Explainability:

Data science and explainable workflows

Gabriela de Queiroz, Chief Data Scientist (gdq@ibm.com)
Saishruthi Swaminathan, Advisory Data Scientist (saishruthi.tn@ibm.com)
AI Strategy and Innovation @ IBM

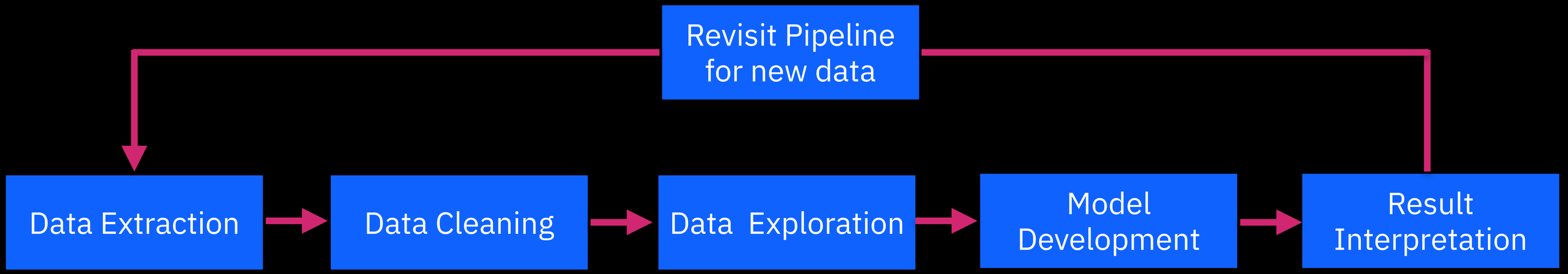
Materials: ibm.biz/rbc-workshop

What is Data Science?

Breaking down the data science definition

Goal	Find solution to the business problem
How?	Transforming problems to well-posed questions
Using?	Mathematics, programming and scientific method
Finally?	Communicate results and its business impact

Data Science Pipeline



Terminology

Dataset

Feature, field, variable, attribute or characteristics

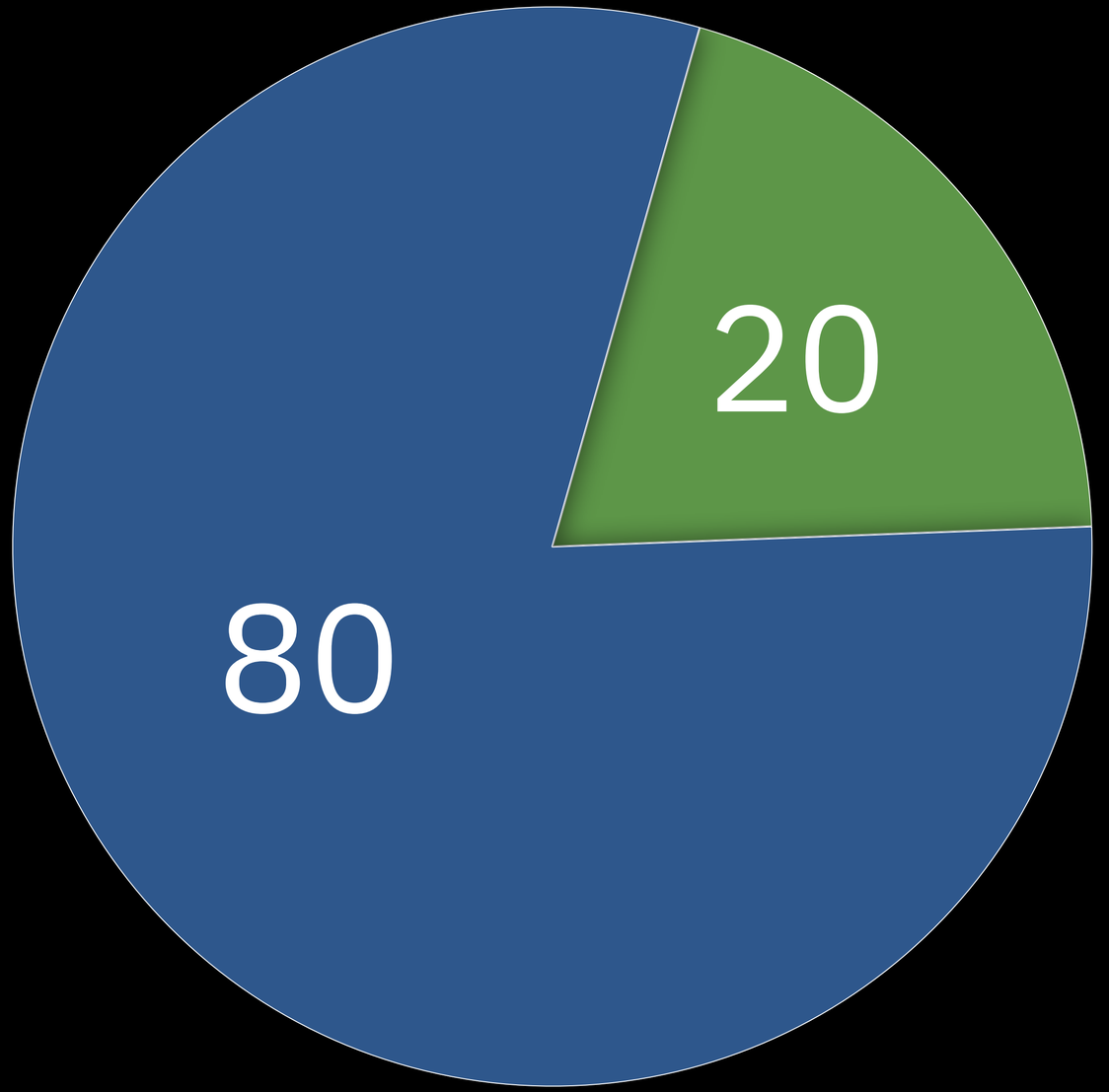
x0	x1	Dealer	Type
5	1	AA	Table
8	3	AA	Table
9	1	AA	Table
5	7	AA	Chair
6	9	AA	Chair
10	9	AA	Chair
20	40	AA	Dining
25	45	AA	Dining

