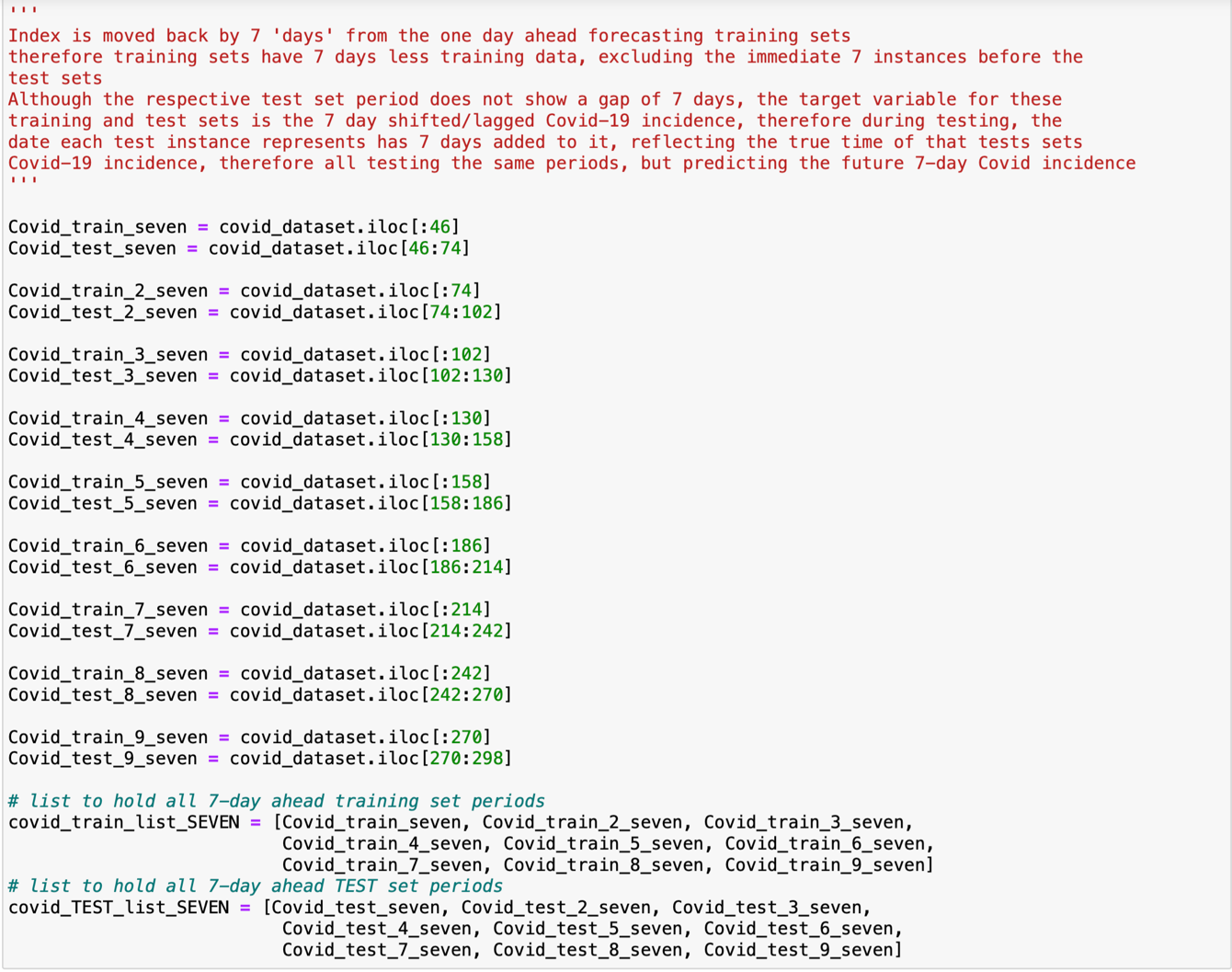
Below shows relevant code snippets for the forecasting exercise, from defining training/test sets, to our feature validation procedure, hyperparameter tuning, walk-forward testing, and our generation of our reported forecasting results

1. Defining the Training and Test set periods for the 1-day ahead, 7-day ahead, and 14-day ahead forecasting exercises respectively

