

# Machine Learning - Introduction



Data Science

# Structured Data (1)

---

- can be stored in as tables (or matrices)
- have a defined interpretation for each column (or limited group of columns)
- can be evaluated (queried) as tables

Structured data can be stored *and* analyzed  
in relational database tables

Data Management Resource: [Forrester Wave - Master Data Management](#)

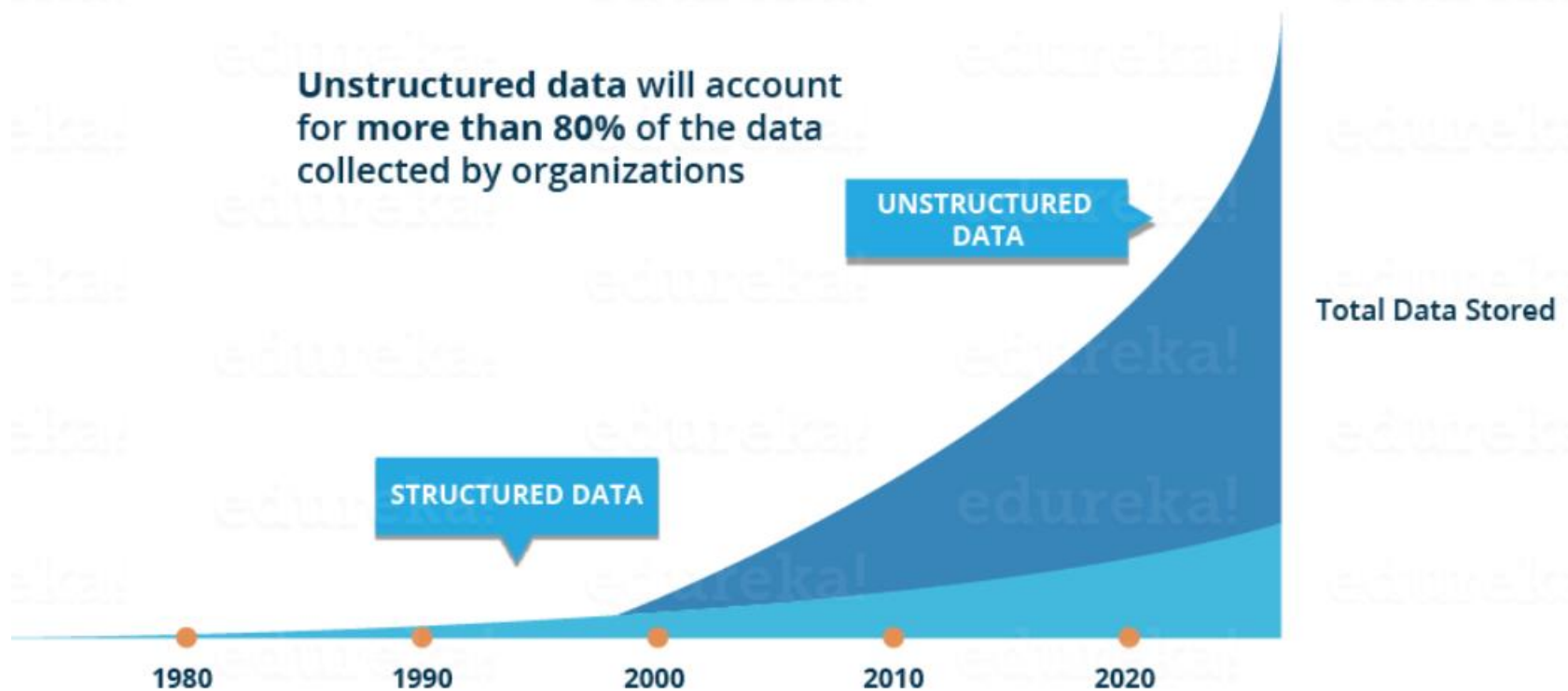
## What is Structured Data?