Data point,
record,
sample,
entity or instance

- Type and Dealer are categorical variables
- x0 and x1 are continuous variables

Why Data Cleaning and Data Visualization are important?

The 80/20 Rule



Importance

- Work with domain experts to get more insights about the data.
- Data cleaning prevents you from building a faulty model.
- Visualizations help in understanding the data and convey analyzed information effectively to stakeholders.

“No one is going to collect data that can fit well with machine learning models”

Common data cleaning steps

- Import and export dataset
- Renaming features
- Changing type of features (e.g., float-> int)
- Selecting features from the dataset
- Filter rows and extract only records needed
- Append / Join tables
- Fill missing values
- Summarize data
- Normalizing and scaling
- Date formatting
- String variables to numeric
- Binning

Data Cleaning - Demo

Problem Statement

Data from location A

x0	x1	Dealer	Type
5	1	AA	Table
8	3	AA	Table
9	1	AA	Table
5	7	AA	Chair
6	9	AA	Chair
10	9	AA	Chair
20	40	AA	Dining
25	45	AA	Dining

Data from location B

x0	x1	Dealer	Type
2	2.0	AA	Table
4	3.0	AA	Table
7	2.0	AA	Table
2	9.0	AA	Chair
2	8.0	AA	Chair
7	8.0	AA	Chair
6	NaN	AA	Table

Handling missing values

General steps for handling missing values:

1. See how many missing data points are there in the dataset

2. Find answer for the below question:

    - Is this value missing because it wasn't recorded or because it doesn't exist?

3. Think how to handle the missing values:

    - Removing data points having missing values (not recommended as you may lose important information).

    - Try filling values.

    - Fill with 0.

Imputation techniques:

1. Continuous variable: Mean, Median, Mode, Linear Regression and Mixed Imputation.
2. Categorical variables: Use 'NA' as a separate level, Logistic Regression, K-Nearest Neighbor, etc.

