# Intro to data science

A really really basic idea about data science and how to go on about it
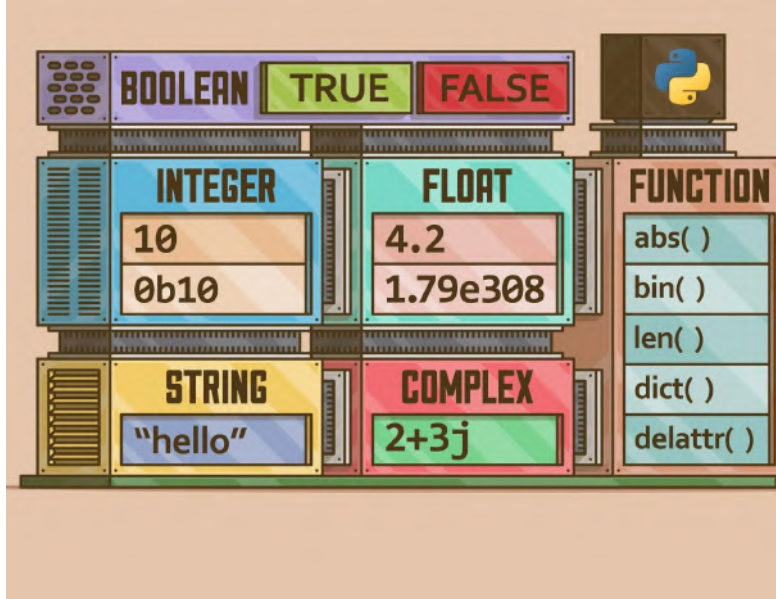
Agenda

- Familiarity with python and prerequisites that you'll need

- What is data science

- how do you even go about solving a problem using data science

- A very light example on ML

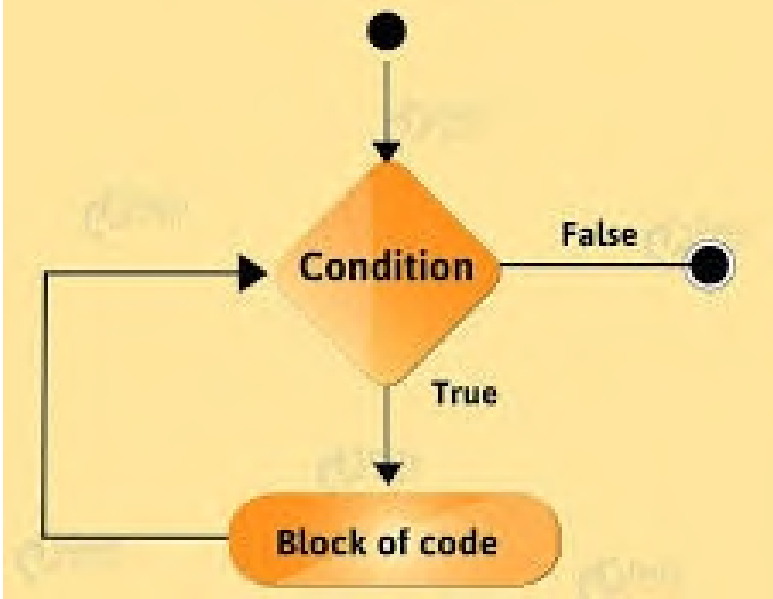- Some cool stuff scientists have done with data science

