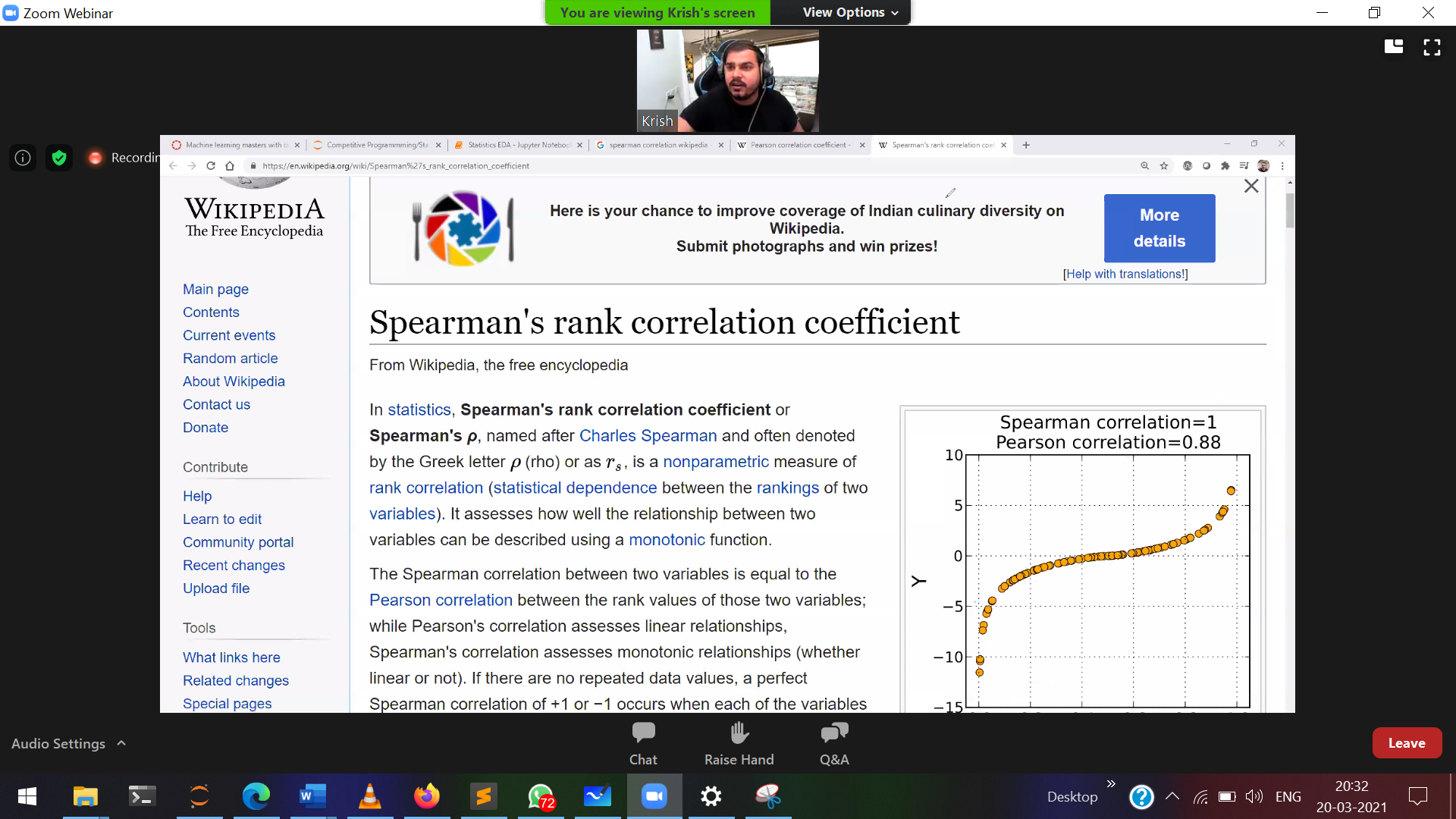
# Pearson Correlation Coefficient

