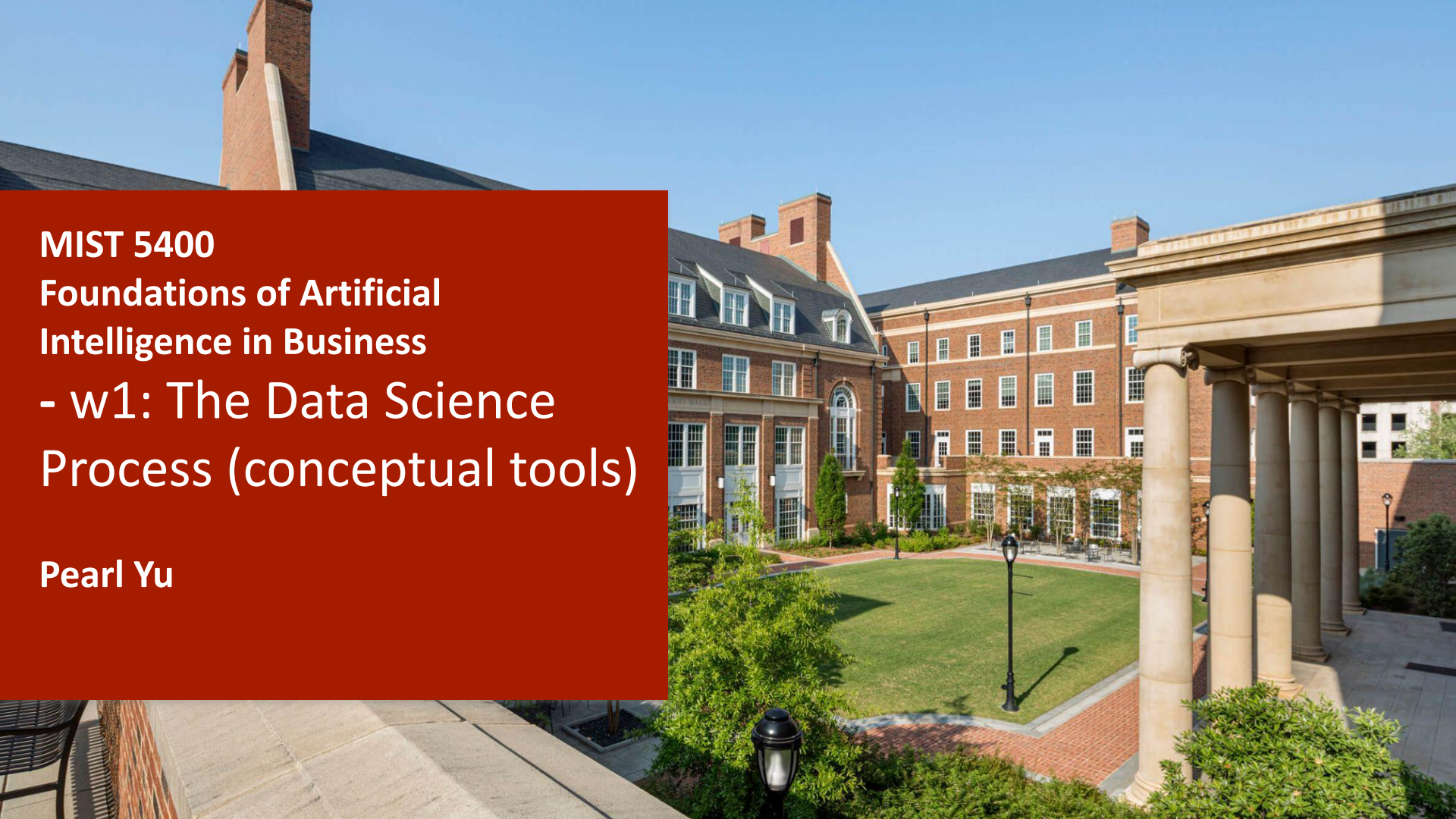


**MIST 5400**
**Foundations of Artificial Intelligence in Business**

# - w1: The Data Science Process (conceptual tools)

**Pearl Yu**

# Data Science Tools

Conceptual tools

- Analytic Space
- Data Science Process
- Predictive Analytics Flows
- Expected Value Framework

Technical tools

- Supervised / Unsupervised
- The models (ML algorithms)
- The training (Optimization)
- The evaluation (Metrics / Overfitting)