Prerequisites

- Beginner to Intermediate Python proficiency
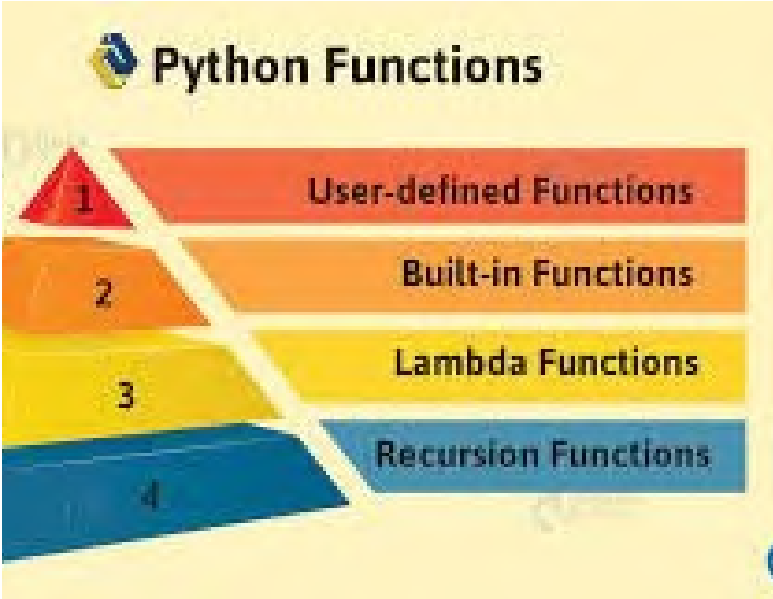
- nah, that's it, really
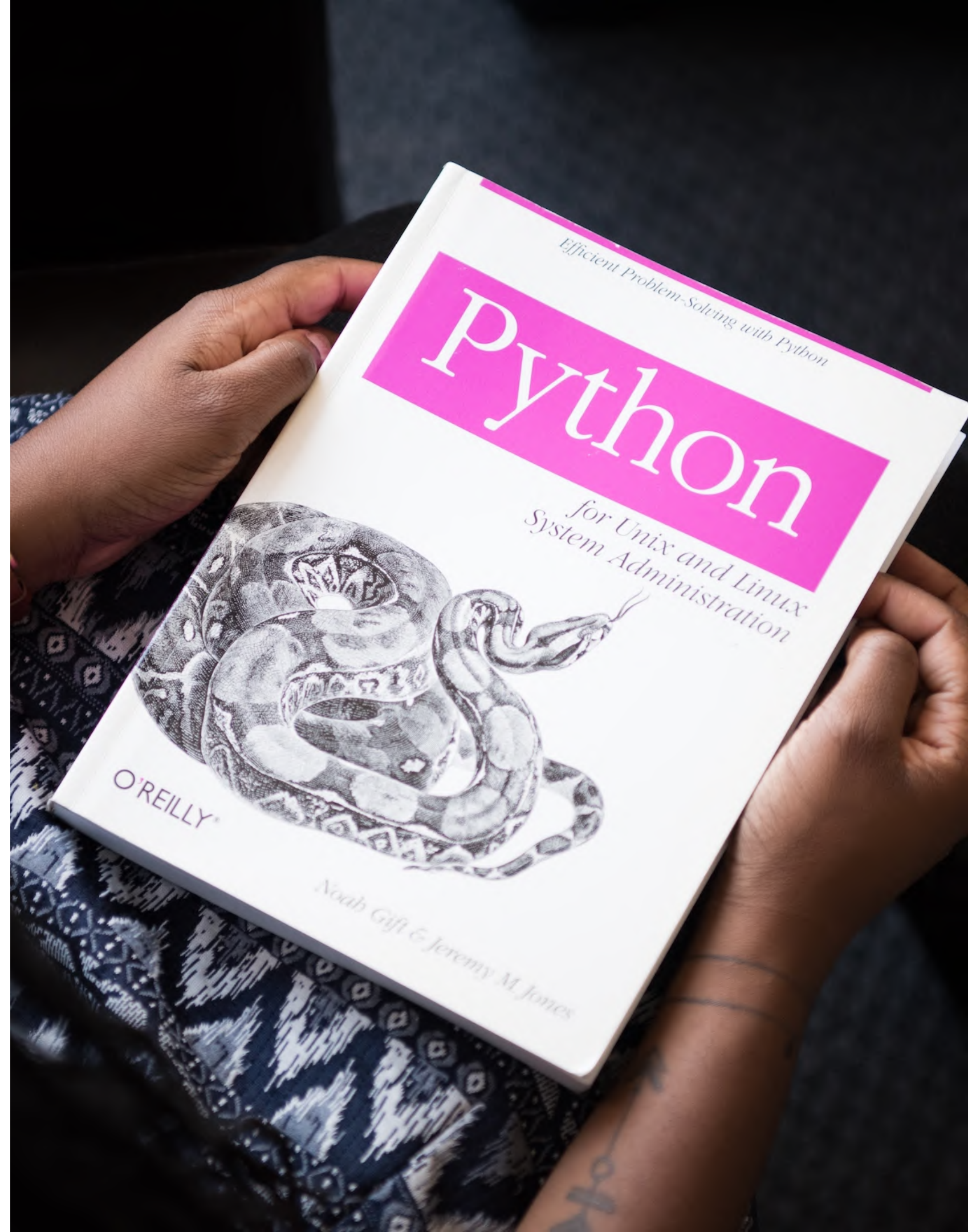
Data types

Conditional statements and loops

Functions

Packages and Libraries

What is data science?????

It is all about 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.
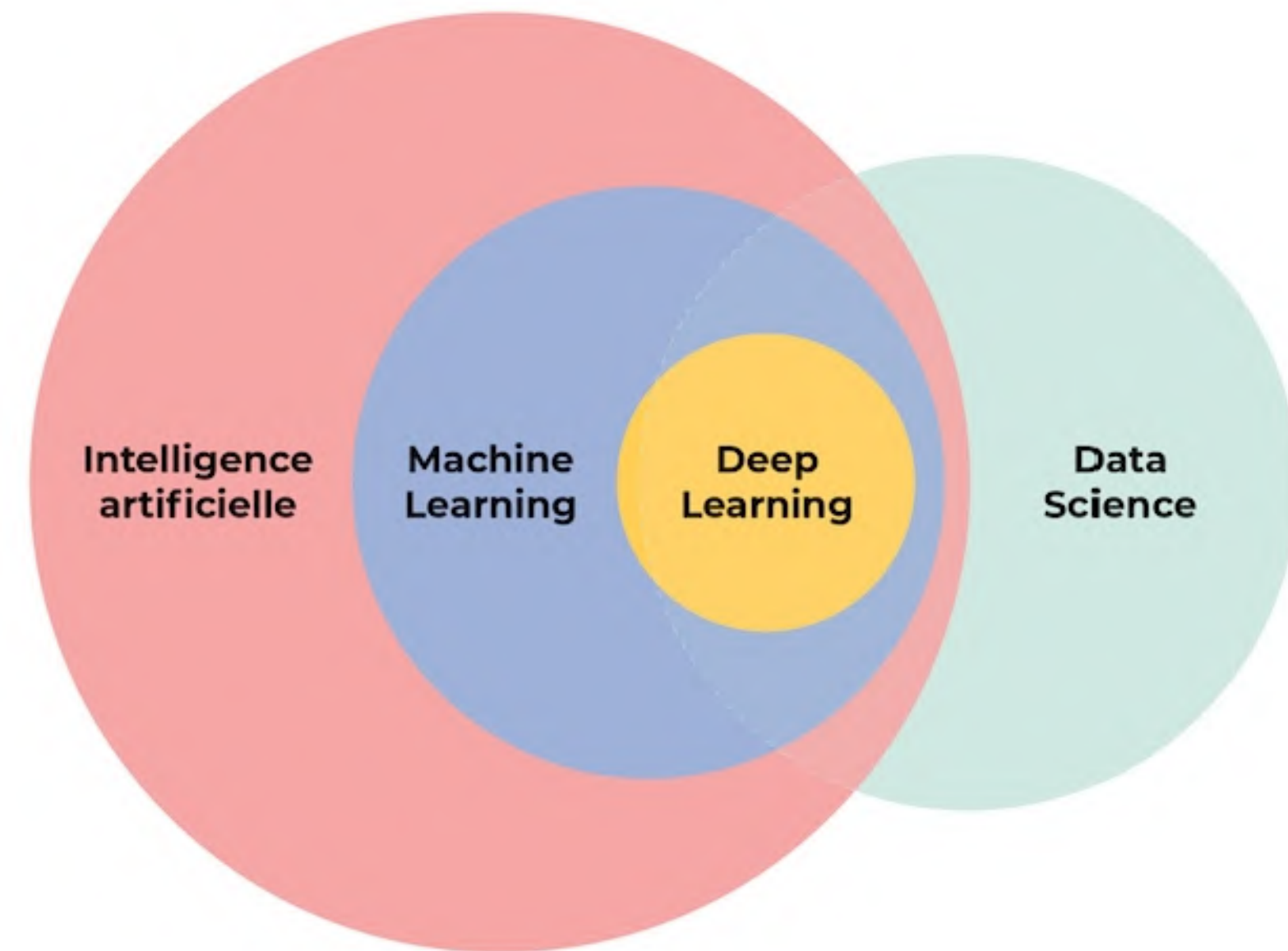
Too wordy? It is.

