we will perform min‑max scaling and compare it to Z score

or Standard Scaling. Okay, here we are,

I am in the Mod3Demo2 notebook, and as before, the first thing I will do is I will connect to a compute free of charge in Google.

Now, as before, the first thing that we need to do is just the same as before, download the dataset, load it into the data dataframe, and that's it.

So, what we will do is just collapse the prep, and a nice way of doing this is just clicking here play,

Run anyway, because yes, trust me, it's me, and that will just run all of the cells all together, and there it is, it already ended, so it's so cool.

We can see we have our data, it's the same data as before, we can see it. If we describe it, it is the same data as before, so you can trust it, and we're up to the normalization, so let's collapse the prep.

And then normalization, again, we are going to do the four steps here.

So, step number one, we're going to import everything. Here, we're going to now import both the Standard Scalar and the min‑max scalar to compare them.

So let's run this. Step number two, instantiate the scalars here.

Again,

we are going to grab an instance of both classes.

Step number three, fit in. On both instances we're going

to run the fit method on the dataframe,

the same dataframe.

And, step number four,

we are transforming method on both scalars to get two datasets, one

for the Standard Scalar and one from the min‑max.

Finally, step "number five" we could say is converting them to their dataframe.

So now we can analyze the results.

Let me close this section and we're up to analyze the results.

If we describe the standard\_df,

we see the same as before, that means mean 0, standard deviation 1.

However, if we check the describe of the minmax\_df,

we see something quite different.

Both from the

uniform\_negative and the uniform\_positive,

we can see that the mean changed to 0.5 and the

standard deviation is around 0.28.

What does this mean?

It means that the min‑max normalizer moved the range,

changed the range of the uniforms from whatever they were to 0,1, and as

they were uniforms and it maintains the distribution,

it basically created the same distribution twice, which is

the uniform distribution from 0 to 1. Afterwards, in the

boxplot and the density plot,

we're going to confirm our theory, but that's basically what happened.

And also, the normal distribution appears to have a mean of 0.5.

Notice that the minimum value is 0 and the maximum value is

1, so let's see what happened there, okay?

Because the variants changed. Let's see if, for example,

the values are proportionally preserved.

So let's see the boxplot, and indeed,

that's the case. We can see the exact same boxplot for my uniform\_negative

and my uniform\_positive, confirming our theory. And we can see that the

normal distribution is still a normal distribution,

but it proportionally changed such that it fits in the interval 0,1. We

can see the same outliers on the top and on the end, which is basically

the extreme values out of two sigmas of the normal distribution, if you

know about that. If you don't,

that's okay, that's not something that you need to do, so do not worry.

So, what we will do is do some boxplots to create some density plots and do

the comparison. So let's go, and here we're going to run four plots before

for both normalizers, and after for each normalizer. And, it will take a

little time because it's doing four plots, and what we can see is

effectively what we were seeing. Here,

we have the before just in case you don't forget, we have the

uniform\_negative, which was negative, and it was uniform,

the uniform\_positive with random positive values and it was

uniform, and the normal distribution of course, 5.

What we can see is that the standard normalization does the same

as before, but the min‑max normalization collapsed both uniforms

into the same uniform from 0 to 1.

Now, here assuming that the shape may look the same, but the actual value is not

the same, so let's keep a notice at the y‑axis scale. Here,

the scale goes from 0 to 0.4, meaning that there's a wide distribution.

However,

here the density goes from 0 to 2.5, this means that actually

it's quite sharp. Also, we can see that the normal

distribution shifted center towards a 0.5,

but it is still a normal distribution, and that is actually quite important.

Finally, what we can see is that although these distributions appear, sorry,

these normalizations appear to be kind of the same, as we know

they are not because the min‑max, for example,

can change their range to 2.5 if you want.

But, one thing that they do have in common is that here we're

adding a super extreme value of 500,000,

we add it to our dataset,

we retransform, so we are using the fit\_transform

method to fit and transform new.

We do the same plots as before, and notice that what happens

basically is that both normal organizations basically collapse at

0, meaning, because there was a very, very,

very large value, meaning that they are completely sensitive to outliers, and

that is something that we will tackle on in the next module.

So, this concludes the demos of this module. A lot, right?

We will close this module by learning what comes next and some key takeaways and tips.