Azure Machine Learning

**Modeling - Feature Engineering
For R-ladies Taiwan**

Speaker : Mia Chang

Azure ML

- **Introduce myself**

#Data Consultant
#Algorithm Research
#Azure Taiwan

Our Target Audience Today...

.

who are just starting out with data science and machine learning
who are looking for ways to systematize their feature engineering efforts

It does not assume mastery of mathematics or statistics. Experience with linear algebra, probability distributions, and optimization are helpful, but not necessary.

Azure ML Modeling - Feature Engineering
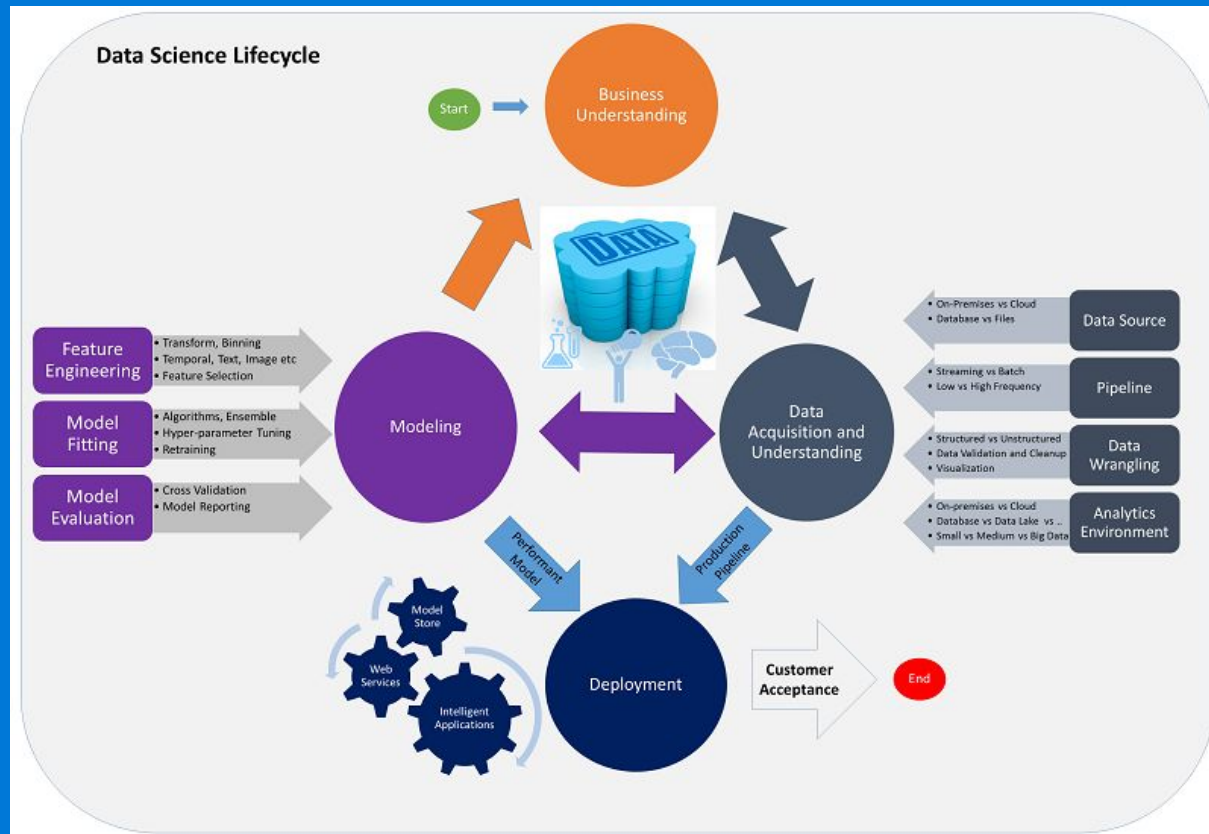
- **Introduction**
  - Team Data Science Process (TDSP)
  - Feature engineering in data science

