

Data science World

Data All Around

Lots of data is being collected

- Web data, e-commerce
- Financial transactions, bank/credit transactions
- Online trading and purchasing
- Social Network



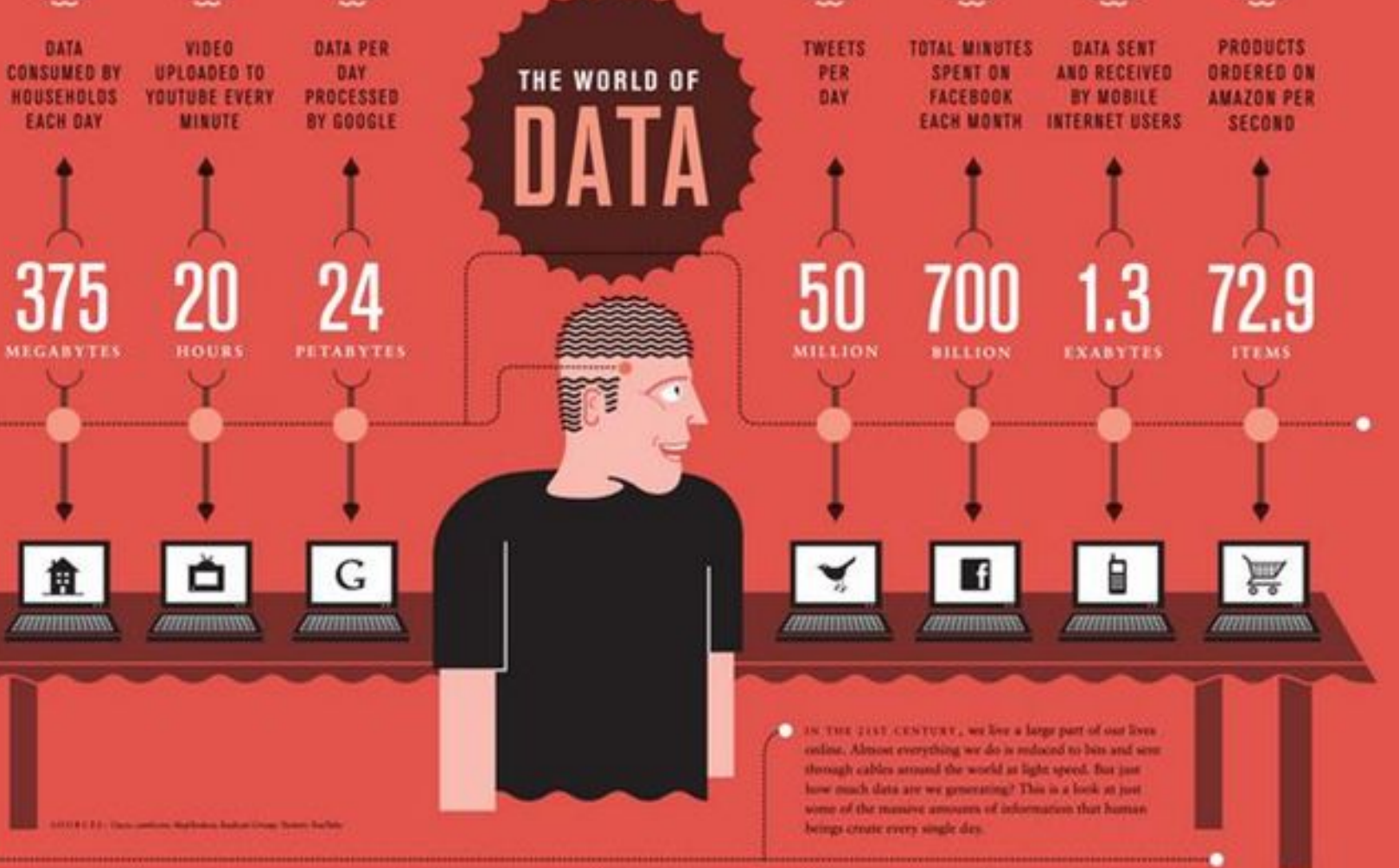


ILLUSTRATION: MATHIEU BOUTIER

What To Do With These Data?