# The roles

Product management

Story telling

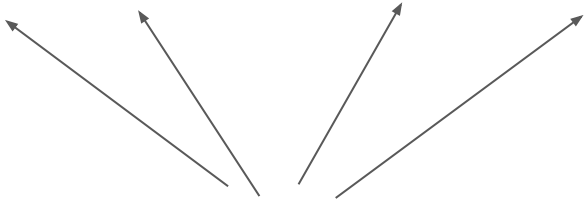
AI/ ML/ Data Engineering

Data Science

# The craft

Realizing value with data science is a **craft.**

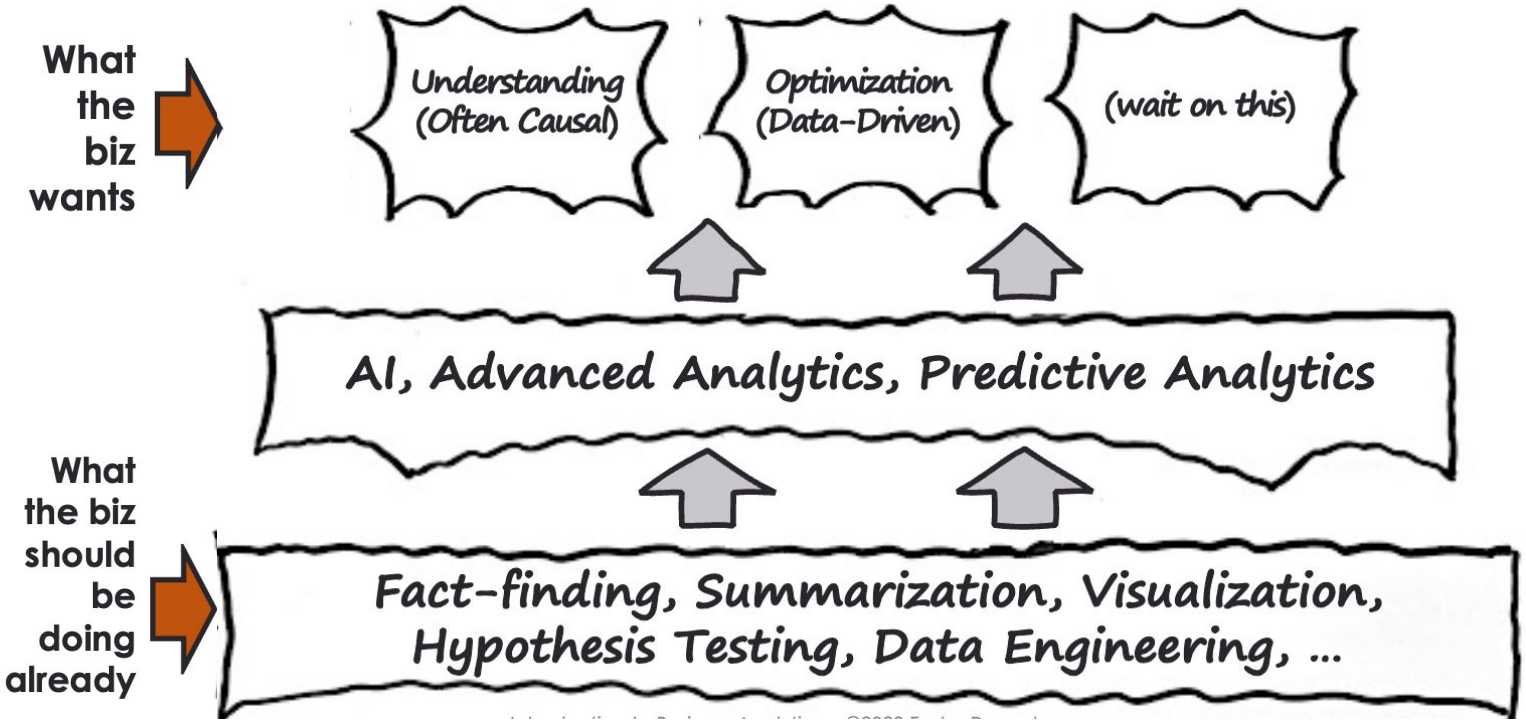
Science + experience + creativity + biz sense