Data science is not ML or AI

Data science is about using data to provide value (money, growth, reputation, etc.) to an organization by analyzing the data.

Machine learning is about using data to make optimized inferences and predictions.

Artificial Intelligence is about using data to impart human-like decision making to machines.

Ever wonder how your Gmail automatically sorts what is spam mail and what isn't?

Cmon, you already know the answer

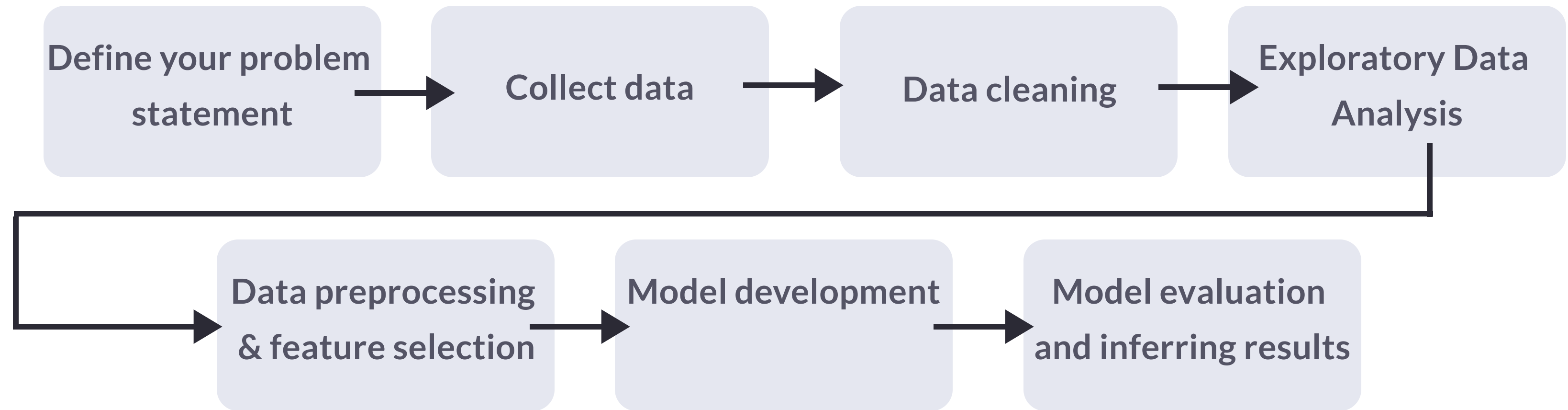That's right, Data science!
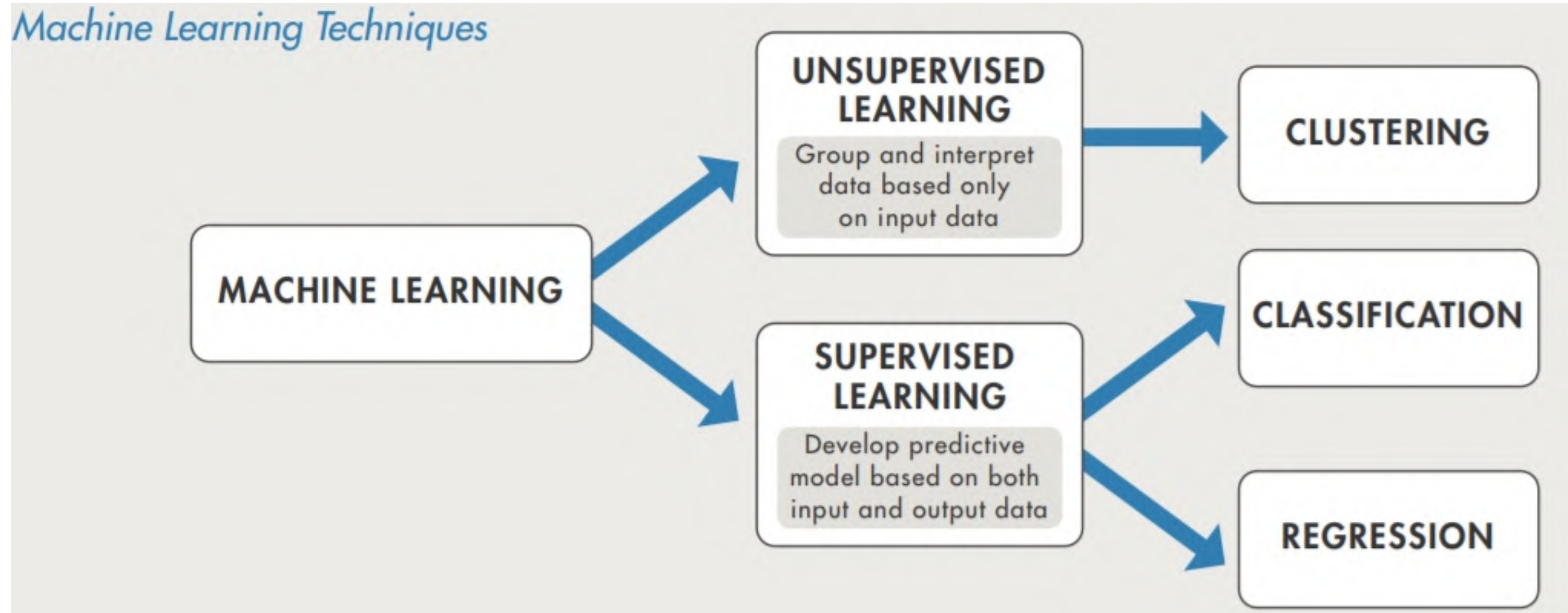
Other cool things that use data science