**Structured data** usually resides in relational databases (RDBMS). Fields store length-delineated data phone numbers, Social Security numbers, or ZIP codes. Even text strings of variable length like names are contained in records, making it a simple matter to search. Data may be human- or machine-generated as long as the data is created within an RDBMS structure. This format is eminently searchable both with human generated queries and via algorithms using type of data and field names, such as alphabetical or numeric, currency or date.

<https://www.datamation.com/big-data/structured-vs-unstructured-data.html>

# Structured and Unstructured Data

More than 80% of data will be unstructured by 2020



# Definition „Data Science“

---

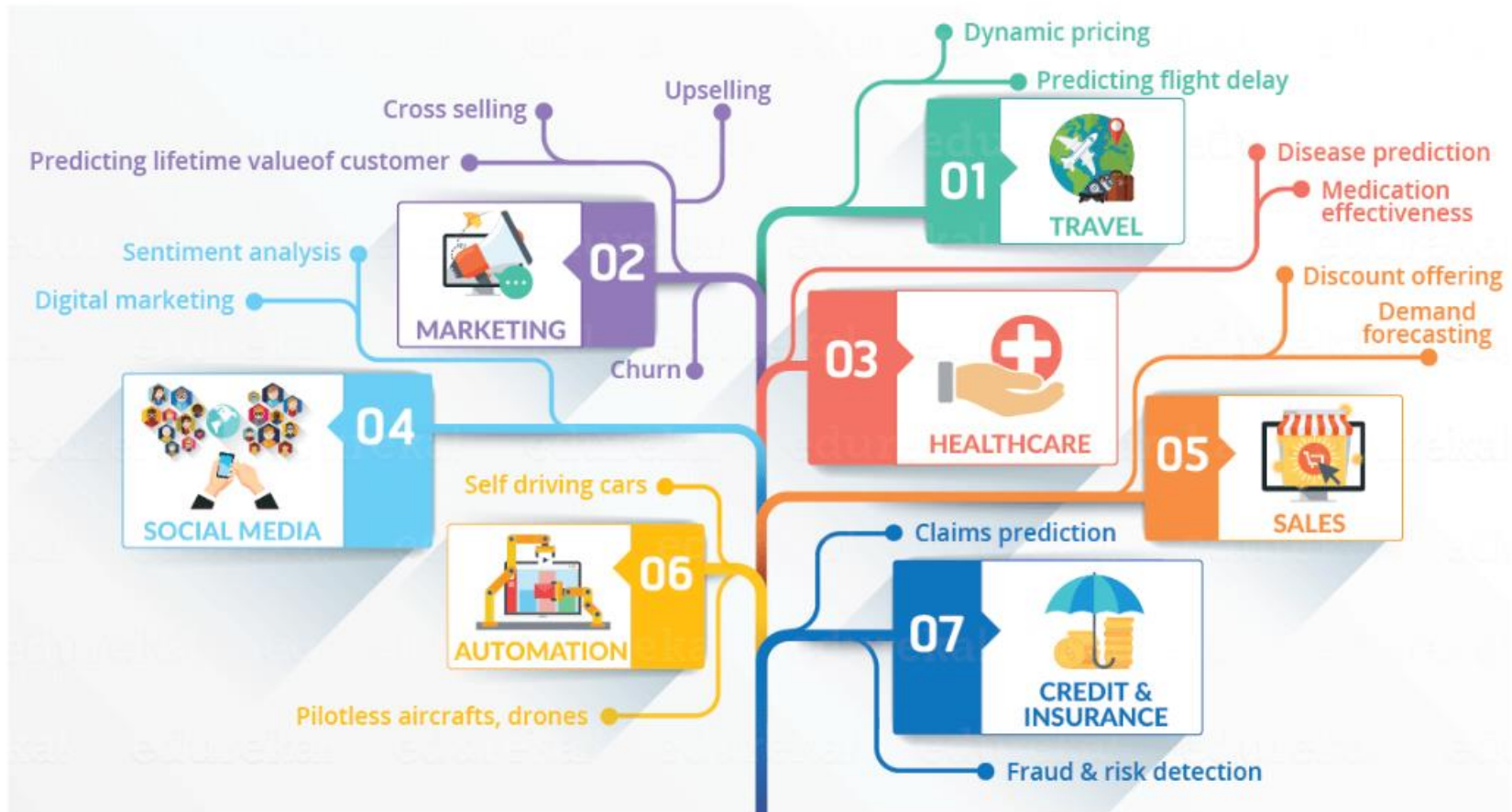
*Data science* is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to **extract knowledge and insights from structured and unstructured data**. *Data science* is the same concept as *data mining* and *big data*: "use the most powerful hardware, the most powerful programming systems, and the most efficient algorithms to solve problems".