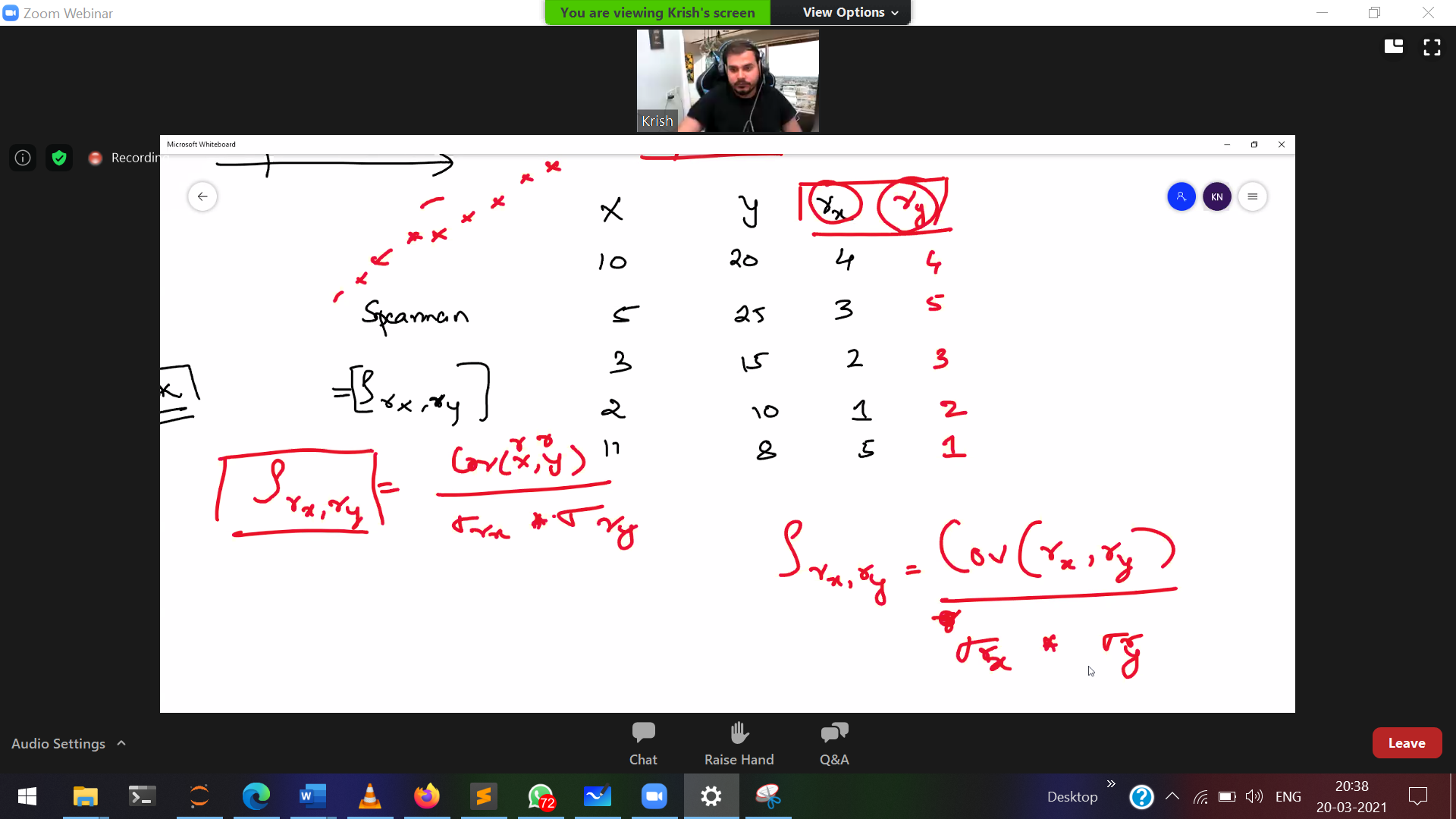
Pearson correlation works very well with linear data.