Other cool things that use data science

- Suggesting movies on Netflix
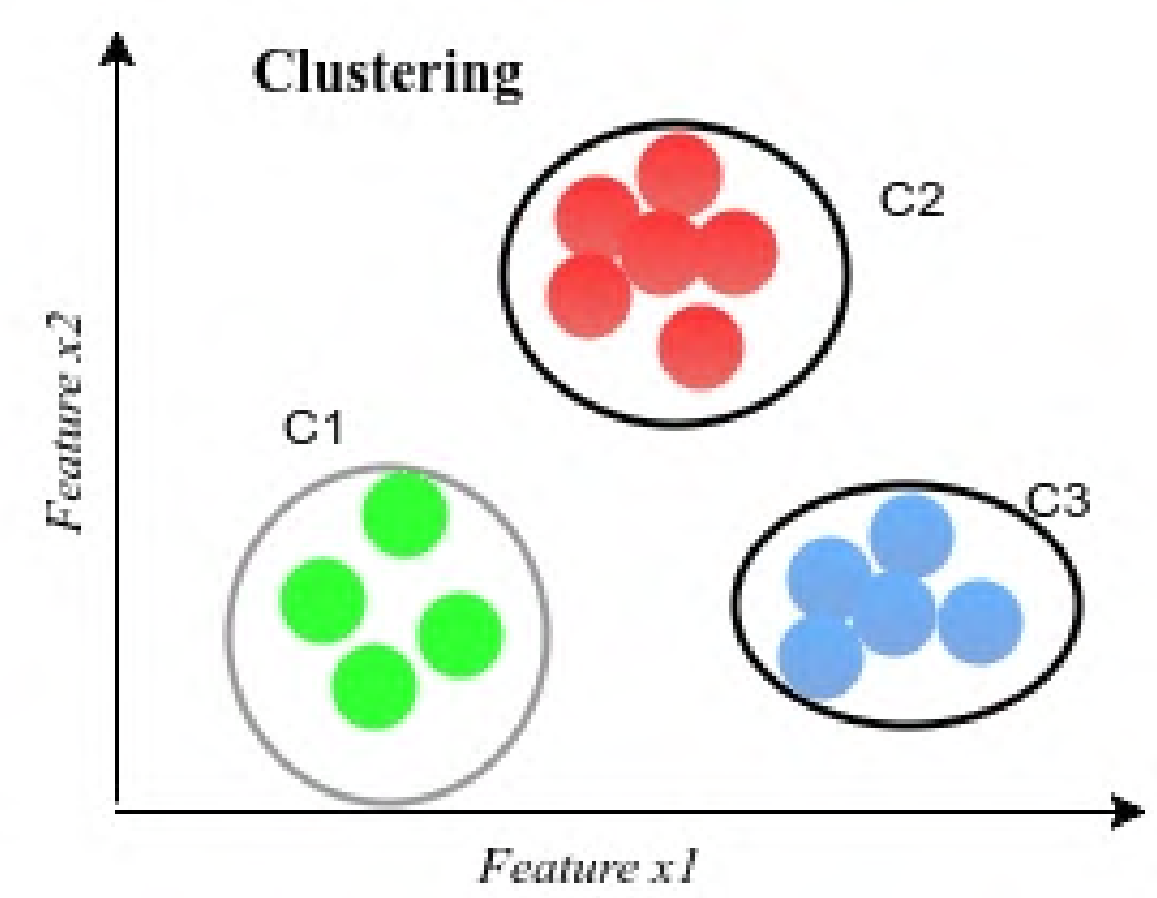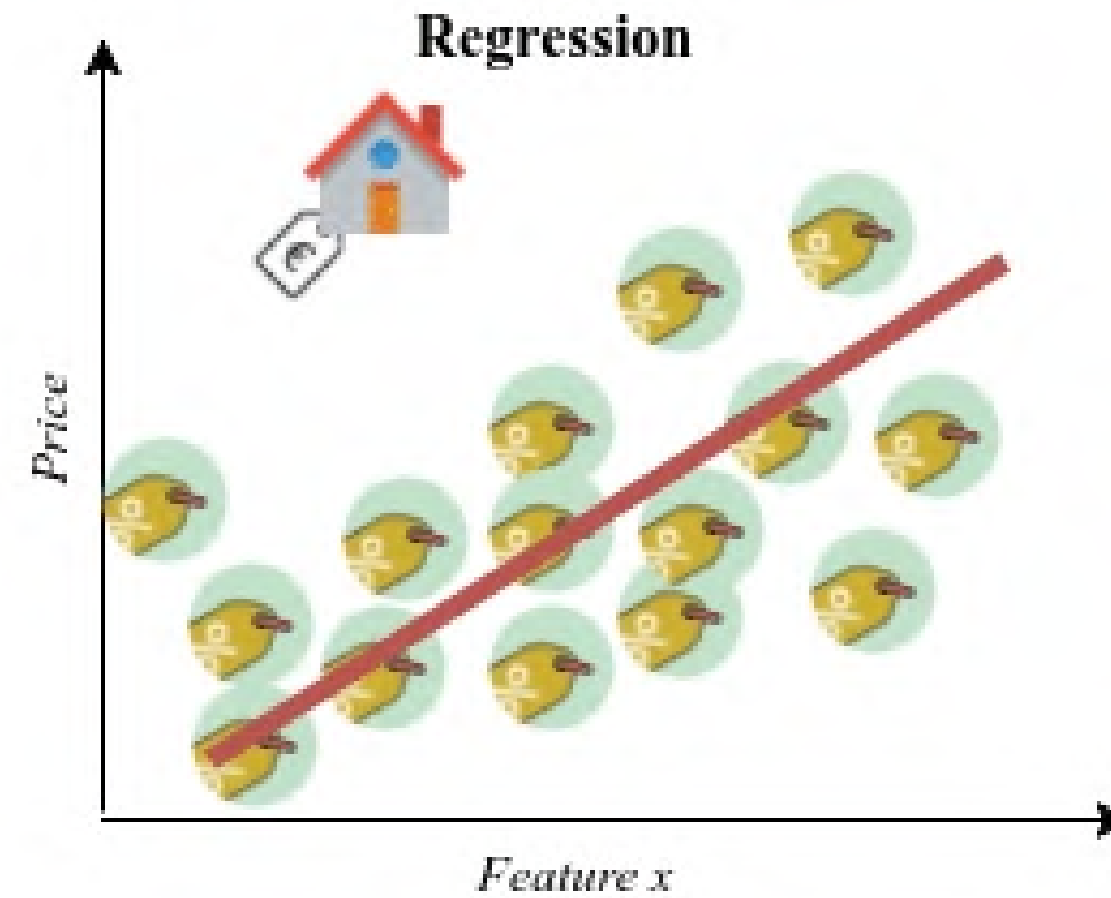- Search engines
- Targeted Advertising