Process

# What do companies use DS for?

- The Analytics Space



# Data Science Process - Customer Churn Problem

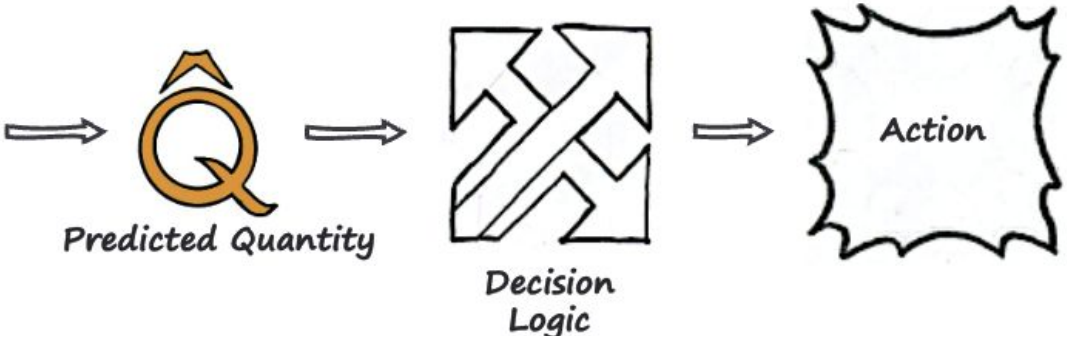
A Data Science Product Manager, has just joined TelCo, one of the largest telecommunication firms. Telco is having a major problem with churn in their wireless business.

In the mid-Atlantic region, 20% of cell-phone customers leave when their contracts expire.

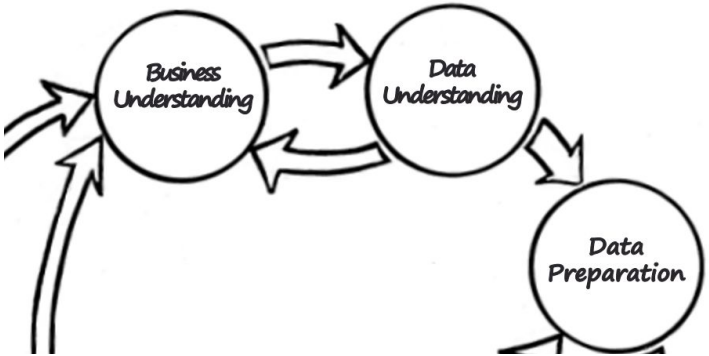
Our task: Devise a precise, step-by-step plan for how the analyst/tech team should use TelCo's vast data resource to decide which customers to target with the special retention offer prior to the expiration of their contracts.