- **Demo**

  - Where to find it in Azure ML?
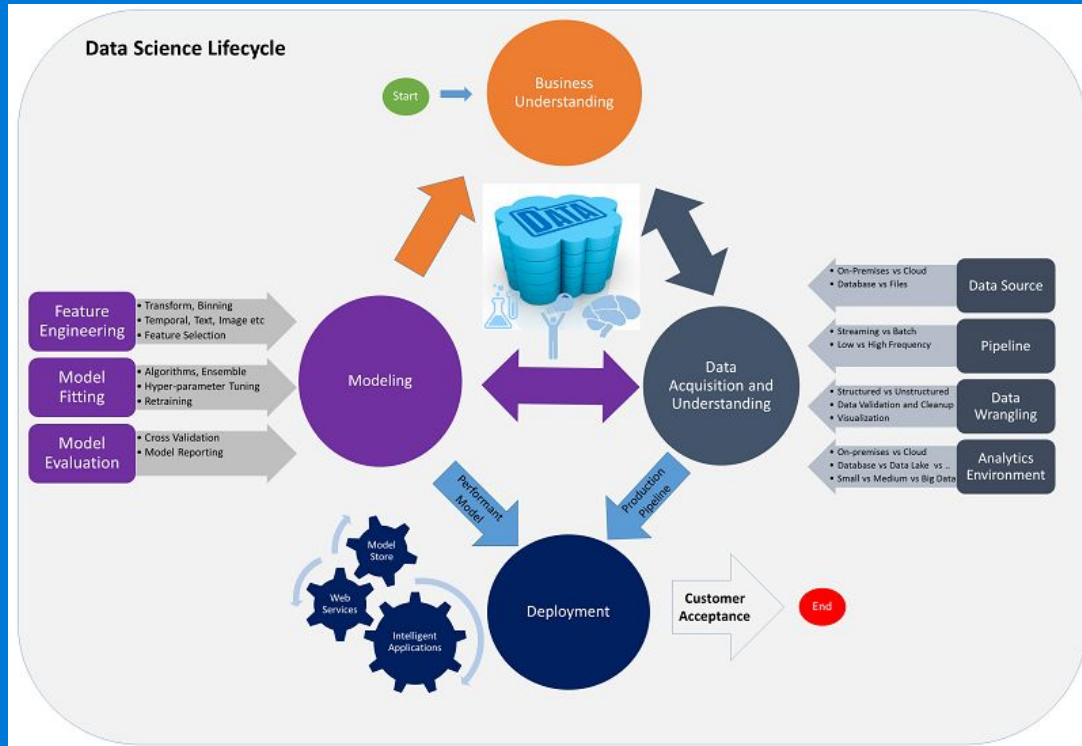 - What if I want to customize my feature engineering process?

- **Recap and Reference**

TDSP Team Data Science Process lifecycle

- Business Understanding
- Data Acquisition and Understanding
- Modeling
- Deployment
- Customer Acceptance
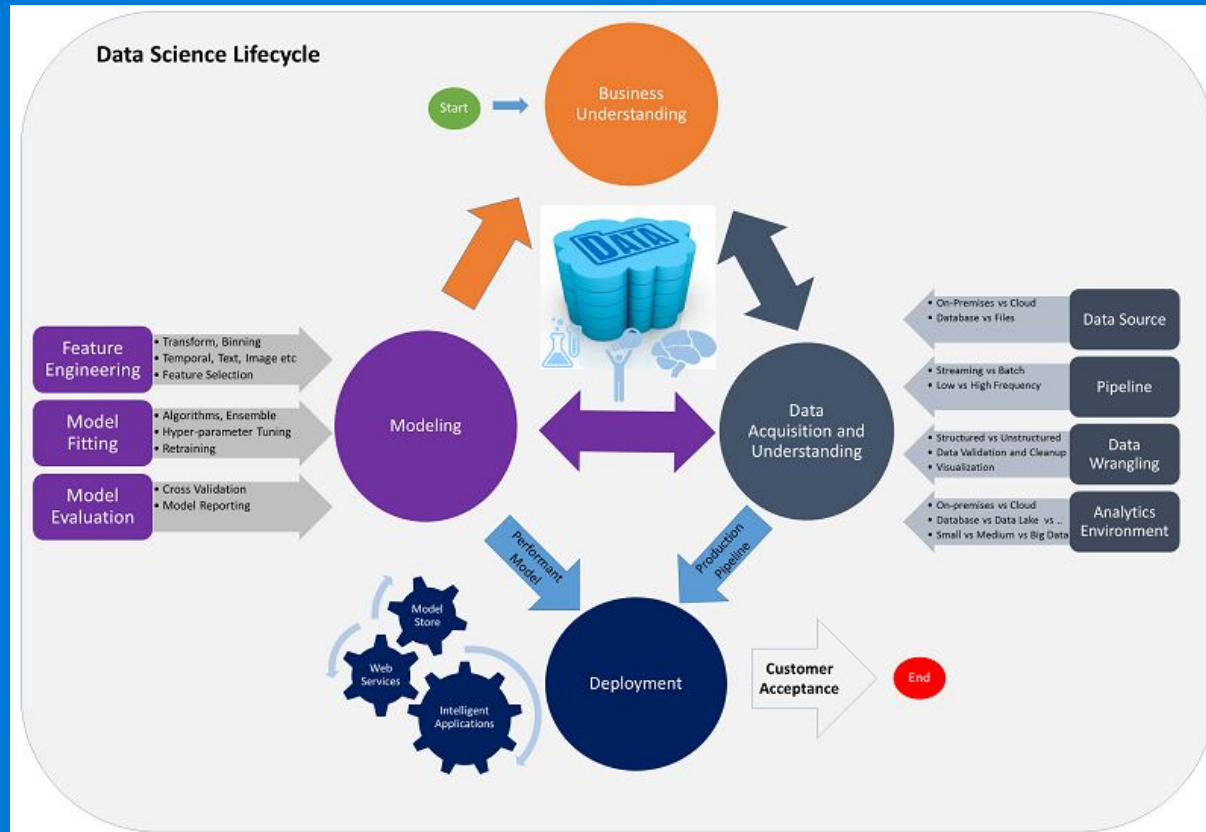
TDSP Team Data Science Process lifecycle