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

2. Below is how we generate average importance scores for each training set

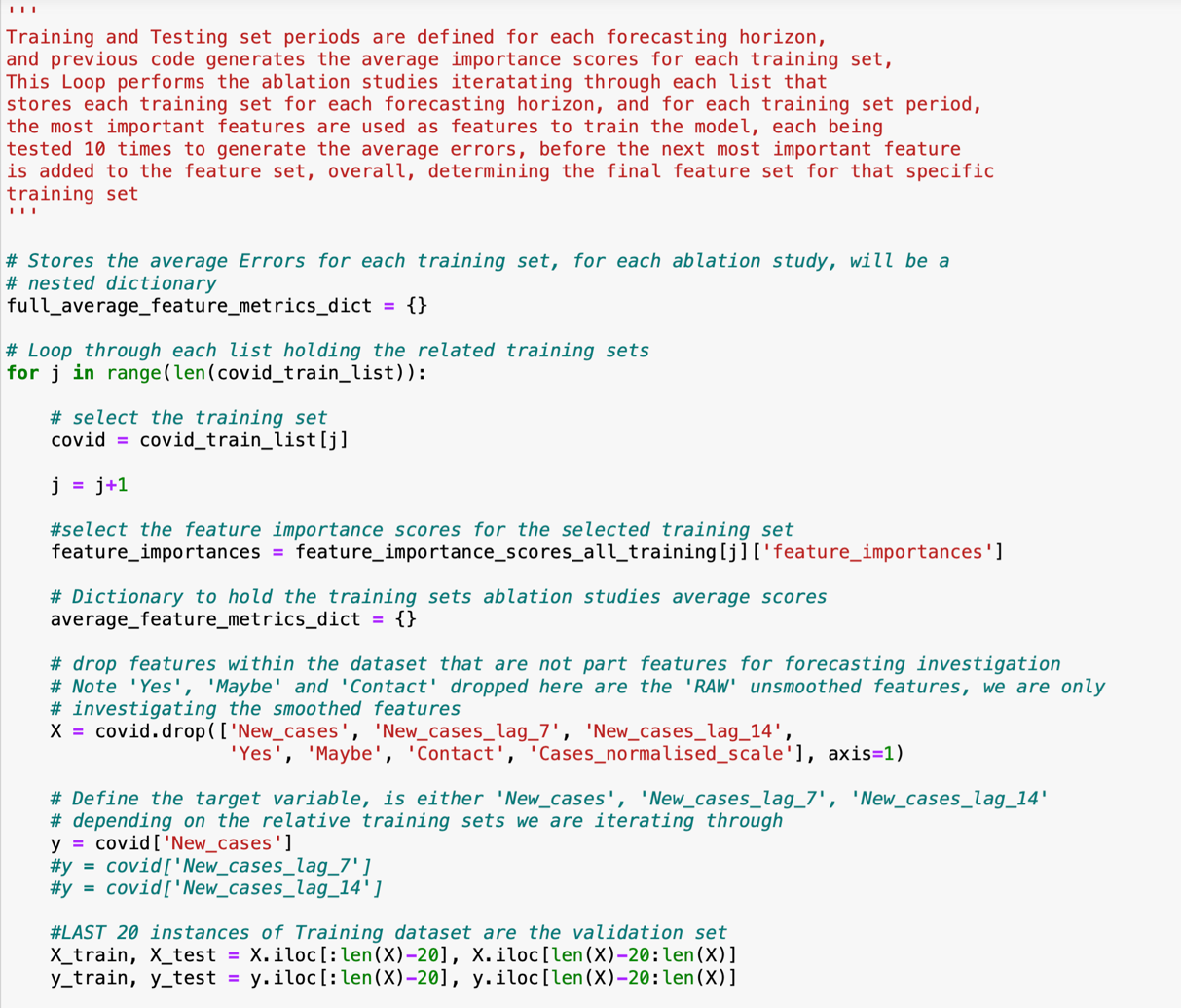
A screenshot of a computer program

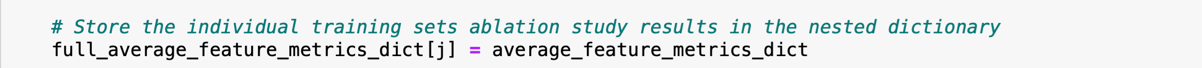
Description automatically generatedA screenshot of a computer program

Description automatically generated

3. Our ablation studies to identify the optimal feature set

A screenshot of a computer program

Description automatically generated



NB: as specified in the methodology – our reddit models must contain at least one reddit feature along with the baseline model to generate our relative comparison against a baseline model, therefore – if ablation studies identify that either no Reddit features of the ‘traditional’ feature does not improve RFR performance for that training period, we still either add the ‘Traditional’ feature, or the most important Reddit feature still.

