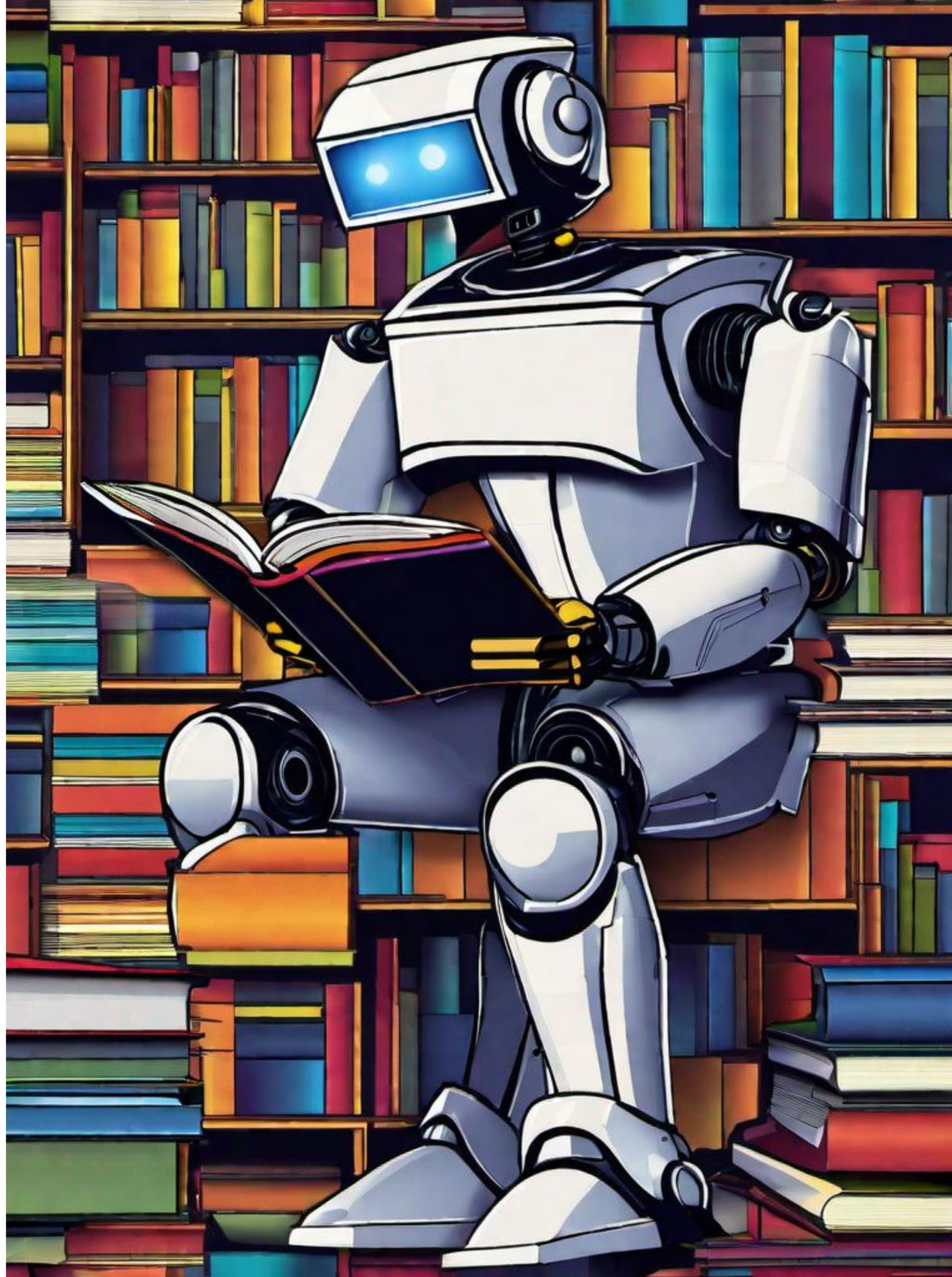
What the hell is Machine Learning

And how can anyone get involved in it?

le wagon



Since 2014

Le Wagon

Has been providing candidates with the skillset necessary to future-proof their career.

A teaching of excellence

Most voted

#1 ranked bootcamp worldwide

LSE HEC PARIS WHU Otto Beisheim School of Management

Official Partners

International campuses

22,000+

People educated at Le Wagon

40

Cities around the world

Focused on practice & product

2300+

Products build

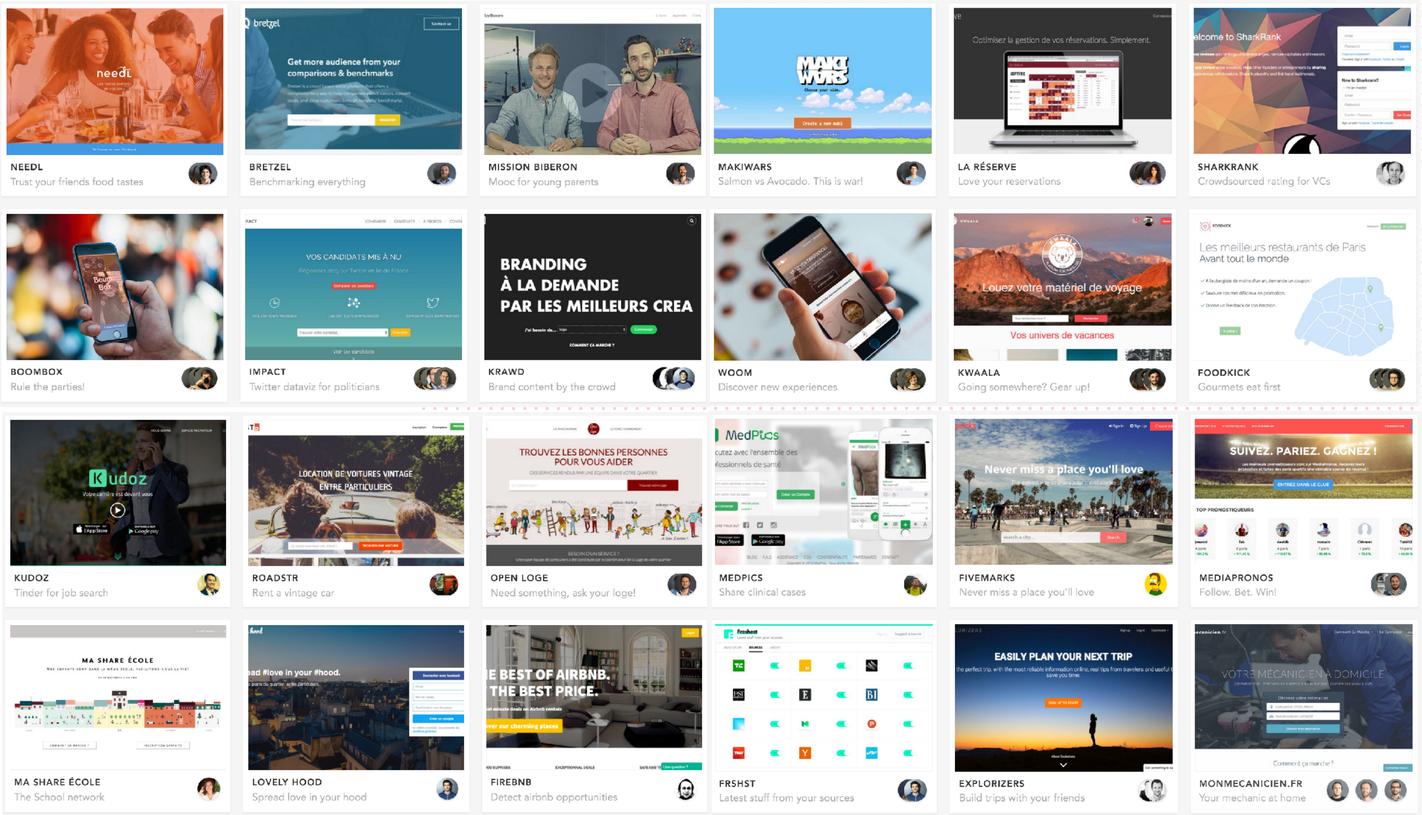
100+

Startups launched

le wagon

Tech is just a tool.

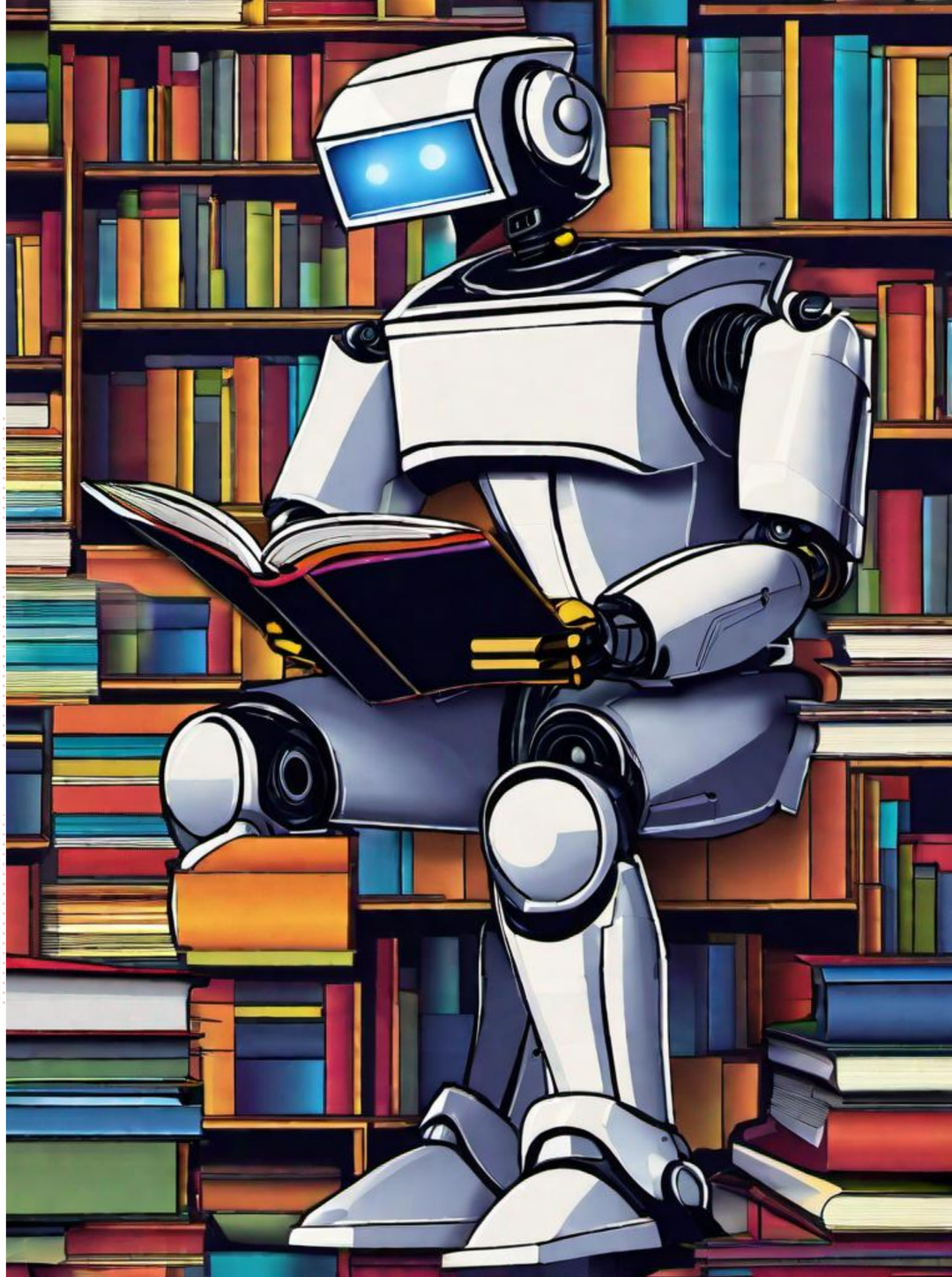
Tools are used for building 🔨



Agenda

- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

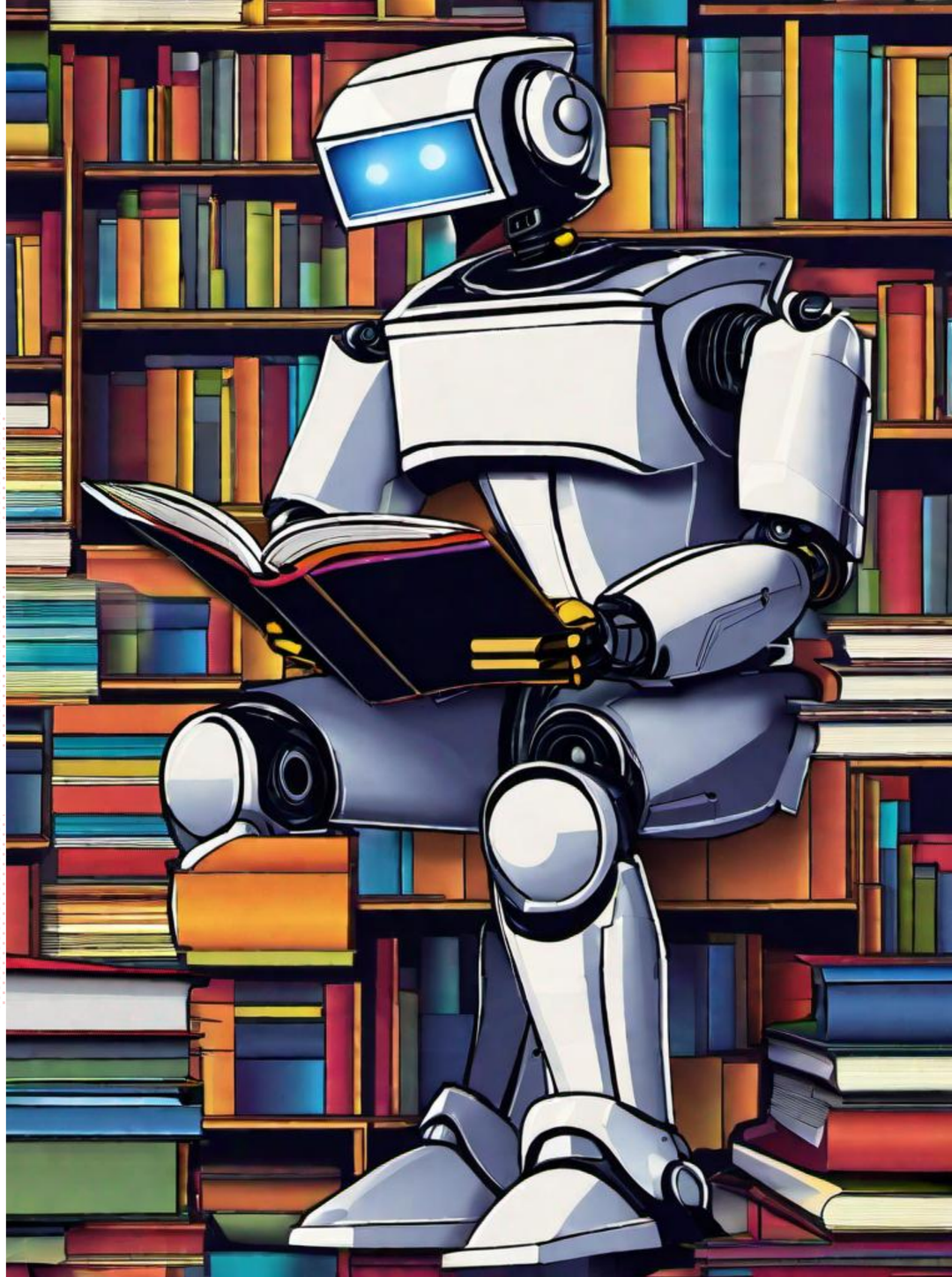
le wagon



Agenda

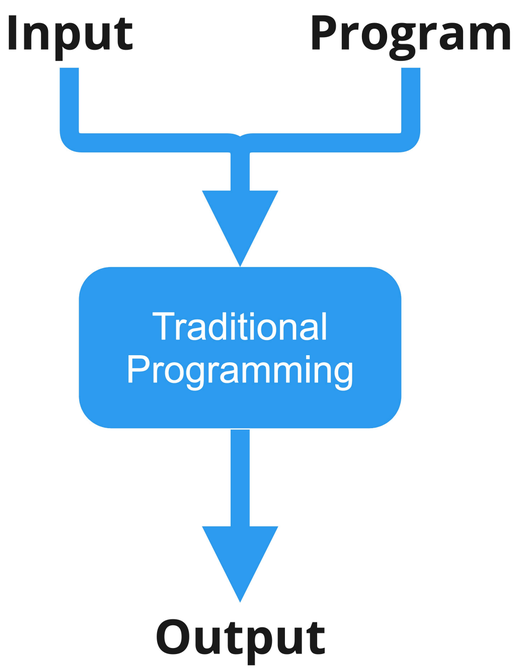
- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

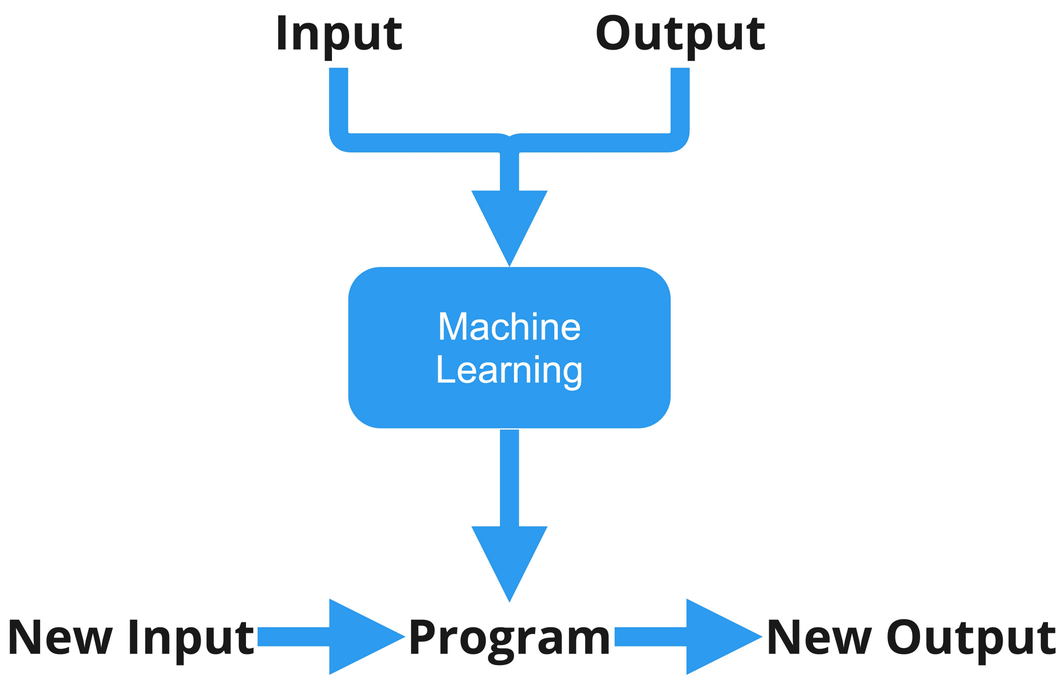
le wagon



“[a] field of study that gives computers the **ability to learn without being explicitly programmed**”

Arthur Samuel (1959)

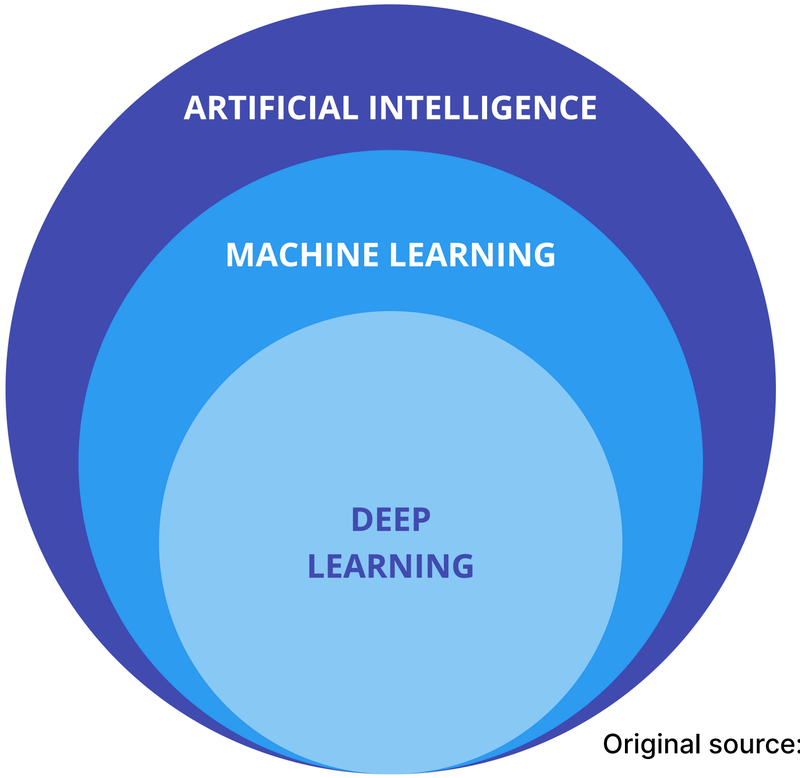




Original source: ScienceDirect

What about terms like

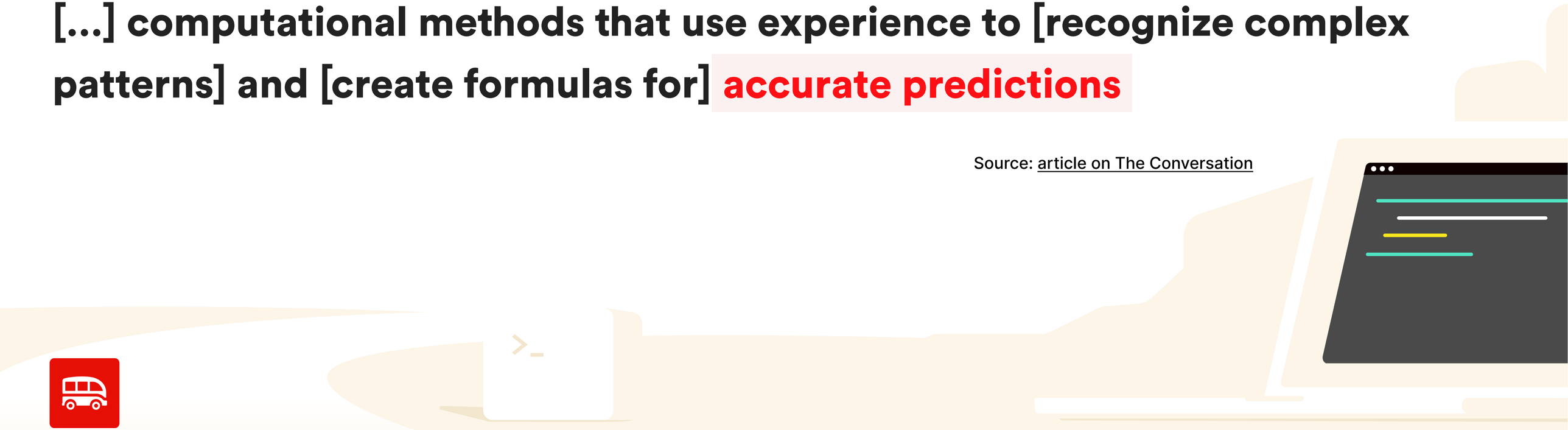
AI or Deep Learning?



Original source: Intel

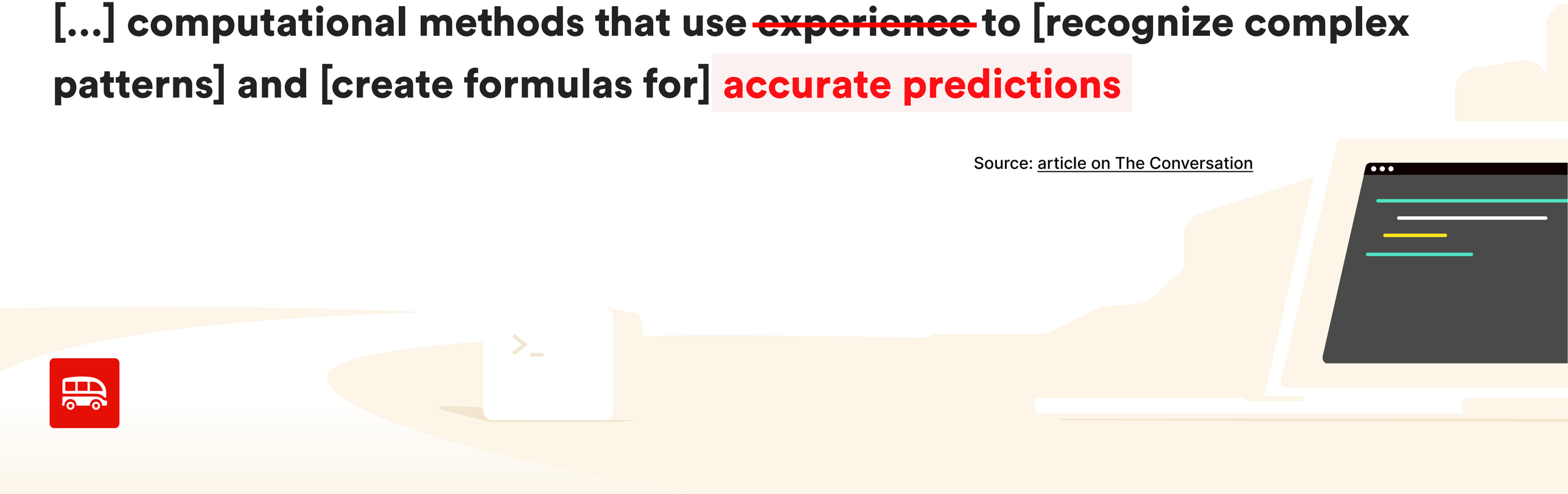
[...] computational methods that use experience to [recognize complex patterns] and [create formulas for] **accurate predictions**

Source: article on The Conversation



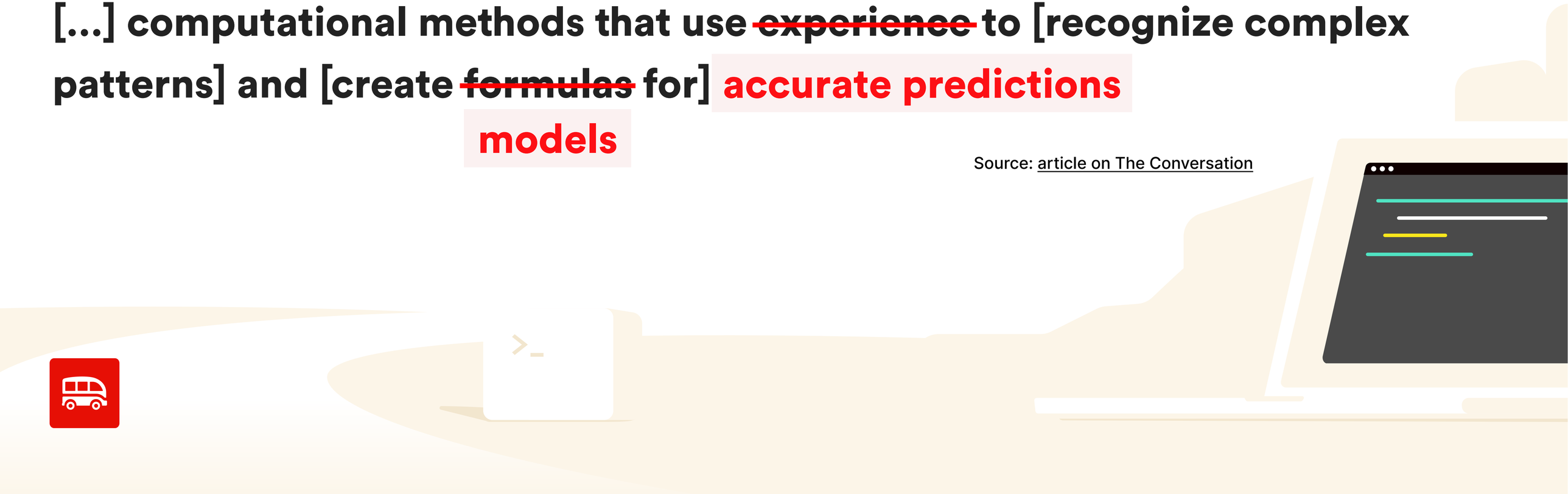
[...] computational methods that use ~~experience~~ data to [recognize complex patterns] and [create formulas for] accurate predictions