▪ Business Understanding

Define Objectives: Work with your customer and other stakeholders to understand and identify the business problems. Formulate questions that define the business goals and that data science techniques can target.
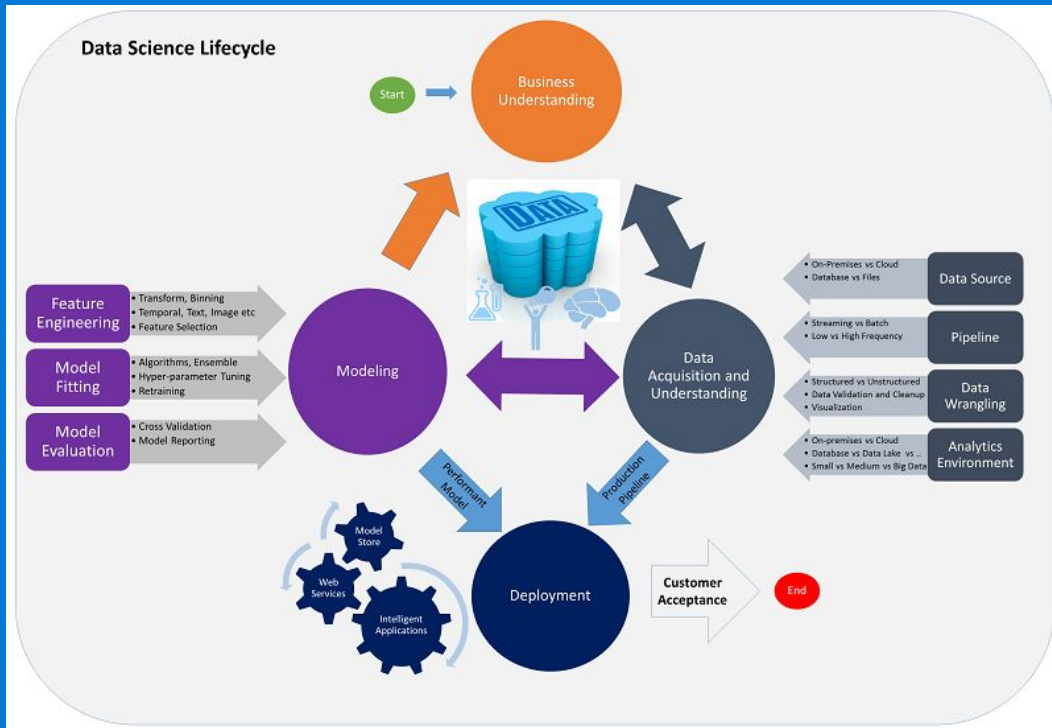
Identify data sources: Find the relevant data that helps you answer the questions that define the objectives of the project.

