**Project ETAV Summary**

**Extract**

To start, I went to Kaggle.com and chose the dataset called *Crossfit Athletes*. <https://www.kaggle.com/datasets/ulrikthygepedersen/crossfit-athletes>

**Transform**

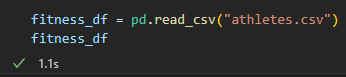
I created a folder to start my project with the csv file inside it, then imported the libraries I would be using, such as numpy, panda, and pyplot. I started the process of creating code cells, each cell performing a specific task.

* One for imports

Text

Description automatically generated

* One for reading the csv file



* And one for each Dataset to be created and observed from the initial finess\_df dataset. This allows for each category.

**Analyze**

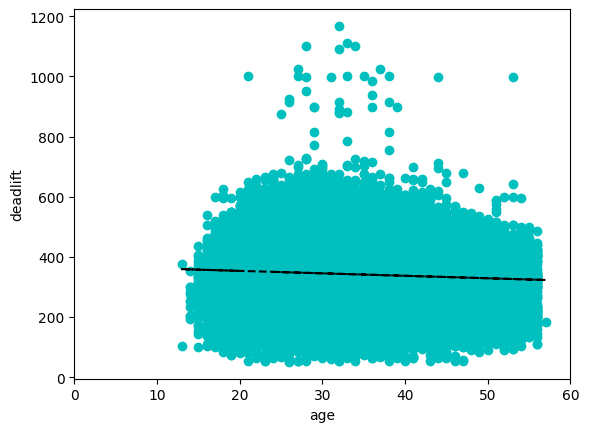
Some of the data in the csv seems to be entered incorrectly, however, with some research we should be able to get a more accurate view of the data.

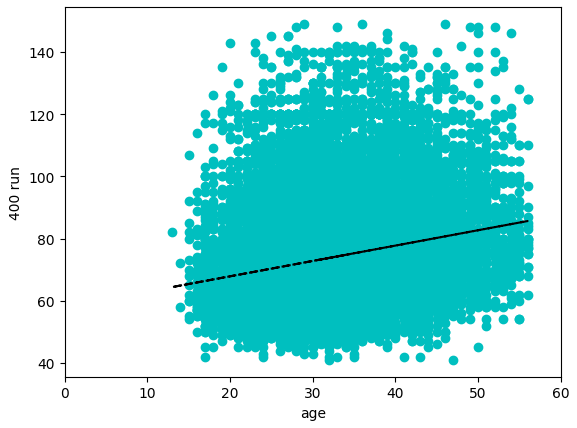
Some of the maximum limits were set by researching world records, since bad data is often outliers laying way beyond those records.

Something to note about the age comparisons:

* The charts show exactly what you might expect from intuition. Overall, as age goes up, deadlift performance, 400m performance, and 5k performance goes down.
* As age goes up, the overall weight of an individual goes up.

**Visualize**

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated