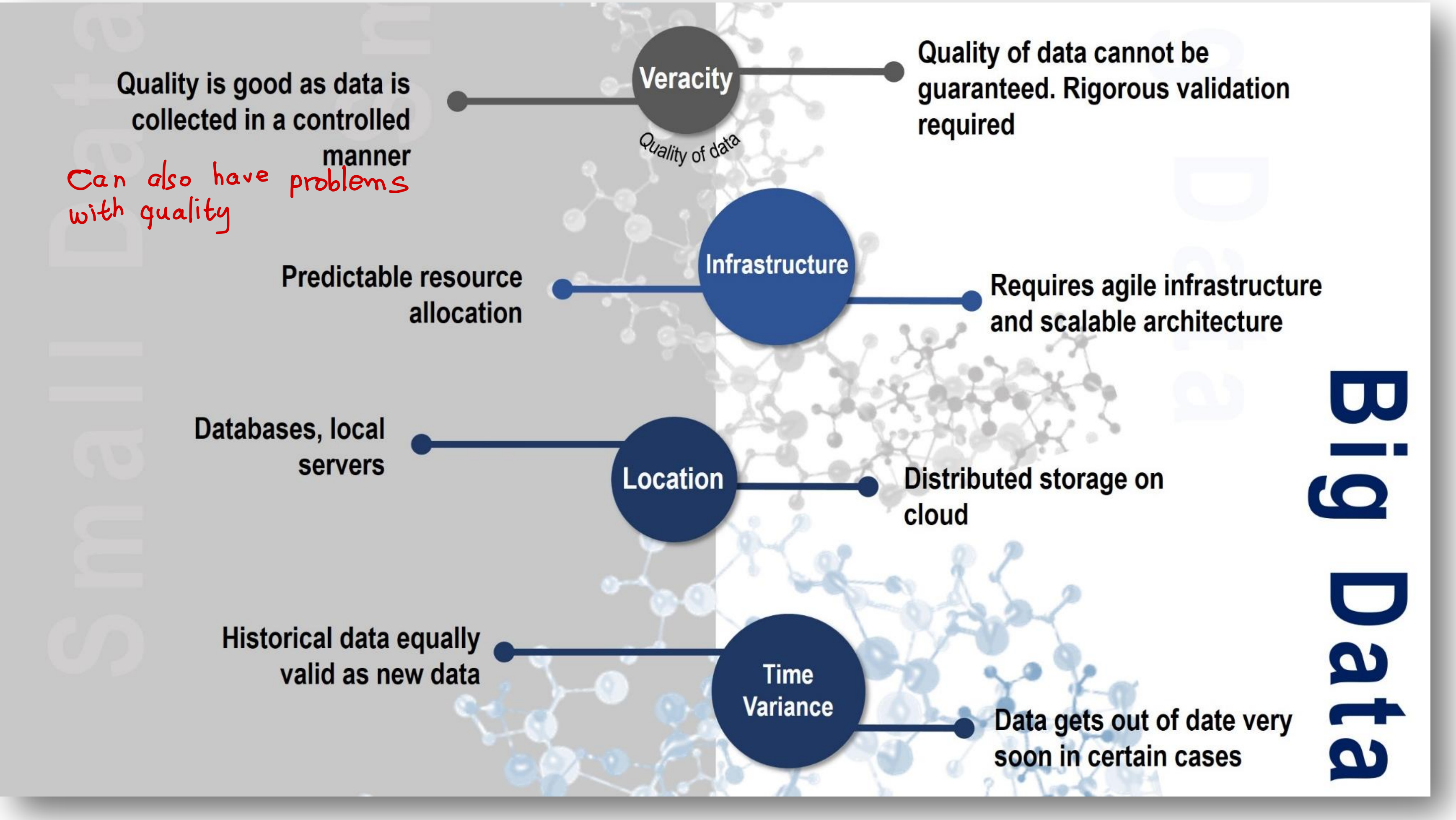
Source

Volume

Big Data
Small Data

small

big



Small Data

Big Data

Veracity
Quality of data

Quality is good as data is collected in a controlled manner

Can also have problems with quality

Quality of data cannot be guaranteed. Rigorous validation required

Infrastructure

Predictable resource allocation

Requires agile infrastructure and scalable architecture

Location

Databases, local servers

Distributed storage on cloud

Time Variance

Historical data equally valid as new data

Data gets out of date very soon in certain cases

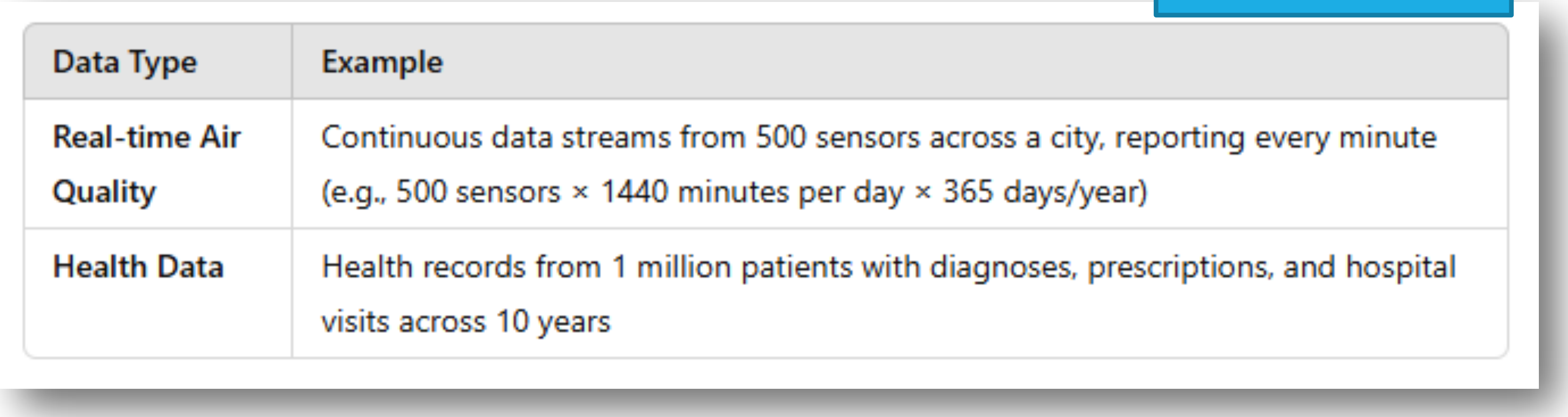
SCI-TECH NEWS

Source

Volume

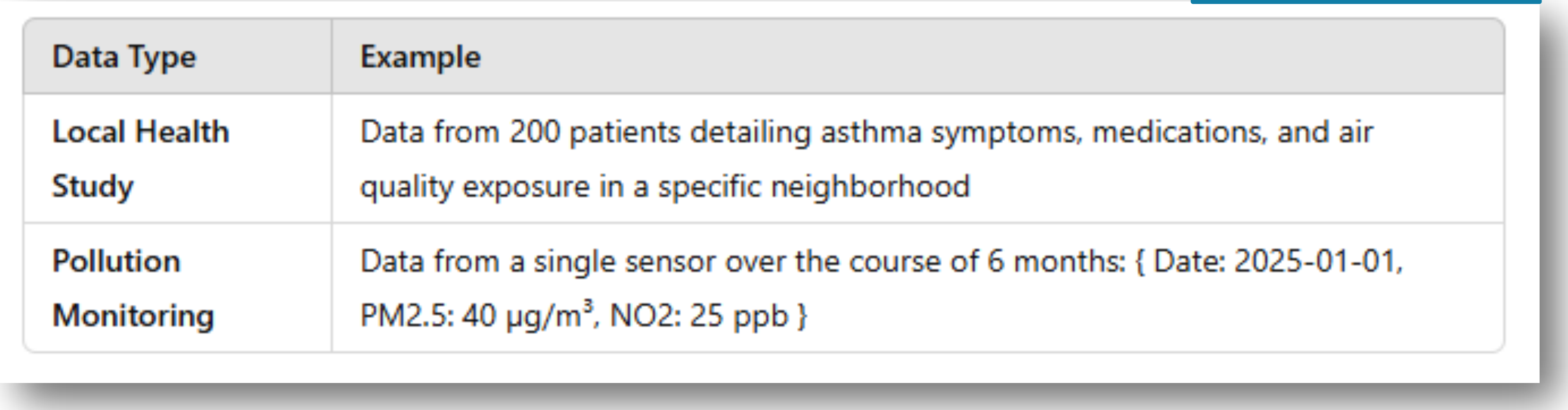
Big Data
Small Data

Big data



Data Type	Example
Real-time Air Quality	Continuous data streams from 500 sensors across a city, reporting every minute (e.g., 500 sensors × 1440 minutes per day × 365 days/year)
Health Data	Health records from 1 million patients with diagnoses, prescriptions, and hospital visits across 10 years

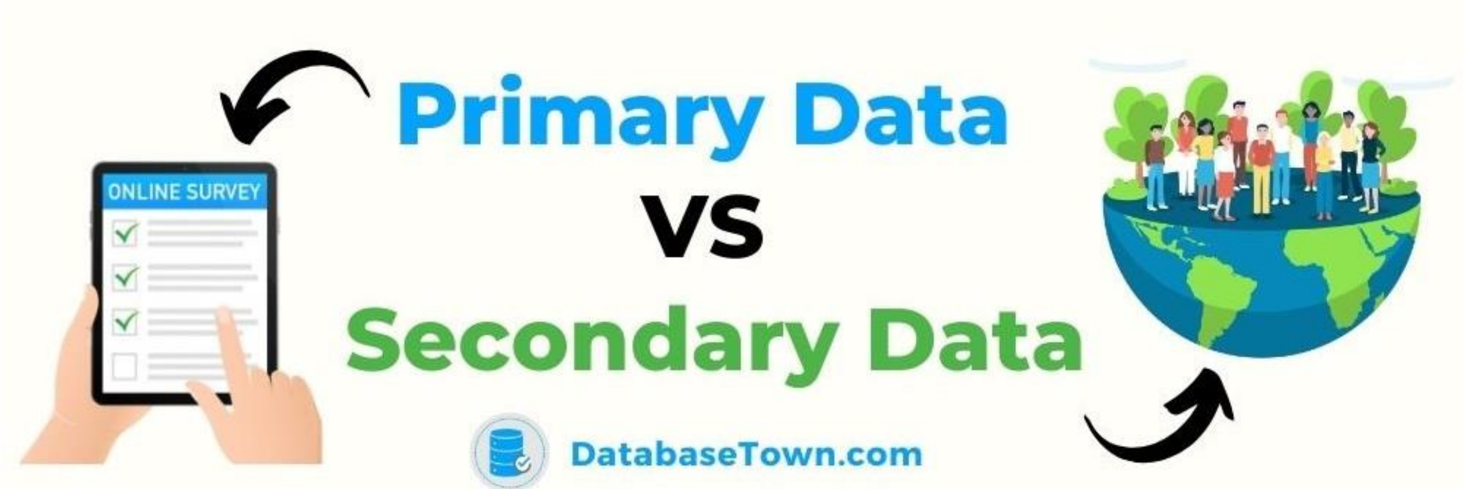
Small data

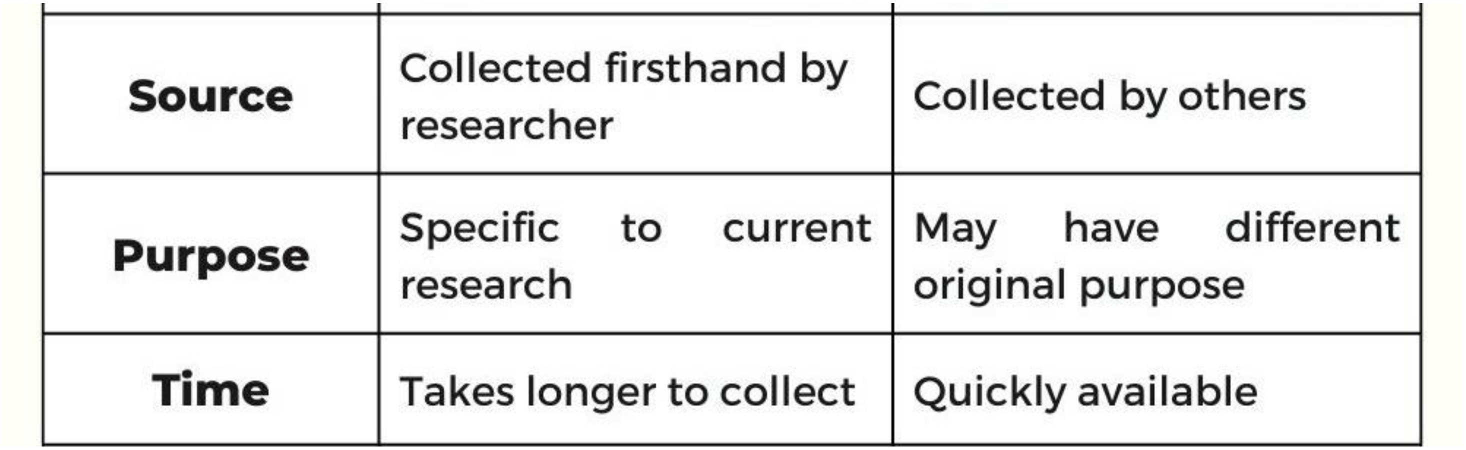


Data Type	Example
Local Health Study	Data from 200 patients detailing asthma symptoms, medications, and air quality exposure in a specific neighborhood
Pollution Monitoring	Data from a single sensor over the course of 6 months: { Date: 2025-01-01, PM2.5: 40 µg/m^3, NO2: 25 ppb }

Collection

Primary Data
Secondary Data





	Primary Data	Secondary Data
Source	Collected firsthand by researcher	Collected by others
Purpose	Specific to current research	May have different original purpose
Time	Takes longer to collect	Quickly available

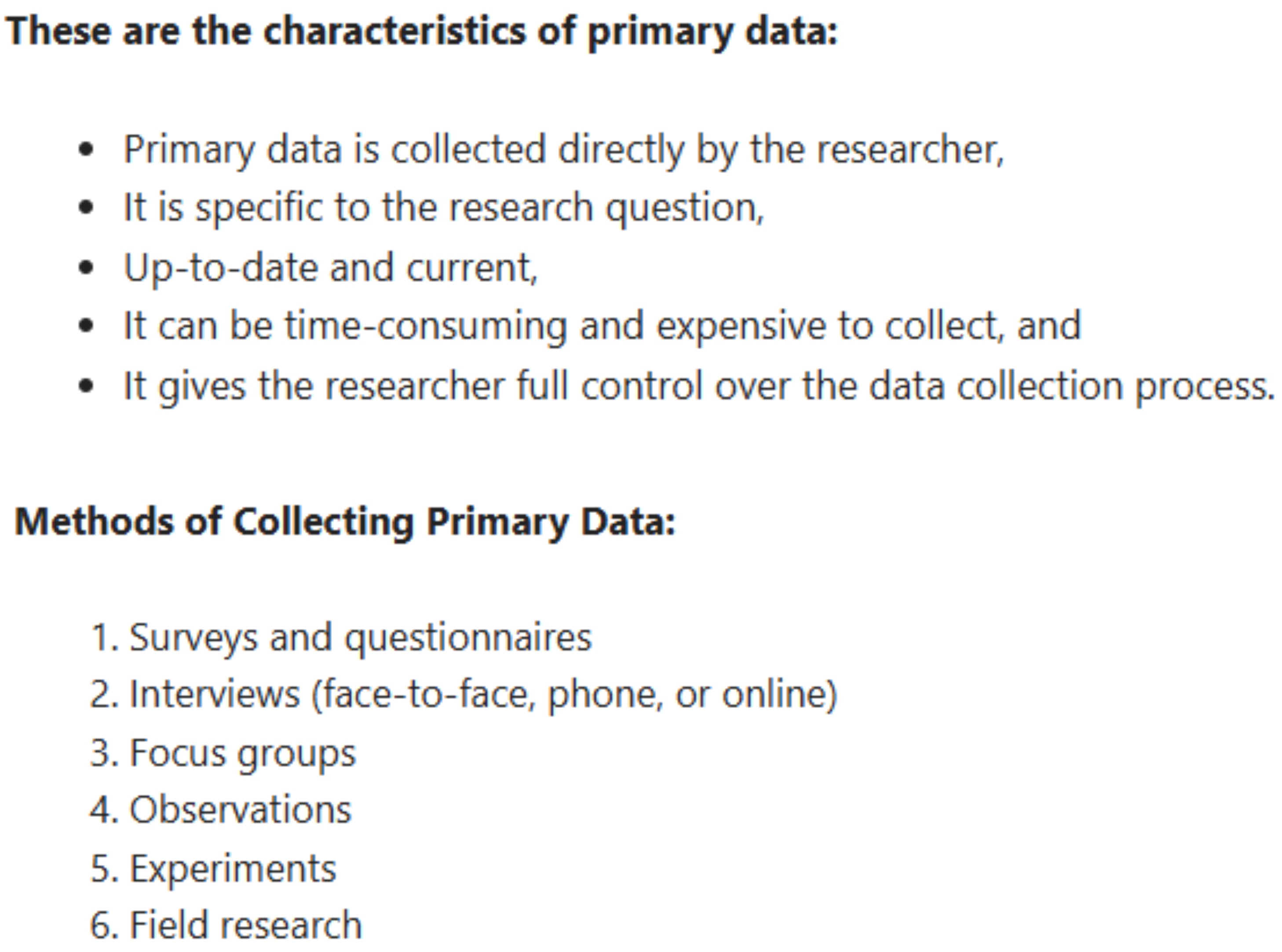
collected within company

collected elsewhere

Source

Collection

Primary Data
Secondary Data

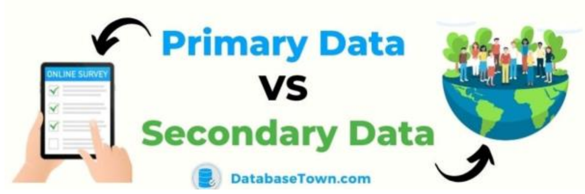


These are the characteristics of primary data:

- Primary data is collected directly by the researcher,
- It is specific to the research question,
- Up-to-date and current,
- It can be time-consuming and expensive to collect, and
- It gives the researcher full control over the data collection process.

Methods of Collecting Primary Data:

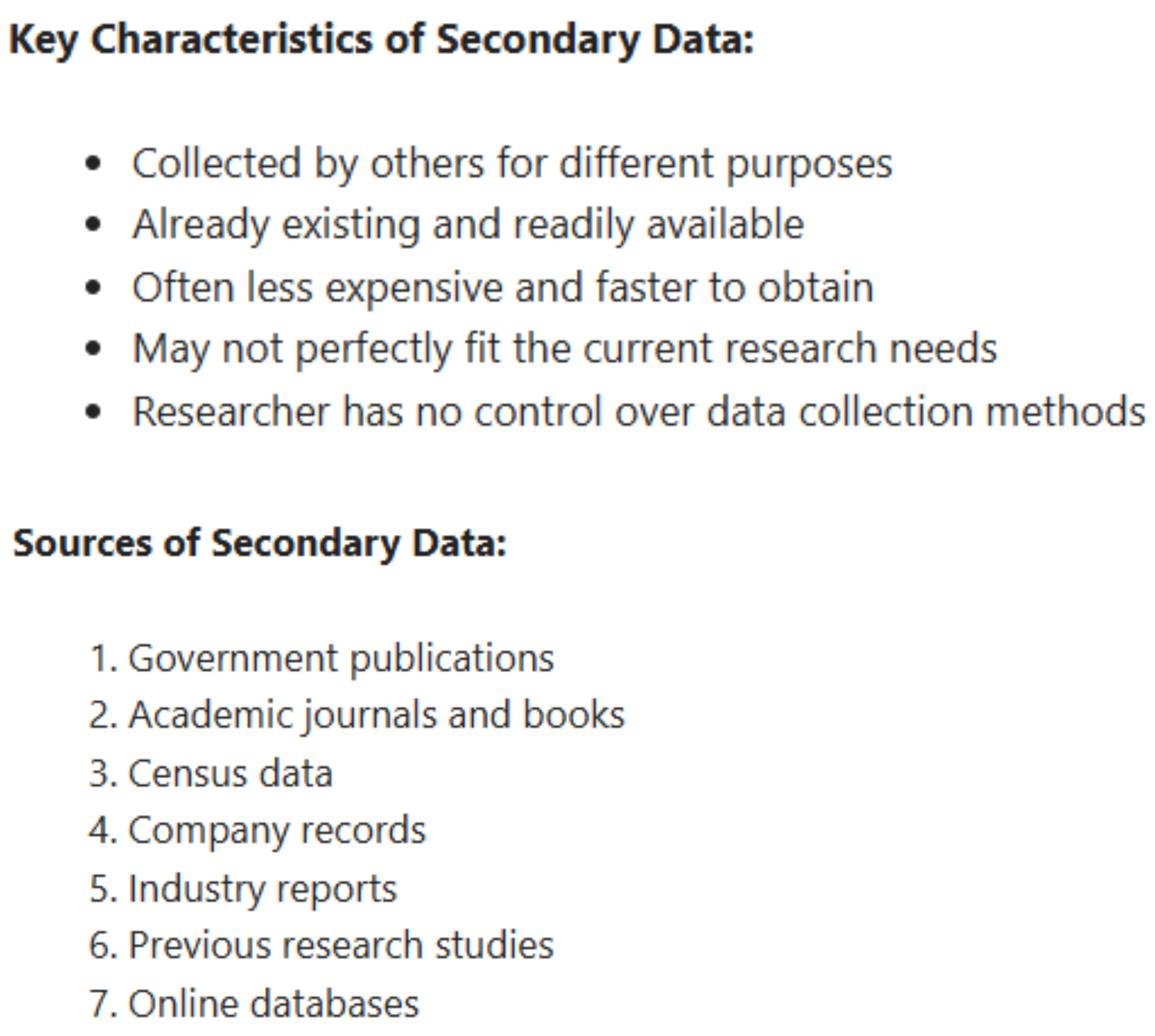
1. Surveys and questionnaires
2. Interviews (face-to-face, phone, or online)
3. Focus groups
4. Observations
5. Experiments
6. Field research



Source

Collection

Primary Data
Secondary Data



Key Characteristics of Secondary Data:

- Collected by others for different purposes
- Already existing and readily available
- Often less expensive and faster to obtain
- May not perfectly fit the current research needs
- Researcher has no control over data collection methods

Sources of Secondary Data:

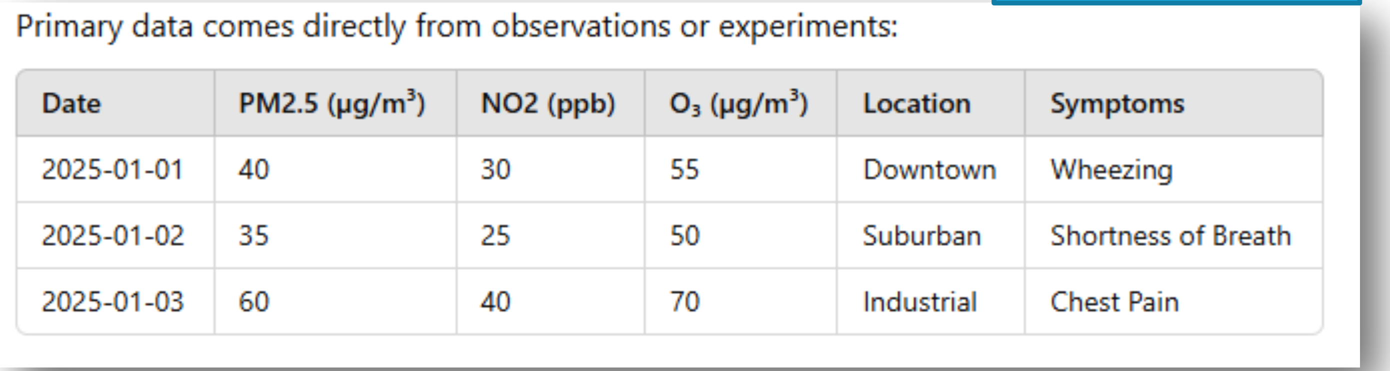
1. Government publications
2. Academic journals and books
3. Census data
4. Company records
5. Industry reports
6. Previous research studies
7. Online databases

Source

Collection

Primary Data
Secondary Data

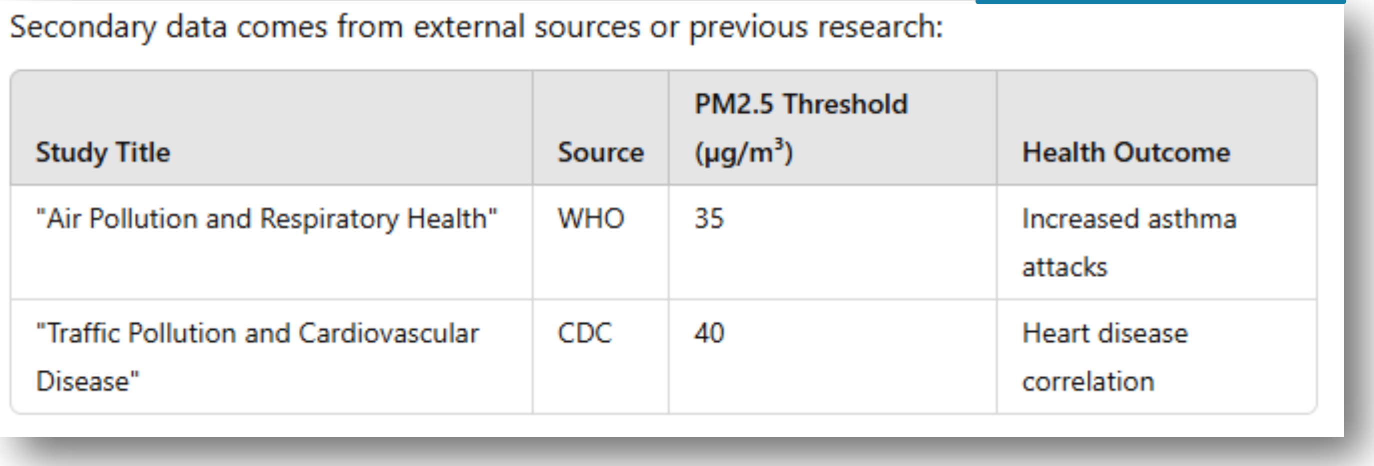
Primary



Primary data comes directly from observations or experiments:

Date	PM2.5 ($\mu g/m^3$)	NO2 (ppb)	O_3 ($\mu g/m^3$)	Location	Symptoms
2025-01-01	40	30	55	Downtown	Wheezing
2025-01-02	35	25	50	Suburban	Shortness of Breath
2025-01-03	60	40	70	Industrial	Chest Pain

Secondary

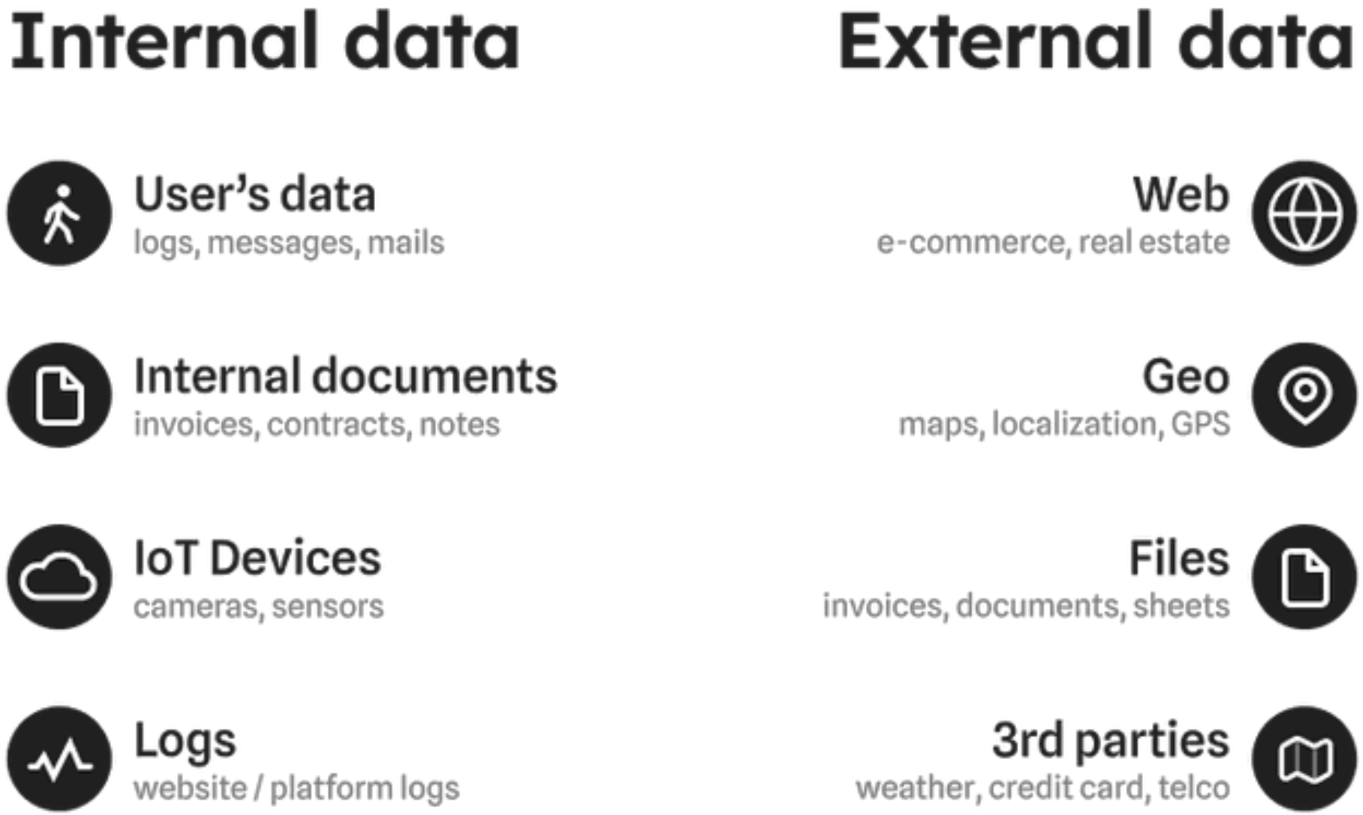


Secondary data comes from external sources or previous research:

Study Title	Source	PM2.5 Threshold ($\mu g/m^3$)	Health Outcome
"Air Pollution and Respiratory Health"	WHO	35	Increased asthma attacks
"Traffic Pollution and Cardiovascular Disease"	CDC	40	Heart disease correlation

Source

Internal Data
External Data



Internal data

User's data
logs, messages, mails

Internal documents
invoices, contracts, notes

IoT Devices
cameras, sensors

Logs
website / platform logs

External data

Web
e-commerce, real estate

Geo
maps, localization, GPS

Files
invoices, documents, sheets

3rd parties
weather, credit card, telco

Internal and External Data: What's the Difference and Why It Matters

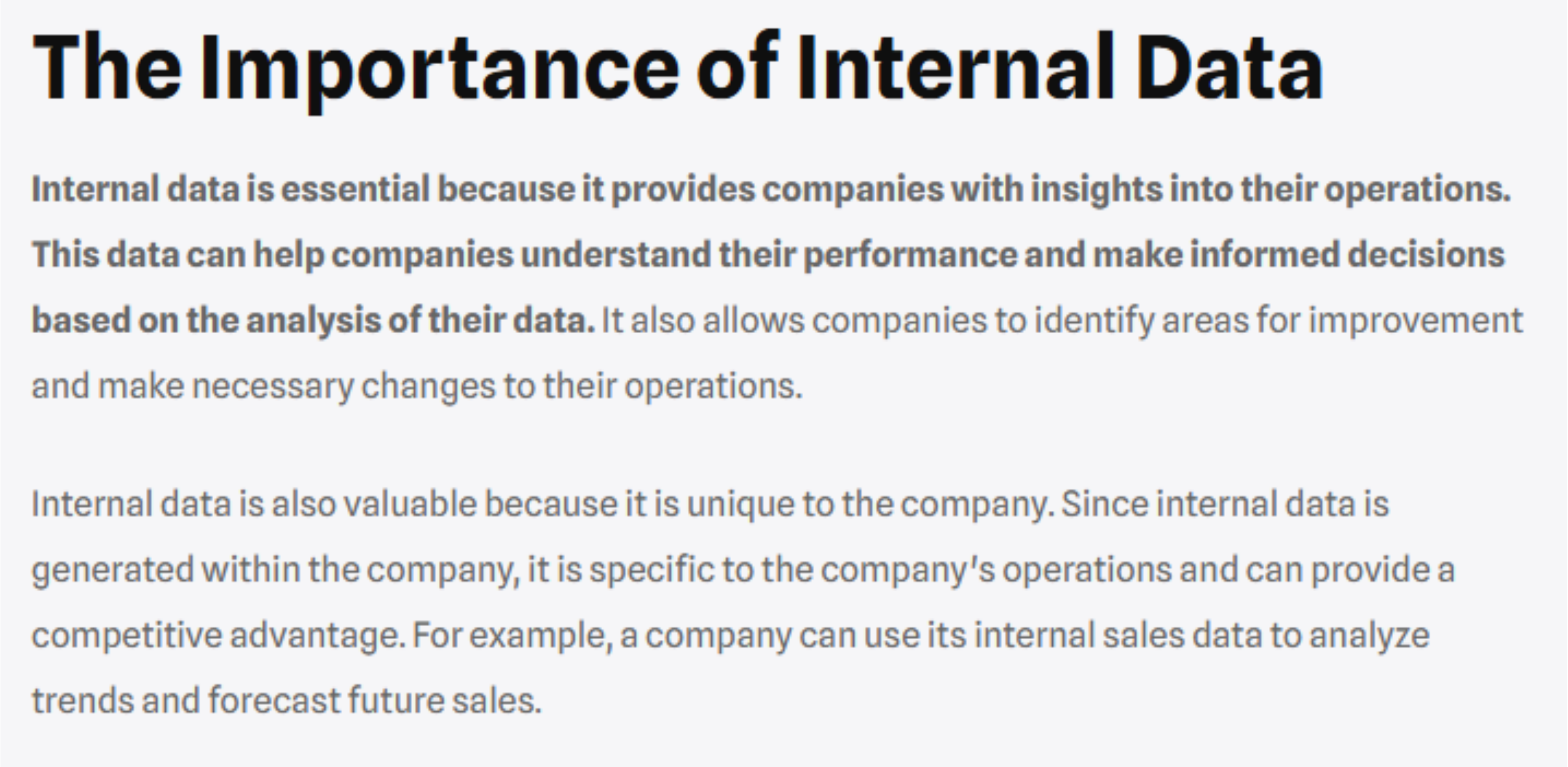
Sebastian Berg
CEO, Co-founder

Product March 15, 2023

Source

Source

Internal Data
External Data



The Importance of Internal Data

Internal data is essential because it provides companies with insights into their operations. This data can help companies understand their performance and make informed decisions based on the analysis of their data. It also allows companies to identify areas for improvement and make necessary changes to their operations.

Internal data is also valuable because it is unique to the company. Since internal data is generated within the company, it is specific to the company's operations and can provide a competitive advantage. For example, a company can use its internal sales data to analyze trends and forecast future sales.

Internal and External Data: What's the Difference and Why It Matters

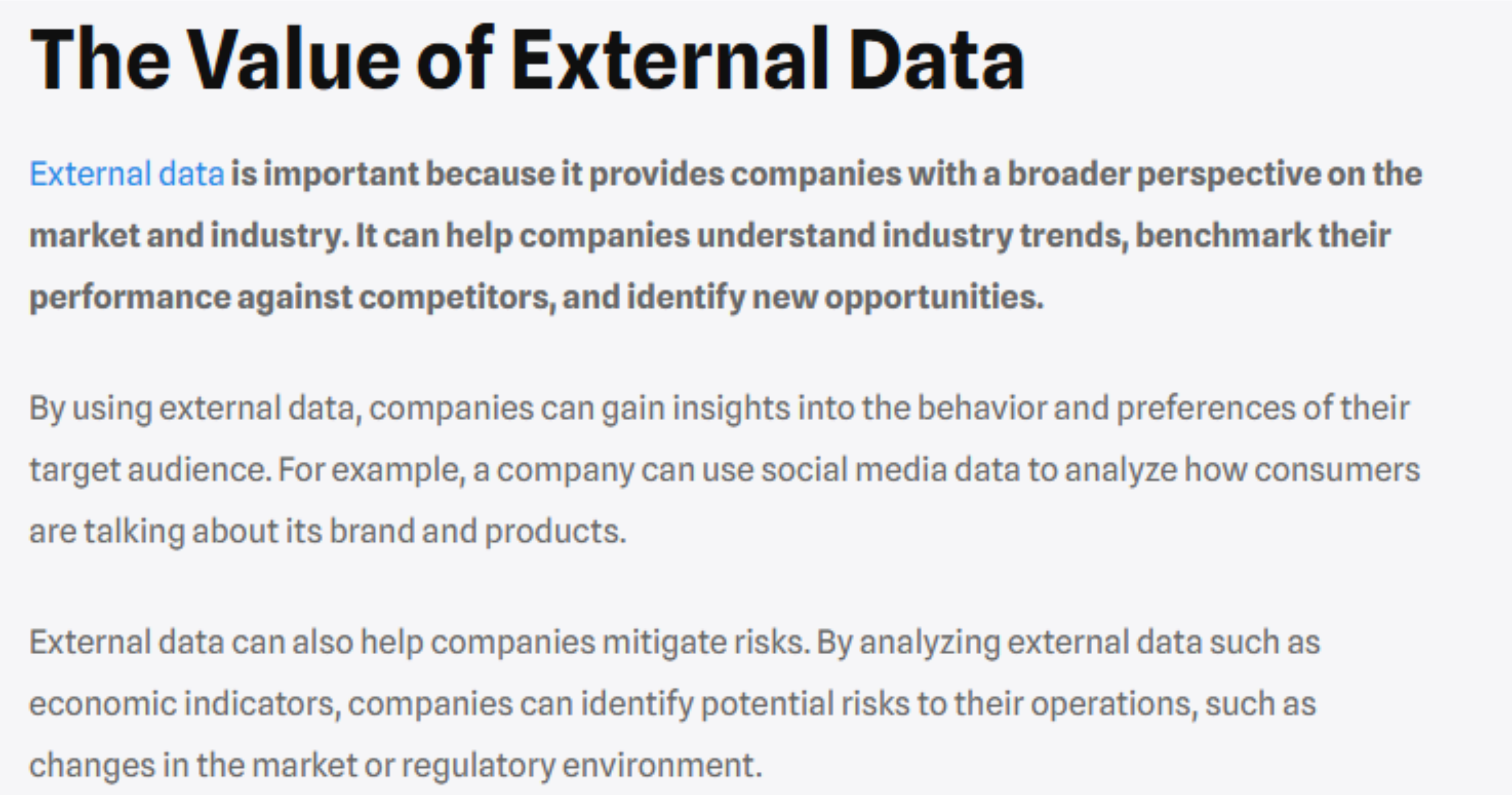
Sebastian Berg
CEO, Co-founder

Product March 15, 2023

Source

Source

Internal Data
External Data



The Value of External Data

External data is important because it provides companies with a broader perspective on the market and industry. It can help companies understand industry trends, benchmark their performance against competitors, and identify new opportunities.

By using external data, companies can gain insights into the behavior and preferences of their target audience. For example, a company can use social media data to analyze how consumers are talking about its brand and products.

External data can also help companies mitigate risks. By analyzing external data such as economic indicators, companies can identify potential risks to their operations, such as changes in the market or regulatory environment.

Internal and External Data: What's the Difference and Why It Matters

Sebastian Berg
CEO, Co-founder

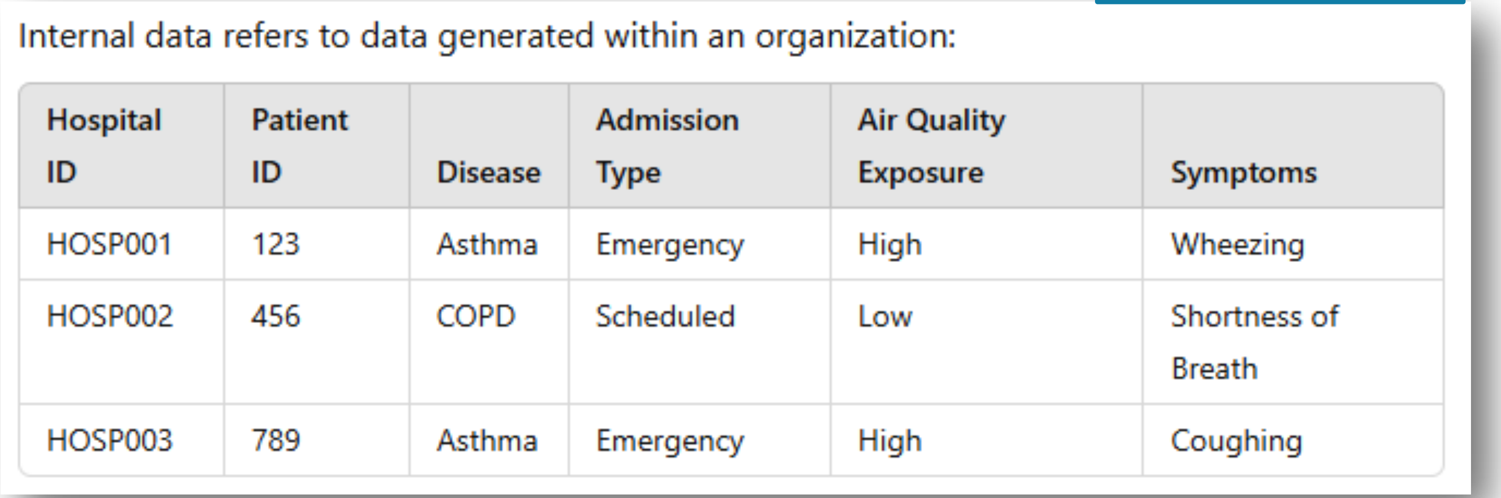
Product March 15, 2023

Source

Source

Internal Data
External Data

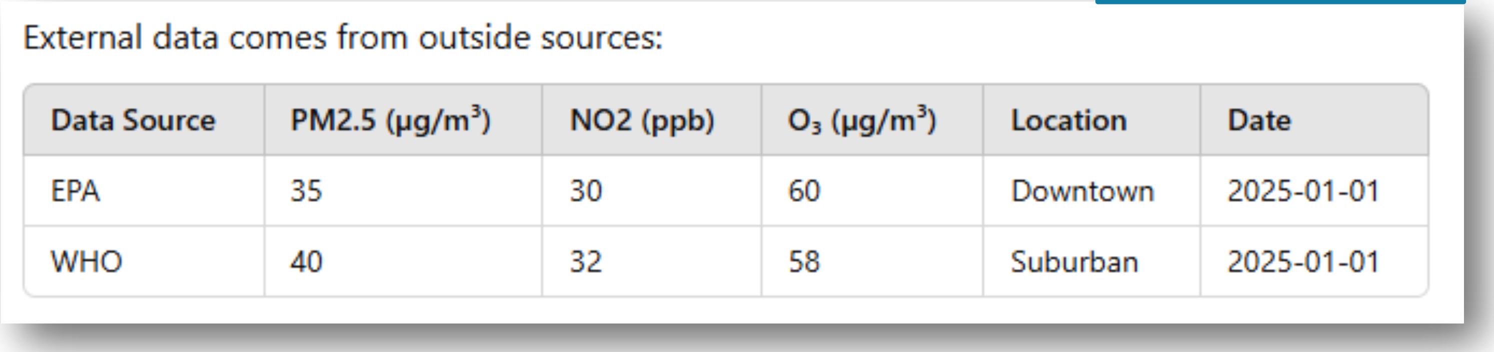
Internal



Internal data refers to data generated within an organization:

Hospital ID	Patient ID	Disease	Admission Type	Air Quality Exposure	Symptoms
HOSP001	123	Asthma	Emergency	High	Wheezing
HOSP002	456	COPD	Scheduled	Low	Shortness of Breath
HOSP003	789	Asthma	Emergency	High	Coughing

External

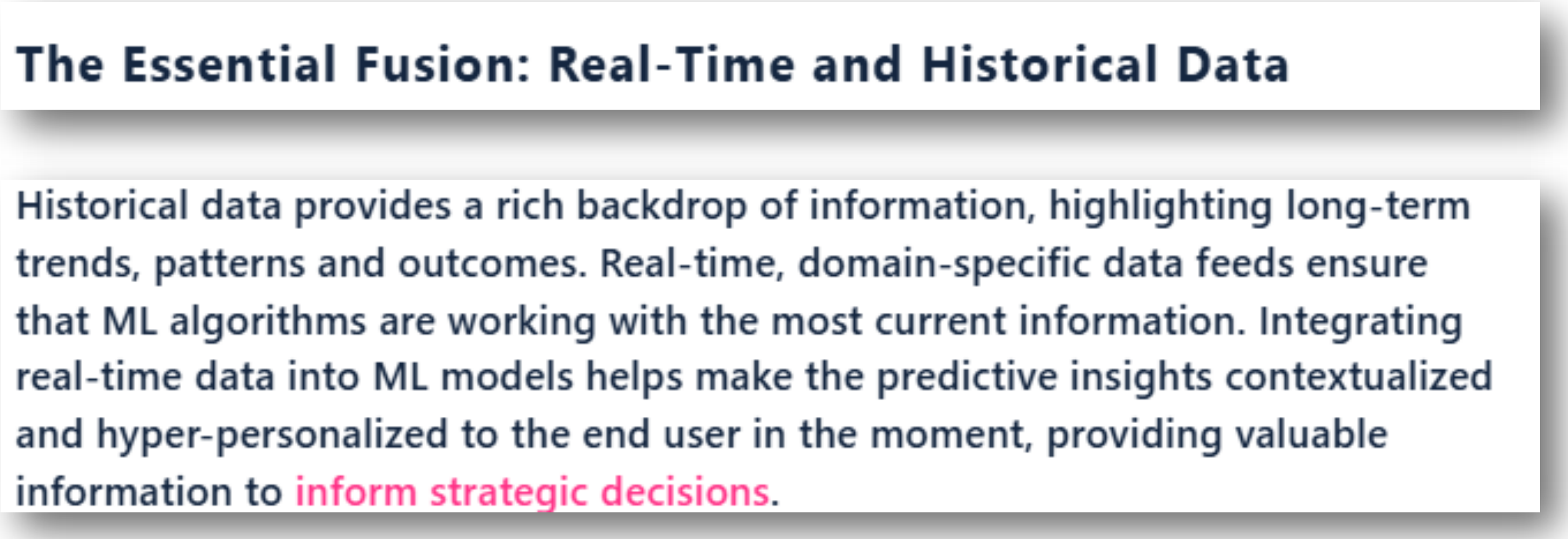


External data comes from outside sources:

Data Source	PM2.5 ($\mu g/m^3$)	NO2 (ppb)	O_3 ($\mu g/m^3$)	Location	Date
EPA	35	30	60	Downtown	2025-01-01
WHO	40	32	58	Suburban	2025-01-01

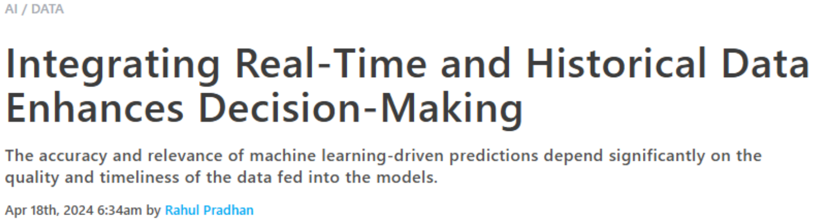
Time

Historical Data
Real-Time Data



The Essential Fusion: Real-Time and Historical Data

Historical data provides a rich backdrop of information, highlighting long-term trends, patterns and outcomes. Real-time, domain-specific data feeds ensure that ML algorithms are working with the most current information. Integrating real-time data into ML models helps make the predictive insights contextualized and hyper-personalized to the end user in the moment, providing valuable information to inform strategic decisions.



AI / DATA

Integrating Real-Time and Historical Data Enhances Decision-Making

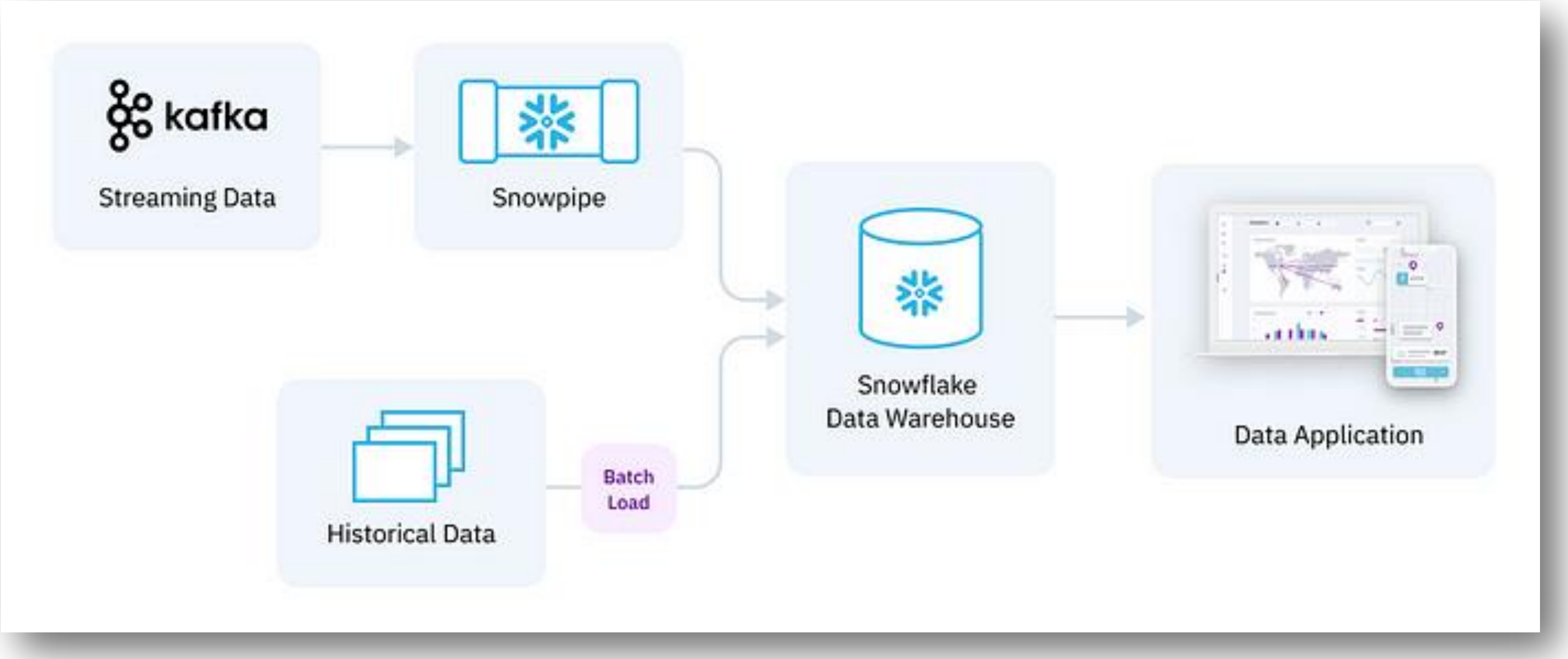
The accuracy and relevance of machine learning-driven predictions depend significantly on the quality and timeliness of the data fed into the models.