TDSP Team Data Science Process lifecycle

- Business Understanding
- Data Acquisition and Understanding

TDSP Team Data Science Process lifecycle

▪ Business Understanding
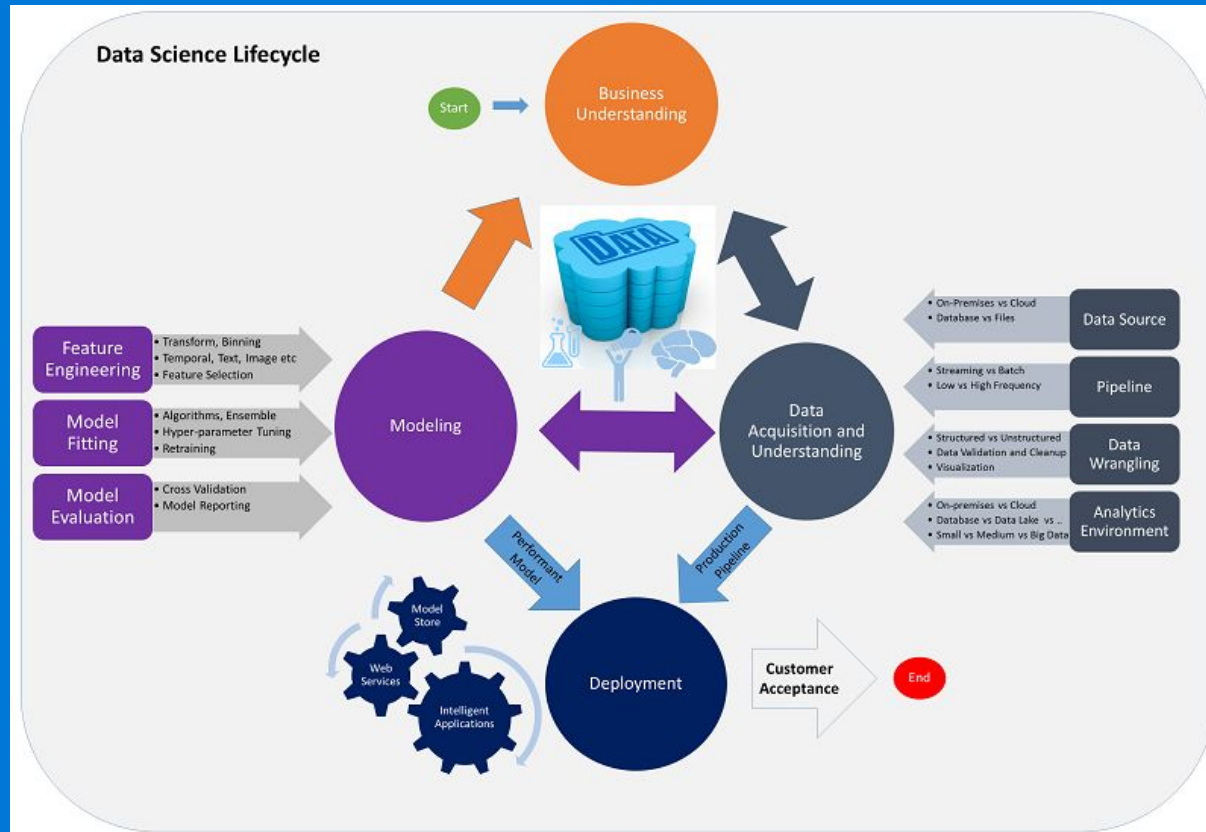▪ Data Acquisition and Understanding

Ingest the data into the target analytic environment.

Explore the data to determine if the data quality is adequate to answer the question.