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science.

Basierend auf: [https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)



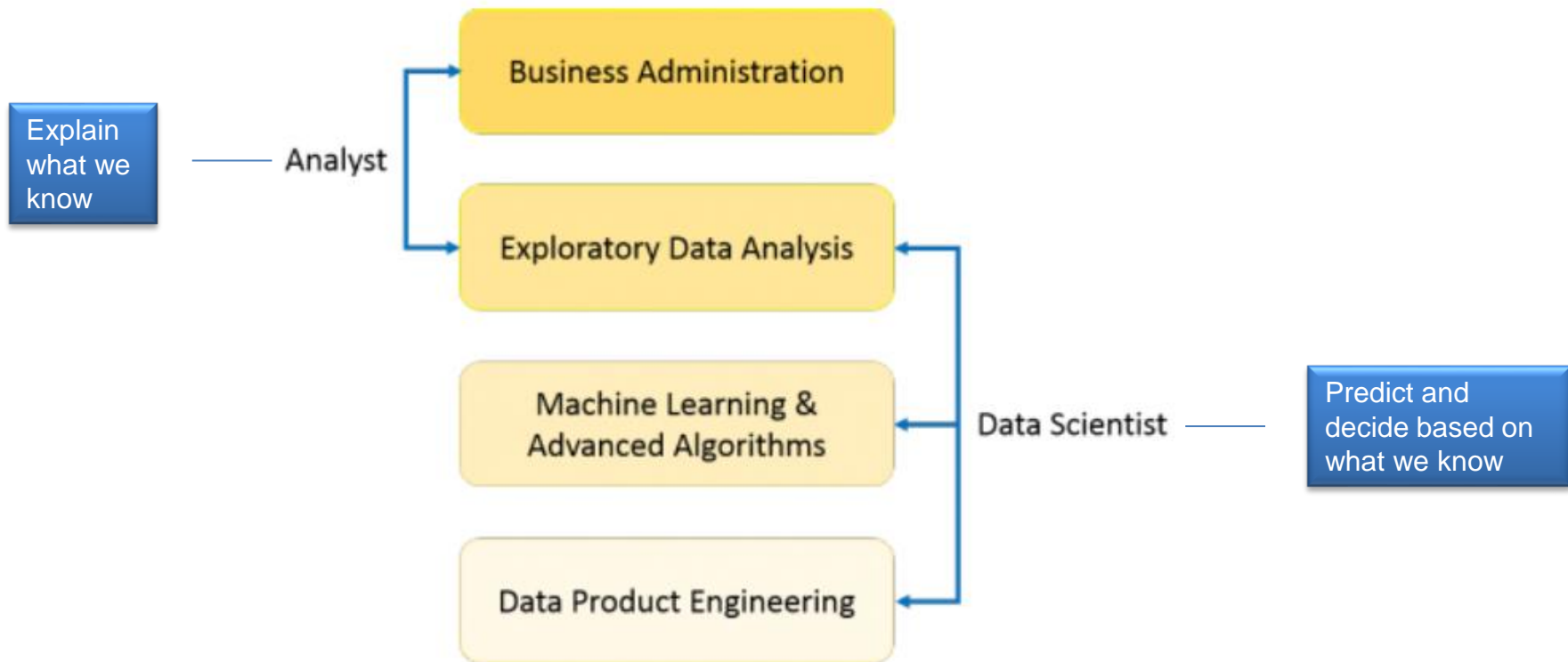
# Use Cases for Data Science



Source: <https://www.edureka.co/blog/what-is-data-science/>

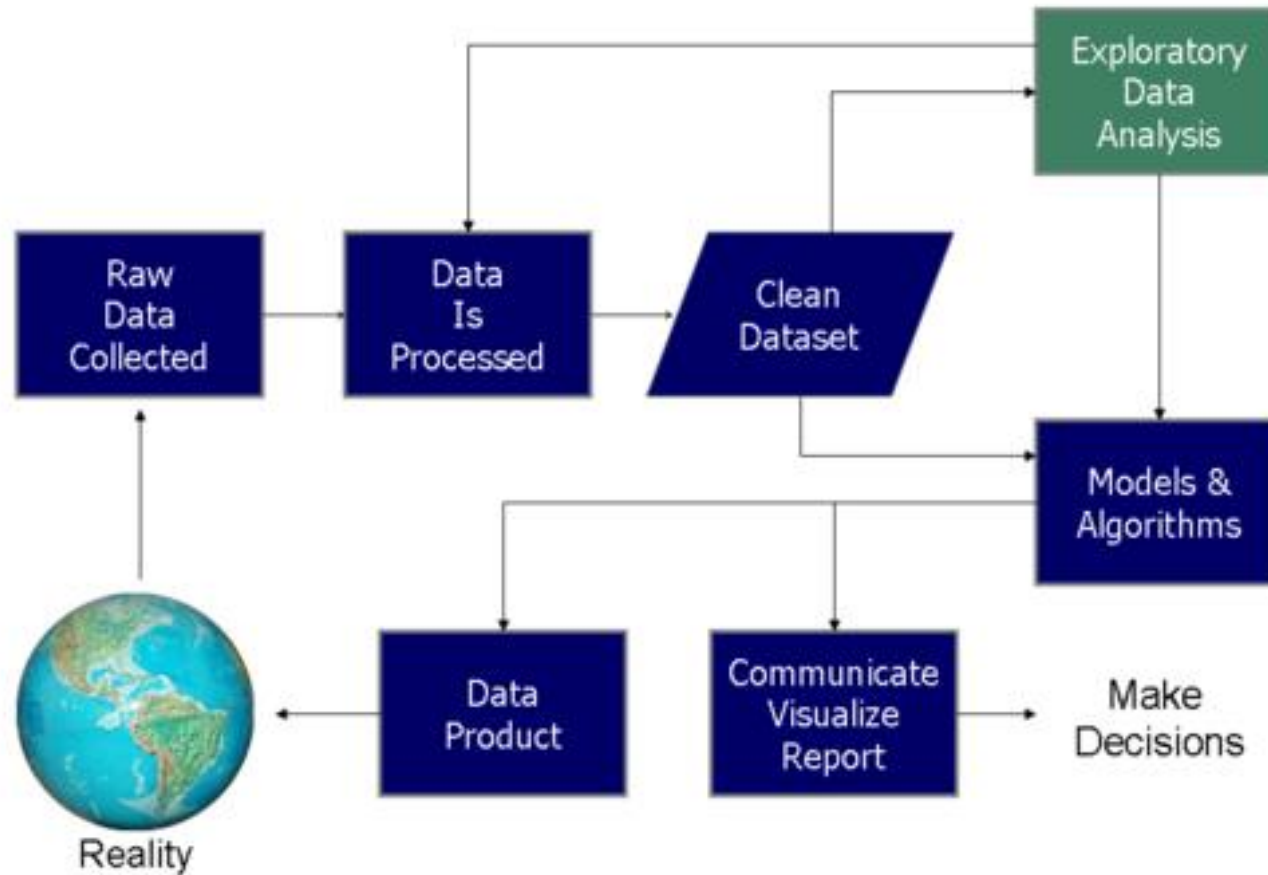
# Data Analyst vs. Data Scientist

## Explain vs Predict/Decide



In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.

Source: <https://www.edureka.co/blog/what-is-data-science/>





<https://blog.iao.fraunhofer.de/data-science-und-kuenstliche-intelligenz-wie-ihre-daten-das-kundenverhalten-vorhersagbar-machen/>