Data can be investigated and carefully analysed, providing insights which enrich our daily lives.

Analyzing data allows us to understand customer preferences, predict trends, and solve problems.

We can find patterns and trends that can guide our decisions.

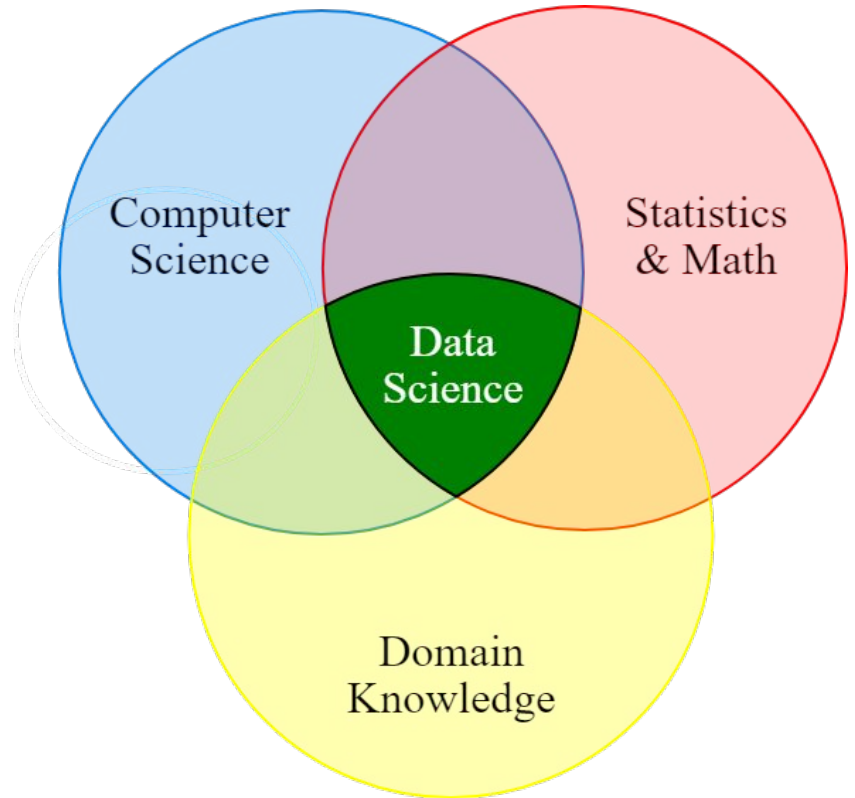
we can make sense of the information we collect.



Extracting meaningful insights = **Data Science**

Data Science

"The data science is an interdisciplinary field, borrowing methods, processes, and algorithms from business, statistics, and computer science to extract information from data."



Careers in data science?

Learning data science offers multiple career options. Some of the common job titles for data scientists include

- Data Scientist
- Business Intelligence Analyst
- Data Mining Engineer
- Data Architect

Careers in data science?

Data Scientist

data enthusiasts who gather and analyze large sets of structured and unstructured data. They analyze, process, and model data and later interpret the results to create actionable plans for companies and organizations.

Business Intelligence Analyst

Business Intelligence Analysts use data to assess the market and find the latest business trends in the industry. This helps to develop a clearer picture of how a company should shape its strategy.

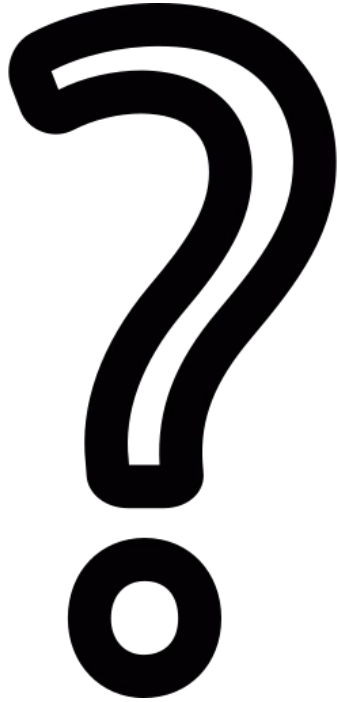
Data Engineer

Data Engineer examines not only the Data for their own business but also that of third parties. In addition to mining data, a data engineer creates robust algorithms to help analyze the data further.

