

ANALYTICS EXCHANGE NYC

LEARNING SUMMIT

WEDNESDAY, OCTOBER 20 - THURSDAY, OCTOBER 21

DATA ANALYSIS

DATA ENGINEERING

DATA GOVERNANCE

Sign up for more events at
bit.ly/anexagenda





LEARNING SUMMIT

WEDNESDAY, OCTOBER 20 - THURSDAY, OCTOBER 21

The Art of Feature Selection & Feature Engineering

Sharon Morris, Data Science Leader, PGIM

Agenda

- Why is feature engineering important?
 - What is feature engineering?
 - Machine learning process
 - Feature engineering for numeric data
 - Feature engineering for categorical data
 - Feature engineering tips
-

Why is feature engineering important?

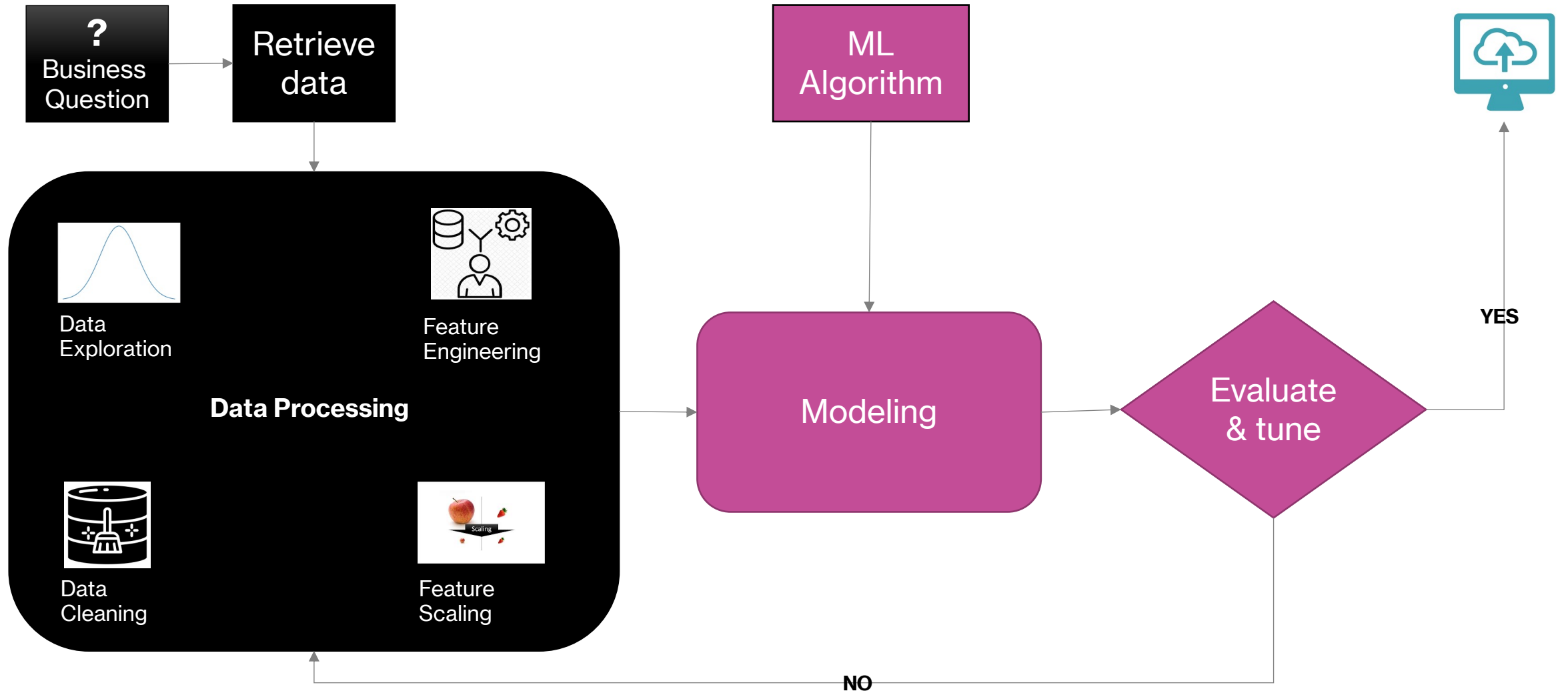
- Create models with that are simpler and easier to interpret and communicate to stakeholders.
- As number of predictors increases, the chance of finding correlations among a predictor or a set of predictors increases. Such correlations among predictors called *multicollinearity* inflates standard errors of coefficients, potentially leading to erroneous conclusions about the relevance of a predictor.
- In cases where $p > n$, traditional estimation techniques will not work.

What is feature engineering?

A **feature** is a numeric representation of raw data.

Feature engineering is the process of using domain knowledge of the data to transform existing features or to create new variables from existing ones, for use in machine learning.

Data science process



Feature Engineering Tips



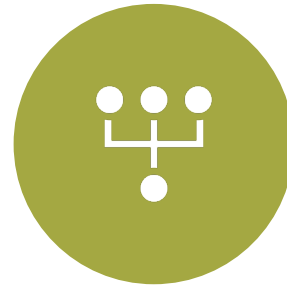
Work with subject matter experts with domain knowledge to understand data fields.



Resample imbalanced data.



Calculate the correlation between features.



Consider the business goals for the model and work backwards.

Featuring engineering for numeric data



Jupyter Notebook