

II. Data Visualization

1. Intro

The initial dataset from the NASA had 422 meteorite classes in it. A high number that did not necessarily reflect the distribution of meteorites. Indeed, as shown in the Phase-1 (Data Wrangling), the top ten meteorite classes encompassed approximately 80% of the total number of meteorites. For that reason, I decided to neglect – for now, the remaining 20% of the data – for a more concrete analysis.

2. Visualization

2.1. Scatter Plot

After extracting the needed segments, I used the Matplotlib Library to generate a Scatter Plot (meteorite classes on the x-axis and their mass on the y-axis) to see the distribution of the mass per class. Surprisingly, with the notable exception of a few outliers, the distribution was quite homogenous among the ten classes, with the majority of the meteorites having a mass range 0-500kg. Here are some conclusions:

- The heaviest meteorite belongs to the H5 class
- The class L6 is more homogenous
- The class CM2 has the lightest meteorites

2.2. Seaborn Implot

I used the Seaborn Library to visualize a sort of a time-series plot showing the distribution of the landings throughout the entire timescale of the dataset. Not surprisingly, most of the landings occurred after the 19th century, which can be explained by the restricted technological advances prior to that time, where only landings that were actually observed by humans were recorded. The third plot shows the mass of the meteorites that landed throughout the timescale. Here are some conclusions:

- Accurate landing records really started in the 19th century
- Recent meteorites seem to be heavier