Common libraries;

* NumPy and Pandas simplify analyzing and manipulating data
* Matplotlib provides attractive data visualizations
* Scikit-learn offers simple and effective predictive data analysis
* TensorFlow and PyTorch supply machine learning and deep learning capabilities

In pandas, *Look carefully at the iloc[0:5] results, and compare them to the loc[0:5] results you obtained previously. Can you spot the difference?*

*The****loc****method returned rows with index*label*in the list of values from*0*to*5*- which includes*0*,*1*,*2*,*3*,*4*, and*5*(six rows). However, the****iloc****method returns the rows in the*positions*included in the range 0 to 5, and since integer ranges don't include the upper-bound value, this includes positions*0*,*1*,*2*,*3*, and*4*(five rows).*

***iloc****identifies data values in a DataFrame by*position*, which extends beyond rows to columns. So for example, you can use it to find the values for the columns in positions 1 and 2 in row 0, like this:*

**loc** is used to locate data items based on index values rather than positions.

Remember, real-world data will always have issues, but this is often a surmountable problem. Remember to:

* Check for missing values and badly recorded data
* Consider removal of obvious outliers
* Consider what real-world factors might affect your analysis and consider if your dataset size is large enough to handle this
* Check for biased raw data and consider your options to fix this, if found