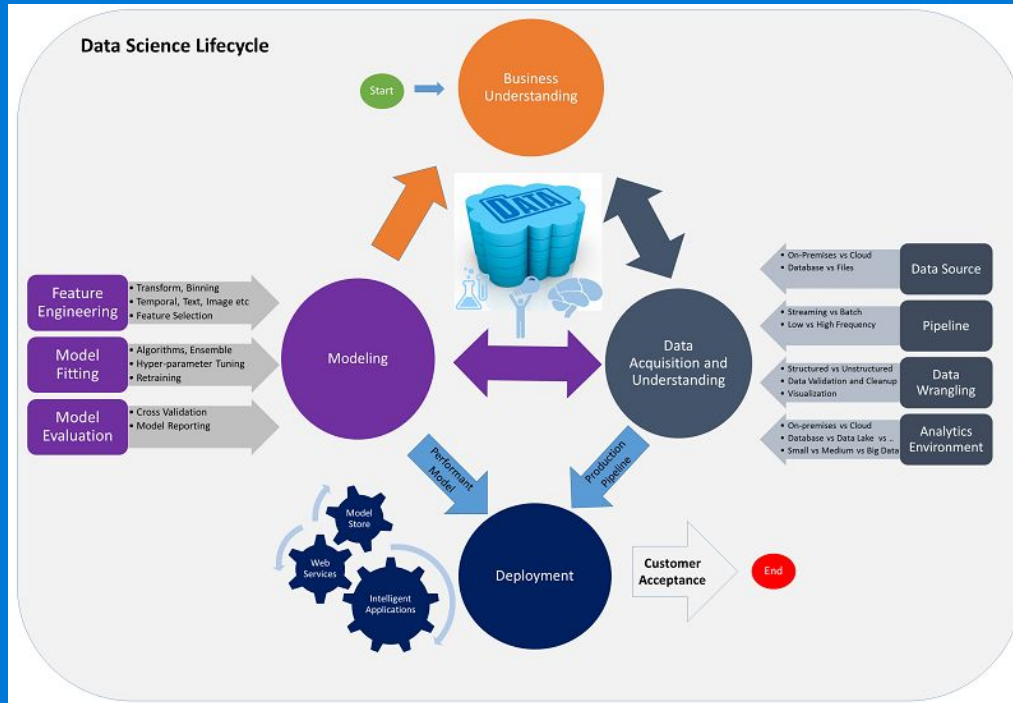Set up a data pipeline to score new or regularly refreshed data.