Steps to solve a problem using data science & ML
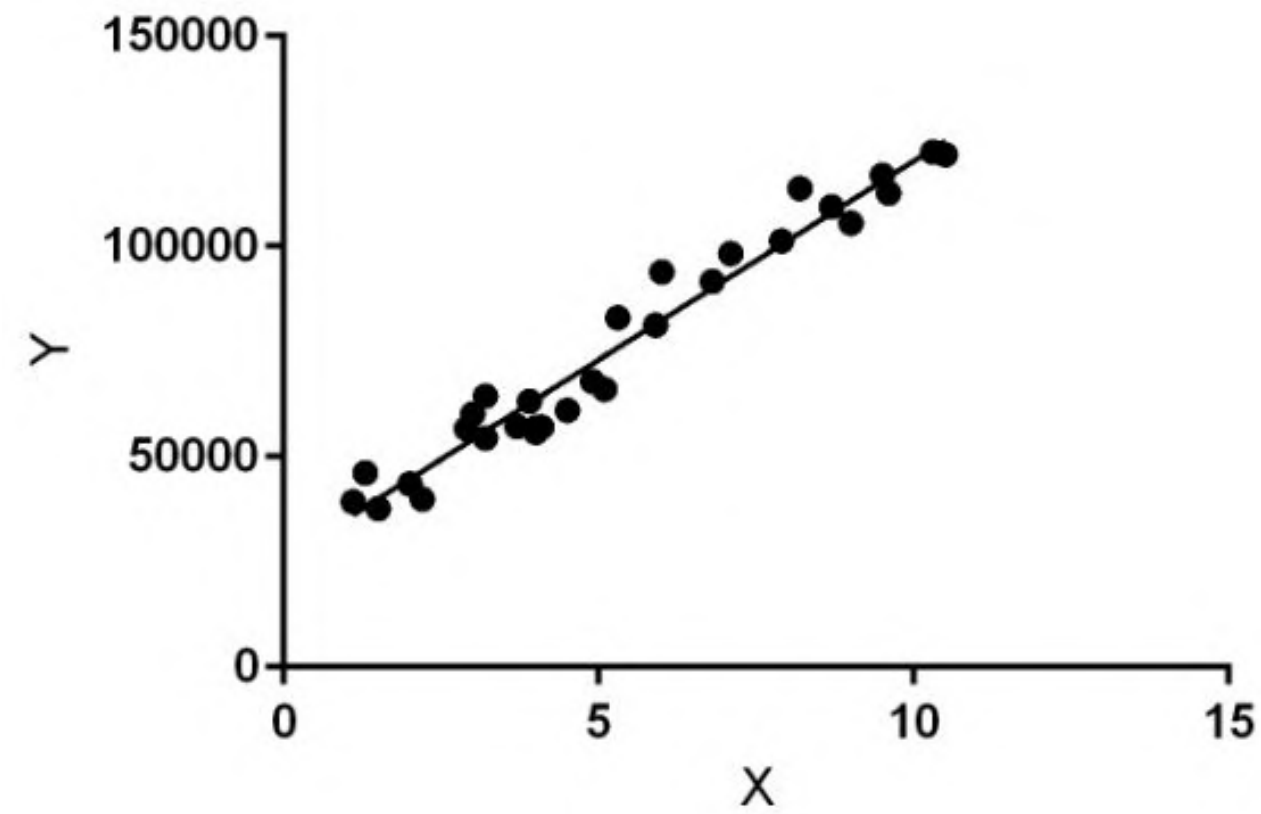
Define your problem statement	→	Collect data	→	Data cleaning	→	Exploratory Data Analysis

Data preprocessing & feature selection	→	Model development	→	Model evaluation and inferring results