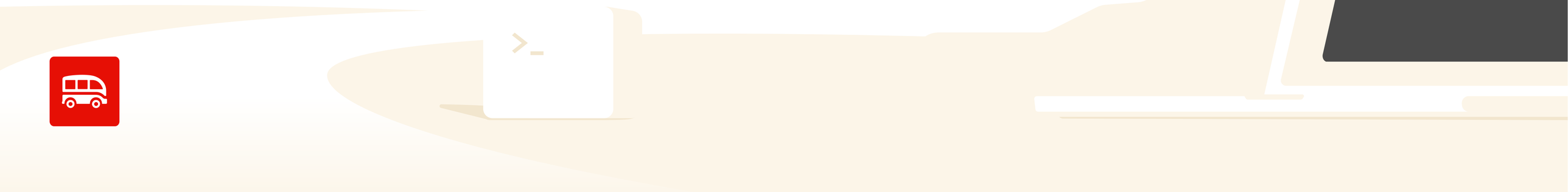
Source: article on The Conversation

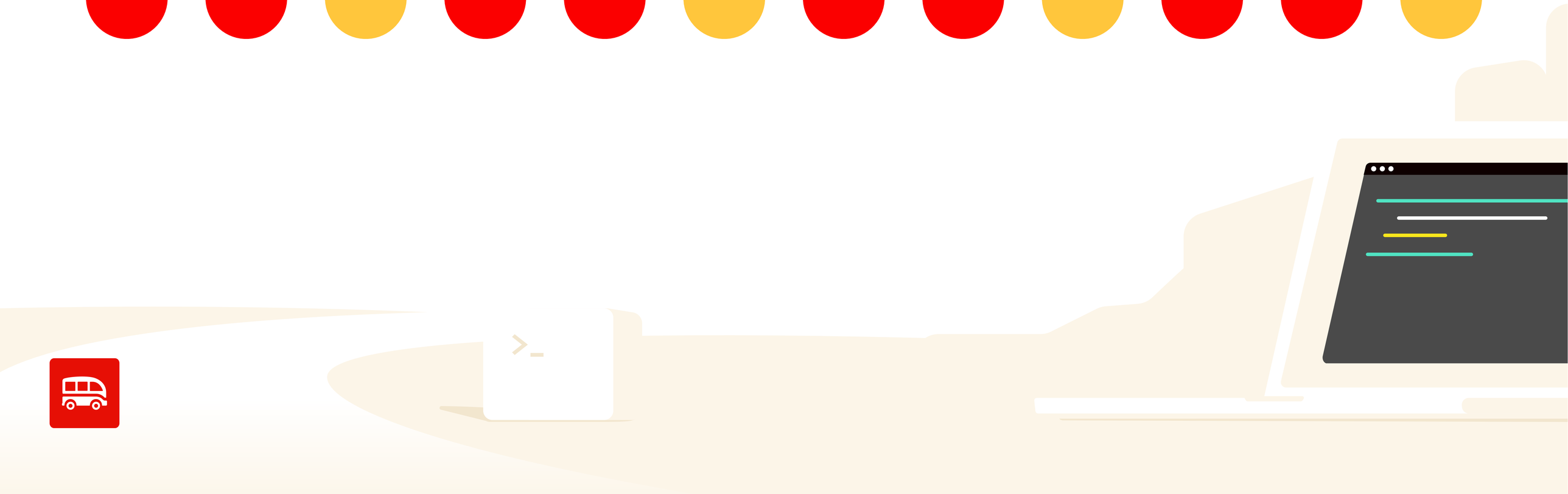


[...] computational methods that use ~~experience~~ data to [recognize complex patterns] and [create ~~formulas~~ models for] accurate predictions

Source: article on The Conversation





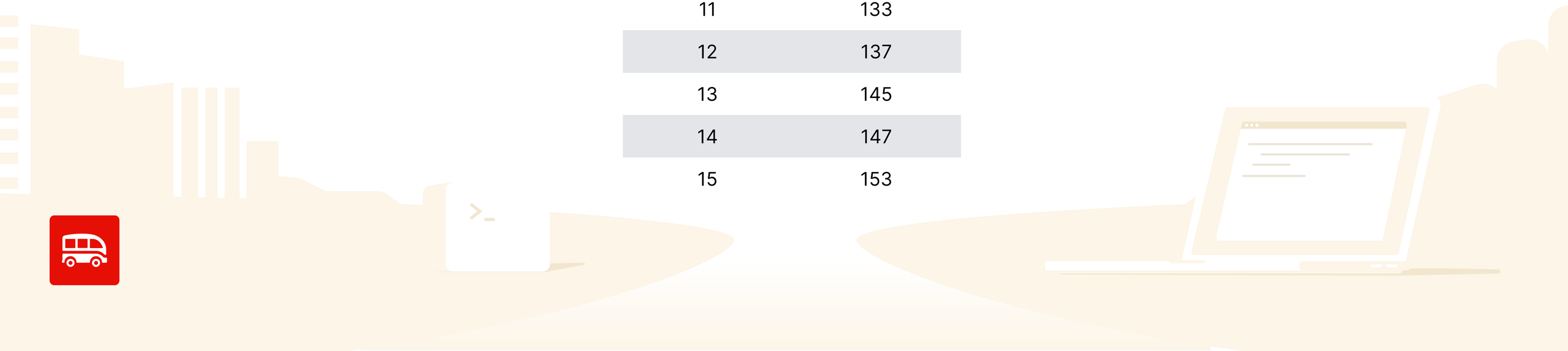


How does Machine Learning work?

Take basic observations

My Age/Height Ratio 📏

Age	Height (cm)
5	111
6	115
7	119
8	122
9	126
10	130
11	133
12	137
13	145
14	147
15	153



Take basic observations

Input

My Age/Height Ratio 📏

Age	Height (cm)
5	111
6	115
7	119
8	122
9	126
10	130
11	133
12	137
13	145
14	147
15	153

Take basic observations

My Age/Height Ratio 📏

Age	Height (cm)
5	111
6	115
7	119
8	122
9	126
10	130
11	133
12	137
13	145
14	147
15	153

Input

Output

Take basic observations

My Age/Height Ratio 📏

Age	Height (cm)
5	111
6	115
7	119
8	122
9	126
10	130
11	133
12	137
13	145
14	147
15	153

Feature

Input

Output

Take basic observations

My Age/Height Ratio 📏

Age	Height (cm)
5	111
6	115
7	119
8	122
9	126
10	130
11	133
12	137
13	145
14	147
15	153

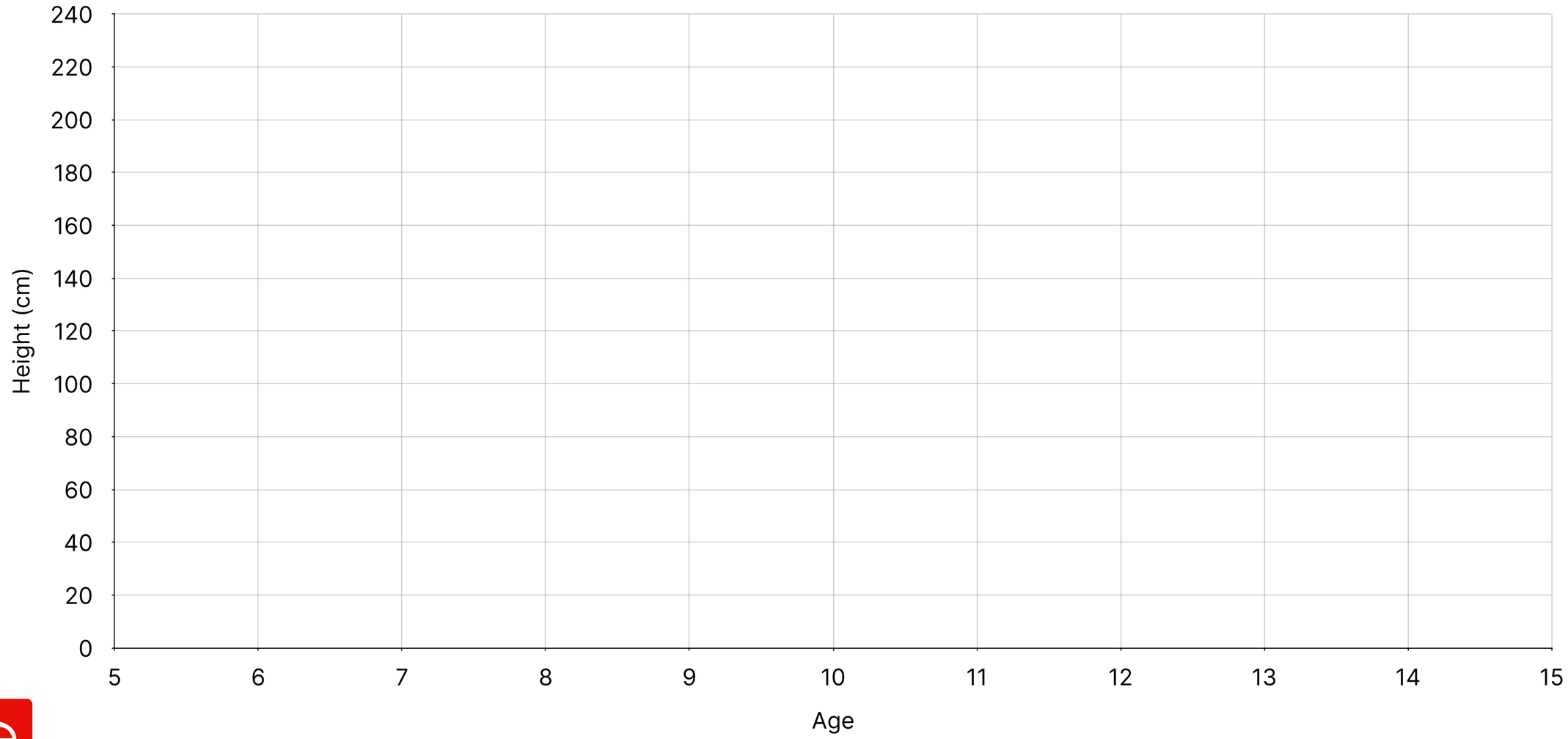
Feature

Input

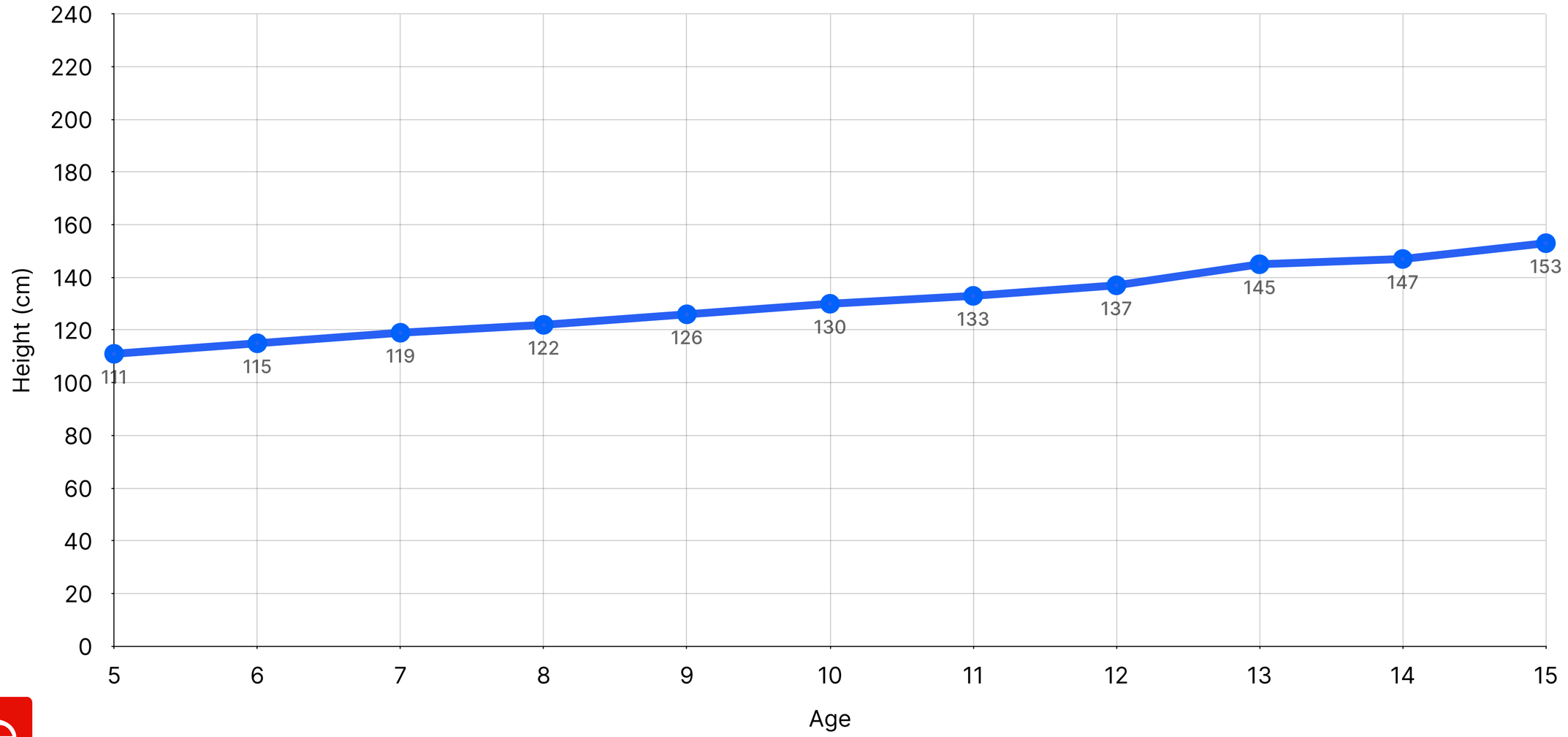
Target

Output

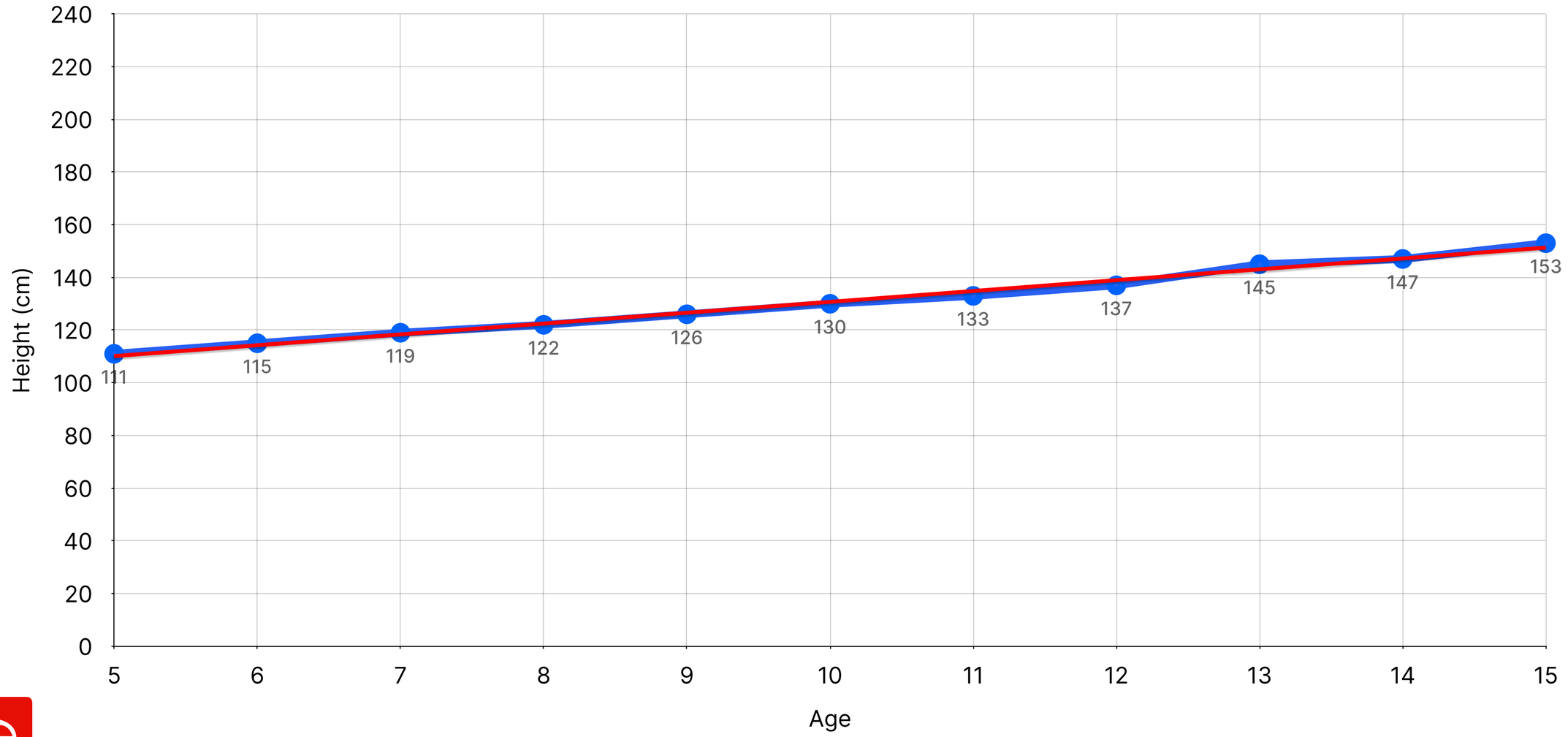
Infer a mathematical function



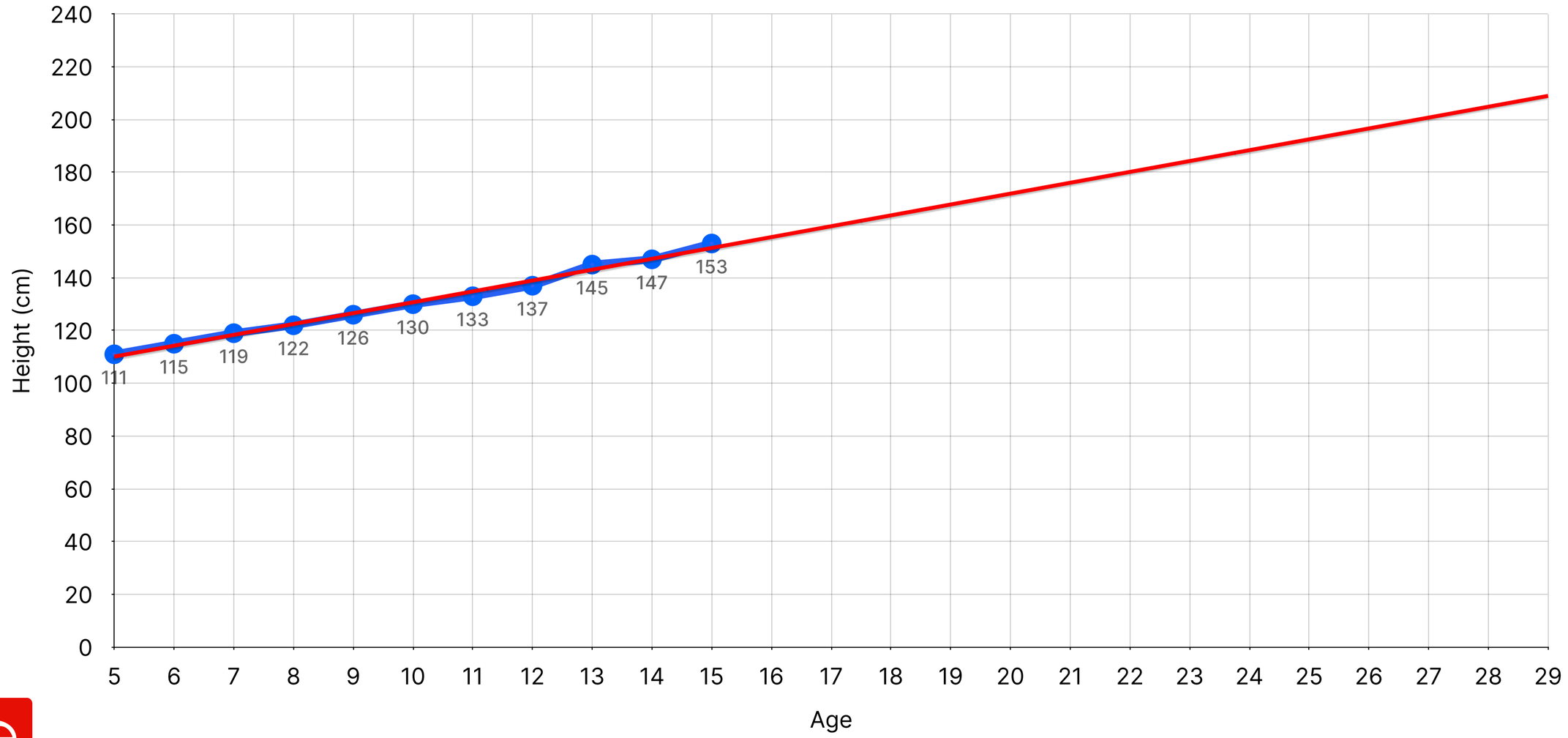
Infer a mathematical function



Infer a mathematical function



Extrapolate new observations

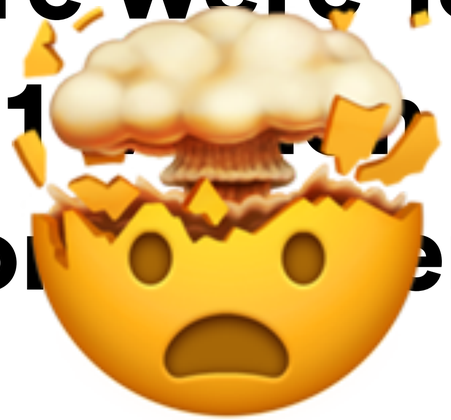


What if there were 150 different features and 1 Million observations for 2 Million different people?

What if there were 150 different features and 1 Million observations for 2 Million different people?

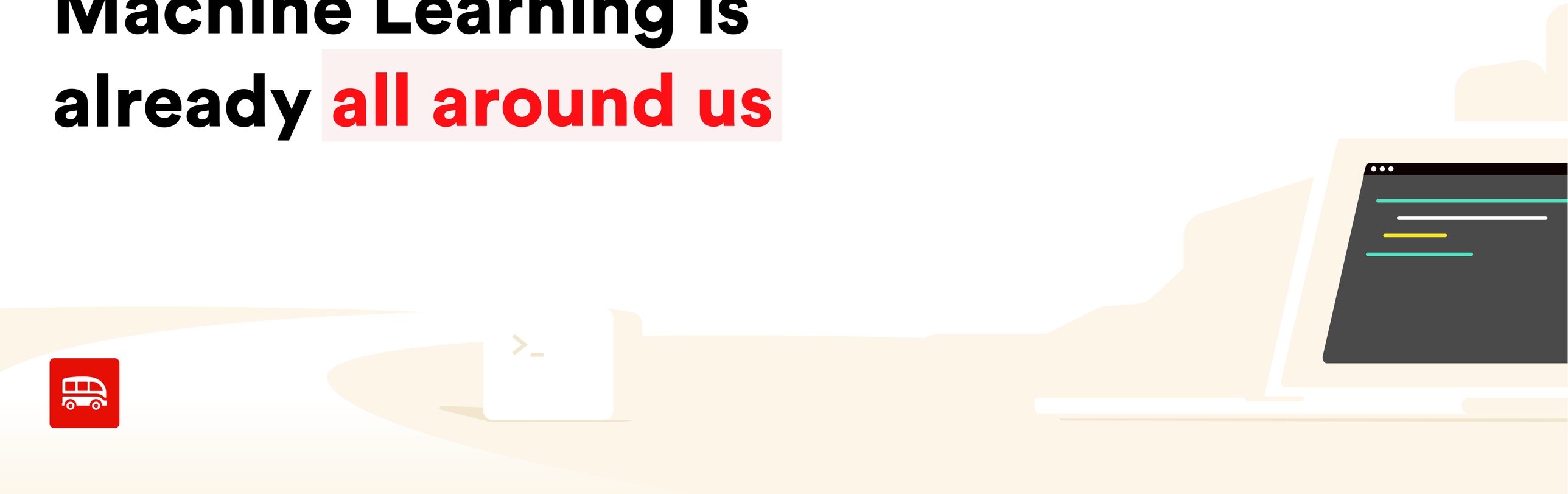
Every single day!

What if there were 150 different features and 1[illegible] observations for 2 Millio[illegible]ent people?



Every single day!