# The Data Science Process - Business Understanding

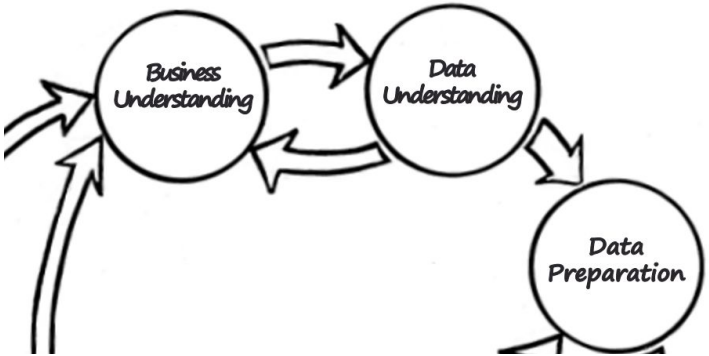
# The Data Science Process

# The Data Science Process

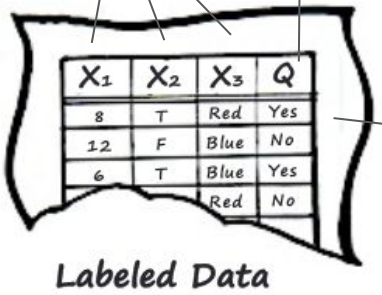


# The Data Science Process




label/target variable

features/attributes



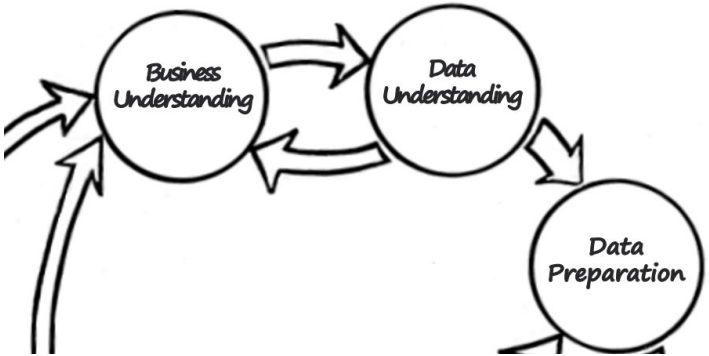

instance

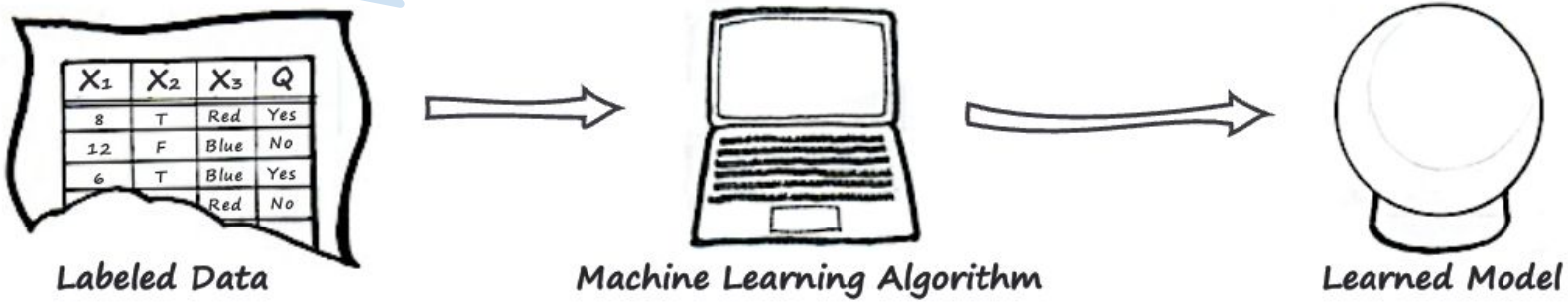
# The Data Science Process

Supervised v.s. Unsupervised learning?

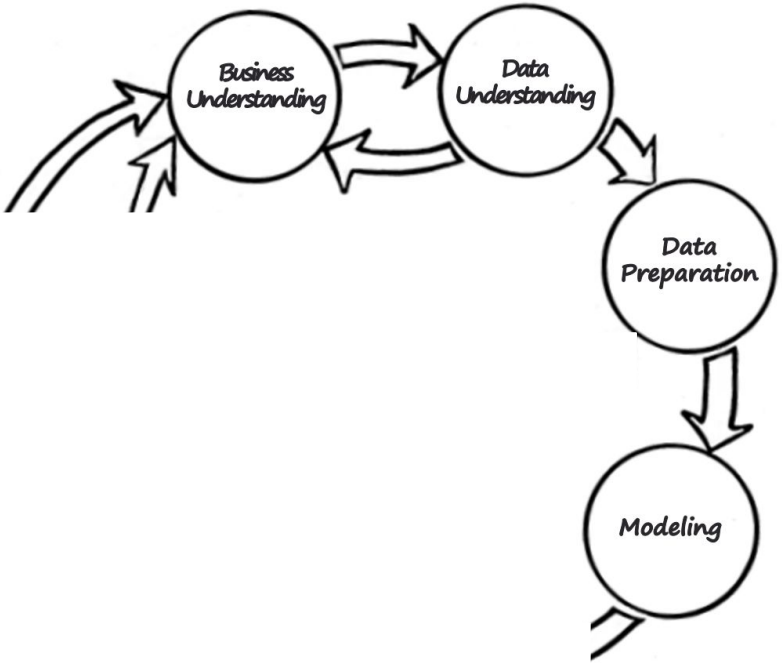
You give your learning algorithm examples to learn from, including the right answers.

After seeing correct examples, the learning algorithm eventually learn to just take the input alone and gives a prediction of Y



