Example: Space Mission Classification

- Build a machine learning model to predict whether space flights in outer space would be a success or a failure given different variables of the missions

- Classification or Regression problem?
  - Model the state of the mission given temperature, nature of the payload, payload target orbit, etc.

Check the dataset

- Take a look at space_mission.csv
- Target (last column): success/failure (1/0)
- Input features include
  - Company: categorical data
  - Temperature (F), Wind speed (mph), Humidity (percentage)
  - Vehicle_type: categorical data
  - Liftoff thrust (kn), Payload to orbit (kg), Rocket height (m), Fairing diameter (m), Payload (kg)
  - Payload orbit: categorical data

Dealing with the data

- Categorical data: Nominal vs. Ordinal
  - Ordinal: small, medium, large -> 0, 1, 2
  - Nominal: red, green, blue -> 1, 2, 3 -> 001, 010, 011
    - Or frequency coding: count how many times each category appears in the dataset
- Incomplete data:
  - Discard the entire entry (row) or discard the corresponding features (column)
  - Let's discard those features that have "NA"

What functions do we need?

- Read the data in
- Separate the data into categorical and numeric
- Remove the "NA" columns
- Replace the categorical text by a numeric value
- Scale the data (done)
- Machine learning (done)
- Testing and reporting (almost done)
  - Accuracy, tp, tn, fp, fn, precision, recall