Machine Learning is already all around us

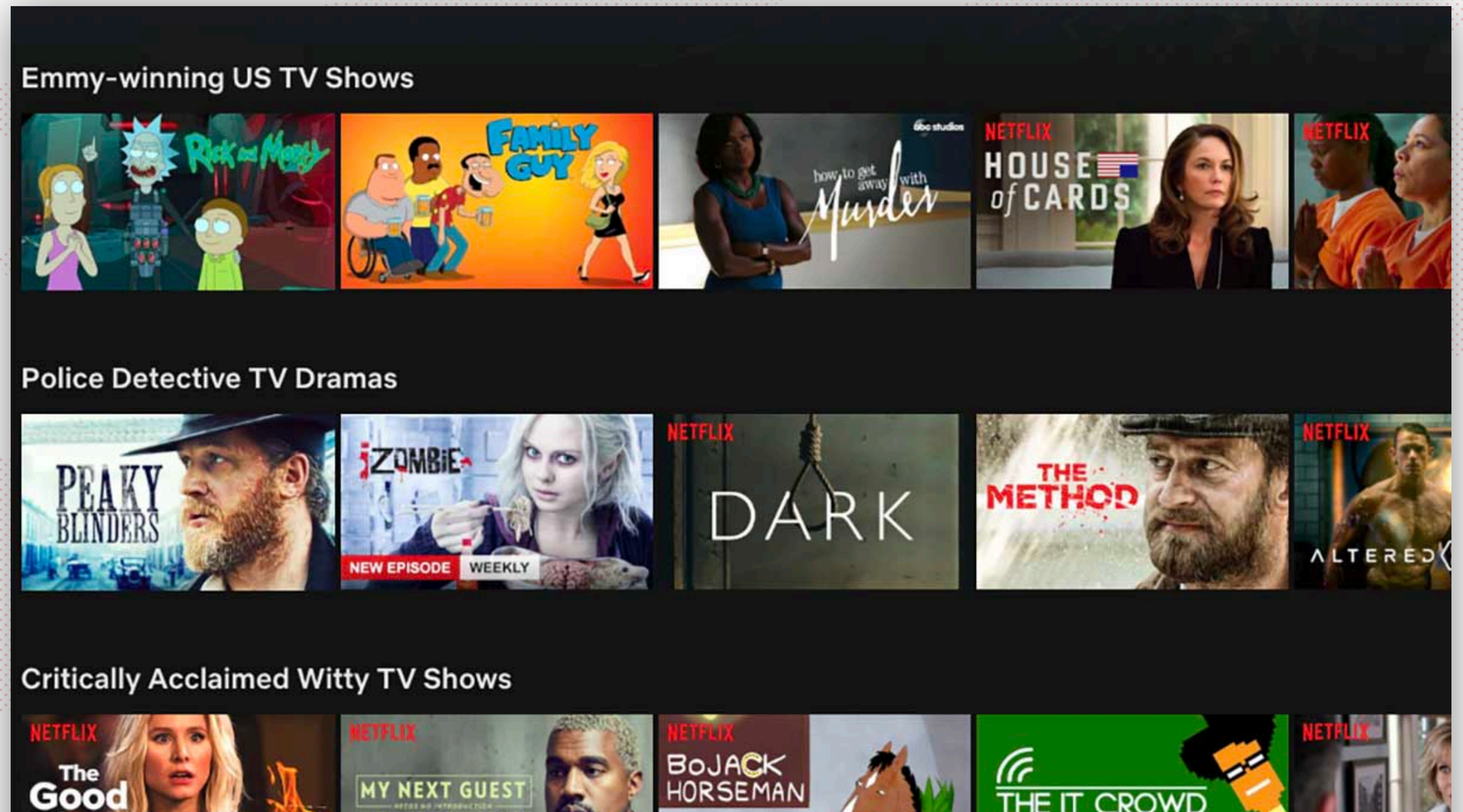


Computer Vision

Classifying and detecting visual data

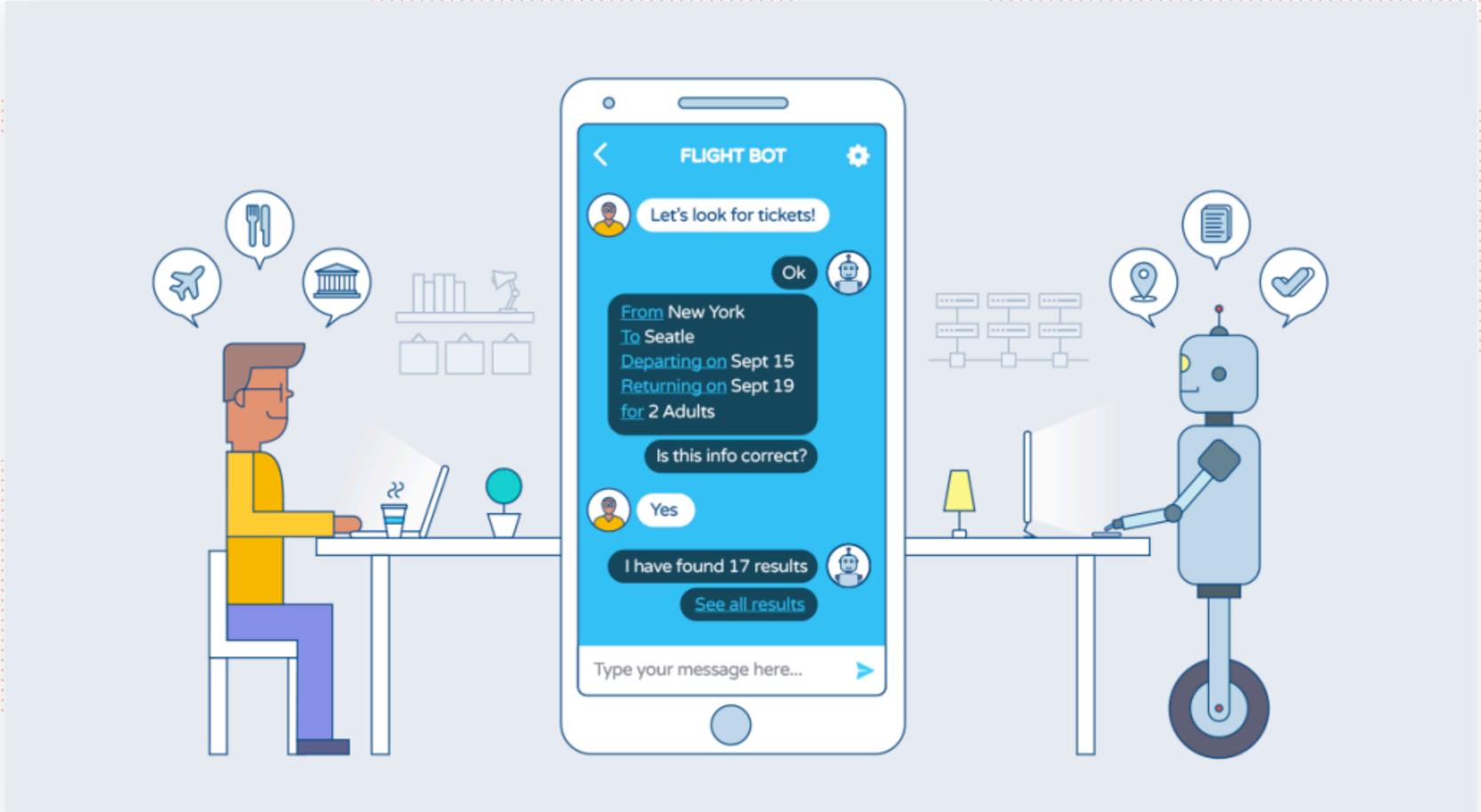
Recommendation Engine

Predicting your next action



Natural Language Processing

Finding meaning in text



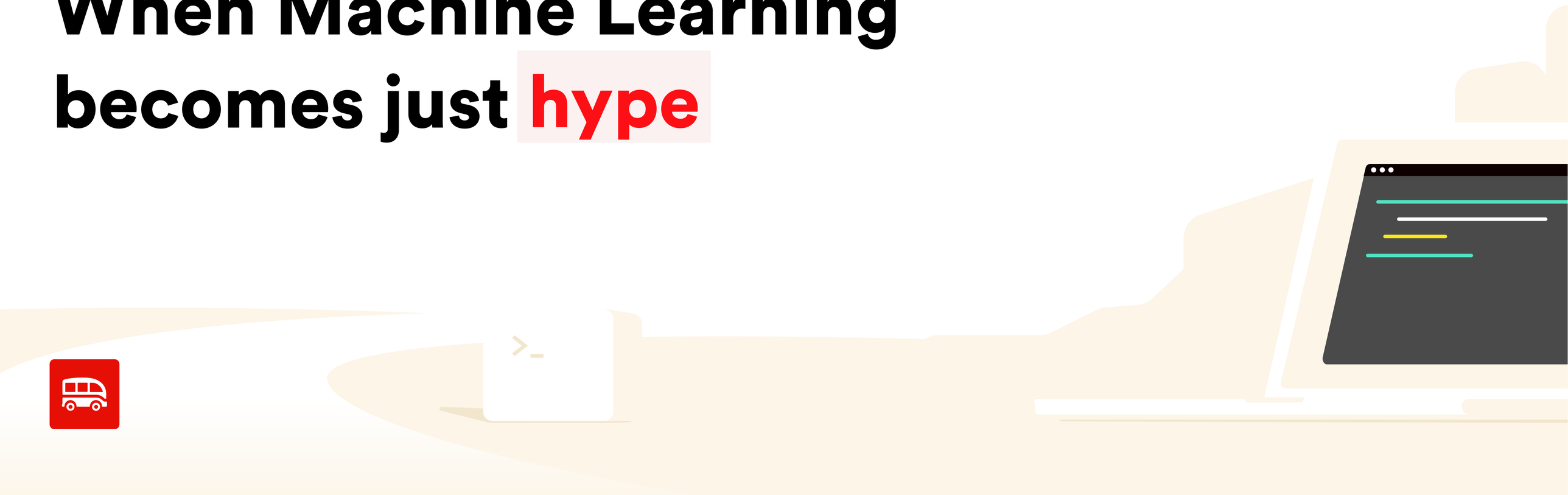
Time Series Analysis

Can past changes predict future changes?

Anomaly Detection

"Reverse-predicting" patterns

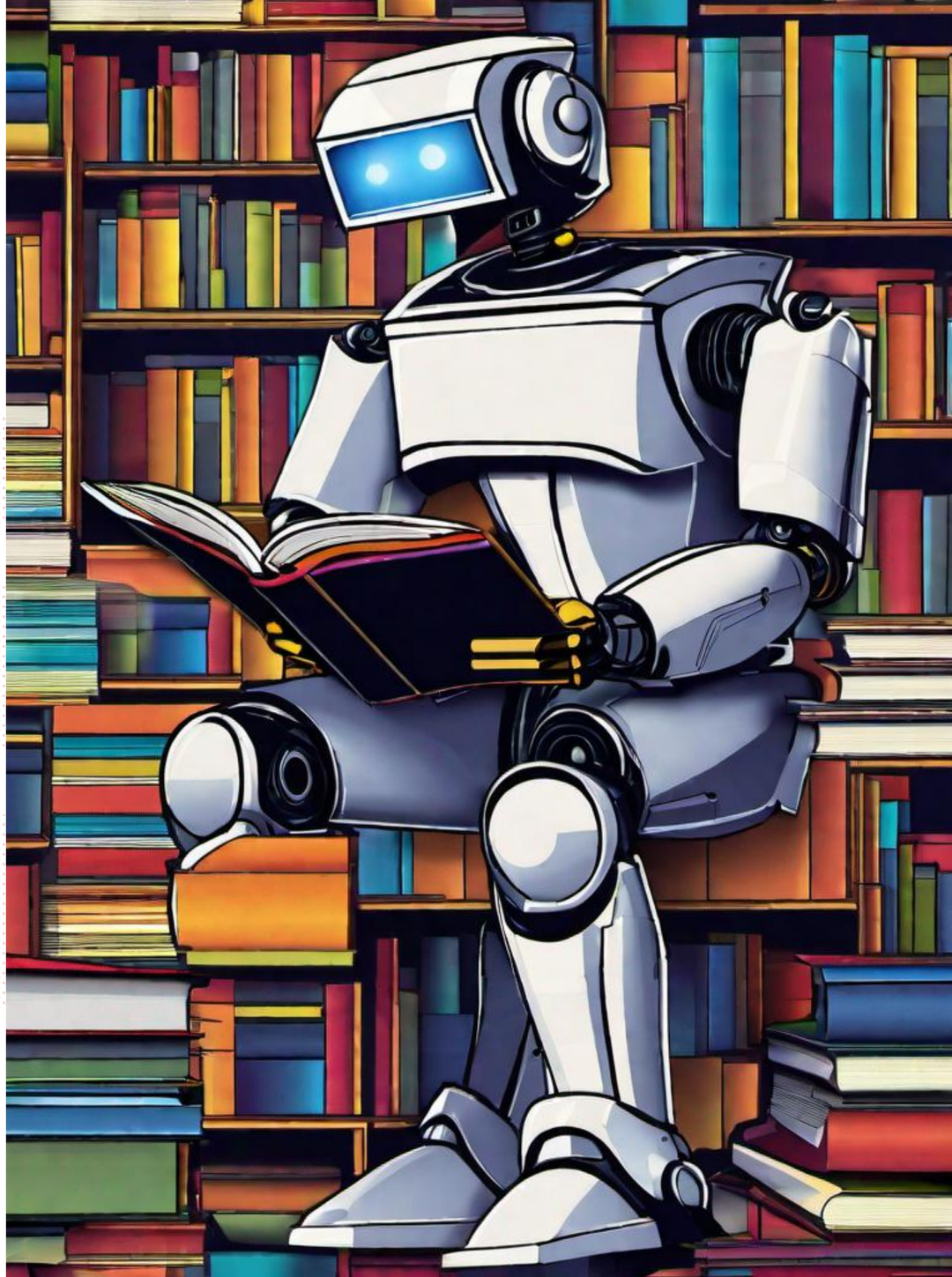
When Machine Learning becomes just hype



Agenda

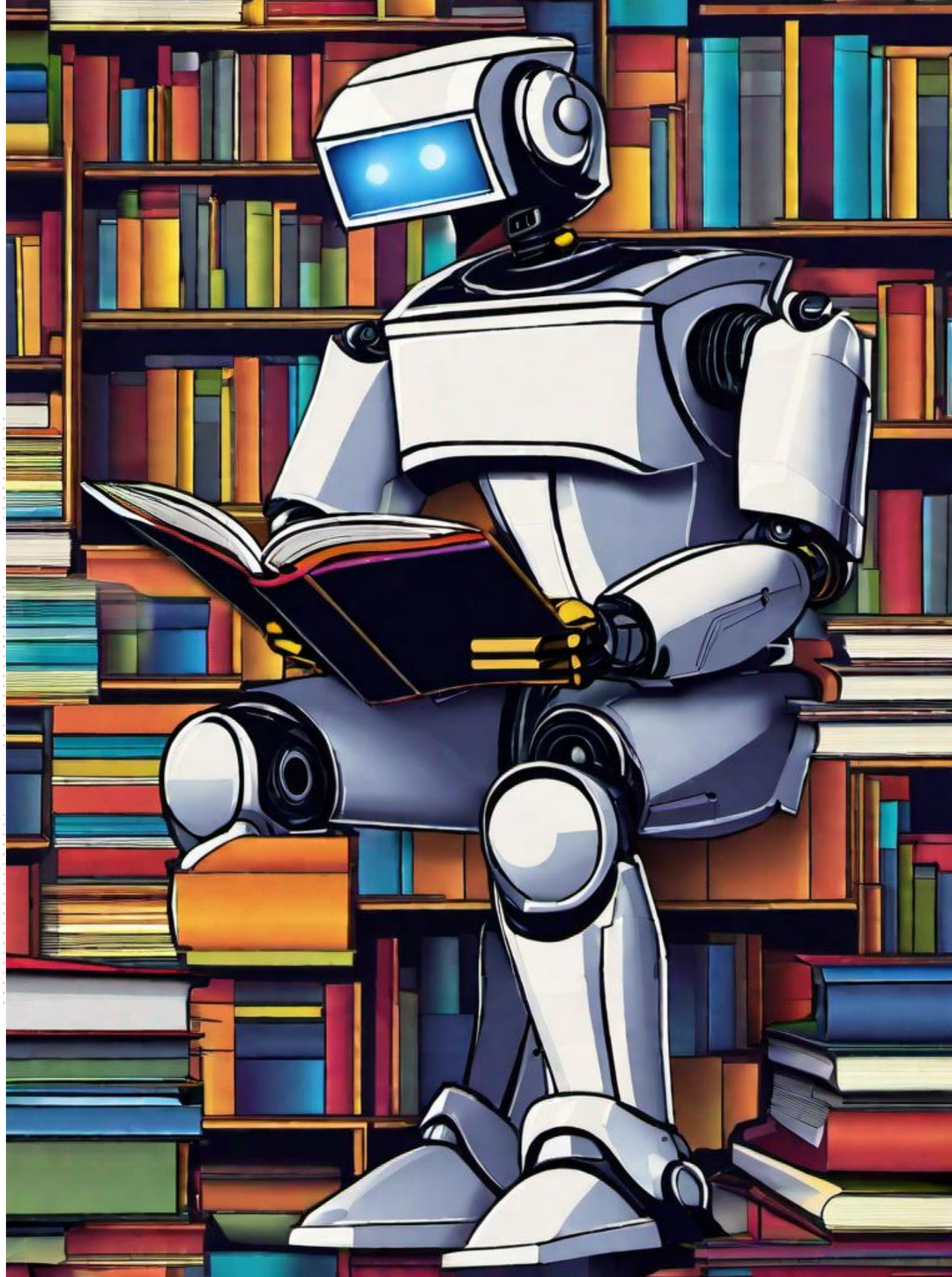
- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

le wagon



Agenda

- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover



le wagon

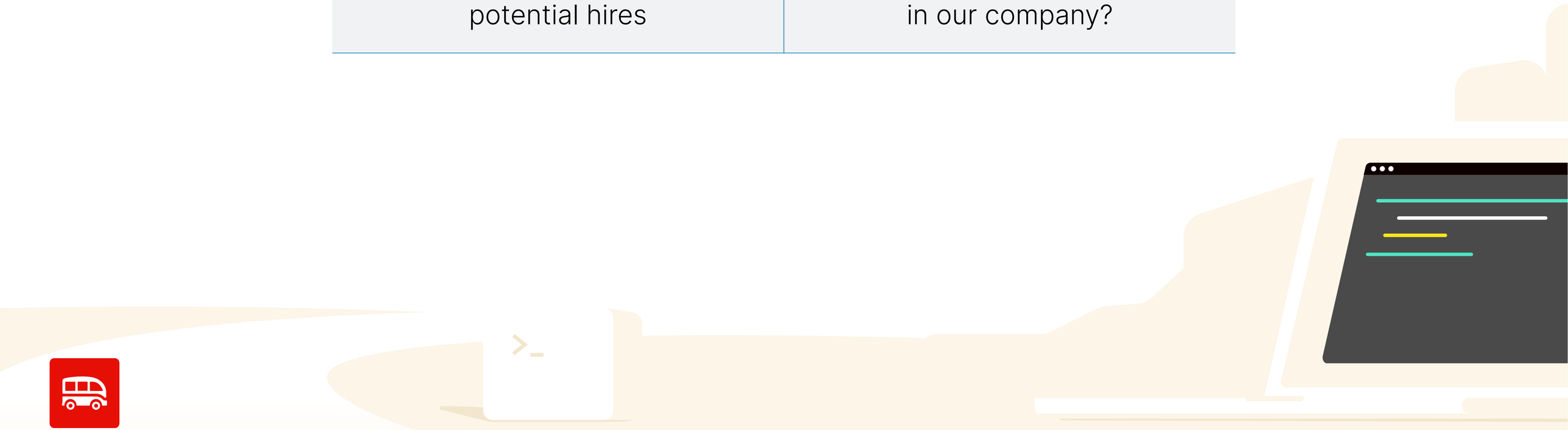
Analytics or Machine Learning?

Analytics or Machine Learning?

Needs ML	Does not need ML

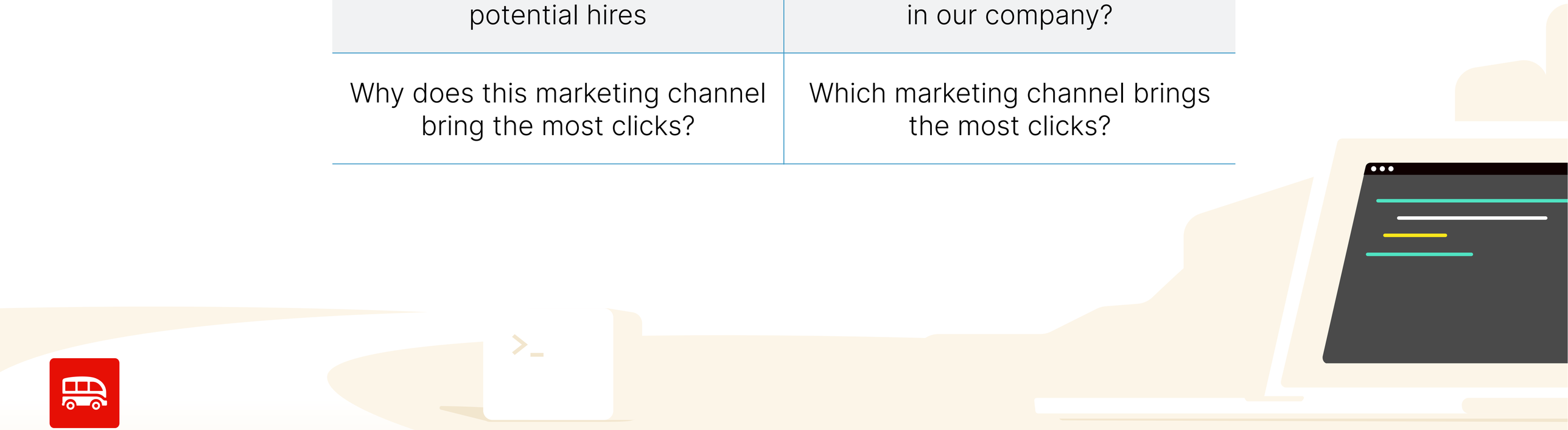
Analytics or Machine Learning?

Needs ML	Does not need ML
I want to predict the salaries of potential hires	Do women and men earn the same in our company?



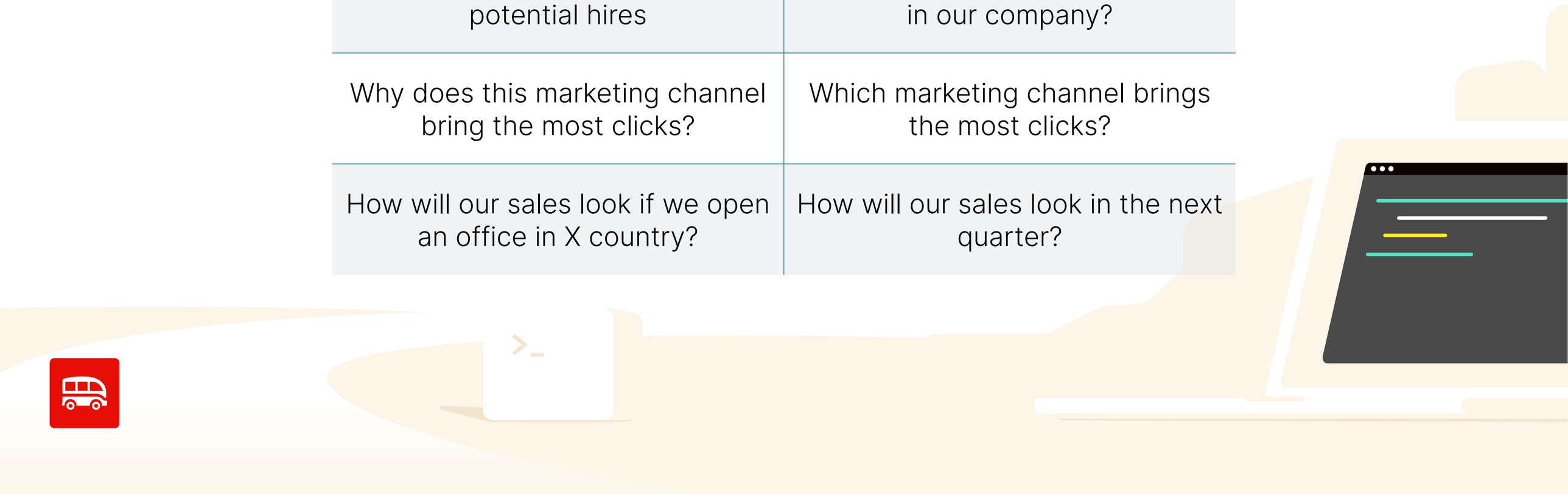
Analytics or Machine Learning?

Needs ML	Does not need ML
I want to predict the salaries of potential hires	Do women and men earn the same in our company?
Why does this marketing channel bring the most clicks?	Which marketing channel brings the most clicks?

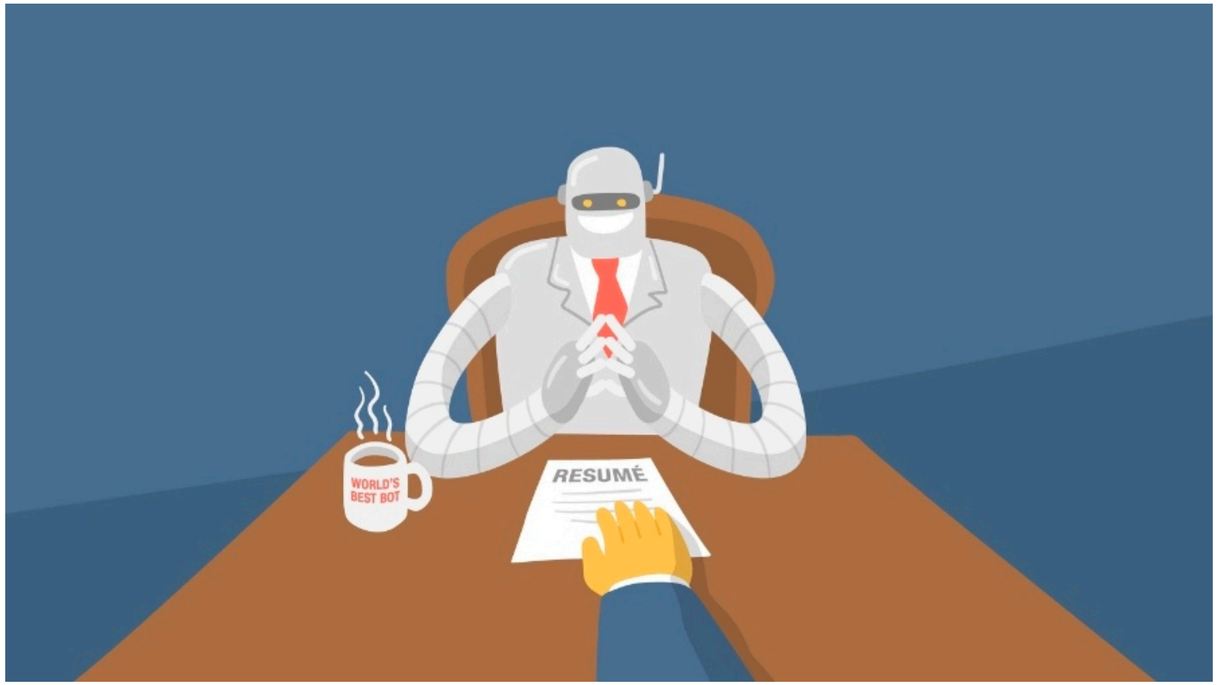


Analytics or Machine Learning?

Needs ML	Does not need ML
I want to predict the salaries of potential hires	Do women and men earn the same in our company?
Why does this marketing channel bring the most clicks?	Which marketing channel brings the most clicks?
How will our sales look if we open an office in X country?	How will our sales look in the next quarter?



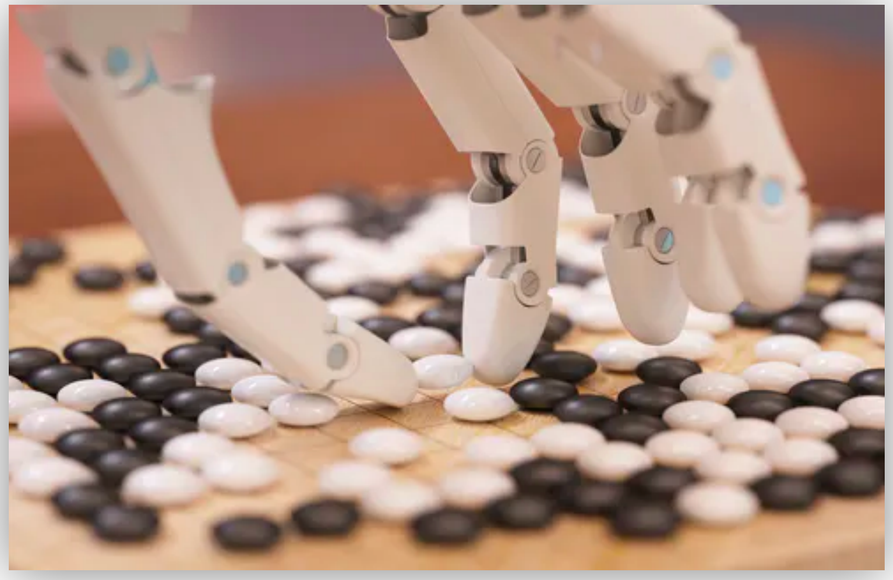
Missing ground truth



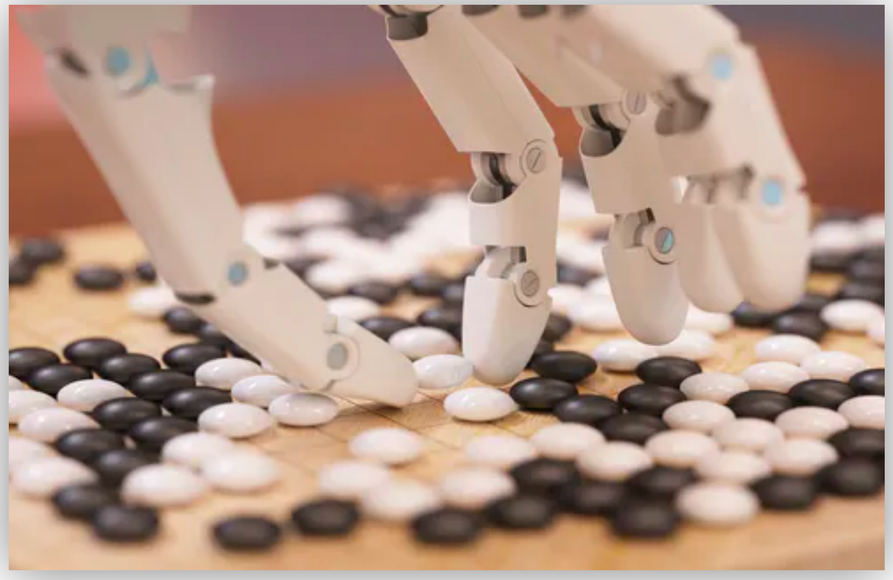
Machine Learning for hiring the perfect candidate!