With 0 slope we cannot find any relationship between x and y.  
magnitude, direction   
non-linear property 🡪 Not captured, only linear property is being noted

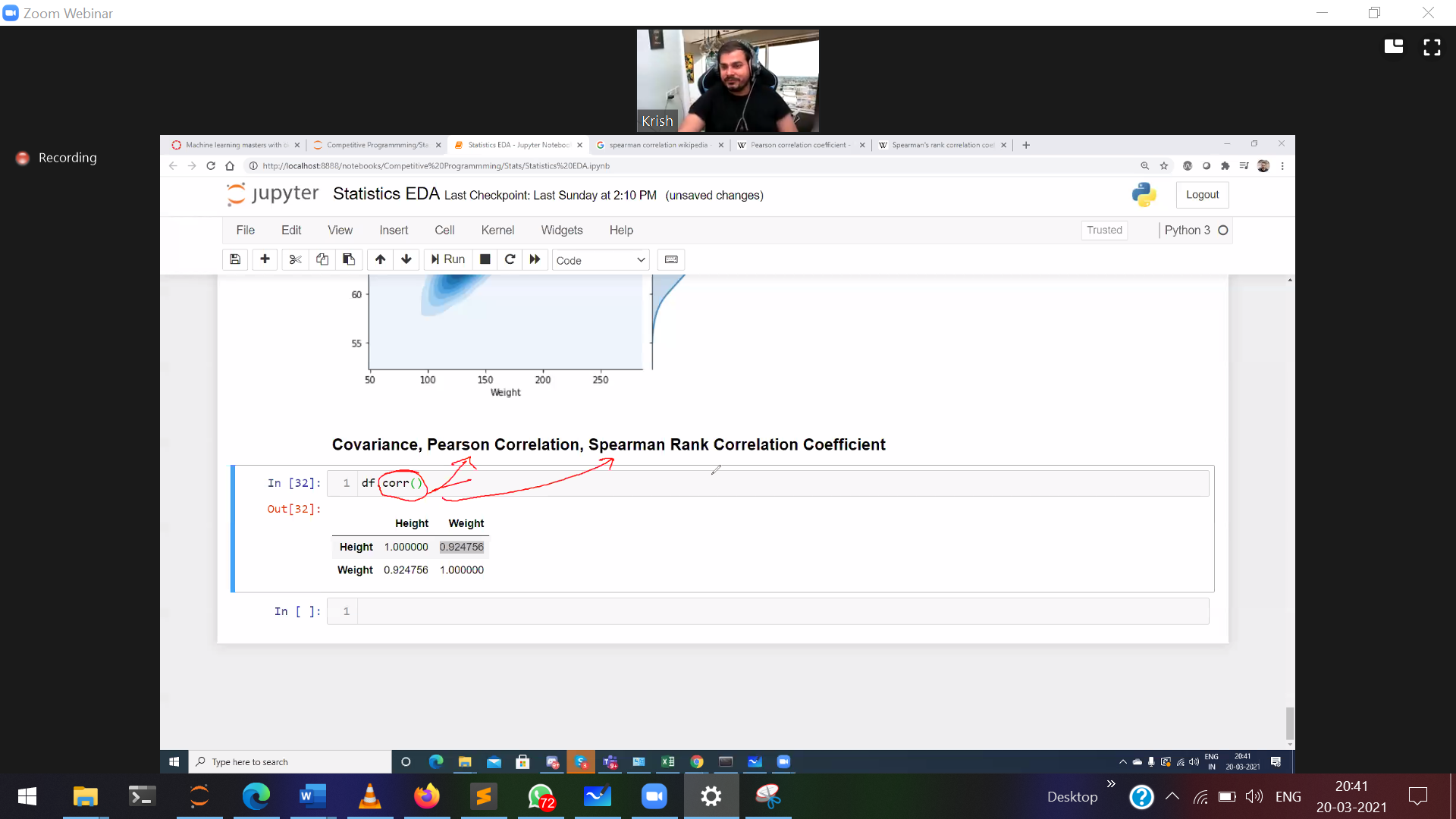
# Spearman’s rank correlation coefficient

Non linear property 🡪 Captured

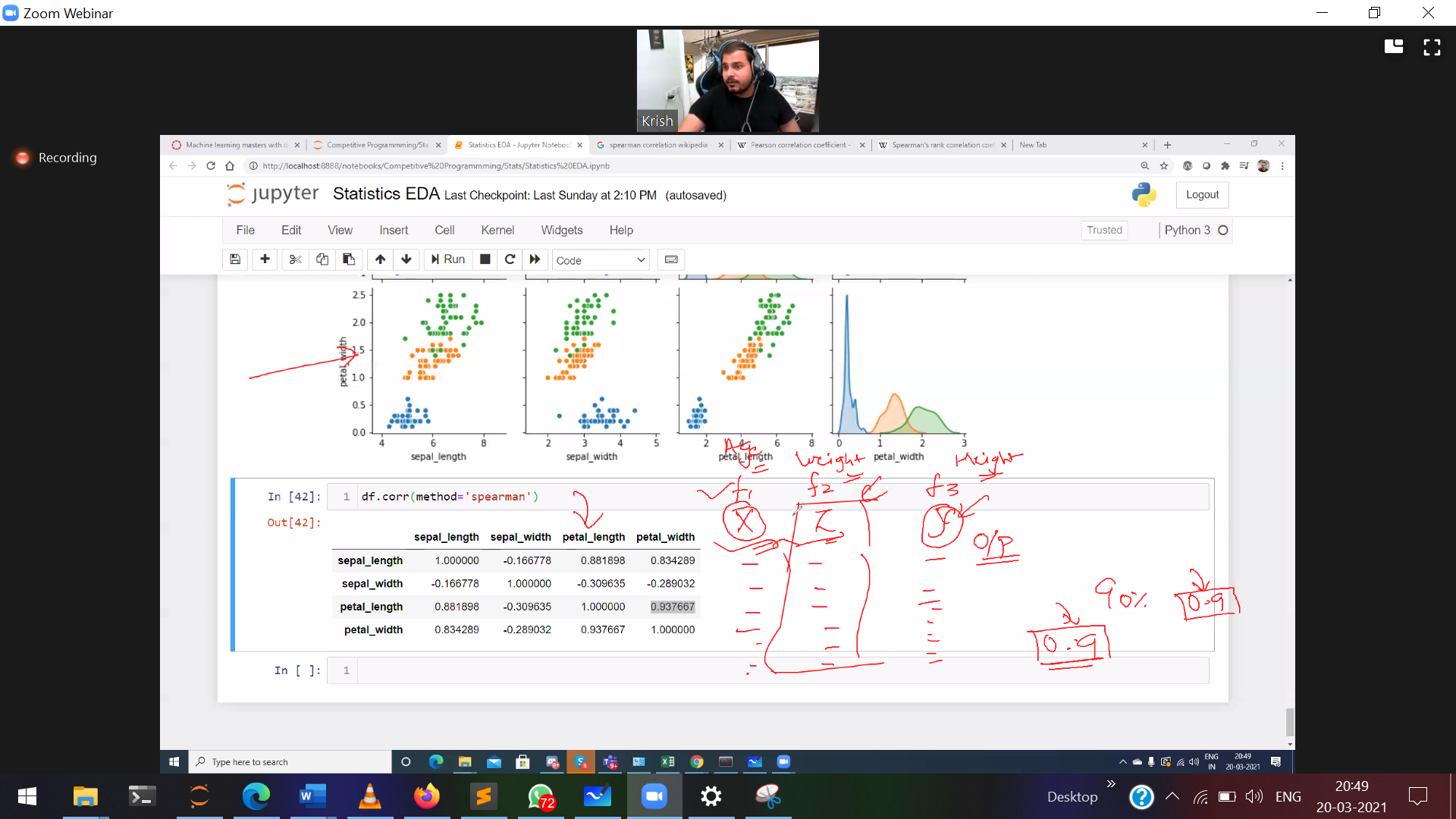




Assignment

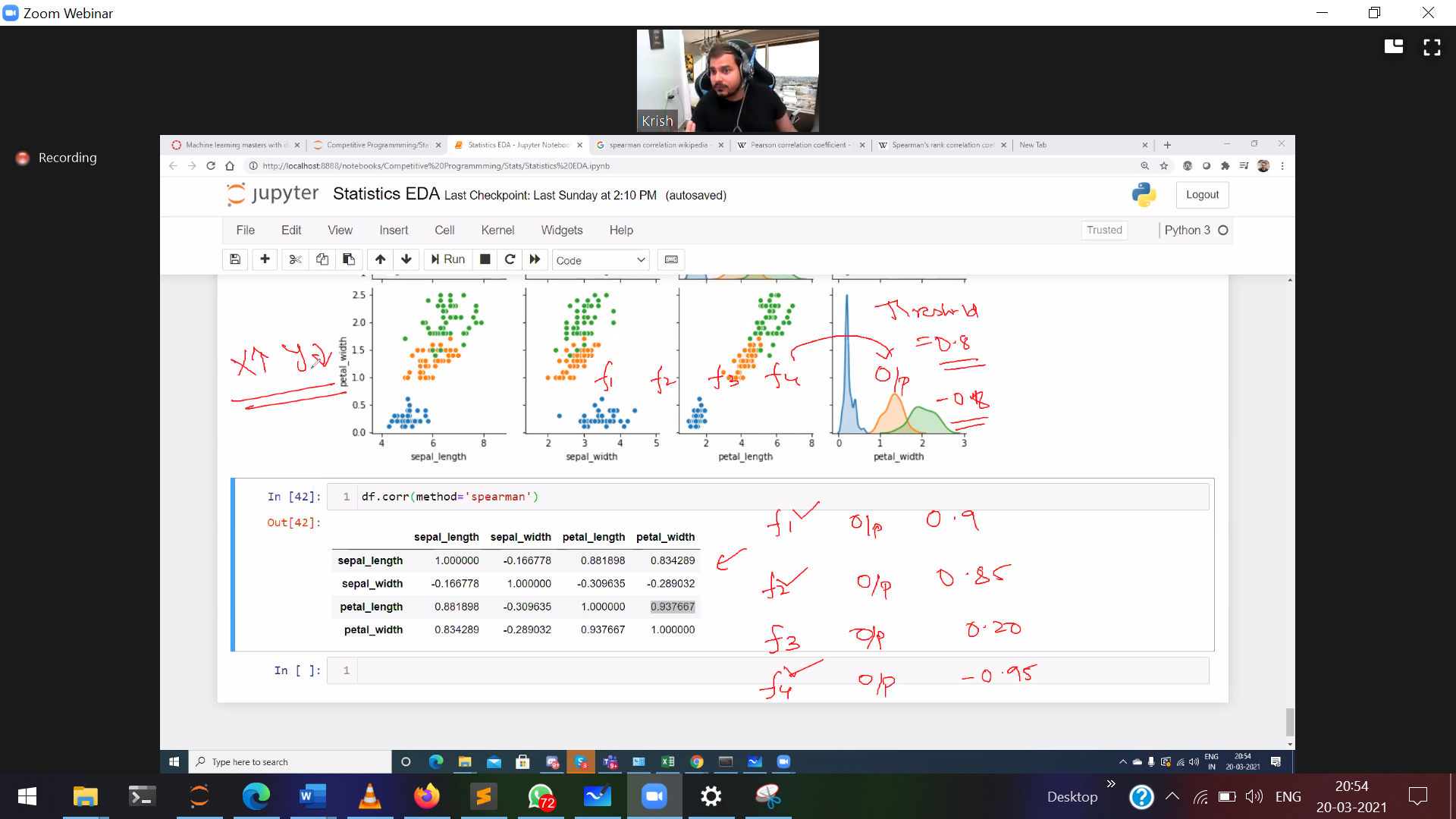


df.corr() 🡪 by default pearsong



Both the features (age and weight) are predicting output (height have almost same correlation (0.9)  
So drop 1 feature, to predict height