Label and one-hot encoding

x0	x1	Dealer	Type
5	1	AA	Table
8	3	AA	Table
9	1	AA	Table
5	7	AA	Chair
6	9	AA	Chair
10	9	AA	Chair
20	40	AA	Dining
25	45	AA	Dining

One-hot Encoding

Length	Breadth	Table	Chair	Dining
5	1	1	0	0
5	7	0	1	0
20	40	0	0	1

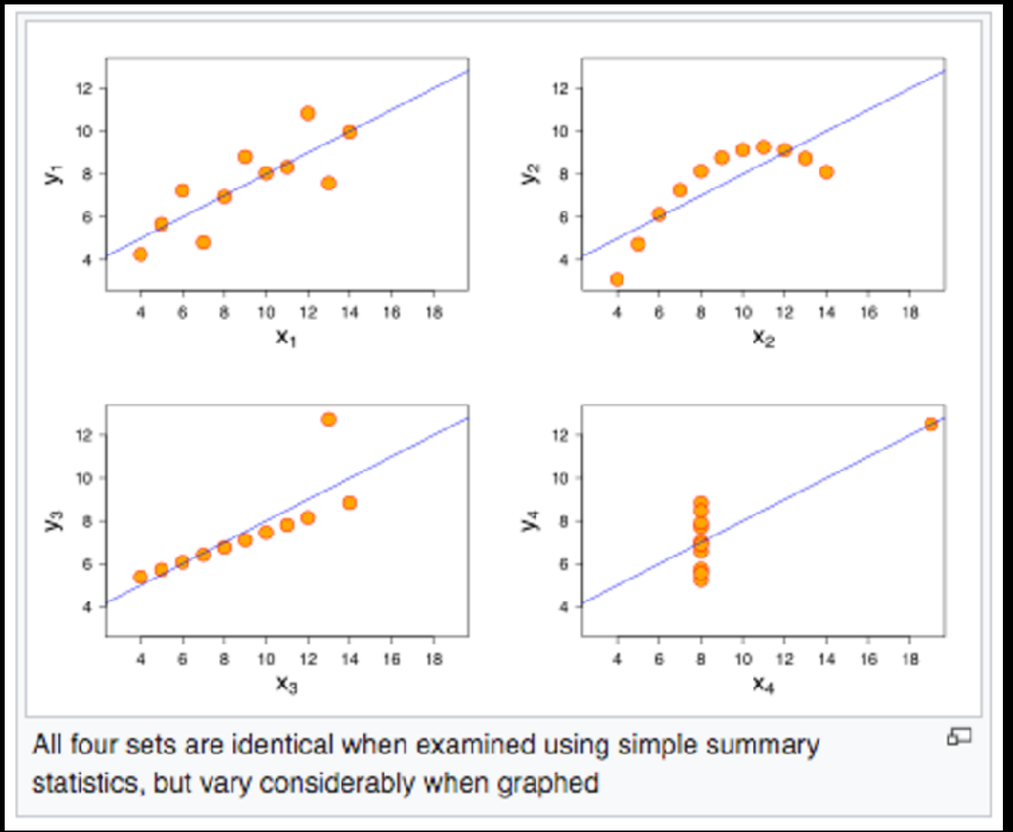
Label Encoding

Length	Breadth	Type
5	1	0
5	7	1
20	40	2

Data Visualization

Anscombe’s Quartet

Property	Value
Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression	0.67

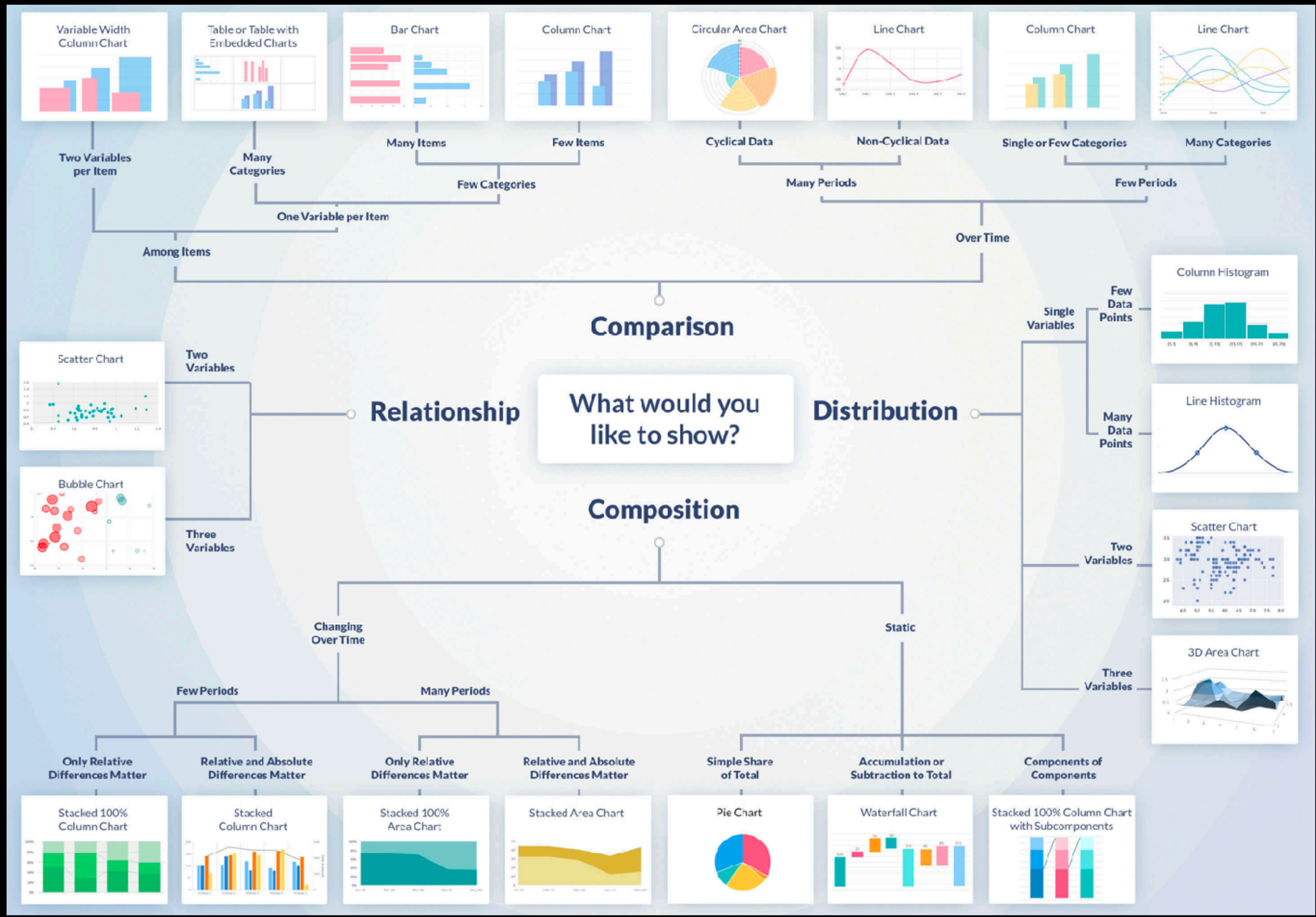


All four sets are identical when examined using simple summary statistics, but vary considerably when graphed

Source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Which graph to use?

Source: https://www.tapclicks.com/guided-visualization/



Popular Data Visualization libraries

- Matplotlib
- Seaborn
- ggplot2
- Bokeh
- pygal
- Plotly

What is Machine Learning?

Read more: https://www.ibm.com/cloud/learn/machine-learning

What is this?

Source: https://www.pexels.com/photo/white-and-grey-kitten-on-brown-and-black-leopard-print-textile-45201/

Breaking down machine learning definition

“Machine learning is the science (and art) of programming computers so they can learn from data,” writes Aurélien Géron in Hands-on Machine Learning with Scikit-Learn and TensorFlow.

ML is a subset of the larger field of artificial intelligence (AI) that “focuses on teaching computers how to learn without the need to be programmed for specific tasks,” … “In fact, the key idea behind ML is that it is possible to create algorithms that learn from and make predictions on data.”

Source: https://www.oreilly.com/ideas/machine-learning-a-quick-and-simple-definition

Types of Machine Learning

1. Supervised Learning

x0	x1	Dealer	Type
5	1	AA	Table
8	3	AA	Table
9	1	AA	Table
5	7	AA	Chair
6	9	AA	Chair
10	9	AA	Chair
20	40	AA	Dining
25	45	AA	Dining

Learning from the previous data which has both predictors (attributes) and response (target).

