**Project: Capstone Project 2: Initial Project Ideas**

1. **Mechanical Fitting Failure Classification:** Title 49 of the Code of Federal Regulations (49 CFR Parts 191, 192) requires gas distribution pipeline operators to submit reports on an annual basis of all hazardous leaks that involve a mechanical fitting (DOT Form PHMSA F-7100.1-2).

The CFR defines a hazardous leak, as well as criteria for submitting reports to the Office of Pipeline Safety. The following kinds of information are collected:

* Key Report information
* Date and Location information
* Type of Mechanical Fitting Involved information
* Apparent Cause of Leak information
* Manufacturer information

PHMSA also uses the data for inspection planning and risk assessment.

**Problem Objective:**

* Identify long and short-term trends at the national, state and operator-specific levels.
* Determine insight into the safety metrics for frequency, causes, and fittings involved.

Data Source : <https://www.kaggle.com/binovi/mechanical-fitting-failure-data>

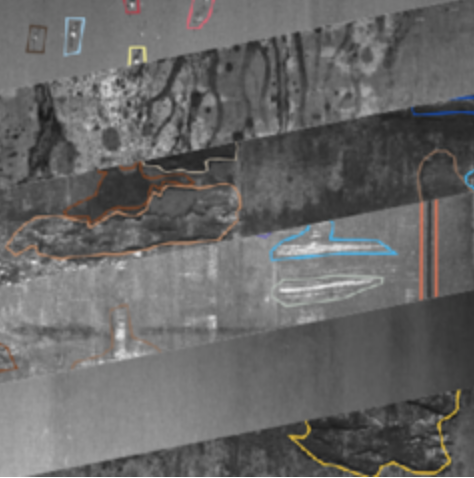
1. **Amazon Fine Food Reviews:** This dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

**Contents**

* Reviews.csv: Pulled from the corresponding SQLite table named Reviews in database.sqlite
* database.sqlite: Contains the table 'Reviews'

**Data includes:**  
- Reviews from Oct 1999 - Oct 2012  
- 568,454 reviews  
- 256,059 users  
- 74,258 products  
- 260 users with > 50 reviews

Data Source: <https://www.kaggle.com/currie32/summarizing-text-with-amazon-reviews/data>

1. **Steel Defect Detection:** [Severstal](https://www.severstal.com/eng/) is leading the charge in efficient steel mining and production. They believe the future of metallurgy requires development across the economic, ecological, and social aspects of the industry—and they take corporate responsibility seriously. The company recently created the country’s largest industrial data lake, with petabytes of data that were previously discarded. Severstal is now looking to machine learning to improve automation, increase efficiency, and maintain high quality in their production.
   1. The production process of flat sheet steel is especially delicate. From heating and rolling, to drying and cutting, several machines touch flat steel by the time it’s ready to ship. Today, Severstal uses images from high frequency cameras to power a defect detection algorithm.
   2. In this competition, you’ll help engineers improve the algorithm by localizing and classifying surface defects on a steel sheet.

Data Source: <https://www.kaggle.com/c/severstal-steel-defect-detection>