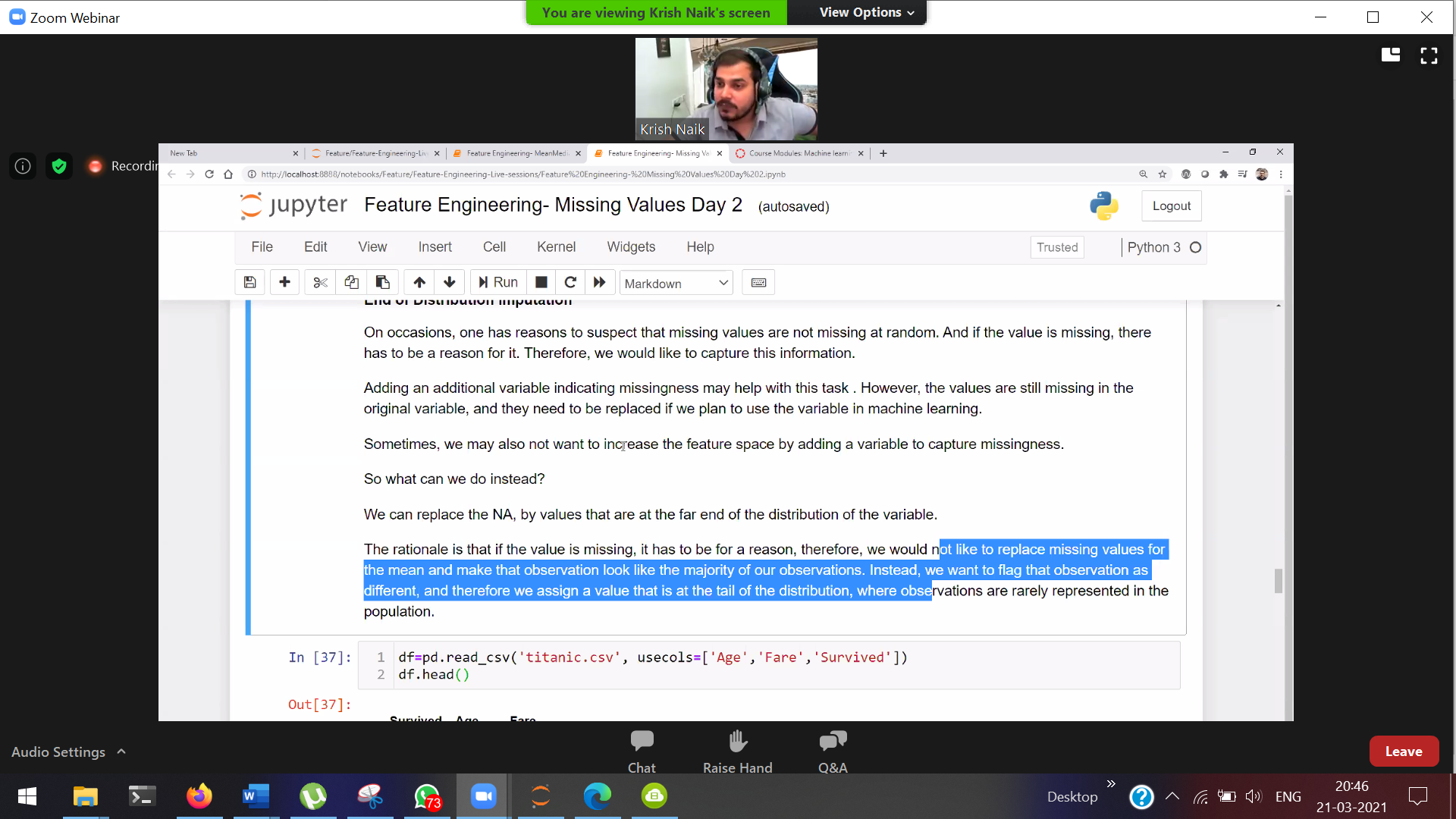
Checking with every independent feature with threshold



Don’t skip -ve correlation, its very very important.   
Set a threshold and sort it away.

Threshold value 🡪 decided only by the domain experty.

**Curse of Dimensionality**



Giving some extra importance for the NaN value, and showing those NaN value so unique by adding the end value.

Confusion Matrix (classification matrix)

Type-1 False positive  
Type-2 False negative