Missing ground truth

Missing ground truth

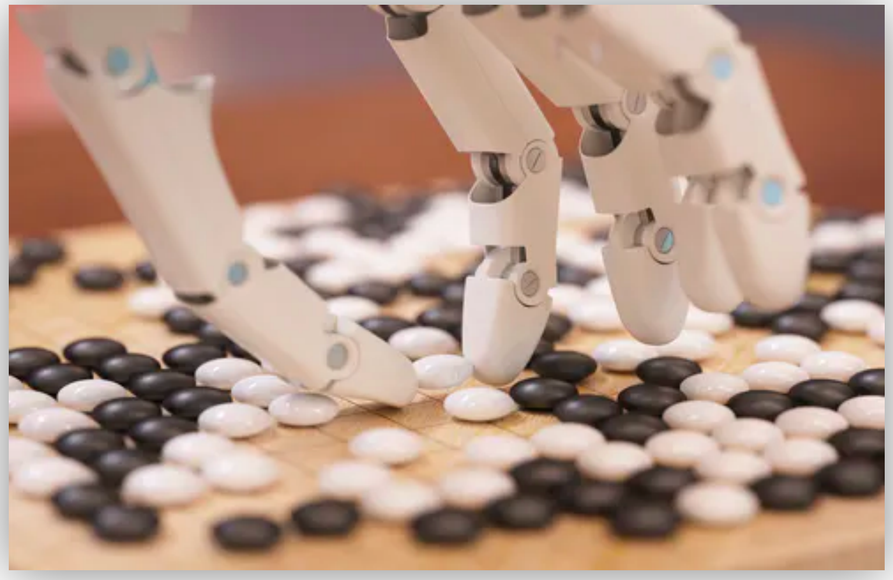


Missing ground truth



vs

Missing ground truth



VS

Things have changed

Things have changed

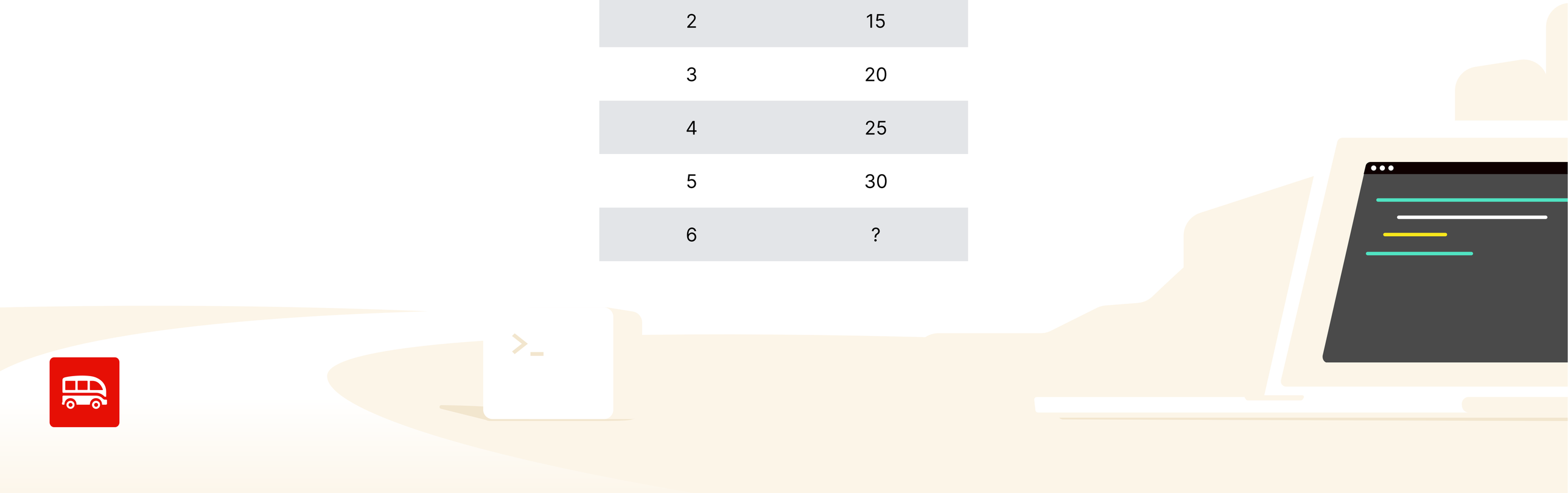
Let's take a drug dose analysis 💊

Things have changed

Let's take a drug dose analysis 💊

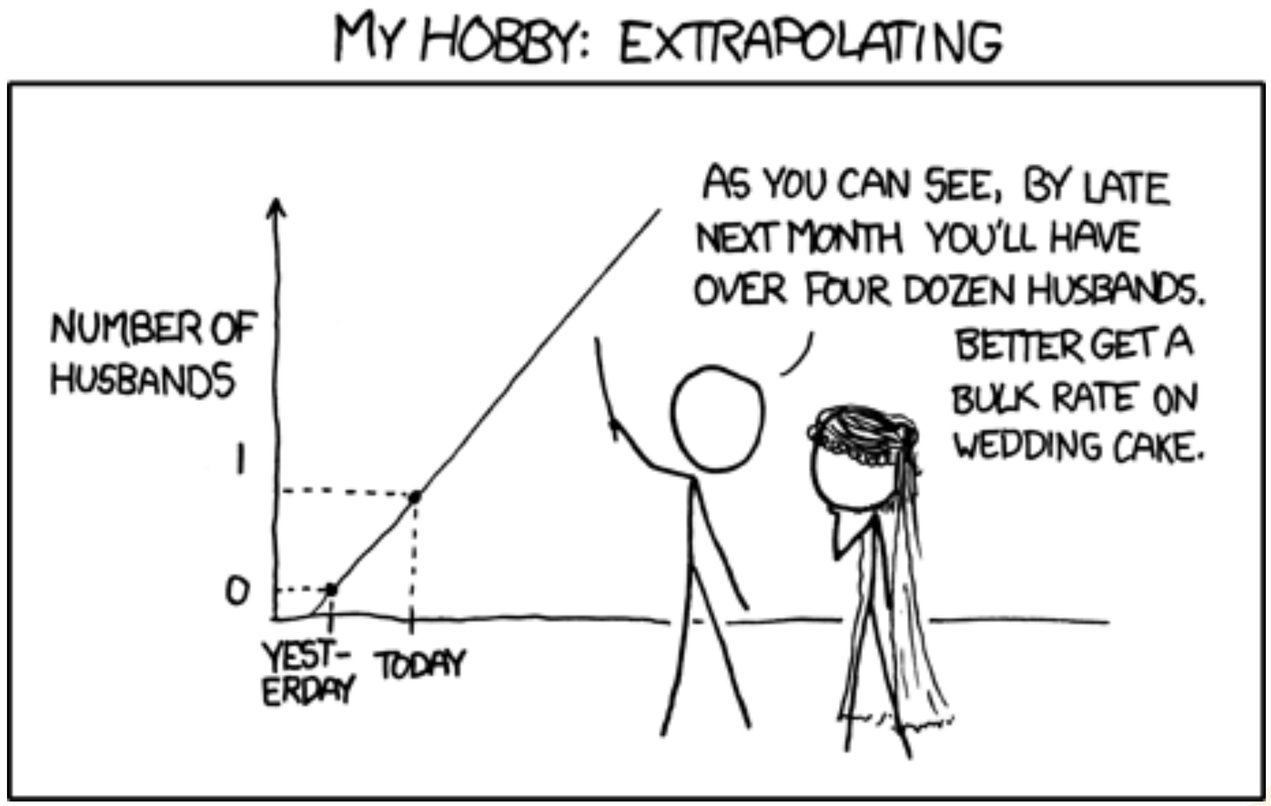
Drug Dosage per Day 💊

Day	Dosage (mg)
1	10
2	15
3	20
4	25
5	30
6	?



Things have changed

Let's take a drug dose analysis 💊



Source: xkcd

Things have really changed

TECH | CORONAVIRUS

The algorithms big companies use to manage their supply chains don’t work during pandemics

The data the algorithms use isn’t reliable

By Nicole Wetsman | Apr 27, 2020, 1:25pm EDT

Listen to this article

SHARE

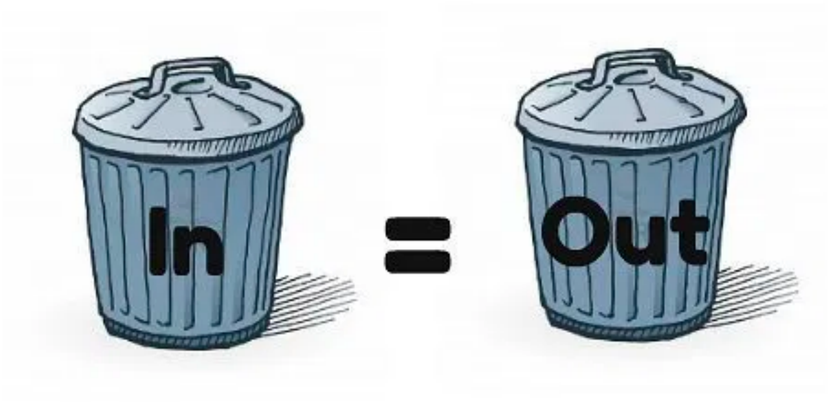
AD

[...] Most firms that think they want advanced AI/ML really just need linear regression on cleaned-up data.

— Tweet by Robin Hanson (@robinhanson)

[...] Most firms that think they want advanced AI/ML really just need linear regression on cleaned-up data.

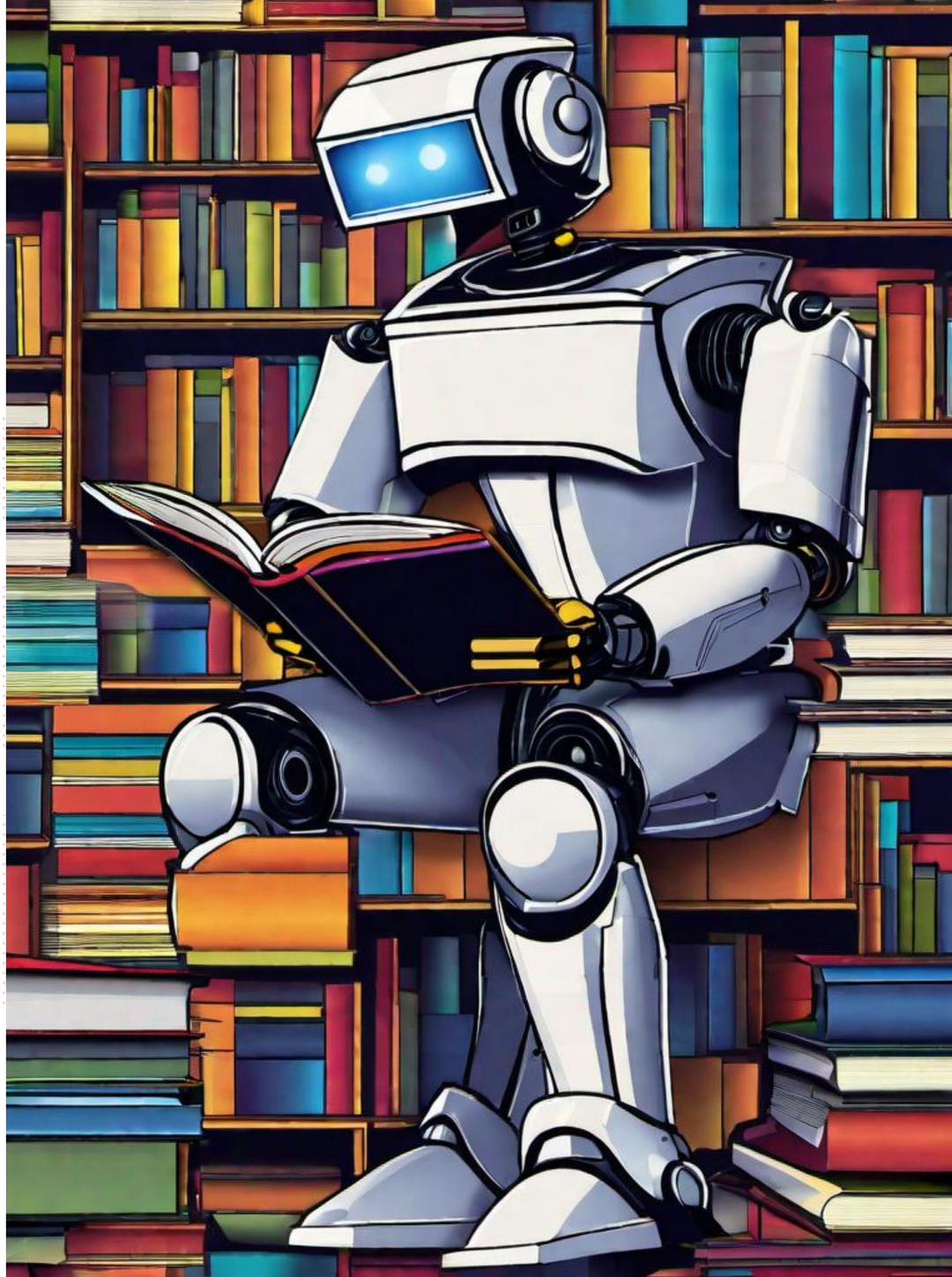
— Tweet by Robin Hanson (@robinhanson)



Agenda

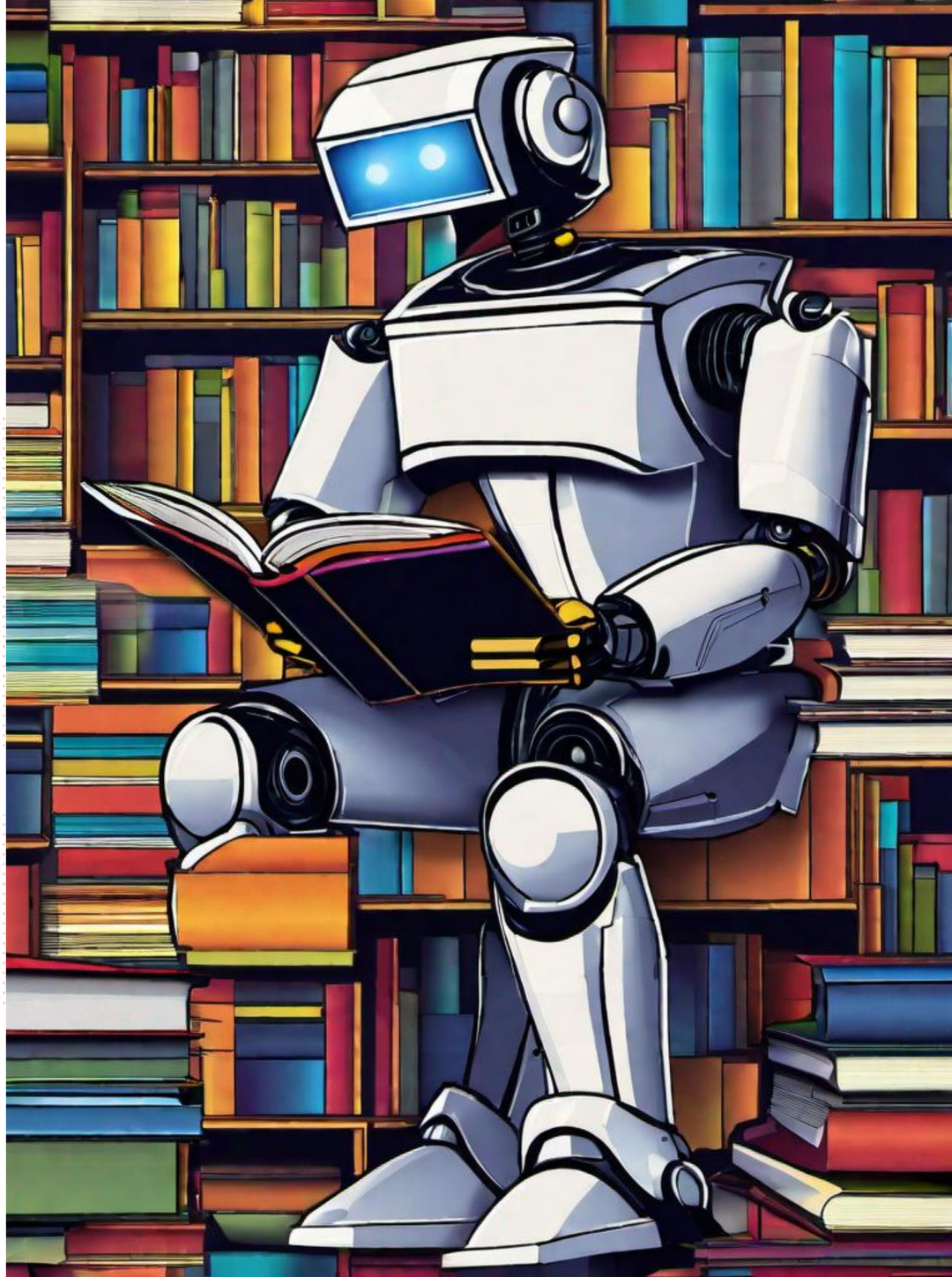
- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

le wagon

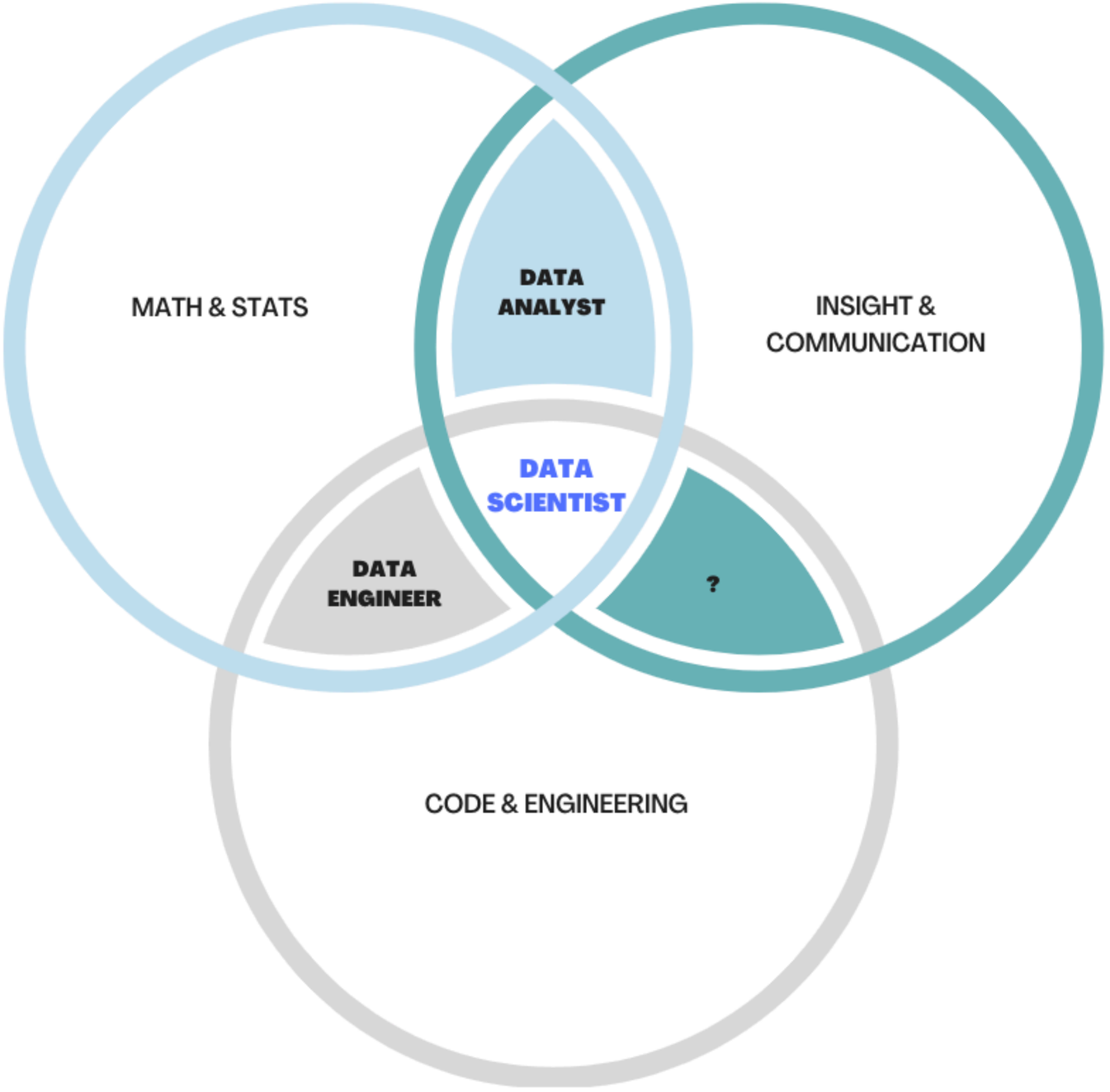


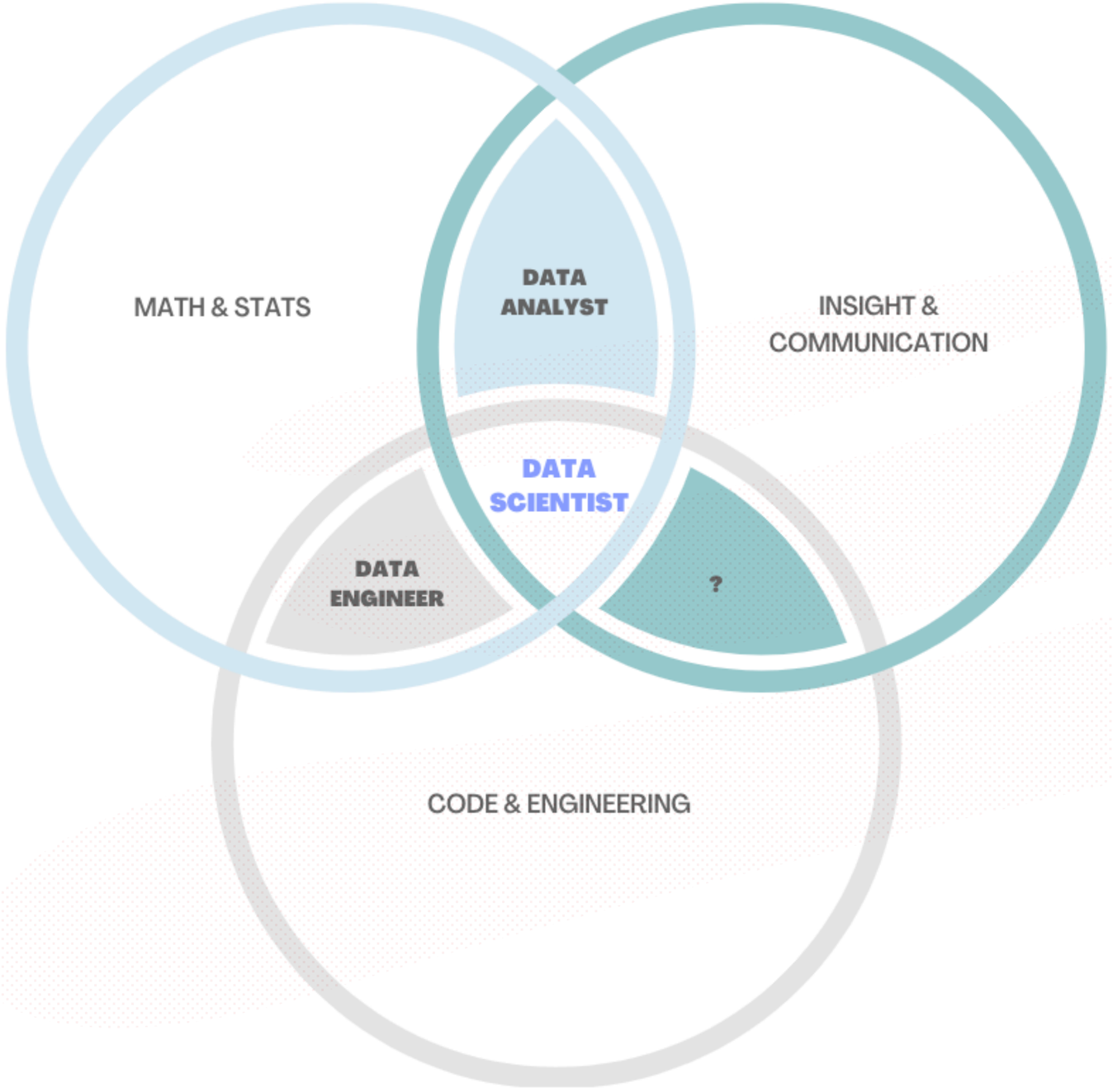
Agenda

- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

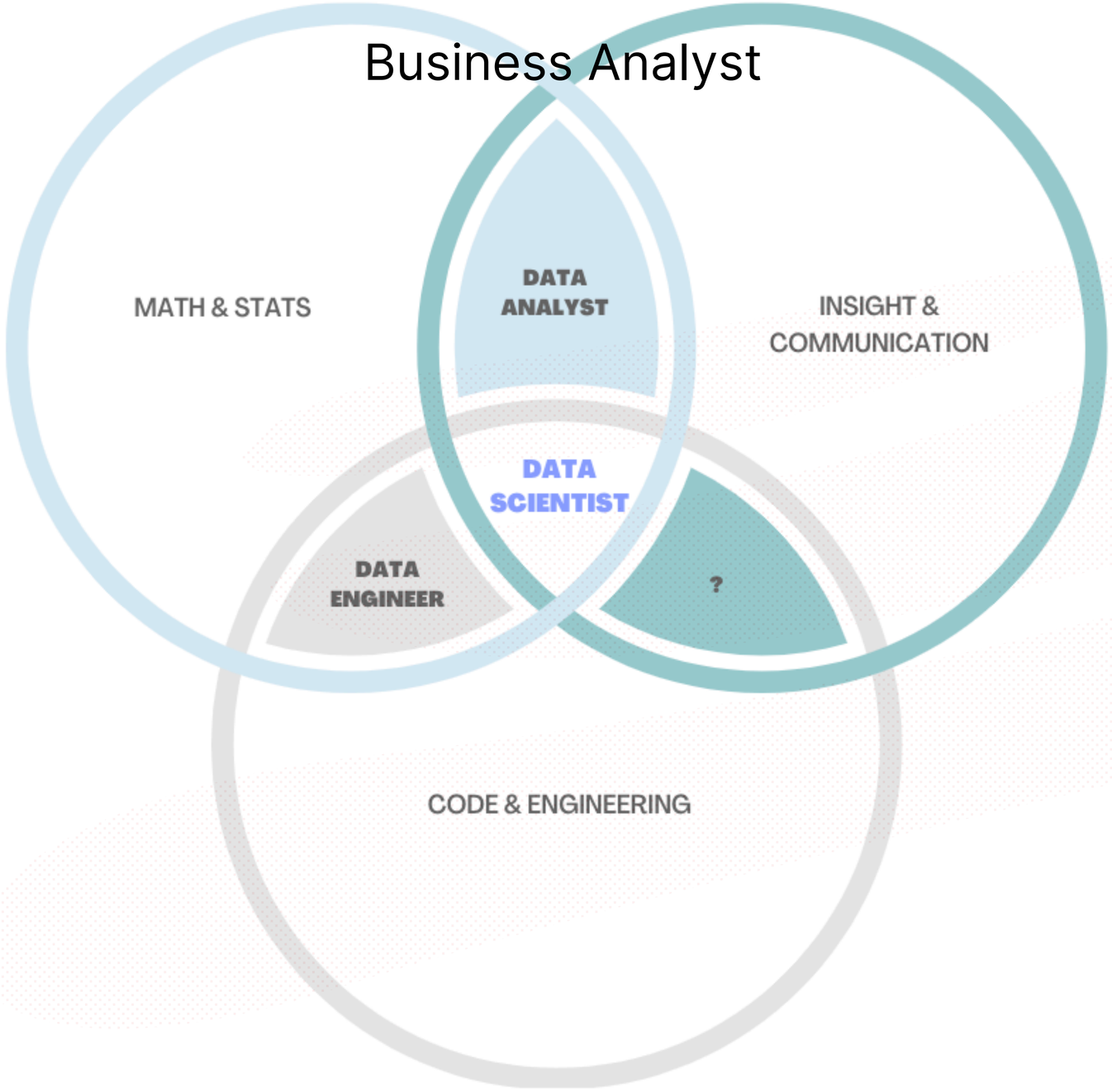


le wagon

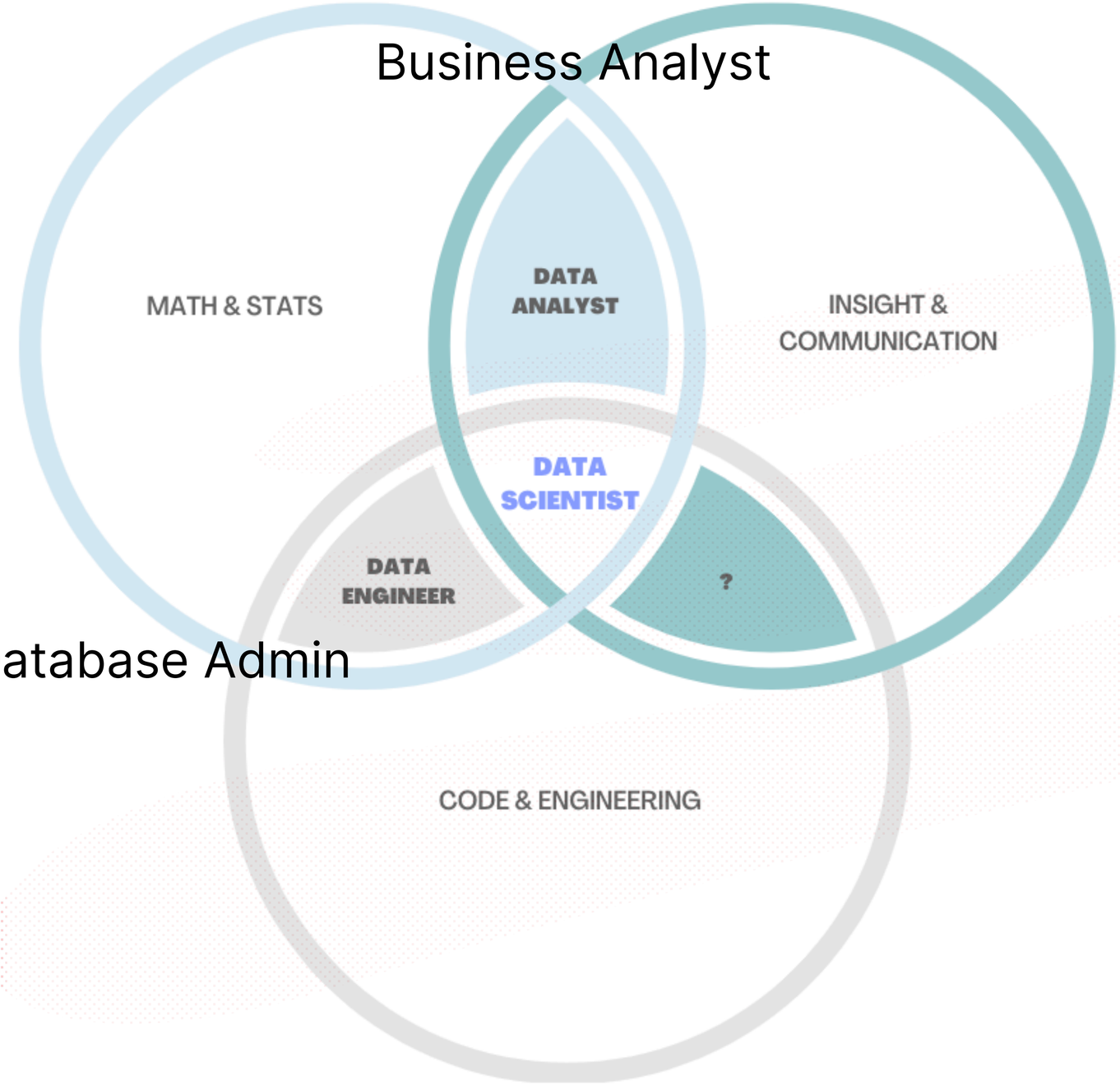

MATH & STATS
INSIGHT & COMMUNICATION
CODE & ENGINEERING
DATA ANALYST
DATA SCIENTIST
DATA ENGINEER
?