Some background to ML and Linear regression
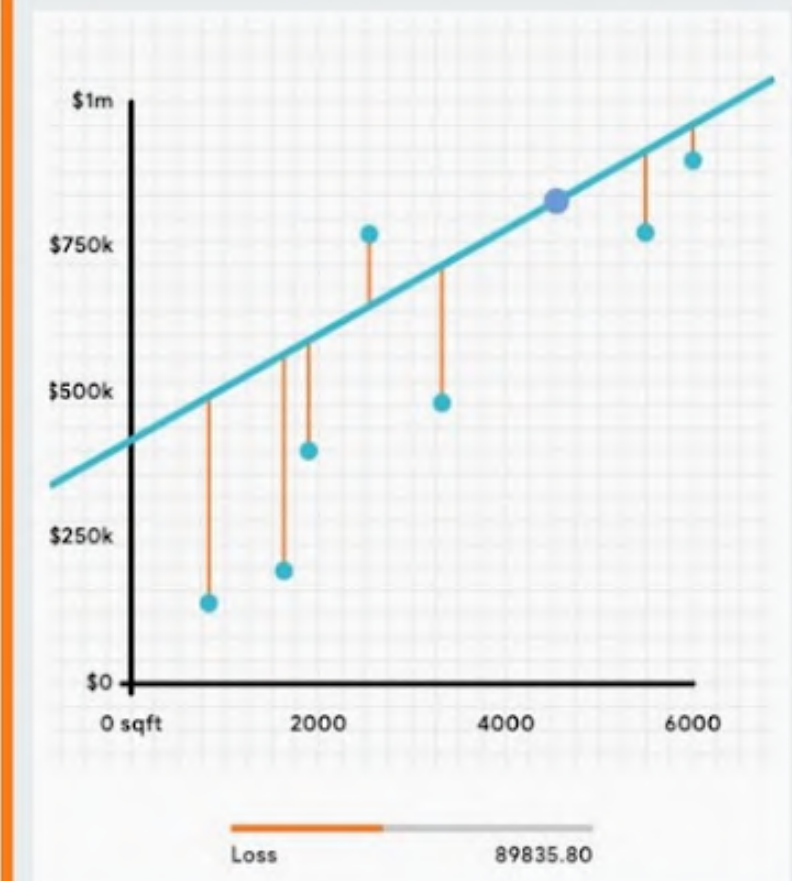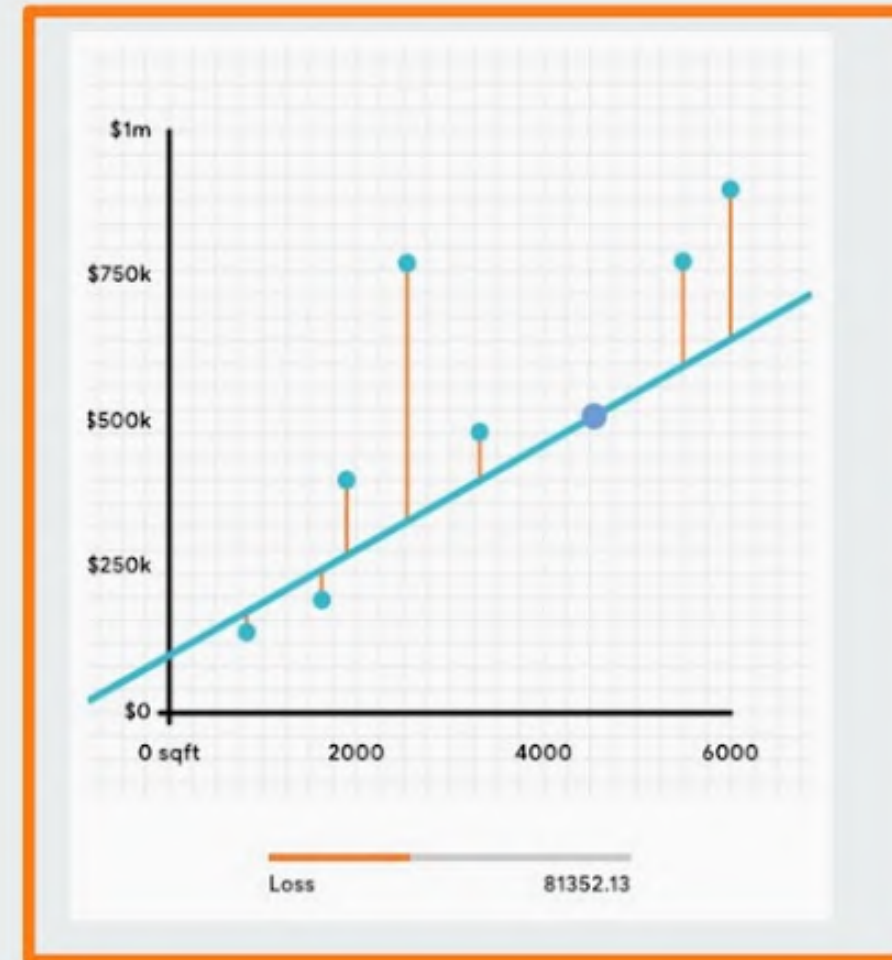