1. Dataset: https://www.kaggle.com/andrewmvd/face-mask-detection

    - I chose this dataset because it represents a comprehensive spectrum of data points relative to what I want to test my model with. In order to best train my model, I must identify what is correct vs what is wrong, and this data set provides a large number of such images with a large variety of ethnicities, ages and mask colours that help strengthen the eventual model I create.

2. Methodology:

    Data Preprocessing

    - With the data at hand, I plan to split the data into three sections: training data, testing data and validation data where each section is further split into two labels 'wearing or not wearing a mask'.
    - Each image will be rescaled to the same height and width

    Model

    - I want to predict whether the individual in the camera that is hooked onto my raspberry pi is wearing a mask or not
    - **Machine Learning Algorithm to use:** CNN (Convolutional Neural Network)
        - **Pros:** less pre-processing of data is required than other machine learning algorithms and mainly used to identify objects in images which is the primary focus of my project
        - **Cons:** computationally demanding, requires large GPUs to process

    - **Evaluation Metric:** Confusion Matrix, Accuracy / logistic loss

3. **Application:**
    - User stands in front of the webcam / camera and the output will either encircle their face with a red box (meaning no mask) or a green box (wearing a mask)