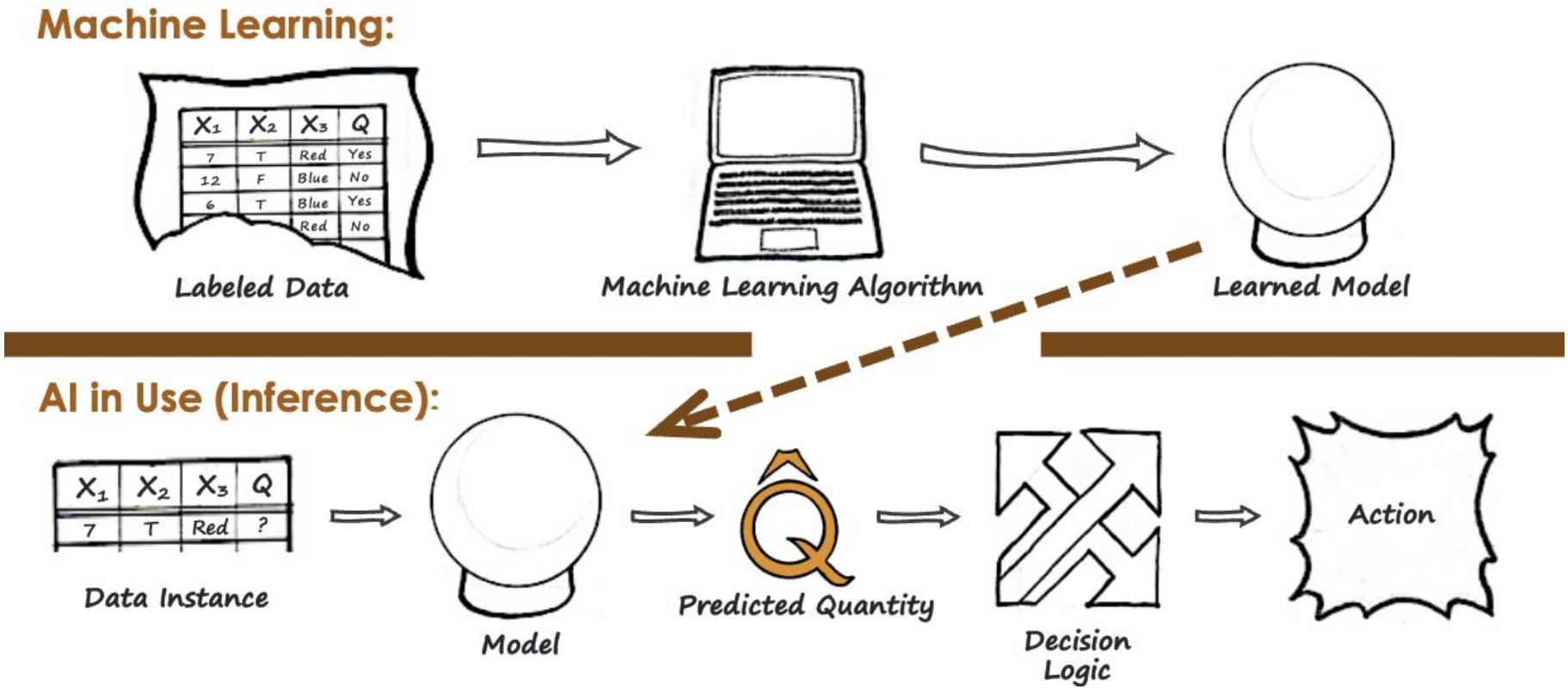



# The Data Science Process

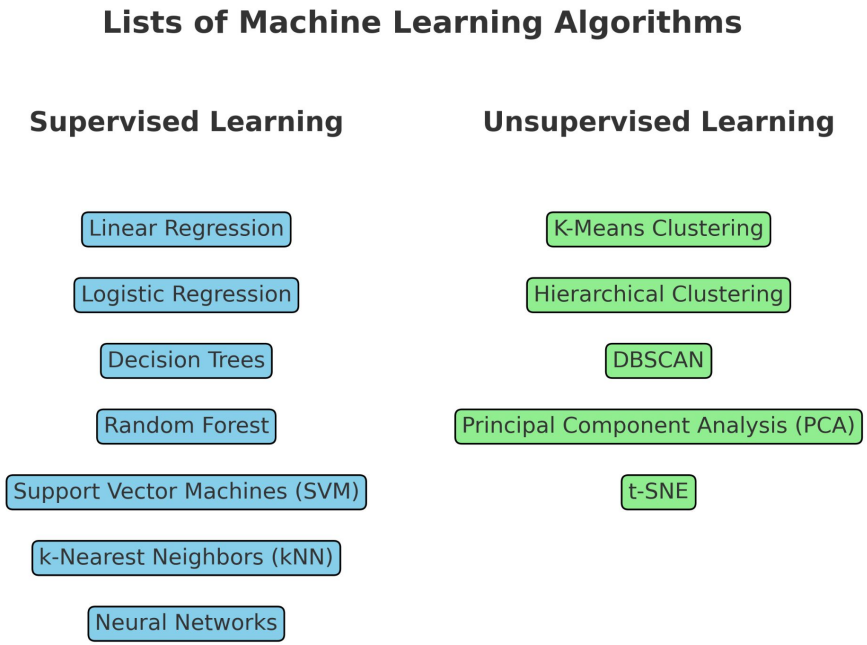


# The Data Science Process

- Predictive Analytics Flow

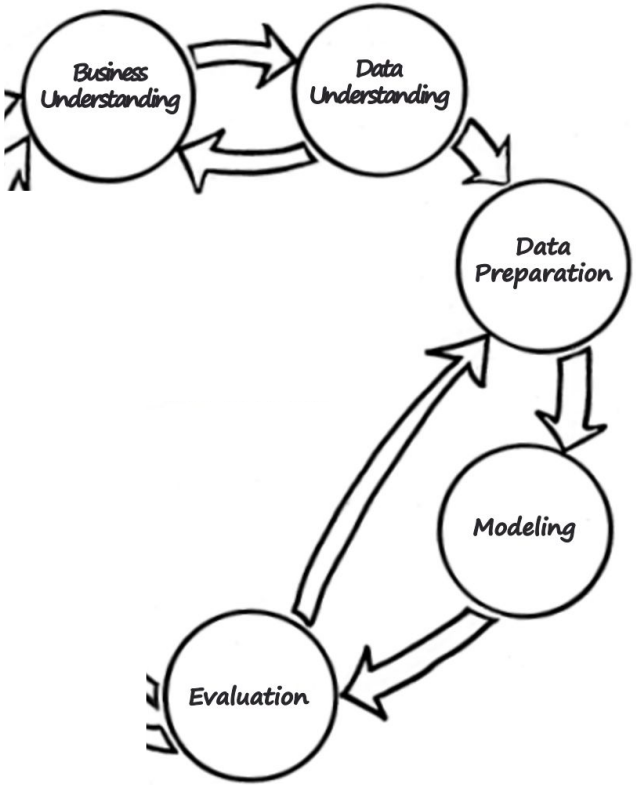