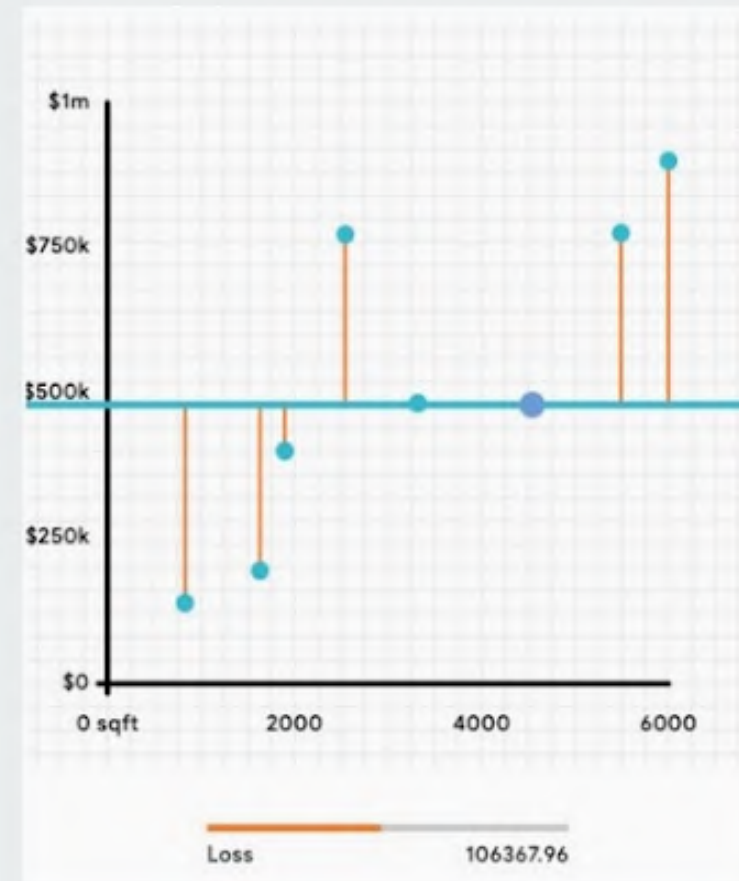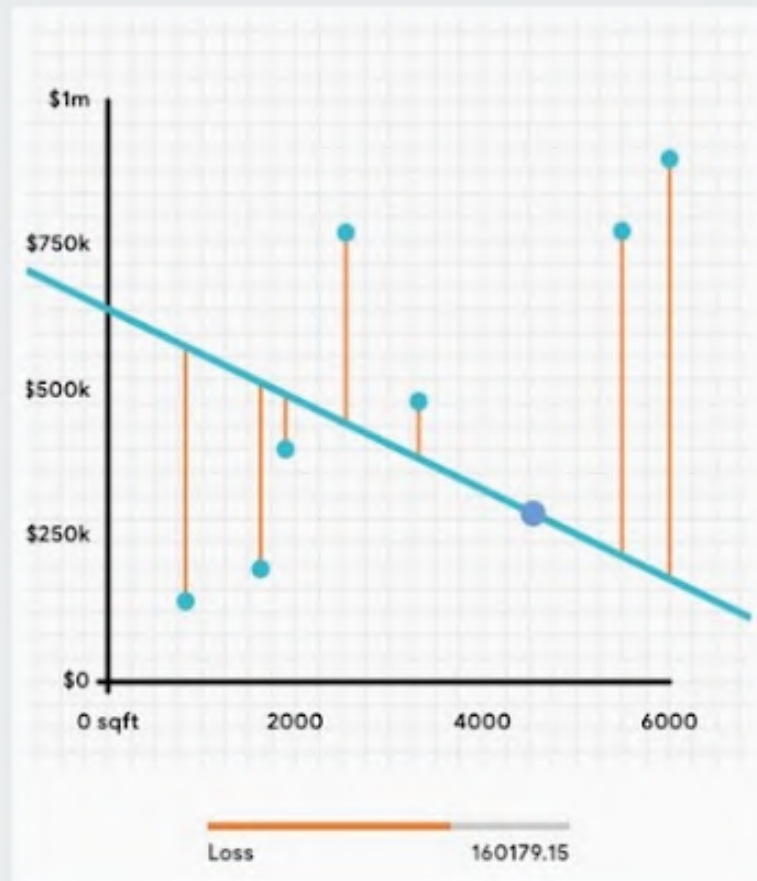
Some examples
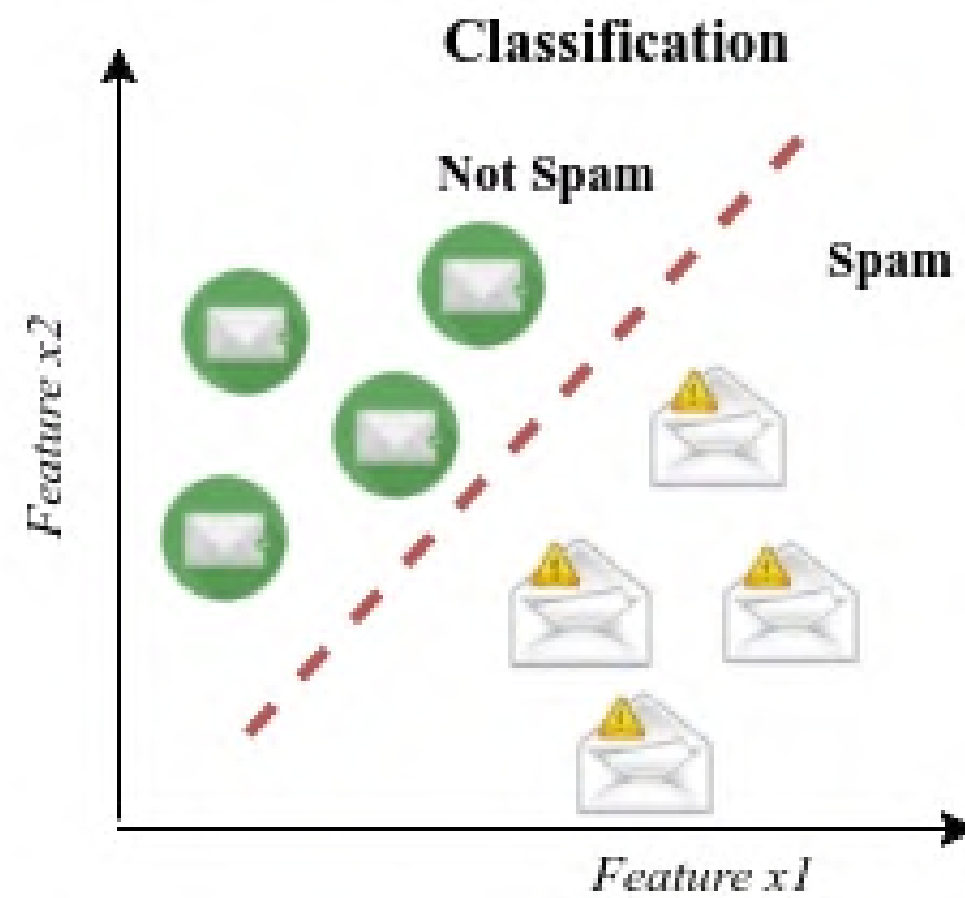
$$Y = mx + c$$

