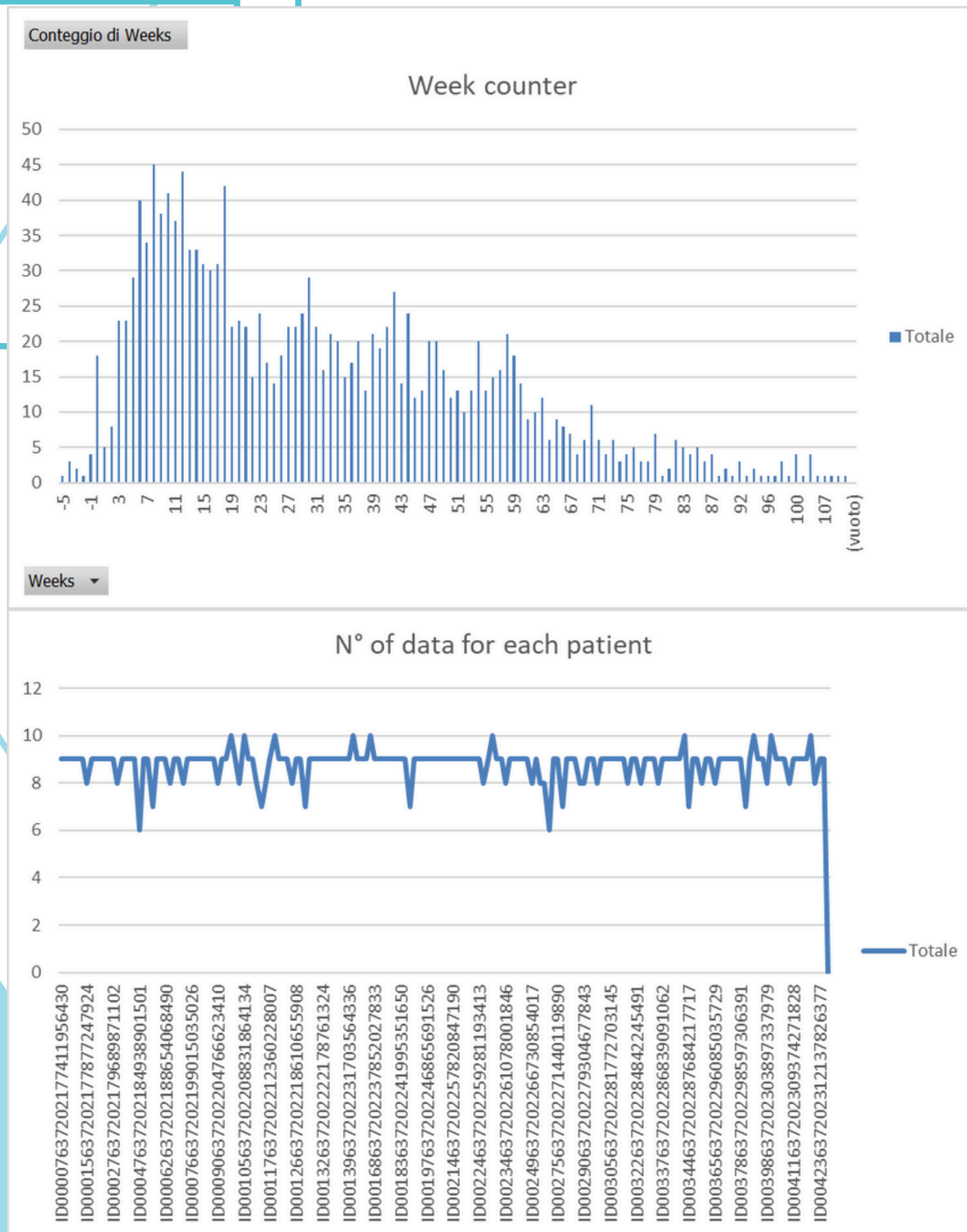
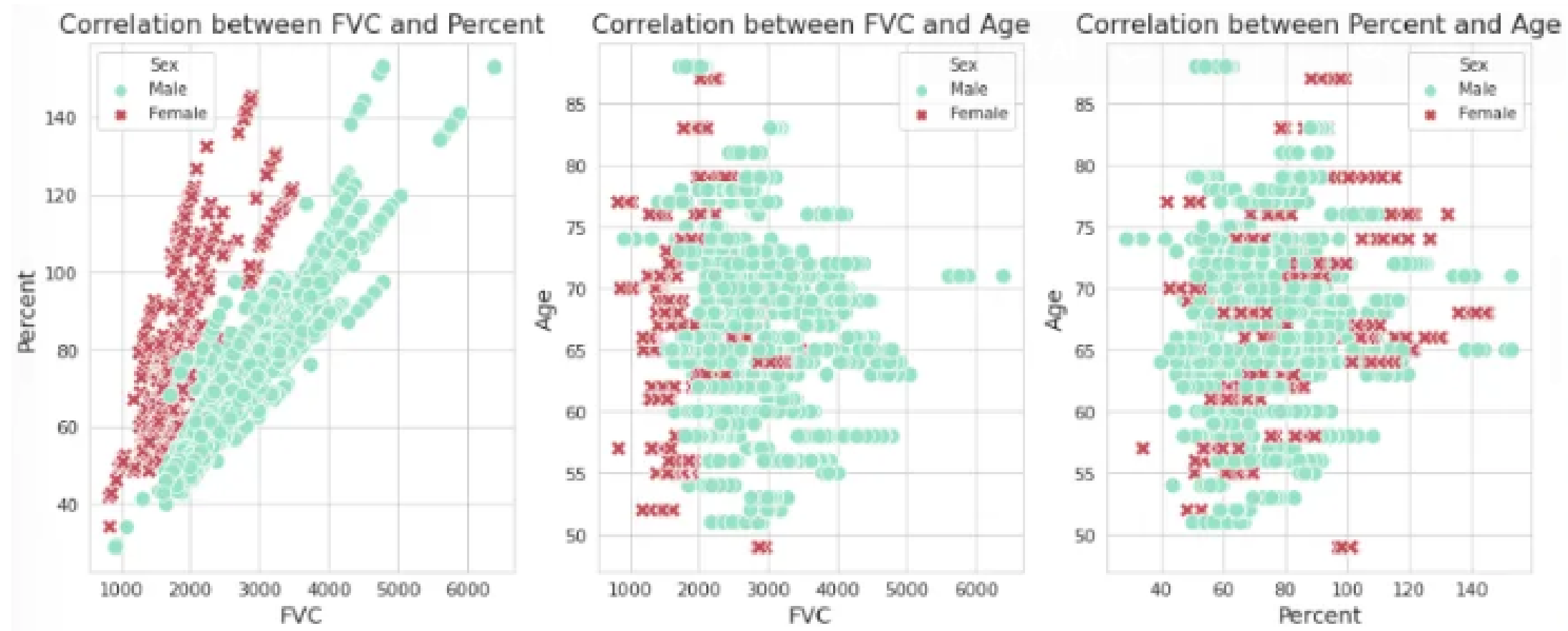