4. Grid search – for each training test, whether ‘Reddit’ or ‘baseline’ training set

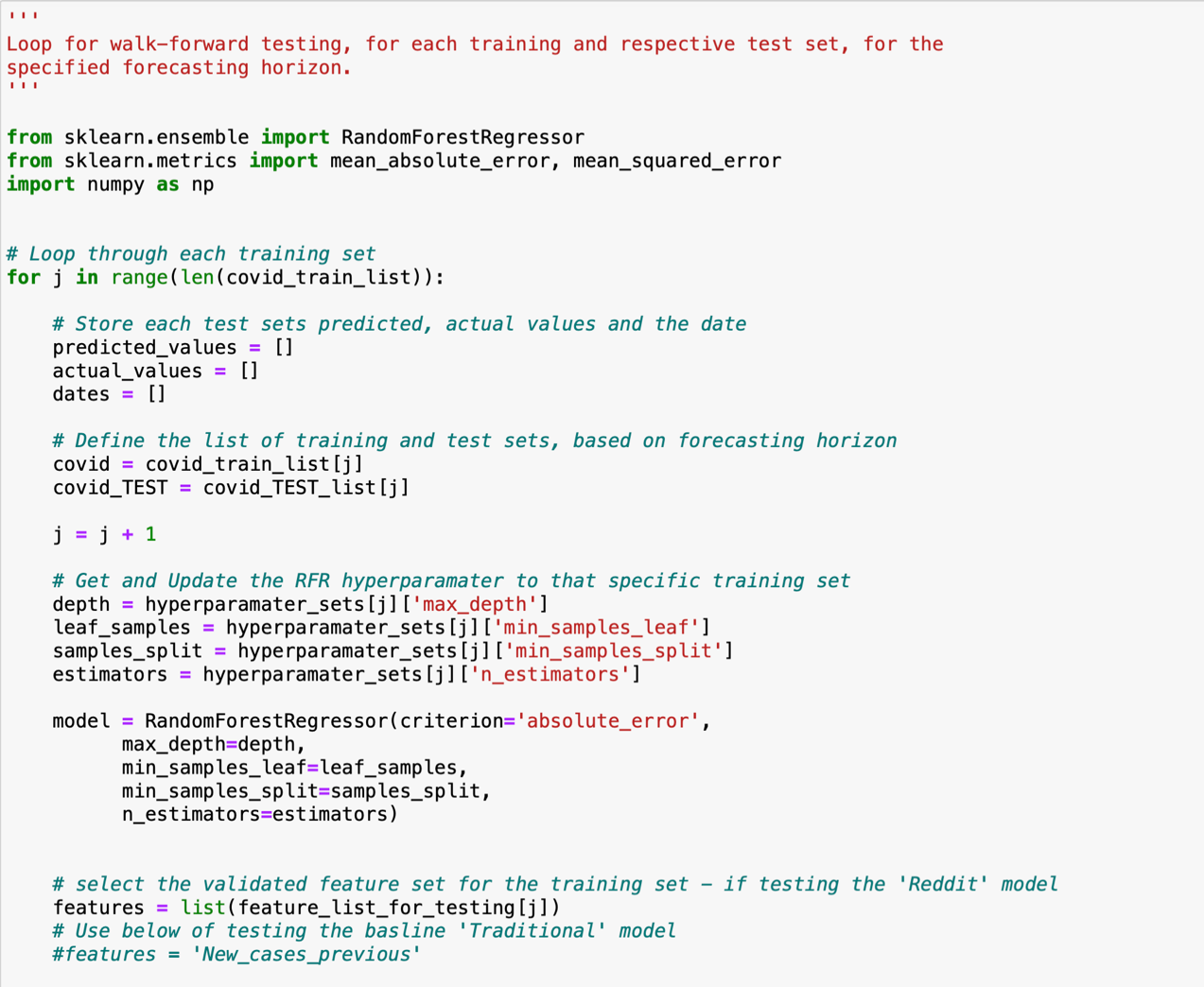
A screenshot of a computer program

Description automatically generated

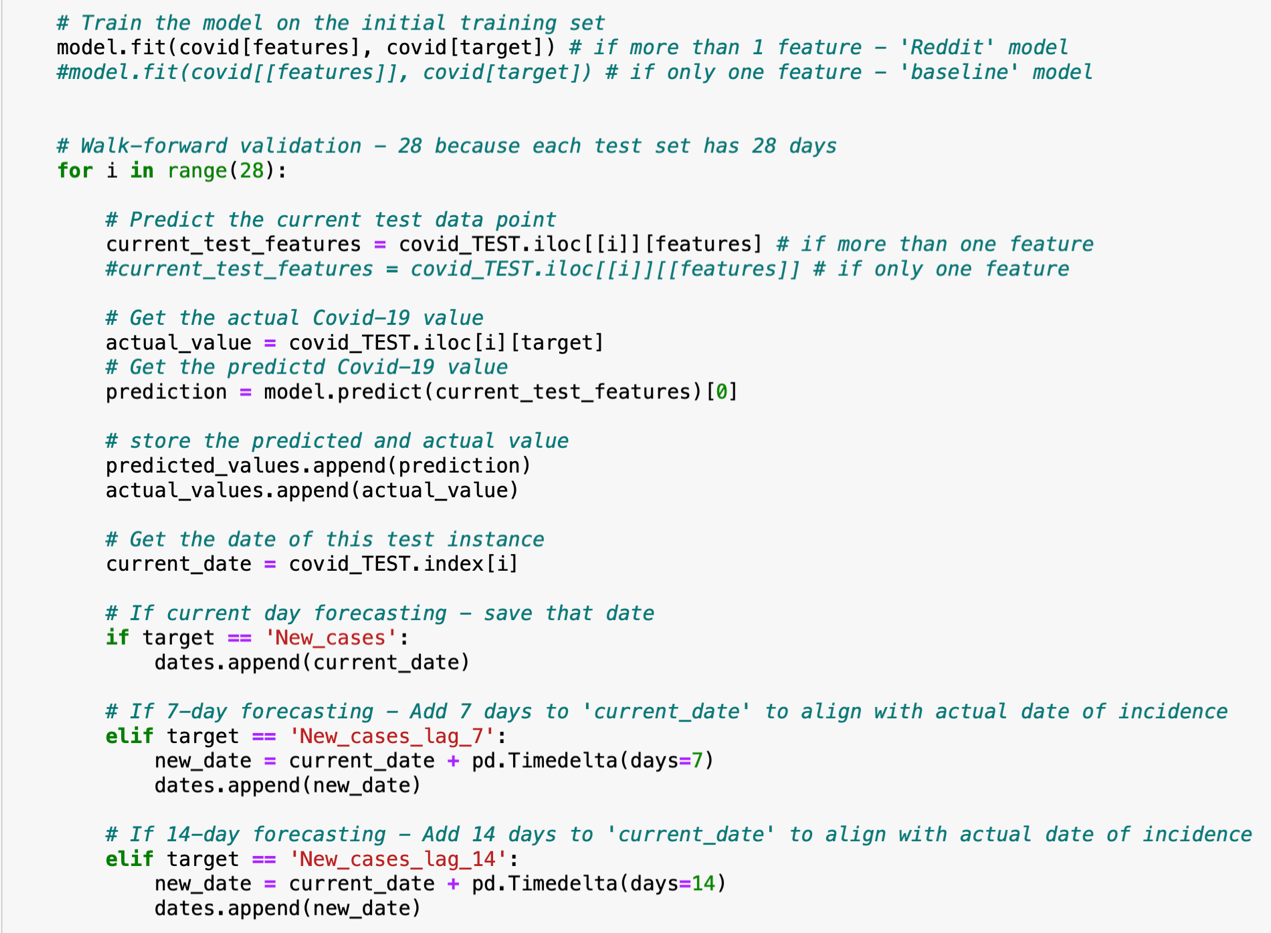
A screenshot of a computer program

Description automatically generated

5. Our walk-forward testing approach



