

This project was something I did for my portfolio. I love Star Wars, and I can watch or read almost anything Star Wars related, so this seemed like a natural fit for me. It also allowed me to dig into things and answer a few questions I've had.

One of the first things I did was see what data I could find that was publicly available for me to use. I found some good data on starwars.fandom.com and downloaded the .csv file to use.

Next, I opened up a Jupyter notebook in VS Code and set up my project. I imported the libraries I expect to use (numpy, urllib, pandas, matplotlib, re), then read the data from the .csv file into Python to explore.

A quick look at the data allowed me to identify some parts I wouldn't be using, so I dropped those columns. I also split some columns so I could look into them independently. I cleaned up some of the problems I noticed, filling null values and obviously bad values with 'unknown'.

I then realized that the capitalization of this data is inconsistent, so I normalized that. I also took the time to rename the original column I split some of the data from earlier, since it now its name did not accurately represent what it contained.

More exploration showed that some other fields were not the proper data type, and some more that contained null values that I needed to replace, so I took care of those. I also separated out other columns and cleaned up some of that data.

After a few more iterations of looking at my data, cleaning, dropping unneeded columns, filling null values, and normalizing the capitalizations and spelling, I was ready to see what I could find.

The visualizations I decided I wanted to focus on were:

1. Lightsaber colors by count
2. Lightsaber colors by percentage
3. Gender breakdown
4. Years of date of death
5. Species

I laid out my grid in Python, then created my visuals.

The first visualization I created was for lightsaber counts. I created a bar chart of the number of lightsabers by color. The largest number of known sabers are blue, with green close behind. This was as I expected, but it was good to confirm.

Alongside that I created a pie chart that shows the same thing by percentage. This reinforces the fact that there are more blue and green among the known lightsaber colors, with blue near 26% and green at 20%.

Next I wanted to see the gender breakdown. It seems like there are far more males than females among the Jedi. It was surprising to see that there are about half as many female Jedi as there are males. I was unaware there were that many female Jedi.

Next to that I placed a stem chart showing the date of death, with a break line at year 0 BBY(Before the Battle of Yavin). I used the stem plot in this case so I could show each individual data point properly.

I then went on to focus on the species of the Jedi. I created a vertical bar chart showing the number of each species. I then modified the chart to add labels above the bars with the values.

These visualizations were interesting and enlightening. While I couldn't draw a lot of definitive solutions from the data, we did learn what the trends were. Some things we learned:

1. Most Jedi used blue or green sabers.
2. If not using blue or green, Jedi were likely to have multiple colors of sabers over time, some of which may have also been blue or green.
3. We found no data on yellow sabers (or other colors), which we know existed.
4. There are twice as many male Jedi as female.
5. Most Jedi die around 19 BBY. This is when Order 66 was implemented, so this makes sense.
6. There are more human Jedi than any other individual species.

7. Other species combined outnumber humans.

What the visualizations tell us is about the Jedi found in the data, not all Jedi ever. There may have been millions of Twi'lek Jedi in the distant past, completely outnumbering humans, for example. We also know there are Sith and other non-Jedi represented here, even though the data refers to them all of them as Jedi. It's incomplete data, but if we assume this is a representative dataset, then we at least can get some ideas about the force users and lightsaber wielders.