Data analysis



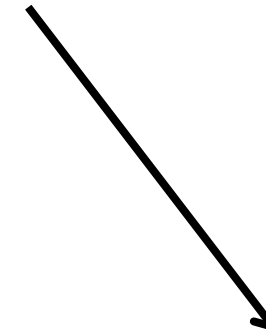
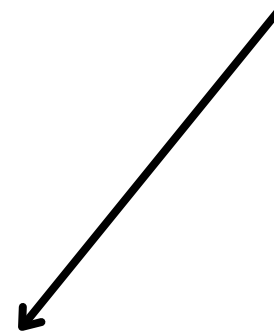
Data analysis



- **PEARSON CORR FVC X PERCENT: 0.672**
- **PEARSON CORR FVC X AGE: -0.09**
- **PEARSON CORR PERCENT X AGE: 0.096**

KAGGLE BEST SOLUTION

Model ensemble



**Linear Regression
CNN**

+

MLP



KAGGLE BEST SOLUTION



1st model

- **Linear Regression** → **Take as input (Weeks,FVC) and determine a slope “a”**
- **Slope “a”** → **Indicates the progression through the weeks**

The slope is then used as a target for the pair (CT scans, data)

- **CNN (EfficientNet)** → **Take as input CT scan and data**

1st model

- **K-fold cross-validation : 5 folds**

Quantiles pre-defined for each fold as [0.8, 0.5, 0.1, 0.1, 0.1]

For each fold:

- **20 CT slices are selected for each patient**

For each slice a slope is determined and the quantile is used to aggregate the per-slice predictions.

The FVC at each week is predicted using the formula:

$$\text{FVC} = \alpha * w + B$$

(B is the intercept calculated based on the known FVC at week 0.

2nd model

- **MLP**

Takes as input only the available data + metadata.

Doesn't take in consideration features extracted from the CT scan.

Model ensemble

The two predictions are combined giving more importance to the 2nd model (0.6).

- **1st model : Gains complex information and patterns from the CT scans**
- **2nd model: Stabilize the predictions**

Final Output :

PatientID - Week - FVC - Confidence