MATH & STATS
DATA ANALYST
INSIGHT & COMMUNICATION
DATA SCIENTIST
DATA ENGINEER
?
CODE & ENGINEERING

Business Analyst

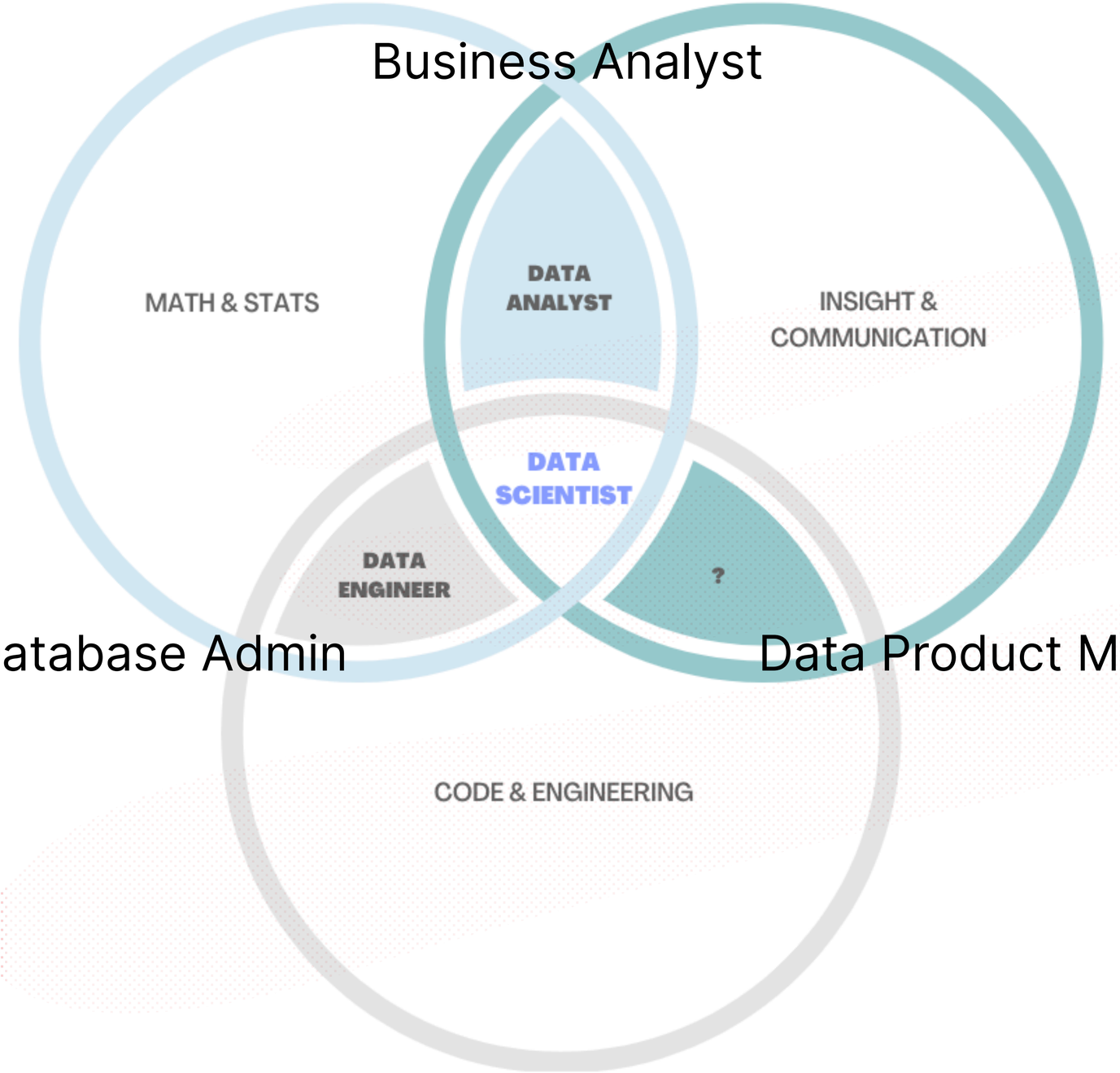


Business Analyst



Database Admin

Business Analyst

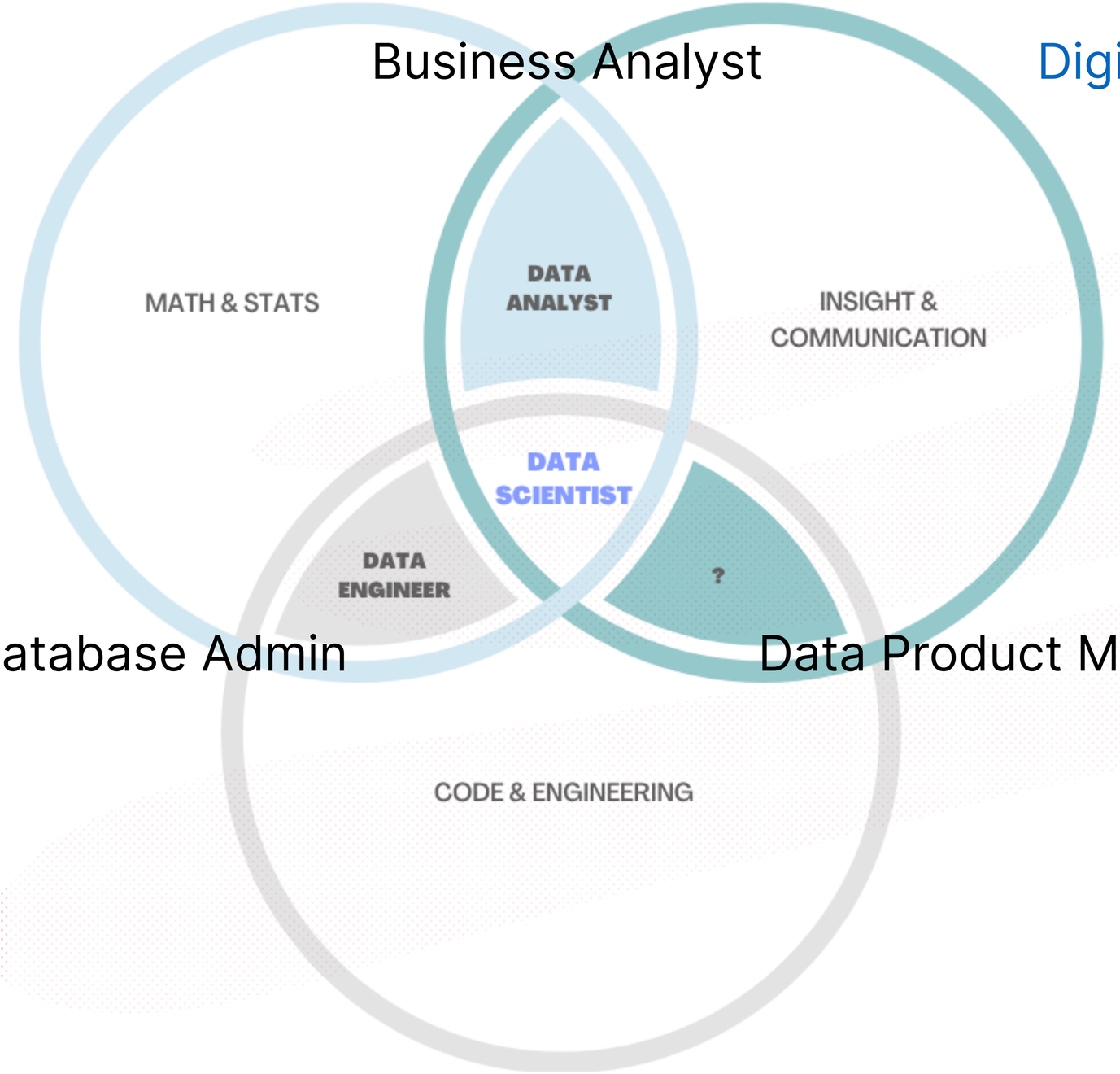


Database Admin

Data Product Manager

Business Analyst

Digital Marketing

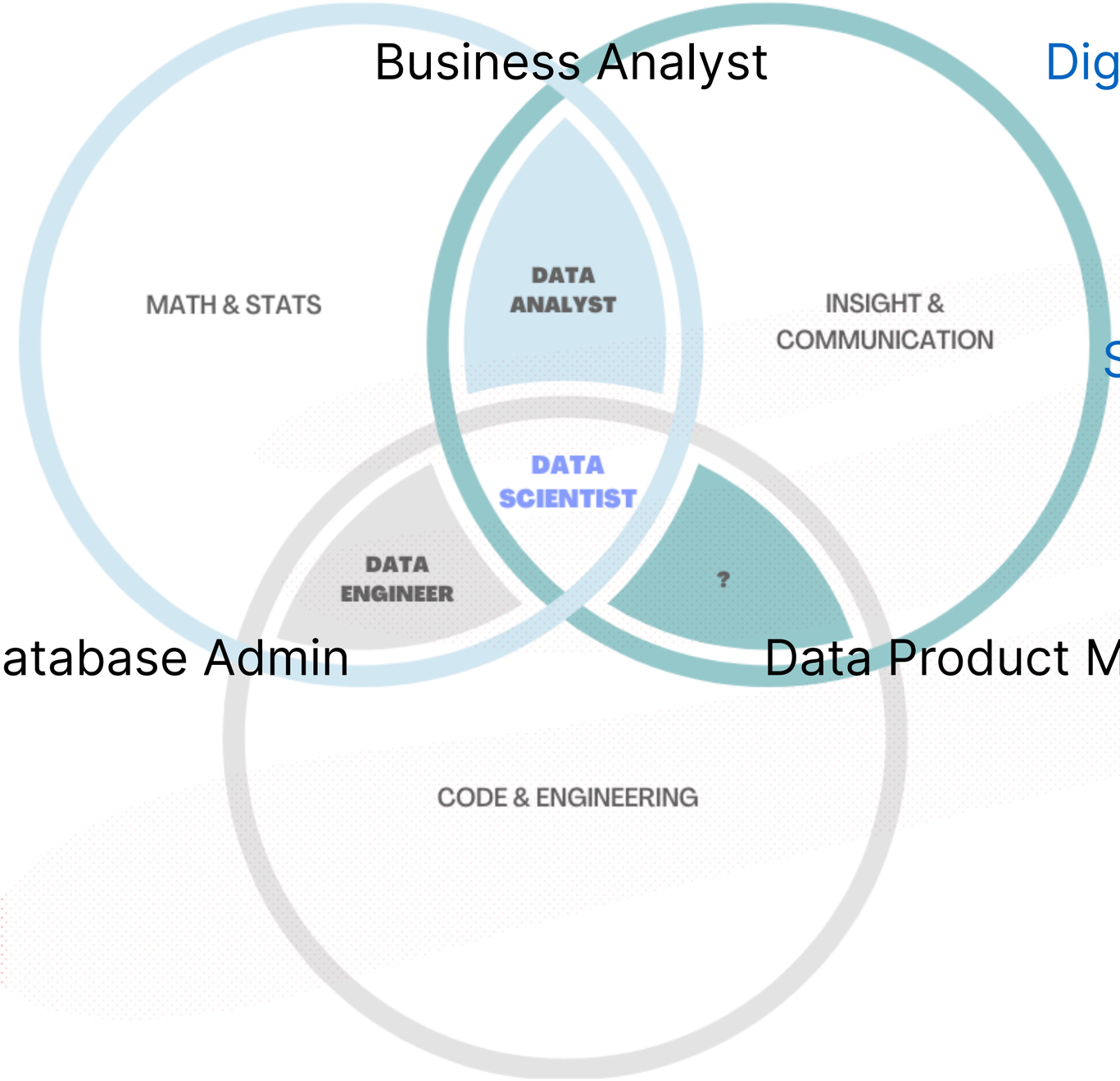


Database Admin

Data Product Manager

Business Analyst

Digital Marketing



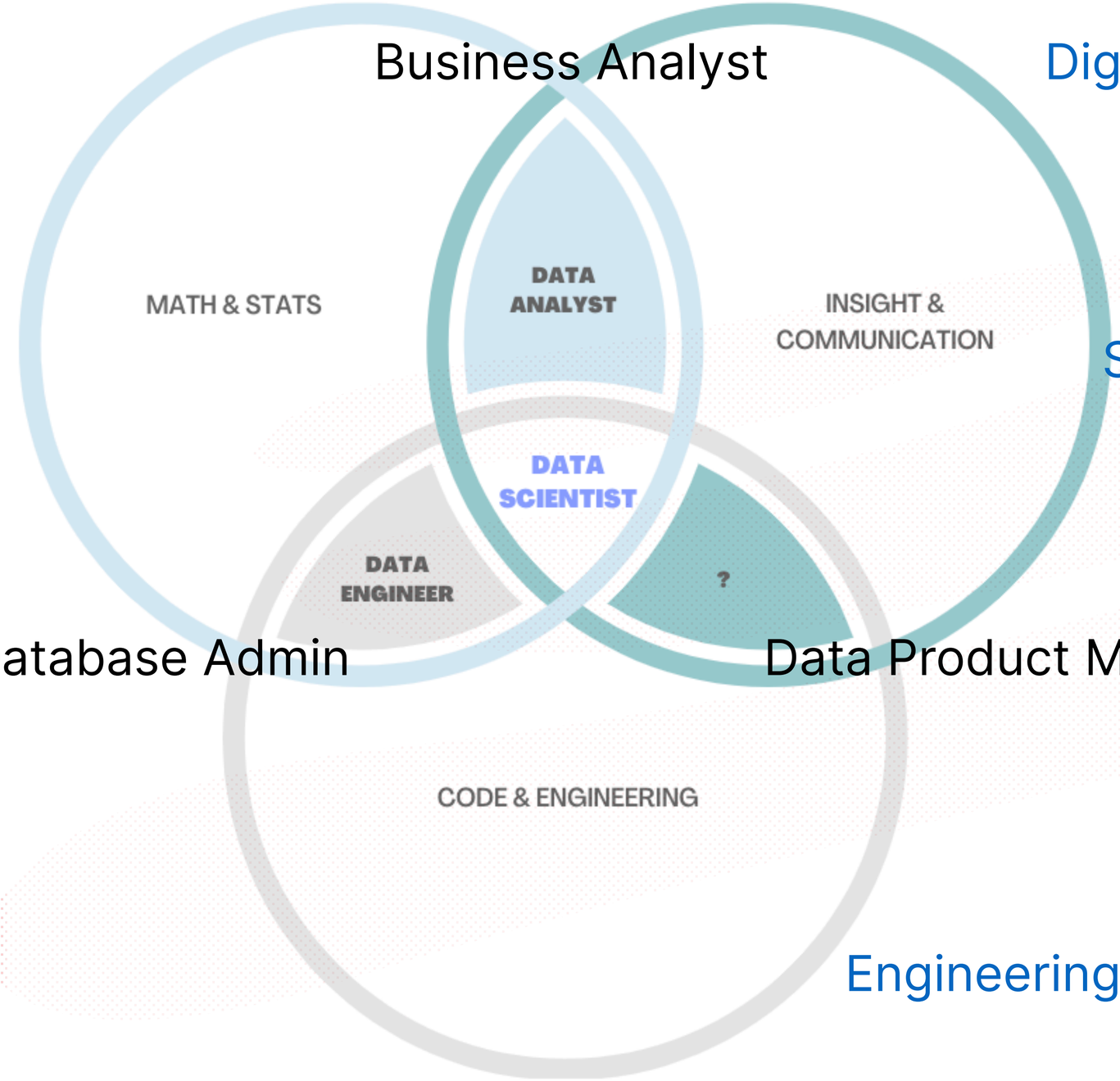
Strategy & Planning

Database Admin

Data Product Manager

Business Analyst

Digital Marketing



Strategy & Planning

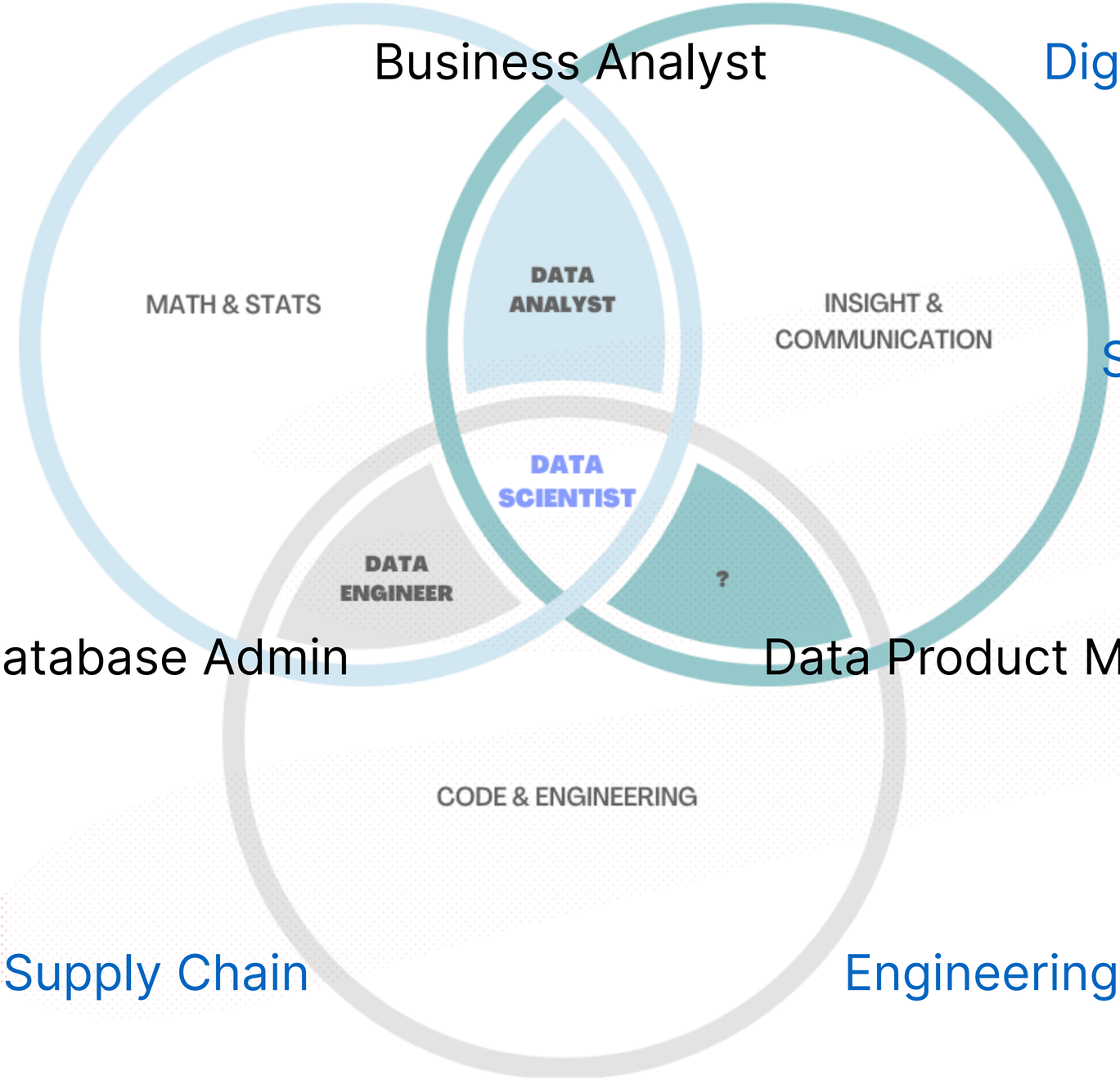
Database Admin

Data Product Manager

Engineering

Business Analyst

Digital Marketing



Strategy & Planning

Database Admin

Data Product Manager

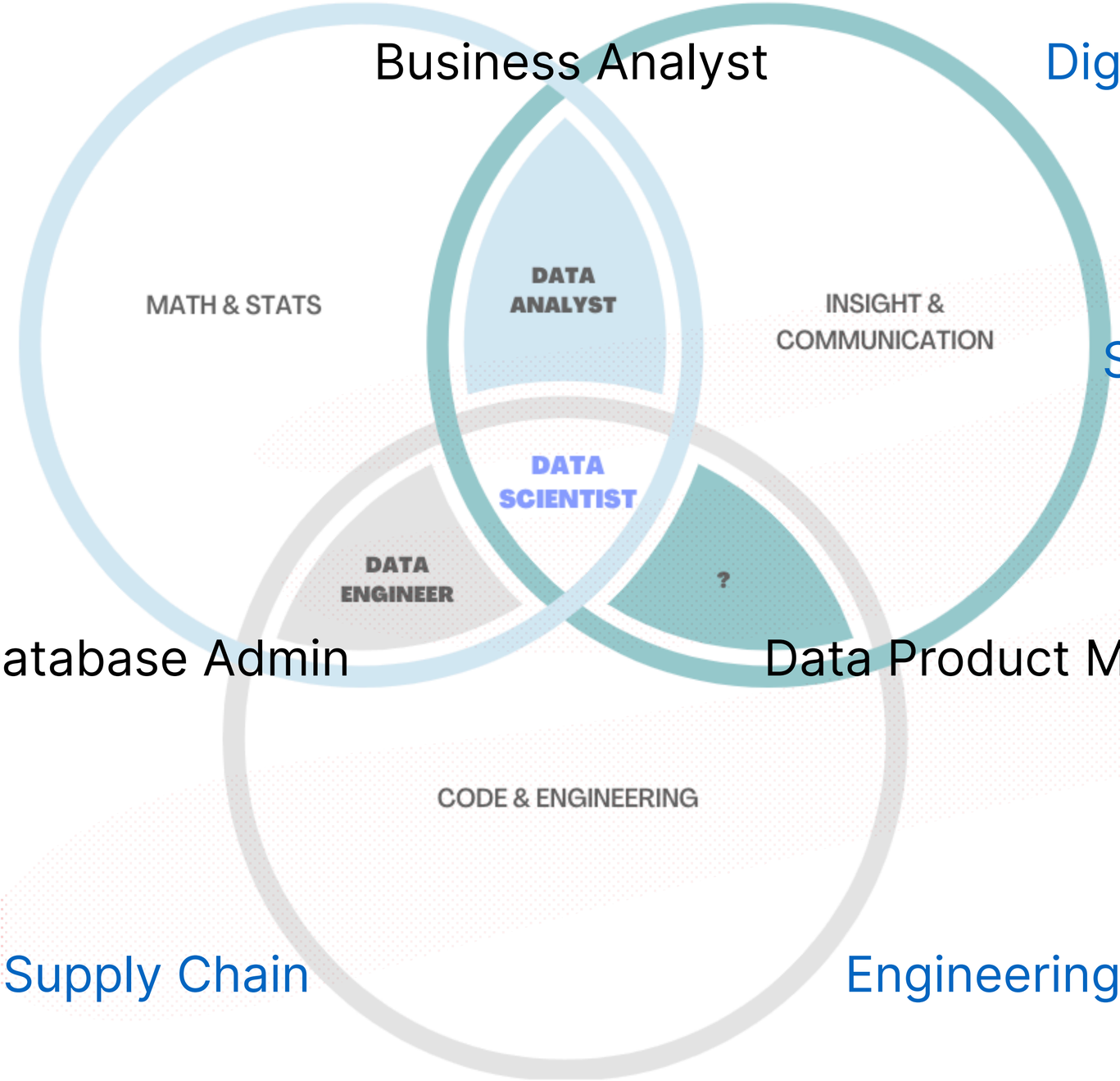
Supply Chain

Engineering

Business Analyst

Digital Marketing

Operations



Strategy & Planning

Database Admin

Data Product Manager

Supply Chain

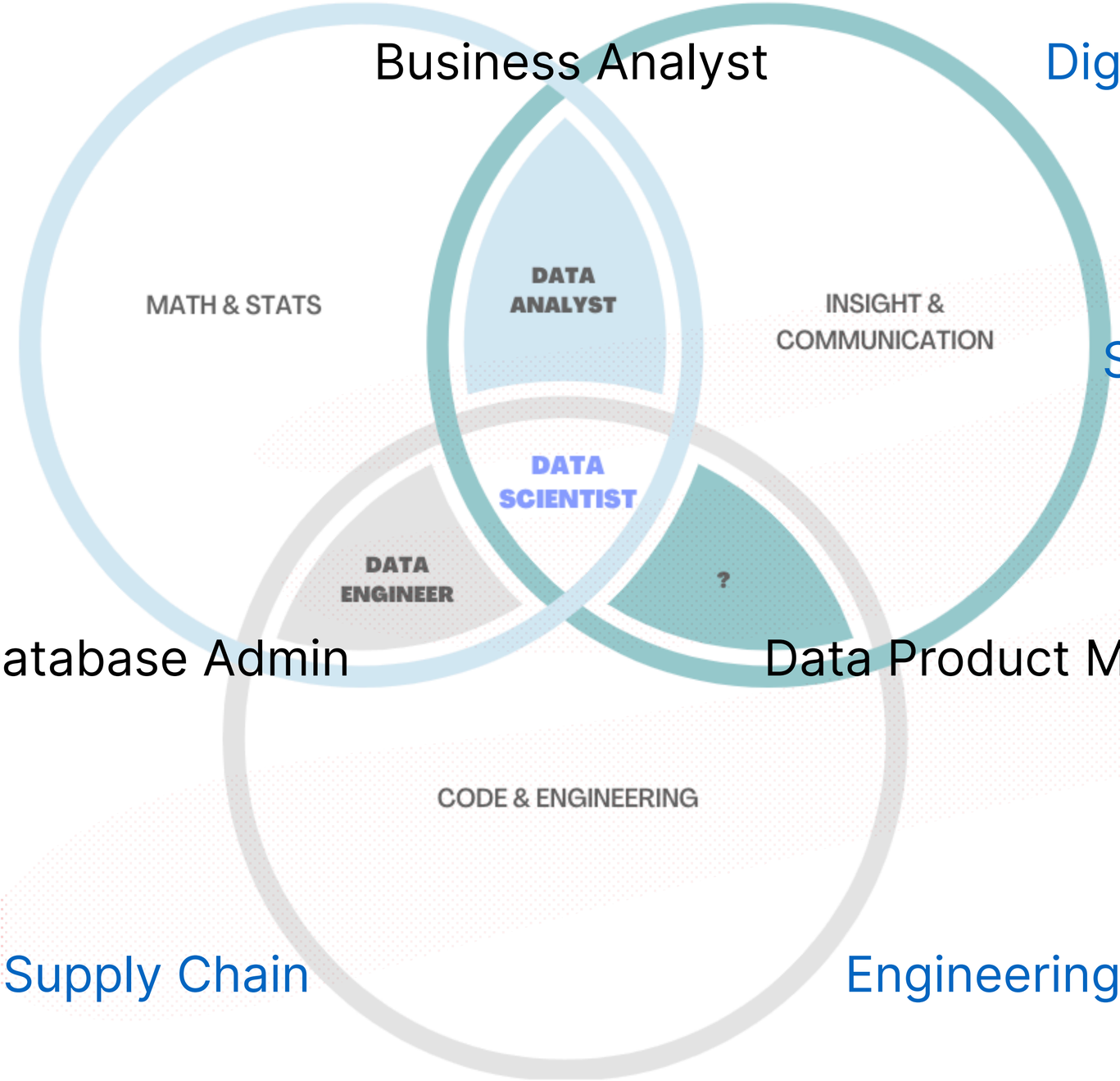
Engineering

Finance & HR

Business Analyst

Digital Marketing

Operations



Strategy & Planning

Database Admin

Data Product Manager

Supply Chain

Engineering

Data Science as a skill

Data Science as a skill

Data Science as a skill

Data Journalism

Film Coloring

Bicycle Building

Data Science as a skill

Data Journalism

Film Coloring

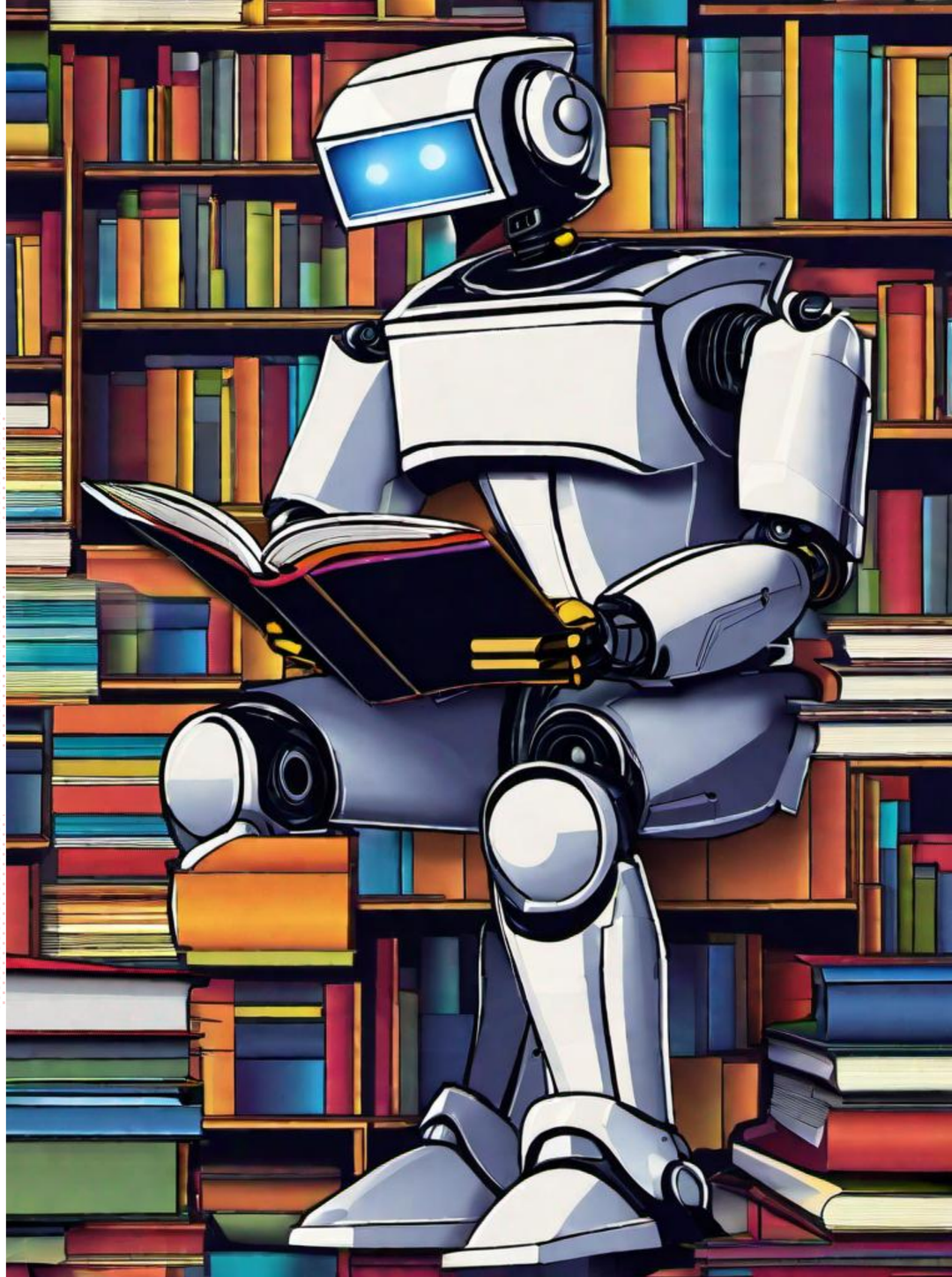
Bicycle Building

Finance, Marketing, Strategy

Agenda

- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

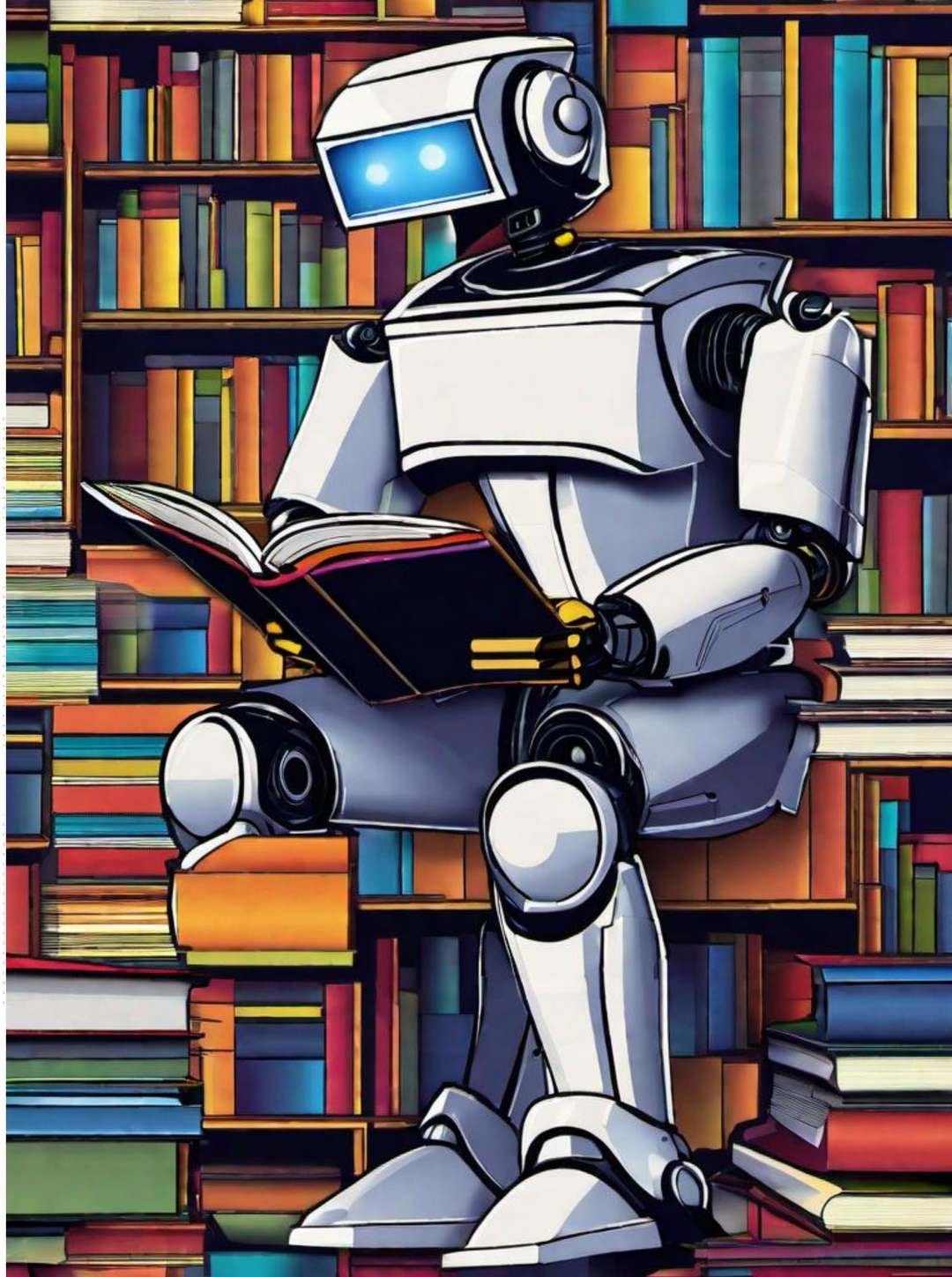
le wagon



Agenda

- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

le wagon



Ready to get nerdy? 🤓

RStudio Connect

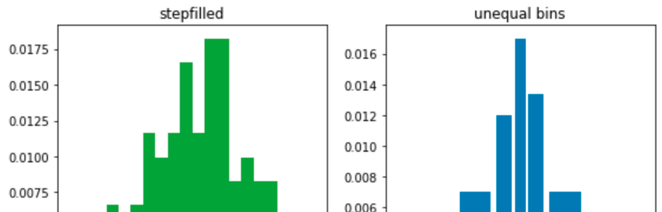
https://colorado.rstudio.com/rsc/jupyter-notebook-visualization/jupyter-static-visualization.html

Python Visualization Libraries

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Matplotlib

In [2]:
```
np.random.seed(0)

mu = 200
sigma = 25
x = np.random.normal(mu, sigma, size=100)

fig, (ax0, ax1) = plt.subplots(ncols=2, figsize=(8, 4))

ax0.hist(x, 20, density=1, histtype='stepfilled', facecolor='g', alpha=0.75)
ax0.set_title('stepfilled')

