**Entry 4 - Get the Data**

In Entry 3, I defined my problem as: Holding all other factors constant, what mass is needed to retain an atmosphere on Mars?

**The Problem**

This sounds like a dataset I'm going to have to create myself.

*Hands on Machine Learning with Scikit-Learn & TensorFlow* recommends automating as much of the data acquisition process as possible, but this is a one-off dataset and the known parameters of planets doesn't change very often, so I'm not going to worry about that in this entry. If I were working on a project where I would connect to the data source again, like a twitter NLP project, I would most certainly want to automate pulling data. Sounds like a project for another dairy entry on a different mini-project.

If I were going to do any automation of this dataset, it would revolve around scraping table data from an HTML page.

**The Options**

The type of entities from which to draw the necessary data is rather limited. There are the planets, moons, and dwarf planets of this solar system and possibly exoplanets of other systems.

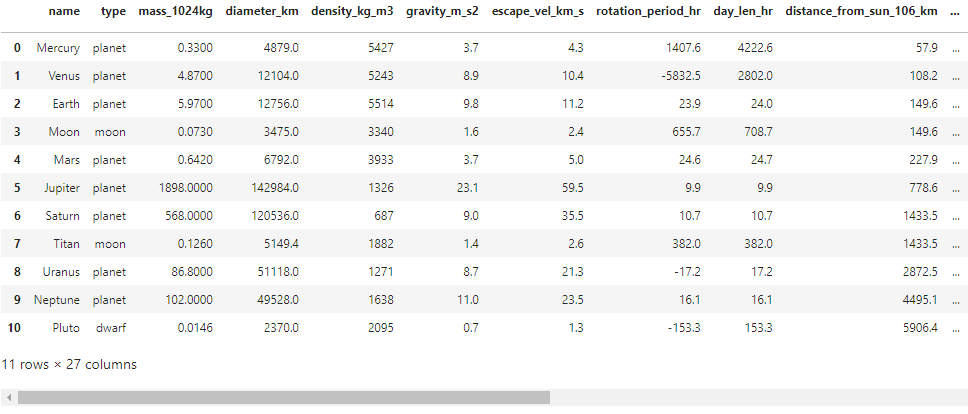
**The Proposed Solution**

Fortunately, I didn't have to comb through information on each and every planetary body individually. The ‘planetary fact sheet’ has many of the features I need. This included 8 planets, 1 moon, and 1 dwarf planet. Starting with this as a base, I gathered 27 features on 11 planetary bodies.

I considered including more moons (Jupiter has 79, Saturn 82, Uranus 27, and Neptune 14), but couldn't find sufficient information on the necessary features. Most importantly was a lack of information on atmospheric mass. The same was true for exoplanets. They're just too far away to have good measurements.

The planetary bodies are: 'Mercury', 'Venus', 'Earth', 'Moon', 'Mars', 'Jupiter', 'Saturn', 'Titan', 'Uranus', 'Neptune', 'Pluto'

The features are: 'name', 'type', 'mass\_1024kg', 'diameter\_km', 'density\_kg\_m3', 'gravity\_m\_s2', 'escape\_vel\_km\_s', 'rotation\_period\_hr', 'day\_len\_hr', 'distance\_from\_sun\_106\_km', 'perihelion\_106\xa0km', 'aphelion\_106\xa0km', 'orbital\_period\_days', 'orbital\_velocity\_km\_s', 'orbital\_inclination\_degrees', 'orbital\_eccentricity', 'obliquity\_to\_orbit\_degrees', 'mean\_temp\_c', 'surface\_pressure\_bars', 'nbr\_moons', 'rings', 'magnetic\_field', 'equatorial\_radius\_km', 'mean\_radius\_km', 'V(1,0) (mag)', 'geometric\_albedo', 'atmospheric\_mass\_kg'



**The Failure**

This step was virtually error free. Other than spending hours looking at a bunch of different sources and trying to find more information on moons and exoplanets, it went pretty smoothly.

The limited number of examples (only 11) may be problematic in future steps. For example, measuring how well the model performs could prove extremely challenging.

**Next Up**

Explore the data

**Sources:**

The planetary body data was retrieved from a variety of sources. There were also a couple of one-off searches for some of the more obscure information, like the V(1,0)(mag) for Titan. These were the major contributors:

* <https://spacemath.gsfc.nasa.gov/Grade67/10Page7.pdf>
* <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/planetary-atmosphere>
* <https://ssd.jpl.nasa.gov/?sat_phys_par>
* <https://ssd.jpl.nasa.gov/?planet_phys_par>
* <https://www.windows2universe.org/our_solar_system/planets_table.html>
* <https://www.windows2universe.org/our_solar_system/moons_table.html>
* <https://nssdc.gsfc.nasa.gov/planetary/factsheet/>
* [https://solarsystem.nasa.gov/moons/saturn-moons/titan/by-the-numbers/![image.png](attachment:image.png](https://solarsystem.nasa.gov/moons/saturn-moons/titan/by-the-numbers/!%5Bimage.png%5D(attachment:image.png))