# 

# Which is the best Operating System for Data Science?

As I continued studying Python, Pandas, Matplotlib, Jupyter Notebooks and so on, I came with a problem on Windows. Some of the libraries and software for data manipulation aren't available in "easy to install" tutorials for Windows OS, or aren't available at all.

Windows has always been my favorite OS, I like it's color mix, toolbar positioning and format, folder icons design and text fonts. But for the work I came across, concerning Data Science, it has brought some dificulties that I'm now willing to handle right now, because it impacts the learning curve on the technologies I want to be aware of.

So, the intent of this article is to bring some of the characteristics related to both Linux and Windows OSs and how perform on each one of them. So let's start!

**1. Speed**

The majority of the world's fastest supercomputers these days run on Linux, at about 90% of the total, compared to just 1% for Windows. One of the reasons is related to the excellent hardware support Linux systems have, for instance, to run Docker containers on NVIDIA Dockers (NVIDIA GPU) a Linux host is need, because there is [no support for Microsoft Windows](https://github.com/NVIDIA/nvidia-docker/wiki/Frequently-Asked-Questions#is-microsoft-windows-supported). Data scientists may come across with large amounts of data, so having an OS that supports the hardware needed to work with that data is an essential for the job.

**2. (Free) Software**

Linux is free and open-sourced, which accounts for a double win. Much of the IT community are open-source enthusiasts and a bunch of tools and applications are available for free according to each need. Data scientists can create and contribute to different projects and make them available for the community, thats why Linux has better support for tools such as [Jupyter Notebook](https://jupyter.org/install) and [IPython](https://ipython.readthedocs.io/en/stable/index.html).

**3. Flexibility**

Linux can be run in an imense amount of devices. For almost every device there is a version of Linux that you can use, such as Ubuntu for desktops and notebooks and Android for tablets and smartphones, which doesn't happens for Microsoft Windows. Raspberry Pi also has it's own special distro, called [Raspbian](https://www.raspbian.org/), with support for optimized instructions for the ARM architecture of the Raspberry Pi's hardware. And because Linux is a lightweight software that does not requires so many GB of RAM and hard drive storage to run the OS, it can be used in 'weaker' and older computers with low hardware capacity.

**4. Presentations and Work Sheets**

I'm a fan of Microsoft Office, specially Excel. Linux has [LibreOffice](https://pt-br.libreoffice.org/), but as far as I'm concerned Excel has many more features and is more optimized than LibreOffice's Calc. So for my case in particular, not using Excel would be hard. One easy solution is to use it in a Windows Virtual Machine (VM), so software compatibility won't be an issue. Both Calc and Excel can be used for data aquisition, storage and manipulation, so having access to them is good for data scientists.

**5. Job Demand and Linux Terminal**

Because of the easy of use for programming projects, Linux is in high demand for data scientists and software engineers. Appart from that Linux is faster and it's terminal helps to search for files, manipulate data and make data visualizations.

**Final Thougths**

Going through all these aspects of Microsoft Windows and Linux OSs it seems a little hard not to see which is the best option for Data Science. As I've mentioned before, Linux is not my everyday use OS, but its advantages may change my preferences. And it is not just a matter of taste anymore, Linux has been making my life easier when learning Data Science-related technologies. I'm not switching from Windows to Linux entirely, specially because I use Excel almost every day. For now an Ubuntu VM will suffice my needs.