Types of Machine Learning

2. Unsupervised Learning

x0	x1	Dealer
5	1	AA
8	3	AA
9	1	AA
5	7	AA
6	9	AA
10	9	AA
20	40	AA
25	45	AA

The goal of this unsupervised machine learning technique is to find similarities in the data point and group similar data points together.

Types of Machine Learning

3. Semi-supervised Learning

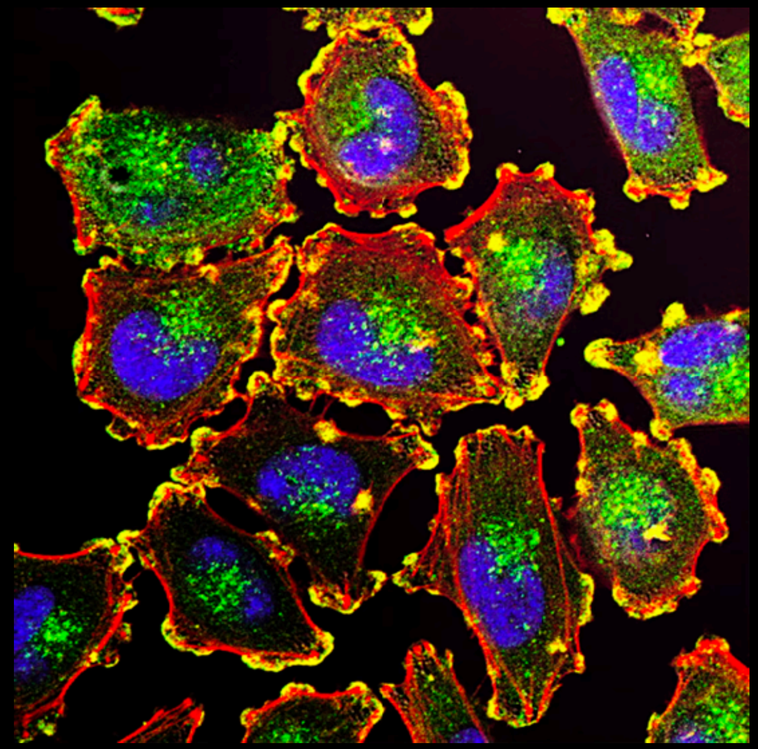
x0	x1	Dealer	Type
5	1	AA	Table
8	3	AA	
9	1	AA	
5	7	AA	
6	9	AA	Chair
10	9	AA	
20	40	AA	Dining
25	45	AA	

Training the system on partially labeled input data. Input data has a lot of unlabeled data and little of labeled data.

Types of Machine Learning

4. Transfer Learning

Reusing a model that was trained for solving one problem and applying it to a different but related problem.



