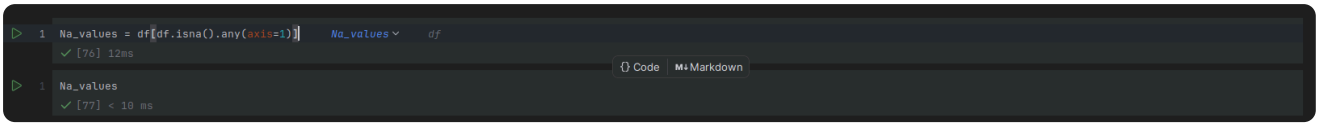
Yuki's Notes

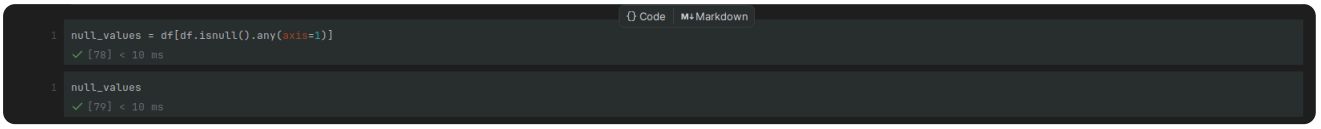
how to clean data

when we import a dataset using pandas the dataset may contain many missing values which may be addressed as NaN or null

so how do we find the Nan or null values in a large dataset ??

we can't possibly search the whole dataset . Right !?
worry no more kiddo we have the perfect thing for you to know the missing values such as NaN or null.
just use

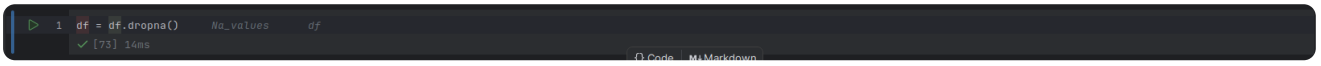




How to actually nuke the missing values ?

the drop function is really helpful in deleting or exterminating the missing values for the missing values you'd want to delete the rows. now ask why rows ? not columns ? because the column is your attribute if you drop the whole attribute we gonna have problems figuring important information about the datsets

so lets use the wizard magic that can automatically clean the missing values



how to actually fix the index after you made the missing values disappear

so you have successfully exterminated those missing values ! well doneee !
but the main issue is theres index like 288 → 290
which would actually mean 289 had missing values and it got exterminated. but the issue is if we don't rearrange the indexes it will remain the same so how do we actually rearrange them. its hard doing manually for each one of them so lets just use another function or wizard magic

we used

```
df.reset_index(drop = True , inplace = True)
```

now the index should be rearranged