# The Data Science Process

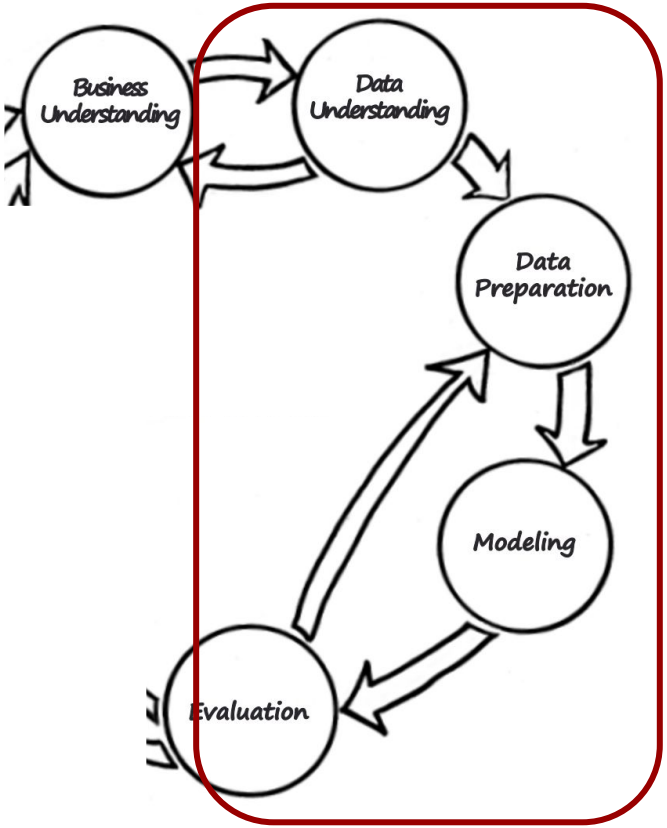



** Asked chatgpt for a list of supervised/unsupervised ML algorithms and give me a .png*

# The Data Science Process



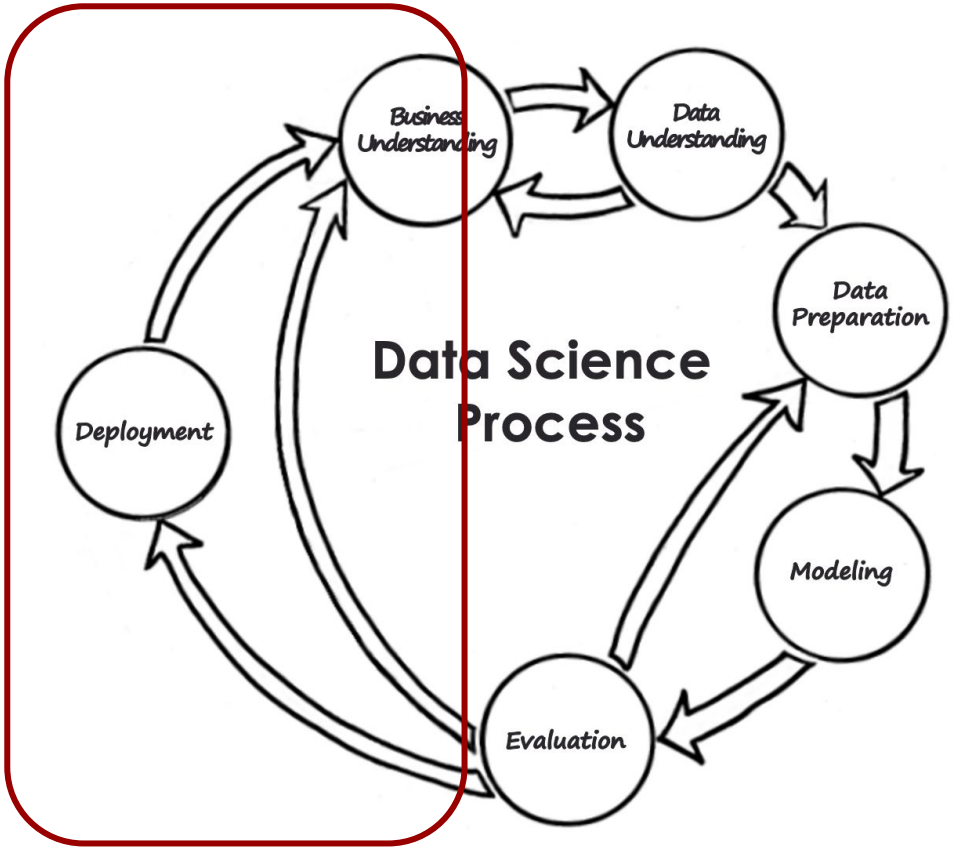