# Create a histogram by providing the bin edges (unequally spaced).
bins = [100, 150, 180, 195, 205, 220, 250, 300]
ax1.hist(x, bins, density=1, histtype='bar', rwidth=0.8)
ax1.set_title('unequal bins')

fig.tight_layout()
plt.show()
```



First, some analysis 🔍

What's the Data?

Understand what we have

- ✓ How many rows and columns?
- ✓ What are the columns we have?
- ✓ What are the averages?
- ✓ What are the minima and maxima?

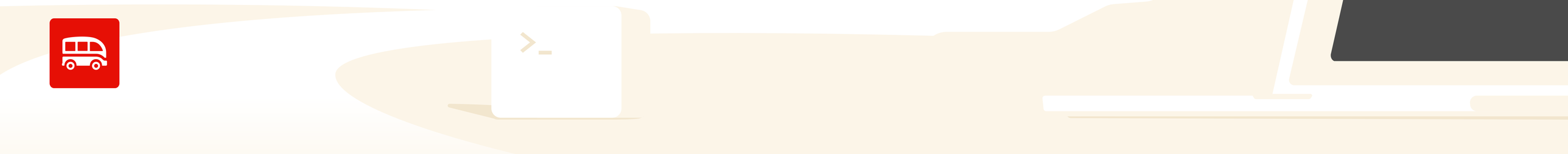
Methods to explore your data:

```
In [ ]: # to get the number of rows, columns
        salaries.shape

In [ ]: # to get the columns and their data types
        salaries.dtypes

In [ ]: # to get a readable summary of your data
        round(salaries.describe())

In [ ]: # to see only one column of the dataset
        salaries["Column Name"]

In [ ]: # to see multiple columns of the dataset
        salaries[["Column Name 1", "Column Name 2"]]
```

Visualize the Data

Get a sense of what it tells you

- ✓ How many of each category do we have?
- ✓ Do any two columns relate to each other?
- ✓ Do some categories influence the output?
- ✓ First step of any Data Scientist!



Visualize the Data

Get a sense of what it tells you

- ✓ How many of each category do we have?

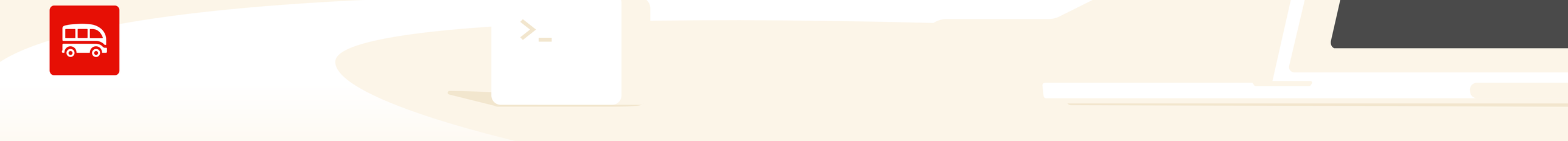
```
In [ ]: # to count the distribution of values in a column
        sns.countplot(data=salaries, x='Column Name')
```

- ✓ Do any two columns relate to each other?

```
In [ ]: # to see the relation of one column to another
        sns.scatterplot(data=salaries, x='Input Column', y='Output Column')
```

- ✓ Do some categories influence the output?

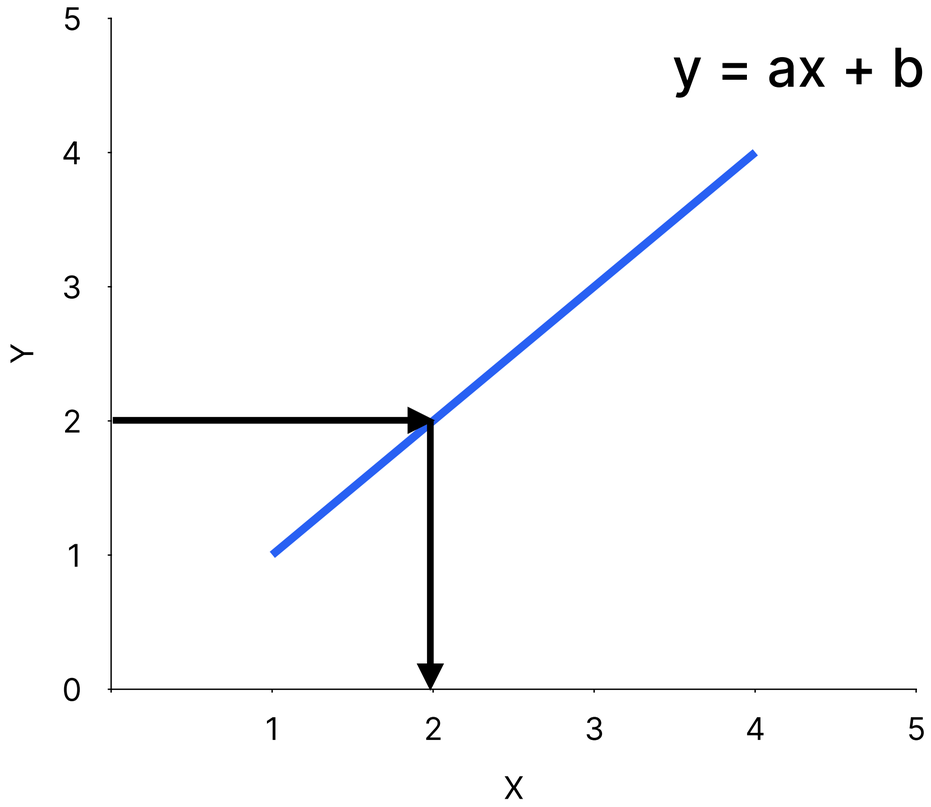
```
In [ ]: # to add a category to the visualization, we use `hue`
        sns.scatterplot(data=salaries, x='Input Column', y='Output Column', hue='Category Column')
```



Challenge 1

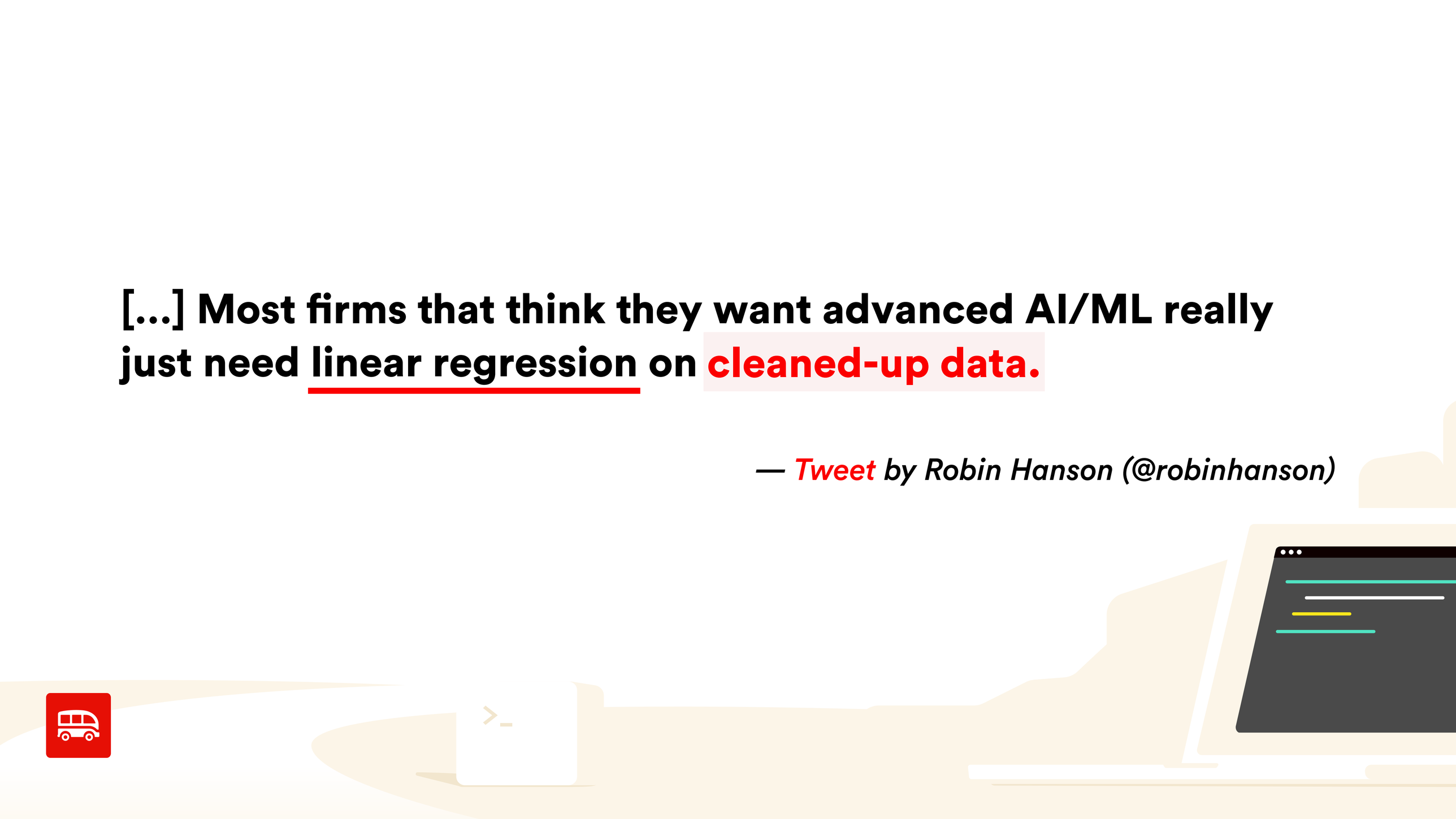
Predicting salaries 💰

Linear Regression

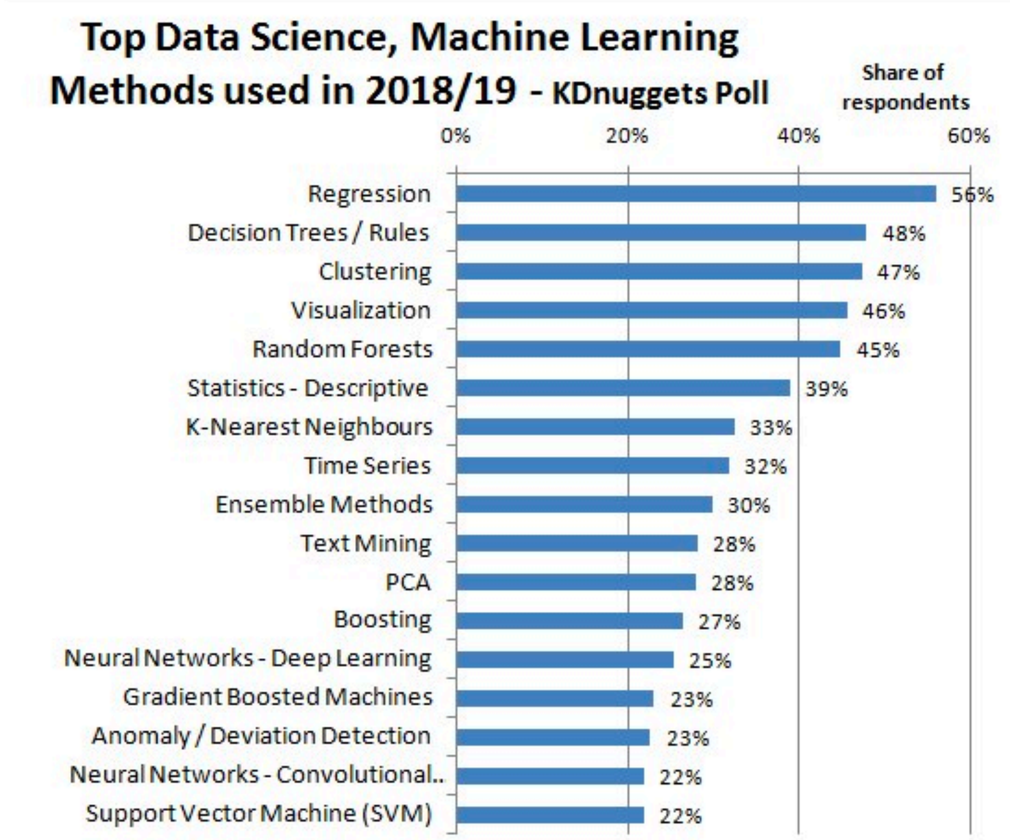


[...] Most firms that think they want advanced AI/ML really just need linear regression on cleaned-up data.

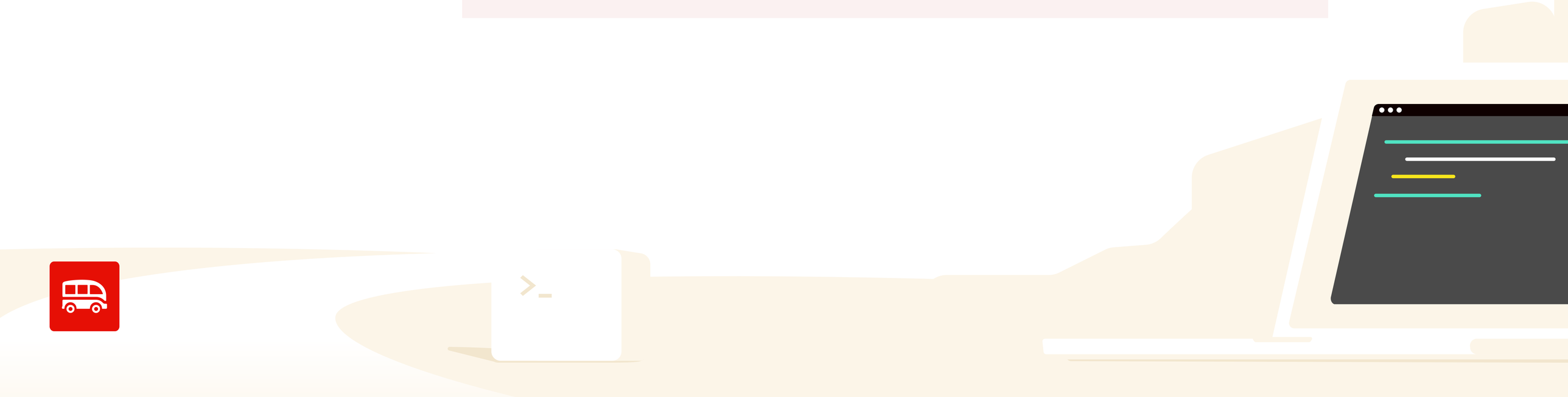
— Tweet by Robin Hanson (@robinhanson)



Why Regression?