Apr 18th, 2024 6:34am by Rahul Pradhan

Source

Time

Historical Data
Real-Time Data



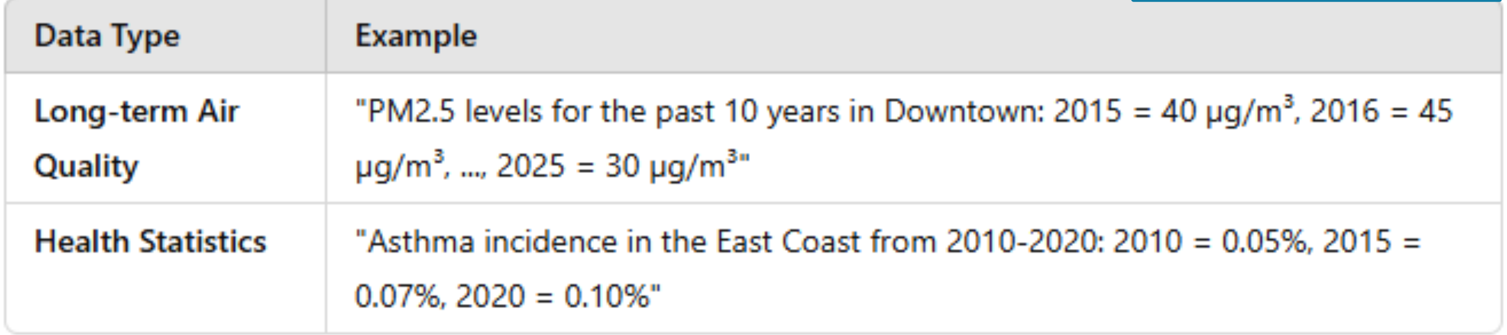
Joining Streaming and Historical Data for Real-Time Analytics: Your Options With Snowflake, Snowpipe and Rockset

Source

Time

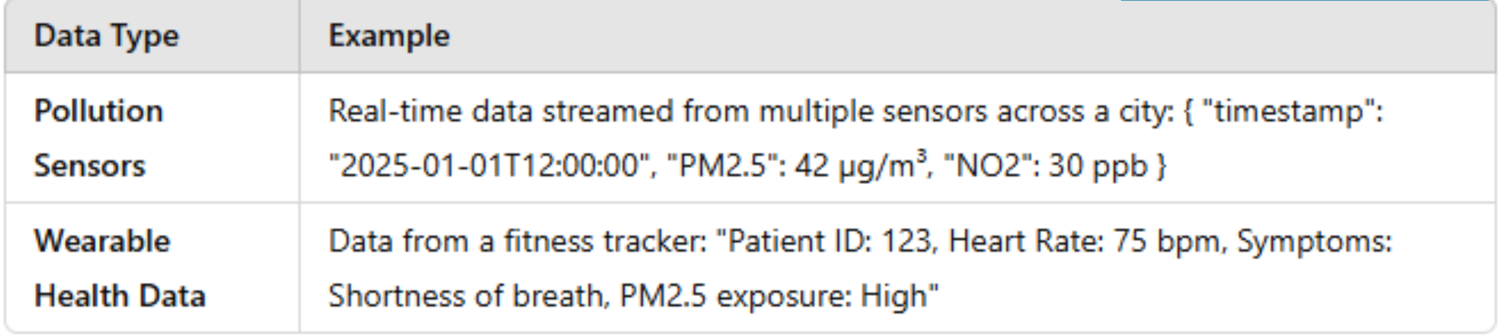
Historical Data
Real-Time Data

Historical



Data Type	Example
Long-term Air Quality	"PM2.5 levels for the past 10 years in Downtown: 2015 = 40 µg/m^3, 2016 = 45 µg/m^3, ..., 2025 = 30 µg/m^3"
Health Statistics	"Asthma incidence in the East Coast from 2010-2020: 2010 = 0.05%, 2015 = 0.07%, 2020 = 0.10%"

Real-time



Data Type	Example
Pollution Sensors	Real-time data streamed from multiple sensors across a city: { "timestamp": "2025-01-01T12:00:00", "PM2.5": 42 µg/m^3, "NO2": 30 ppb }
Wearable Health Data	Data from a fitness tracker: "Patient ID: 123, Heart Rate: 75 bpm, Symptoms: Shortness of breath, PM2.5 exposure: High"

Measurement

Quantitative (Numerical)
Qualitative (Categorical)

Format

Structured
Unstructured
Semi-Structured

Volume

Big Data
Small Data

Collection

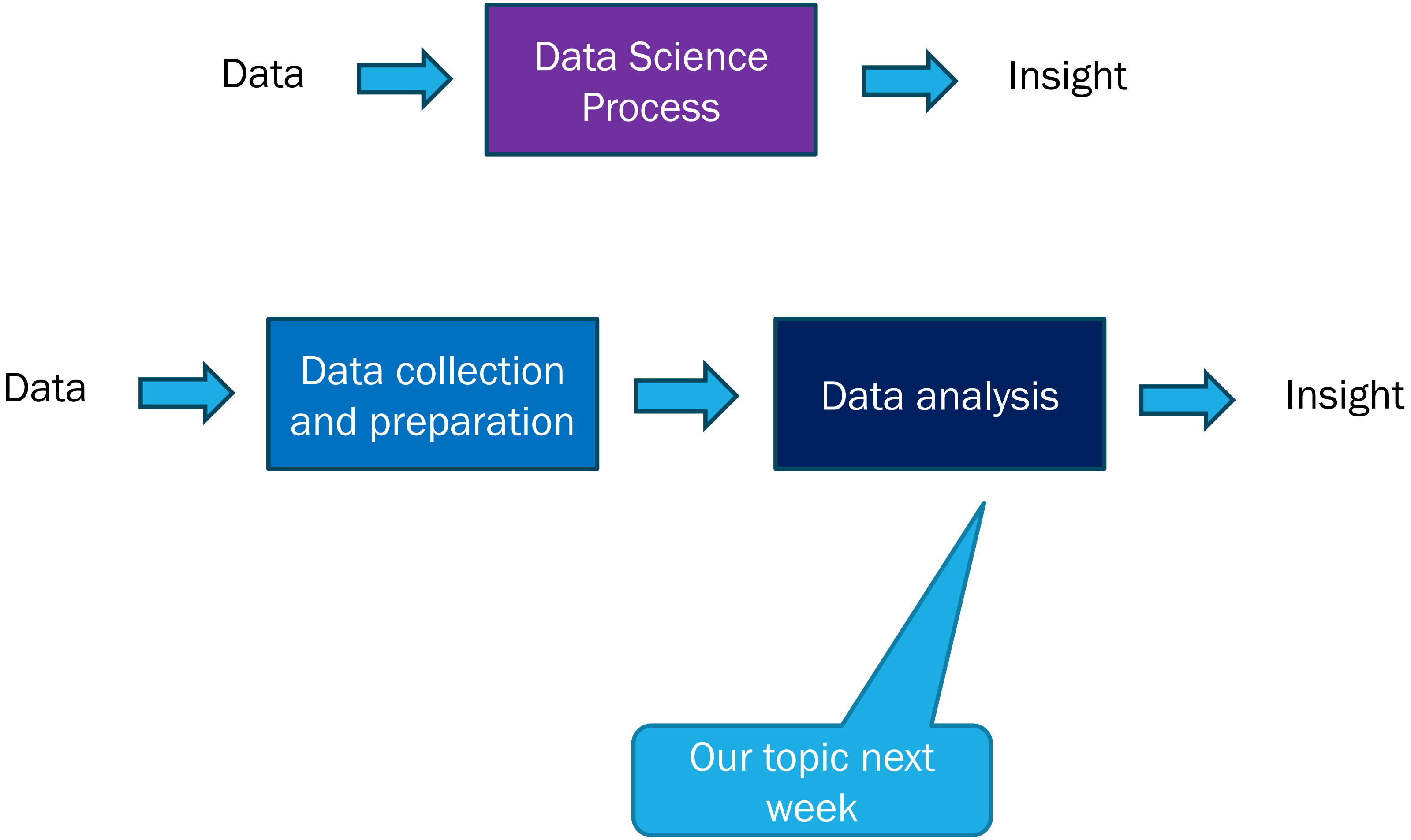
Primary Data
Secondary Data

Source

Internal Data
External Data

Time

Historical Data
Real-Time Data



Data → Data Science Process → Insight

Data → Data collection and preparation → Data analysis → Insight

Our topic next week

DERIVING INSIGHTS FROM DATA

- Types of insights
- Types of data