**Data Science Project on President Heights**

Report

If you are a beginner in Data Science you must solve this project, as you will learn a lot about working on Data,that comes from a csv file or any other formats.

This data is available in the file heights.csv uploaded in git hub, which is a simple comma-separated list of labels and values.

We’ll use the Pandas package to read the file and extract this information (note that the heights are measured in centimeters):

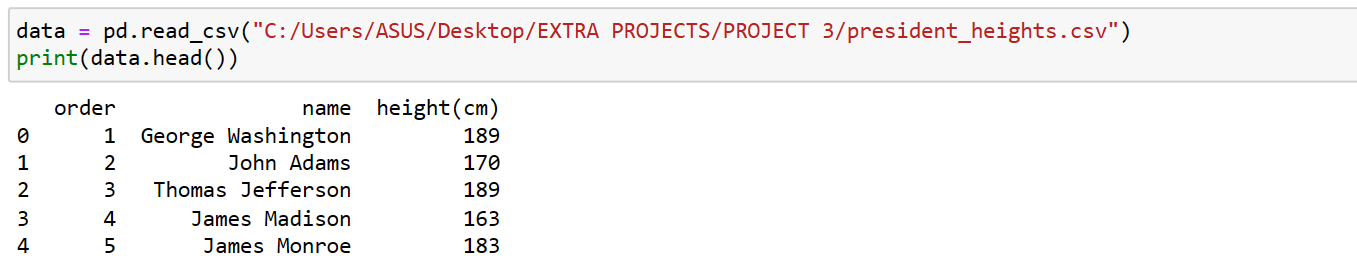


Fig 1

Now that we have this data array, we can compute a variety of summary statistics:

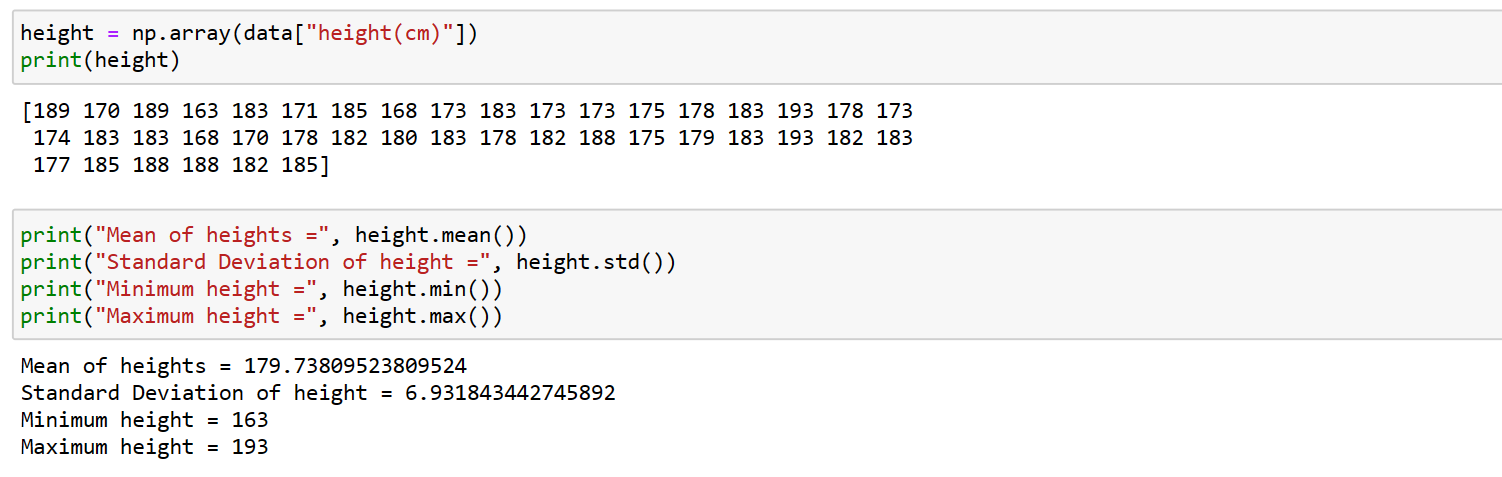


Fig 2

Note that in each case, the aggregation operation reduced the entire array to a single summarizing value, which gives us information about the distribution of values. We may also wish to compute quantiles:

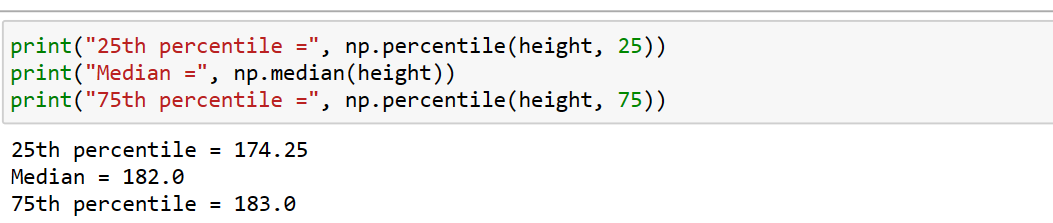


Fig 3

We see that the median height of US presidents is 182 cm, or just shy of six feet. Of course, sometimes it’s more useful to see a visual representation of this data, which we can accomplish using tools in Matplotlib:

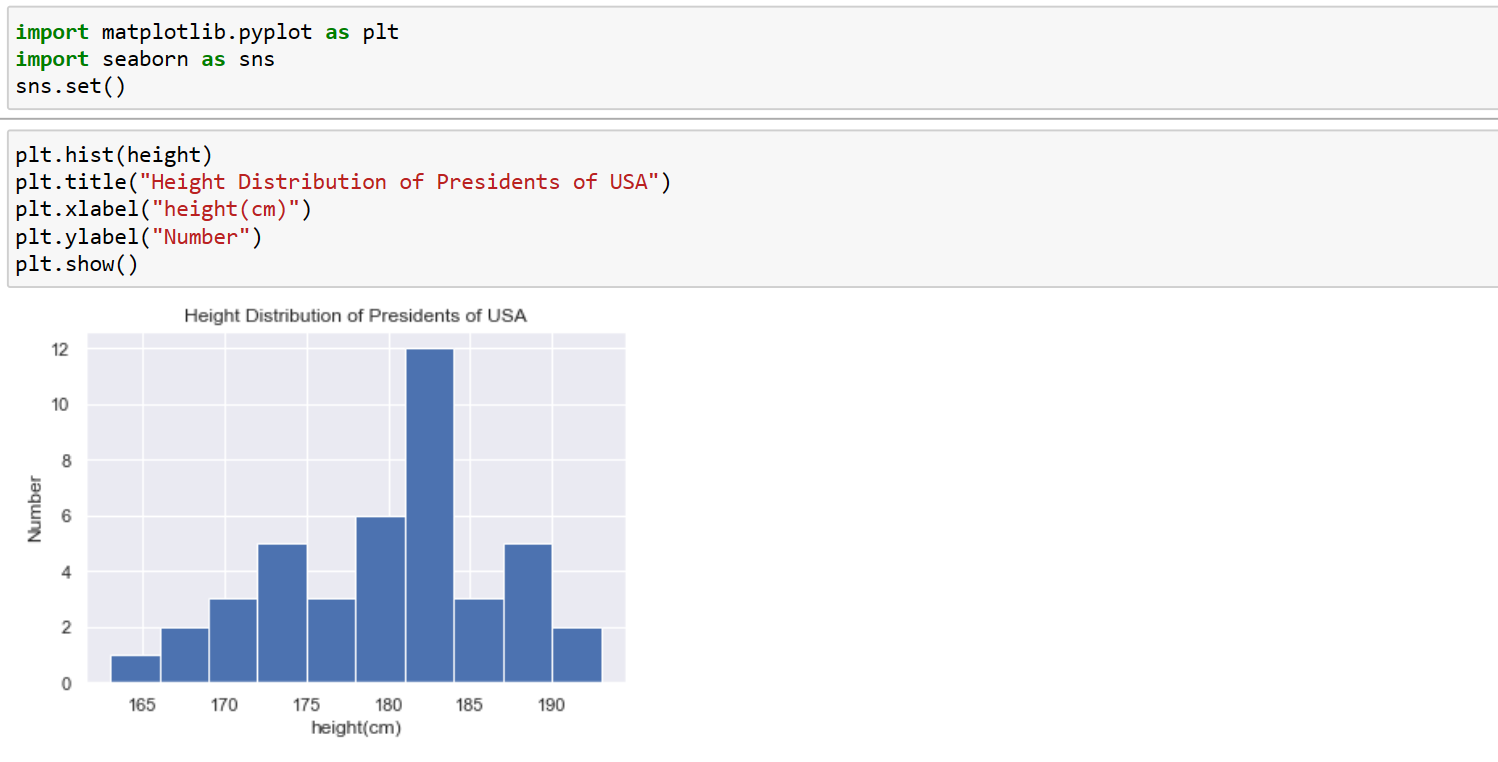


Fig 4

These aggregates are some of the fundamental pieces of exploratory data science that we’ll explore in more depth in later coming projects.