Data Architect

Data Architects work closely with users, system designers, and developers to create a blueprint that data management systems use to centralize, integrate and maintain the data sources.

How data science helps us?



Simply stated, data science helps us answer different types of questions from data. Some common questions to ask from data are:

- Which class does this belong to - A or B?
- Is this an outlier?
- What will probably be the value of this variable?
- What should be done now?

Which class does this belong to - A or B?



The answers to some questions can only be from a definite number of options.

For example,

Q: Will it rain today?

A: Yes/No

Q: Will the weather be hot or cold?

A: Hot/Cold

To make such predictions, we use a family of algorithms called **classification algorithms**.

Is this an outlier?

In some cases, the objective is to find outliers or anomalies in data that is otherwise mostly consistent. Some examples of anomaly detections are:

Q: Is this email normal or spam?

Q: You are checking your car tyre pressure. Is the reading normal?

The algorithms that are used for these types of questions are called **anomaly detection algorithms**.



What will probably be the value of this variable?



There are scenarios in which we must predict numerical values of a variable based on historic data. Some examples are:

Q: How much rainfall will we receive this year?

A: 100 mm

Q: How many runs will the winning team score?

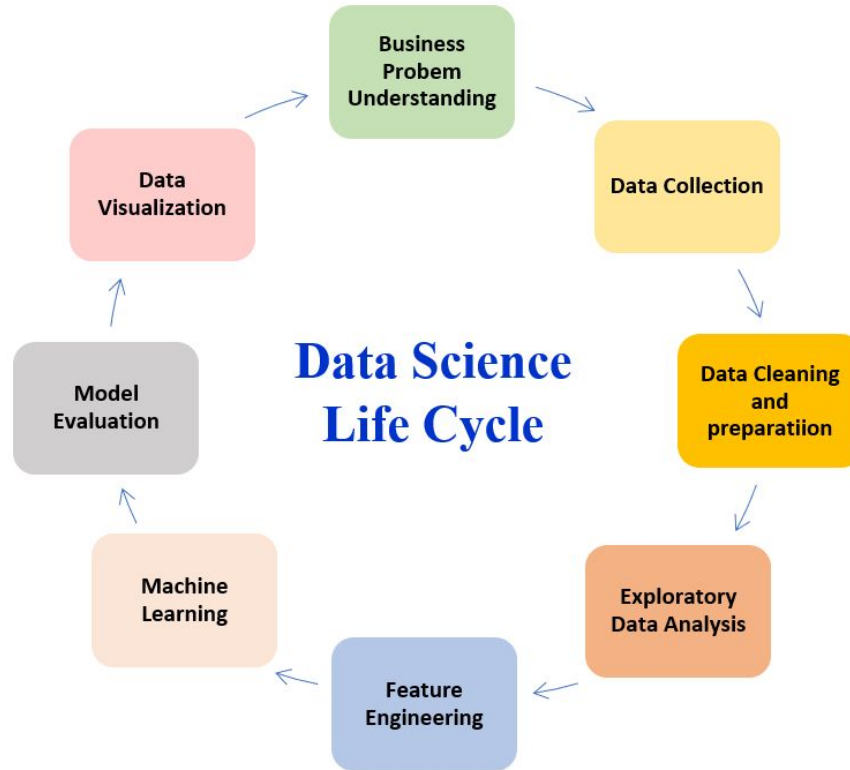
A: 320

The kind of algorithms that can predict these values are called **regression algorithms**.

•



Data science lifecycle



Data science tools

