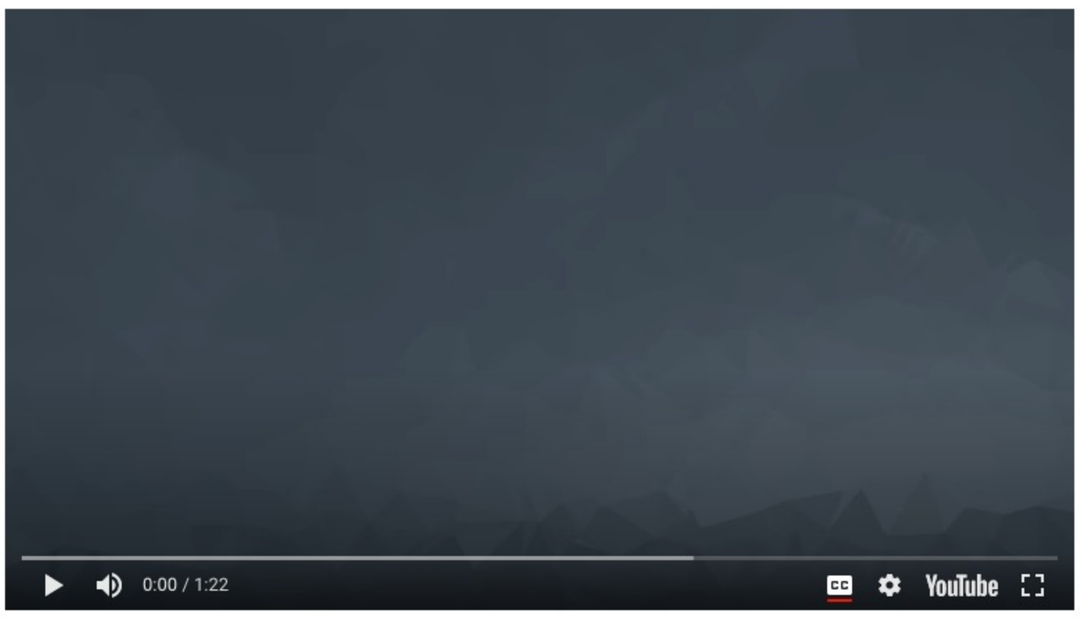


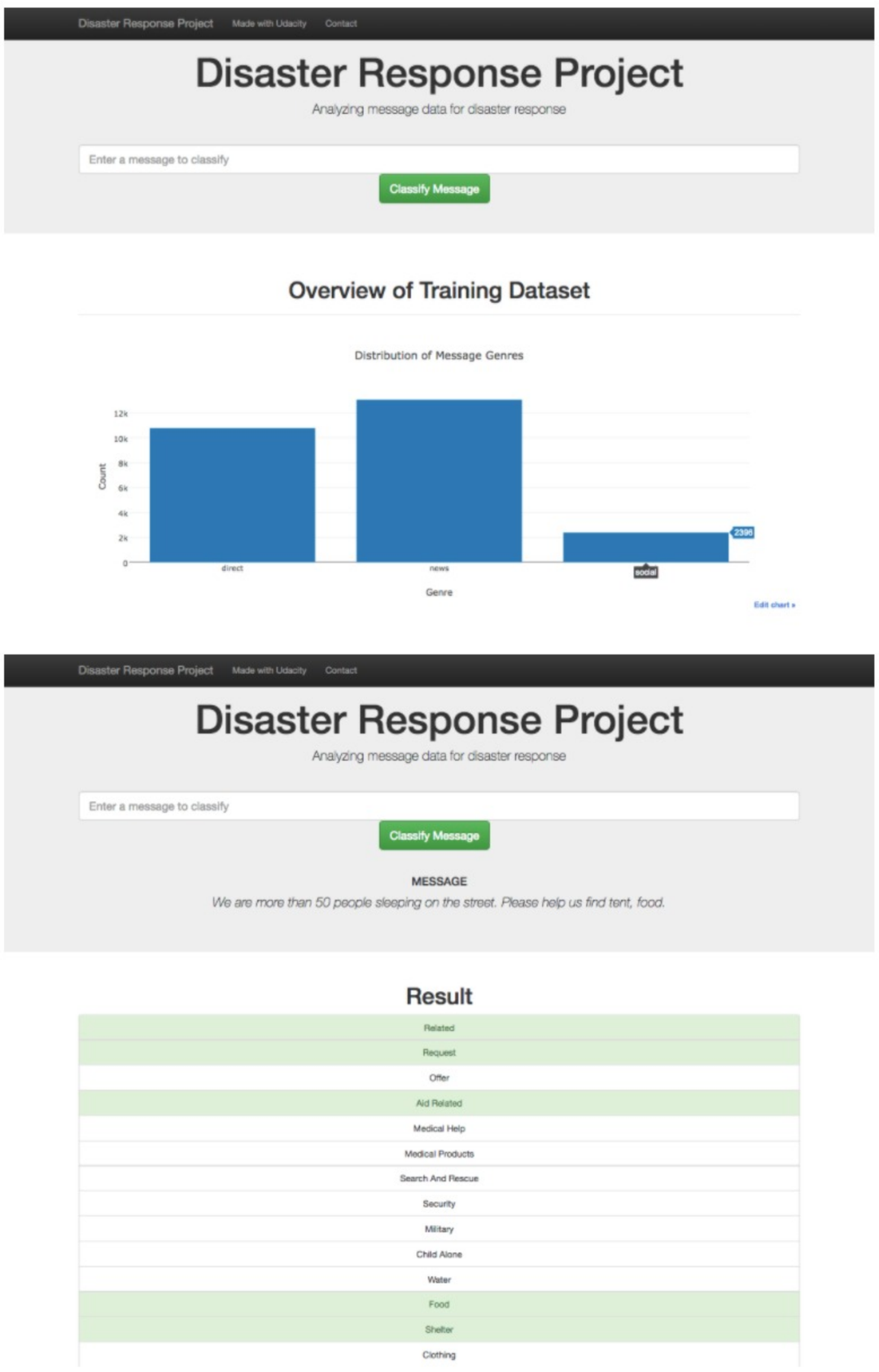
Project Overview

In this course, you've learned and built on your data engineering skills to expand your opportunities and potential as a data scientist. In this project, you'll apply these skills to analyze disaster data from Figure Eight to build a model for an API that classifies disaster messages.

In the Project Workspace, you'll find a data set containing real messages that were sent during disaster events. You will be creating a machine learning pipeline to categorize these events so that you can send the messages to an appropriate disaster relief agency.

Your project will include a web app where an emergency worker can input a new message and get classification results in several categories. The web app will also display visualizations of the data. This project will show off your software skills, including your ability to create basic data pipelines and write clean, organized code!

Below are a few screenshots of the web app.



Project Components

There are three components you'll need to complete for this project.

1. ETL Pipeline

In a Python script, `process_data.py`, write a data cleaning pipeline that:

- Loads the `messages` and `categories` datasets
- Merges the two datasets
- Cleans the data
- Stores it in a SQLite database

2. ML Pipeline

In a Python script, `train_classifier.py`, write a machine learning pipeline that:

- Loads data from the SQLite database
- Splits the dataset into training and test sets
- Builds a text processing and machine learning pipeline
- Trains and tunes a model using GridSearchCV
- Outputs results on the test set
- Exports the final model as a pickle file

3. Flask Web App

We are providing much of the flask web app for you, but feel free to add extra features depending on your knowledge of flask, html, css and javascript. For this part, you'll need to:

- Modify file paths for database and model as needed
- Add data visualizations using Plotly in the web app. One example is provided for you

Github and Code Quality

Your project will also be graded based on the following:

- Use of Git and Github
- Strong documentation
- Clean and modular code

Follow the RUBRIC when you work on your project to assure you meet all of the necessary criteria for developing the pipelines and web app.

NEXT