Source: Photo by National Cancer Institute on Unsplash

Demo

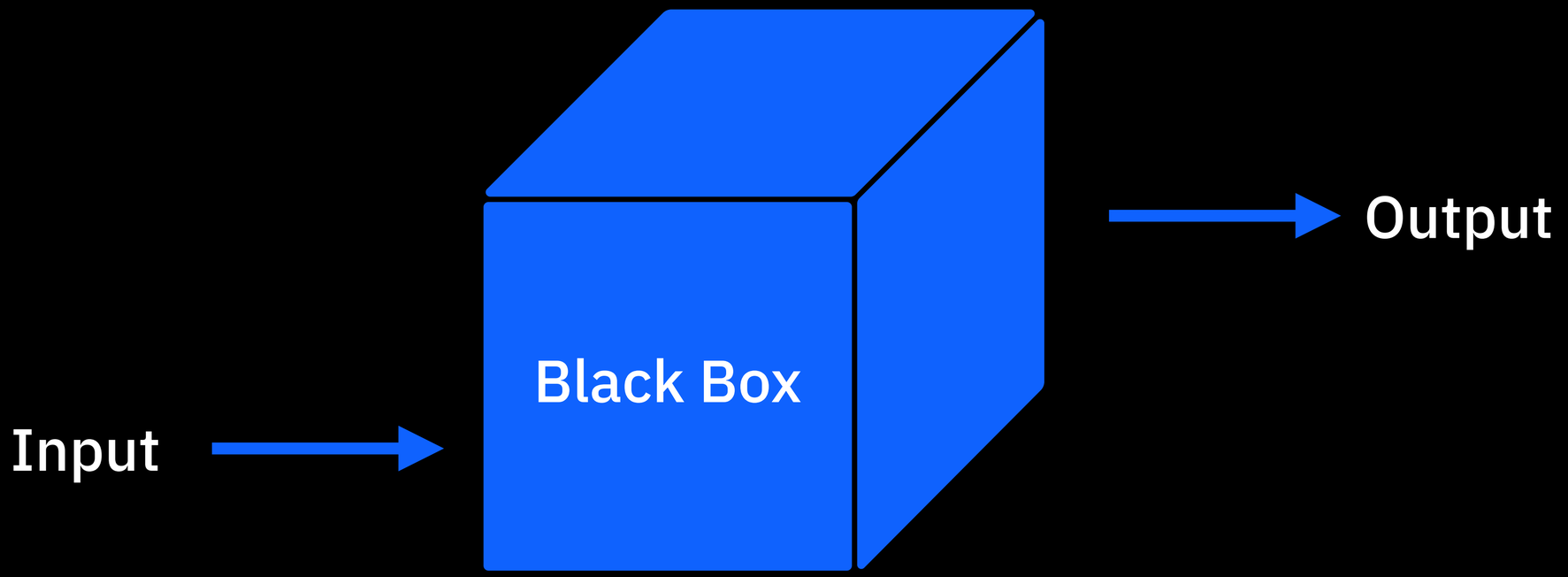
Data Science Basics

ibm.biz/rbc-workshop

Watson Studio

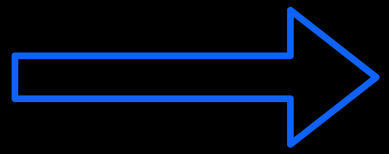
Explainability

The black box problem

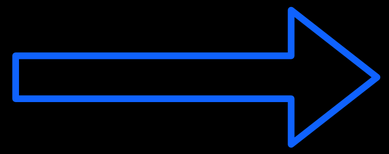


Internal behavior of the code is unknown

Dataset



AI/ML
Model



Decision
or Action

User

- Why does the model decide or do that?
- How sure is the model about this decision?
- How can I claim in case of error?
- How can I be sure there are no biases?
- Why should I trust the model?

How does a model work?

What is driving the decisions?

Can I trust the model?

Key Stakeholders

Data Scientist

- Understand the model
- Debug it
- Improve its performance

Business owner

- Understand the model
- Evaluate fit for purpose
- Agree to use

Model Risk

- Challenge the model
- Ensure its robustness
- Approve it

Regulator

- Check its impact on consumers
- Verify reliability

Consumer

- “What is the impact on me?”
- “What actions can I take?“

Reference: https://arxiv.org/pdf/2009.11698.pdf

AI Explainability 360

↳ (AIX360)

https://github.com/IBM/AIX360

AIX360

AIX360 toolkit is an open-source library to help **explain** AI and ML models and their predictions.

This includes 3 classes of algorithms: local post-hoc, global post-hoc, and directly interpretable explainers for models that use image, text, and structured/tabular data.

Toolbox

Local post-hoc
Global post-hoc
Directly interpretable

Demo

AIX 360 - interactive experience

http://aix360.mybluemix.net

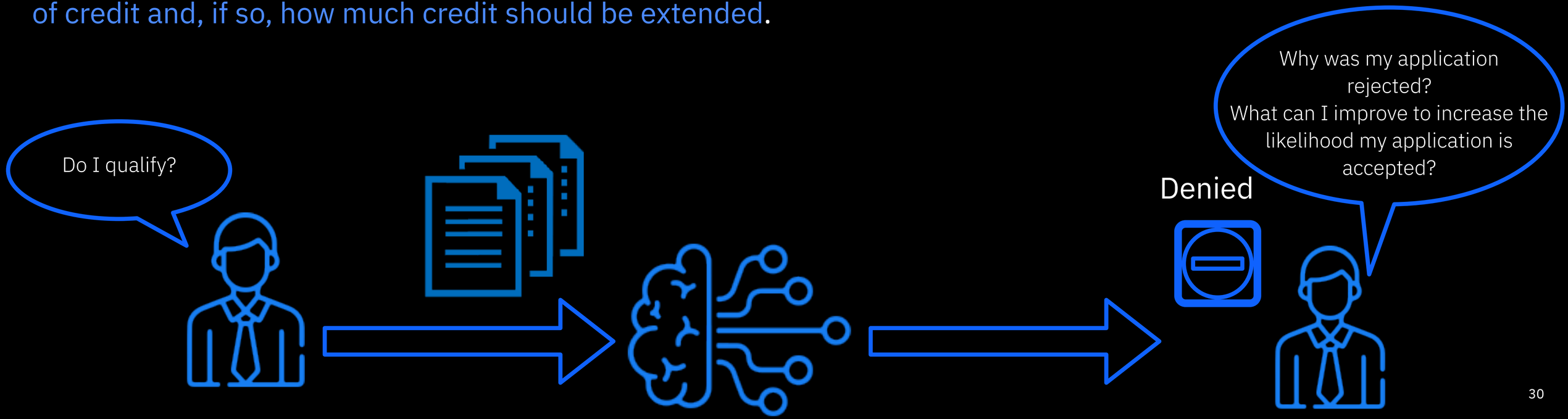
FICO Explainable Machine Learning Challenge

Use Case:

The customers in this dataset have requested a credit line in the range of $5,000 - $150,000.

The fundamental task is to use the information about the applicant in their credit report to predict whether they will make timely payments over a two-year period. This is the machine learning task that we focus on.

The machine learning prediction is then used by loan officers to decide whether the homeowner qualifies for a line of credit and, if so, how much credit should be extended.



Break

10 minutes

Demo

Explainability

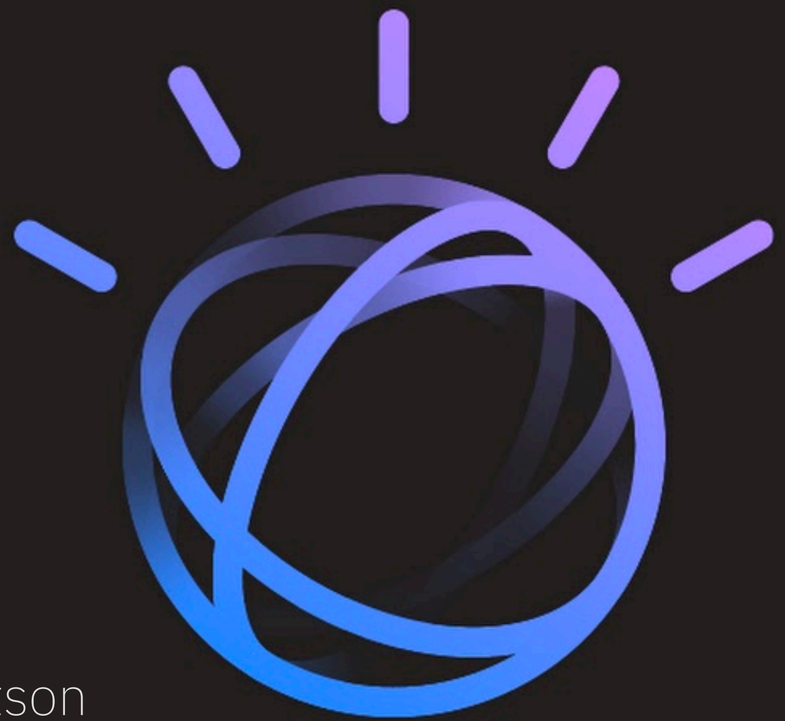
ibm.biz/rbc-workshop

Now it is your turn

ibm.biz/rbc-workshop

Thank you!

Gabriela de Queiroz, Chief Data Scientist (gdq@ibm.com)
Saishruthi Swaminathan, Advisory Data Scientist (saishruthi.tn@ibm.com)
AI Strategy and Innovation @ IBM



IBM Watson