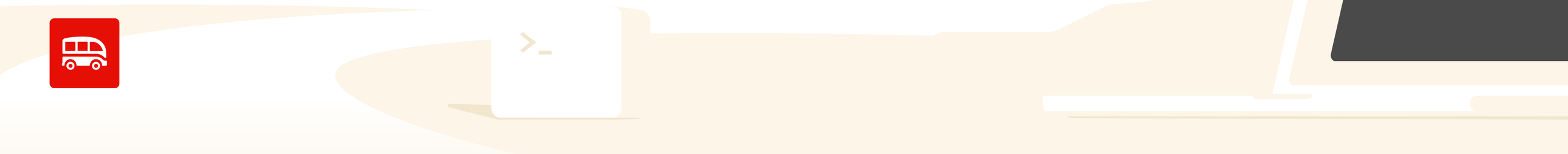


Visual demo on setosa.io! 🎥



Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data



Road to Machine Learning

1. Select the **features** and **targets**

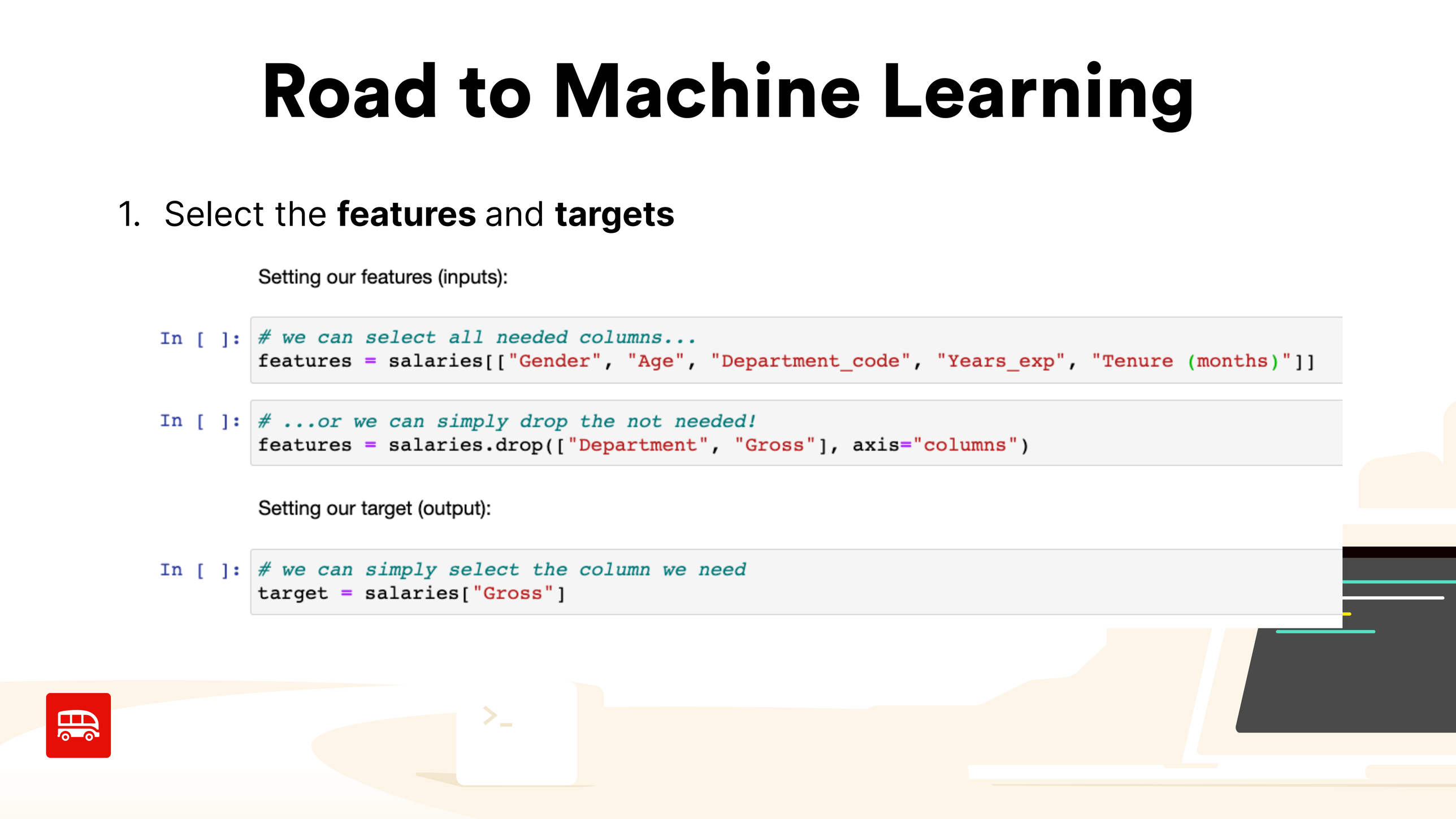
Setting our features (inputs):

```
In [ ]: # we can select all needed columns...
        features = salaries[["Gender", "Age", "Department_code", "Years_exp", "Tenure (months)"]]
```

```
In [ ]: # ...or we can simply drop the not needed!
        features = salaries.drop(["Department", "Gross"], axis="columns")
```

Setting our target (output):

```
In [ ]: # we can simply select the column we need
        target = salaries["Gross"]
```



Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data

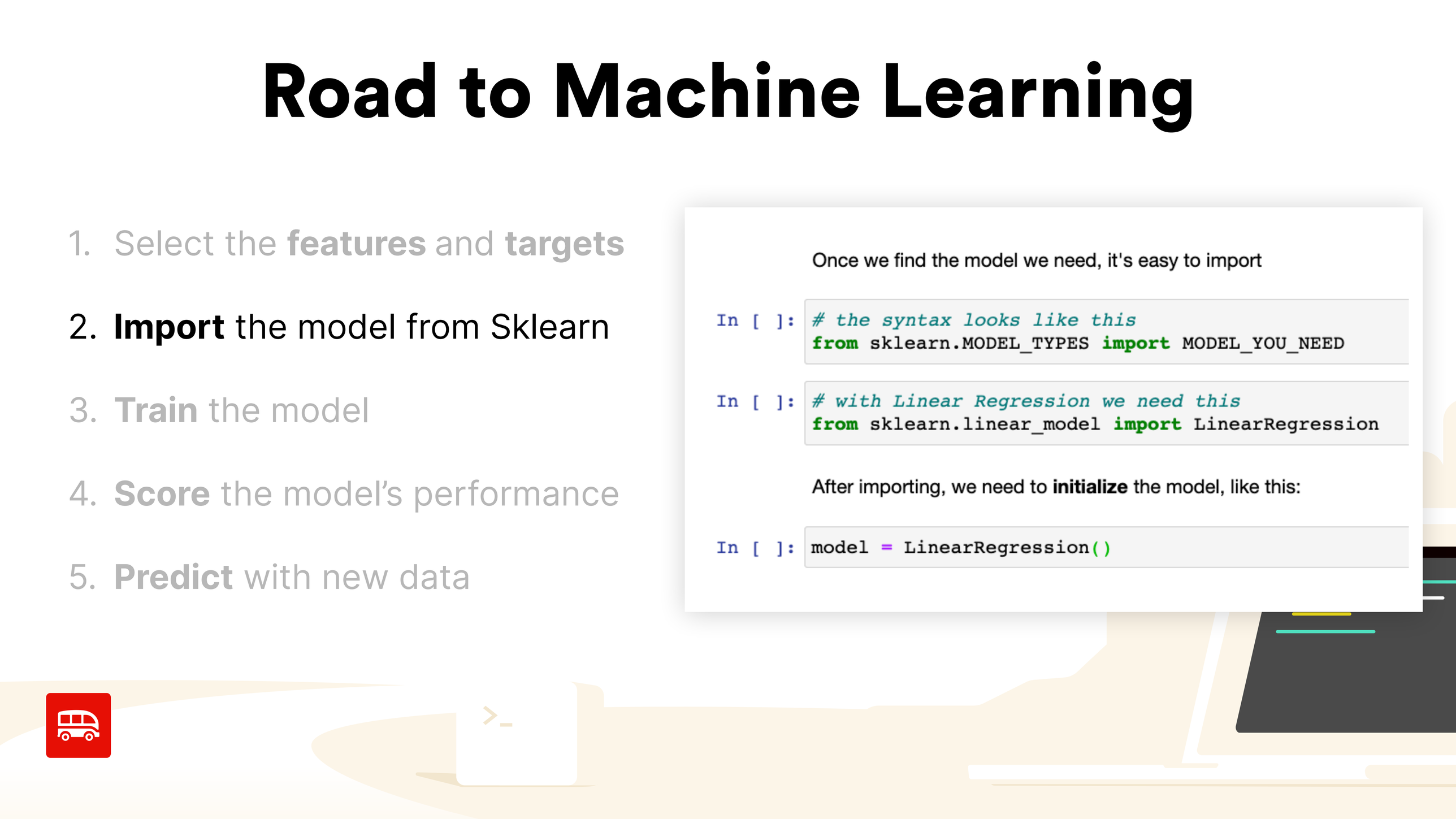
Once we find the model we need, it's easy to import

```
In [ ]: # the syntax looks like this
        from sklearn.MODEL_TYPES import MODEL_YOU_NEED
```

```
In [ ]: # with Linear Regression we need this
        from sklearn.linear_model import LinearRegression
```

After importing, we need to **initialize** the model, like this:

```
In [ ]: model = LinearRegression()
```



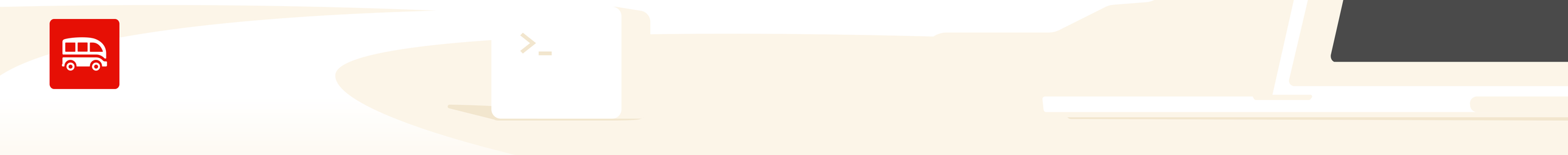
Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data

To train the model, we use the `.fit` method:

```
In [ ]: model.fit(features, target)
```

The `model` then finds the **best fitting** line between features and target



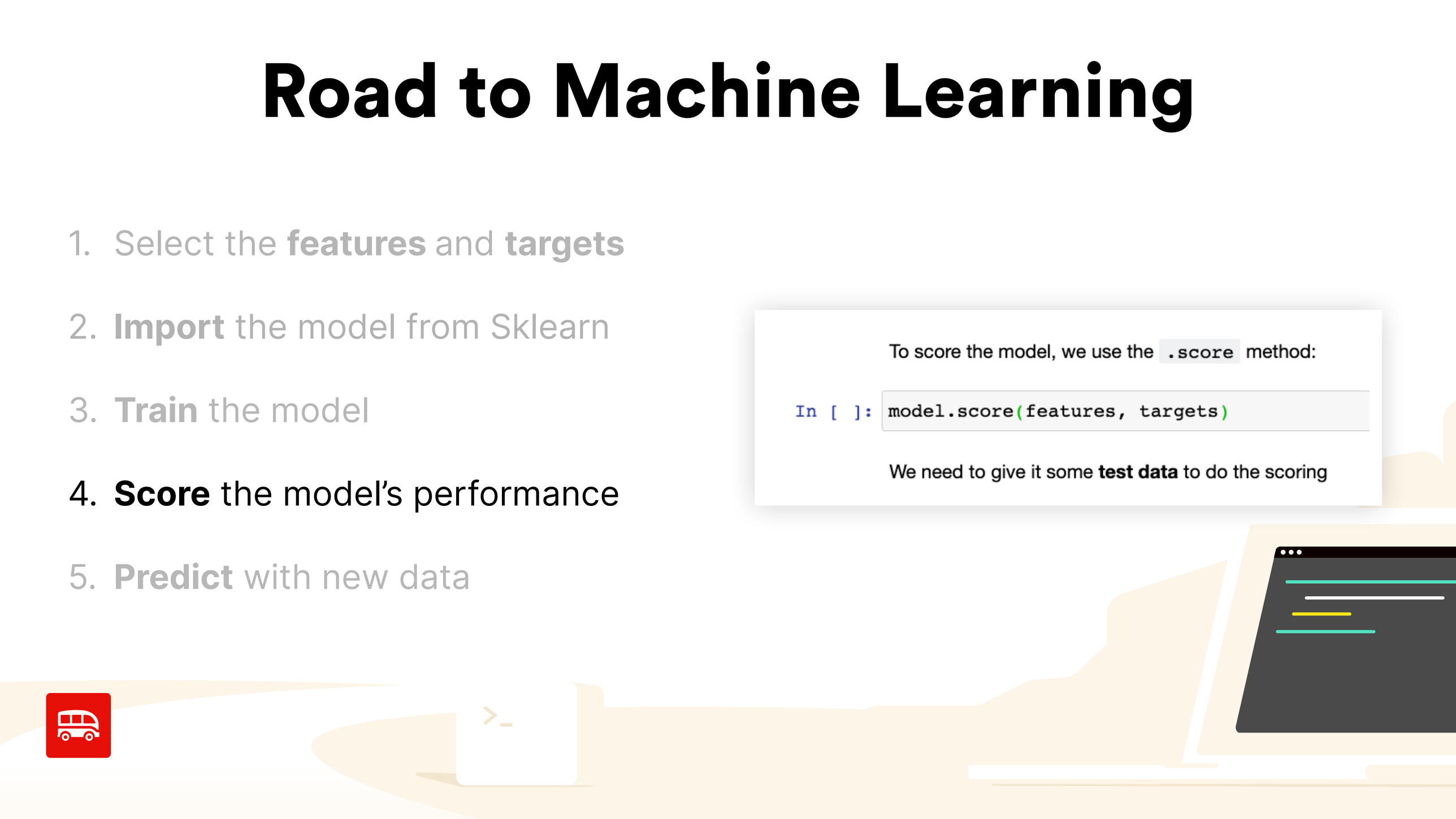
Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data

To score the model, we use the `.score` method:

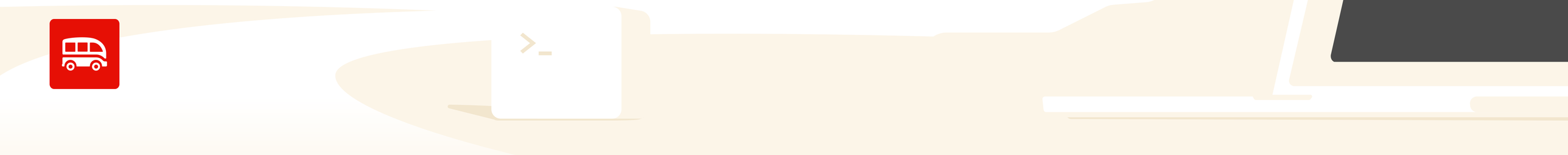
```
In [ ]: model.score(features, targets)
```

We need to give it some **test data** to do the scoring



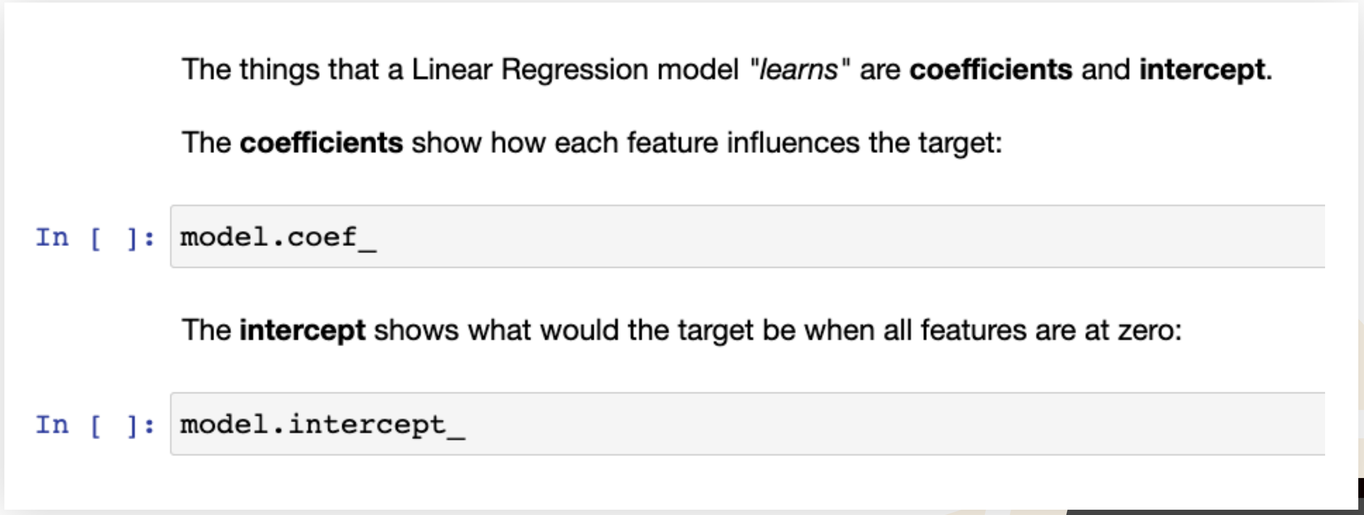
Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data



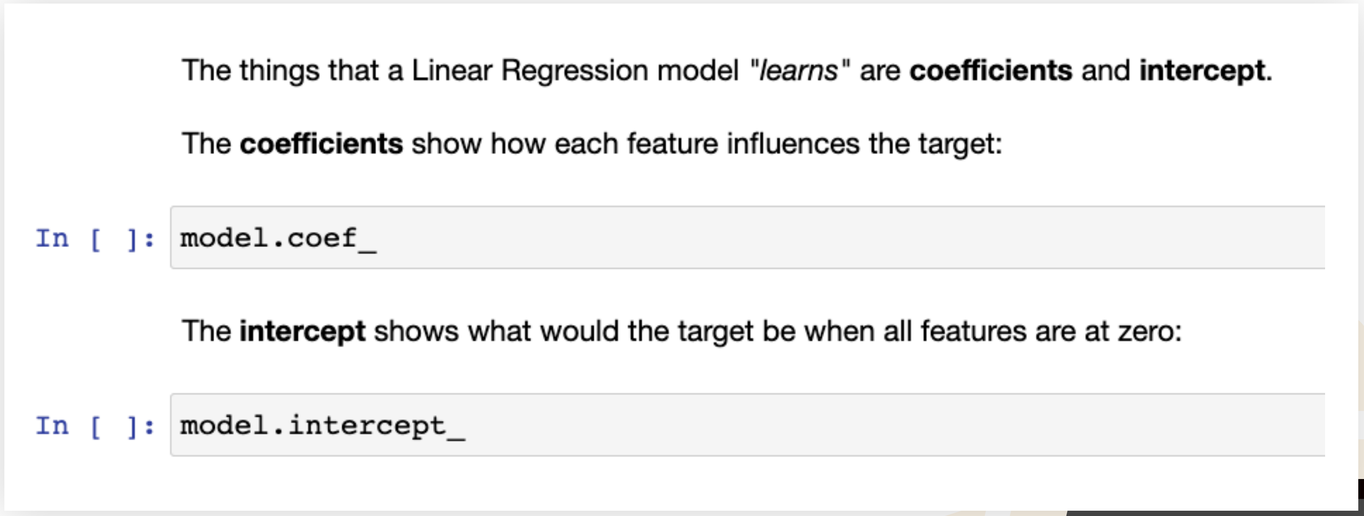
Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data



Road to Machine Learning

1. Select the **features** and **targets**
2. **Import** the model from Sklearn
3. **Train** the model
4. **Score** the model's performance
5. **Predict** with new data
6. **Explain** the model



That easy?! 😮

Yes...but no

That easy?! 😮

Yes...but no

```
model.fit(features, target)
```

```
model.score(features, targets)
```

That easy?! 😮

Yes...but no

```
model.fit(features, target)
```

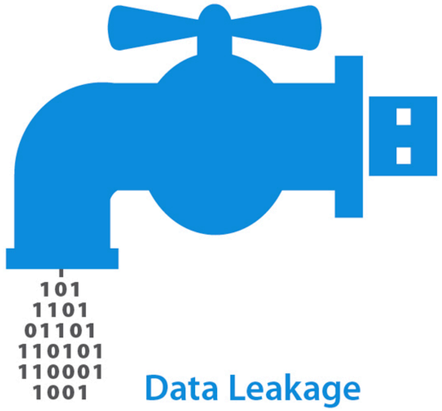
```
model.score(features, targets)
```

That easy?! 😮

More on this in challenge 2

```
model.fit(features, target)
```

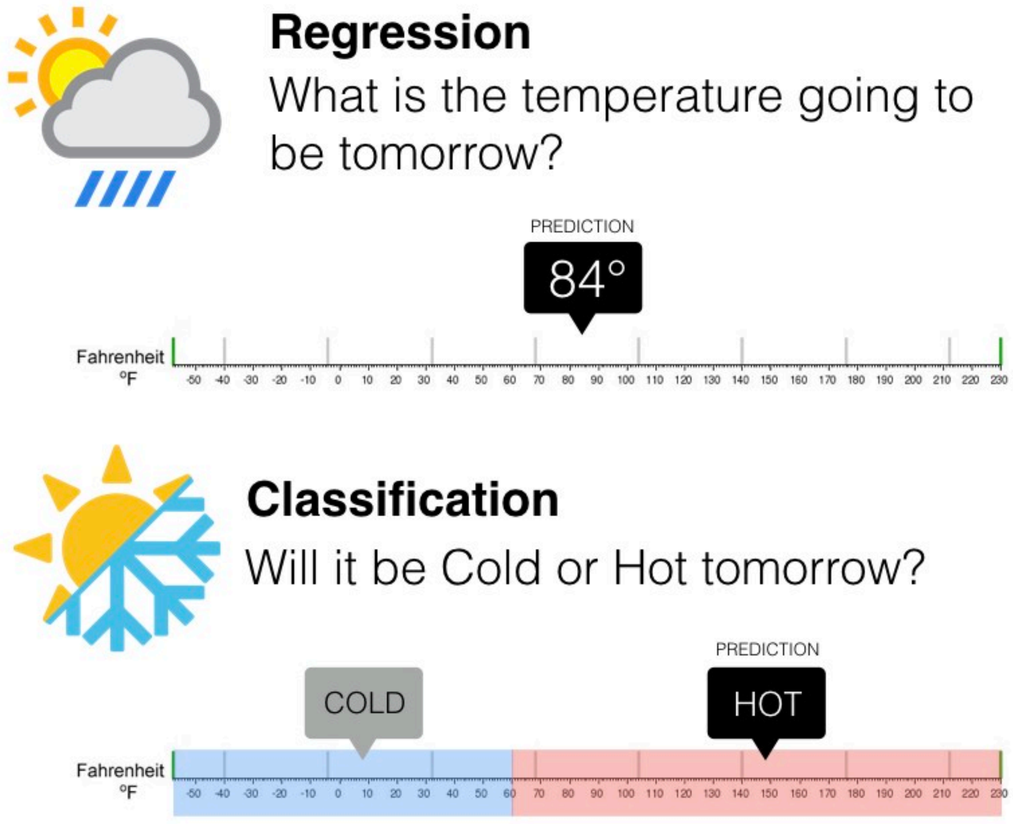
```
model.score(features, targets)
```



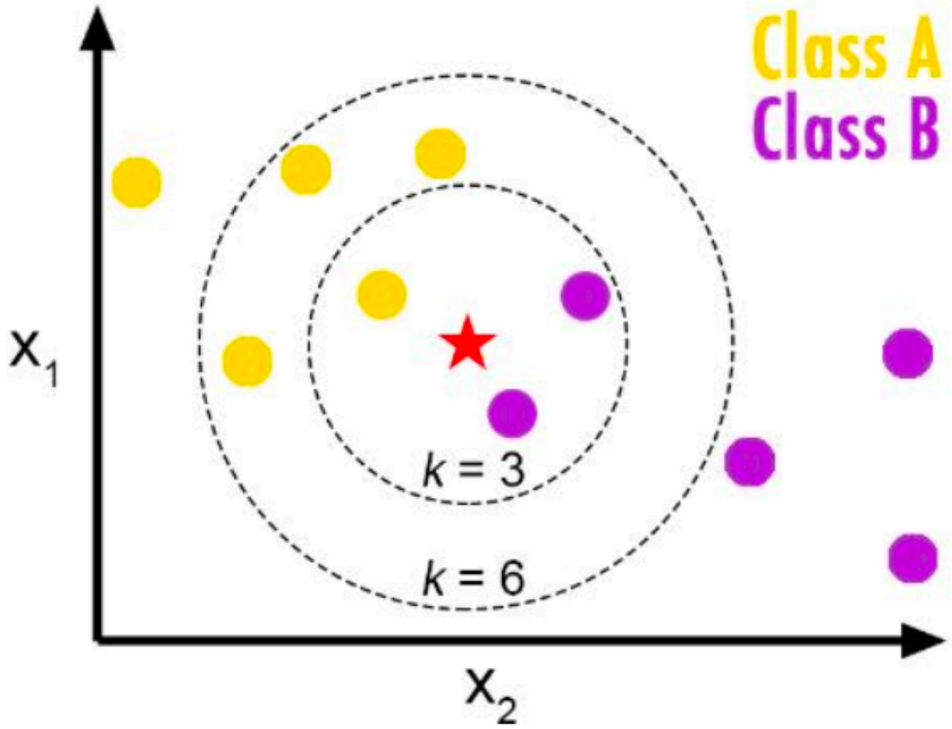
Optional Challenge 2

Customer churn

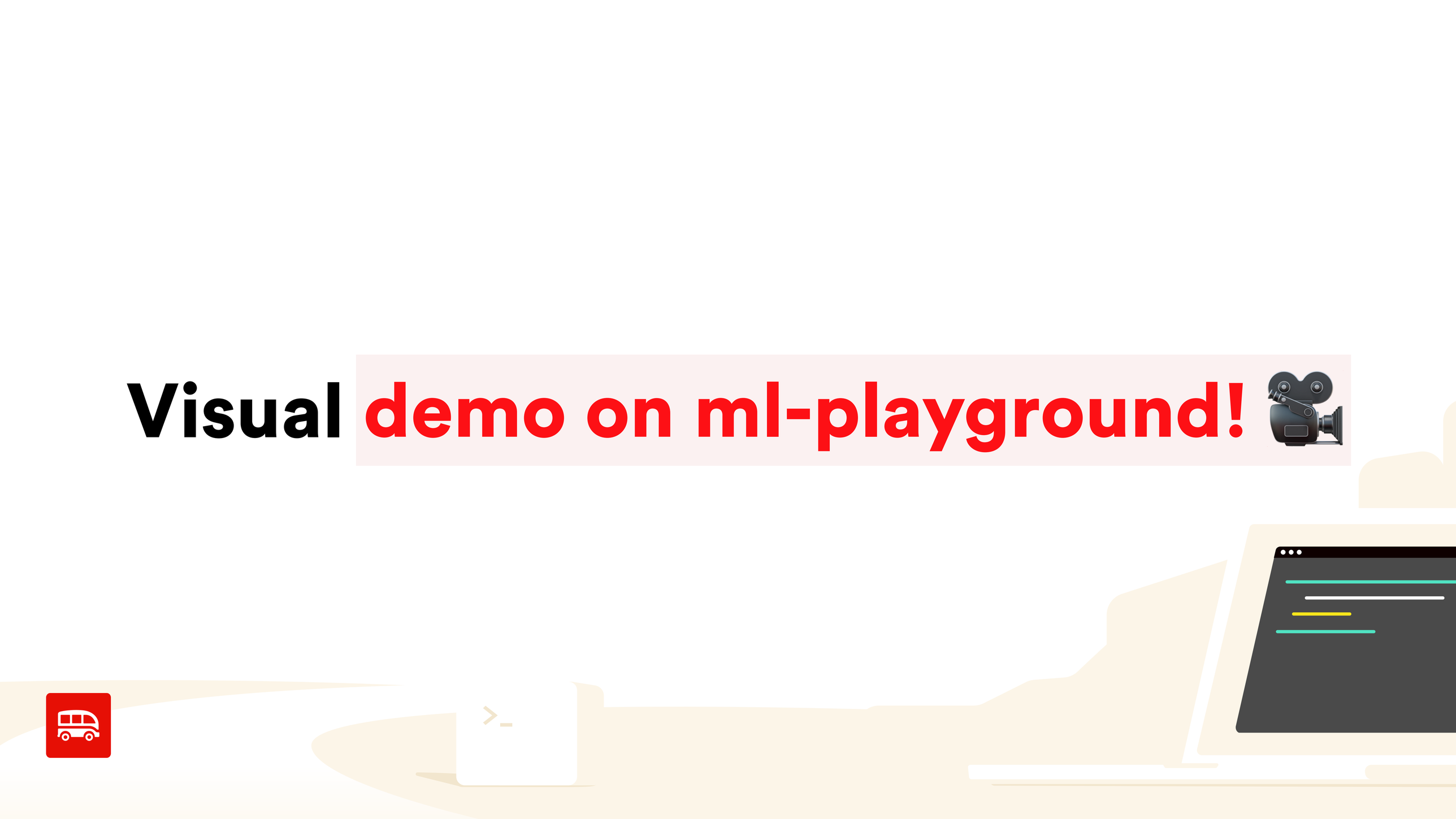
Regression VS Classification



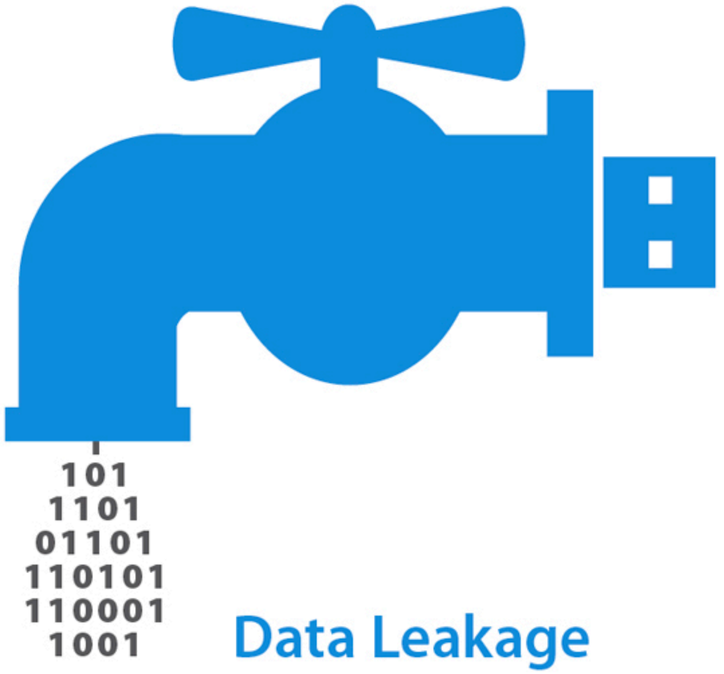
K-Nearest Neighbors



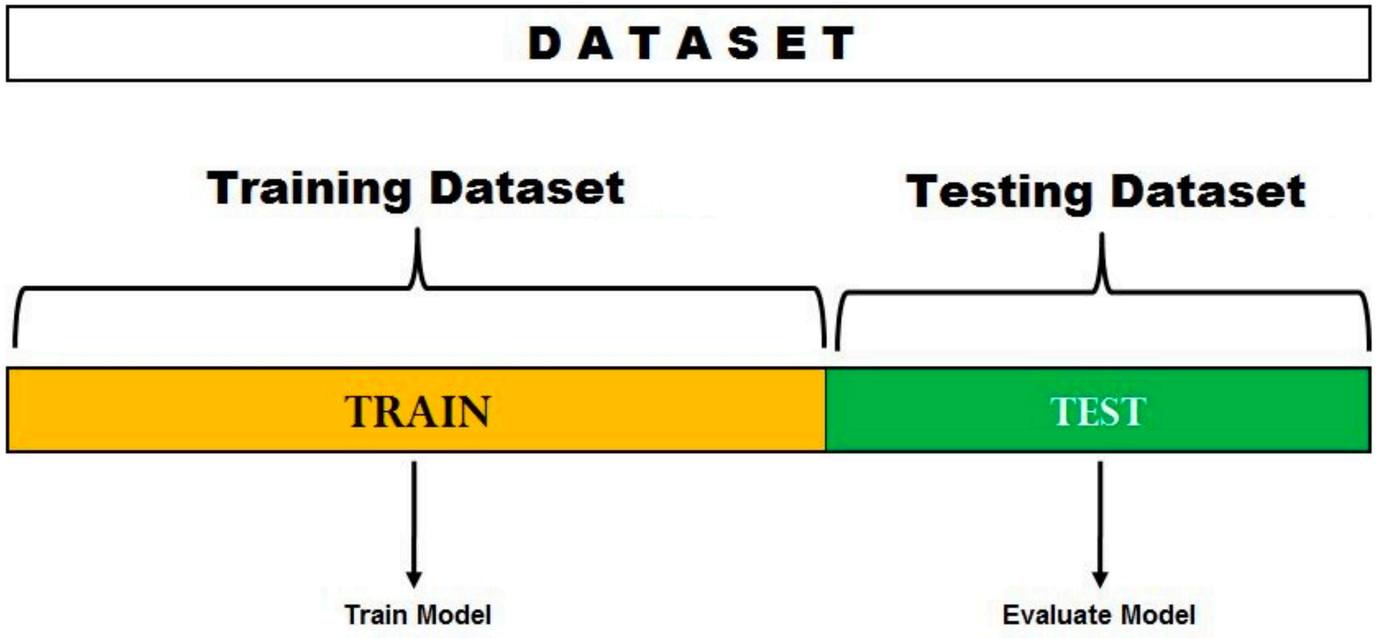
Visual demo on ml-playground! 🎥



We have a leak



Holdout Method



Congrats! 👏

Now you know

le wagon

Congrats! 👏

Now you know

- ✓ What the concept of Machine Learning is

Congrats! 👏

Now you know

- ✓ What the concept of Machine Learning is
- ✓ The basics of Jupyter Notebook - the #1 tool of any Data Scientist ⚛️