# The Data Science Process




The ‘science’ (technical) part we’ll talk about next week.

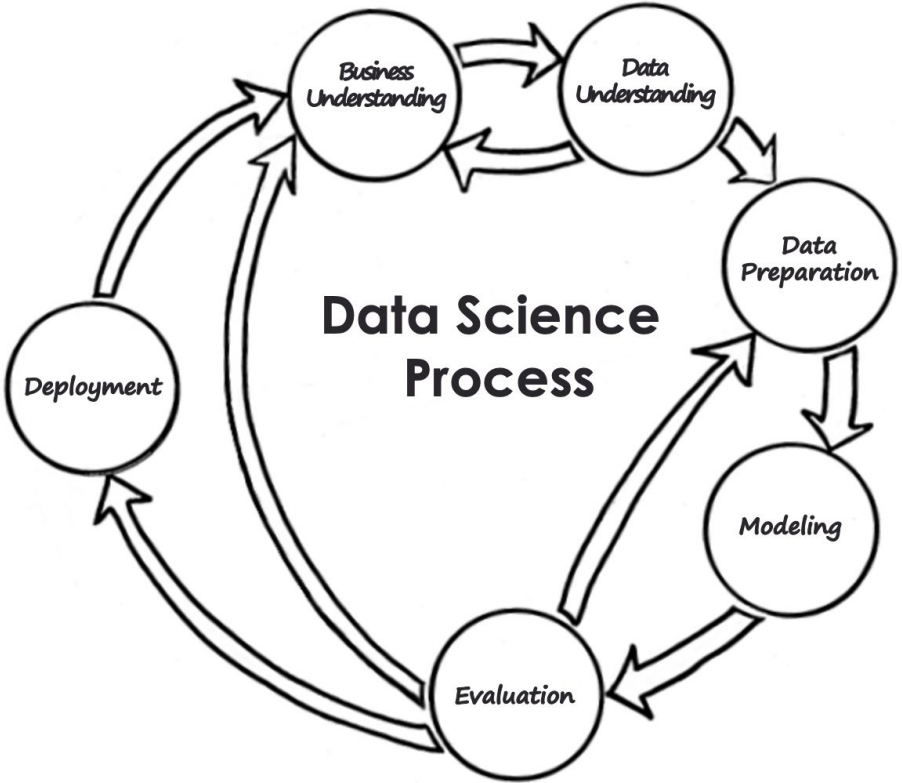
# The Data Science Process

The managing & ‘in-production’ part we’ll talk in the second half of this course.