TDSP Team Data Science Process lifecycle

- Business Understanding
- Data Acquisition and Understanding
- Modeling

TDSP Team Data Science Process lifecycle

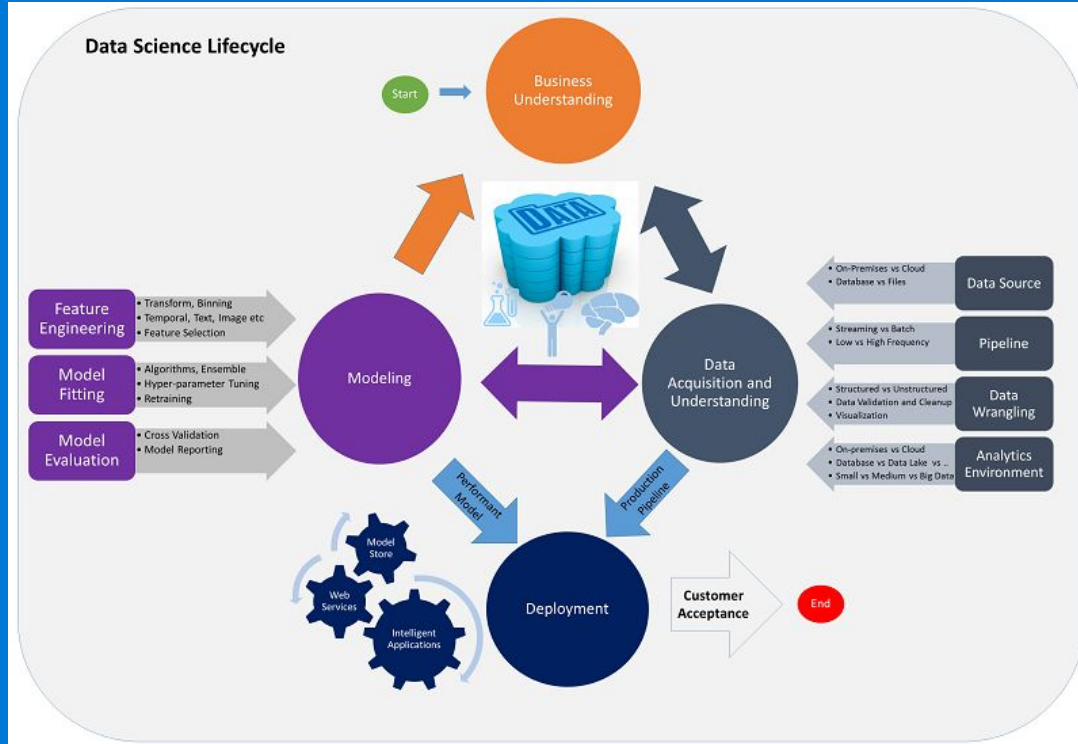Data Science Lifecycle

- Business Understanding
- Data Acquisition and Understanding
- Modeling
    Goals
        Optimal data features for the machine learning model
    How to do it
- Deployment
- Customer Acceptance

TDSP Team Data Science Process lifecycle

▪ Business Understanding
▪ Data Acquisition and Understanding
▪ Modeling
     Goals
     How to do it
          Feature Engineering
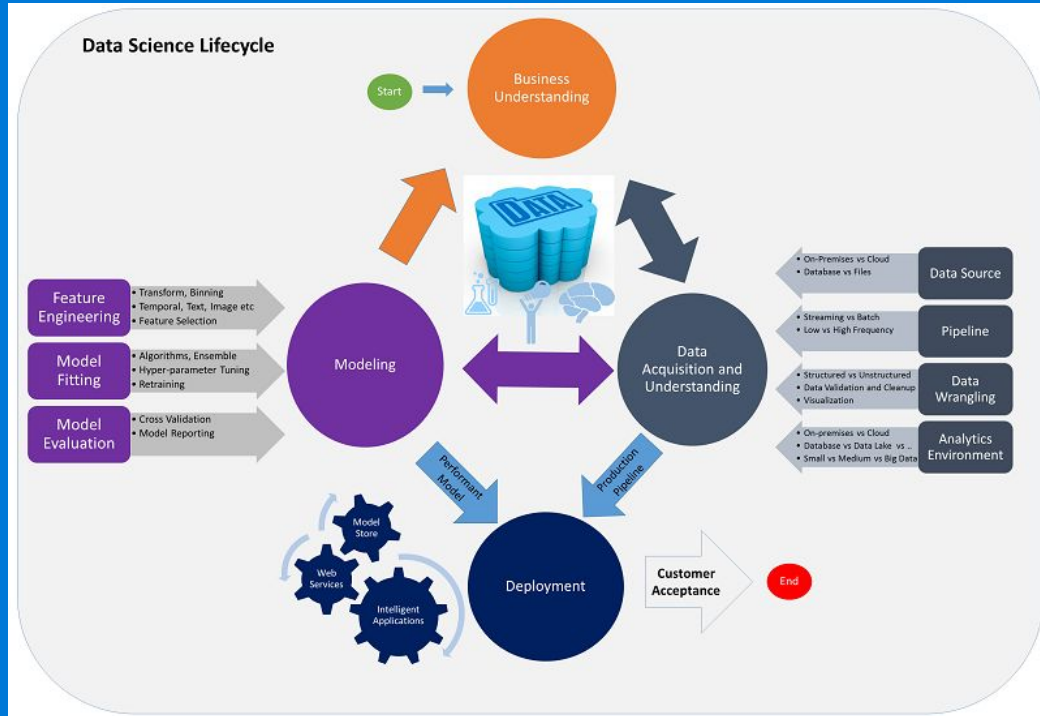   Model training
          Suitable for production
▪ Deployment
▪ Customer Acceptance

TDSP Team Data Science Process lifecycle

▪ Business Understanding
▪ Data Acquisition and Understanding
▪ Modeling
▪ Deployment

Operationalize the model

TDSP Team Data Science Process lifecycle

Data Science Lifecycle

▪ Business Understanding
▪ Data Acquisition and Understanding
▪ Modeling
▪ Deployment
▪ Customer Acceptance

System validation: confirm the deployed model and pipeline are meeting customer needs.