le wagon

Congrats! 👏

Now you know

- ✓ What the concept of Machine Learning is
- ✓ The basics of Jupyter Notebook - the #1 tool of any Data Scientist ⚛️
- ✓ How to use Python libraries - of which there are thousands 🤩

le wagon

Congrats! 👏

Now you know

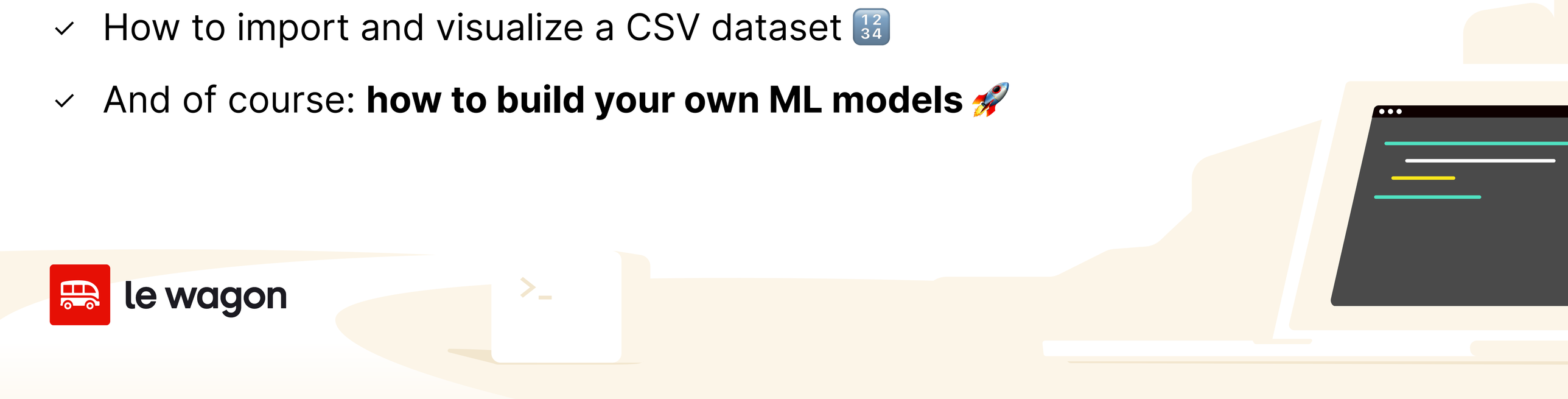
- ✓ What the concept of Machine Learning is
- ✓ The basics of Jupyter Notebook - the #1 tool of any Data Scientist ⚛️
- ✓ How to use Python libraries - of which there are thousands 🤩
- ✓ How to import and visualize a CSV dataset 🔢

le wagon

Congrats! 👏

Now you know

- ✓ What the concept of Machine Learning is
- ✓ The basics of Jupyter Notebook - the #1 tool of any Data Scientist ⚛️
- ✓ How to use Python libraries - of which there are thousands 🤩
- ✓ How to import and visualize a CSV dataset 🔢
- ✓ And of course: **how to build your own ML models** 🚀

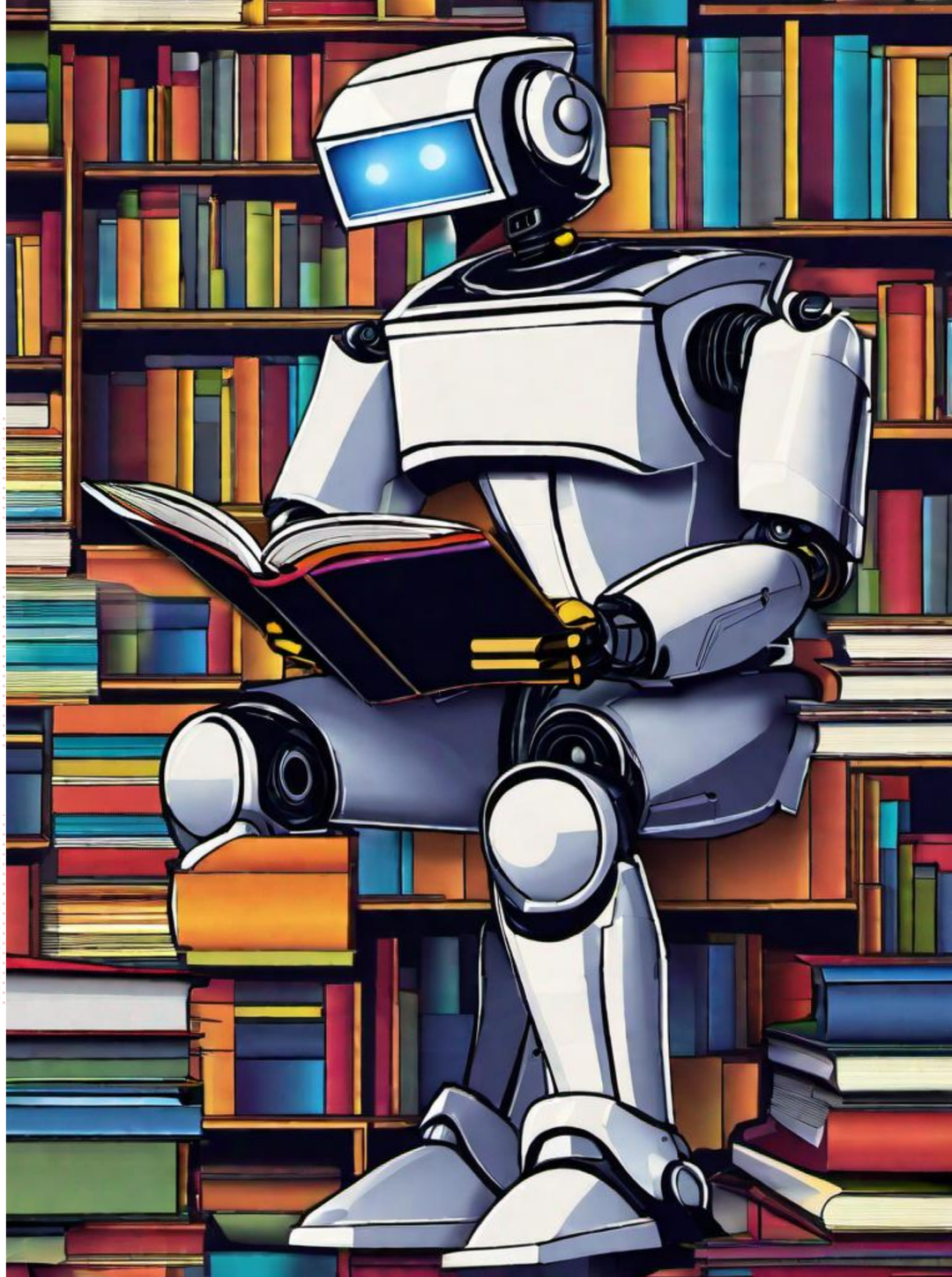


le wagon

Agenda

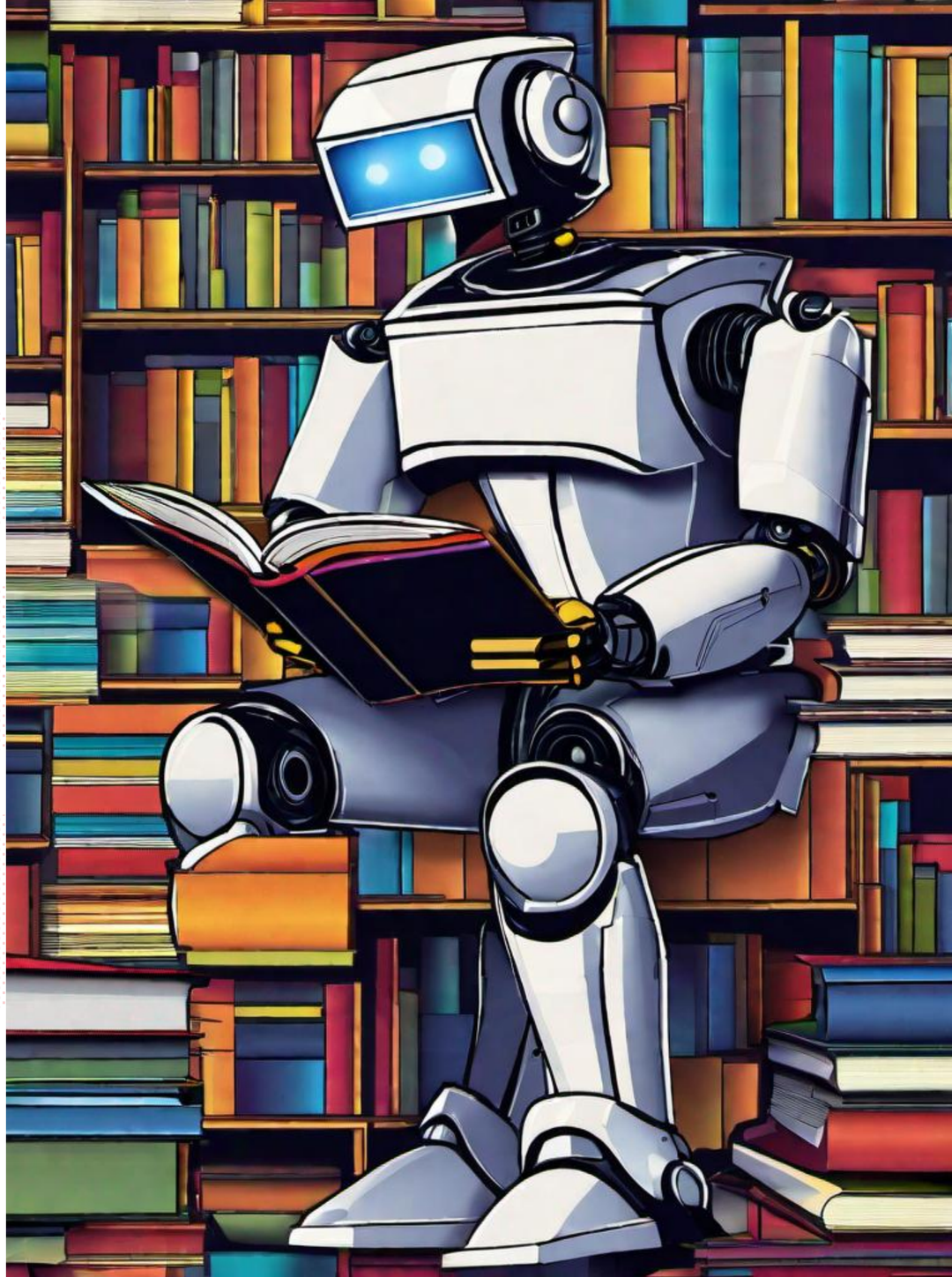
- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover

le wagon



Agenda

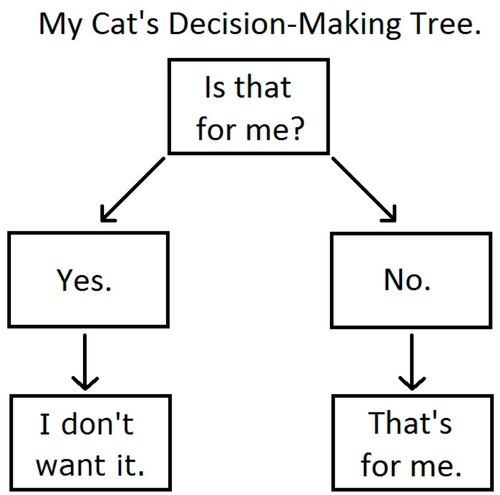
- ✓ What is Machine Learning
- ✓ What is NOT Machine Learning
- ✓ Who are the people building ML
- ✓ Let’s code our own models! 🚀
- ✓ What we didn’t cover



le wagon

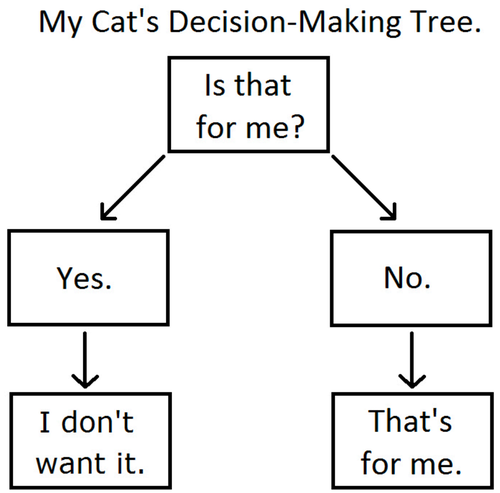
Other model types

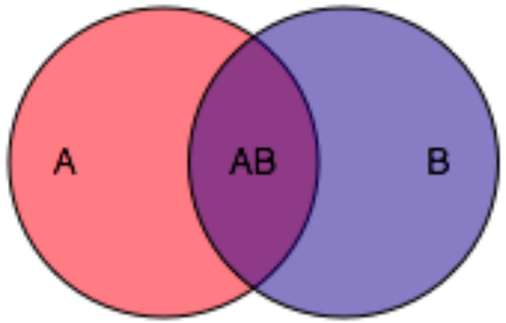
Other model types



Decision tree based

Other model types

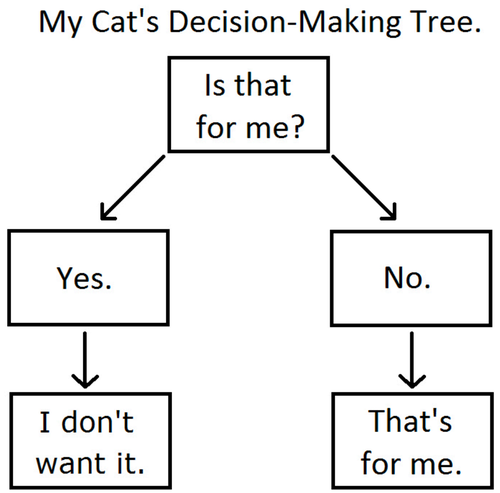


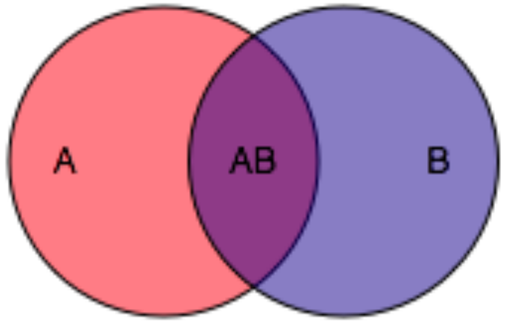


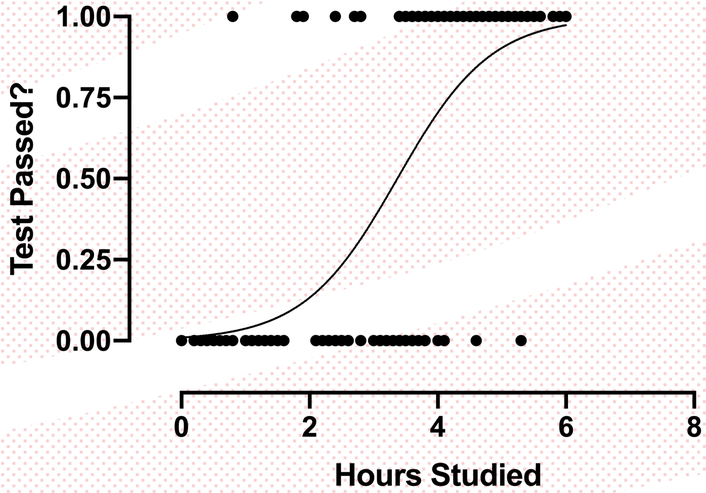
Decision tree based

Probabilistic

Other model types





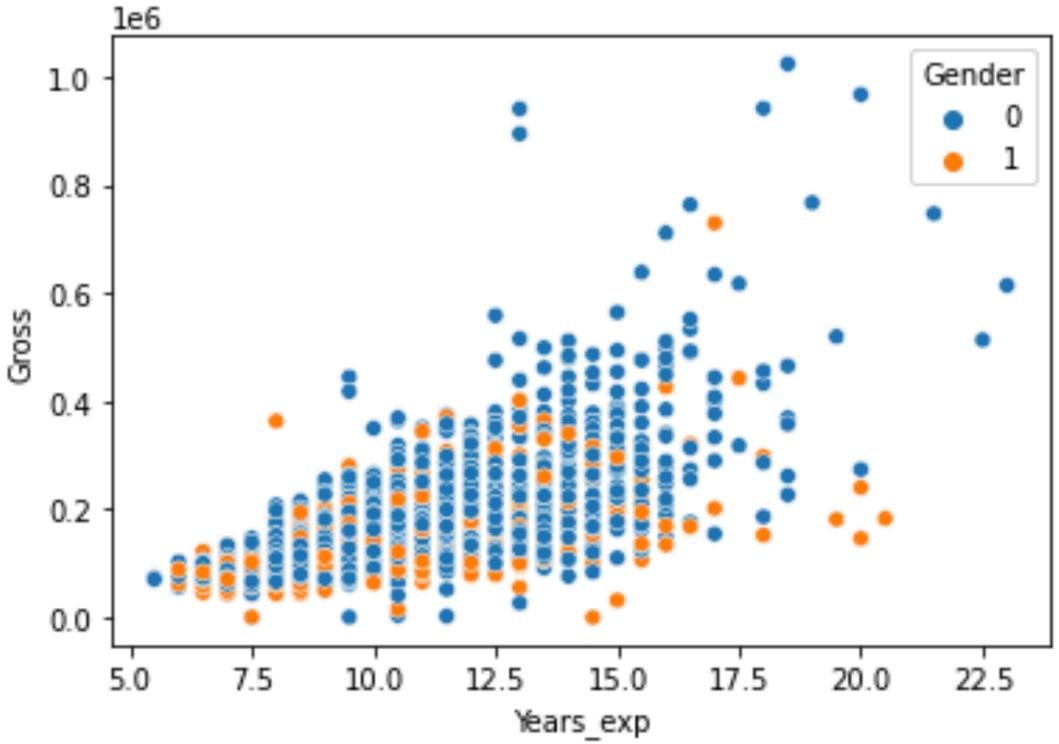


Decision tree based

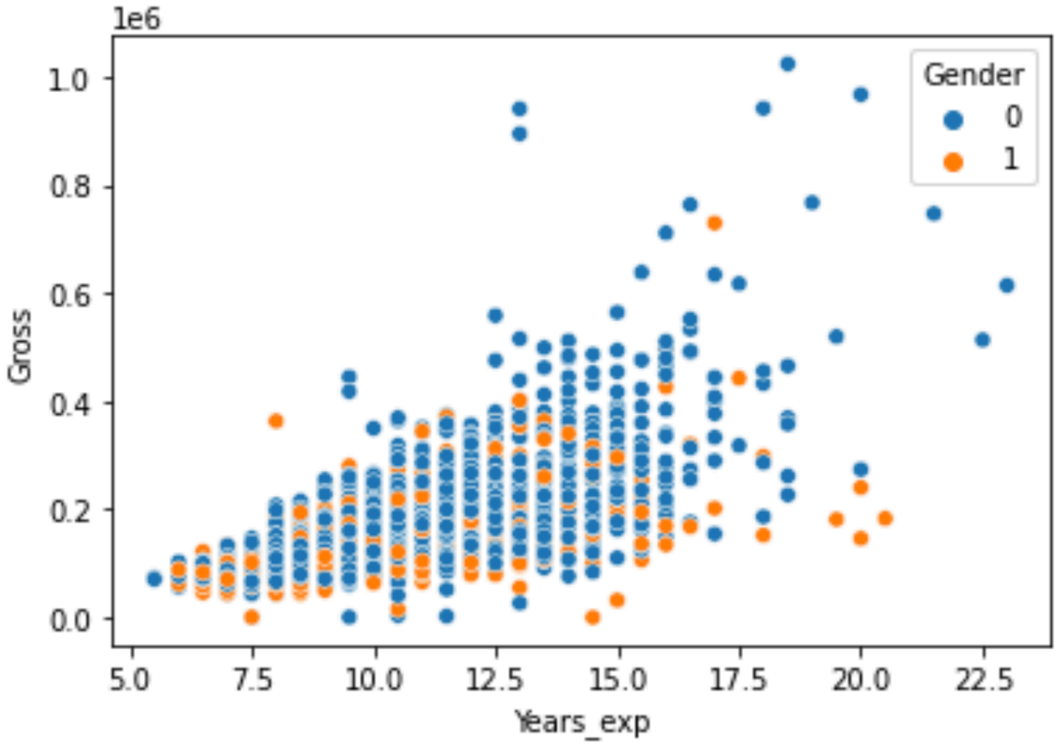
Probabilistic

Logistic

Balance and bias



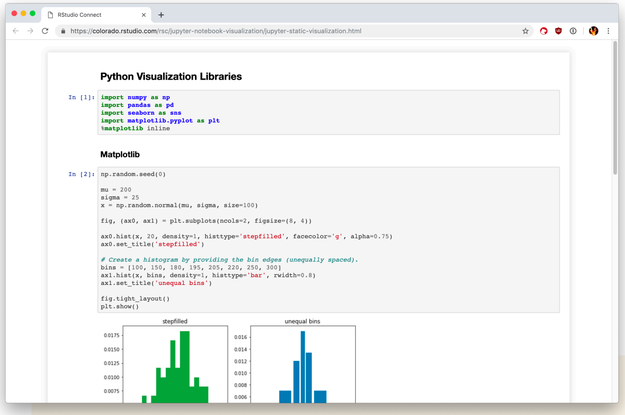
Balance and bias



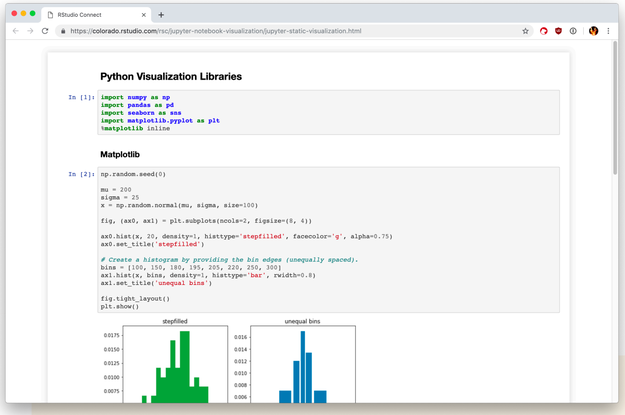
Reality or poor data?

Making a product

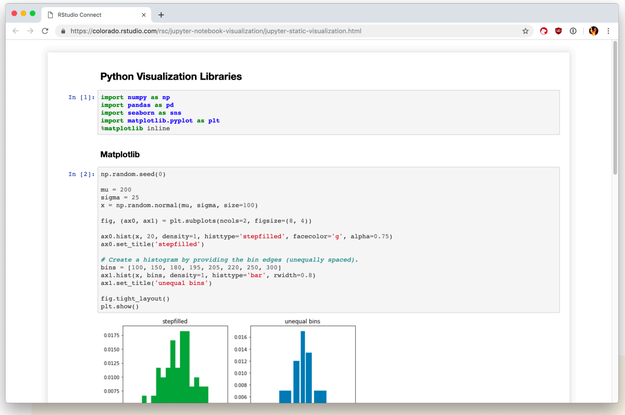
Making a product

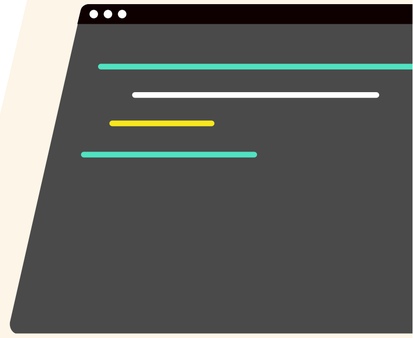


Making a product



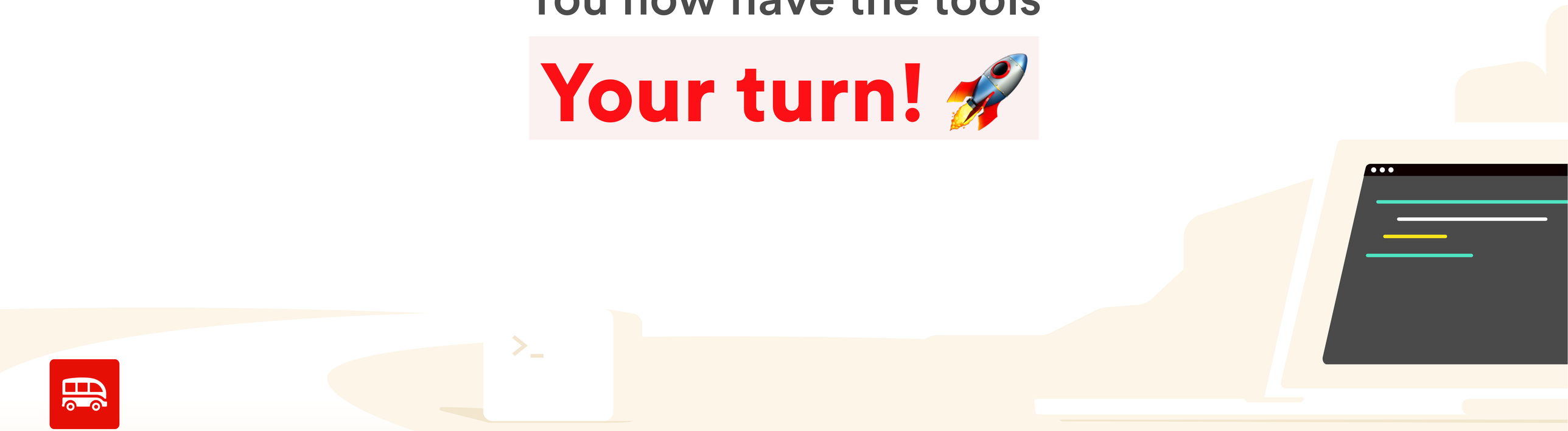
Making a product





You now have the tools

Your turn! 🚀



Thank you! 💻

Want to go further? Head over to

Le Wagon!