A computer code with text

Description automatically generated with medium confidence

A screenshot of a computer program

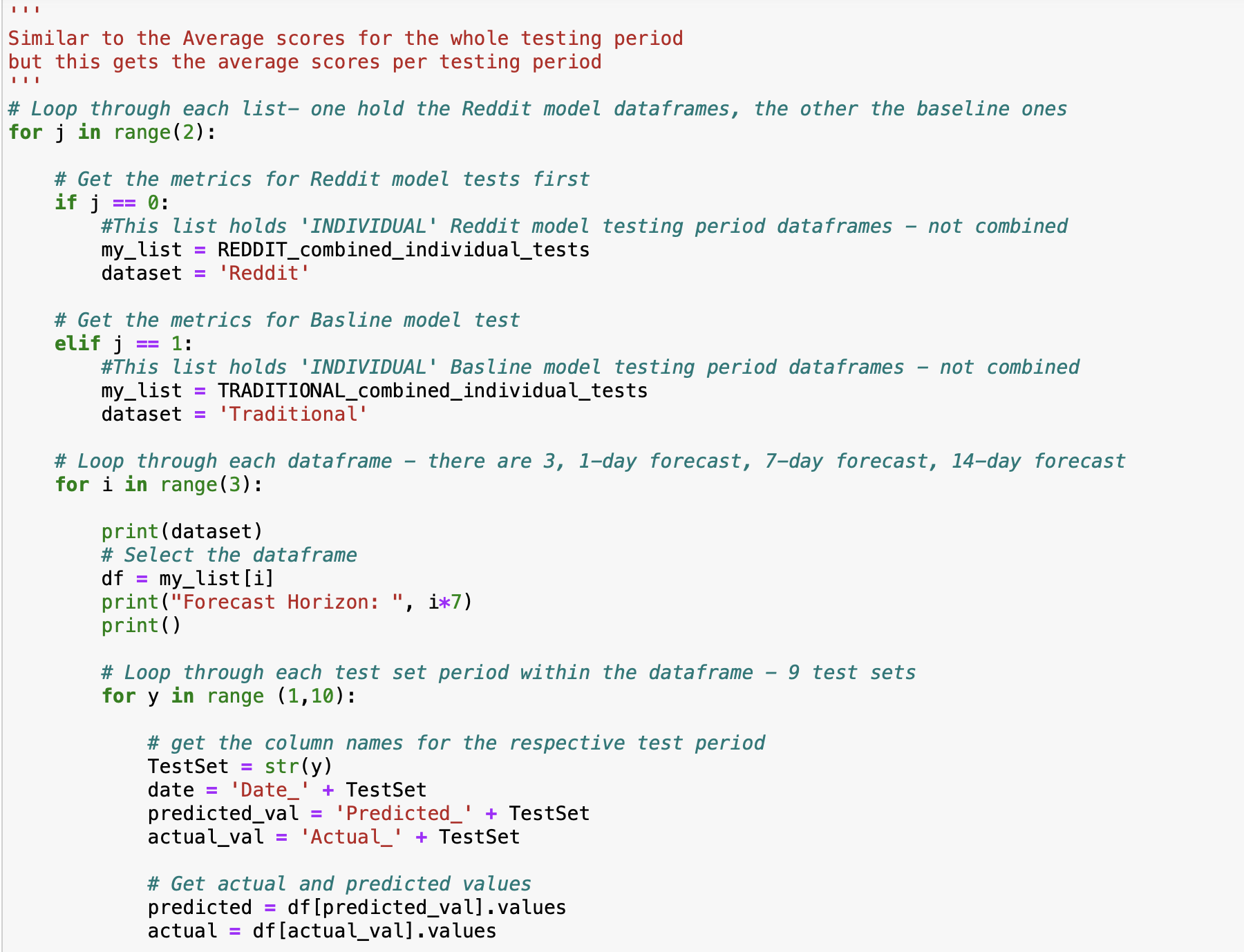
Description automatically generated

6. How we generate average performance metrics for entire testing period   
- before this – our dataframes are combined to make it suitable for this, so there is only one column for predicted and actual values covering the entire testing period

A screenshot of a computer program

Description automatically generated

7. How we generate average performance metrics per testing period



A screenshot of a computer program

Description automatically generated