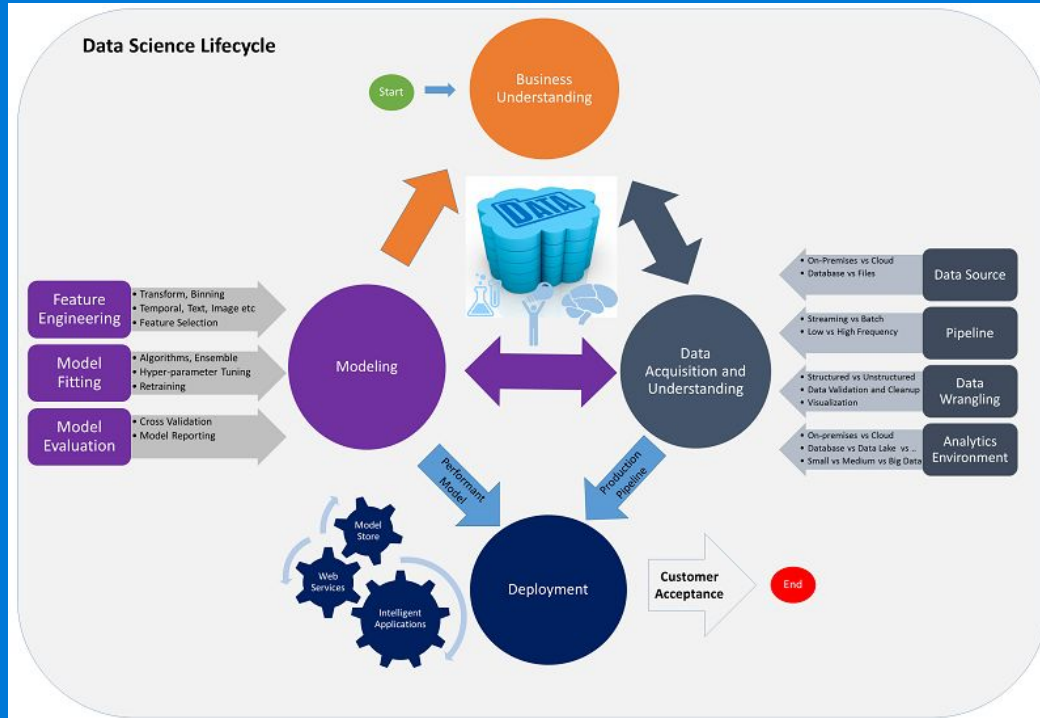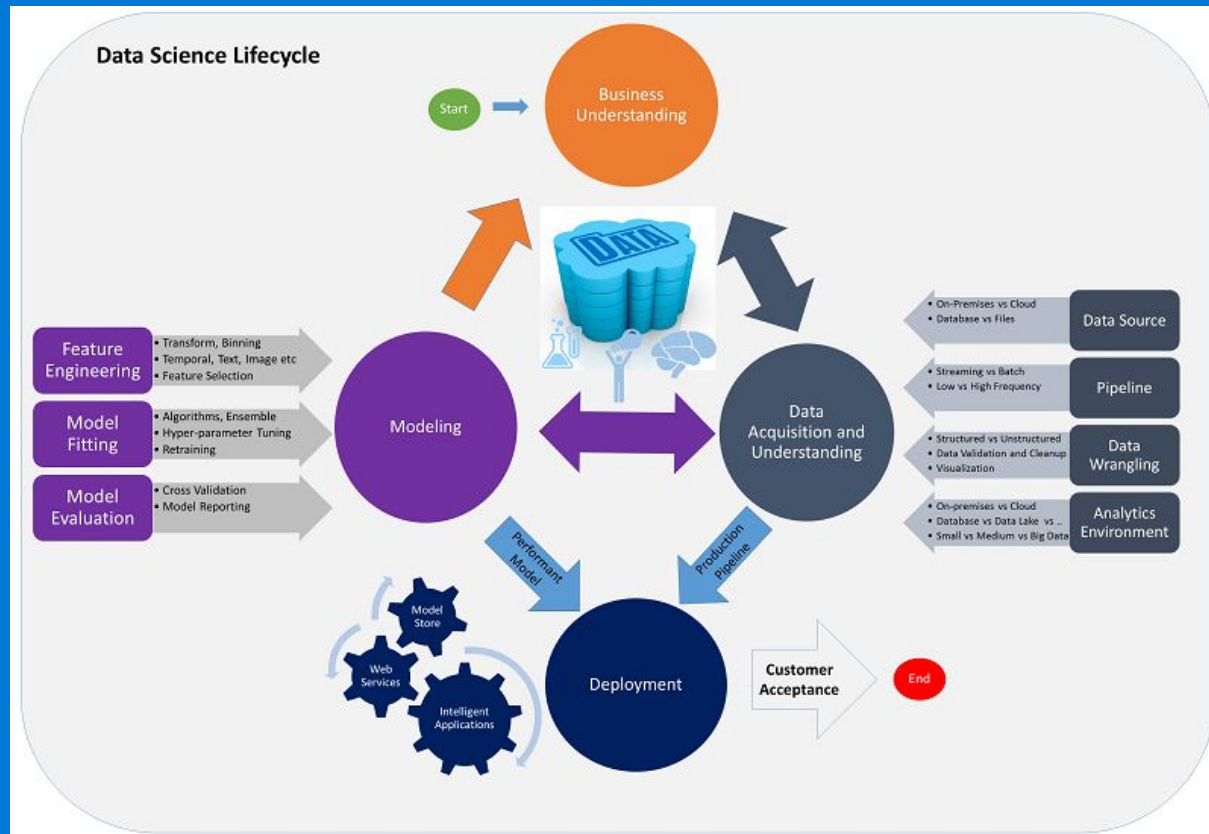Project hand-off: to the entity that will run the system in production.

