Project Details:

* **Flight delay** – Determine whether or not a flight will be delayed.
  + You work for a travel booking website that seeks to improve the customer experience for people whose flights were delayed.
  + The company wants to let customers know, when they book their flights, if their flights will be delayed due to weather.
  + You need to use ML to identify whether a flight will be delayed due to weather.
* **Recommendation engine** - Recommend content based on Amazon.com customer reviews.
  + You work for a startup that delivers on-demand video streaming services to users.
  + The company wants to introduce recommendations based on each user’s viewing history.
  + You need to use ML to create a recommendation engine.
* **Fraud detection** - Identify fraudulent activities using Amazon SageMaker.
  + You work for a multinational bank.
  + Over the last few months, there has been a significant uptick in the number of customers experiencing credit card fraud.
  + You need to use ML to identify fraudulent credit card transactions before they have a larger impact on your company.

*Note:* While you won’t be asked to explicitly share out on Modules 0-6, you should still take notes and think critically. Getting into the ML mindset at the very beginning and taking notes on your thoughts, questions, and ideas will help you immensely when it comes time to present or apply what you learned to your ML problem at work.

# **Step 1: Problem formulation and getting started**

Project selection:

## Questions to consider:

1. What is the business problem?
2. What is the business goal?
3. What is the business metric?
4. Why is the business problem appropriate for machine learning?
5. What type of ML should be used?
6. Reframe the business problem as a machine learning problem:

# **Iteration I**

# **Step 2: Data exploration and data preprocessing**

# Data preprocessing and visualization

## Questions to consider:

1. Did you have to make any assumptions about the data?
2. What does exploratory data analysis and visualization tell you about the data?
3. What techniques did you use to clean and preprocess your data?

# **Step 3: Model training**

## Questions to consider for model training:

1. What percentage of the data should be training, validation, and test?
2. Did you randomize the split? If not, how might that impact the model?
3. What algorithm should be used?

# **Step 4: Model evaluation**

## Questions to consider for model evaluation:

1. What metric did you choose to evaluate your model?
2. How did your model do on your chosen metric?
3. What did you learn from your evaluation metric?

# **Iteration II**

# **Step 5: Adding Feature engineering and hyperparameter optimization**

## Questions to consider:

1. Which features engineering features were used and why?
2. What AWS SageMaker hyperparameter tuning did you use and what was the impact on the model?
3. What is the correlation between metric and individual hyperparameters?

# **Final Thoughts**

1. What is the original business success metric and the current model performance?
2. If you had more time what would you do?
3. What are the key lessons you learned during this project?