# The Data Science Process

- Oh it's kind of an iterative process.
  Let's revisit business understanding.



# The Data Science Process

- The expected value framework

**Action:**
If I have a customer:

Send offer E[profit| send offer] = Pr(stay| send offer)*(Value of customer - offer cost)
+Pr(churn|send offer)*( 0 - offer cost)

Don't send offer E[profit| Not send offer] = Pr(stay| no offer)*(Value of customer - 0)
+Pr(churn|no offer)*( 0 - 0)

# The Data Science Process

- The expected value framework

We could predict this too, or let's assume it's the current plan price

**Action:**
If I have a customer:

Send offer E[profit| send offer] = Pr(stay| send offer)*(Value of customer - offer cost)
+Pr(churn|send offer)*( 0 - offer cost)
= 1 - Pr(stay| send offer)

Let's assume it's decided already.

= 1 - Pr(churn| no offer)

Don't send offer E[profit| Not send offer] = Pr(stay| no offer)*(Value of customer - 0)
+Pr(churn|no offer)*( 0 - 0)

So, the unknown quantities to predict (target variables)?

# The Data Science Process

- The expected value framework

We could predict this too, or let's assume it's the current plan price

**Action:**
If I have a customer:

Send offer E[profit| send offer] = Pr(stay| send offer)*(Value of customer - offer cost)
+Pr(churn|send offer)*( 0 - offer cost)
= 1 - Pr(stay| send offer)

Let's assume it's decided already.

= 1 - Pr(churn| no offer)

Don't send offer E[profit| Not send offer] = Pr(stay| no offer)*(Value of customer - 0)
+Pr(churn|no offer)*( 0 - 0)

So, the unknown quantities to predict (target variables)?

**Decision logic:**
E[profit | send offer] - E[profit |no offer] > a threshold, send offer.
How to decide the threshold: Could be 0, could be based on the budget limit, etc.

# The Data Science Process

- The expected value framework

* Asked ChatGpt to turn my hand-written notes into a prettier, readable graph.

# Key Takeaways

- The analytics space:
  - Business use AI to explore / understand / optimize
  - We start with summaries, descriptive understanding then build (AI) models to achieve these purposes.
- The data science process:
  - ALWAYS start from business understanding
  - Data → Model learning → evaluation, this sciency part
  - The deployment, the management, in-production part.
  - It could be an iterative cycle.
- The predictive analytics flows to guide the business understanding / task formulation:
  - Inference is using the trained/learnt model to make predictions.
  - There'll be some predicted quantity, and DECISION LOGIC.
- The Expected value framework allows considerations of costs/benefits of decisions into task formulation,

# MegaTelco Case

Your task is to devise a precise, step-by-step plan for how the analyst/tech team should use MegaTelCo's vast data resource to decide which customers to target with the special retention offer prior to the expiration of their contracts. Be specific as to what data to use and how to use them, and specifically how the team should decide on the set of customers to target to best reduce churn for a particular incentive budget.

# MegaTelco Case

Predicted Quantity?

How likely a customer is to churn?
How likely a customer is to take the offer if churns?

Decision logic?

To send the offer or not.
[after the contract expires]

# MegaTelco Case

What would be an instance?

A customer

What would the action?

To send the offer or not.
[after the contract expires]

# MegaTelco Case

What would be an instance?

A customer

What would the action?

To send the offer or not.

# Model Performance Analytics

**Meeting Time:** 08/14/2022 - 12/02/2022 Tue, Thu 3:55 PM - 5:10 PM

**Etiquette:** Attendance is required. Let me know if you can't make it.

- Fitting the data and overfitting the data, holdout testing, cross-validation, learning curves
-

# Logistics

**Meeting Time:** 08/14/2022 - 12/02/2022 Tue, Thu 3:55 PM - 5:10 PM

**Etiquette:** Attendance is required. Let me know if you can't make it.

- **Model Performance Analytics II: Profit, Lift, ROC analysis, expected value framework, domain knowledge validation**
-

# Logistics

**Meeting Time:** 08/14/2022 - 12/02/2022 Tue, Thu 3:55 PM - 5:10 PM

**Etiquette:** Attendance is required. Let me know if you can't make it.

- A key skill in "Analytical Engineering" is decomposing the business problem into subproblem