TDSP Team Data Science Process lifecycle

- Business Understanding
- Data Acquisition and Understanding
- Modeling
- Deployment
- Customer Acceptance

Feature Engineering in Data science

where domain expertise and creativity can pay high dividends

Why Data Scientists take so much time on Preprocessing,
(Data Cleaning and Feature Engineering)?

Feature engineering sits right between "data" and "modeling"
in the machine learning pipeline for making sense of data.

What is feature in Data Science?

A Feature -
a peice of information that is potentially useful for **prediction**.
Where domain know how can contribute its power.
A feature is a numeric representation of raw data.

What do we usually do when doing feature engineering?
▪ how to represent text data or image data,
▪ how to reduce dimensionality of auto-generated features,
▪ when and how to normalize

Feature Engineering in Data science
where domain expertise and creativity can pay high dividends

For example

- **Numerical continously data > bend category data**
    - huminity + temperature > good / bad day
    - age > kid / adult
    - age + gender > Cangotowar

- And there are some features that are NOT relevant to what we want.
    - Student ID > Grads
    - age / gender > the behavior of watching drama

Feature Engineering Process
1. Brainstorming features
2. Decide what feature to create
3. Create the feature
4. Study the impact of features on model
5. Iterating the feature if it's useful
6. Repeat 1-3

Feature Engineering in Data science

where domain expertise and creativity can pay high dividends

Feature engineering attempts to
increase the predictive power of learning algorithms by creating features from raw data

Feature engineering (特徵工程設計) : This process attempts to create additional relevant features from the existing raw features in the data, and to increase the predictive power of the learning algorithm.從資料中的現有原始特性建立其他相關特性，以及增加學習演算法的預測功效。

Feature selection (特徵選取) : This process selects the key subset of original data features in an attempt to reduce the dimensionality of the training problem.嘗試選取主要的原始資料特性子集，排除不相關、多餘或高度相關的特性，以縮小定型問題的維度。

Deep Learning in feature engineering.

Azure ML Modeling - Feature Engineering

▪ Feature Engineering - Creating Features from Your Data
>> Data Transformation - Quantize Data (Group Data Into Bins)
>> Text Analytics - Feature Hashing

▪ Feature Selection - Filtering Features from Your Data
>> Feature Selection -   Filter Based Feature Selection
Fisher Linear Discriminant Analysis
Permutation Feature Importance

2 Demo

- Where to find it in Azure ML?

- What if I want to customize my feature engineering?
  Retail Forecasting: Step 3 of 6, feature engineering

3 Recap and Reference

Azure Machine Learning

Modeling - Feature Engineering

- Use when prediction
- Optimal data features for the machine learning model
- Numerical continously data > bend category data
- Feature Engineering and Feature Selection
- Demo

- Azure ML Feature Engineering – Convert to Indicator Values
- Feature Engineering and Preprocessing
- Beginning Azure ML Part 6 - Feature Engineering
- Neural Networks 2: machine learning = feature engineering

Microsoft

Mia Chang

#Data Analyst, Data Consultant

#Algorithm Research

#Co-Founder - Azure Taiwan Community

Email:mia5419@gmail.com