Linear Regression is a machine learning algorithm based on **supervised learning**.

Regression models a target prediction value based on independent variables.

Line of best fit

Line of best fit represents the algorithm which predicts the values with the least error.
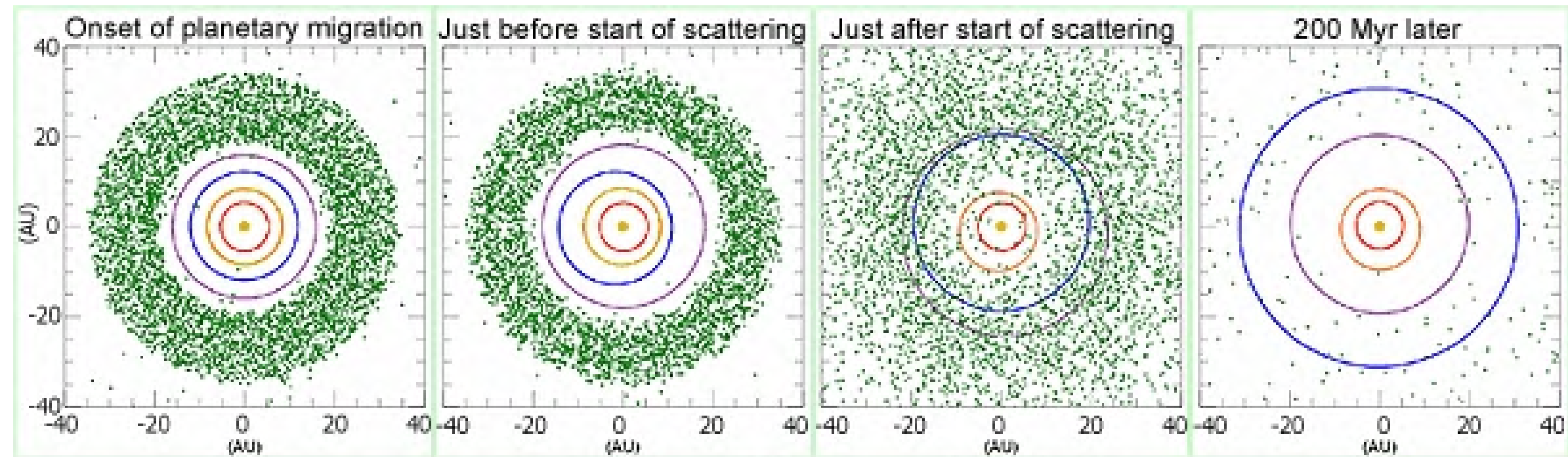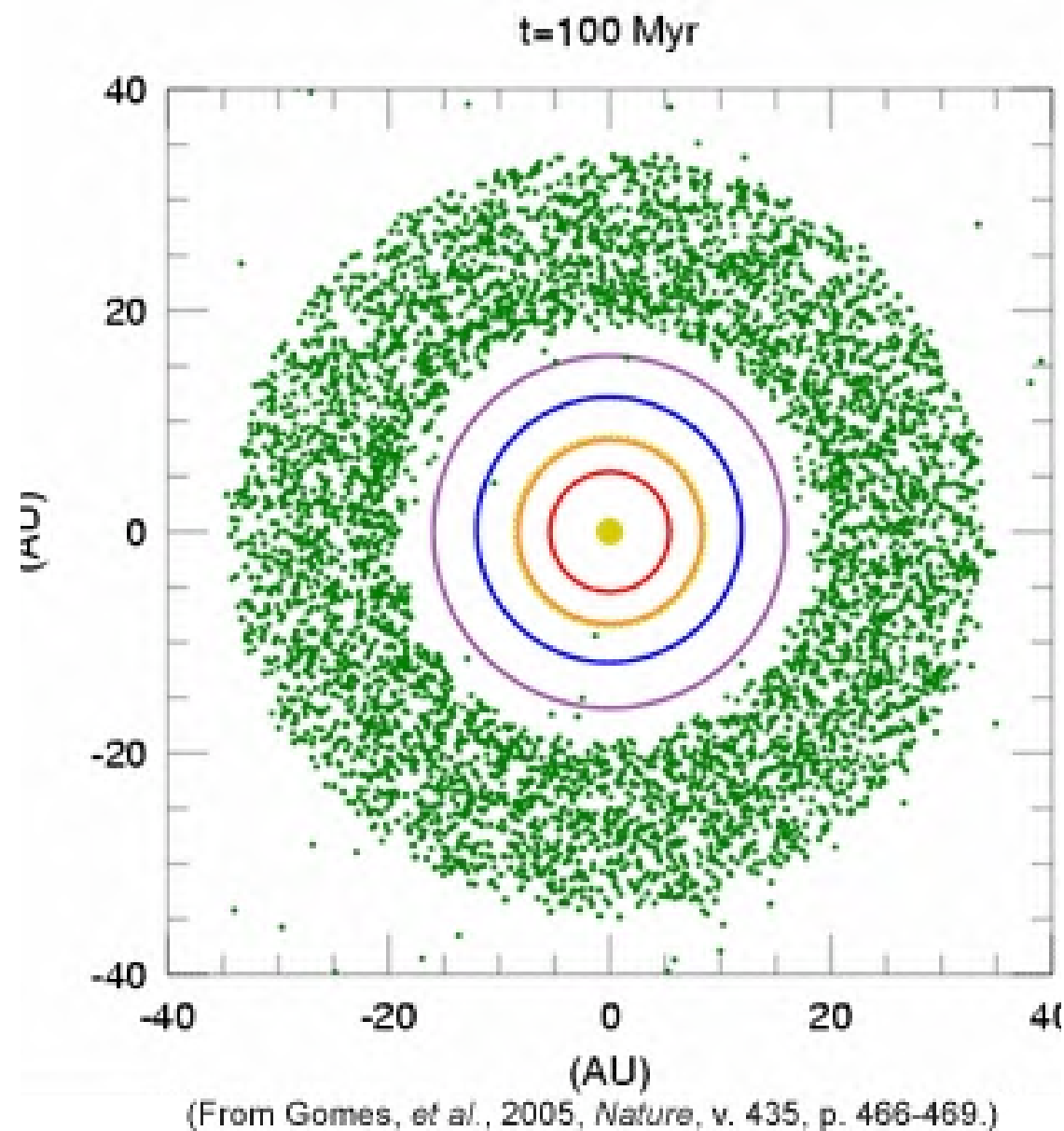
Code!

Finally some code

Some awesome feats scientists have achieved using data science

This is of course, related to space.

Formation of our solar system

(From Gomes, et al., 2005, Nature, v. 435, p. 466-469.)

From Gomes, R., H. F. Levison, K. Tsiganis, and A. Morbidelli (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. Nature, v. 435, p. 466-469.

Wanna read more?

The collision between the Milky Way Galaxy and the Andromeda Galaxy.

Star spaghettification simulation

Thank you!